

## 英文読解速度テストに用いたテキストの難易度を考慮した テスト有効性の検証

岡本隆志\*

南條浩輝\*

吉見毅彦\*

小谷克則†

### 1 はじめに

近年、英語を使用する機会が増えつつあり、より実践的な英語能力が求められている。その獲得支援として英語学習支援システムの構築が考えられている。このような英語学習者支援システムの構築の基礎資料として英語学習者コーパスが必要となる。代表的な外国語学習者コーパスには、Izumi et al.2004[1], Wen et al.2008[2], Granger et al.2009[3], Meures et al.2010[4] などがあるが、その多くが学習者の作文・発話を収集したコーパスであり、読解コーパスは少ない。これに対し、日本語を母語とする英語学習者の4技能(読解, 聴解, 作文, 発音)を測定するテストの結果を収集しコーパス(以下, 4技能コーパス)が作成された[5, 6]。

本研究では、4技能コーパスのうち読解コーパスの有効性、すなわち、その読解コーパスを収集するために行われた日本語を母語とする英語学習者の英語読解速度を測定するテスト(以下, 読解速度テスト)の結果の有効性を検証する。これまでに読解コーパスの有効性の検証を行ったものに、小谷らの研究[7]がみられる。ただし、検証において、読解速度テストに用いたテキストの難易度が考慮されていない。これに対し、本研究では4技能コーパスの読解コーパスを対象とし、テキストの難易度を考慮して検証を行う。テキストの難易度別に検証することで難易度とテストの有効性の関係を明らかにできる可能性があり、日本人英語学習者の学習・教育にどのような難易度のテキストを使うのが適切であるのかを明らかにできると考える。本研究で読解能力として用いている読解速度は小谷らの研究[7]で読解能力の評価指標として有効であることが確認されており、読解速度は重要な評価指標として利用されている。この読解コーパスの有効性が確認されることで、学習者が英文テキストを読んだ時に理解で

きていない英文を探すシステムなどに利用することが考えられる。

### 2 読解コーパスの構築

先行研究[5, 6]で次の方法により読解速度テストが実施され、読解コーパスが作成された。

#### 2.1 読解速度テストの受験者

読解速度テストの実施時期から1年以内にTOEICを受験したことがある受験者を募集し、日本語を母語とする大学生90人(性別は男性48人, 女性42人)を受験者とし、平均年齢は21.6歳であり、最年少者は19歳, 最年長者は40歳であった。受験者の英語学習期間は、53ヶ月~243ヶ月であった。受験者のうちTOEICスコアが730点~985点の30人を上級クラスに、500点~725点の30人を中級クラスに、280点~495点の30人を下級クラスに分類された。

#### 2.2 テストのテキストと設問

読解速度テストに用いたテキストはVoice of America (VOA)のサイト(<http://www.voanews.com>)から収集されたニュース記事である。使用された記事の詳細を表1に示す。記事はすべて約350語から成る。難易度の観点で分類されており、2つの平易な記事と2つの難解な記事の2種類が用意されている。平易な記事は英語学習者向けの特別なセクションから収集された、慣用句を含まない1500単語の基本語彙からなる簡単なテキストである。難解な記事は英語母語話者向けのセクションから収集されたテキストである。表1より、平易な記事と難解な記事では一文あたりの文

\*龍谷大学理工学研究科

†関西外国語大学外国語学部

表 1: 読解速度テストに使用された記事の長さ

難易度	記事の長さ		文の長さ			
	単語数	文数	平均 (語)	分散	最短 (語)	最長 (語)
平易	335	25	13.4	4.7	7	23
平易	356	25	14.2	4.2	5	22
難解	350	15	23.3	7.2	12	37
難解	348	15	23.2	10.1	11	42

の長さ (単語数の平均) が異なり, 難解な記事の方が文が長いことがわかる。

各記事に関する設問としては 4 択の多肢選択式の問題 5 問が用いられた。読解速度テストであるにも関わらず, 設問が設けられた理由は, 記事を読んだときに適当に読んでいる可能性 (記事の内容を理解していない可能性) を判断する 1 つの基準とするためである。

## 2.3 テストの実施方法

TOEIC スコアを持つ日本人英語学習者に読解速度テストを受験してもらった。このテスト結果とテストに使用された記事, 受験者の TOEIC スコアを集められたものを読解コーパスとしている。受験者には 4 つの記事を一文ずつ読んでもらい, 辞書や参考書の使用を禁止し, 記事を構成する文を 1 度だけ読むという制限を設けていた。さらに, 各記事に対して 13 分の制限時間を設けた。1 つの記事を読み終えるごとに設問に答えてもらっていた。一文ごとに読解時間を測定し, 記事の読解時間と単語数から 1 分間に受験者が読んだ単語数を読解速度 (Words Per Minute (WPM)) として算出された。

## 2.4 読解速度テストの結果

読解速度テスト結果の平均値, 中央値, 最小値, 最大値, 標準偏差を表 2 に示す。全記事での読解速度の度数分布図を図 1 に, 平易な記事での読解速度の度数分布図を図 2 に, 難解な記事での読解速度の度数分布図を図 3 に示す。

表 2: 読解速度テストの結果

	全記事	平易な記事	難解な記事
平均値	104.9	107.2	101.1
中央値	99.8	103.8	97.0
最小値	47.9	36.0	48.2
最大値	293.8	282.5	312.6
標準偏差	37.6	40.4	37.1

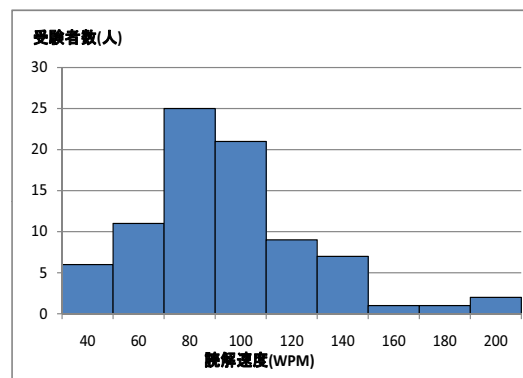


図 1: 全記事での読解速度のヒストグラム

## 3 読解速度テストの有効性の検証

記事の難易度別に有効性を確認するため, 妥当性と信頼性について検証を行う [8, 9]。

### 3.1 検証対象

検証対象は, 全記事を使用したテスト結果と平易な記事, 難解な記事をそれぞれ使用したテスト結果の合計 3 種類のテスト結果である。受験者ごとに, 読解速度を求めてテストの有効性を検証した。具体的には, 受験者ごとの全記事 (80 文) の読解速度の合計を全記事での全文数の 80 で割った結果を 1 件のデータとした。平易な記事の場合は, 受験者ごとの平易な記事 (50 文) の読解速度の合計を平易な記事での全文数の

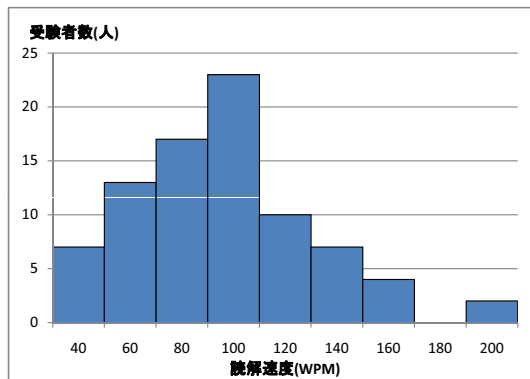


図 2: 平易な記事での読解速度のヒストグラム

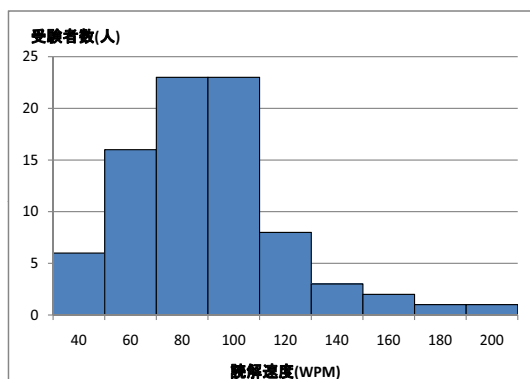


図 3: 難解な記事での読解速度のヒストグラム

50 で割った結果を 1 件のデータとした。難解な記事の場合は、受験者ごとの難解な記事 (30 文) の読解速度の合計を難解な記事での全文数の 30 で割った結果を 1 件のデータとした。検証を行う前に外れ値を取り除いた。具体的には、読解速度が 1000WPM 以上の場合と TOEIC のリーディングセクションの得点 (以下、TOEIC 読解スコア) が 300 点以上あるにも関わらず読解速度が 70WPM 以下の場合を外れ値とした<sup>1</sup>。これは受験者が記事をほとんど読んでいない可能性があると考えられるためである。

### 3.2 パラメトリック手法の前提条件の検証

パラメトリック手法の前提条件は正規性と等分散性であるので、正規性の検証のため有意水準 5% でコルモゴロフ・スミルノフ検定を行った結果、全記事での読解速度のデータは正規分布でないといえない ( $K=1.10$ ,  $p=0.08$ ) ことがわかった。平易な記事

<sup>1</sup> 英語母語話者の読解速度は 200~300WPM 程度であるとされる [10]

での読解速度のデータは正規分布でないといえない ( $K=0.91$ ,  $p=0.17$ ) ことがわかった。難解な記事での読解速度のデータは正規分布でないといえない ( $K=1.14$ ,  $p=0.07$ ) ことがわかった。

次に、TOEIC スコアに基づき分類した上級クラス・中級クラス・下級クラスでの平均読解速度分布の母分散が等しいか調べるため有意水準 5% でバートレットの方法を行った結果、全記事での読解速度のデータは分散が等しくない ( $\chi^2=12.05$ ,  $p<0.05$ ) ことがわかった。平易な記事での読解速度のデータは分散が等しくない ( $\chi^2=12.41$ ,  $p<0.05$ ) ことがわかった。難解な記事での読解速度のデータは分散が等しくない ( $\chi^2=7.76$ ,  $p<0.05$ ) ことがわかった。

等分散性を確認できなかったので、本研究ではノンパラメトリック手法を用いて検証を行う。

### 3.3 妥当性の検証

テストの妥当性は、あるテストが測定対象とする受験者の能力をどの程度測定できているかによって検証される。テストの妥当性は着眼点により、構成概念妥当性と基準関連妥当性に分けられる。

#### 3.3.1 構成概念妥当性の検証

構成概念妥当性は、テストが構成概念を適切に測定しているかどうかにより検証される。その検証法の一つである異集団法では、テストが異なる構成概念を保持する受験者を区別できれば、そのテストの構成概念妥当性が確立されたと考えられる。本稿の場合、英語学習者を TOEIC スコアにより習熟度別に区別することから、TOEIC スコアにより示される英語能力が構成概念と位置づけられる。そこで、読解速度テストの結果に基づいて受験者を習熟度別に弁別できるか (クラス間に統計的有意差が確認できるか) どうかを検証する。

構成概念妥当性を調べるためにクラスカル・ウォリス検定を行った結果、全記事での習熟度間における読解速度に統計的に有意な差が確認できた ( $H(2, N=83)=32.97$ ,  $p<0.05$ )。平易な記事での読解速度に統計的に有意な差が確認できた ( $H(2, N=83)=39.93$ ,  $p<0.05$ )。難解な記事での読解速度に統計的に有意な差が確認できた ( $H(2, N=83)=14.94$ ,  $p<0.05$ )。次

表 3: 読解速度の多重比較の結果

	全記事	平易な記事	難解な記事
上級・中級間	3.45*	3.78*	2.32
上級・下級間	5.48*	5.89*	3.71*
中級・下級間	2.68*	3.29*	1.71

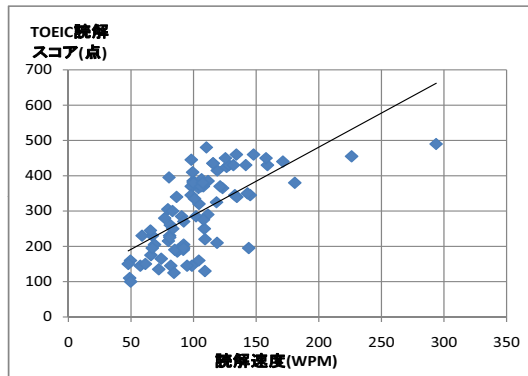


図 4: TOEIC 読解スコアと全記事での読解速度の散布図

にスティール・ドゥワスの方法で多重比較を行った結果を表 3 に示す。検証の結果、全記事と平易な記事ですべてのクラス間で有意差（有意水準 5%）があるといえることがわかった。難解な記事では、上級・下級間で有意差（有意水準 5%）があるといえることがわかった。よって、全記事と平易な記事で構成概念妥当性が確認できた。

### 3.3.2 基準関連妥当性の検証

基準関連妥当性は、あるテストの結果と妥当性が確立されたテストの結果が類似しているかどうかによって検証される。本稿で用いた読解速度テストは TOEIC のリーディングセクションと同様に読解能力を測定するテストである。したがって、TOEIC 読解スコアと類似した結果になれば、読解速度テストに基準関連妥当性があるとみなせる。

TOEIC 読解スコアと全記事での読解速度の散布図を図 4 に、TOEIC 読解スコアと平易な記事での読解速度の散布図を図 5 に、TOEIC 読解スコアと難解な記事での読解速度の散布図を図 6 に示す。基準関連妥当性を調べるためにスピアマンの順位相関係数の有意性検定を行った結果、全記事では相関係数は 0.70 であり、統計的に有意な相関が確認できた ( $F=8.90, p<0.05$ )。平易な記事では相関係数は 0.77 であり、統計的に有意な相関が確認できた ( $F=10.91, p<0.05$ )。難解な記

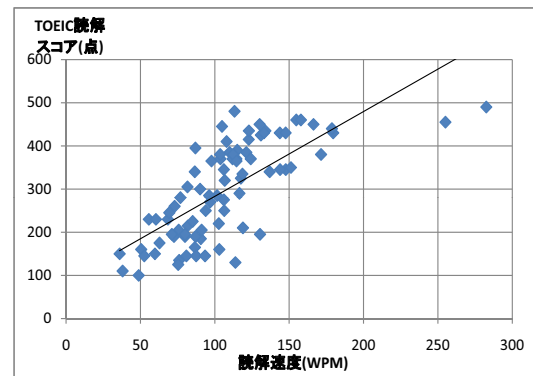


図 5: TOEIC 読解スコアと平易な記事での読解速度の散布図

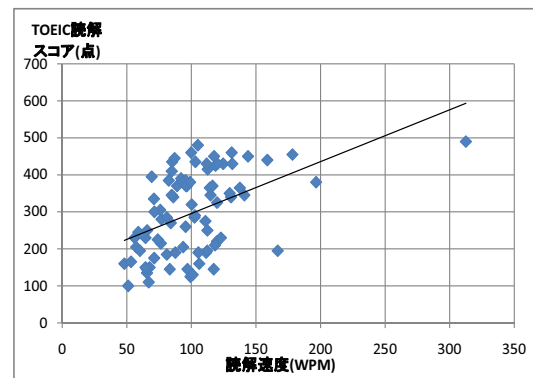


図 6: TOEIC 読解スコアと難解な記事での読解速度の散布図

事では相関係数は 0.48 であり、統計的に有意な相関が確認できた ( $F=4.94, p<0.05$ )。よって、すべてで基準関連妥当性が確認できた。

### 3.4 信頼性の検証

テストの信頼性は、ある受験者が同じレベルのテストを受験したときに、同程度のテスト結果になるかどうかによって検証される。本稿では、内部一貫性法によるクロンバックの  $\alpha$  信頼性係数を算出する。一般に、クロンバックの  $\alpha$  信頼性係数が 0.7 以上であれば、テストの信頼性は高い [8]。

クロンバックの  $\alpha$  信頼性係数を求めた結果、全記事での信頼性係数は 0.943、平易な記事での信頼性係数は 0.981、難解な記事での信頼性係数は 0.873 と 0.7 以上であったので、すべてで信頼性はあるとみなせることがわかった。

### 3.5 検証の考察

検証の結果, 全記事と平易な記事では, 妥当性と信頼性が確認できたため, 有効性があることが確認できた。難解な記事では, 信頼性は確認できたが, 妥当性が確認できなかったため, 有効性は確認できなかった。これらのことから, 日本人英語学習者の学習・教育支援に平易な記事が使えることがわかった。

## 4 結論

本稿では, 読解速度テスト結果(コーパス)の有効性を, テストに用いた記事の難易度別に検証した。

英語母語話者向けの難解な記事は英語学習者向けの平易な記事に比べ入手が容易である。よって, それで有効性が認められることが, 今後のデータ収集にとって有益と考えられる。しかし, 今回の検証では, 難解な記事での有効性は確認できなかった。今回は記事数が4つと少なかったことから, データ数を増やしてより精密に検証を行っていく必要がある。

## 謝辞

本研究は, 科研費 22300299 の助成を受けて行われた。

## 参考文献

- [1] 和泉絵美, 井佐原均, 内元清貴. 日本人 1200 人の英語スピーキングコーパス. アルク, 2004.
- [2] Qiufang Wen, Maocheng Liang, and Xiaoqin Yan. *Spoken and Written Corpus of Chinese Learners(SWECCL)2.0*. Foreign Language Teaching and Research Press, Beijing, China, 2008.
- [3] Sylvaine Granger, Estelle Dagneaux, Fanny Meunier, and Magali Paquot. *International Corpus of Learner English, version 2*. Presses Universitaires de Louvain, Louvain-la-Neuve, Belgium, 2009.
- [4] Detmar Meurers, Niels ott, and Ramon Ziai. Compiling a task-based corpus for the analysis of learner language in context. In *proceedings of Linguistic Evidence 2010*, pp.214-217, 2010.
- [5] Katsunori Kotani, Takehiko Yoshimi, Hiroaki Nanjo, and Hitoshi Isohara. Compiling learner corpus data of linguistic output and language processing in speaking, listening, writing, and reading. In *International Joint Conference on Natural Language Processing(IJCNLP)*, pp.1418-1422, 2011.
- [6] Katsunori Kotani, Takehiko Yosimi, Hiroaki Nanjo, and Hitoshi Isahara. Statistical analysis of a learner corpus integrating linguistic output and language processing data in speaking, listening, writing, and reading. In *Corpora and Language Technologies in Teaching, Learning and Research*, pp.21-28, 2011.
- [7] 小谷克則, 吉見毅彦, 九津見毅, 佐田いち子, 井佐原均. 英文読解速度テストの有効性の検証. *情処科学技術レターズ*, Vol.5 pp.407-411, 2006.
- [8] J.D. ブラウン, 和田稔. 言語テストの基礎知識. 大修館書店, 1999.
- [9] 尾崎茂. 言語テスト学入門. 大学教育出版, 2008.
- [10] R.P Carver. Optimal rate of reading prose. In *Reading Research Quarterly XVIII*, pp.56-88, 1982.