

英語学習支援のための聴解コーパスの有効性の検証

上田翔太[†]南條浩輝[†]吉見毅彦[†]小谷克則[‡]

1 はじめに

近年、インターネットの普及・発展や、国際化・情報化社会が進み、コミュニケーションの手段として外国語を習得する必要性が高まっている。日本でも、日本人が外国語を習得するために教育現場における外国語学習の支援が求められている。そこで、外国語学習支援システムの構築などが考えられているが、そのための基礎資料として日本人外国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、これらは学習者の読解・作文・発話データを収集したコーパスである。聴解データを収集した聴解コーパスとして、Kotani et al. 2011[5][6] があるが、その有効性は確認されていない。

本稿では、このコーパスの有効性を検証し、英語学習支援へのコーパスの利用可能性に関する情報をコーパス利用者に提供する。聴解コーパスの利用例として、英語音声を入力とし、その聴解難易度を出力する聴解難易度測定システムなどが考えられる。

2 聴解コーパスの構築

Kotani et al. 2011[5][6] では、次のような手続きで聴解理解度テストを実施し、その結果を聴解コーパスとしている。

2.1 テストの受験者

テストの受験者は日本語母語話者 90 人である。性別は男性 48 人、女性 42 人である。平均年齢は 21.6 歳である。英語学習期間は、53ヶ月から 243ヶ月である。テスト受験者は全員、テスト受験時期から 1 年以内に TOEIC 受験経験がある。

テスト受験者をその TOEIC スコアにより、習熟度別に 3 クラスに分けている。TOEIC スコア 735 点以上の 30 名を上級クラス、470 ~ 730 点の 30 名を中級

クラス、465 点以下の 30 名を初級クラスとしている。受験者の TOEIC スコアの分布を図 1 に示す。

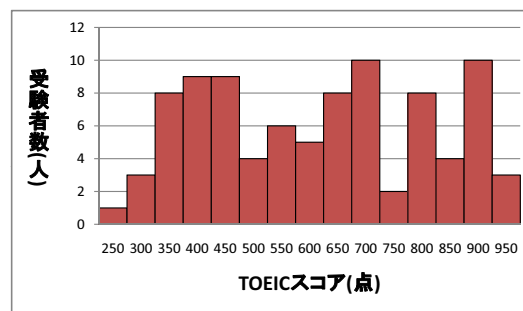


図 1: 受験者の TOEIC スコア分布

2.2 テストに用いたテキスト

テストに用いたニュース記事は Voice of America (VOA) (<http://www.voanews.com>) から集めた 4 記事である [5]。4 記事は難易度により、平易な記事 2 記事と難解な記事 2 記事に分けられる。記事はおよそ 350 語からなり、難易度により読み上げ速度が異なる。使用したテキストの長さを表 1 に示す。

平易な記事には VOAs Special English の記事を用いている。VOAs Special English は、短い単純な文で構成され、基本語彙 1500 語だけで記述されている。また、慣用句は含まれていない。読み上げは明瞭にゆっくりと行われ、読み上げ速度は、英語母語話者にとって通常速度 (1 分間に 250 音節程度) の 3 分の 2 程度である。

難解な記事には VOA editorials を用いている。VOA editorials は幅広い論点から見た異なる主張を述べる社説である。読み上げは英語母語話者にとって通常速度で行われる。

2.3 テストの実施方法

受験者は、英語母語話者が読みあげた 4 つのニュース記事を聴く。記事を聴き終わると記事の内容に関する設問に答える。設問は 4 択問題であり、各記事につき 5 問から成る。4 つのテストの正答率の平均を、受験者の聴解理解度の値としている。各記事とその聴解理解度を聴解コーパスとする。

[†]龍谷大学 理工学研究科[‡]関西外国語大学 外国語学部

表 1: 聴解理解度テストに使用したテキストの長さ

記事の難易度	記事の長さ		文の長さ			
	語の数	文の数	文の平均長 (語)	標準偏差	最短の文 (語)	最長の文 (語)
平易	358	25	14.32	4.8	5	22
平易	341	25	13.6	3.7	6	20
難解	356	15	23.7	8.9	8	39
難解	352	15	23.5	7.6	9	39

2.4 テストの結果

聴解理解度テストの結果の、平均値、中央値、最小値、最大値、標準偏差を表 2 に示す。テストの聴解理解度の分布を図 2 から図 4 に示す。

表 2: 聴解理解度テストの結果

	全記事	平易な記事	難解な記事
平均値	0.53	0.53	0.54
中央値	0.50	0.50	0.50
最小値	0.15	0.20	0.10
最大値	0.95	0.90	1.00
標準偏差	0.15	0.16	0.19

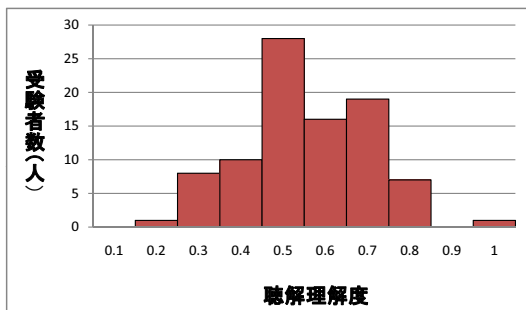


図 2: 全記事を使用した場合の聴解理解度分布

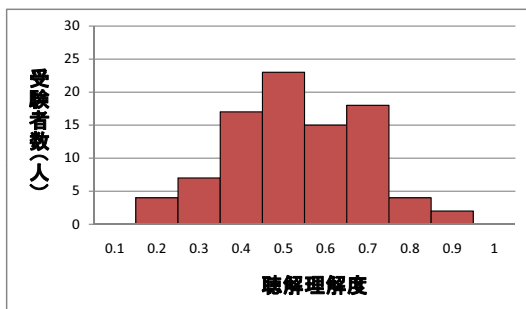


図 3: 平易な記事を使用した場合の聴解理解度分布

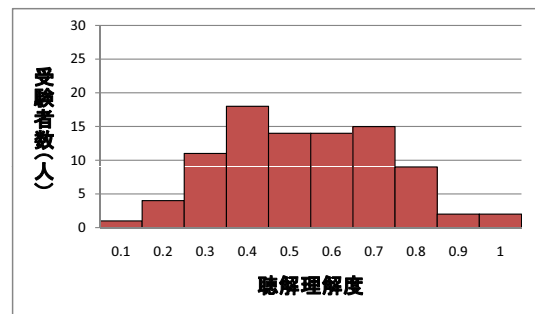


図 4: 難解な記事を使用した場合の聴解理解度分布

3 聴解コーパスの有効性の検証

本研究では、各記事の聴解理解度をテスト項目と考えて古典的テスト理論 [7] による検証を行う。妥当性と信頼性が確立された場合、有効性が確立される。そこで、妥当性と信頼性の検証を行う。

本研究では以下の 3 つのテスト結果を検証対象として検証を行う。

- ・全記事を使用したテスト結果
- ・平易な記事を使用したテスト結果
- ・難解な記事を使用したテスト結果

3.1 パラメトリック手法の前提条件の検証

まず、有効性の検証をパラメトリック検定で行えるかどうかを見るために正規性・等分散性の検定を行う [8]。

テスト結果の正規性を調べるために、コルモゴロフ・スミルノフ検定を用いて、各記事群における上級・中級・下級クラスからなる結果の分布に正規性があるか有意水準 5% で検定を行った。その結果、全記事 ($K=0.98$, $p=0.13$) のテスト結果の分布は正規分布と異なるといえないことがわかった。平易な記事 ($K=1.31$, $p=0.03$)、難解な記事 ($K=1.25$, $p=0.04$) のテスト結果の分布は正規分布と異なることがわかった。

テスト結果の等分散性を調べるためにバートレットの方法を用いて、各記事群における上級・中級・下級ク

ラスの母分散が等しいか有意水準 5%で検定を行った。その結果、全記事 ($\chi^2(2, N=90)=17.49, p<0.05$), 平易な記事 ($\chi^2(2, N=90)=50.51, p<0.05$), 難解な記事 ($\chi^2(2, N=90)=6.59, p<0.05$) で等分散性がないことがわかった。

パラメトリック検定適用の前提条件を満たさなかったため、本研究ではノンパラメトリック手法で検定を行う。

3.2 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、基準関連妥当性に分けられる。

3.2.1 構成概念妥当性の検証

構成概念妥当性とは、テストが測定対象である構成概念 (英語聴解能力) を正しく測定できているかを示すものである。本研究では、クラスカル・ウォリスの順位和検定により構成概念妥当性の検証を行った。

その結果、全記事 ($H(2, N=90)=39.19, p<0.05$), 平易な記事 ($H(2, N=90)=39.53, p<0.05$), 難解な記事 ($H(2, N=90)=24.52, p<0.05$) 全てで有意差が見られた。

次にスティール・ドゥワスの方法で多重比較を行い、それぞれのクラス間の有意差を調べた。t 検定統計量の値を表 3 に示す。

表 3: スティール・ドゥワスの方法での多重比較におけるクラス間有意差の t 検定統計量と検定結果 (* $p<0.05$)

	全記事	平易な記事	難解な記事
上級・中級クラス間	5.02*	5.38*	3.34*
上級・初級クラス間	5.49*	5.49*	4.67*
中級・初級クラス間	2.02	0.44	1.95

表 2 より、全記事、平易な記事、難解な記事全てで、上級・中級クラス間と上級・初級クラス間に有意差があった。このことより、現時点では中級の受験者と初級の受験者を判別できるとはいえないことがわかった。

3.2.2 基準関連妥当性の検証

基準関連妥当性とは、テスト結果と既に妥当性が確立されているテストの結果が類似しているかを示すものである。本研究では、各受験者のテストの結

果と TOEIC のリスニングセクションの得点 (以下、TOEIC 聴解スコア) が類似しているかを調べた。テスト結果と TOEIC 聴解スコアの散布図を図 5 から図 7 に示す。

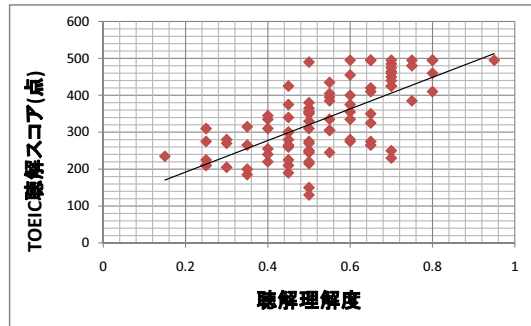


図 5: 全記事を使用した場合の聴解理解度と TOEIC 聴解スコアの散布図

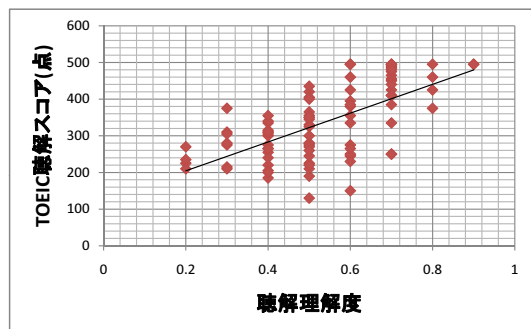


図 6: 平易な記事を使用した場合の聴解理解度と TOEIC 聴解スコアの散布図

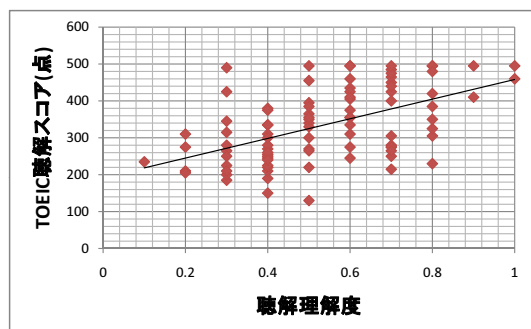


図 7: 難解な記事を使用した場合の聴解理解度と TOEIC 聴解スコアの散布図

スピアマンの順位相関係数を求め、さらにその相関係数の有意性の検定を行った。結果を表 4 に示す。

相関係数を求めた結果、全記事、平易な記事、難解な記事全てで、テスト結果と TOEIC のスコアに強い相関がみられた。また、検定の結果、全記事 ($F(1, 88)=8.47, p<0.05$), 平易な記事 ($F(1, 88)=7.69, p<0.05$), 難解な記事 ($F(1, 88)=5.91, p<0.05$) 全てで有意差が見られた。よって基準関連妥当性が確認された。

表 4: スピアマンの順位相関係数と有意性の検定結果 (* $p < 0.05$)

全記事	0.67*
平易な記事	0.63*
難解な記事	0.53*

3.3 信頼性の検証

テストの信頼性は、ある受験者が同一のテストを受験したとき同程度の結果になるかどうかによって検証される。本研究では、内部一貫性法であるクロンバック 信頼性係数により信頼性を検証した。クロンバック 信頼性係数は式 (1) によって求めた。k はテストの項目数、 S_j は各テスト項目のテスト結果の標準偏差、 S_Y は全体のテスト結果の標準偏差である。

$$= \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k S_j^2}{S_Y^2} \right) \quad (1)$$

クロンバック 信頼性係数を求めた結果を表 5 に示す。

表 5: クロンバック 信頼性係数

全記事	0.621
平易な記事	0.470
難解な記事	0.380

信頼性の検証結果より、全記事の場合では信頼性の高いテストの目安とされる値 0.7 をやや下回ったが、一定の信頼性はあるとみなせることがわかった。

4 結論

本研究では、聴解コーパスの有効性を、古典的テスト理論により検証した。妥当性の検証の結果、中級の受験者と初級の受験者を判別できるといえなかったが、ある程度の妥当性は見られた。信頼性の検証の結果、全記事を用いた場合は一定の信頼性はあるとみなせるとわかった。これらより、全記事を用いた場合のデータは英語学習支援に利用できる可能性があることがわかった。

今後は、コーパスデータを利用した英語学習支援システムの構築や、コーパスの改良として受験者数や記事数を増やすことなどを行っていく予定である。

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. In *ALC*, 2004.
- [2] Q. Wen, M. Liang, and X. Yan. Spoken and written corpus of chinese learners(swecc). 2.0. Beijing, China: Foreign Language Teaching and Research Press., 2008.
- [3] Sylvaine Granger, Estelle Dagneaux, Fanny Meunier, and Magali Paquot (Eds.). International corpus of learner english: version 2. Louvain-la-Neuve, Belgium: Presses Universitaires de Louvain, 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 Yosimi, Hiroaki Nanjo, and Hitoshi Isahara. 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 2011)*, 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] J.D. ブラウン. 言語テストの基礎知識. 大修館書店, 1999.
- [8] 森敏昭, 吉田寿夫. 心理学のためのデータ解析テクニカルブック. 北大路書房, 1990.