

オーセンティック・リスニング・テキストの
CEFR リスニングのレベル判断における諸問題

Issues Arising from the Application of the CEFR Listening
Reference Level Descriptions to Authentic Listening Texts

根岸雅史

Masashi NEGISHI

東京外国語大学

Tokyo University of Foreign Studies

Abstract

The Common European Framework of Reference for Languages (CEFR) is spreading its influence beyond Europe, and is being used for teaching and assessment in many parts of the world. In the application of the CEFR Listening Reference Level Descriptions to authentic listening texts, several issues have been raised. Negishi (2008) showed that most of the assessments of the levels of difficulty of the listening texts correspond to the average test difficulty of their test items, and that main features in the listening descriptors, such as linguistic difficulty, speed, etc. contribute to the difficulty, whereas minor features such as accent contribute to a less extent. In this article, the author attempts to reinterpret Negishi (2008). The results show that non-native judges tend to rate the levels of listening texts based on “linguistic” features if the text is above their level, and based on “content” features if it is below their level. They also indicate that text length is not necessarily the major feature that decides the level of the text, although it is closely related to the average difficulty of the test items, and that it is only when there is a great difference in text length that the judges take this feature into consideration in deciding the levels of authentic texts.

Keywords

CEFR, Reference Level Descriptions, Text Length

1. CEFRのレベル記述

Common European Framework of Reference for Languages (CEFR) は、欧州の言語の共通枠組みとしてスタートしたが、今日では、その影響力は欧州にとどまらず、世界に広がりつつある。Negishi (2006), 根岸 (2006), 中島・永田 (2006) などによれば、このCEFRが欧州だけでなく日本という文脈においても適用がほぼ可能であるということが明らかになってきている。事実、Cambridge ESOLのNick Saville氏などは、近い将来CEFRのうちのEuropeanという語は削除されるかもしれないと述べている。このような状況の中で、

CEFRは世界の国と地域の中で、言語の学習・指導・評価に関する具体的な判断の際の拠り所となりつつある。

ただし、CEFRは欧州のすべての言語の共通枠組みとして考えられているために、個別言語への言及はなく、レベル記述はcan-do statementsによっており、パフォーマンス準拠である。そのため、CEFRの利用にあたっては、それぞれの言語で、たとえば、レベルごとのcan-do statementsがどのような語彙・文法によって実現されるかを考えなければならない。

英語に関しては、現在、ケンブリッジ大学を中心にEnglish Profileというプロジェクトが進行中である。この概要については、以下の通りである。

Building on existing resources such as the Common European Framework of Reference for Languages and the Breakthrough, Waystage, Threshold and Vantage specifications, a multi-disciplinary team is working to produce Reference Level Descriptions for English. These will provide a uniquely detailed and objective analysis of what levels of achievement in language learning actually mean in terms of the grammar, vocabulary and discourse features that learners can be expected to have mastered at each level.

(Council of Europe. *English Profile: Reference level descriptions for English*. Available:<http://www.englishprofile.org/index.html> [2008, October])

このEnglish Profileは、発表技能、とりわけ、ライティングの学習者コーパスを作成することで、各レベルの学習者の言語的な記述を行おうとしている。将来、このコーパスの分析が進めば、各レベルの学習者の特徴が、母語の違いや学習履歴などのさまざまな観点から明らかになっていくだろう。

しかしながら、CEFRにはなにも発表技能だけのレベル記述があるわけではなく、受容技能のリスニングやリーディングのレベル記述も含まれている。言語教育場面では、これらの受容技能に関する具体的なレベル記述も発表技能と同様に重要なはずである。そこで、本稿では、このうちのリスニングについてフォーカスを置いて、そのレベル判断における諸問題について考察を試みる。

まず、CEFR Common Reference Levels: self-assessment gridのListeningに関する記述 (Council of Europe, 2001, pp.26-27) を見てみる。

A1

I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.

A2

I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family

information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.

B1

I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.

B2

I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.

C1

I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.

C2

I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.

これらの記述の中には、「言語的難易度 (e.g. A1: familiar words and very basic phrases)」「トピックの馴染みの度合い (e.g. B2: provided the topic is reasonably familiar)」「興味の度合い (e.g. B1: current affairs or topics of personal or professional interest)」「理解の詳細さ (e.g. B1: understand the main points of clear standard speech)」「長さ (e.g. B2: understand extended speech)」「議論の複雑さ (e.g. C1: when it is not clearly structured and when relationships are only implied and not signalled explicitly)」「発話速度 (e.g. C2: when delivered at fast native speed)」「訛り (e.g. C2: provided I have some time to get familiar with the accent)」「明瞭さ (e.g. A1: when people speak ... clearly)」「ノイズ (e.g. A2: clear ... announcements)」といった要素が含まれている。ちなみに、Buck (2001) は、リスニングの難易度に影響を与えるテキストの特徴として、「言語的特徴」「明示性」「構成」「内容」という枠組みを提案しているが、上述の要素は、これらにほぼ含まれていると言える。

これらの要素は個別に見れば難易度は上がる(たとえば、馴染みのある語彙は、ない語彙より難易度が低いし、議論が複雑であればあるほど理解は困難になる)と考えられる。確かに、この原則自体には異論を挟む余地はなさそうであるが、現実にはそれぞれの難易度が手と手を取り合って増していくわけではない。つまり、「議論は複雑ではない」が「訛り」が強いという組み合わせがあったり、「長さは短い」が「早口である」という組み合わせがあったりする。こうした現実には、CEFRのような枠組みの適用にあたり、困難を引き起こすものと思

われるが(Council of Europe [2001]の中には, CEFRのこうした問題への解決策は明確には示されていないと思われる), 現実の場面で, どのようにレベル判断が行われるのか, そこに含まれる諸問題を根岸(2008)のデータを再解釈することで考察してみたい。

2. CEFRリスニング・テキストの難易度判断に関する諸問題

2.1 根岸(2008)の調査の概要

Horváth and Pižorn (2005)による *Listening* は, British Council の英語教育書シリーズ *INTO EUROPE* のなかの1冊で, シリーズ・エディターの Charles Alderson の指導のもと, ハンガリーとスロバキアの英語教師たちが作成したものである。ここには現実の生活から録ったオーセンティック・テキスト(authentic texts recorded from real life)が CD で提供されている。本書のもっとも特徴的な点は, それぞれの録音に対して, CEFR の分類に基づくレベル分けがなされている点である。この判断は, 以下の通り, 英語教育とアセスメントの専門家によってなされている。

As part of the piloting and analysis of these listening tasks, 14 experts in language testing and language education more generally assessed the level of the tasks in terms of the Council of Europe Framework. The judgements given by these experts were then compared with the actual difficulty that the tasks had shown when they were tried out on learners in school. As a result each task has been given a provisional level (A1, A2, etc.) on the Council of Europe Framework and this level is marked on the task, and the Contents pages in Appendix 2. (Horváth and Pižorn, 2005, p.32)

もちろん, “a provisional level (A1, A2, etc.) on the Council of Europe Framework” とあるように, この判定レベルは暫定的なものであり, これを絶対視することは危険かもしれない。しかし, このような形の英語のテキストに専門家の CEFR のレベル判断がついたものは, 筆者の知る限りでは, 今日見当たらない。

Horváth and Pižorn (2005)には, short-answer questions, multiple-choice tasks, true/false tasks, completion tasks, matching tasks というリスニング・タスクが含まれているが, 根岸(2008)では, テスト方法の影響を避けるためにテスト方法をさまざまな統計的分析にかけやすい multiple-choice tasks に統一した。各問題の詳細は表1の通りである。

表1 Contents of the CDs (Horváth and Pižorn; 2005, Appendix 2)

大問番号	タイトル	CEFR レベル	テスト項目数	長さ (分:秒)
12	Shuttle launch	C1	7	3:51
13	Sleepy Hollow	A2	5	1:54
14	Nicole Kidman	A2	6	3:05
15	The jazz singer	B2	6	2:29
16	Vision regained	B2	6	4:07
17	Home of the future	C2	5	2:02

さらに、これに対応するアンケートも実施した。このアンケートでは、音声テキストごとに CEFR のレベルを判断させると同時に、CEFR のリスニングの能力記述に含まれている要素について、very difficult (4) - difficult (3) - easy (2) - very easy (1)の4段階で判断させた。これらの要素は、Buck (2001)を参考にし、「言語的難易度(語彙)」「言語的難易度(文法・文構造) (Buck, 2001, p.152)」「トピックの馴染みの度合い」「興味の度合い」「理解の詳細さ」「長さ」「議論の複雑さ」「発話速度」「訛り」「明瞭さ」「ノイズ(Buck, 2001, p.187)」とした。たとえば、「訛り」であれば、訛りが強くてわかりづらくて非常に困難を生じさせていると判断すれば、very difficult (4)を選択するという具合である。

調査参加者は、大学1～4年生、大学院博士前期課程および後期課程学生 82 名。このうち、テスト問題を受験したのは 82 名、アンケートに回答したのは大学院生 23 名で、2ヶ月間にわたり筆者によるリスニングの能力記述枠組みおよび CEFR についての講義とワークショップを受講した。

2.2.1 レベル判断に関する結果

表2からわかるように、テストの平均正答率および項目応答理論(IRT)による項目難易度の順は、欧州の専門家による CEFR に基づくリスニングの難易度の判断の順とほぼ対応していた。ただし、12番(C1)と17番(C2)は若干の逆転が見られた。また、15番と16番はどちらも B2というレベルであるが、平均正答率および項目難易度にかかなりの幅が見られる。

表2 大問別平均正答率および IRT 項目難易度

大問番号	12	13	14	15	16	17
CEFR レベル	C1	A2	A2	B2	B2	C2
平均正答率	0.49	0.87	0.90	0.76	0.56	0.48
IRT 平均項目難易度	0.61	-0.70	-0.83	-0.22	0.41	0.60

本研究では、CEFR の A1から C2までのレベルをそれぞれ1から6に変換して数量的分析を行った。欧州の専門家の CEFR のレベル判断と日本の大学院生による CEFR のレベル判定との間には 0.92 というきわめて高い相関があることから、難易の相対的判断はかなり一致していることがわかる(表3参照)。しかしながら、それぞれのレベル判断の平均値(日本の大学院生による判断 4.13, 問題作成者側による判断 3.83)の差は 0.30 となっており、これは日本の大学院生の方が上のレベル(難しい)と判断していることを表している。

表3 2つのレベル判定度平均得点率の相関

	大学院生の判断	欧州の専門家の判断	大問別平均得点率
大学院生の判断	1.00		
欧州の専門家の判断	0.92	1.00	
平均得点率	-0.63	-0.85	1.00

また、表4を見ると、日本の大学院生の判断は、欧州の専門家の判断とはかなりずれているものもあることがわかる。特にロシア語訛りで話される15番は、日本の大学院生による判断とはぶれている。これは、CEFRでは訛りに言及しているのはC2であるため、この要素から見れば15番はC2と判断されてもおかしくはないが、そのほかの要素の多くは容易(easy)と判断されている。つまり、どの要素を中心に判断したかで、結果が異なってきていると言える。

表4 大学院生によるCEFRレベル判断 (人)

大問番号 CEFR レベル判断	12 (C1)	13 (A2)	14 (A2)	15(B2)	16(B2)	17(C2)
C2	1	0	0	<u>5</u>	0	5
C1	<u>9</u>	0	1	4	6	<u>6</u>
B2	4	2	4	<u>5</u>	<u>7</u>	3
B1	1	<u>10</u>	<u>7</u>	1	2	1
A2	0	<u>1</u>	<u>3</u>	0	0	0
A1	0	2	0	0	0	0

*グレーの網掛けは、レベル判断が一致しているもの。下線は最頻値

2.2.2 レベル判断に関する再解釈

日本人大学院生は英語の学習者でもあり、そのため、自分のレベルを軸にして、そこからの距離感を元にテキストの難易度を判断しているということが伺える。今回のテスト・データおよびこれまでのケンブリッジ英検のデータなどから、今回の回答者である日本人大学院生の大半はB2レベルであるということがわかっている。このことから、自分のレベルより上のテキストについての判断は、ある程度正確に行っているが、自分のレベルより下のテキストについては難しいと判断していることがわかる。とりわけ、A2をB1と判断しているということは、Basic UserレベルとIndependent Userレベルとの判断の違いであり、その影響は大きいと言える。

この原因について、もう少し詳細に考察してみようと思う。第2言語話者である判断者は、当然のことながら、レベル判断に際して自らの理解度をその拠り所とするであろう。自分のレベルよりも上のテキストの理解にあたっては、理解できない部分が言語的な要素を中心に分析されるが、自分のレベルより下のテキストの理解では、言語的な処理はほぼ自動的に行われているために、内容的な特徴を中心にテキストのレベル判断を行ってしまうものと

思われる。母語話者がリスニング・テキストのレベル判断する場合も、内容をもとに行うと予想されるということと相通じるものがあるかもしれない。13番と14番は、それぞれSleepy Hollowという昔話と映画俳優Nicole Kidmanの半生についての放送である。確かに言語的に考えれば、これらのテキストは、A2のI can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance.という記述に当てはまるのかもしれない。しかし、それぞれのテキスト内容面のレベル記述は、A2では、I can catch the main point in short, clear, simple messages and announcements.となっているが、B1では、I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest となっており、B1のレベル記述の方が、昔話とドキュメンタリーとはむしろ相性がいいのかもしれない。

2.3.1 CEFRリスニング・レベルと各要素、とりわけ「長さ」との関係に関する結果

大問平均正答率とCEFRに含まれるリスニングの様々な要素との相関係数(表5)からは、大問平均正答率は「言語的難易度(語彙)」「言語的難易度(文法・文構造)」「トピックの馴染みの度合い」「理解の詳細さ」「議論の複雑さ」「発話速度」との相関が高いことがわかる。これらの要素は、どのようなスクリプトの聞き取りにも関わるものであり、リスニングの難易度に恒常的に影響を及ぼすと考えられる。これに対して、「長さ」「訛り」「ノイズ」との相関係数は全般に低く、レベルの判断としては決定的な影響を与えるわけではないと言えるだろう。

表5 リスニングの要素ごとのレベル判断と、2つのCEFRのレベル判断
およびテストの平均得点率との相関

	CEFRレベル 院生判断平均	CEFRレベル もとの定義	平均得点率
言語的難易度(語彙)	0.93	0.97	-0.84
言語的難易度(文法・文構造)	0.85	0.87	-0.71
トピックの馴染みの度合い	0.88	0.88	-0.69
興味の度合い	0.65	0.64	-0.38
理解の詳細さ	0.98	0.94	-0.72
長さ	0.50	0.45	-0.55
議論の複雑さ	0.92	0.87	-0.68
発話速度	0.91	0.96	-0.83
訛り	0.60	0.32	0.12
明瞭さ	0.56	0.58	-0.41
ノイズ	0.07	0.35	-0.35

*グレーの網掛けは、±0.7以上の相関係数

このうち、「訛り」や「ノイズ」の影響は局所的だ(つまり、極端な「訛り」や無視できないほどの「ノイズ」はどのテキストにもあるわけではない)が、「長さ」はあらゆるテキストに伴う要素で

あり、もう少し全体的な影響力があってもよさそうに思える。そこで、以下では、「長さ」に関して、さらに詳細に見ていこうと思う。表2で難易度の逆転のあった12番(C1)と17番(C2)について、リスニングの要素別難易度判断を見てみる(表6参照)。12番と17番に若干であるが逆転が見られた原因としては、「長さ」の違いが考えられる。12番は長いと「感じられて」いるだけでなく、実際に、17番に比べて倍近く的时间がある。CEFRのリスニングの様々な要素のうち「長さ」以外はすべて17番のほうが困難であると判断されているにもかかわらず、12番と17番が逆転したということは、「長さ」がリスニングの難易度に大きな影響を及ぼしていると言えるだろう。

表6 リスニングの要素別難易度判断とt検定

大問番号	12	17	P 値
CEFR レベル	C1	C2	
言語的難易度 (語彙)	2.96	3.13	0.19
言語的難易度 (文法・文構造)	2.35	2.74	0.00
トピックの馴染みの度合い	3	3.3	0.15
興味の度合い	2.57	3.22	0.00
理解の詳細さ	3	3.13	0.48
長さ	3.48	2.17	0.00
議論の複雑さ	2.61	2.74	0.42
発話速度	3	3.04	0.81
訛り	1.78	1.96	0.38
明瞭さ	2.09	2.74	0.00
ノイズ	1.13	3.52	0.00

* グレーの網掛けは、t検定(両側)で0.05レベルで有意なもの

2.3.2 CEFRリスニング・レベル記述における「長さ」に関する再解釈

「長さ」は、表5の相関係数の低さから見れば、それ自体は、いくつかある要因の1つに過ぎず、レベル判断の決定的要因にはなっていないが、明らかな「長さ」の違いがある場合は、他の要素が仮に逆転しているような場合でも、長いテキストのほうを難しいと判断することが表6からわかる。また、テストの項目難易度に関しても、極端な「長さ」の違いは、明らかに難易度を押し上げると言える。おそらく、テキストの「長さ」は、聞き手の認知的負荷を高めるために、レベルの判断者にも「難しい」という印象を与えると同時に、テスト受験者のパフォーマンスにも影響を及ぼしていると考えられる。

CEFRリスニングのレベル記述においては、「長さ」は、それぞれのレベルに相当するテキスト・タイプの「本質的な長さ」を反映して記述されていると思われる。CEFRリスニングのレベル記述では、「長さ」に関する記述は、A2の“short, clear, simple messages and announcements”の部分とB2の“I can understand extended speech and lectures ...”およびC1の“I can understand extended speech ...”という部分に現れるだけであるが、これらにおいては、いずれもそれぞれのテキスト・タイプと「長さ」の関係は不可分となっ

ている。つまり、メッセージやアナウンスは元々短いものであるし、スピーチや講義は本来ある程度のまとまった長さがあるということである。おそらく、こうした意味での「長さ」によるテキスト・タイプのレベル判断はそれほどぶれないと思われるが、ナレーションやインタビューなどの一部を切り取ったものは、それがテストであれ、教材であれ、本来の長さを備えておらず、レベルの判断が厄介である。

3. 総括

本稿による、根岸(2008)の再解釈の結果をまとめる。英語のテキストに関して、CEFR のリスニングの能力記述に基づいたレベル判断を英語の非母語話者が行う場合、自分より上のレベルのものには、言語的な要素をもとに判断する傾向があるが、自分より下のレベルのものには、判断者自身の言語的な処理が自動化されているために、言語的な要素はあまり意識せず、内容的な要素をもとに判断する傾向があるようである。つまり、CEFR のリスニングのレベル記述には、言語的な要素から内容的な要素までさまざまなものが含まれているが、判断者のレベルにより、判断基準となる要素が異なっている可能性があるということである。この意味では、今回はたまたま評価者のレベルより下だったリスニング・テキストの内容がやや難しめであったために、評価者がレベルを上判断してしまったが、もしこれが言語的には難しくても内容的に容易なものであれば、また判断は異なったかもしれない。今回の評価者は日本の大学院生であったが、今後日本人英語教師が CEFR を使うときにもこの問題は考慮しなければならない。評価者がこの点に焦点を置いた訓練を受けることで、多くの要素をバランスよく利用できるようになるのかどうかは、今後の課題であろう。

テキストの「長さ」が、テストの項目難易度に及ぼす影響とレベル判断に及ぼす影響とは分けて考えたほうがよさそうである。テキストの長さが長ければ、テストの項目難易度は高くなると思われるが、レベル判断は必ずしもそれに対応したものとはなっていないようだ。CEFR のリスニングのレベル判断において、通常はテキストの長さは主要な判断基準とはなっていないが、圧倒的にどちらかが長いようだと、そのレベル判断を押し上げる要因となるかもしれない。また、本来的にはまとまった長さのあるテキストのレベルは難易度が高いが、そこから抜粋されたようなテキストの判断は難しく、判断がぶれる可能性がある。

現実のリスニングのテキストは、CEFR リスニングのレベル記述のすべての要素が手と手を取り合って上昇していくような性格のものではなく、一方のレベルが低くても、もう一方のレベルが高いというような組み合わせも頻繁にある。今後、オーセンティック・テキストをさらに分析することで、どのような要素がレベル判断として有効に機能しているかを実証的に検証していかなければならないであろう。

参考文献

- Buck, G. (2001). *Assessing Listening*. Cambridge: Cambridge University Press.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Council of Europe. *English Profile: Reference level descriptions for English*.
Available:<http://www.englishprofile.org/index.html> [2008, October]
- Horváth, K. F. and Pižorn, K. (2005). *INTO EUROPE: Listening*.
Budapest: British Council.
- Negishi, M. (2006). How much do we have in common with the Common European Framework of Reference? — The applicability of the CEFR to an IRT-based English proficiency test in Japan?, *Readings in Second Language Pedagogy and Second Language Acquisition in Japanese Context*, pp.83-100
- 中島正剛・永田真代 2006. 「CEFR の日本人外国語学習者への適応可能性」『外国語教育研究』 No.9, pp.5-24.
- 根岸雅史 2006. 「CEFR の日本人外国語学習者への適用可能性の向上に向けて」『言語情報学研究報告 14 第二言語習得理論に基づく言語教育と評価モデル』 pp.79-101.
- 根岸雅史 2008. 「CEFRリスニングレベルの決定要因を探る」『現場型リサーチと実践へのアプローチ: 英語教育・英語学習研究 金谷憲教授還暦記念論文集』 桐原書店.