

# 教科書を規準とする難易度推定

## Readability Assessment Based on Textbook Corpora

藤田 早苗      藤野 昭典      小林 哲生  
Sanae Fujita      Akinori Fujino      Tessei Kobayashi

NTT コミュニケーション科学基礎研究所  
NTT Communication Science Laboratories

To absorb knowledge from the books is very important for child development. Therefore, to recommend books of suitable readability, we aim to create readability measures from infants to elder children. We have proposed the method to create readability assessment based on picture books as criterial corpora targeting children of preschool age. In this paper, we employ textbook corpora as criterial corpora, we show that our proposed method is also effective for textbooks. In the future, we are planning to combine these readability measures and build a consistent readability measures from infants to elder children.

### 1. はじめに

絵本を含めた本からの言語情報は、子供の言語発達における重要なインプットである。しかし、子供にとって難しすぎる本は意欲をそぐ可能性があり、逆に簡単すぎても物足りない可能性がある。そのため、子供の読解力や言語能力に応じた適切な推薦が重要である。

そこで、我々は子供の発達や興味にあわせて適切な難易度の本やテキストを推薦することを目的とし、幼児期から学童期、あるいはそれ以上まで一貫した難易度推定を行えるようにすることを目指している [13]。[13] では、未就学児を対象とした絵本の方が教科書より難易度が低いと仮定し、幼児期から、学童期まで一貫した難易度測定を行うことを試みた。しかし、小学校低学年の教科書は必ずしも絵本より難しいとは限らないことがわかったため、まず絵本だけを規準データとして難易度を推定する方法を提案した [12]。その結果、提案手法は絵本では、 $R = 0.929$  という高い相関で難易度を推定できた。本稿では、教科書を規準データとして利用した場合にも提案手法が有効であることを示す。

何を規準として難易度を規定するかは、一般向けか [14]、外国人学習者向けか [4, 10]、子供向けか [7, 11] など、想定する対象によって異なると考えられる。また、同じ子供向けであっても、教科書を規準コーパスとするのか [7, 9, 11]、教科書以外の規準を考えるのか [1, 3, 6, 7]、など様々な方法が考えられる。しかし、教育への利用を考えた場合、学年や年齢に応じた難易度がわかることは重要であり、教科書は学年に対応しているという利点がある。また、教科書は国の方針にしたがって統制されているため、規準データとして適切だと考えられる。そこで、本稿では、学童期以上の子供を対象として想定し、教科書を規準データとして、学年に対応した難易度推定を行う。

### 2. 先行研究

教科書を基準とした難易度測定を行う研究には、佐藤ら [9, 14] と、柴崎・玉岡 [11] (以下、柴崎方法) の研究がある。

佐藤らは、小学校から大学までの教科書を用いて、13 段階

の難易度を推定する難易度測定システム「帯」を構築し、公開している<sup>\*1</sup>。難易度は、1 から 6 が小学 1 年から 6 年、7 から 9 が中学 1 年から 3 年のように対応付けられる。難易度の規準には、小中高大の教科書 127 冊から抽出した 1478 サンプル、約 100 万字のコーパスを用いている。帯では、まず、それぞれの難易度に対する尤度を、連続する 2 文字の生起確率 (文字 bigram) に基づいて計算し、得られた尤度のうち、最大の尤度をとる難易度を求める難易度としている。また、難易度に順位関係が存在し、難易度に対する尤度は緩やかな曲線を描くことが期待されることから、尤度の値をクラス間でスムージングしている。帯では文字 bigram のみを用いるため、形態素解析や品詞体系に影響されないという利点がある一方、構文的な複雑さを反映できないという問題がある。

柴崎ら [11] は、小学 1 年から中学 3 年までの国語教科書中のテキストを基に、学年による文章の難易度の測定方法を提案している。利用したテキストは、国語の散文の教材のみ、243 テキスト、約 58 万文字、約 2 万文である。柴崎らは、複数の特徴量について調査した結果、テキスト全体のひらがなの割合と 1 文の平均述語数が、学年を推定するための有意な独立変数となることを示した。柴崎らが導出した、学年を推定するための重回帰式は (1) の通りである。

$$Y = -0.145X_1 + 0.587X_2 + 14.016 \quad (1)$$

ここで、 $Y$  = 学年、 $X_1$  = テキスト全体のひらがなの割合、 $X_2$  = 1 文の平均述語数 である。柴崎方法では、述語数によって文の複雑さは反映されているが、語彙的なむずかしさが反映されないという問題がある。

### 3. データ

規準コーパスには、現代日本語書き言葉均衡コーパス<sup>\*2</sup>(以下、BCCWJ) に含まれる小・中学の教科書を用いた。BCCWJ には高校教科書も含まれるが、本稿では利用しない。高校教科書の場合、学習する学年が明確に決まっておらず、小・中学の教科書と同様に学年に対応するクラスを付与できなかったためである。表 1 に、教科書のデータサイズと含まれる教科の内訳を示す。

連絡先: 藤田 早苗, NTT コミュニケーション科学基礎研究所,  
〒 619-0237 京都府相楽郡精華町光台 2-4, Tel: 0774-93-5331, Fax: 0774-93-5345, fujita.sanae@lab.ntt.co.jp

\*1 <http://kotoba.nuee.nagoya-u.ac.jp/sc/obi2/>, 本稿では、obi ver. 2.304 を利用

\*2 <http://www.ninjal.ac.jp/kotonoha/>

表 1: データサイズ: BCCWJ 小・中学校教科書

学 年	クラ ス	文字数		語数		文数		異なり語数		ファイ ル数	教科(数)
		Sum.	Ave.	Sum.	Ave.	Sum.	Ave.	Total	Ave.		
小1	1	4,441	493.4	2,470	274.4	271	30.1	184	95.9	9	国語(2), 数学(3), 生活(2), 芸術(2)
小2	2	7,116	711.6	4,219	421.9	398	39.8	124	105.1	10	数学(4), 生活(1), 芸術(5)
小3	3	24,723	1301.2	13,955	734.5	1,087	57.2	725	193.7	19	国語(4), 数学(3), 理科(4), 社会(1), 芸術(7)
小4	4	25,322	1688.1	14,996	999.7	1,040	69.3	666	219.8	15	国語(3), 数学(3), 理科(2), 社会(2), 芸術(5)
小5	5	41,080	2054.0	24,727	1236.4	1,576	78.8	1,002	265.8	20	国語(4), 数学(3), 理科(4), 社会(4), 芸術(5)
小6	6	61,579	2932.3	37,421	1782.0	2,483	118.2	2,887	403.5	21	国語(4), 技術家庭(3), 数学(7), 理科(2) 社会(3), 芸術(2)
中1	7	26,395	2399.5	16,179	1470.8	1,000	90.9	878	344.9	11	国語(6), 数学(2), 芸術(3)
中2	8	113,489	4053.2	69,185	2470.9	3,640	130.0	3,607	538.5	28	国語(3), 数学(2), 理科(6), 社会(13), 芸術(4)
中3	9	85,277	4060.8	50,312	2395.8	2,866	136.5	2,065	482.4	21	国語(2), 技術家庭(6), 数学(2), 理科(5) 社会(2), 芸術(4)
合計		389,422	2528.7	233,464	1516.0	14,361	93.3	12,138	335.6	154	国語(28), 技術家庭(9), 数学(29), 理科(23) 生活(3), 社会(25), 芸術(37)

ただし、語数は MeCab ver. 0.996 (<http://mecab.sourceforge.net/>, UniDic 版) による解析結果をカウント。文字数、語数は空白を除いた数。

## 4. 提案手法

難易度は利用されている語と構文的な複雑さの両面から規定されると考えられる。[12]では、利用されている語と文の複雑さの両方を反映するための方法を提案し、絵本を対象とした場合には非常に高い精度で難易度推定を行うことができることを示した。本稿では [12] の提案手法を基本的に踏襲し、教科書においても提案手法が有効であることを示す。以下に、[12] の提案手法を簡単に紹介する。

### 4.1 学習器

学習器には  $SVM^{RANK}$  [2]<sup>\*3</sup> を利用する。 $SVM^{RANK}$  は Support Vector Machine (SVM, [8]) を用いたランキング学習用の学習器であり、順位付けがはっきりしている訓練データ同士による一対比較を繰り返すことでランキングのモデルを学習する。例えば、各テキストが 9 つのクラスのいずれかに属する場合、 $9 > 1, 8 > 1, \dots, 2 > 1$  のように、訓練データ中の各クラスのテキストのそれぞれ同士を一対比較し、各テキスト対のスコアの大小関係を学習する。

予測結果として出力されるスコアは、クラスを直接表すものではないが、スコアでソートすることにより、ランキングすることができる。例えば、テキスト  $A$  のスコア  $>$  テキスト  $B$  のスコアの場合、テキスト  $A$  の方がテキスト  $B$  より難易度が高いという予測となる。

さらに本稿では、訓練データのスコアからクラスの境界を推定する。具体的には、クラス  $i$  とクラス  $i+1$  に分ける場合、クラス  $i$  に含まれるテキストのスコアの最大値  $max_i$  と、クラス  $i+1$  に含まれるテキストのスコアの最小値  $min_{i+1}$  との中間値を閾値  $th$  とし、 $th$  よりスコアが小さければクラス  $i$ 、大きければクラス  $i+1$  とする。

### 4.2 特徴量

[12]では、文の複雑さを反映するための特徴量として、1文に含まれる文節数の平均値(以下、平均分節数)と、1文に含まれる語数の最大値(以下、最大語数)を利用している。本稿では、これらに加え、全文字に対するひらがなとカタカナを合わせた割合も利用する実験を行う。絵本では、ひらがなとカタカナ以外ほとんど出現しないため、[12]では用いていないが、小学生以上では効果がある可能性があると考えたためである。

次に、出現する語の難しさを反映する方法について述べる。出現する語を特徴量に反映する最も単純な方法は、出現する語

の出現回数を数えて特徴量とする方法だろう。しかし、語の種類は非常に多いため、そのまますべてを特徴量として利用することは困難である。たとえば、[7]は、全対象テキストで一対比較を繰り返して難易度順にソートする手法を提案している。[7]は出現する語を特徴量として利用しているが、一対のテキスト毎に特徴ベクトルを構築することで特徴量の爆発を抑えている。しかし、全テキストのソートに時間がかかることと、同じ特徴量で統一的に比較できないという問題点がある。

提案手法では、単語  $n$ -gram を用いて各クラスにおける言語モデルを構築し、各テキストの各クラスに対する尤度を特徴量として利用する。語そのものを利用するのではなく、数値に直すことで、統一的な比較を行うことができる。ここで、尤度を計算する際、各語の出現頻度ではなく、 $tf \cdot idf$  を利用すると、より高い精度で推定できる。

具体的な計算方法は次の通りである。まず、クラス  $i(1..i..n)$  における単語  $w_j$  の生起確率  $P_i(w_j)$  を、式 (2) で求める。

$$P_i(w_j) = \frac{f(w_j, D_i)}{\sum_{k=1}^n f(w_j, D_k)} \quad (2)$$

ここで、 $D_i$  は、クラス  $i$  に含まれる文書集合を示し、 $f(w_j, D_i)$  は、 $D_i$  における単語  $w_j$  の出現頻度である。次に、 $P_i = 0$  となる場合に尤度が計算できなくなることを避けるため、すべての  $P_i$  が 0 でなくなるまで、式  $P_i(w_j) = \frac{P_{i-1}(w_j) + P_{i+1}(w_j)}{2}$  を繰り返し適用して補正する。

各クラスにおける言語モデル  $M_i$  におけるテキスト  $T$  の尤度  $L(M_i|T)$  は式 (3) で計算する。また、頻度の代わりに  $tf \cdot idf$  で重み付けする場合の式は (4) の通りである。

$$L(M_i|T) = \sum_{w_j \in T} f(w_j, T) \log P_i(w_j) \quad (3)$$

$$L(M_i|T) = \sum_{w_j \in T} \frac{f(w_j, T)}{\sum_l f(w_l, T)} \times \log \frac{D}{df_j} \times \log P_i(w_j) \quad (4)$$

ここで、 $f(w_j, T)$  は、テキスト  $T$  中での語  $w_j$  の出現頻度、 $\sum_l f(w_l, T)$  は、テキスト  $T$  に含まれる語の数、 $D$  は全テキスト数、 $df_j$  は語  $w_j$  の出現するテキストの数を表している。

上述の説明は単語 unigram で行ったが、本稿では、さらに、単語 bigram, 単語 trigram を用いた実験も行う。

ここで、尤度が最大となるクラスを、求めるクラスとしても良く、それだけでも高い精度が得られる。ただし、その場合、構文的な難しさを反映できないため、得られた尤度を特徴量と

\*3 [http://www.cs.cornell.edu/people/tj/svm-light/svm\\_rank.html](http://www.cs.cornell.edu/people/tj/svm-light/svm_rank.html)

表 2: 実験結果

方法	±0 (%)	±1 (%)	$\kappa w$	$R$		
帯 [9]	66	42.9	133	86.4	0.73	0.905
柴崎方法 [11]	56	36.4	116	75.3	0.622	0.851
単語 n-gram : 尤度 (頻度, 式 (3)) 最大のクラスを選択						
uni-freq	51	33.1	85	55.2	0.365	0.652
bi-freq	115	74.7	147	95.5	0.886	0.967
tri-freq	151	98.1	152	98.7	0.984	0.989
単語 n-gram : 尤度 ( $tf \cdot idf$ , 式 (4)) 最大のクラスを選択						
uni-tfidf	109	70.8	139	90.3	0.836	0.934
bi-tfidf	149	96.8	153	99.4	0.986	0.996
tri-tfidf	152	98.7	153	99.4	0.993	0.997
提案手法: $SVM^{RANK}$						
uni/bi-tfidf + sta1	137	89.0	152	98.7	0.956	0.988
uni/bi-tfidf + sta2	135	87.7	152	98.7	0.951	0.987
tri-tfidf + sta1	151	98.1	153	99.4	0.991	0.997
tri-tfidf + sta2	150	97.4	153	99.4	0.988	0.996

sta1 には平均文節数と最大語数, sta2 にはひらがな + カタカナ割合も利用.

して他の特徴量と共に学習に利用することでよりロバストな難易度推定モデルを構築することを目指す.

$SVM^{RANK}$  では特徴量の値は連続値で与えることができる. しかし, 本稿で利用する特徴量の値には非常にばらつきがあることから, 特徴量を正規化して与えることとした. 具体的には, 各特徴量の平均  $m$  と標準偏差  $s$  を計算し, 特徴量の値  $x$  を  $x' = (x - m)/s$  のように正規化してから与える.

## 5. 実験結果と分析

本章では, 前章で紹介した規準コーパスに対し, 先行研究, および提案手法の適用実験を行う. 難易度のクラスと学年は, 表 1 の通り, クラス 1 が小学 1 年, クラス 7 が中学 1 年, のように対応する.

表 2 に実験結果を示す. ここで,  $\pm 0$  は正しいクラスを推定できた数と割合 (的中率),  $\pm 1$  は前後 1 クラスずれて推定された場合も正解とした場合の数と割合を示す. また, 重み付き  $\kappa w$  係数 (Weighted  $\kappa w$ ), 正解クラスと推定されたクラスの相関係数 ( $R$ ) も示した.

表 2 上部に, まず, 帯 [9] の配布モデルによる推定結果を掲載した. 推定結果は,  $R = 0.905$  であり, [9] で最も良い結果として報告されている値 ( $R = 0.94$ ) には及ばないものの, 高い精度で推定することができた.

次に, 柴崎らの式 (1) を適用した結果を掲載した. ただし, 式 (1) では,  $Y$  は離散値ではなく連続値として得られるため, 小数第一位の四捨五入によっていずれかのクラスに振り分けられている. つまり,  $Y = 1.4$  と得られれば, クラスは 1 とした. さらに,  $Y < 0.5$  の場合には, クラス 1 とした. 柴崎ら [11] の実験では, 決定係数は  $R^2 = 0.791$  と報告されている. 本稿のデータに適用した場合,  $R^2 = 0.724201$  ( $R = 0.851$ ) となり, 報告されている値ほどではないが, こちらも相関は高い.

帯と柴崎方法では, 帯の方が精度が高いが, 本稿の対象コーパスには, 様々な教科の教科書が含まれるため, 国語の散文のみを学習に利用している柴崎方法に不利な設定といえる.

表 2 の中部には, 式 (3), (4) によって尤度を計算し, 最も尤度が高くなったクラスを求めるクラスとした場合の結果を示した. 評価は leave-one-out cross-validation により行った. ここで, 'uni-freq' を除き, 尤度を利用した結果はすべて, 先行研究より高い精度を得られた. 特に, 頻度より  $tf \cdot idf$  で重み付けした方が精度が高い. 低学年より高学年の方が出現する語数

自体が多い (表 1) ため, 特に低学年の場合, 頻度で重み付けをすると, より高い学年の尤度の方が高くなる傾向があるが,  $tf \cdot idf$  で重み付けを行う場合, 対象テキストに特徴的な語に対する重みを大きくすることが出来るため, 出現する語数が少ない点を補正できたのだと考えられる.

また,  $n$  が小さい方が  $tf \cdot idf$  で重み付けを行う効果が高く, 'uni-tfidf' より, 'bi-tfidf', 'tri-tfidf' を利用するほうが 20% 以上も精度が高くなっている. つまり, 出現する語の特徴を捉えるには, bigram 以上が望ましい.

表 2 の下部には提案手法の結果を示した. 評価は leave-one-out cross-validation で行っている. なお, 線形 SVM を利用し, コストパラメータは評価対象を除いた学習データの 5-fold-cross-validation によって毎回決定した.

提案手法の内, 'uni/bi-tfidf+sta1' は, 絵本で最終的に採用したのと同じ特徴量を利用している. 'uni-tfidf' による精度は 'bi-tfidf' 等より低いが, 言語モデルに出現しない語をなくして頑健にする目的で 'uni-tfidf' も採用している. ただし, 'tri-tfidf+sta1/2' のモデルの方が, +10% 近く精度が高い. また, sta2 はひらがなとカタカナの割合も利用したモデルだが, 利用しないモデルの方が +1.3% 精度が高くなった. 提案手法で最も精度が高かったモデルは 'tri-tfidf+sta1' であり, 的中率 ( $\pm 0$ ) 98.1%,  $R = 0.997$  だった. この結果は, 'tri-tfidf' の結果より若干低いが両者に有意差はない. また, 前述のように, 提案手法は構文的な複雑さを反映できるため, よりロバストだという利点がある. なお, 本稿では, 先行研究との比較のため,  $R$  の値を中心に分析したが,  $\kappa w$  の方がクラス推定の正確さを比較するという点では差がわかりやすく, 表 2 でも, 手法ごとの差がより大きくあらわれている.

表 3 に提案手法 'tri-tfidf+sta1' によりクラス毎に分けた場合の詳細な結果を示す. 本稿で利用した規準コーパスには様々な教科が含まれるにも関わらず, 非常に正確に学年を推定できていることがわかる. 絵本 [12] は 4 クラスで実験しており, 一概には比較できないが, 絵本より精度よく推定できている. これは, 教科書の方が, 出現する語数が多いこと, 教科書がよく統制されていることなどが理由として考えられる.

## 6. 特徴量の妥当性分析

4.2 節で述べた様に, 文の複雑さを表す特徴量として, 平均文節数と最大語数を利用した. また, 全文字に対するひらがなとカタカナを合わせた割合も利用した実験を行った. 各クラスにおけるこれらの値の平均と分散を表 4 に示す. 表 4 から, 平均分節数と最大語数は, クラスが高くなるに従って増加し, ひらがなとカタカナを合わせた割合は逆に減少する傾向があることが見て取れる. これらはほぼ単調に増減しているが, 中学などの一部のクラスで逆転減少も起こっている. 言語モデルだけを利用した場合より, やや精度が低くなったのはそのためだと考えられる.

次に, 言語モデルの妥当性を定性的に確認する. 言語モデルを構築すると, 各 n-gram が各学年においてどのような出現傾向にあるかを可視化して確認できるという利点がある. 例として, 図 1 に, 「つかう」「使う」「運動」「長い」について, 式 (2) で求めた生起確率を図示する. 「つかう」と「使う」は, ひらがな表記が漢字表記の違いたが, 「つかう」は小学校低学年で多く出現し, 高学年からはほとんど出現しない. 一方, 「使う」は, 低学年ではほとんど出現しないが, 高学年以降生起確率が高くなる. また「運動」は学年と共に多く出現するようになるが, 「長い」は各学年でほぼ均等に出現してい

表 3: 提案手法 (tri-tfidf + stal : SVM<sup>RANK</sup>)

	1	2	3	4	5	6	7	8	9	計	±0	(%)	±1	(%)
1	9	0	0	0	0	0	0	0	0	9	9	100.0	9	100.0
2	1	8	1	0	0	0	0	0	0	10	8	80.0	10	100.0
3	0	0	19	0	0	0	0	0	0	19	19	100.0	19	100.0
4	0	0	0	15	0	0	0	0	0	15	15	100.0	15	100.0
5	0	0	0	0	20	0	0	0	0	20	20	100.0	20	100.0
6	0	0	0	0	0	20	0	1	0	21	20	95.2	20	95.2
(小計)										(94)	(91)	(96.8)	(93)	(98.9)
7	0	0	0	0	0	0	11	0	0	11	11	100.0	11	100.0
8	0	0	0	0	0	0	0	28	0	28	28	100.0	28	100.0
9	0	0	0	0	0	0	0	0	21	21	21	100.0	21	100.0
合計	10	8	20	15	20	20	11	29	21	154	151	98.1	153	99.4

$\kappa w = 0.991, R = 0.997, RMSE = 0.197$

表 4: 特徴量のクラス毎の平均と分散

クラス	平均文節数		最大語数		ひらがな+カタカナ割合	
	平均	分散	平均	分散	平均	分散
1	4.0	1.3	16.0	5.6	95.3	6.3
2	4.2	0.7	20.2	5.4	87.8	10.9
3	4.7	1.0	29.4	7.5	83.1	8.2
4	5.3	1.4	31.1	7.5	78.7	6.2
5	5.4	1.1	39.4	11.1	72.3	5.8
6	5.4	1.2	41.3	10.4	69.6	5.9
7	6.3	1.2	42.8	7.8	65.7	3.7
8	6.9	1.3	50.6	15.7	64.9	5.3
9	6.6	1.4	49.6	12.8	66.0	6.1

る。このように、語やその表記によって、どういった学年で出現しやすいか、という傾向をよく捉えることができている、本稿でも有効に働いたと考えられる。

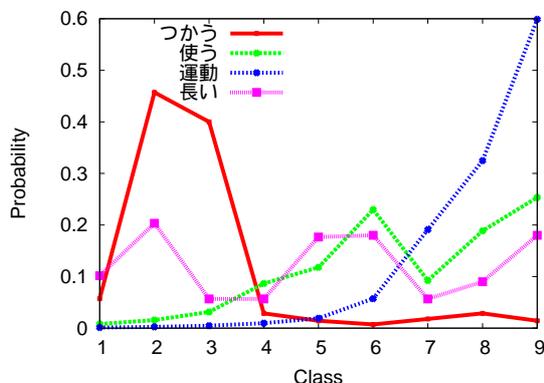


図 1: 生起確率 (unigram): 式 (2) で計算

## 7. まとめと今後の課題

本稿では、教科書を規準データとした難易度推定方法を提案した。教科書を規準データとして利用する先行研究には、帯 [9] や、柴崎らの方法 [11] があり、いずれの推定結果も学年との相関は高い。しかし、帯の場合、文字 bigram だけを利用しているため、構文的な複雑さを反映できない。柴崎らの方法の場合、逆に、どのような語が出現しているかといった情報を反映できない。

そこで、本稿では、出現する語と、文の難しさを反映する特徴量の両方を組み込むことのできる手法を提案した。基本的な手法は、[12] で絵本を対象として提案した手法と同様である

が、教科書に対しても有効であり、先行研究より推定精度が高いことを示した ( $R = 0.997$ , 的中率 ( $\pm 0$ ) 97.4%)。

今後は、高校以上を対象とする文章の難易度推定方法の検討や、絵本と教科書の両方を規準データとして統合する方法の検討を行い、それにより、幼児期から学童期、あるいはそれ以上の難易度を含めた一貫した難易度推定モデルの構築を目指したい。また、一般的な難易度の推定だけでなく、子供の興味や発達に応じたリコメンド方法について検討していきたい。

## 参考文献

- [1] Kevyn Collins-Thompson and James P Callan. A language modeling approach to predicting reading difficulty. In *Proceedings of HLT-NAACL 2004*, pp. 193–200, 2004.
- [2] Thorsten Joachims. Training linear svms in linear time. In *Proceedings of KDD-2006* pp. 217–226, 2006. ACM.
- [3] Yi Ma, Eric Fosler-Lussier, and Robert Lofthus. Ranking-based readability assessment for early primary children’s literature. In *Proceedings of NAACL-2012*, pp. 548–552, 2012.
- [4] Sarah E. Petersen and Mari Ostendorf. A machine learning approach to reading level assessment. *Computer Speech & Language*, Vol. 23, No. 1, pp. 89–106, 2009.
- [5] Satoshi Sato, Suguru Matsuyoshi, and Yohsuke Kondoh. Automatic assessment of japanese text readability based on a textbook corpus. In *Proceedings of LREC-2008*, 2008.
- [6] Sarah Schwarm and Mari Ostendorf. Reading level assessment using support vector machines and statistical language models. In *Proceedings of ACL-2005*, pp. 523–530, 2005.
- [7] Kumiko Tanaka-Ishii, Satoshi Tezuka, and Hiroshi Terada. Sorting texts by readability. *Association for Computational Linguistics*, Vol. 36, No. 2, pp. 203–227, 2010.
- [8] Vladimir Naumovich Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., 1995.
- [9] 小島健輔, 佐藤理史, 藤田篤. 文字 bigram モデルを用いた日本語テキストの難易度推定. 言語処理学会第 15 回年次大会 (NLP-2009), pp. 897–900, 2009.
- [10] 李在鎬. 大規模テストの読解問題作成過程へのコーパス利用の可能性. 日本語教育学会論文誌, Vol. 148, pp. 84–98, 2011.
- [11] 柴崎秀子, 玉岡賀津雄. 国語教科書を基にした小・中学校の文章難易度学年判定式の構築. 日本教育工学会論文誌, Vol. 33, No. 4, pp. 449–458, 2010.
- [12] 藤田早苗, 小林哲生, 南泰浩, 杉山弘晃. 幼児を対象としたテキストの対象年齢推定方法. 2015. (under review).
- [13] 藤田早苗, 小林哲生, 平博順, 南泰浩, 田中貴秋. 絵本を基にした対象年齢推定方法の検討. JSAI-2014, 3D4-4. 2014.
- [14] 佐藤理史. 均衡コーパスを規準とするテキスト難易度測定. 情報処理学会論文誌, Vol. 52, No. 4, pp. 1777–1789, 2011.