

国立国語研究所学術情報リポジトリ

日本語教師向けWebコンテンツ「つくば語彙チェッカー」の概要

メタデータ	言語: Japanese 出版者: 公開日: 2023-11-24 キーワード (Ja): キーワード (En): 作成者: 岩崎, 拓也, 波多野, 博顕, 伊藤, 秀明 メールアドレス: 所属:
URL	https://doi.org/10.15084/0002000179

日本語教師向け Web コンテンツ「つくば語彙チェッカー」の概要

岩崎 拓也 (筑波大学 人文社会系)
波多野 博顕 (筑波大学 人文社会系)
伊藤 秀明 (筑波大学 人文社会系)

Overview of the Web Content "Tsukuba Vocabulary Checker" for Japanese language teachers

Takuya Iwasaki (Faculty of Humanities and Social Sciences, University of Tsukuba)
Hiroaki Hatano (Faculty of Humanities and Social Sciences, University of Tsukuba)
Hideaki Ito (Faculty of Humanities and Social Sciences, University of Tsukuba)

要旨

本発表では、現在構築中の「つくば語彙チェッカー」についての紹介を行う。「つくば語彙チェッカー」は、web上で形態素解析を実行し、「リーディング・チュウ太」「日本語教育語彙表 Ver1.0」「日本語文法項目用例文データベース『はごろも』ver.3」「EDR 電子化辞書」といったデータベースをもとに、入力したテキストの頻度や語彙レベルを一括して確認することができる日本語教師向けのコンテンツである。また、この「つくば語彙チェッカー」は、テキスト中の任意の箇所に空欄補充問題などを作成することが可能な機能も搭載している。本発表では、これらの紹介に加えて、「つくば語彙チェッカー」の使いやすさ向上のためにどのような改修を行っているか、今後どのような機能を実装予定なのかについても紹介する。

1. はじめに

筑波大学グローバルコミュニケーション日本語・日本事情遠隔教育拠点（以下、日本語・日本事情遠隔教育拠点）は、文部科学省による教育関係共同利用拠点の認定拠点として、2010年より日本語教育に関するコンテンツ及びリソースの開発・運用・情報発信を行っている。本発表で紹介する「つくば語彙チェッカー」は、日本語・日本事情遠隔教育拠点が2021年3月より「リーディング・チュウ太」（元東京国際大学 川村よし子氏を中心に開発）の運用を引き継いだことを契機として開発が始まった。

「リーディング・チュウ太」は、1990年代末の公開から25年ほどが経つコンテンツである。入力テキスト中の単語と漢字を旧日本語能力試験の出題基準に合わせてレベル判定を行う機能や、単語の意味を多言語で提示する辞書機能など、その有益さから現在では「日本語教育のデファクトスタンダードの1つ」（北村2013）として、日本語教育機関のみならず地方行政団体など多くの機関で利用されている。また、現在では「日本語教育語彙表」や「機能語用例データベース『はごろも』」など、日本語教育用に構築された他のアノテーション・リソースに基づくwebコンテンツも公開されており、多面的なアプローチで電子テキストの解析が可能な状態になっている。

一方、上記の各コンテンツは、開発・公開時の経緯等から、当然ながらそれぞれが独立したものとして公開されている。そのため、現状では同じテキストを複数のアノテーション・リソースにあたって一度に解析することができない。また、操作方法や出力結果の表示とい

ったユーザビリティの面でも異なっているため、ユーザーはそれぞれのコンテンツの特徴を理解した上で利用する必要がある。また、伊藤・山田（2021）が日本語教師および日本語学習者を対象に「リーディング・チュウ太」のユーザビリティ調査を行った結果、シンプルな作りであってもどのように使うのかということの理解や、コンテンツ独特の用語（宝箱、道具箱など）の使い方がユーザーの負担を高めることを示している。このことから、現代のユーザーを意識したユーザビリティの改善も重要である。

以上の背景から、日本語・日本事情遠隔教育拠点では、川村よし子氏から「リーディング・チュウ太」、日本語学習辞書支援グループ（2015）から「日本語教育語彙表 ver1.0」、「はごろも」プロジェクト（2015）から「機能語用例データベース『はごろも』ver.3」の各データを提供いただくことで、様々なデータベースを一括して検索できるコンテンツ「つくば語彙チェッカー」の開発を進めている。多様なアノテーション・リソースを横断的にかつ簡便に利用できるツールの開発は、日本語教育関係だけでなく多方面の機関で有益なコンテンツとなるだろう。

2. 「つくば語彙チェッカー」の特徴

図1に「つくば語彙チェッカー」の現在のトップページのスクリーンショットを示す（なお、今回使用した図は全て現在開発中のコンテンツから引用したものである）。「つくば語彙チェッカー」では、左側にさまざまなタブがあり、そこから各機能に移動することができる。メインの機能としては、形態素解析をはじめとする語彙チェック機能と、問題・テスト作成機能の二つがある。以下では、これらの機能についての紹介を行う。

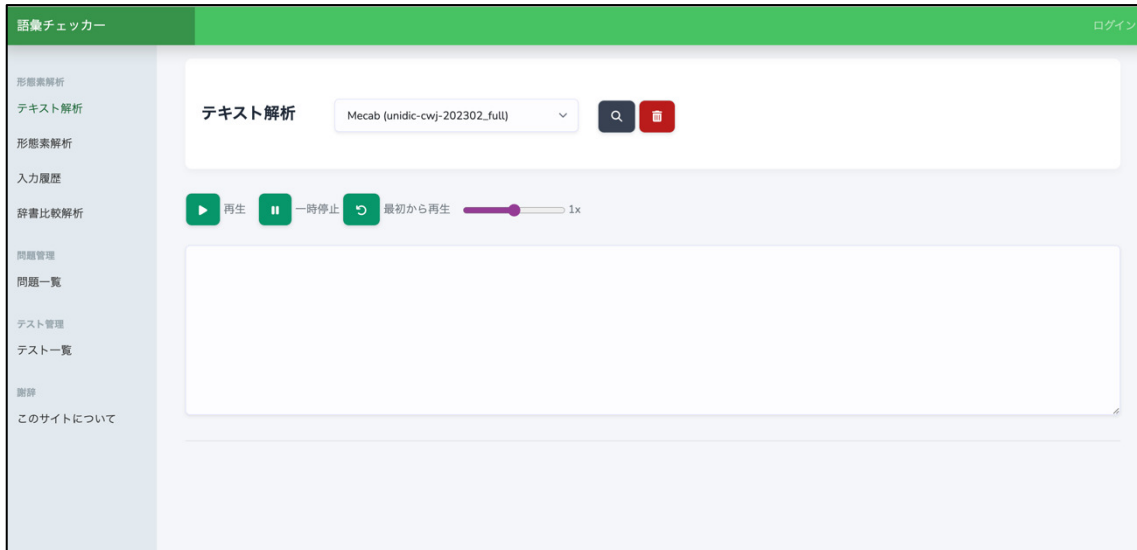


図1 「つくば語彙チェッカー」のトップ画面

2.1 語彙チェック機能

「形態素解析」という枠内には、「テキスト解析」「形態素解析」「入力履歴」「辞書比較解析」という四つのタブが存在する。まず、テキスト解析機能について説明する。

「テキスト解析」は、入力したテキストに形態素解析を施した上で「リーディング・チュウ太」「日本語教育語彙表 Ver1.0」「日本語文法項目用例文データベース『はごろも』ver.3」「EDR 電子化辞書」の各データベースに基づき、レベル判定や意味の参照を一括で行う機能である。この際、現在はレベル判定リストとの照合が形態素ベースになるため、派生語や

複合語の一部でオリジナル版と異なる結果となる。この不整合は今後の開発で改善していく予定である。

「テキスト解析」では、「MeCab+unidic-cwj-2023.02」と「Chasen+ipadic-2.7.0」という形態素解析器と形態素解析用辞書の組み合わせを選択して解析ができる。「Chasen+ipadic-2.7.0」を選択可能にしている理由は、もともと「リーディング・チュウ太」が「Chasen+ipadic-2.7.0」によって動作するものであったため、これに合わせた解析結果を出力できるようにするためである。そのため、通常は「MeCab+unidic-cwj-2023.02」による解析で問題ないと思われる。

テキスト解析を行った後の解析結果画面を図2に示す。この解析結果のページでは、入力したテキストの「リーディング・チュウ太」と「日本語教育語彙表 Ver1.0」における語彙レベルのクロス集計表が画面中央に表示される。また、その上には、各データベース名のタブがあり、これをクリックすることによって、画面右側に各レベルや頻度に属する形態素情報が表示される。必要に応じてチェックボックスをクリックして選択することで、まとめてCSV形式で出力結果をダウンロードすることも可能である。

それ以外にも、一括解析の結果を参照したうえで、テキスト中の語を適切なものに修正し、再度解析することが何度でも可能である。テキストから合成音声を作成する機能も備えているため、例えば、想定したレベルに合ったテキストが作成できた場合、それを音声素材として学習者に聞かせることも可能である。



図2 テキスト解析機能の画面

画面左側の機能一覧から移動できる形態素解析機能では、指定した解析器と辞書をもとに形態素解析を実行し、CSV形式で結果をダウンロードできるようになっている(図3)。解析結果の詳細をすぐに確認したい場合や、テキストから属性情報(品詞・語種など)別でリストを作りたい場合などに活用できる。



図3 形態素解析機能の画面

そのほかの機能として、「辞書比較解析機能」も搭載されている。この機能は、解析された形態素との照合結果を各リソース（「リーディング・チュウ太」「日本語教育語彙表 Ver1.0」「日本語文法項目用例文データベース『はごろも』ver.3」「EDR 電子化辞書」）に基づいて表示するとともに、その品詞と頻度も併せて一つのリストとして表示するものである（図4）。各形態素がテキストに出現した順で上から表示され、品詞や語彙のレベルを横断的に可視化することに特化している。

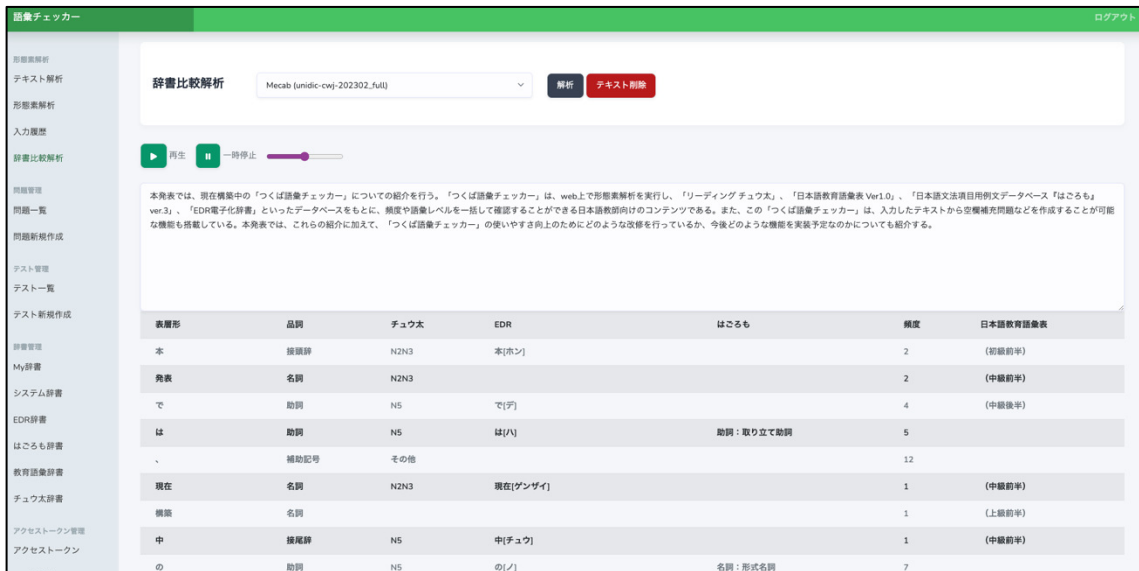


図4 辞書比較解析機能の画面

2.2 穴埋め問題作成機能

この「つくば語彙チェッカー」の特徴的な機能の一つに、穴埋め問題作成機能が搭載されている点が挙げられる。任意のテキストを、テキスト解析機能で形態素解析をかけ、その結果を踏まえて穴埋め問題を作成することができる。テキスト解析の結果の画面（図3）の下には、図5のように穴埋め問題作成とテスト名の一覧のスペースが存在する。穴埋め問題

を作成するためには、穴埋めにしたいテキスト部分をマウスオーバーするだけで作成することができる。また、マウスオーバーの順番がどの順番であろうとも、テキストのはじめから（1）（2）のように自動で順番を揃えてくれるようになっている（図6）。

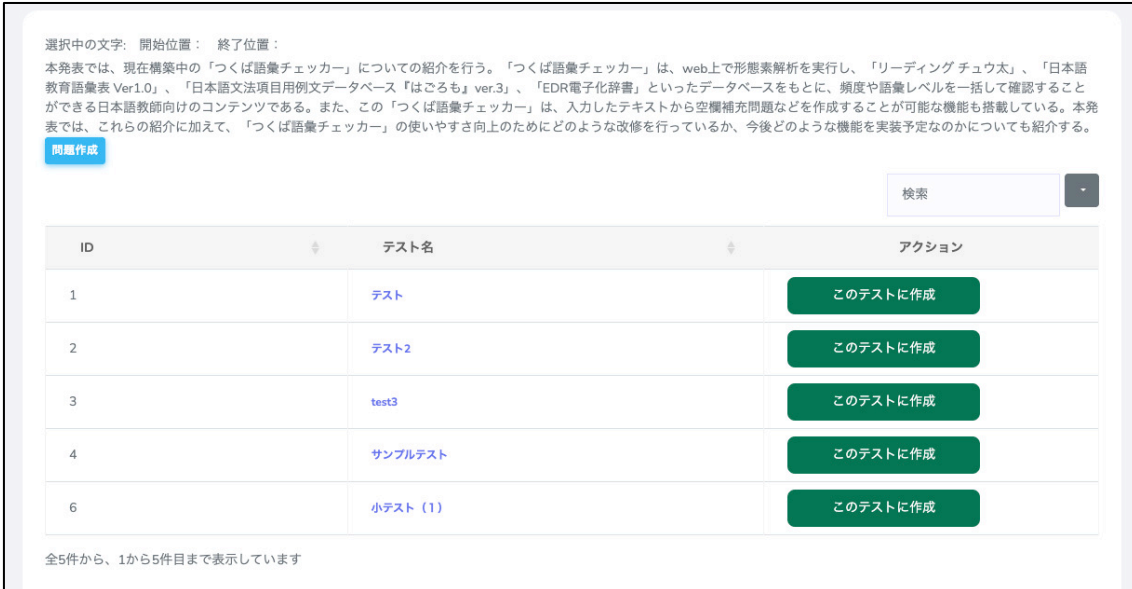


図5 穴埋め問題作成機能の画面



図6 穴埋め問題作成機能による問題の作成画面

こうして作成した穴埋め問題は、あらかじめ作成しておいたテストに組み込むことが可能となっている（このテスト作成機能については現在開発中である）。この方法以外にも、左側にある【問題作成】ボタンをクリックし、【問題詳細】ページに移動して、穴埋め問題の作成、検索用タグの付与、作成した問題／解答の印刷を行うこともできるように計画中である。なお、この機能については現在開発中であるため、今後の改修によって仕様が変更する可能性がある点に注意されたい。

3. 現在行っている「つくば語彙チェッカー」の改修について

このように、「つくば語彙チェッカー」では、主に日本語教師のテキスト解析補助のために平易な形態素解析機能をもととした日本語教育関連のデータベースの一括検索を可能とする機能を搭載し、穴埋め問題を中心としたテスト作成機能を有する web コンテンツである。当初は、図 1 のように簡易的なボタンとレイアウトによる UI デザインであり、開発者にしか取り扱いができない、癖のあるデザインとなっていた。そこで、想定されるユーザーである日本語教師にとって使いやすくするために、2名の現役日本語教師と日本語教育を専攻とする大学院生に協力してもらい、使用感を確かめてもらったうえで、感想を提出してもらい、改修を試みている。

現在までには、「リーディング・チュウ太」の解析結果と同じように、解析結果のクロス集計表をデフォルトで表示されるように改修を行った（日本語教育語彙表の解析結果も併せて表示するように変更）。また、検索結果からデータベースごとの結果へのアクセスは、当初は右上に配置されていたが、分かりにくいとのことだったため、画面中央へ配置変更した（図 7）。今回の改修によって追加された機能は、すでに紹介したとおり、一度入力したテキストを再度利用できるように入力履歴のタブの作成、また、入力履歴を検索できるウィンドウの追加である。また、すでに紹介した辞書比較解析機能も新たに追加した機能の一つである。今後は、解析結果を一括して CSV でダウンロードする機能などの機能の追加に加え、テスト作成機能など新機能の搭載も予定されている。



図 7 「つくば語彙チェッカー」のトップページ（改修前後の比較）

また、先述の通り、現状では形態素解析の結果をそのままレベル判定用のリソースと照合させている。そのため、オリジナル版と一部で結果の不整合が起きている。例えば、派生語の「高さ」は、解析結果のままでは「高」と「さ」に分けられるため、各形態素について判定リストと照合してしまう（細かく切りすぎてしまう問題）。派生語・複合語が判定リストにあった場合、それらを一語として処理しレベル判定を行う必要があるため、今後はこのような接続に係る処理を進めていく予定である。

4. まとめ：「つくば語彙チェッカー」公開に向けて

以上、本発表では、現在構築中の「つくば語彙チェッカー」の特徴を紹介し、現在行っている改修についても報告した。この「つくば語彙チェッカー」の本公開は 2023 年度末を予定している。なお、今後は「リーディング・チュウ太」が寺・北村・落水（1996）の読解支援システム「dictlinker」を発展させて開発する際に、機能の充実だけではなく、「使ってもらうための雰囲気作りも重要であることを学んだ」（北村 2013）と述べているように、今回の発表に間に合わなかったテスト機能の追加などの機能の改修を行いつつ、UI 改修についても予定をしている。これらの機能が完成し、日本語教師をはじめとした日本語教育関係者に使用してもらうことで、これまでに日本語教育界に培われてきた言語資源リソースを発展させていきたい。

謝 辞

本研究は川村よし子氏、日本語学習辞書支援グループ、「はごろも」プロジェクトの長年の研究およびデータのご提供により成り立っております。この場を借りて深く御礼申し上げます。

文 献

- 伊藤秀明・山田智久 (2021). 「日本語学習デジタルリソースのコストパフォーマンスについての検討」『筑波大学シンポジウム 日本語語彙辞書を利用した新たな研究』2021年11月13日 口頭発表.
- 北村達也 (2013). 「日本語読解学習支援システム「リーディング・チュウ太」」『甲南大学紀要 知能情報学編』6:2, pp.243-253.(https://www.konan-u.ac.jp/hp/kitlab/pub/konan_kiyou2013.pdf よりダウンロード可能).
- 寺朱美・北村達也・落水浩一郎 (1996). 「WWW ブラウザを利用した日本語読解支援システム」『日本科学教育学会 年会論文集』20, pp.103-104.(https://doi.org/10.14935/jssep.20.0_103 よりダウンロード可能).
- 日本語学習辞書支援グループ (2015). 『日本語教育語彙表 ver.1』,<http://jhlee.sakura.ne.jp/JEV/> (2023年7月27日確認).
- 「はごろも」プロジェクト (2015). 『機能語用例データベース「はごろも」 ver.3』,<http://hgrm.jpn.org> (2023年7月27日確認).