

# 学習リソースの推薦を目的としたアンダスタンダビリティ推定手法の検討

## Study of understandability estimation method for the purpose of recommendation of learning resources

藤崎 優理<sup>\*1</sup>, 三好 康夫<sup>\*2</sup>, 鈴木 一弘<sup>\*2</sup>, 塩田 研一<sup>\*2</sup>, 岡本 竜<sup>\*2</sup>

Yuri FUJISAKI<sup>\*1</sup>, Yasuo MIYOSHI<sup>\*2</sup>, Kazuhiro SUZUKI<sup>\*2</sup>, Ken-ichi SHIOTA<sup>\*2</sup>, Ryo OKAMOTO<sup>\*2</sup>

<sup>\*1</sup> 高知大学大学院総合人間自然科学研究科理学専攻

<sup>\*1</sup> Graduate School of Humanities and Social Sciences, Kochi University

<sup>\*2</sup> 高知大学理学部

<sup>\*2</sup> Faculty of Science, Kochi University

あらまし：我々はこれまで、学習リソースの推薦を目的として、リソースの難易度推定手法の開発を行ってきた。本研究では、リソースで扱われている内容が同等で難しさが同じであるにもかかわらず、読むのにあたり、とっつきやすさや分かりやすさに違いがあることに着目する。我々はこの違いをアンダスタンダビリティの差として捉え、推定手法について検討する。

キーワード：アンダスタンダビリティ, リーダビリティ, 数学, 難易度推定, 情報推薦

### 1. はじめに

我々はこれまで、学習リソースの推薦を目的として、リソースの難易度推定をリソースに書かれた内容を一切使用しないで行う手法の開発を行ってきた(1)。これにより、学習リソースにどの程度の難しい概念を含んでいるかという意味での難易度は、ある程度の精度で推定できるようになった。しかし、文章中に含まれる難解な概念が同じで難しさが同等であるはずの複数の学習リソースを読み比べてみると、内容がすぐ理解できる読みやすいものと、理解しにくく読みにくいものがある。例えば、微分積分の意味を解説する学習リソースがあった時、図を使わずに難解な数式と文章のみで説明しているものと、図やグラフを用いてコップと水や移動速度と距離の関係等の例を織り交ぜて説明しているものとは、後者の方が理解しやすいということである。また、後者の方が読むのにあたりとっつきやすく、初学者に対して学習リソースを推薦するのであれば、後者を薦めるべきである。

そこで我々はこの違いをアンダスタンダビリティの差として捉え、本研究は学習リソースのアンダスタンダビリティ推定を目的とする。本稿では、学習リソースとして数学に関する解説文を対象に、アンダスタンダビリティを推定する手法の検討を行う。まずは手始めに、文章のテキスト情報のみを用いた手法の検討から行うため、本稿では図、グラフ、数式の内容については取り扱わない。

### 2. リーダビリティとアンダスタンダビリティ

#### 2.1. リーダビリティ

読みやすさを対象とした研究においては、リーダビリティという尺度を用いることが多い。リーダビリティ推定に関する先行研究には、例えば柴崎ら(2)や李(3)のようなリーディング教材の推薦を目的としたものがある。李は日本語教育における読解教育のための日本語文章難易度判定システムをWeb上に公開しており、誰でもフォームに文章を入力するだけでリーダビリティ値を推定することができる。

多くの先行研究のリーダビリティの定義は「文章の読みやすさ」のように厳密に定めておらず、各々の目的に応じた指標に基づきリーダビリティを推定している。基本的にリーダビリティの指標には、語彙の難しさや構文の複雑さ、1文の情報量等が用いられるが、リーディング教材推薦を目的としたものは、使われている漢字の難しさや文字種(ひらがな、カタカナ、漢字)の割合等の指標が用いられる。

#### 2.2. アンダスタンダビリティ

我々は、文章中に含まれる難解な概念が同じで難しさが同等であるはずの複数の学習リソースの読みやすさの差を何らかの尺度で表したい。リーダビリティの先行研究のように語彙の難しさを指標として読みやすさを求めようとする、例えば微分積分の意味を解説する学習リソースには全て「微分」「積分」という高校2年生で習う数学用語が含まれてい

るということで、全て同等の読みやすさと推定されてしまうと考えられる。柴崎らは「語彙の難易を意味や概念で考えると、それは readability (読みやすさ) の問題ではなく、わかりやすさ (understandability) の問題になる」と述べて研究対象としていないが、我々の目的においては語彙の難易を意味や概念で考えなければならず、わかりやすさの問題を扱わなければならない。そこで本研究ではアンダスタンダビリティという尺度を定義し用いることとする。

本研究のアンダスタンダビリティの定義は、文章のわかりやすさとし、文章をどれほどわかりやすく解説しようとしているかを示すものとする。アンダスタンダビリティを推定する指標として、噛み砕き度ととっつきやすさを提案する。

### 3. アンダスタンダビリティ推定手法の検討

#### 3.1. 噛み砕き度

難しい概念を噛み砕いた表現に言い換えると、初心者にとってわかりやすくなる。そこでアンダスタンダビリティの指標として噛み砕き度を用いる。ここでは、ある概念をそれより少し易しい同分野の概念で表現したものを噛み砕き表現とし、文章中の最も難解な語彙に対する噛み砕き表現の噛み砕き具合を噛み砕き度とする。

数学に関する解説文を対象にした噛み砕き度を推定し数値化するアイデアとしては、数学用語の関係を表すネットワークを用いる方法を考案中である。図1は中学3年間に学ぶ単元の関係をネットワークにしたものである。このネットワークは、どの単元の知識を元に次の新しい単元を学ぶかを示したもの(ただし、これは説明のためのイメージ例であり、正確な関係を示したものではない)である。例えば、1

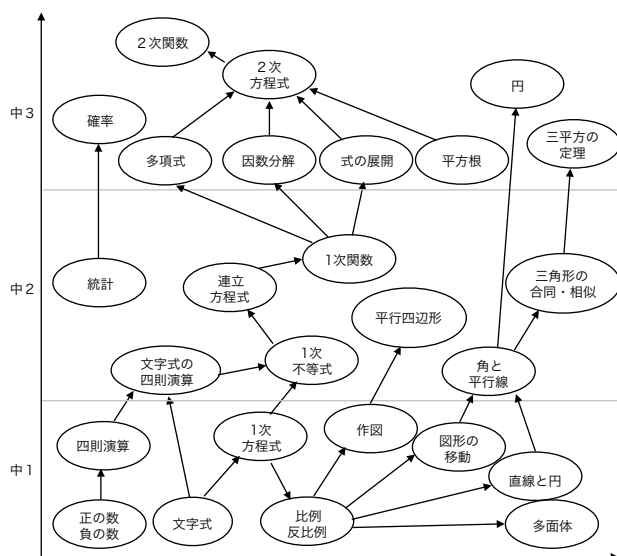


図1 中学数学単元ネットワークのイメージ

次不等式の内容を噛み砕いて解説するには1次方程式と文字式の四則演算の内容を用いて解説すればよいということを示している。従って、1次不等式の解説文中に、1次方程式や文字式の四則演算の解説文に出てくる特徴と同じ特徴が出てくると、噛み砕き度が高い解説文であると推定できるのではないかと考えている。

#### 3.2. とっつきやすさ

噛み砕き度では難しい概念をその分野の少し易しい概念を用いて説明することを評価しているが、たとえ話や事例を使って解説することで初心者にわかりやすくなることもある。さらに、身近な例で説明されると話に親しみやすくとっつきやすい。そこで、アンダスタンダビリティの指標としてとっつきやすさも用いる。とっつきやすさは、文章中のたとえ話や事例の交わり具合とする。

数学に関する解説文を対象にしたとっつきやすさを推定し数値化するアイデアとして、数学用語との関連の低い語彙に注目する。数学に関する解説文中に数学用語と無関係な語彙が使われている場合、比喩表現を用い身近なものを例とした解説であると判断し、とっつきやすさを推定できるのではないかと考えている。

### 4. おわりに

本稿では、文章中に含まれる難解な概念が同じで難しさが同等であるはずの複数の学習リソースの読みやすさの差を表すための尺度としてアンダスタンダビリティを定義し、推定手法の検討を行った。今後は提案した手法を用いて推定システムを実装し、評価を行いたい。

#### 謝辞

本研究の一部は、JSPS 科研費 25330364 の助成を受けた。

#### 参考文献

- (1) Miyoshi, Y., Suzuki, K., Shiota K. and Okamoto, R.: Evaluation of Difficulty Estimation for Learning Materials Recommendation, Proc. of ICCE2014, pp.71-76, 2014.
- (2) 柴崎秀子, 沢井康孝: 国語教科書コーパスを応用した日本語リーダビリティ構築のための基礎研究, 電子情報通信学会信学技報, NLC2007-32, pp.19-24 (2007)
- (3) 李在鎬: 日本語文章難易度判別システム, <http://jreadability.net/> (2016年6月現在) (n.d.)