

小学生向け英語教材の分析における潜在意味解析の利用可能性

— *Hi, friends!* の物語文を題材とした事例 —

Applicability of latent semantic analysis on English teaching materials used in
Japanese elementary schools: A case study of a story in *Hi, friends!*

名畑目 真 吾

Shingo NAHATAME

概要

潜在意味解析 (latent semantic analysis; LSA) とは、単語や文が表す概念間の意味的な関連度をコーパスと統計分析に基づいて算出する手法である。本稿は、潜在意味解析によって評価されるこのような意味的な関連度が、絵本など小学生向けの英語教材の特徴を表す指標の1つとして利用され得る可能性を論じたものである。まず、理論的基盤として、言語習得における意味のある文脈の重要性、教材としての絵本の活用、そして潜在意味解析の基本的な概念と活用方法を述べた。さらに、外国語活動の教材として広く用いられている *Hi, friends!* に含まれる物語文 (桃太郎) を対象として、潜在意味解析によって可能な分析、得られた結果の解釈、分析上の留意点などを実際の分析データに基づいて検討した。

キーワード：潜在意味解析，テキスト分析，児童英語，外国語活動，絵本，*Hi, friends!*

Abstract

Latent semantic analysis (LSA) uses statistical computations to measure semantic relatedness between concepts represented by words or sentences in a large corpus of text. This paper discusses the applicability of LSA on English teaching materials used in elementary schools. First, it describes the theoretical assumptions, including the importance of learning language in a meaningful context, use of picture books as teaching materials, and the basic concept and operation of LSA. It then introduces a case study where a story in *Hi, friends!* (a widely used English teaching material in Japanese elementary schools) is investigated. The study demonstrates the use of LSA to analyze the story's text and interpretation of the results, and discusses aspects to be considered in the analysis.

Keywords: latent semantic analysis, text analysis, teaching English at elementary school, picture books, *Hi, friends!*

目次

- 1 はじめに
- 2 理論的基盤
 - 2.1 言語習得における「意味のある文脈」の重要性
 - 2.2 児童英語教育における絵本の活用
 - 2.3 潜在意味解析の概要と言語理解研究への応用
 - 2.4 小学生向け英語教材の分析における潜在意味解析の利用可能性
- 3 分析事例
 - 3.1 対象
 - 3.2 方法
 - 3.3 結果と考察
- 4 まとめと示唆
- 5 潜在意味解析による教材分析の留意点と今後の展望

1 はじめに

平成 20 年の小学校学習指導要領の改訂により、公立小学校の高学年（第 5 及び第 6 学年）を対象に外国語活動が導入され（文部科学省，2008），平成 23 年度から年間 35 時間が全面的に実施されることとなった。そして，平成 28 年度中に答申が予定されている次期学習指導要領により，高学年における外国語活動の教科化，中学年（第 3 及び第 4 学年）を対象とした外国語活動の実施が事実上決定している（文部科学省，2016a）。このような動きの中で，文部科学省が中学年向けに英語絵本を作成するなど（文部科学省，2016b），小学生向けの英語教材は今後飛躍的に増加していくことが予想される。

しかし，教員は数ある教材についてそれぞれの特徴を把握し，その中から何らかの視点や基準をもって効果的と思われるものを選択しなくてはならない。また，教材作成者においても，直観や経験だけでなく，理論や客観的な根拠に基づいて教材を開発したり，その特徴を分析したりすることが望まれる。そこで本稿は，小学生向けの英語教材の特徴を明らかにするアプローチの 1 つとして，潜在意味解析と呼ばれる手法の利用可能性を論じ，実際の分析事例を示すことを目的とした。以下，その理論的基盤として，言語習得における意味のある文脈の重要性，児童英語教材としての絵本の活用，及び潜在意味解析の概要と言語理解研究への応用について述べる。

2 理論的基盤

2.1 言語習得における「意味のある文脈」の重要性

言語習得には理解可能な大量のインプット（聞くこと、読むこと）が必要であるという点は、多くの研究者において合意が得られている（白井，2012）。しかし、これは言語学習者に対して単に多くのインプットを与えればよいということではなく、単語や文が意味のある文脈の中で提示されることが重要なのである。このことはすべての言語学習者に対して同様であり、子どもが外国語を学ぶ場合もその例外ではない（アレン玉井，2011）。アレン玉井（2011）では、子どもの外国語学習において意味のある文脈とともにインプットを与える重要性を以下のように述べている。

子どもは言語を分析し、その構造を理解する力をまだ十分に持ち合わせていません。彼らは意味を想像しながら、文脈全体を理解していきます。従って、分かりやすいだろうと、1文レベルや単語レベルで言葉を教えていくと、子ども達は文脈がないので言葉の意味を想像できなくなります。1～2回のやり取りしかないような簡単なダイアログでは、子どもはどのような文脈で言語が使用されているのか想像できません。部分を積み重ねるだけで、全体が理解できるわけではありません。子どもは言語活動の内容に興味を持ち、全体の意味を理解したうえで、初めて部分に注意を向けていきます。(p.110)

この「意味のある文脈」という考え方は、ホールランゲージアプローチと呼ばれる教授法の概念に基づいている。このアプローチでは、意味は個々の単語や文にあるものではなく全体の中にあると考え、自然な環境の中で文脈を通してことばを提示し、指導していく方法をとる。たとえば、リタラシー（読み書き）指導においては、子どもたちが絵本を読むことや教室内に多くの印刷物を掲示することなどが推奨されている（アレン玉井，2010）。このようなホールランゲージアプローチに加え、コンテンツベースアプローチ（内容重視のアプローチ）、アクティビティベースアプローチ（活動重視のアプローチ）などの要素を取り入れながら、アレン玉井（2010，2011）は物語や昔話を中心とした英語授業の在り方を提案している。

絵本や物語に限らず、読むこと（リーディング）全般においても「意味のある文脈」という考え方は重要である。リーディングの処理にはボトムアップ処理とトップダウン処理の2つがあり、ボトムアップ処理には文字や単語の認識、文構造の把握などが含まれるのに対し、トップダウン処理にはテキストの主題を把握したり、背景知識や文脈を利用して内容を予測しながら読んでいくことなどが含まれる（Nuttall，2005）。つまり、1文1文を正確に理解するだけでなく、テキストを読んでまとまりのある理解を構築したり、英

文を読み進めながら理解を修正したりするには、前後の文脈を含めて意味を構築していくことが不可欠なのである(卯城, 2009)。

このように言語習得や言語能力の発達においては、「意味のある文脈」の中で言語材料を理解していくことが重要な役割を果たす。

2.2 児童英語教育における絵本の活用

ここでは、前節で述べたホールランゲージアプローチの1つの在り方である絵本の活用について、児童英語教育の観点から述べる。このアプローチの考えに基づけば、絵本は児童に対してまとまった英語を提示でき、意味のある文脈とともにインプットを与えられることにその最大の意義がある。

絵本の活用については、日本でも外国語活動導入の前後からその関心度が高まっている。現在全国的に用いられている外国語活動の共通教材である *Hi, friends! 2* (文部科学省, 2012a) においても「桃太郎」の絵本が収録されており、*Hi, friends! 2* の指導編 (文部科学省, 2012b) では、外国語活動において絵本を用いる意義として、以下のようなことが述べられている。

児童に聞かせる工夫の1つとして、絵本の読み聞かせが考えられる。絵本の絵から情報を読み取り、状況を理解しながら、児童は相手の話を聞くことになるため、「聞いてわかる」体験をさせやすい。また、選ぶ絵本の内容によって、現実には起こり得ないことを絵本の世界で体験することもできる。さらに、昔話の中には、生きていく知恵や教訓的なことが組み込まれている場合もある。このようなことを踏まえ、外国語活動でも、外国語による絵本の読み聞かせを行うことが考えられる。(p.32)

Hi, friends! 2 による指導は児童が英語を読むことを目指したものではないが、絵本の読み聞かせによって、絵などの付加情報を用いながら教師が語る物語の状況や内容全体を理解させようという目的が、上記の引用部分から示唆される。

また、畑江 (2012) はリーディングの第一歩として絵本の活用を推奨しており、音声を中心とした小学校での指導と、読むことが本格的に導入される中学校での指導の円滑な接続に絵本の活用が貢献すると述べている。そのほかにも、異文化に触れる、英語の音に慣れる、世界を広げる、自分の意見を考えさせたり表現させたりするなど、子どもの英語習得における絵本活用の様々な意義が主張されている (外山, 2010; リーパー, 2010)。

絵本活用への関心が高まるとともに、近年では絵本を用いた小学校での授業実践報告も多く行われている (e.g., 大川, 2014; 松浦・伊東, 2012)。これらの報告では、絵本活用によってリーディングへの興味・関心の向上に効果があること、学年 (発達段階) に応

じた絵本の活用方法、児童が好む英語絵本などが授業実践に基づいて述べられている。

このように、小学生に英語を教える際に絵本を活用することについて、意味のある文脈とともにインプットを与えられることのほか、様々な意義が多くの研究者や実践者によって支持されている。

2.3 潜在意味解析の概要と言語理解研究への応用

本節では、本研究で教材のテキスト分析に用いる、潜在意味解析 (latent semantic analysis; LSA) について述べる。LSA は心理学の分野で広く用いられてきた手法であり (Landauer, McNamara, Dennis, & Kintsch, 2007), 単語や文によって表される概念間の意味的な関連度を算出するものである。例えば、「大学」という単語と「学校」「入学する」「卒業する」といった単語の概念に関連があることは、直感的に理解できるだろう。LSA では、そのような意味的関連度をコーパス (書籍等の言語材料を電子化したデータベース) と統計手法に基づいて客観的に評価する。

LSA では、単語や文などの分析対象が、大規模な言語コーパスによって形成される高次元の意味空間内で数値列 (ベクトル) として表象され、それらの間のコサインが算出される。このコサインの値が概念間の距離を表し、意味的な関連度とされる。この意味的関連度は理論的には -1.0 から +1.0 までの値を取るが、実質は 0 に近いほど無関連であり、1.0 に近づくほど概念間の意味的な関連度が高いと解釈される。実際にはより複雑な計算プロセスが必要とされるが、詳細な手続きと原理は猪原・楠見 (2011) を参照されたい。

LSA の利点の 1 つは、算出される意味的関連度が比較対象に含まれる語の重複や類義・階層関係を反映しているだけでなく、同じ文脈で直接的・間接的に共起し易い語という観点から、意味的関連度を測定できることである。このことにより、単なる語の重複や類義関係よりも、さらに精緻な意味的関連度の表示が可能となる。

また、単語や文など、様々な単位を分析対象とできることも LSA の利点の 1 つである。単語どうしの意味的関連性は、その類義関係などから直感的にも判断しやすいが、文以上の単位で意味的な関連性を判断するのは難しい。しかし、LSA では単語対単語の意味的関連度だけでなく、文対文、パラグラフ対パラグラフのような相対的に大きな概念間での意味的関連度の算出が可能である。さらに、単語対文、文対パラグラフのような異なる単位間での意味的関連度の算出も可能である。

LSA のウェブアプリケーション (<http://lsa.colorado.edu/>) を用いた分析例をもとに、さらに説明を加える。このウェブ上には複数の分析ツールが含まれているが (Nearest Neighbor, Matrix Comparison, Sentence Comparison, One to Many Comparison, Pairwise Comparison), ここでは基本的な分析ツールについてのみ説明する。まず、Nearest Neighbor を使った分析では、ターゲットとなる単語と意味的関連度の高い語のリストの

提示が行われる。たとえば、表 1 は “university” をターゲットとして、それと意味的関連度が高い語を提示したものである（ここではコーパスに 100 回以上現れている高頻度語に限定）。この例では、“college” のようなターゲットの類義語だけでなく “graduated”, “graduate”, “institute”, “attended” などターゲットと同じ文脈で共起し易いと考えられる語も高い意味的関連度を持つと評価されていることが分かる。

表 1 LSA (Nearest Neighbor) による意味的関連度の分析の例

意味的関連度の値	語
0.72	harvard
0.70	college
0.64	graduated
0.56	graduate
0.56	institute
0.54	universities
0.50	attended

注. ターゲット語を “university” とした場合。

Sentence Comparison という分析ツールでは、単語よりも大きな文単位での分析が可能である。表 2 は、Bill Martin Jr. の著作として知られる絵本 *Brown bear Brown bear What do you see?* の冒頭 4 文について、各文と隣接する文の意味的関連度を算出したものである。例えば、文 1 と文 2 ではそれらの意味的関連度が 0.12 と評価されている。

表 2 LSA (Sentence Comparison) による意味的関連度の分析の例

意味的関連度	文
0.12	1: Brown Bear, Brown Bear, What do you see?
	2: I see a red bird looking at me.
0.82	3: Red Bird, Red Bird, What do you see?
0.28	4: I see a yellow duck looking at me.

その他の分析ツールでは、Matrix comparison を用いれば複数の分析対象について最大ペア数の比較を一括して行うことが可能である。また、One to Many Comparison では、ある分析対象とその他複数の対象との比較を同時に行うことができる。

このような単語や文間の意味的関連度を算出する LSA は、主に母語による文章理解研究を中心として用いられてきた。これらの研究では、LSA で評価される文どうしの意味的関連度はその読解時間や内容の記憶、一貫性に対する読み手の評価など、文章理解における人間の様々な認知活動に影響を与えていることが示されている (e.g., Todaro, Millis, & Dandotkar, 2010; Wolfe, Magliano, & Larsen, 2005)。

また、近年では第二言語 (second language; L2) 習得研究の分野においても LSA が活用されている。Crossley, Salsbury, McCarthy, and McNamara (2008) では、1 年を通じ

て収集された L2 学習者の発話データについて、発話間の意味的関連度を LSA によって分析した。その結果、学習者の L2 に対する習熟度が時間経過とともに向上するにつれて、学習者による発話間の意味的関連度の値も高くなったことを報告している。つまり、L2 に習熟するほど、前の発話と意味的に関連した発話をするようになったということである。

また、第二言語習得研究の分野では、テキスト分析への LSA の活用が多く行われている。LSA をテキスト分析に用いることで、単語や文の長さのみに基づく伝統的なリーダビリティの指標 (e.g., *Flesh-Kincaid Grade Level*) では測定できない「意味的関連度」というテキストの特徴を分析することができる。たとえば、Crossley, Louwerse, McCarthy, and McNamara (2007) は ESL (*English as a Second Language*) で用いられる初級者向けのテキストのうち、簡素化された (学習者向けに語彙や命題を調整した) テキストとオーセンティックな (学習者向けに特別な操作のされていない) テキストを比較し、前者よりも後者において隣接する文間の意味的関連度が低いことを報告している。さらに、Crossley and McNamara (2008) は中級者向けのテキストを同様の観点で分析し、ほぼ同じ結果を報告している。また、Crossley, David, and McNamara (2012) では初級、中級、上級のテキストについて、レベルが上がるごとに隣接する文間の意味的関連度が低くなることを報告している。これらのテキスト分析研究では、テキストの難易度が高まるにつれて、文間の意味的関連度が低くなるということが共通して示されている。このように、LSA によって算出される文間の意味的関連度は、L2 学習者向けのテキストの特徴やレベルを反映することが明らかにされている。

国内では、名畑目 (2012) において LSA を用いた検証が先駆的に行われている。この研究では、実用英語検定 (英検) の空所補充問題に用いられるテキストを対象とした分析を行い、英検の上位の級になるほど、空所を含む文とその前後の文との意味的関連度が低くなることを明らかにした。難易度が高いと想定される文章ほど隣接する文間の意味的関連度が低くなるという結果は、上記の Crossley らの研究と一致するものである。

このように、LSA は単語や文間の意味的関連度を算出するために用いられ、算出された関連度は、母語及び第二言語による文章理解やテキスト特徴の一側面を反映することが多くの研究によって示されてきた。

2.4 小学生向け英語教材の分析における潜在意味解析の利用可能性

2.1 節では、言語習得において意味のある文脈とともにインプットを与える重要性を述べた。それは子どもの外国語習得においても同様であり、その 1 つの方法として絵本の活用がある。2.2 節で述べたように、児童を対象とした英語指導において絵本を活用する意義は様々あるが、英語をまとまりとして提示できることがその最大の利点である。しか

し、これまでの研究は絵本を用いた指導の方法やその成果、児童の絵本に対する好みなどを報告したものが多く、絵本の特徴を客観的に分析・評価した研究はこれまでほとんど行われていないのが現状である。

英語を意味のある文脈とともに提示することが絵本を用いる大きな利点であることを踏まえると、ある絵本がどの程度意味のある文脈を提示しているのかを客観的に評価することができれば、その意義は大きいだろう。この意味のある文脈の評価方法について、その可能性の1つが2.3節で述べたLSAであると筆者は考える。LSAは単語や文が表す概念間の意味的な関連度を単語の共起情報なども含めて算出し、その関連度はテキストの一貫性や読みやすさにも関わっている。そのため、LSAによって算出される単語や文の意味的関連度に基づけば、「意味のある文脈」という観点から絵本を客観的に評価・分析できるのではないだろうか。¹

このような考えに基づいて、本研究では小学生向けの絵本に対してLSAによってどのような分析が可能か、どのように結果が解釈されるか、そしてどのような分析上の問題点や留意点があるかを、以下で事例とともに検討していくこととした。分析事例に用いる題材は、外国語活動の共通教材として全国的に用いられているという点から、*Hi, friends! 2*に含まれる物語文(桃太郎)を取り上げることとした。

3 分析事例

3.1 対象

外国語活動の共通教材 *Hi, friends! 2* の Lesson 7 に含まれる物語文「桃太郎」を分析対象とした。児童用のテキストには物語のいくつかの場面を表した絵(計22枚)と登場人物の最低限の発言(セリフ)が文字として記載されているのみであるが、教員用の指導編ではデジタル教材あるいは教員によって読み聞かせできるように、さらに多くのセリフが記載されている(図1参照)。

児童用テキストに記載されたセリフと指導編に記載されたセリフのどちらも分析対象となり得るが、LSAを実行するには分析対象にある程度の文脈量が必要であること、実際に児童が得るインプットは指導編に記載されたセリフも含めたものであることを踏まえ、本分析では指導編に記載されたセリフに基づく物語文を分析対象とした。結果として、対象となる物語全体には344語、119文が含まれた。1文あたりの平均語数は2.7語であり、非常に短い文が多く含まれていることが分かる。また、文単位ではなく発話単位(場面が切り替わるか、別の登場人物が発言をするまでのセリフを1つの発話とする)で換算すると、物語は冒頭文も含め合計で74の発話が含まれていた。物語に含まれる発話は表3に示すとおりである。

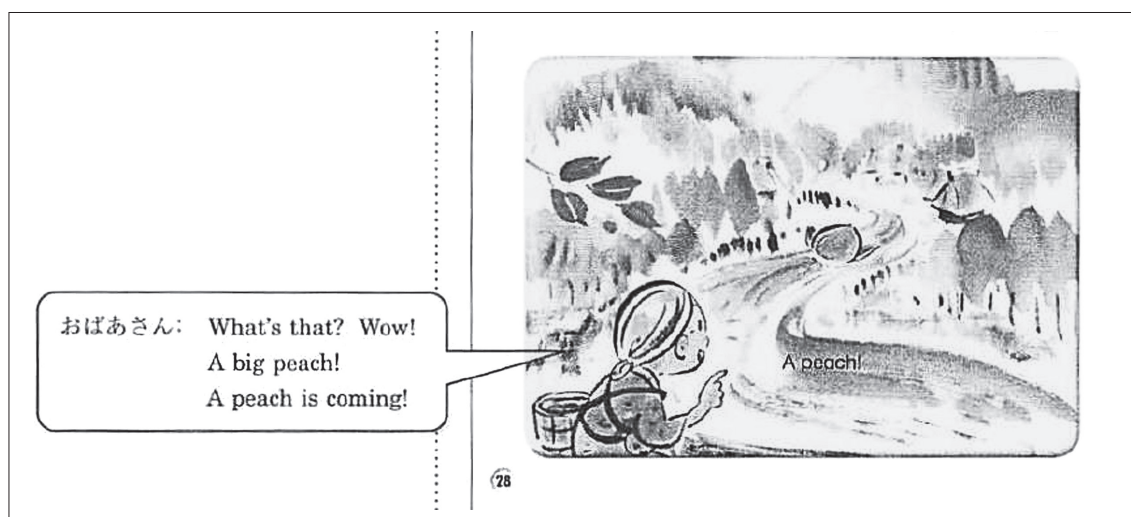


図 1. *Hi, friends! 2* 指導編 Lesson 7 の抜粋.

表 3 *Hi, friends! 2* の物語文「桃太郎」に含まれる発話

発話番号 (登場人物)	内 容
U1 (冒頭文)	An old man and an old woman lived happily.
U2 (おばあさん)	Take care. See you.
U3 (おじいさん)	See you later.
U4 (おばあさん)	What's that? Wow! A big peach! A peach is coming!
U5 (おじいさん)	Look! A boy! A little boy!
U6 (おばあさん)	Yes, yes. A boy from the peach!
U7 (桃太郎)	Hello, Grandpa and Grandma.
U8 (おじいさん)	Momotaro is a good boy.
U9 (おばあさん)	Yes, very good and cute.
U10 (赤鬼)	Arrrgh! I'm strong.
U11 (青鬼)	Arrrgh! I'm strong, too.
U12 (赤鬼 & 青鬼)	We are strong!
U13 (桃太郎)	I'm going to Onigashima.
U14 (おじいさん)	Very dangerous.
U15 (おばあさん)	Very dangerous. Please stay here.
U16 (おじいさん)	Please stay here.
U17 (おじいさん)	You are strong.
U18 (おばあさん)	Yes, you are strong and brave!
U19 (桃太郎)	Yes, I'm strong and brave!
U20 (おじいさん)	A flag. Here you are.
U21 (おばあさん)	Kibidango. Here you are.
U22 (桃太郎)	Thank you. Thank you very much, Grandpa and Grandma.
U23 (桃太郎)	I'm strong! I'm brave! Don't worry.
U24 (おばあさん)	Take care!
U25 (おじいさん)	Take care!
U26 (桃太郎)	See you soon.
U27 (猿)	Hello, Momotaro.
U28 (桃太郎)	Hi, Monkey. How are you?
U29 (猿)	I'm good. How are you?
U30 (桃太郎)	I'm very fine.
U31 (猿)	What's this?
U32 (桃太郎)	It's kibidango.
U33 (猿)	A kibidango, please.
U34 (桃太郎)	OK. Here you are.
U35 (猿)	Thank you.
U36 (桃太郎)	You're welcome. We are good friends. Let's go to Onigashima.
U37 (猿)	OK. I'm strong and brave. Let's go!

(次ページへ続く)

表 3 *Hi, friends! 2* の物語文「桃太郎」に含まれる発話 (続き)

発話番号 (登場人物)	内 容
U38 (犬)	Hello, Momotaro.
U39 (桃太郎)	Hi, Dog. How are you?
U40 (犬)	I'm good. How are you?
U41 (桃太郎)	I'm very fine.
U42 (犬)	What's this?
U43 (桃太郎)	It's kibidango.
U44 (犬)	A kibidango, please.
U45 (桃太郎)	OK. Here you are.
U46 (犬)	Thank you.
U47 (桃太郎)	You're welcome. We are good friends. Let's go to Onigashima.
U48 (犬)	OK. I'm strong and brave. Let's go!
U49 (雉)	Hello, Momotaro.
U50 (桃太郎)	Hi, Bird. How are you?
U51 (雉)	I'm good. How are you?
U52 (桃太郎)	I'm very fine.
U53 (雉)	What's this?
U54 (桃太郎)	It's kibidango.
U55 (雉)	A kibidango, please.
U56 (桃太郎)	OK. Here you are.
U57 (雉)	Thank you.
U58 (桃太郎)	You're welcome. We are good friends. Let's go to Onigashima.
U59 (雉)	OK. I'm strong and brave. Let's go!
U60 (桃太郎)	Look! Onigashima!
U61 (猿)	Oh, yes. Onigashima.
U62 (犬)	We are strong.
U63 (雉)	We are brave!
U64 (桃太郎)	Yes, we are strong and brave. And we are good friends.
U65 (桃太郎・犬・猿・雉)	We are strong and brave!
U66 (鬼たち)	We are strong!
U67 (鬼たち)	Sorry. Sorry. We are sorry.
U68 (桃太郎)	That's OK. Be good! We are good friends!
U69 (鬼たち)	Yes, yes, we are good friends.
U70 (桃太郎)	Grandpa and Grandma, we are here.
U71 (おじいさん)	Oh, Momotaro. My good boy. We are happy.
U72 (おばあさん)	Oh, Momotaro. My brave boy. We are happy.
U73 (村人たち)	We are proud of you!
U74 (桃太郎たち)	We are very happy!

3.2 方法

上記の物語における「意味のある文脈」を LSA によって分析するにあたって、まず分析対象とする単位を考える必要がある。分析単位には、単語、文、発話と複数の候補が考えられる。今回分析対象とする物語文は 1 文が非常に短く、1 文あたりが伝える意味の量が少ないことを考慮すると、意味的関連度を算出するにあたっては 1 文よりもさらに大きな単位での分析が望ましく、ここでは発話単位で分析を行うこととした。

また、発話間の意味的関連度を算出する方法にも、複数のパターンがある。たとえば、ある発話と (a) その直前の発話、あるいは (b) それまでに行われたすべての発話などの比較対象が考えられる。本稿は「小学生向けの英語物語文を LSA で分析する場合にどのような解釈が可能か」をできるだけ分かり易く示すことが主要な目的の 1 つであるため、最も解釈が明確になる「ある発話と直前の発話との意味的関連度」を算出することと

した。

具体的な分析は、LSA のウェブアプリケーションを用いて行った。対象となる物語文を発話ごとに区切り、Matrix Comparison によってすべての発話間の意味的関連度を算出した。そして、算出された値のうち、直前の発話との意味的関連度にあたる部分のみを抜粋した。分析のベースとなるコーパスは、多くの先行研究では大学レベルのものを用いているが、今回は小学生向けの英語教材が分析対象であることを踏まえ、小学校3年生レベル (General Reading up to 3rd grade) のものを用いた (コーパスの詳細は Landauer et al., 2007 を参照)。また、以下の結果の考察の過程で、ウェブアプリケーションの機能の1つである Near Neighbor (2.3 節参照) を用いた分析も行った。

3.3 結果と考察

冒頭文を除く各発話とその直前の発話の意味的関連度 ($N = 73$) の箱ひげ図を以下に示す。

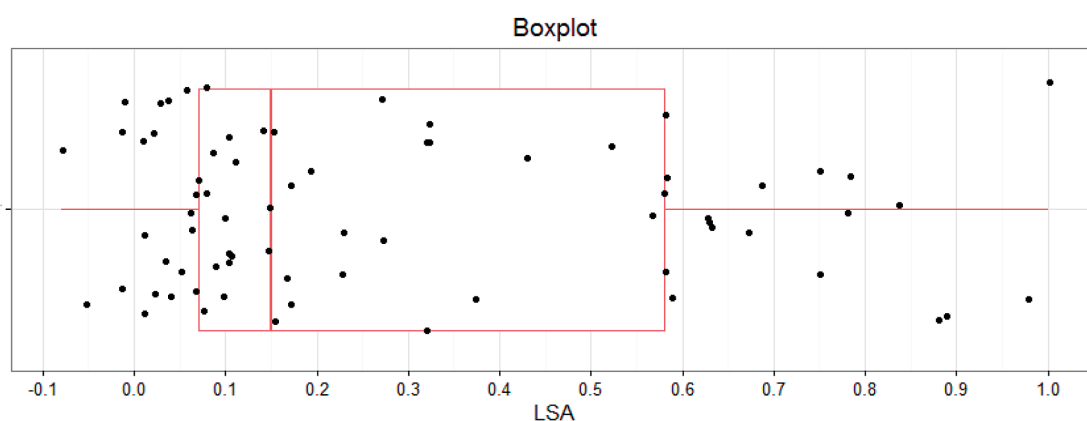


図2. 「桃太郎」における発話間の意味的関連度の分布.

発話間の意味的関連度の値は 0.10 前後に集中しているものの、1.00 まで点在しており、かなり広く分布しているといえる。つまり、直前の発話との意味的関連度は、発話によって大きく異なるということが分かる。

このデータの平均、95% 信頼区間 (Confidence Interval; CI), 標準偏差を表 4 に示す。さらに、それらの値を先行研究で報告されている値と比較するために、いくつかの先行研究についても同様の統計値を表 4 に記載した。これらと比較すると、本事例によって得られた意味的関連度の平均値は、先行研究と同じくらいかそれより高いという推定になる。たとえば、Crossley et al. (2007), Crossley and McNamara (2008) と比較すると、本事例で得られた値はオーセンティックなテキストよりも簡素化されたテキストに値が近いといえるが、さらにそれよりも高く推定されている。同様に、本事例の値は Crossley

et al. (2012) では初級者用のテキストの値に最も近いが、それよりもかなり高く推定されている。これらの先行研究では主に中学生～大学生の学習者を対象としたテキストを扱っているため、それらよりも平易な小学生向けの絵本教材において意味的関連度が高く推定されるのは、難易度が高い文章ほど意味的関連度の値が低くなるという先行研究の知見と整合する結果である (Crossley et al., 2007, 2012; Crossley & McNamara, 2008)。また、名畑目 (2012) との比較でいえば、本事例で得られた値は英検の準2級で用いられた文章で得られている値に近い。

表4 本事例と先行研究で報告されている意味的関連度の値の比較

	<i>M</i>	95% CI	<i>SD</i>
桃太郎 (本事例)	0.29	[0.22, 0.36]	0.29
Crossley et al.'s (2007) study Simplified texts	0.19	[0.17, 0.21]	0.07
Crossley et al.'s (2007) study Authentic texts	0.16	[0.13, 0.19]	0.07
Crossley and McNamara's (2008) study Simplified texts	0.21	[0.20, 0.22]	0.07
Crossley and McNamara's (2008) study Authentic texts	0.19	[0.18, 0.20]	0.07
Crossley et al.'s (2012) study Beginner texts	0.17	[0.16, 0.18]	0.07
Crossley et al.'s (2012) study Intermediate texts	0.15	[0.14, 0.16]	0.07
Crossley et al.'s (2012) study Advanced texts	0.13	[0.12, 0.14]	0.06
名畑目 (2012) 英検準2級	0.28	[0.21, 0.35]	0.13
名畑目 (2012) 英検2級	0.23	[0.17, 0.29]	0.14
名畑目 (2012) 英検準1級	0.16	[0.10, 0.22]	0.13
名畑目 (2012) 英検1級	0.16	[0.12, 0.20]	0.09

注. 先行研究の95%CIの値は、標本数と平均値、標準偏差に基づいて筆者が計算により算出したものである。また、先行研究で報告されているのは、隣接する文間の意味的関連度であることに留意されたい。

しかしながら、本事例において得られている標準偏差の値は、先行研究で得られている値と比較してかなり大きいことに留意されたい。これは、先に述べたように、得られた意味的関連度の値のばらつきが大きいことを示している。そのため、本事例における意味的関連度の高い発話や低い発話を具体的に上げ、どのような特徴があったのかを観察していくことが必要である。

まず、0.7以上のような非常に高い意味的関連度を持つものは、数は多くないもののいくつか見られている。これらを実際に見てみると、発話間で同じ単語や表現が繰り返して用いられているものが多い。たとえば、“Look! A boy! A little boy!” [U5] – “Yes, yes. A boy from the peach!” [U6] (関連度 = 0.75) や “Yes, we are strong and brave. And we are good friends.” [U64] – “We are strong and brave!” [U65] (関連度 = 0.78) がその例である。高校生や大学生など年齢の高い学習者向けの教材であれば、このような同じ単語や表現の単純な繰り返しが使われるということはあるまいだろう。そのため、小学生向けの教材でこのような高い関連度の値を持つ発話が見られることは、分析結果が教材の特徴をとらえているともいえる。

また、平均に近い値が得られた発話では、“Hello, Momotaro.” [U27] – “Hi, Monkey.

How are you?” [U28] (関連度 = 0.23) が例として挙げられる。非常に短いやりとりであり、発話間で語彙や表現の繰り返しも見られていないものの、平均に近い意味的関連度が得られていた。これは、“Hello,…” – “Hi, … How are you?” というやりとりの典型性・共起性が高いことが意味的関連度に影響しているためと推察される。この発話に類似したものとして、“Monkey”の部分が“Dog”“Bird”という他の動物の名前になっている発話があるが ([U39, U50]), それらの発話では直前の発話との意味的関連度が低くなっていた (関連度 = 0.07, 0.09)。しかしながら、“Momotaro”を含めた人物や動物の名前を除いた“Hello,…” – “Hi, … How are you?” というやりとり自体の関連度は 0.25 と推定され、上記の考察を裏付けるものである。

一方で、0.10 を下回るような意味的関連度が低い発話を見てみると、場面が大きく転換している発話であったり (e.g., “What’s that? Wow! A big peach! A peach is coming!” [U4] – “Look! A boy! A little boy!” [U5], 関連度 = 0.09), 発話だけでは唐突なやり取りになっていたり (“Yes, yes. A boy from the peach!” [U6] – “Hello, Grandpa and Grandma.” [U7], 関連度 = 0.01), 発話が短いために発話を持つ意味が限定的になっているような場合がある (e.g., “A kibidango, please.” [U33, U44, U55] – “OK. Here you are.” [U34, U45, U56], 関連度 = 0.06)。しかしながら、これらのいくつかは場面を表す絵、読み上げる教員の声のトーンや表情、ジェスチャーなどによって物語の展開が理解できるような部分である。また、今回扱っているのが直前の発言との意味的関連度であるため、それより前の部分の発言を踏まえれば、意味的関連度が高くなるような場合もあるかもしれない。つまり、意味的関連度が低い発話であっても、必ずしもそのことが問題となるわけではないことに留意されたい。

しかしながら、発話に含まれる語彙や表現に基づいて、2つの発話をより意味的に関連したものにすることは可能である。たとえば、“I’m strong! I’m brave! Don’t worry.” [U23] – “Take care!” [U24] というやり取りの意味的関連度は 0.07 であるが、その理由として、先述したように2つ目の発話が短いために発話の持つ意味が限定的になっている可能性がある。そのため、2つ目の発話を拡大し、前の発話と意味的に関連する内容を含めたい。このような場合、ウェブアプリケーションの Near neighbor という分析ツールを使えば、発話に含まれる単語と意味的関連が高い語を表示することができる。たとえば、1つ目の発話に含まれる“strong”“brave”と意味的関連が高い語 (高頻度語に限定) を表示すると、“strong”については“giant”, “muscle”, “enough”が、“brave”については“afraid”“frightened”“kill”“fight”などがある。これを踏まえて、たとえば2つ目の発話を“You have big muscle. Don’t be afraid, Momotaro. Take care!” とすると 0.42 まで意味的関連度が上がる。このような LSA の一部の機能を使った手順・工夫により、発話どうしをより意味的に関連させたものにすることが可能である。

4. まとめと示唆

本稿の目的は、小学生向けの英語教材の分析における潜在意味解析 (LSA) の利用可能性を論じ、実際の分析事例を示すことであった。まず、理論的基盤として、言語習得における「意味のある文脈」の重要性と教材としての絵本の活用、及び LSA の基本的な概念や活用法を述べた。そして、それらに基づき、LSA によって評価される単語や文が表す概念間の意味的関連度が、教材としての絵本や物語文における「意味のある文脈」の指標の1つとなる可能性を論じた。

分析事例については、外国語活動の共通教材として全国的に用いられている *Hi, friends!* に含まれる物語文 (桃太郎) を対象として、LSA によってどのような分析が可能か、どのように結果が解釈されるか、そしてどのような分析上の問題点や留意点があるかを実際の分析データに基づいて検討した。この分析事例から、小学生向けの英語教材の分析における LSA の活用法とその利点は以下のようにまとめられる。

1. 絵本や物語における発話や文間の意味的関連度を数値で客観的に評価し、特定の教材が持つ意味的関連度の高低を判断できる。
2. 特定の教材の中で発話や文間の意味的関連度のばらつき、及び意味的関連度が高い部分や低い部分を明らかにできる。
3. 教材の中で発話や文間の意味的関連度が低い部分に関しては、意味的関連度を高める単語を含めるなどして、より意味的関連度の高い文脈を提供できる可能性がある。

1 点目については、絵本や物語が提供する文脈の意味的関連度を主観的な判断ではなく、数値によって客観的に判断できるというのが LSA を使用する大きな意義である。ただし、意味的関連度の高低についての絶対的な基準はないため、特定の教材で得られた値は先行研究で得られている値や他の教材との比較などによって、その高低が判断される必要があるだろう。たとえば本事例では、桃太郎の物語文における発話間の意味的関連度の平均は、英検の1級、準1級、2級の空所補充読解問題における文間の意味的関連度よりも高く、準2級のそれに近いと判断できる。また、他の小学生向けの教材と比較を行えば、発話や文間の意味的関連度の観点から教材を選定することも可能である。

2 点目については、同じ教材の中でも文や発話によって意味的関連度の高低に幅があるかどうかを検討できる。本事例では、直前の発話との意味的関連度は発話によって大きく幅があった。このように各意味的関連度にばらつきがあった場合、値が高い部分や低い部分を詳細に観察することで、その特徴を推察することが可能である。特に意味的関連度が低い部分を特定した場合、意味的関連度が低くとも絵などの非言語情報で内容の一貫性が保たれているのか、もしくはさらに単語や表現、挿絵などを加えることで一貫性を高める

必要があるのかなどを判断することが必要になる。このような情報は、小学生向けの絵本や物語の作成や開発、改善において貴重な情報となるだろう。

3点目については、本事例で示したように、LSAのウェブアプリケーションの機能の一部を使えば、ある特定の発話や文のペアについて意味的関連度を高める単語を特定することができる。そのような単語や表現を場面に合わせて含めることで、発話・文間の意味的関連度の値を高めることが可能である。このような意味的関連度を高めるためのアプローチは、2点目で述べたような教材の開発や改善に役立つだけでなく、絵本や物語の読み聞かせなどで教師が発する教材以外の単語や表現を考える際の手助けにもなるだろう。

上記では、小学生向けの英語教材の分析におけるLSAの利用可能性をまとめたが、LSAの利用にあたっては分析上の留意点や限界点もある。この点については、今後の展望と合わせて、以下の節で詳細に述べることとする。

5. 潜在意味解析による教材分析の留意点と今後の展望

本稿で扱ったLSAによる小学生向けの英語教材の分析は、多分に探索的な性質を持つものである。そのため、今回得られた知見が今後どのように発展していくべきか、LSAを用いる際の留意点や限界点も含めながら、その展望を述べる必要がある。

まず、本事例は分析の方法や結果の解釈を端的に示すため、1つの物語文のみを分析対象としていた。先行研究で行われているような教材分析を行うならば(Crossley et al., 2007, 2012; Crossley & McNamara, 2008)、多数のテキストについて同様の分析を行い、その値をテキストの特徴ごとに比較するようなアプローチを今後検討する必要がある。たとえば、特定の学年を想定されて書かれた物語文や、レベル別のシリーズ教材(Graded readers)などをその対象にできるであろう。このようなアプローチによって、今回用いた*Hi, friends!*の物語文が持つ文脈の意味的関連度が、他の小学生向けの英語教材と比較してどの程度のものなのかを特定できる。

また、今回用いた*Hi, friends!*の物語文で扱われている単語や表現、文はかなり平易なものである。具体的には、1文が非常に短く、その中に日本語の固有名詞も頻繁に含まれる。*Hi, friends!*の物語文を対象とした本事例における意味的関連度の平均値は、中学校～大学生までの英語テキストを対象とした先行研究で報告されている値よりも概ね高く推定されていたが、名畑目(2012)における英検の準2級を対象とした値とは同程度に推定されていた。文章の難易度や対象となる学習者の発達段階を踏まえれば、本事例では準2級よりも高い意味的関連度が得られることが想定されるが、実際はそうならなかった。その理由の1つとして、上記のように*Hi, friends!*の物語文では言語材料が簡素化されすぎており、一部の発話では意味内容が限定的になっていたために想定よりも関連度が低く

なったことが考えられる。LSAを最大限に活用するのであれば、文や発話を持つ情報量がある程度豊かな物語文を対象とすることが望ましいのかもしれない。

さらに、今回行ったLSAの算出方法はあくまで一例であることに留意されたい。本稿では、先述のように、LSAによるテキスト分析の結果がどのように解釈されるかを端的に示すために「ある発話と直前の発話との意味的関連度」を算出していた。しかし、隣接する発話だけではなく、離れた発話とどのくらいの意味的関連度があるのか、発話が含まれる場面全体とどのくらいの意味的関連度があるのかも重要な指標となり得る。隣接する文・発話のような局所的な意味的関連度だけでなく、離れた文・発話のような大局的な関連度も評価対象とすることで、ホールランゲージアプローチに基づく「意味のある文脈」という考え方をより適切に反映できるのではないか。今後、そのような分析方法の違いによって、算出される意味的関連度がどのように異なるかの検討が必要である。

最後に、LSA自体の限界点として、その分析には文の統語構造や接続詞で表される接続関係などは考慮されていないことに留意されたい。つまり、LSAによって算出される意味的関連度は、あくまでテキスト特徴の一側面を反映しているにすぎない。統語構造や接続関係なども含めた多面的なテキスト特徴を分析対象とする場合は、Coh-Matrix (McNamara, Graesser, McCarthy, & Cai, 2014) と呼ばれるコンピューターツールを用いてその分析が可能である。今後、小学生向けの英語教材を対象として、このような多面的なテキスト分析のアプローチを行っていく予定である。

謝辞

本研究はJSPS科研費(研究活動スタート支援)JP15H06571の助成を受けたものである。

注

1. アレン玉井 (2011) では、言語特徴としてだけではなく、子どもたちの背景知識や興味なども含めて「意味のある文脈」という表現を使用し、その重要性を説いている。本研究で言及している「意味のある文脈」とは、そのような様々な側面のうち、言語特徴としての文脈に特化して述べているものである。

引用文献

- アレン玉井, 『ストーリーと活動を中心とした小学校英語』, 東京, 小学館集英社プロダクション, 2011
- アレン玉井, 『小学校英語の教育法—理論と実践』, 東京, 大修館書店, 2010
- 猪原敬介・楠見孝, “潜在意味解析に基づく概念間類似度の心理学的妥当性: 言語統計解析アプローチの効用と限界”, 心理学評論, 54 巻, 2011, pp.101-122
- 卯城祐司, 『英語リーディングの科学』, 東京, 研究社, 2009
- 大川陽子, “小学校英語活動における英語絵本の活用に関する研究—児童の発達段階に応じた英語絵本の活用—”, 鳴門教育大学小学校英語教育センター紀要, 5 巻, 2014, pp.31-40
- 白井恭弘, 『英語教師のための第二言語習得論入門』, 東京, 大修館書店, 2012
- 外山節子, 『英語の絵本活用マニュアル』, 東京, コスモピア, 2010
- 名畑目真吾, “Latent Semantic Analysis (LSA) による空所補充型読解テストの解明: 文レベルの意味的関連度を観点として”, EIKEN BULLETIN (英検研究助成報告書), 24 巻, 2012, pp.42-58
- 畑江美佳, “小学校外国語活動における「読む」ことへの第一歩としての絵本の活用”, 『融合文化研究』, 18 巻, 2012, pp.2-13
- 松浦友里・伊東英, “小学校外国語活動における英語絵本の導入効果に関する実践研究”, 岐阜大学カリキュラム開発研究, 29 巻, 2012, pp.94-101
- 文部科学省, 『小学校学習指導要領解説 外国語活動編』, 東京, 東洋館出版, 2008
- 文部科学省, 『Hi, friends! 2』, 東京, 東京書籍, 2012a
- 文部科学省, 『Hi, friends! 2 指導編』, 東京, 東京書籍, 2012b
- 文部科学省, “次期学習指導要領等に向けたこれまでの審議のまとめについて (報告)”, 文部科学省, 2016a, 入手先 <http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo3/004/gaiyou/1377051.htm>, (参照 2016-10-22)
- 文部科学省, “小学校の新たな外国語教育における補助教材 (Hi, friends! Story Books) 作成について (第 3・4 学年用)”, 文部科学省, 2016b, 入手先 <http://www.mext.go.jp/a_menu/kokusai/gaikokugo/1362148.htm>, (参照 2016-10-22)
- リーパー・すみ子, 『アメリカの小学校では絵本で英語を教えている: 英語が話せない子どものための英語習得プログラム ガイデッド・リーディング編』, 東京, 径書房, 2011
- Crossley, S.A., David, A., & McNamara, D.S. “Text simplification and comprehensible input: A case for an intuitive approach” *Language Teaching Research*, 16, 2012, pp.89-108.
- Crossley, S.A., Louwse, M., McCarthy, P.M., & McNamara, D.S. “A linguistic analysis of simplified and authentic texts” *The Modern Language Journal*, 91, 2007, pp.15-30.
- Crossley, S.A., & McNamara, D.S. “Assessing second language reading texts at the intermediate level: An approximate replication of Crossley, Louwse, McCarthy, and McNamara (2007).” *Language Teaching*, 41, 2008, pp.229-409.
- Crossley, S. A., Salsbury, T., McCarthy, P. M., & McNamara, D. S. “LSA as a measure of second language natural discourse.” *Proceedings of the 30th annual conference of the Cognitive Science Society*, 2008, pp.1906-1911.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. *Handbook of Latent Seman-*

- tic Analysis*. New York, NY: Routledge, 2007
- McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. *Automated evaluation of text and discourse with Coh-Metrix*. New York, NY: Cambridge University Press, 2014
- Nuttall, C. *Teaching reading skills in a foreign language*. Oxford, UK: Macmillan Education, 2005
- Todaro, S., Millis, K., & Dandotkar, S. "The impact of semantic and causal relatedness and reading skill on standards of coherence" *Discourse Processes*, 47, 2010, pp.421-446.
- Wolfe, M. B. W., Magliano, J. P., & Larsen, B. "Causal and semantic relatedness in discourse understanding and representation." *Discourse Processes*, 39, 2005, pp.165-187.