

中日機械翻訳における離合詞の処理手法

范 莉 馨[†] 任 福 繼^{††}
宮 永 喜 一[†] 枡 内 香 次[†]

中国語の単語には、二文字以上からなるものが多い。そのうち語素間の結びつきが弱く、その間に他の部分(挿入成分と呼ぶ)を挿入できる単語がある。このように語素を連結してもよく、離してもよい使い方ができる単語を離合詞とよぶ。自然言語理解、特に機械翻訳において、上述の離合詞を正しく認識できなければ、中国語文の形態素解析、さらに構文解析に対しては大きな障害になると考えられる。本論文では、教科書など大量の実例文から離合詞および関連情報を抽出し、この情報を分析し、離合詞の構造上の特徴および挿入成分を検討してから、中日機械翻訳における離合詞の処理手法を提案する。さらに離合詞を含む300文を用いて翻訳実験を行った。その結果、離合詞処理の正解率が約93%であり、本手法の有効性が確かめられた。

A Processing Method of Chinese “LIHECI” in Chinese-Japanese Machine Translation

FAN LIXIN,[†] REN FUJI,^{††} YOSHIKAZU MIYANAGA[†] and KOJI TOCHINAI[†]

There are many Chinese words which consist of two or more Chinese characters. Among them, there are words in which other constituents can be inserted because the connection between the two characters is weak. If two characters which make up a word, can form words individually as well as in combination, we call these words “LIHECI”. If these “LIHECI” can’t be understood correctly, it will be a big obstacle to the morphological analysis and sentence analysis of Chinese sentences in natural language processing and machine translation. In this paper, a method for processing “LIHECI” in Chinese-Japanese machine translation is presented, and some characteristics of “LIHECI” are also considered.

1. ま え が き

最近、自然言語処理における柔軟性という概念が、東洋の哲学、文化などの影響を受けて提案されている¹⁾。すなわち、言語には、系統または規則的な面と無系統または無規則的な面の両面があり、また離散的、連続的という両面もあると考えられる。

しかし、言語処理の道具であるコンピュータは0と1の世界で動作し、基本的に固定的な規則しか受け付けない。それで“基本的”、“大体”などの語が持つ意味的「柔軟性」は直接にコンピュータで処理されない場合が極めて多く、この問題を解決することが言語処理の重要な課題であると考えられている。

ここでは、上記の問題に属する一例として、我々の開発している中日機械翻訳において、中国語の「詞

(ci)/単語」,特に「離合詞(liheci)」を考える。中国語の単語には、二文字以上からなるものが多い。そのうち語素間の結びつきが弱く、その間に他の成分(挿入成分と呼ぶ)を挿入できる単語がある。このように語素を連結してもよく、離してもよい使い方ができる単語を離合詞とよぶ⁴⁾。自然言語理解、特に機械翻訳では、上述の離合詞を正しく認識できなければ、中国語文の形態素解析、さらに構文解析に対しては大きな障害になると考えられる^{5),6)}。

本論文では、教科書など大量の実例文から離合詞および関連情報を抽出し、この情報を分析し、離合詞の構造上の特徴および挿入成分を検討してから、中日機械翻訳における離合詞の処理手法を提案する。さらに、離合詞を含む300文を用いて翻訳実験を行った。その結果、離合詞処理の正解率が約93%であり、本手法の有効性が確かめられた。

以下、2章では中日機械翻訳システムにおける離合詞、3章では離合詞の分類、4章では挿入成分、5章では中日機械翻訳における離合詞の処理手法、6章で

[†] 北海道大学工学部

Faculty of Engineering, Hokkaido University

^{††} 広島市立大学情報科学部

Faculty of Information Sciences, Hiroshima City University

は実験および考察について述べる。

2. 中日機械翻訳システムにおける離合詞

著者らは先に中日機械翻訳システムを構築するための中国語品詞の定義および分類を行った^{2),3)}。その後、さらに広範囲にわたる多数の文の検討により、表1に示した特徴をもつ中国語の「詞」が多数出現することが見いだされた。特にこれらの語を現在のシステムで処理すると、表2に示したような誤結果が出やすい、あるいは単語として認識できない場合もある。

表1, 2から次の2つの点に分かる。

a. 動詞の「投 tou/投げる」と名詞の「稿 gao/原稿」を連結して用いる場合は1つの単語として「投稿する」という1つの意味を1つの文にして表しており、単に語素の意味を2つ合わせて表している「原稿を投げる」意味ではない。

b. 離して用いる場合はその間に挿入した成分によって意味も違う。この場合、その間に挿入した成分の

表1 離合詞「投稿」の例
Table 1 Examples for liheci [tougao].

No.	中国語離合詞例	日本語訳
01	投 稿	投稿する
02	投 了 稿	投稿した
03	投 着 稿	投稿している
04	投 過 稿	投稿したことがある
05	投 完 稿	投稿し終えた
06	投 上 稿	採用された
07	投 開 稿	投稿し始める
08	投 起 稿	投稿し始める
09	投 得 了 稿	投稿できる
10	投 不 了 稿 来	投稿できない
11	投 半 天 稿	随分長い間投稿する
12	投 下 几 稿	ちょっと投稿する
13	投 次 稿	ひとつ投稿する
14	投 回 稿	ひとつ投稿する
15	投 点 几 稿	すこし投稿する
16	投 不 少 稿	たくさん投稿する
17	投 的 稿	投稿したのだ

表2 誤り訳語の例
Table 2 Examples of mistaken translation.

No.	中国語	誤日本語訳
01	投 了 稿	原稿を投げた
02	投 着 稿	原稿を投げている
03	投 過 稿	原稿を投げた事がある
⋮	⋮	⋮
17	投 的 稿	投げた原稿

意味を加えて表すため連語の一種として扱われる。

自然言語理解、特に機械翻訳において、上述の係り受け関係を正しく認識できなければ、中国語文の形態素解析、さらに構文解析に対しては大きな障害になる。

勿論、中国語では、すべての単語は離合詞であるわけではない。例えば「労働 laodong/働く」が離合詞ではない。それで、文字「劳 lao」と「働 dong」の間に他の成分を挿入することができず、普通の単語（離合詞以外の単語を意味する）と同じように、単語の後ろに他の付加成分を付加すべきである。すなわち、過去を表すと、「劳了働 laoledong」ではなく、「労働了 laodongle」である。進行状態を表すと、「劳着働 laozhedong」ではなく、「労働着 laodongzhe」である。この場合、単語「労働」の原形は分離できないので、処理するのが容易であると考えられる。

それで、上述のような離合形態の単語群を全部辞書に登録すれば問題を解決できると考えられるが、実際には不可能である。このような係り受け関係を含む機械翻訳は日本でも中国でも十分に研究されていないので、今後の重要な研究課題であると考えられる。

我々は、このような特徴をもつ単語に対しては既に通常単語の範囲を越えて、新しい定義を賦与する必要があると考えられる。本論文では上例のような語素間の結びつきが弱く、その間に他の成分を挿入できる単語、すなわち語素を連結してもよく、離してもよい使い方ができる単語を離合詞という。例えば下例中の「畢業 biye/卒業する」と「就業 jiuye/就職する」は離合詞である。

例文(1): 他成績不好, 能畢業?

意味: 彼は成績がよくないのに, 卒業できるか?

例文(2): 他成績不好, 畢不了業.

意味: 彼は成績がよくないので, 卒業できない.

例文(3): 畢業以後就業有問題?

意味: 卒業してから就職するのに問題があるか?

例文(4): 恐怕馬上就不了業.

意味: たぶんすぐには就職できないだろう.

3. 離合詞の分類

本システムでは、離合詞に含まれる2語素の属性は「岩波中国語辞書」と「新華字典」を参考して決める。また、分類作業を便利にするため、次のいくつかの分類ルールを設けている。

(1) 多義性をもつ語素は離合詞全体の意味によっ

て多義性を解消，組み合わせパターンを決める．例えば，

- “鬧病 (naobing)”: a. 病気になる→v+n ○
- b. 騒いで病む→v+v ×

ここで，記号「○」は正しい分類を，また「×」は誤った分類を意味する．また，vは動詞，nは名詞，a（後出）は形容詞を表す．

(2) 多属性のもつ語素は離合詞全体の意味によってその属性を決める．例えば，“招生 (zhaosheng)/学生を募集する”の“生”は，

- a. 名詞の“人/学生”を取る →v+n ○
- b. 動詞の“生/生む”を捨てる→v+v ×

(3) 深い意味をもつ離合詞2語素の属性はそれぞれの表層の意味によって組み合わせパターンを決める．

例えば，

- “吃醋 (chicu)”: a. やきもち ×
- b. 酢を飲む→v+n ○

(4) 形容詞の後ろに後置詞“儿”を付けると，名詞として処理する．例えば，

- a. “叫好 (jiaohao)”: v+a →よいと呼ぶ，
- b. “叫好儿 (jiaohaor)”: v+n →喝采を送る．

以上に書いてある分類ルールによって，表3に示す6冊の教科書から抽出した616語の離合詞に対して統計実験を行った．抽出された離合詞を付録1に示している．

表4では求めた616語離合詞の両語素間にある8種の構造パターンおよび出現頻度を示している．そのうち，「v'」と「a'」は統計の手法に基づいて求めた，結果補語になれる動詞と形容詞を意味する．また中国語文法により，離合詞は以下の4種類に属している．

- ①動賓関係； ②動補関係；
- ③並列関係； ④主述関係．

表5では結果補語になれる常用の38個の動詞を示している．

表3 抽出作業用の資料
Table 3 Collected data for the experiment.

No.	書名	著者名	出版社	年/版	ページ数
01	実用現代漢語語法	劉月華, 他	中国外語数学与研究出版社	86.10/第2版	628
02	現代漢語語法教程	陳国梁	中国西安交通大学出版社	86.5/第2版	404
03	中国語文法の基礎	三野昭一	日本三修社	87.5/第12版	196
04	中国語離合詞500	鹿綜世, 他	日本東方書店	90.11/第1版	238
05	新中国語3	李培文, 他	日本中華書店	91.6/第12版	295
06	文法からの中国語の入門	香坂順一	日本光生館	91.9/第4版	239

表4 離合詞構造パターンおよび出現頻度
Table 4 The liheci structural patterns and frequency.

種類	意味	No.	パターン	例/日本語の意味	数量	頻度	数量	頻度
動賓関係	賓語(目的語)となる語素の多くは単独で単語として用いられた場合, 名詞となるものであるが, 形容詞, 動詞になるものもある.	01	v+n	安心/安心する, 報名/申し込む;	474	76.9	513	83.3
		02	v+a	帮忙/手伝う, 折苦/苦労する;	11	1.8		
		03	v+v	管用/効果がある, 發燒/熱が出る;	28	4.6		
動補関係	補語となる語素は単独で単語として用いられた場合, 動詞になるものが最も多く, その次が形容詞になるものである.	04	v+v'	折散/バラバラする, 超過/越える;	48	7.8	91	14.8
		05	v+a'	提高/レベルを向上する, 抓紧/無駄を作らない;	43	7.0		
並列関係	二つの動詞的な語素が並列されてできあがっている.	06	v+v	記録/記録する, 拍照/写真を取る	8	1.3	8	1.3
主述関係	数が少ない, また使われる率も高くない.	07	n+v	口服/口で従う, 心服/心服する;	2	0.3	4	0.6
		08	n+a	心慌/落ち着きを失う, 心軟/気弱だ;	2	0.3		
合計					616	100	616	100

表6では結果補語になれる常用の37個の形容詞を示している。

表4からわかるように、動賓関係中のパターン「v+n」の出現頻度が一番高く約76.9%であった。また、動賓関係と動補関係を合わせて全体の98.1%であった。本論文で提案する手法は主に動賓関係と動補関係に基づく検討されたものである。次に、離合詞の両語素間に挿入できる成分および特徴について述べる。

4. 挿入成分

本章では、よく使われる5つの挿入成分を選び出し、各挿入成分の代表的な例を用いてその特徴を簡単に説明する。表8に挿入成分および離合詞の構造の一部を示している。

そして次に示した各例文では、 を付けるのは離合詞の両語素または離合詞の日本語訳であり、 を付けるのは両語素間に挿入された成分またはその日本語訳であることを意味する。

4.1 可能不可能を表す成分

4-1-1. 「得 de/不 bu」: 主観的条件から見ても、客観的条件から見ても、可能であること、または不可能であることを表す。「得」は肯定形式であり、「不」は否定形式である。

例文①: 你的技術超得過他嗎?

訳文①: あなたの技術は彼をしのぐことができるか?

例文②: 我超不過他。

訳文②: 私は彼を越えられない。

4-1-2. 「得了 deliao/不了 buliao」: 主観的条件から見て可能であること、または不可能であることを表す。「得了」は肯定形式であり、「不了」は否定形式である。

例文③: 他成績不好, 畢得了業嗎?

訳文③: 彼は成績がよくないのに, 卒業できるか?

例文④: 他成績不好, 畢不了業。

訳文④: 彼は成績がよくないので, 卒業できない。

4-1-3. 「得 de+結果補語/不 bu+結果補語」: 動作を行うと、ある結果が得られるかあるいは得られないかを表す。よく使われる結果補語には「着 zhao」, 「開 kai」, 「住 zhu」などがある。

例文⑤: 今天和王先生見得着面嗎?

訳文⑤: 今日王先生と会うことができるか?

例文⑥: 今天和王先生見不着面。

訳文⑥: 今日王先生と会うことができない。

4-1-4. 「得 de+方向補語/不 bu+方向補語」: 主観的条件、客観的条件からみてある願望や目的が実現できるかどうかを表す。つまり、離合詞の中に挿入される方向補語はすべて派生義を表している。よく使われるものに「上 shang」, 「下 xia」等がある。

表5 動詞結果補語 (v') 一覧

Table 5 Resultant complements of verb.

No.	原形	ピンイン	No.	原形	ピンイン
01	成	cheng	20	跑	pao
02	出	chu	21	入	ru
03	錯	chuo	22	散	san
04	倒	dao	23	傷	shang
05	到	dao	24	上	shang
06	掉	diao	25	死	si
07	定	ding	26	碎	sui
08	丟	diu	27	通	tong
09	懂	dong	28	透	tou
10	動	dong	29	完	wan
11	給	gei	30	為	wei
12	過	guo	31	下	xia
13	翻	fan	32	醒	xing
14	会	hui	33	着	zhao
15	見	jian	34	住	zhu
16	尽	jin	35	走	zou
17	進	jin	36	醉	zui
18	開	kai	37	做	zuo
19	了	liao	38	作	zuo

表6 形容詞結果補語 (a') 一覧

Table 6 Resultant complements of adjective.

No.	原形	ピンイン	No.	原形	ピンイン
01	長	chang	20	緊	jin
02	臭	chou	21	快	kuai
03	粗	cu	22	寬	kuan
04	大	da	23	滿	man
05	淡	dan	24	慢	man
06	低	di	25	明	ming
07	短	duan	26	清	qing
08	对	dui	27	輕	qing
09	多	duo	28	少	shao
10	肥	fei	29	熟	shu
11	高	gao	30	瘦	shou
12	够	gou	31	透	tou
13	慣	guan	32	細	xi
14	光	guang	33	鮮	xian
15	好	hao	34	小	xiao
16	黑	hei	35	窄	zhai
17	紅	hong	36	重	zhong
18	壞	huai	37	准	zhun
19	黃	huang			

例文⑦：時間已経過了，還報得上名？

訳文⑦：期限がもう過ぎたのに，まだ申し込めるか？

例文⑧：報不上名了。

訳文⑧：申し込めない。

4.2 結果補語

一般の動詞と結果補語とを結ぶ場合は，動詞の後ろに結果補語を置くが，離合詞の場合は，それを構成する2つの語素の間に結果補語を挿入する。

4-2-1. 「完 wan」：動作が終わったことを表す。

例文⑨：他游完泳就去商店了。

訳文⑨：彼は泳ぎ終わると買物に行った。

4-2-2. 「好 hao」：動作が終わり，しかもその結果がよかったことを表す。

例文⑩：他們排好隊了。

訳文⑩：彼らはきちんと並び終わった。

4-2-3. 「開 kai」：動作あるいは思考活動について，動作・状態が始まり，しかもそのまま続くことを表す。

例文⑪：小王剛要講話，他們就起開哄了。

訳文⑪：王君が話を始めようとしたとき，彼らは騒ぎ始めた。

4.3 方向補語

すべて派生義を表す。

4-3-1. 「上 shang」：三種類の意味を表す。①ある物がある所に存在させること，②達成しにくい目的を達成したこと，③動作・状態が始まり，しかも続くことを表す。離合詞の意味によって，“上”がどの意味を表すかが決定される。

例文⑫：早就插上秧了。

訳文⑫：とっくに田植えをした。

例文⑬：他總算念上書了。

訳文⑬：彼はやっと学校に入った。

例文⑭：他們第一次見面就聊上天了。

訳文⑭：彼らは初めて会ったのにもうおしゃべりし始めた。

4-3-2. 「起 qi…来 lai」：離合詞の中間に“起”を挿入し，“来”を離合詞の後ろにつける。動作・状態を開始し，しかも継続することを表す。

例文⑮：人們都鼓起掌来。

訳文⑮：人々はみんな拍手し始めた。

4-3-3. 「下 xia…来 lai」：離合詞の中間に“下”を入れ，“来”を離合詞の後ろにつける。非動作動詞についてはある状態が始まり継続すること，あるいは始

まった継続しそうであることを表す。動作動詞についてはこの動作によって事物を固定されたり分離させたりすることを表す。

例文⑯：停下車来等他。

訳文⑯：車を止めて彼を待ってください。

4.4 動態助詞

動態助詞“着”，“了”，“過”は離合詞の中間に挿入しなければならない。

4-4-1. 「着 zhe」：動作が継続する状態を表す。

例文⑰：我讀着書。

訳文⑰：私は本を読んでいる。

4-4-2. 「了 liao」：動作・状態の実現あるいは完成を表す。

例文⑱：聽說他已經辭了職。

訳文⑱：彼はもう退職したそうです。

4-4-3. 「過 guo」：これまでにある経験をしたことを表す。

例文⑲：我到東京以後就沒搬過家。

訳文⑲：私は東京に来てから引越したことがな
い。

4.5 構造助詞“的 de”

離合詞の中間に構造助詞“的”を単独で挿入すると，動作・状態がすでに実現したこと，あるいは完成したことを表す。そして文全体の意味の焦点が，①いつ，②どこで，③誰が，④どのようにして，の説明の上に置かれることになる。

例文⑳：她哪年結的婚？

訳文㉑：彼女は何年に結婚したのですか？

5. 中日機械翻訳における離合詞の処理手法

上述したように，高精度の中日機械翻訳システムを実現するため，離合詞を正しく認識するのは極めて重要である。本章では離合詞の処理手法として1つの試みを述べる。

5.1 離合詞テーブル

離合詞辞書は通常の辞書と独立に構築される。本論

表 7 離合詞テーブルの構造
Table 7 Organization of liheci table.

前語素 (LW 1)	後語素 (LW 2)	挿入成分 の所属	離合詞の 構成条件	離合詞 パターン	その ほか
安	心	A	略	v+n	—
超	過	B	略	v+v'	—
拍	照	C	略	v+v	—
⋮	⋮	⋮	⋮	⋮	⋮

文ではこの離合詞辞書を離合詞テーブルと呼び、その構造を表7に示す。

表7では前語素は離合詞の頭部であり、後語素は離合詞の尾部である。また辞書容量を節約するため、各々の離合詞の挿入成分から幾つかの類別に帰納して処理する。これを挿入成分欄に記入する。本来、この欄に挿入成分の全部のリストを記入すべきであるが、その量は多い。例えば表1に示した“投稿”の場合は少なくとも16種の挿入成分がある。また、大量の実例の検討によって挿入成分に対して若干の類別に分けられるので、本システムではA、Bなど類別だけ登録されている。例えば、 $A = \{3, 4, 14, 25\}$ 。ただし $\{ \}$ は集合の記号であり、 $\{ \}$ 中の数字は表8における離合詞の構造の番号である。

離合詞の構成条件欄に、前語素と後語素から離合詞になる条件を記録している。離合詞パターン欄に本離合詞の種類を登録している。

5.2 訳語関数テーブル

訳語関数とは離合詞を含む中国語文に対応する日本語表現である。例えば、離合詞“投稿”を含む中国語“投稿”について、下記<1>式で表示できる。

<LW1> 了 <LW2> ……<1>

<1>式ではLW1、LW2はそれぞれ離合詞の前、後語素である。本例ではLW1は“投”であり、LW2は“稿”である。

なお、“投稿稿”に対応する日本語表現は“投稿した”であり、これを形式化して<2>式で表される。

<JW> <し> た ……<2>

<2>式ではJWは離合詞自身(離合詞の原形を示す)に対応する日本語である。また、<2>式を<1>式に対応する訳語関数と呼ぶ。JWは“投稿する”である。明らかにJWと<2>式の“した”と連接すれば、動詞の語尾変化ルールに従って変換を行わなければならない。詳細の説明はここで省略したが、訳語関数テーブルの構造を表8に示す。探索の索引として挿入成分を使う。

5.3 処理の手順

中日機械翻訳における離合詞の処理手法としては主に以下の手順で行う。

(1) 離合詞の確認

離合詞の前語素が見つめられると、これをスタックに入れる。そして離合詞テーブルにおける登録されたこの前語素に対応する後語素を取り出し、これを探索の目標として引続き文末まで探索する。もし文中には離合詞テーブルにおける同様な前語素がない場合(同じ前語素からなる異なる離合詞が存在している)、この前語素を含む離合詞が存在しないことを意味するので、離合詞の処理を中止する。もし探索が成功すれば

表8 挿入成分および訳語関数テーブルの一部
Table 8 Organization of translation functions.

No.	挿入成分	No.	離合詞の構造	対応する日本語の表現
1	可能不可能を表す成分	1	<LW1> 得 <LW2>	<JW> ことができる
		2	<LW1> 不 <LW2>	<JW> できない
		3	<LW1> 得了 <LW2>	<JW> ことができる
		4	<LW1> 不了 <LW2>	<JW> できない
		5	<LW1> 得 <着> <LW2>	<JW> ことができる
		6	<LW1> 不 <着> <LW2>	<JW> できない
		7	<LW1> 得 <開> <LW2>	<JW> ことができる
		8	<LW1> 不 <開> <LW2>	<JW> できない
		9	<LW1> 得 <住> <LW2>	<JW> ことができる
		10	<LW1> 不 <住> <LW2>	<JW> できない
		11	<LW1> 得 <上> <LW2>	<JW> ことができる
		12	<LW1> 不 <上> <LW2>	<JW> できない
		13	<LW1> 得 <下> <LW2>	<JW> ことができる
		14	<LW1> 不 <下> <LW2>	<JW> できない
2	結果補語	15	<LW1> 完 <LW2>	<JW> <し> 終わる
		16	<LW1> 完 <LW2> 了	<JW> <し> 終わった
		17	<LW1> 好 <LW2>	きちんと <JW>
		18	<LW1> 好 <LW2> 了	きちんと <JW> <し> 終えた
		19	<LW1> 開 <LW2>	<JW> <し> 始める
		20	<LW1> 開 <LW2> 了	<JW> <し> 始めた
3	方向補語	21	<LW1> 上 <LW2>	<JW> <し> た
		22	<LW1> 起 <LW2> 来	<JW> <し> 始めた
		23	<LW1> 起 <LW2> 来了	<JW> <し> 始めた
		24	<LW1> 下 <LW2> 来	<JW> <て> 来た
		25	<LW1> 下 <LW2> 来了	<JW> <て> 来た
4	動態助詞	26	<LW1> 過 <LW2>	<JW> <し> たことがある
		27	<LW1> 了 <LW2>	<JW> <し> た
		28	<LW1> 着 <LW2>	<JW> <て>
5	構造助詞	29	<LW1> 的 <LW2>	<JW> <し> たのだ

注: <LW1> = 離合詞の前語素;
<LW2> = 離合詞の後語素;
<LW> = <LW1> + <LW2> = 中国語の離合詞;
<JW> = 離合詞に対応する日本語訳;

ば、さらに中間に挿入した成分および他の条件を用い、最終確認を行う。中間に挿入した成分と他の条件とも満足すると、離合詞が確認される。この処理は表7と表8を用いて行われる。

(2) 離合詞への還元

離合詞は中間に挿入した成分によって分離されるので、(1)で離合詞の構造が確認されてから、中間成分を抜き、元の離合詞に還元する必要となる。

(3) 離合詞自身の訳語

(2)中に得た品詞(離合詞自身)を用い、中日辞書を調べて離合詞自身と対応する日本語単語を取り出す。これは表8のJWである。

(4) 日本語表現の生成

訳語関数テーブルを使い、挿入成分を索引として日本語表現を生成する。

(5) 文全体に組み込み

上記の処理は離合詞を含む部分だけを処理したが、これを文全体の処理に組み込まなければならない。

以上の手順をまとめて離合詞の処理の流れを図1に示す。

図1では、RETURN1は離合詞の存在しない出口であり、RETURN2は訳語関数テーブルにデータが不足である出口であり、RETURN3は離合詞を含む表現を正しく処理できる出口である。

6. 実験および考察

本論文で提案した離合詞処理手法の有効性を確認するため、小規模の実験を行った。以下、その結果および考察について述べる。

6.1 実験

実験では、付録1に掲載している離合詞から無作為に50語を抽出し、離合詞テーブルを手作業で構成した。また、訳文関数26個を登録した。表3の文献から登録した離合詞を含む300文を実験対象文として抽出した。現在のシステムでは原文の解析機能がとても不足であることを予想したが、全文の結果が出た文の数は213であり、正しい訳文は101文であった。このまま離合詞処理手法の評価ができないので、入力原文の修正を行った。修正の原則としては離合詞を含む表

現の部分を修正せず、その他の要素、例えば、長い並列句、複雑な補語文、慣用句などを簡略化また省略する。これにより、300文を正しく翻訳した。離合詞を含む表現を正しく処理したのは283文で、正確率が約93%であった。

6.2 考察

実験結果から、離合詞の処理は極めて良好であると判断される。特に、誤った文の特徴を分析して、訳文関数3個を追加した後、翻訳を行った結果、291文が正しく処理された。以上により、本論文で提案した離合詞の処理手法の有効性が確認された。

誤ったのはほとんど離合詞の両語素間に挿入した成分に関わる。以下、その一例を用いて検討しようとする。

例文a：王君這回吃了虧。

本例文では、「吃虧 chikui/損をする」は離合詞である。翻訳する際に、中国語「吃虧(属性はv)」と日本語の訳「損をする(属性はv)」を中日辞書に登録し

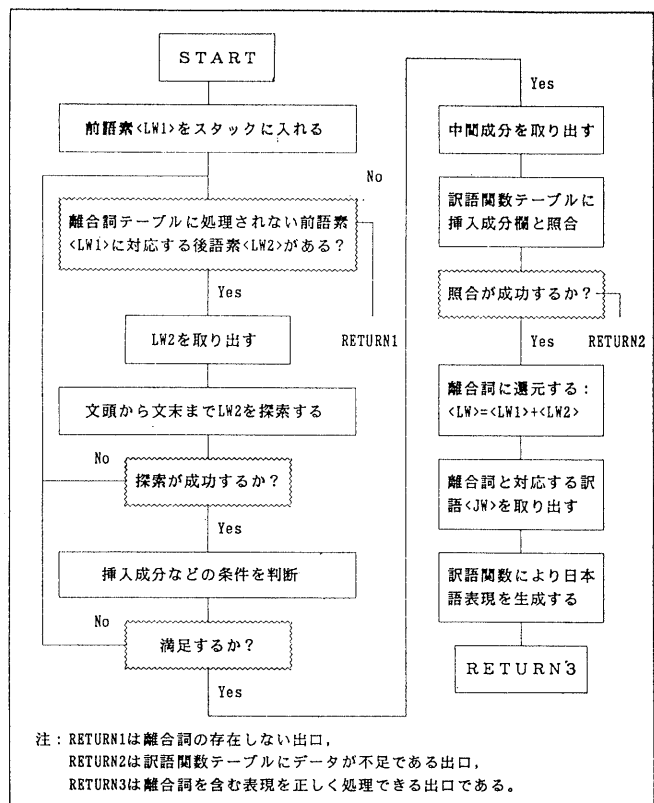


図1 離合詞の処理の流れ
Fig. 1 Flowchart of the algorithm.

ている。また、挿入成分「了」により、その訳文関数は表 8 中の No. 27 に対応しているため、実験の結果としては、以下の訳文が得た。

訳文 a : 王君は今回損をした。

これは正しい訳文であるが、同じ離合詞「吃虧」からなった例文 b を翻訳する際に、問題になる。

例文 b : 王君這回吃了一個大虧。

この文は正しく翻訳できなかった。その処理過程を追跡すると、以下のことが分かった。「吃 chi」は離合詞テーブルの〈LW1〉に登録したので、離合詞テーブルから「吃〈LW1〉」に対応する〈LW2〉を取り出す。勿論「虧 kui」に登録したので、〈LW2(虧)〉と例文の「虧」と照合が成功した。しかし、次の処理が失敗した。これは挿入成分の条件(例文の挿入成分は「了一個大 leyi geda」である)が満足できなかったからである。

本例文の訳文としては次のとおりである。

訳文 b : 王君は今回大きな損をした。

一般的に、もし離合詞の処理で失敗したら、離合詞と認めず通常的手法で処理すべきだろうと考えられるが、本例文に対しても正しい訳文が得られない。これは「吃」に対応する日本語単語は「食べる」なので、意味が間違えるからである。今回の実験ではこのような例が多くなかったが、広範囲にわたる実験を行う際、さらに中間挿入成分に関する検討が必要であると考えられる。

また、本システムでは精度を高めるために1つの開放的な検索システムを考えている。その概念図を図 2 に示す。

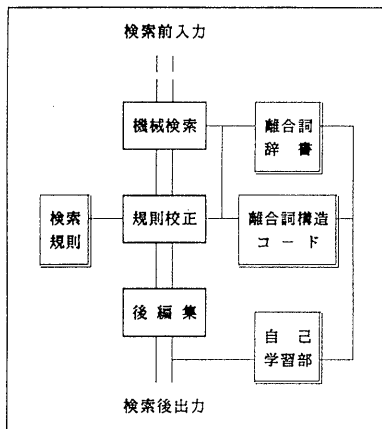


図 2 開放的な検索システム概念図
Fig. 2 Concept of the system.

すなわち、検索後の“後編集部”では検索の結果に対して人手により校正を行う。同時に、システムは間違いやすいところに標識を付ける。その後“自己学習部”を設ける。“自己学習部”とは校正した離合詞の構造特徴から新しい知識を抽出し、“離合詞辞書”と“離合詞構造コード”に記入することである。“検索規則”と“離合詞構造コード”はもう1つのプログラムにより管理し、随時に修正を行うことができる。これにより、開放的な目的を実現し、離合詞の処理精度を高めることができると考えられる。

7. おわりに

本論文では言語の柔軟性に着目して、単語の定義に対して再検討を行い、離合詞という概念および中日機械翻訳における離合詞を含む中国語表現の処理手法を提案した。従来、言語学界では単語の分類を詳細に行ってきたが、中国語の場合、いまなお 10% 以上の単語はすでに設定した分類に入れることができない。特に上述した離合詞という概念は 90 年代の初期、中国言語学界に提出されたばかりで、自然言語理解とくに機械翻訳においては、詳細な分析、研究はまだ開始されていないので、離合詞に対する検討は極めて重要であると考えられる。

本来、単語と非単語あるいは慣用表現の間にはもともとは明確な限界がなく、単語はただ理解と生成を便利にするため設定した最小の形式単位であると考えられる。我々は高精度の中日機械翻訳システムを構築するために、システムでは意味理解あるいは訳文生成を行うとき障害になる単語(離合詞)を1つにまとめて、特別の処理を行っている。

本論文で提案した手法の要点としては、1) 離合詞の構造特徴を検討し、離合詞の分類を行ったこと、2) 実例を用いて文の意味的表現に重要な役割をもつ離合詞の中間挿入成分の類別をまとめたこと、3) 中日機械翻訳においては訳文関数を用いて離合詞を含む中国語表現を日本語表現に変換すること、との3点が挙げられる。

実験により、離合詞を含む表現の正解率は約 93% であった。この結果により、本論文で提案した離合詞処理手法の有効性が確認された。多数の文を対象とし、また広範な専門分野にわたる大規模な実験は行っていないが、教科書、文献などに記載されているような代表的な離合詞はほとんどすべて網羅しているので実用的には充分であると考えられる。

なお、挿入成分の長い文については現在のシステムでまだ処理することができない。また、システム全体としての翻訳能力はまだ充分ではないと考えられる。今後、これらの問題を解決するため、中間挿入成分の系統的に抽出および訳語関数と挿入成分の分類などのより詳細な検討を行う予定である。

参 考 文 献

- 1) 史 有為：柔性觀念在自然言語処理中の体现，IDG 計算機世界，*Communications China Computer-world*, pp. 85-87 (1992).
- 2) 范 莉馨，任 福継，宮永喜一，栃内香次：中日機械翻訳における構文構造の特徴に着目した訳文の生成，電子情報通信学会言語理解とコミュニケーション研究会，信学技報 NLC 91, AI 91-30, pp. 25-32 (1991).
- 3) 范 莉馨，任 福継，宮永喜一，栃内香次：中日機械翻訳における中国語複合語の自動合成について，情報処理学会論文誌，Vol. 33, No. 9, pp. 1103-1113 (1992).
- 4) 鹿 宗世，李 清華，大瀧幸子：中国語離合詞 500，東方書店 (1990).
- 5) 范 莉馨，任 福継，宮永喜一，栃内香次，機械翻訳における中国語の離合詞について，1993年電子情報通信学会春季大会論文集 D-118, p. 6-122 (1993. 3).
- 6) 范 莉馨，任 福継，宮永喜一，栃内香次：中日機械翻訳における離合詞の処理手法，電子情報通信学会言語理解とコミュニケーション研究会，信学技報 NLC 93-35, pp. 33-40 (1993).

(平成 5 年 8 月 10 日受付)

(平成 6 年 5 月 12 日採録)



范 莉馨 (正会員)

1963 年生。1984 年北京郵電大学無線通信専攻卒業。同年、中国郵電部郵電科学研究院に勤務。助理工程師。1991 年北海道大学大学院工学研究科電子工学専攻修士課程修了。1994 年同大学院博士課程修了。工学博士。CSK 研究員，大連理工大学計算機技術研究所助教授兼務。自然言語処理，機械翻訳に関する研究に従事。電子情報通信学会，ACSEJ 各会員。



任 福継 (正会員)

1959 年生。1982 年北京郵電大学電信工部卒業。1985 年同大学院計算機応用専攻修士課程修了。同年同大計算機工部講師。1986 年中国科学院博士課程，1987 年中退来日。1991 年北海道大学工学研究科電子工学専攻博士課程修了。工学博士。CSK 研究員を経て，現在広島市立大学情報科学部助教授。1992 年より大連理工大学計算機学部教授兼務。計算機科学，人工知能，自然言語処理等の研究に従事。人工知能学会，AAMT，ACSEJ 各会員，ACSEJ 会長。



宮永 喜一 (正会員)

1956 年生。1981 年北海道大学大学院工学研究科電子工学専攻修士課程修了。工学博士。現在，北海道大学工学部電子工学科助教授。並列計算機システム，デジタル信号処理等の研究に従事。電子情報通信学会，日本音響学会，IEEE 各会員。



栃内 香次 (正会員)

昭和 14 年生。昭和 37 年北海道大学工学部電気工学科卒業。昭和 39 年同大学院工学研究科修士課程修了。現在同工学部電子工学科教授。工学博士。自然言語処理，音声情報処理および信号処理プロセッサなどの研究に従事。電子情報通信学会，日本音響学会各会員。

附表 常用離合詞一覽

No.	離合詞	No.	離合詞	No.	離合詞	No.	離合詞
01	碍事aishi	02	安家anjia	03	安心ianxin	04	安心2anxin
05	拔河bahe	06	拔尖儿bajianr	07	把关baguan	08	罢工bagong
09	罢官baguan	10	摆阔baikuo	11	拜年bainian	12	搬家banjia
13	帮忙bangmang	14	报警baojing	15	报名baoming	16	报幕baomu
17	报数baoshu	18	报信baoxin	19	报帐baozhang	20	备课beike
21	备料beiliao	22	闭幕bimu	23	毕业biye	24	变心bianxin
25	变质bianzhi	26	表态biaotai	27	并肩bingjian	28	拨款buokuan
29	播音buoyin	30	播种buozhong	31	补考bukao	32	补课buke
33	补票bupiao	34	参军canjun	35	操心caoxin	36	插手chashou
37	插秧chayang	38	插嘴chazui	39	拆散chaisan	40	搀假chanjia
41	唱戏changxi	42	超过chaoguo	43	吵架chaojia	44	撤职chezhi
45	称心chengxin	46	称职chengzhi	47	成材chengcai	48	成家chengjia
49	成名chengming	50	乘凉chengliang	51	吃醋chicu	52	吃饭chifan
53	吃惊chijing	54	吃苦chiku	55	吃亏chikui	56	抽空choukong
57	出差chuchai	58	出厂chuchang	59	出场chuchang	60	出丑chuchou
61	出口chukou	62	出名chuming	63	出气chuqi	64	出神chushen
65	出师chushi	66	出事chushi	67	出血chuxue	68	出院chuyuan
69	除根儿chugenr	70	除名chuming	71	传话chuanhua	72	传令chuanling
73	喘气chuanqi	74	串门儿chuanmenr	75	闯祸chuanhuo	76	吹风chuifeng
77	吹牛chui niu	78	辞行cixing	79	辞职cizhi	80	凑数coushu
81	存货cunhuo	82	存款cunkuan	83	存心cunxin	84	错车cuoche
85	搭伴儿dabanr	86	达到dadao	87	搭手dashou	88	答茬儿dachar
89	打靶daba	90	打败dabai	91	打春dachun	92	打赌dadu
93	打盹dadunr	94	打架dajia	95	打猎dalie	96	打通dalong
97	打仗dazhang	98	打针dazhen	99	打字dazi	100	带头daitou
101	带路dailu	102	担心danxin	103	当家1dangjia	104	当家2dangjia
105	倒霉daomei	106	捣鬼daogui	107	捣乱daoluan	108	得意deyi
109	点名dianming	110	点头diantou	111	掉队diaodui	112	掉色diaose
113	跌跤diejiao	114	顶事dingshi	115	顶用dingyong	116	定案dingan
117	定亲dingqin	118	定性dingxing	119	订婚dinghun	120	丢脸diulian
121	丢人diuren	122	懂事dongshi	123	动工donggong	124	动身dongshen
125	动手1dongshou	126	动手2dongshou	127	动手3dongshou	128	读书1dushu
129	读书2dushu	130	赌气duqi	131	赌钱duqian	132	对号duihao
133	多事duoshi	134	多心duoxin	135	多嘴duozui	136	发财facai
137	发愁fachou	138	发呆fadai	139	发狠fahen	140	发慌fahuang
141	发火儿fahuor	142	熬夜aoye	143	出门chumen	144	倒车daoche
145	发烧fashao	146	发誓fashi	147	翻案fanan	148	翻脸fanlian
149	放假fangjia	150	放心fangxin	151	费事feishi	152	费心feixin
153	分工fengong	154	负责fuze	155	改行gaihang	156	干杯ganbei
157	捣鬼gaogui	158	革命geming	159	隔音geyin	160	鼓掌guzhang
161	挂号guahao	162	拐弯guaiwan	163	管用guanyong	164	过关guoguan
165	过火guohuo	166	过节guojie	167	过年guonian	168	害怕haipa
169	合影heyong	170	狠心henxin	171	滑冰huabing	172	化妆huazhuang
173	怀孕huaiyun	174	还价huanjia	175	还债huanzhai	176	换班huanban
177	换个儿huanger	178	灰心huixin	179	回话huihua	180	发言fayan
181	回信huixin	182	回嘴huizui	183	汇款huiquan	184	会餐huican
185	会客huike	186	集邮jiyou	187	记功jigong	188	记过jiguo
189	记录jilu	190	加班jiaban	191	加油jiayou	192	兼职jianzhi
193	剪彩jiancai	194	见面jianmian	195	讲话jianghua	196	讲课jiangke
197	讲学jiangxue	198	交班jiaoban	199	交差jiaochai	200	交卷jiaojuan

附表 (つづき)

No.	離合詞	No.	離合詞	No.	離合詞	No.	離合詞
201	叫好儿jiaohao	202	叫苦jiaoku	203	叫门jiaomen	204	接班儿jiebanr
205	揭底jiedi	206	揭短jieduan	207	结伴jieban	208	结仇jiechou
209	结婚jiehun	210	结业jieye	211	结账jiezhang	212	解饿jieer
213	解乏jiefa	214	解恨jiehen	215	解闷jiemen	216	解聘jiepin
217	解气jieqi	218	解手jieshou	219	解围jiwei	220	借债jiezhai
221	尽力jinli	222	经手jingshou	223	揪心jiuxin	224	就伴儿jiubanr
225	就业jiuye	226	救火jiuhuo	227	救急jiuji	228	救命jiuming
229	鞠躬jugong	230	捐款juankuan	231	绝食jueshi	232	开饭kaifan
233	开工kaigong	234	开锅kaiguo	235	开花kaihua	236	开会kaihui
237	开课kaike	238	开课2kaike	239	开门kaimen	240	开门2kaimen
241	开头儿kaitour	242	开学kaixue	243	开眼kaiyan	244	开业kaiye
245	开张kaizhang	246	看病kanbing	247	看见kanjian	248	口服koufu
249	夸口kuakou	250	旷课kuangke	251	亏本kuiben	252	拉平laping
253	来信laixin	254	劳驾laojia	255	落色luose	256	离婚lihun
257	理发lifa	258	立功ligong	259	联欢lianhuan	260	聊天liaotianr
261	领情lingqing	262	领先lingxian	263	留级liuji	264	留念liunian
265	留神liushen	266	留学liuxue	267	留影liuying	268	露面儿lumianr
269	露馅儿luxianr	270	录音luyin	271	骂街majie	272	买帐maizhang
273	卖命maiming	274	卖劲儿maijinr	275	满额maner	276	毛腰maoyao
277	冒名maoming	278	冒险maoxian	279	迷路milu	280	命名mingming
281	命题mingti	282	摸底modi	283	摸门儿momenr	284	募捐mujun
285	纳闷儿namenr	286	纳税nashui	287	闹病naobing	288	闹鬼naogui
289	闹事naoshi	290	逆稿儿nigaor	291	逆风nifeng	292	念经nianjing
293	念书1nianshu	294	念书2nianshu	295	努力nuli	296	拍板paiban
297	拍板2paiban	298	拍手paishou	299	拍照paizhao	300	排版paiban
301	排队paidui	302	排戏paixi	303	盘货panhuo	304	盘腿pantui
305	跑步paobu	306	赔本peiben	307	赔款peikuan	308	赔礼peili
309	碰杯pengbei	310	碰见pengjian	311	碰头pengtou	312	拼命pinming
313	评奖pingjiang	314	破案poan	315	破产pochan	316	破例poli
317	起草qicao	318	起床qichuang	319	起哄qihong	320	起火qihuo
321	起头qitou	322	牵线qianxian	323	签到qiandao	324	签名qianming
325	签字qianzi	326	抢先qiangxian	327	请假qingjia	328	请客qingke
329	求婚qihun	330	求情qiuqing	331	求人qiuren	332	劝架quanjia
333	缺德quede	334	缺勤queqin	335	缺席quexi	336	让步rangbu
337	让座rangzuo	338	绕远儿raoyuanr	339	惹祸rehuo	340	惹事reshi
341	认错rencuo	342	认输renshu	343	认帐renzhang	344	任职renzhi
345	入境rujing	346	入迷rumi	347	入学ruxue	348	入院ruyuan
349	撒谎sahuang	350	散步sanbu	351	散会sanhui	352	扫地saodi
353	扫兴saoxing	354	煞车shache	355	傻眼shayan	356	伤心shangxin
357	上班shangban	358	上当shangdang	359	上课shangke	360	上税shangshui
361	上学1shangxue	362	上学2shangxue	363	上瘾shangyin	364	生病shengbing
365	生气shengqi	366	升班shengban	367	升级shengji	368	升学shengxue
369	生火shenghuo	370	省事shengshi	371	省心shengxin	372	失火shihuo
373	失眠shimian	374	失效shixiao	375	失信shixin	376	失学shixue
377	失业shiyue	378	失约shiyue	379	失职shizhi	380	施肥shifei
381	识字shizi	382	使坏shihuai	383	使劲shijin	384	收工shougong
385	受潮shouchao	386	受罚shoufa	387	受害shouhai	388	受贿shouhui
389	受奖shoujiang	390	受惊shoujing	391	受苦shouku	392	受凉shouliang
393	受骗shoupian	394	受气shouqi	395	受穷shouqiong	396	受伤shoushang
397	授训shouxun	398	受灾shouzai	399	受罪shouzui	400	授奖shoujiang
401	梳头shutou	402	输血shuxue	403	摔交shuaijiao	404	睡觉shuijiao
405	顺心shunxin	406	说服shuofu	407	说话shuohua	408	说理shuoli

付表 (つづき)

No.	離合詞	No.	離合詞	No.	離合詞	No.	離合詞
409	说情shuoqing	410	死心sixin	411	送礼songli	412	送行songxing
413	帐suanzhang	414	随便suibian	415	随心suixin	416	抬杠taigang
417	抬头taitou	418	谈话tanhua	419	探亲tanqin	420	叹气tanqi
421	烫发tangfa	422	逃学taoxue	423	讨好几taohaor	424	讨厌taoyan
425	题词tici	426	提成ticheng	427	提高tigao	428	提价tijia
429	提议tiyi	430	挑刺儿tiaocir	431	跳班tiaoban	432	跳舞tiaowu
433	听话tinghua	434	听见tingjian	435	停工tinggong	436	通车tongche
437	通话tonghua	438	通信tongxin	439	同班tongban	440	同事tongshi
441	同学tongxue	442	偷懒toulan	443	投稿tougao	444	投票toupiao
445	投资touzi	446	退步tuibu	447	退票tuipiao	448	退烧tuishao
449	脱身tuoshen	450	托福tuofu	451	违法weifa	452	问好wenhao
453	握手woshou	454	误事wushi	455	洗澡xizao	456	下班儿xiabanr
457	下课xiake	458	下台xiatai	459	下乡xiaxiang	460	现眼xiayan
461	享福xiangfu	462	象样xiangyang	463	泄密xiemi	464	泄气xieqi
465	卸车xieche	466	谢幕xiemu	467	心服xinfu	468	心慌xinhuang
469	心软xinruan	470	行贿xinghui	471	行礼xingli	472	休假xiujia
473	压服yafu	474	延期yanqi	475	养病yangbing	476	要命yaoming
477	用功yonggong	478	用力yongli	479	用心yongxin	480	游行youxing
481	游泳youyong	482	遇难yunan	483	晕车yunche	484	砸锅zaguo
485	栽脏zai zang	486	遭殃zaoyang	487	造反zaofan	488	造孽zaonie
489	造谣zaoyao	490	增产zengchan	491	扎根lzhagen	492	扎根2zhagen
493	扎针zhazhen	494	沾边zhanbianr	495	沾光zhanguang	496	张嘴lzhangzui
497	张嘴zhangzui	498	掌舵zhangduo	499	掌勺zhangshao	500	仗势zhangshi
501	掌权zhangquan	502	招标zhaobiao	503	招兵zhaobing	504	招股zhaogu
505	招生zhaosheng	506	招手zhaoshou	507	着火zhaohuo	508	着急zhaoji
509	着凉zhaoliang	510	找钱zhaopian	511	找事lzhaooshi	512	找事2zhaooshi
513	照办zhaoban	514	征兵zhengbing	515	争光zhengguang	516	值班zhiban
517	执政zhizheng	518	止步zhibu	519	争气zhengqi	520	中毒zhongdu
521	中奖zhongjiang	522	中暑zhongshu	523	中意zhongyi	524	种地zhongdi
525	种花zhonghua	526	抓紧zhuajin	527	抓药zhuayao	528	转车zhuanche
529	转脸zhuilian	530	转身zhuanshen	531	转弯儿zhuawan	532	转学zhuaxue
533	转业zhuanye	534	转正zhuanzheng	535	转向zhuangxiang	536	装蒜zhuangsuai
537	撞车zhuangche	538	走调儿zoudiaor	539	走路zoulu	540	走神儿zoushenr
541	走样儿zouyangr	542	走运zouyun	543	坐牢zuolao	544	作案zuotan
545	作弊zuobi	546	作废zuofei	547	做伴儿zuobanr	548	做东zuodong
549	做工zuogong	550	做梦zuomeng	551	做主zuozhu	552	打倒dadao
553	说明shuoming	554	打开dakai	555	听见tingjian	556	遇见yujian
557	碰见pengjian	558	拿住nazhu	559	记住jizu	560	搬开bankai
561	开开kaikai	562	离开likai	563	站住zhanzhu	564	站稳zhanwen
565	想起xiangqi	566	留下liuxia	567	看出kanchu	568	想到xiangdao
569	批示pishi	570	批注pizhu	571	通报tongbao	572	上来shanglai
573	上去shangqu	574	下来xialai	575	下去xiaqu	576	进来jinlai
577	进去jinqu	578	出来chulai	579	出去chuqu	580	回来huilai
581	回去huiqu	582	过来guolai	583	过去guoqu	584	起来qilai
585	照相zhaoxiang	586	划船huachuan	587	唱歌changge	588	刮风guafeng
589	踢球tiqi	590	跳高tiaogao	591	跳远tiaoyuan	592	停车tingche
593	考试kaoshi	594	报仇baochou	595	打球daqiu	596	开车kaiche
597	上山shangshan	598	上车shangche	599	上楼shanglou	600	上街shangjie
601	下楼xialou	602	变调biandiao	603	调动diaodong	604	干活ganhuo
605	关心guanxin	606	集合jihe	607	遇到yudao	608	构成goucheng
609	隔行gehang	610	费力feili	611	回头huitou	612	发话fahua
613	发言fayan	614	留心liuxin	615	迟到chidao	616	翻案fanan