The University of Osaka
Institutional Knowledge Archive

Title	レベル別ベトナム語教材作成支援システムの構築
Author(s)	清水, 政明; グエン, ティ ゴック トー
Citation	外国語教育のフロンティア. 2018, 1, p. 55-62
Version Type	VoR
URL	https://doi.org/10.18910/69778
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

レベル別ベトナム語教材作成支援システムの構築

Construction of a Web-Based System that can Evaluate the Level of Vietnamese Words and Sentences in Teaching Materials

清水 政明、グエン・ティ・ゴック・トー

Abstract

This study aims to aid in the development of a web-based system that can evaluate the level of words in Vietnamese sentences and will be used for teaching and learning Vietnamese. The system will show users the percentage of words containing each level of the Vietnamese vocabulary. The person who develops of teaching materials can use this system to check the level of words that are contained in a certain text and can arrange them to make the text more appropriate for the learners' proficiency. To divide the vocabulary into 3 levels is based on teaching materials that are published in Vietnam and other countries. However, the user evaluations of the system allow us to reconsider the 3-level program to make the system more useful for a wide range of learners.

キーワード：ベトナム語、教材作成、レベル判定、レベル別語彙

1. はじめに

本稿では、学習者個々人の習得度に応じたレベル別の教材を提供する上で有用と思われるウェブベースのツールを構築するための基礎的研究を行う。類似のシステムは日本語教育の分野等で既に実現されているが[1)]、ベトナム語のような日本において教育機会の比較的少ない言語（Less Commonly Taught Languages、以下LCTL）を対象としたシステムは管見の限りまだ存在せず、それを実現するにはベトナム語に対応可能な自然言語処理の技術が必須である。従来、我が国においては日本語や英語等多くの学習者の抱える言語を対象とする自然言語処理分野の研究は比較的進んでおり、それに基づく各種学習支援システムは、いくつか例を見出すことが可能である。一方、ベトナム本国に目を移すと、例えばハノイ工科大学等を中心とする研究機関やIT系企業において、ベトナム語を対象とした自然言語処理技術が急速な進歩を遂げており、それらの研究者と協力することにより本稿で提案するシステムを実現させることは極めて大きな意義を有するものと考える。以下、システム構築の手順と開発、実装、評価の概要を紹介する。

2. レベル別学習語彙の選定

本稿で提案するシステムを実現するには、まず基礎データとなるレベル別語彙リストを作成する必要がある。当該リストについては、筆者自身が作成した既存のものを利用することが可能であるが[2)]、今回のシステム構築に際してはその精度を上げる必要がある。既存のものは、国内外の各種教科書、語彙リスト、文法書等を参考に作成したものであるが、日本人学習者を対象に作成されたとはいえ、例えば漢語系語彙の扱い等、依然多くの問題を残すものである3)。また、昨今喫緊の課題である日本在住のベトナムにルーツを持つ児童生徒への母語教育への対応も重要な課題であり、多様な学習者の言語能力測定に応用可能なものである必要がある。例えば、ベトナム教育・訓練省による「海外在住ベトナム人のためのベトナム語教育プログラムの公布」（教育・訓練省決定：26/2006/QD-BGDDT）に示されたレベルと評価枠を参考に、ベトナム国内の同分野の研究動向にも目を向けつつ作業を進める必要がある。また、海外在住ベトナム人に対する母語維持の問題に関しては、カナダ、オーストラリアを筆頭にいくつかの国で既に長い実績の歴史があり、例えばオーストラリアのVictorian Curriculum and Assessment Authorityによるベトナム語の教育、評価の手法[4)]等は高い質を有するもので参考に値する。それらを参考にしつつ、従来のリストと同様レベルをA, B, Cの3段階に分け、文のレベルを評価する基礎とする必要がある。そこでこれまで同様にベトナム本国の語学能力評価基準に関する政府機関（特に教育・訓練省）及び研究機関の文書、並びに各種研究成果を参照しつつ、より日本人に相応しいリストを作成するべく語彙の収集と各語彙のレベルについて検討し、その成果をテキストレベル判定システムに組み込んだ。

3. システム構築

以上の作業を経て作成されたレベル別語彙リストを利用して、文レベル判定システムのプロトタイプを開発した。ベトナム語を対象とした当該システムと類似のシステムとしてまず挙げられるのは日本語文の自動レベル判定システムである「リーディング･チュウ太」である。当該システムは、ユーザの入力した日本語の文章中に含まれる語彙の中に、日本語能力試験出題基準所収の級別語彙がどの程度含まれているかを瞬時に表示し、文全体のレベルを判定するシステムである。同時に、所収漢字に関してもレベル別に色分けして表示することが可能な仕様となっている。ベトナム語の正書法は、ラテン文字を基礎としつつ種々の補助記号を付加する「クオックグー（国語）」と呼ばれるもので、日本語のような多種の文字が混在することはなく漢字の問題も生じない点、日本語に比べてシステム構築は容易である。しかし、ラテン文字を基礎とするとはいえ複雑な補助記号を伴う正書法を処理するのは決して容易な問題ではなく、Unicodeをベースに文字を処理するための一定の知識と技術が必要である。よって、少なくとも日本国内でベトナム語を対象とした自

然言語処理技術の開発に本格的に取り組む個人・団体は管見の限り皆無に等しい。そのように限られた条件下でのシステム構築の手法について以下に説明する。

3.1 ベトナム語構文解析プログラムの利用

当該システムを実装するためには、まずレベル判定の対象となる入力された文を、単語毎に分割することが可能な構文解析（parsing）のプログラムが必要である。幸いベトナム語は孤立型言語に属し、語そのものが屈折変化することはないので、構文解析とは言いつつも、文の構造を正確にとらえることは今回のシステム構築においては必須ではなく、専ら語と語の切れ目が見極められれば当座の用は足りる。そこで、構文解析プログラムを構築するための大規模なベトナム語テキストコーパス及びレキシコンを有し、精度の高い構文解析技術を実現した研究室の協力を仰ぐ必要があり、筆者の知る限りではハノイ工科大学（Đại học Bách Khoa Hà Nội）及びその卒業生が活躍する企業が数少ない候補の一つとなった。そこで情報通信学部（Viện Công Nghệ Thông Tin và Truyền Thông）で日本語教育に携わるスタッフを通じて、自然言語処理を専門とするスタッフにアクセスし、最終的には彼ら自身が経営するIT系企業であるDatasection Vietnam（以下、DSV[5)]）に辿り着いた。DSVは独自開発の精度の高い構文解析プログラムを有しており、ハノイ工科大学で培われた豊富な知識とリソースが惜しみなく地道に実用化されていることを知って奮い立った。

3.2 レベル別語彙との同定

レベル別語彙リストと構文解析プログラムが揃ったら、後は解析の対象となる文を解析した結果とレベル別語彙リストの同定を自動的に行い、そのマッチした結果を計算するプログラムを作成する必要がある。今回はその語彙リスト同定、計算プログラムの構築、及びウェブ上でアクセス可能なプログラムの構築とサーバ管理を全てパッケージとしてDSVに依頼することとした。以上のプロセスを図示すると図1の通りとなる。

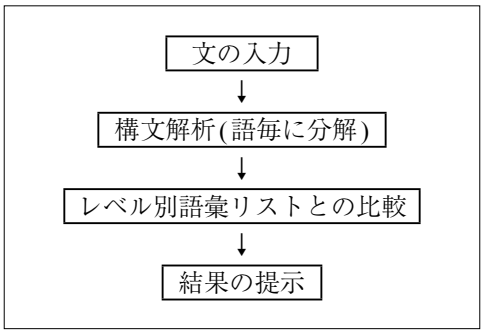


図1 システムの流れ

3.3 分析結果の提示

以上のプロセスを経て得られた結果を、最終的にユーザに分かりやすい形で提示する必要がある。また、本稿の提案するプログラムは広いユーザの利用に供することを目的としているので、ウェブ上での利用が可能である必要がある。それらを加味したものとして、図2に示すようなGUIを介してウェブ上でのアクセスが可能なプログラムが完成した。ユーザは「Here you can analyze text」のボックスに分析対象となるテキストを入力することとなる。そして、「Analyze」ボタンをクリックすると、図3の分析結果画面が表示される。現時点ではA, B, Cの3段階に色分けされた語彙が、文毎に分けて表示された各文中に示され、右のコラムに文章中に含まれる各レベルの語彙リストと各語彙の頻度、文全体の語彙に対する各レベルの語彙の割合が表示される。ユーザはこれに基づいて文のレベルを判定することになる。

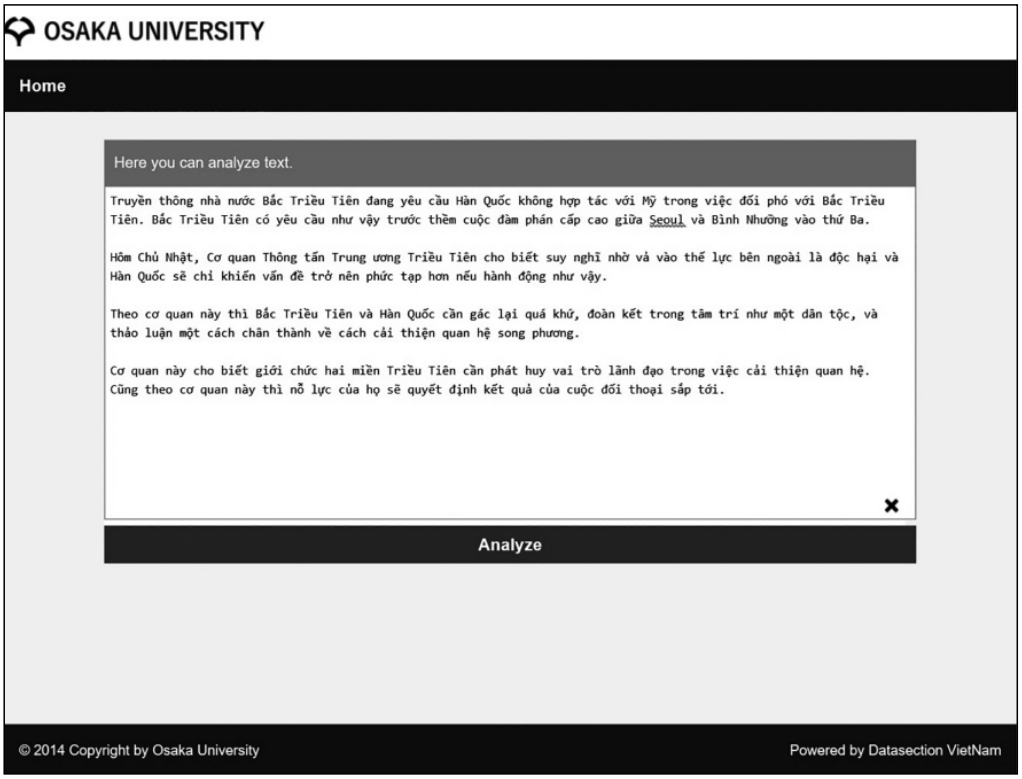


図2 文入力のイメージ

図3 分析結果の提示

4. システムの実装と評価

以上の過程を経て構築されたシステムがウェブ上で常時アクセスし利用することが可能となった。以下がその公開URLである。

ベトナム語文レベル判定システム公開URL http://osk-v1.datasection.com.vn/

システムの動作に関しては、これまで特に大きな問題は生じておらず、極めてスムーズな動作が確保されている。そこで日本国内でベトナム語教育に携わる教員、研究者に試用を依頼したところ、そこで出された改善案は、専ら収録語彙の選択とそのレベル分けの問題に集約された。現状のシステムでは、レベルA所収語彙数が2,694語、レベルBが3,578語、レベルCが732語という均整を欠いた単語セットとなっている。それぞれ初級、中級、上級という位置づけでレベル分けを行ったものであるが、特にレベルA初級語彙の語彙数が多いという声がしばしば聞かれたので、更にレベル分けを細分化し、入門、初級、中級、上級という4段階の語彙レベル分けを目下検討している。語彙セットが完成したら、図4に示すページから管理者自身が自由に所収語彙を調整することが可能である。

5. システムの利用法

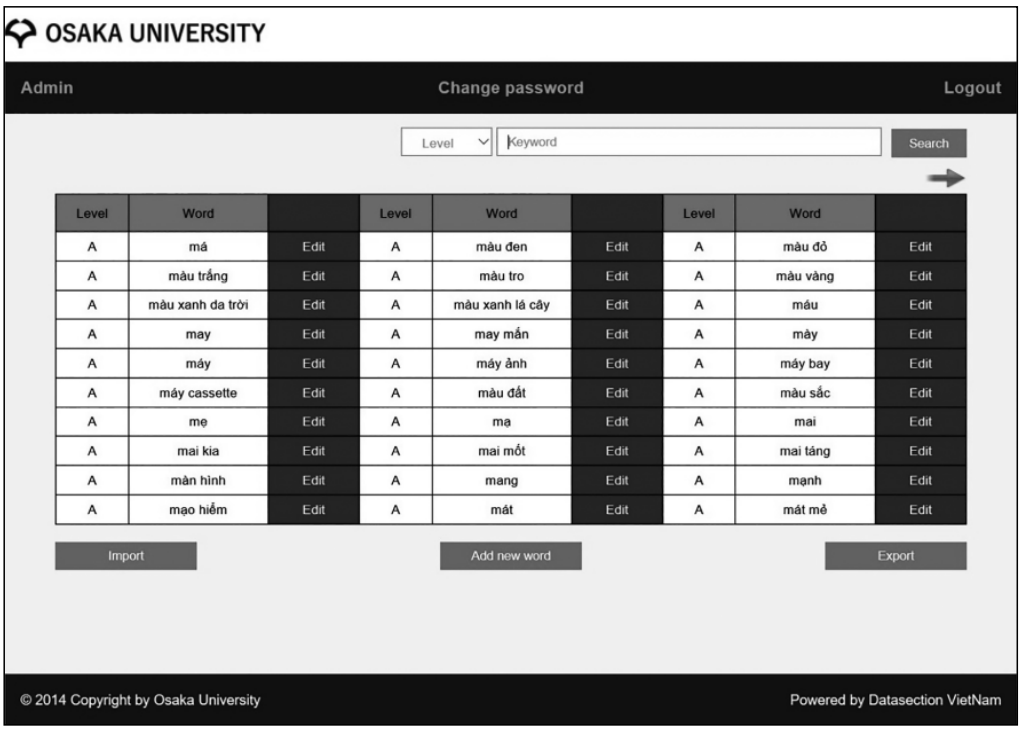
上記評価に基づいたシステム改良を進めると同時に、以下に本システムの考えられる用途を列挙してみたいと思う。

5.1 レベル別ベトナム語教材作成

当該システムを利用してまず考えられる産物は、レベル別教材の作成である。教師が独自に作成する文章だけでなく、既存の文章を利用した教材を作成する場合にも、当該教材の電子テキストを本システムで評価し単語のレベルを確認することにより、必要であれば語彙のレベルを調整し、各レベルに相応しい語彙に統一することが可能となり、各レベルに相応しいコンテンツの教材を作成することができる。

5.2 ベトナム語検定試験問題の作成

次いで、現在社会的ニーズが高まりつつあるベトナム語検定試験の問題作成にも利用することが可能である。特に検定試験の場合、その出題項目を厳格に規定することが求められるので、それを一々手作業で確認、調整するという極めて手間のかかる作業をある程度自動化することが可能となる訳である。



Level	Word		Level	Word		Level	Word	
A	má	Edit	A	màu đen	Edit	A	màu đỏ	Edit
A	màu trắng	Edit	A	màu tro	Edit	A	màu vàng	Edit
A	màu xanh da trời	Edit	A	màu xanh lá cây	Edit	A	máu	Edit
A	may	Edit	A	may mắn	Edit	A	mày	Edit
A	máy	Edit	A	máy ảnh	Edit	A	máy bay	Edit
A	máy cassette	Edit	A	màu đất	Edit	A	màu sắc	Edit
A	mẹ	Edit	A	mạ	Edit	A	mai	Edit
A	mai kia	Edit	A	mai mốt	Edit	A	mai táng	Edit
A	màn hình	Edit	A	mang	Edit	A	mạnh	Edit
A	mạo hiểm	Edit	A	mát	Edit	A	mát mẻ	Edit

図４ 収録語彙調整画面

5.3 学習者によるアウトプットのレベル判定

最後に、ベトナム語学習者による作文データを入力することにより、学習者自身のアウトプットがどのレベルにあるかを判定することが可能となる。日本語を母語とするベトナ

ム語学習者の産出文のレベルを判定することはもちろんであるが、その語彙セットを調整することにより、日本在住のベトナムにルーツを持つ子供たちの母語による産出能力を測ることも可能であることを付言しておく。

6. おわりに

本研究は、筆者等によるこれまでの研究成果を踏まえ、日本人ベトナム語学習者に有用な教材を作成、またその学習成果を評価するテスト等を作成する上で有用なベトナム語テキストのレベル判定システムを構築するための基礎的研究を行った。従来の研究により作成されたレベル別語彙リストを利用し、一部改変したリストのデータをシステムに組み込んだ上で、教材あるいはテストに利用する文章がどのレベルに属するものかを判定するシステムを実装することに成功した。国内では同様のシステムが日本語を対象に実現されているが、ベトナム語のようなLCTLでは実用的に構築されたものがまだ極めて稀である。そこで、ベトナム語を対象とした自然言語処理の専門家と共同で当該システムの構築を実現した。従来研究レベルでベトナム語構文解析プログラムの提案はあったものの、本システムに利用可能なレベルでの実用性の高いものが極めて稀であった。その技術と本稿の提案する発想との理想的な融合が本システムの実現を可能にしたものと自負する次第である。

謝辞

本研究は科研費基盤研究(C)「レベル別ベトナム語教材作成支援システム開発のための基礎的研究」（課題番号25370714、研究代表者清水政明、平成2013 ～ 2015年）によるものである。本システムの構築に当たり、構文解析プログラム、各種プログラムの作成及びサーバの利用をご快諾頂いたDatasection Vietnamに深謝申し上げる次第である。

注

1) 代表的な例は「リーディング・チュウ太」(http://language.tiu.ac.jp/)である。本稿が提案するシステムも 多くの部分をその着想に負うている。第3節で詳細を紹介する。
2) 当該語彙リスト及びレベル別文法項目リストが、科研費基盤研究(C)「日本人学習者を対象としたベトナム語検定試験制定のための基礎的研究」（課題番号22520562、代表清水政明、2010 ～ 2012年度）により作成された。
3) 参考にした教科書・語彙集の類は以下の通りである。Nguyễn Văn Huệ (chủ biên), 2008, *Giáo trình tiếng Việt cho người nước ngoài* (Tập 1, 2), Nxb. Đại học Quốc gia thành phố Hồ Chí Minh; Binh Nhu Ngo, 2010, *Continuing Vietnamese, Tuttle Publishing*; Robert B. Jones & Huynh Sanh Thong, 1960, *Introduction to Spoken Vietnamese*, American Council of Learned Societies; 上田広美編 (2003)『東南アジア大陸部言語調査票（カンボジア語, ラオス語, タイ語, ベトナム語）』東京外国語大学外国語学部。

4) http://www.vcaa.vic.edu.au/Pages/vce/studies/lote/vietnamese/vietnamindex.aspx#H2N1000 B
5) 企業の詳細については右サイトを参照されたい。（http://datasection.com.vn/）

参考文献

川村よしこ

2009 『チュウ太の虎の巻―日本語教育のためのインターネット活用術』くろしお出版、東京。

2012 「日本語読解学習システム『リーディング・チュウ太』の取り組み」『ウェブマガジン「留学交流」』19。

Carol Anne Spreen

2002 *New Technologies and Language Learning: Cases in the Less Commonly Taught Languages (Technical Report)*, University of Hawaii Press.