

日本語語彙特性反応型あなうめテスト作成器 JACKET の開発と応用

松下達彦（東京大学）、得丸智子（開智国際大学）、
劉瑞利（中山大學）、中島明則（フリーランス）

Development of a Lexical Trait Responsive Cloze Test Creator (JACKET) and its Application

Tatsuhiko MATSUSHITA, University of Tokyo

Satoko TOKUMARU, Kaichi International University Ruili LIU, Sun Yat-sen University

Akinori NAKAJIMA, Freelancer

要旨：本研究では、語彙データベースを応用し、特定の語彙特性を持つ語や文字などについて穴をあけるウェブツールを試作した。ウィンドーにテキストを貼り付け、条件指定して作成ボタンを押すだけで作成でき、ダウンロード、加工もできる。区切り単位として、語/文字/節/文を指定でき、語の選択では、頻度レベル、品詞、語種、語彙ジャンル、穴あけ間隔などを指定できる。文字の場合は、文字種、頻度レベル、穴あけ間隔などを指定できる。また、語/文字/節/文に共通の指定項目として、フリガナの有無、穴にした部分を選択肢として残すかどうかを指定できる。語彙力や文法力の評価、自律学習や授業への応用、語彙リスト作成、未知語の意味推測への研究など、多方面での応用が期待される。

キーワード：語彙特性、穴埋め、クローズテスト、テスト自動生成、自律学習

1. はじめに

テストは学習者の知識や学習内容の理解度の評価に不可欠だが、テスト項目の作成は知識と時間を必要とし、教師の大きな負担である。テストの自動生成は教師の労力の節約に役立つだけでなく、学習者の自律学習の助けにもなる。特に、e-learning システムの大規模化に伴い、近年、テスト自動生成システムに関する研究が注目を集める（Kurdi, et al., 2020）が、日本語教育においては、テスト自動生成システムに関する研究も開発もまだ少ない。そこで、本研究では、日本語の各種の語彙特性に応じて穴埋めテストを自動的に作成するツールである Japanese Cloze Key Test (JACKET) を開発した。以下、JACKET の開発とその利用法について説明し、日本語教育の現場や研究での応用可能性について述べる。

2. 主な先行研究

2.1 穴埋めテスト

穴埋めテストとは、出題文の一部を空欄（穴）にして内容を解答させる問題形式である。代表的な穴埋めテストとして、クローズテスト（cloze test）（Taylor, 1953）や C テスト（C-test）がある。クローズテストには、規則的に n 番目の語（通常 5~10 語ごと）を削除

する機械的クローズテストと、語の機能などを考慮しながら削除する意図的クローズテストの 2 種類がある (Tremblay, 2011)。採点方法についても、正答を削除された語だけに限定する正語法と文脈上問題がなければ正答とする適語法がある。C テストは、クローズテストを改善したテストであり、文章の 2 番目の文から、1 語おきに単語の後半を機械的に削除することで作成し、正答は 1 つだけに限定できる (Klein-Braley & Raatz, 1984)。

クローズテストは第二言語 (L2) 学習者の L2 能力を測るテストとしてよく使われている (Aitken, 1977 ; Oller, 1973 ; Tremblay, 2011 など)。クローズテストがどのような能力を測定しているのかは明確ではないが、L2 能力を測定するうえで妥当性、信頼性のあるテストとして認められている (Aitken, 1977)。さらに、Tremblay (2011) は研究目的を実現するための実用的なツールでもあるとも論じている。日本語教育においても、クローズテストはブレースメントテストとして妥当性と信頼性の高いテストであることが確認されている (小川, 1993 など)。

一方、クローズテストの難易度は文章の難易度以外に、対象とする語の種類 (内容語か機能語か)、文法項目など様々な要素に影響される (長, 2008)。対象項目の種類を指定できるテストが簡単に自動作成できれば、項目の影響を研究できるであろう。

2.2 テスト自動生成システムの開発

テキストから自動的に問題文を生成する研究が、本格的に進められるようになったのは最近 20 年ほどである。IT 技術の発達に伴い、学習効果の評価や自律学習、自己評価などを目的としたテスト自動生成システムの開発が盛んになっている (Kurdi et al., 2020)。そのうち言語教育向けの開発は主に読解力や文法・語彙知識の測定に焦点を当てて行われている。英語読解問題の自動生成を試みた Huang & He (2016) や、英語または中国語の語彙知識を測るための多肢選択式問題の自動生成システムを開発した Susanti et al. (2017)、Liu et al. (2018) などが挙げられる。ただし、これらはウェブ上で公開されていない。

日本語教育では、テスト自動生成システムの構築に関する研究もシステム開発も英語などに比べ非常に少なく、課題も多い。例えば、豊田ほか (2019) は、ウェブ上で収集したデータから、漢字の読み、変換などの問題の自動生成を試みたが、語彙レベルや問題の難易度の設定ができないことで実用化には多くの課題が残されていると述べている。

3. JACKET の開発

先行研究では、日本語の穴埋めテストにおいて、対象項目の語彙レベルや語彙の種類 (機能語か内容語かなど) によってテストの妥当性や信頼性に異なる影響があるが、それらが明らかでないことがわかった。そこで、本研究では、語彙や文字のデータベースを組み込み、特定の語彙特性を持つ語や文字などの条件が指定できる穴埋めテスト自動生成システム (ウェブツール) JACKET を開発した。

3.1 JACKET の機能

システムには 1)穴開け問題を作成する機能、2)作成した問題を編集する機能、3)作成した問題に回答する機能、4)画面から穴開け箇所の問題を判定し正解を表示する機能がある。

1)ではウィンドーに任意のテキストを貼り付ける。システムは入力された文章を解析し、文字、語、節、文の 4 つについて指定条件で文章中に穴開けをした問題を作成し、2)で作成済みの問題を手動で編集、保存できる。編集時に原文と問題（穴開け箇所を■に置換したもの）をテキストファイルでダウンロードできる。保存や編集が不要であれば、2)を飛ばしてウェブ上で 3)に進み、作成済みの問題呼び出し、問題を解くことができる。そして 4)で正語法で採点し、正解を表示できる。正解をデータベースから取得して判定し、正答数と正答率を表示するとともに、正答の入力欄を緑、誤答の欄をピンクで着色する。不正解の問題については、正解はすぐには表示されず、次に「再挑戦」ボタンを押すと不正解の問題に再解答できる。「再挑戦」せずに、「正解」ボタンを押せば正解が表示される。

1)で指定できる条件は以下の通りである。まず、区切り単位として、語/文字/節/文から選択する。語は UniDic の短単位であり、接辞も含んでいる。節は句点（。）または読点（、）で区切られている単位であり、文は句点で区切られている単位である。

語（接辞を含む）の選択では語彙頻度レベル（1000 位単位、1K～20K, 21K+）⁽¹⁾、旧日本語能力試験語彙レベル、品詞、語種（和語、漢語等）、ジャンル（学術共通語彙、文系共通学術語彙、理系共通学術語彙、文理混合学術語彙、文芸語彙、その他）を指定できる。

文字を単位とする場合は、文字種（ひらがな/カタカナ/漢字/その他）、漢字頻度レベル（100 位単位、1C～20C, 21C+）、旧日本語能力試験漢字レベル、を指定できる。

また、語/文字/節/文に共通の指定項目として、フリガナの有無（本文、および穴埋め問題）、選択肢条件（穴にした部分を選択肢として残すかどうか）がある。語または文字を指定した場合は、穴あけ間隔（0～99、[0]を指定すると条件に合うものをすべて指定、デフォルトは[3]）、句点前後の穴あけ（個別の条件に加えて、句点の前または後を指定して穴あけ）を指定できる。節、文を区切り単位として指定した場合は指定できるのは共通項目の指定のみである。

3.2 JACKET のシステム概要

3.2.1 コンピュータ資源

システムは、以下の共通環境を用いた

The screenshot shows the 'JACKET Creator for Teachers and Researchers of Japanese' interface. At the top, it says '日本語あなうめテスト Japanese Cloze Key Test Ver0.01 作成サイト (教員・研究者用)'. Below this, there's a large text area for input. To the right of the text area are buttons for '作成' (Create) and 'クリア' (Clear). Below the text area, there are several sections for specifying conditions: '語' (Word) with a dropdown for frequency level, '品詞' (Part of speech), '語種' (Language type), and 'ジャンル' (Genre). There are also checkboxes for 'フリガナ' (Romaji) and '穴あけ間隔' (Gap interval). At the bottom, there are buttons for '作成' (Create) and 'クリア' (Clear), and a section for '問題番号' (Question number) and '問題編集' (Edit question).

図 1 JACKET の初期画面（実際はカラー）

WEB システムとして実装している。

開発言語：Python2.7 および JavaScript、DB：MYSQL 5.0、WEB フレームワーク：Flask、O/R マッパー：SQLAlchemy、WEB フォーム拡張：WTForms、JavaScript ライブラリ：jQuery、CSS フレームワーク：Bootstrap。

3.2.2 文書処理ツール等

システムで利用している文書処理ツール等は以下のとおりである。

形態素解析エンジン：MeCab 0.98、日本語辞書：UniDic 1.311 (MeCab 用)、ベースワード語彙辞書：「日本語を読むための語彙データベース (VDRJ) Ver.1.1」(松下, 2011) (語彙素毎の頻度、レベル、語種、ジャンル、品詞などの情報を持つ)、ベースワード文字辞書：「現代日本語文字データベース (CDJ) Ver.2.0」(松下, 2014) (文字毎の頻度、文字種などの情報を持つ)。

3.2.3 システムで使用している DB テーブル

語彙辞書、文字辞書、問題 (作成済みの問題と正解) を格納する。

4. JACKET の応用

4.1 教育への応用

まず、A)学習者自身が例えば電子テキストの読解に取り組み、それをさらに自分で貼り付けて穴埋めの練習をすることができる。文法学習であれば助詞や助動詞を中心に抜く、外来語や学術語彙などの特定の属性に絞って抜くなど、節や文のレベルで抜いて文脈の流れと節や文の関係を考えるなど、様々な工夫が可能である。B)教師はこれを授業内に組み込んで実施することができるし、文法力や語彙力の評価にも使える。また、C)選択肢を残すという指定をすれば、穴埋めだけではなく、テキストに含まれる特定の語彙を抽出する作業にも使えるので、語彙リストの作成にも役立つ。

A)は学習者の自律学習ツールとして使うということである。特に文字や語彙の学習は、学習者自身の教室外での学習に任せやすいが、語や文字は単独で存在するのではなく、多くの場合、文章や会話の中で使われるため、「文章全体の意味や文脈の流れの中で適切な語や文型が定まるプロセスを、学習者が繰り返し体験すること」が重要だと思われる⁽²⁾。そのためには穴埋めテストの形式での反復学習が有効だと考えられる。

B)については、開発者の一人が授業で試用したが、すでにいくつかの利点や改善点が見えてきている。教師がテストを作成するのと同じツールで学習者が練習問題を作成できる点は学習者の動機づけを高める上で有効なようである。また、即座に自動採点されゲーム感覚で手軽に学習できることは、結果的に学習時間を増加させ、学習を促進するように思われる。また、本文ふりがな機能は学習者レベルに応じて学習項目以外の部分をわかりやすくできるため、便利である。一方、印刷して教室で使用する場合、選択肢の番号で解答できれば、漢字を書く能力に影響されない結果が得られるので、選択肢に自動的に番号を付す機能が備わっていれば便利である。試用を繰り返してフィードバックを得て、条件設

定の方法や画面を学習者にとってわかりやすいものにしていくことは今後の課題である。

C)については、現在、任意のテキストからレベルに合う語彙を抜き出せるツールとして、「リーディングチュー太」(Kawamura et al., 1997)、「読解基本語彙チェッカ」(本田, n.d., 2019)、「J-LEX」(菅長・松下, 2013)、「jReadability 日本語文章難易度判別システム」(Lee & Hasebe, 2013)があるが、いずれも重要度や頻度レベルを指定できるだけで、品詞や語種や語彙ジャンルの指定はできない。その点、JACKET はさまざまな語彙特性に応じて任意のテキストから条件に合う語彙を取り出せるので、語彙リスト作成に活用できる。

4.2 研究への応用

研究面では、まずは語彙力や文法力の評価ツールとして機能するであろう。2.1 で述べたように、L2 習得研究においても有効なツールである。穴あけの間隔や語彙特性によって、どの程度、難易度が変わるかを調べることでそのものが、文脈から未知語の意味などを推測する能力の研究となる。そのためのツールとしての応用も期待される。また、選択肢を残すか残さないかで正答率が変わるが、その程度がどの程度なのかについても妥当性・信頼性の検証のために調べる必要がある。また、空欄を文字単位にするか語単位にするかもテスト研究としては興味深い。このように様々な条件を変えることで、意味推測や語の産出に働く要素を明らかにでき、同時に、このテストの妥当性検証ができると思われる。

5. おわりに

本稿では、日本語語彙特性反応型あなうめテスト作成器 JACKET につき、先行研究、機能やシステム概要について説明し、教育や研究への応用可能性と現時点で判明している改善点などについて述べた。今後も試用を続けて改善したい。

付記：本研究は科研費（課題番号 25284096、19H01270、研究代表者：渡部倫子）の助成を受けた。本システムは 2021 年 5 月現在 <http://160.16.101.253/test/fibt/> で公開している。

注

- (1) K は英語で 1000 を表し、例えば 5K は 5 番目の 1000 語（4001～5000 位）を示す。
- (2) 開発者はこの点につき、「暗在性哲学（Philosophy of the Implicit）」(Gendlin, 1991) に基づく言語学習法の一つの具現化であるとも考えている。

参考文献

- 小川多恵子. 1993. 「プレースメントとしてのクローズテスト」『筑波大学留学生センター 日本語教育論集』8: 201-213.
- 長加奈子. 2008. 「Cloze Test の難易度に影響を与える要因について」『福岡女学院大学短期大学部紀要・英語英文学』44: 33-49.

- 菅長洋一・松下達彦. 2013 「日本語テキスト語彙分析器 J-LEX」
<http://www17408ui.sakura.ne.jp/index.html>
- 豊田哲也・島田めぐみ・保坂敏子. 2019 「Web テキストを用いた日本語学習問題自動生成システムの構築」『東アジア日本語教育・日本文化研究』 22: 137-148.
- 本田ゆかり. 2019 「コーパスに基づく「読解基本語彙 1 万語」の選定」『日本語教育』 172: 118-133.
- 本田ゆかり. n.d. 「読解基本語彙チェッカ」 <http://yukari.overworks.jp/>
- 松下達彦. 2011 「日本語を読むための語彙データベース (VDRJ) Ver.1.1 (研究用)」
<http://www17408ui.sakura.ne.jp/tatsum/database.html#vdrj>
- 松下達彦. 2014 「現代日本語文字データベース (CDJ) Ver.2.0 (研究用)」
<http://www17408ui.sakura.ne.jp/tatsum/database.html#cdj>
- Aitken, K. G. 1977. Using cloze procedure as an overall language proficiency test. *TESOL Quarterly*, 11(1): 59-67.
- Gendlin, E. T. 1991. *Thinking beyond patterns: Body, language, and situations*. P. Lang.
- Huang, Y., & He, L. 2016. Automatic generation of short answer questions for reading comprehension assessment. *Natural Language Engineering*, 22(3): 457-489.
- Kawamura, Y., Kitamura, T., & Hobara, R. 1997. 「日本語読解学習支援システム リーディンググチュウ太」 <https://chuta.cegloc.tsukuba.ac.jp/>
- Klein-Braley, C., & Raatz, U. 1984. A survey of research on the C-Test. *Language testing*, 1(2): 134-146.
- Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. 2020. A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, 30(1): 121-204.
- Lee, J. & Hasebe, Y. 2013. jReadability 日本語文章難易度判別システム.
<https://jreadability.net/sys/>
- Liu, M., Rus, V., & Liu, L. 2018. Automatic Chinese multiple choice question generation using mixed similarity strategy. *IEEE Transactions on Learning Technologies*, 11(2): 193-202.
- Oller, J. W., Jr. 1973. Cloze tests of second language proficiency and what they measure. *Language Learning*, 23(1): 105-118.
- Susanti, Y., Tokunaga, T., Nishikawa, H., & Obari, H. 2017. Controlling item difficulty for automatic vocabulary question generation. *Research and Practice in Technology Enhanced Learning*, 12(25): 1-16.
- Taylor, W. L. 1953. "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 30(4): 415-433.
- Tremblay, A. 2011. Proficiency assessment standards in second language acquisition research. *Studies in Second Language Acquisition*, 33(3): 339-372.