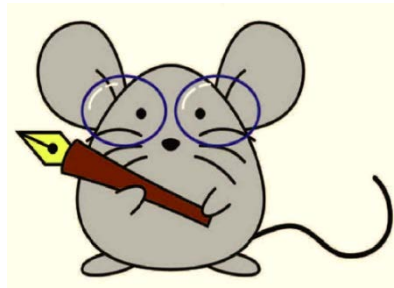


日本語読解支援システム 『リーディング・チュウ太』の 難易度判定ツール



東京国際大学
川村よし子

日本語読解支援システム 『リーディング・チュウ太』の 難易度判定ツール

- 1. 先行研究
- 2. 「単語チェッカー」の概要
- 3. 「単語チェッカー」の運用実験
- 4. 「単語チェッカー」と「語彙チェッカー」
- 5. 今後の課題

はじめに

- 目的：
「日本語教育語彙表」（以下「語彙表」）を組み入れた
単語の**難易度判定ツール**の開発
- 日本語読解支援システム『リーディング・チュウ太』
 - ◎辞書ツール
 - ◎難易度判定ツール
 - 語彙チェッカー
 - 漢字チェッカー

（旧日本語能力試験出題基準に準拠）

- ◎読解教材バンク （川村 2009）

はじめに

- 目的：
「日本語教育語彙表」（以下「語彙表」）を組み入れた
単語の**難易度判定ツール**の開発
- 日本語読解支援システム『リーディング・チュウ太』
 - ◎ 辞書ツール
 - ◎ **難易度判定ツール**
 - 語彙チェッカー
 - 漢字チェッカー

（旧日本語能力試験出題基準に準拠）

- ◎ 読解教材バンク （川村 2009）

1. 先行研究

◎背景

- 日本語能力試験は2010年に改訂
- 新しい出題基準は非公開
- CEFR・JF日本語教育スタンダードに準拠した教育



単語の難易度を示す新たな指標が必要

◎『リーディング・チュウ太』の模索

- 旧日本語能力試験出題基準に準拠⇒語彙チェッカー
- 単語親密度（NTT日本語の語彙特性）を指標（川村・北村 2008）
⇒単独で使われることの少ない語に問題あり
例：する・くる・これら・どの・くださる
- 新聞（朝日新聞14年分）の出現頻度を指標（川村 2009）
- 新聞のIDF（出現の偏り）を指標（北村・富岡・川村 2009）
⇒いずれも初級の単語が漏れてしまう
例：たぶん・水曜・どなた・おととい・あちら

2. 「単語チェッカー」の概要

学習辞書科研が作成した「日本語教育語彙表」

「現代日本語書き言葉均衡コーパス (BCCWJ)」および
「日本語教科書コーパス」に準拠

- ① 約1万8000語の見出し語を選定
- ② 現代日本語の使用実態と学習者の必要を反映
- ③ 語彙レベルとしては、6レベルを設定

語彙表の学習者レベル	初級前半	初級後半	中級前半	中級後半	上級前半	上級後半
レベル別単語数	424	794	2306	6471	6392	1562
当該レベルまでの累積語数	424	1218	3524	9995	16387	17949
単語チェッカーの表示 Level	Level 5	Level 4	Level 3	Level 2	Level 1	その他

「語彙表」の上級後半が1562語のみため、

「単語チェッカー」では上級後半も「その他」として分類

2. 「単語チェッカー」の概要

学習辞書科研が作成した「日本語教育語彙表」



「語彙チェッカー」の仕様を踏襲

①形態素解析

MeCabによって単語に分割

②単語のレベル（難易度）

「語彙表」に準拠

③レベルによる色分け

④レベル別単語リスト

⑤レベル別単語数

チュウ太のレベルチェッカー

開発中のため予告なく仕様が変更されることがあります。

様々な情報をアナウンスするためのブログを開設しました(2013/4/15)。 <http://chutainfo.blogspot.jp/>

- 2014.04.07 [ICJLE2014採択](#)
- 2014.04.07 [音アプリ開発に関する技術講習会](#)
- 2014.04.07 [エディタのセミナー](#)

入力した文章を単語ごとにレベル分け・色分け・カウントします。

↓文章を入力して下さい。

1. はじめに

本発表では、単語の難易度判定ツール「単語チェッカー」の開発について報告する。この「単語チェッカー」は、日本語読解学習支援システム『リーディング・チュウ太』の支援ツールの一つとして、学習辞書科研との連携によって開発したツールである。

『リーディング・チュウ太』は、1999年から Web (<http://language.tiu.ac.jp/>) 上で公

goi kaigo goi+bun keizaigo **tango** reset

freq fami mlDF ylDF freq6



- goi: 旧日本語能力試験出題基準に基づく分類
- kaigo: 介護用語のレベルに基づく分類
- goi+bun: goiの機能に加え、一文中の平均文字数および文章のレベルを表示
- keizaigo: 経済用語のレベルに基づく分類
- tango: 筑波大学との共同研究による新基準に基づく分類

1. はじめに

本発表では、単語の難易度判定ツール「単語チェッカー」の開発について報告する。この「単語チェッカー」は、日本語支援システム『リーディング・チュウ太』の支援ツールの一つとして、学習辞書科研との連携によって開発したツールである。『リーディング・チュウ太』は、1999年からWeb(<http://language.tiu.ac.jp/>)上で公開され、支援環境を提供しているが、ここには、旧日本語能力出題基準(以下「出題基準」)に準拠して語彙や漢字のレベルを判定する語彙チェッカーや漢字チェッカーが搭載されている(川村2009)。この語彙チェッカーは、多くの日本語関係者が教材作成や研究に利用している。ところが、日本語能力試験は2010年に改訂され、以降、単語や漢字に関する新しい出題基準は明らかにされていない。また、近年、CEFRやJF日本語教育スタンダードに準拠した新しい形の教育も行われるようになってきた。そのため、各レベルの学習者がどのような単語を習得しておく必要があるのかについて新たな指標が求められている。一方、川村・北村(2009)では単語親密度、川村・北村・富岡(2009)では、新聞やYahoo知恵袋、北村・川村(2010)では日本語教科書における単語の出現頻度を活用して様々な難易度判定ツールの開発を進めてきたが、日本語学習者のための単語の難易度判定ツールとしては、語彙チェッカーに代わるものがないというのが現状であった。こうした中、日本語教育の視点から個々の単語の難易度の新たな基準を示すものとして、学習辞書科研が作成した「日本語教育語彙表」(以下「語彙表」)は大きな意味を持っている。今回開発した「単語チェッカー」は、入力された文章を、自動で単語に区切り、入力文中のすべての単語のレベルを「語彙表」をもとに判定するツールである。本発表では、この「単語チェッカー」の概要を報告するとともに、その機能を「出題基準」に準拠した「語彙チェッカー」と分析することによって、「語彙表」の特徴を明らかにすることを目的としている。

総数	語彙総数	level5	level4	level3	level2	level1	その他	助詞,助動詞,接続詞	記号類
484	381	45	14	75	48	23	33	143	103
127.0%	100.0%	11.8%	3.7%	19.7%	12.6%	6.0%	8.7%	37.5%	27.0%
(172)	(143)	(20)	(9)	(30)	(31)	(11)	(14)	(28)	(29)
120.3%	100.0%	14.0%	6.3%	21.0%	21.7%	7.7%	9.8%	19.6%	20.3%

- Level5 45 (20)
 - 新しい 2
 - どの 1
 - 新聞 1

- ある 1
- 上 1
- 漢字 3
- 年 2
- ない 1
- この 3
- Level4 14 (9)
 - 度 4
 - 親 1
 - もと 1
 - こと 2
 - 試験(する) 2
 - なる 1
 - 持つ 1
 - 大きな 1
 - 中 1
- Level3 75 (30)
 - もの 2
 - 形 1
 - 利用(する) 1
 - 以降 1
 - 学習(する) 6
 - 能力 2
 - 表 4
 - 必要(な)だ 1
 - ため 2
 - 科 2
 - 代わる 1
 - 各 1
 - 多く 1
 - 語彙 9
 - 者 3
 - 読解(する) 2
 - 研究(する) 1
 - はじめ 1
 - 今回 1
 - 単語 13




3. 「単語チェッカー」の運用実験

- ① 語彙表に準拠してレベル判定できているか
- ② 語彙表の判定基準で単語のレベル判定がうまくできるか

3. 「単語チェッカー」の運用実験

- ① 語彙表に準拠してレベル判定できているか
- ② 語彙表の判定基準で単語のレベル判定がうまくできるか

- ある 1
- 上 1
- 漢字 3
- 年 2
- ない 1
- この 3
- Level4 14 (9)
 - 度 4
 - 親 1
 - もと 1
 - こと 2
 - 試験(する) 2
 - なる 1
 - 持つ 1
 - 大きな 1
 - 中 1
- Level3 75 (30)
 - もの 2
 - 形 1
 - 利用(する) 1
 - 以降 1
 - 学習(する) 6
 - 能力 2
 - 表 4
 - 必要(な)だ 1
 - ため 2
 - 科 2
 - 代わる 1
 - 各 1
 - 多く 1
 - 語彙 9
 - 者 3
 - 読解(する) 2
 - 研究(する) 1
 - はじめ 1
 - 今回 1
 - 単語 13

- **Level 23 (11)**
 - 頻度 1
 - 概要 1
 - 密度 1
 - 搭載(する) 1
 - 難易 4
 - ツール 6
 - スタンダード 1
 - 判定(する) 5
 - 連携(する) 1
 - 指標 1
 - 分析(する) 1
- **その他 33 (14)**
 - 富岡 1
 - の 2 
 - 川村 4
 - 研 2
 - 太 2
 - 北村 3
 - すべて 1 
 - 改訂(する) 1
 - おく 1 
 - 知恵袋 1
 - 準拠(する) 3
 - リーディング・チュウ 2
 - こうした 1 
 - チェッカー 9
- **助詞,助動詞,接続詞 143 (28)**
 - の 23
 - られる 1
 - か 1

3. 「単語チェッカー」の運用実験

① 語彙表に準拠してレベル判定できているか

問題点

a. 異表記への対応

「全て」「置く」等がひらがな表記された場合**その他**になる

⇒対応：「すべて」「おく」を異表記として登録

ただし、買う・書く (Level5) 等で「かう」「かく」を異表記登録すると「飼う」 (Level4) 「掻く」 (Level3) がひらがな表記された場合にはLevel5と判定されてしまう。

b. MeCabの分析への対応

「こうした」「の」等が**その他**になる

⇒対応：「こうした」をリストに追加登録

⇒対応：「の」を名詞としてもリストに追加登録

3. 「単語チェッカー」の運用実験

① 語彙表に準拠してレベル判定できているか

② 語彙表の判定基準でレベル判定がうまくできるか

親密度・新聞の単語頻度・新聞の単語IDFで問題となった単語はすべて語彙表にあり適切なレベル判定ができる

Level5

くる

どの

する

おととい

水曜

Level4

あちら

これら

たぶん

どなた

Level3

くださる

⇒ただし、「ください」が、Level3になってしまう

<要検討>

4. 単語チェッカーと語彙チェッカー

4.1 レベル別単語数と累積語数の比較

表1 「語彙表」のレベル別単語数と累積語数

語彙表の学習者レベル	初級前半	初級後半	中級前半	中級後半	上級前半
レベル別単語数	424	794	2306	6471	6392
当該レベルまでの累積語数	424	1218	3524	9995	16387
単語チェッカーの表示Level	Level 5	Level 4	Level 3	Level 2	Level 1



表2 「出題基準」のレベル別単語数と累積語数

出題基準の級	4級	3級	2級	1級
級別単語数	728	681	3626	2974
当該級までの累積語数	728	1409	5035	8009
語彙チェッカーの表示Level	Level 4	Level 3	Level 2	Level 1

⇒語彙表では中級後半で、すでに累積語数が9995語

4. 単語チェッカーと語彙チェッカー

4.2 単語のカバー率の比較

次の文章を「単語チェッカー」「語彙チェッカー」にかけ
おのおののリストのカバー率を調査する

- 〈入力文〉 a. 2014年3月11日の『朝日新聞』のトップニュース10件
b. 2014年3月11日のNHKのWebニュース10件

表3 単語チェッカーと語彙チェッカーの分析結果

	単語チェッカー（語彙表）	語彙チェッカー（出題基準）	
	初級—中級後半のカバー率	2級-4級のカバー率	1級-4級のカバー率
朝日新聞	85.30%	79.40%	85.20%
NHKニュース	87.10%	82.40%	87.40%

⇒中級後半までのカバー率は、2-4級のカバー率よりも高い

⇒ただし、「出題基準」は2級までの累積語数は5035語

「語彙表」は中級後半までで9995語

⇒1-4級（8009語）のカバー率と比較するとほぼ同じ

4. 単語チェッカーと語彙チェッカー

結論：

1. 語彙表の収録語の選定方法は学習者にとって有用
2. 中級後半までに約1万語の単語を習得できるのであれば、語彙表を学習に利用することによって出題基準の1級に相当するカバー率の語彙学習が可能になる

課題：

1. 学習者は中級後半までに何語学習しなければならないのか
or 中級後半までに何語学習することが可能なのか
⇒ 語彙表の利用には十分な配慮と教え方の工夫等が必要
2. 「語彙表」の中級前半の単語数が2306語であるのに対して、中級後半の単語数が6471語と言うのはバランスが悪い
⇒ 「語彙表」の改訂時には、この点について十分な吟味が必要

4. 単語チェッカーと語彙チェッカー

4.1 レベル別単語数と累積語数の比較

表1 「語彙表」のレベル別単語数と累積語数

語彙表の学習者レベル	初級前半	初級後半	中級前半	中級後半	上級前半
レベル別単語数	424	794	2306	6471	6392
当該レベルまでの累積語数	424	1218	3524	9995	16387
単語チェッカーの表示Level	Level 5	Level 4	Level 3	Level 2	Level 1

表2 「出題基準」のレベル別単語数と累積語数

出題基準の級	4級	3級	2級	1級
級別単語数	728	681	3626	2974
当該級までの累積語数	728	1409	5035	8009
語彙チェッカーの表示Level	Level 4	Level 3	Level 2	Level 1

5. 今後の課題（単語チェッカー）

- ① 異表記を吟味して対応

例：すべて・おく

- ② MeCabの「クセ」に合わせて語彙表への追加登録

例：の（名詞-非自立）

- ③ 用法の異なりによるレベルの異なりへの対応

例：「本発表」の「本」

⇒ 品詞ごとに対応

- ④ 語彙表の複合語の扱いへの対応（要協議）

例 1 <複合語 1> 衆議院・議員・衆議院/議員

例 2 <複合語 2> 終身/刑・終身/雇用・終身/雇用/制

例 3 <複合動詞> 飲み/込む・教え/込む

⇒これによって、学習語彙数を減らすことも可能

⇒必要があればMeCabの辞書に追加登録

謝辞

- 本研究の一部は、2014年年度科学研究費基盤(B) 課題番号24320096によるものである。
- 単語チェッカーの開発に関しては、甲南大学の北村達也氏の協力を得た。

ここに記して感謝の意を表する。

参考文献

- 川村よし子 (2009) 『チュウ太の虎の巻』 くろしお出版.
- 川村よし子 (2009) 「単語親密度と頻度情報を活用した文章の難易度判定」 『ヨーロッパ日本語教育』 Vol. 13, 132-139.
- 川村よし子・北村達也 (2008) 「文章の難易度判定のための単語親密度チェッカーの開発」 『日本語教育方法研究会誌』 Vol. 15 No. 2, 24-25.
- 北村達也・富岡洋介・川村よし子 (2009) 「IDFを用いた単語レベル判定システムの構築と検証」 『日本語教育方法研究会誌』 Vol. 16 No. 1, 52-53.
- 工藤拓 (2006) MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.net>

- 現在「単語チェッカー」は、甲南大学のサーバー上で公開しています。

<http://basil.is.konan-u.ac.jp/chuta/level/>

- 解析上の問題点やレベル判定の問題点等に気づいた方は、是非 kawamura@tiu.ac.jp にご連絡ください。



ご清聴ありがとうございました。