

同形異義語を対象とした中国語学習システムの構築 Development of Chinese Language Learning System for Homograph

西田 昌史[†] 周 元晨[†] 綱川 隆司[†] 西村 雅史[†]
Masafumi Nishida Genshin Shu Takashi Tsunakawa Masafumi Nishimura

1. はじめに

近年、コンピュータ及びインターネット技術の普及に伴い、コンピュータを用いた外国語の学習システムが開発され、人々が学習支援システムにより外国語を学習することが容易になった。中国語においては、日本の小学校で学ぶ必修漢字（約 1600 字）に着目し、日本人による中国語学習のために「日中漢字データベース」の開発が行われてきた[1]。他にも、日本語と中国語の意味差異を対象としたエージェント指向の漢字学習システムについても研究が行われてきた[2]。さらに、発音の学習については、非ネイティブ話者のための中国語有気音と無気音を識別する方法と特徴量の事後確率に基づいたスコアリング方法などの研究が行われた[3]。

本研究では、日本人にとって間違いやすい同形異義語に着目した中国語学習システムを提案する。本システムでは、各漢字の子音と母音の発音を学び、ユーザが学習した後に音声認識を用いてユーザの発音を評価する。その際、音声認識で認識された単語のピンイン情報をユーザに提示し、間違えて認識された場合に、この認識された単語のピンイン情報を見て学習する。また、音声の高さの情報（中国語の韻律情報）を Wavesurfer で可視化したものを提示する。さらに、システムから提示された文章の翻訳課題を行うことで、ユーザが同形異義語を学習することができる。

2. 提案システム

2.1 日中漢字の同形異義語のデータベース

これまでにも日中漢字のデータベースは作成されてきた。しかし、日中の漢字の中には、同じ漢字が中国語にも日本語にも使われているが、その中に表記が同じであっても意味が異なる漢字がたくさんあり、従来のデータベースではこれらについて十分に検討されていなかった。

本研究では、これらの漢字を対象にデータベースを作成した。具体的には、文献[4]を用いて日中の漢字の同形異義語の中からよく使われ、よく誤解しやすい 100 個の漢字をまとめ、データベースを作成した。そして、中国語の各漢字に特定なピンインがあり、ピンインなどにより発音が変わるため、本システムでは、同形異義語を対象とした漢字のピンイン音声とピンインの子音と母音の情報をデータベースに登録してある。これらの音声情報で発音の学習ができる。提案システムを図 1 に示す。

日本語漢字	日本語意味	中国語漢字	中国語意味	日本語読み	ピンイン	母音1	子音1	母音2	子音2	number
結束	unity	結束	完了	棚孔	jie2shu4	j	sh	ie	u	1
丈夫	strong solid	丈夫	男	牢固	zhang4fu1	zh	ang	f	u	1
上手	well	上手	意味がない	在行	shang4sh...sh	sh	ang	sh	ou	1
切手	Stamp	切手	手を切る	邮票	qie1shou3	q	ie	sh	ou	1
勉強	study	勉強	強制する	学习	mian3qia...m	m	ian	q	iang	1
好	好み, 好き	好	いい	喜欢	hao3	h	ao			1
娘	daughter	娘	母	女儿	niang2	n	iang			2
家内	wife	家内	家の中	老婆	jia1nei4	j	ia	n	ei	2
息子	son	息子	利益の単位	儿子	er2zi3		er	z	i	2
愛人	lover	爱人	妻	老婆	ai4ren2	ai	r	en		2
手紙	letter	手紙	テッシュ	邮票	shou3zhi3	sh	ou	zh	i	2

図 1 日中漢字の同形異義語のデータベース

2.2 音声認識を用いた発音評価

本システムでは、音声認識技術により同形異義語の発音を学習する。日中漢字の同形異義語のデータベースをもとにユーザが学習したい同形異義語を検索でき、ピンインなどの音声情報と音声の高さなどの韻律情報が表示されて学習する。これらを学習した後に、ユーザは音声認識の技術により直接自分で発音の正しさを確認することができる。音声認識には、baidu voice recognition API [5]を用いた。その後、認識された単語の情報をユーザに表示するので、音声認識の結果が間違った場合に、認識された単語ピンインを見て、発音の違いを知ることができる。

2.3 単語に着目した翻訳評価

提案システムで発音の学習をした後に、同形異義語の意味差を学習するために、翻訳評価の機能をシステムに実装した。翻訳評価では、主に同形異義語の意味を学習するので、ユーザはシステムが用意した日中の同形異義語が入った中国語の文章を日本語に訳し、システムはユーザが訳した日本語の文を評価する。この評価方法は、正しい日本語の単語が日本語の文中に含まれているかどうかをチェックする。例えば：“工作结束了”という中国語の文章があって、“工作”は“仕事、ジョブ、アルバイト”の意味、“结束了”は“終わった、完了だった、終了だった”の意味である。訳された日本語の文中に“仕事、ジョブ、アルバイト”と“終わった、完了だった、終了だった”などの単語が入っているかどうかを確認することで、翻訳が正しく行われているかどうか評価を行う。

[†] 静岡大学 Shizuoka University

3. 評価実験

3.1 実験方法

本実験では、日本人の被験者4名と日本語ができる外国人の被験者4名で、意味学習と発音学習の実験を行った。意味学習については、10種類の同形異義語を対象として、これらの同形異義語を使った中国語の文章を日本語に訳してもらった。まず、被験者に学習をせずに文章の翻訳を行ってもらった。次に、システムに登録している意味情報を学習してもらい、もう一度文章の翻訳を行って評価した。

発音学習については、まず10種類の同形異義語のピンインを見せて三回発音をしてもらい、音声認識で評価を行った。次に、これらの同形異義語のピンイン情報と音声の高さなどの情報を見て学習してもらった。その後、もう一度三回発音して音声認識で評価した。ここで、間違っ

3.2 実験結果

8名の被験者で実験を行った結果を表1と表2に示す。表1は10種類の同形異義語についての実験の翻訳正答率である。表2は各単語で学習前後に分けて三回発音して、発音の評価を行った結果である。表2にある数字は[学習前の認識正答回数/学習後の認識正答回数]である。

表1 意味学習結果

被験者	学習前(%)	学習後(%)
A(日本人)	10	80
B(日本人)	20	90
C(日本人)	30	80
D(日本人)	0	70
E(マレーシア)	80	100
F(韓国)	30	50
G(インドネシア)	0	20
H(ベトナム)	50	70

表2 発音学習結果（学習前／学習後）

	A	B	C	D	E	F	G	H
好	3/3	3/3	3/3	3/3	3/3	3/3	3/3	3/3
娘	3/2	2/2	2/3	3/3	1/2	0/0	1/1	3/3
老婆	2/2	3/3	3/3	2/3	3/3	2/1	3/3	3/3
方便	3/3	3/2	0/2	2/2	3/3	1/3	3/3	1/1
交代	2/3	1/3	2/3	2/3	3/3	3/3	3/3	3/3
一同	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
改行	2/1	3/3	2/2	3/3	3/3	1/3	1/1	2/1
看病	0/0	2/1	0/0	0/2	3/3	0/0	2/1	0/0
今朝	3/2	3/3	3/3	1/3	3/3	2/1	2/2	2/3
先生	0/0	0/3	3/3	1/2	2/3	1/0	1/3	2/2

3.3 考察

意味学習については、表1の結果から学習前の平均正答率は23%、学習後の平均正答率は70%となった。表1の結果より、提案システムで学習することで翻訳の正答率が明らかにようになっていた。同形異義語に注目した単語の学習は文章の理解に役に立つことがわかった。そして、日本人の被験者と日本人以外の被験者を比較すると、学習前の正答率は日本人の被験者が低かったため、同形異義語が入っている文章は日本人にとって誤解しやすいことがわかった。

発音学習については、表2の結果から学習前の平均正答率は61%、学習後の平均正答率は67%であった。表2の結果より、提案システムで学習することで若干正答率が上がる傾向があるが、学習前後の正答率の差は小さかった。学習前後の正答率を比較すると、正答率が上がる割合は約21%で、正答率が下がる割合は約13%である。また、多くの被験者の正答率が低い単語と正答率が高い単語はほとんど同じであるため、発音しやすい単語と発音しにくい単語があると考えられる。そして、“一同”の単語の発音は全員が認識できなかった。この単語はよく似た発音の単語“一桶”に認識された。これは音声認識で認識されにくい単語であると考えられる。

4. おわりに

本研究では、日中の同形異義語を対象にデータベースを作成し、翻訳の課題をシステムが提示することで同形異義語を学習できるシステムを提案した。また、本システムでは同形異義語の発音情報（ピンインや母音、子音など）と音声の高さの情報を提示することで発音の学習を行うことができ、音声認識技術を用いて発音の評価を行った。

評価実験の結果、提案システムにより同形異義語の意味の学習を行うことで、翻訳の正答率は学習前に比べて大きく改善した。また、提案システムにより同形異義語の発音の学習を行うことで、発音の正答率は学習前に比べて若干改善した。以上の結果から、日中の同形異義語を学習する際に、提案システムが有効であることが明らかになった。

今後は、発音の学習方法の改善やより多くの同形異義語に対して多くの被験者を対象に評価を行う必要がある。

参考文献

- [1]須曾野 仁志, 下村 勉, 王 博, 彭 宇, “日本人中国語学習のための日中漢字データベース”, 日本科学教育学会年會論文集, 31 (2007).
- [2]刘 玉琴, 緒方 広明, 越智 洋司, 矢野 米雄, “日中の意味差に着目したエージェント指向漢字学習支援システム—Anckle”, 電子情報通信学会論文誌, Vol. J82-D-II, No. 10 (1999).
- [3]田嘉 鹏, 三輪 譲二, “中国語発音教育のための有気音と無気音の識別とスコアリング”, 電子情報通信学会論文誌, D-II Vol. J85-D-II, No. 5, pp. 842-850 (2002).
- [4]郭 明輝, 磯部 祐子, 谷内 美江子 “日中同形異義語 1500”, 国際語学社出版 (2011).
- [5]baidu voice recognition API : <http://ai.baidu.com/docs#/ASR-Online-Java-SDK/d7497018>