

日英対訳絵本の語彙から見た 日本語フレームネットと英語フレームネット

小原 京子*

ohara@hc.st.keio.ac.jp

*慶應義塾大学・理研 AIP

大久保佳子**

ookubo@jsa.co.jp

**日本システムアプリケーション

1. はじめに

本稿では、日英対訳絵本を題材に、日本語フレームネット (Japanese FrameNet, JFN) と英語フレームネット (FrameNet, FN) 上で認知的観点から見た基本語・基本概念が網羅されているかを見るために両者の語彙カバー率 (JFN/FN 上の意味フレームに語彙がどの程度登録済みか) と意味フレームカバー率 (語の文脈に適した意味・用法が既に意味フレームとして JFN/FN 上に定義済みか) を調査し比較する。また、日英対訳絵本上で対応する日本語と英語の語彙が喚起する意味フレームの一致・不一致について考察を行う。

幼児を対象として書かれた対訳絵本のテキストを分析することにより、認知的な観点からの基本概念が各言語においてどのような語彙・構文で表現されるかをみることが出来る。また、そこでみられる両言語の意味フレームや構文の不一致は両言語の基本的な違いを反映していると考えられる。ただし、本稿では語彙の分析を対象とし、文レベルの構文については議論しない。

日本語フレームネットプロジェクトでは、近年 BCCWJ コーパスの書籍・白書のジャンルのコアデータに対して全文テキストアノテーションを行ってきた。BCCWJ 書籍コアデータに出現する自立語 (動詞・形容詞・形状詞・名詞・副詞・接続詞) の各々の文における語義に対する意味フレームカバー率を調査したところ、延べ語数比で 81.3%であった (小原 2011)。

小原・大久保 (2018) では、BCCWJ 書籍・白書コアデータを対象とした全文テキストアノテーション結果の評価として、日英対訳絵本の日本語版 (原文) テキストに出現する日本語の自立語に関する JFN の語彙カバー率と意味フレームカバー率を調査し、今後の JFN データベース拡張に結びつけることを目指した。

本稿では、小原・大久保 (2018) で用いた日英対訳絵本の英語版対訳テキストに出現する英語の自立語に関する FN の語彙カバー率と意味フレームカバー率を、小原・大久保 (2018) と同様の方法で調査し、日本語のそれらと比較した結果を報告する。FN の語彙カバー率と意味フレームカバー率を調査した研究は本稿が初めてである。調査の結果、小原・大久保 (2018) の JFN に関する指摘と同様に、FN も語彙カバー率については改善の余地があることが判明した。また、対応する日英語の語句が喚起する意味フレームの一致・不一致を考察することで、アノテーションの難易度がある程度選別することが可能であることが明らかとなった。

第2節では日英対訳絵本に出現した日英の語彙が、認知的な観点から見た基本語であるかどうかを見るため、それらの語の難易度を検討する。第3節では JFN と FN の語彙カバー率と意味フレームカバー率を比較する。第4節では、日英対訳絵本上の対応する日本語と英語の語句のペアが喚起する意味フレームの一致・不一致について考察を行う。

分析の対象とした日英対訳絵本は『英語対訳つき とべ! アンパンマン1』である。漫画形式の構成で、各コマのセリフに対応する英語対訳がついている (やなせ 1991)。ストーリーは 23 話あり、すべての会話数 (コマ数に相当) は 652 であった。

2. 日英対訳絵本に出現する日英語彙の難易度

今回用いた日英対訳絵本の日本語原文テキストと英語翻訳テキストに出現する用言が認知的な観点から見て基本語とみなされるかを調べるため、それらの語の難易度を調査した。ここで用

言とは、日本語に関しては形状詞・形容詞・動詞・名詞・形状詞可能・名詞・サ変可能・名詞・サ変形状詞可能、英語は be 動詞以外の動詞と形容詞である。

日本語は、『新教育基本語彙』(阪本一郎 1984)を採用した。教育基本語彙として挙げられるものは多くあるが、語数がきわめて多いことが採用の理由である。日本語原文テキストに出現する用言の85% (異なり語¹) が小学校3年生以下で習得する語彙であり、日本語原文テキストに出現したが『新阪本教育基本語彙』に掲載されていない用言は全体の7%であった。

次に、英語について調査を行った。英語に関しては日本語の『新教育基本語彙』のような大規模な語彙リストが調査期間内に入手できなかったため、*Collins First Dictionary* を基本的な指標として用いることとした。また、補助資料として、CDI、Readingrockets Basic Spelling List、Flocabulary の kindergarden~grade6 までのリスト、cehs.unl.edu の Vocabulary List (For preschool or school age)なども合わせて参照した。

Collins First Dictionary は、「4歳以上対象」と明記されている通り就学前の子供向けの辞書である。その辞書においては、英語翻訳テキストに出現する71.1% (異なり語)の語が載っていた。さらに、*Collins First Dictionary* 以外の上記の語彙リスト (いずれも小学生レベルの語彙リスト)のみに掲載のある語が18語あり、*Collins First Dictionary* にのみに掲載された語と合わせると英語翻訳テキストに出現する用言の79.4%がこれらに記載されていた。日本語に比べるとやや低い数値ではあるが、指標となる *Collins First Dictionary* の対象年齢は、4歳以上就学前の子供向けであること、それに対して日本語の『新教育基本語彙』の最低基準であるAは小学生レベルであること、また両者の登録用語数の差 (*Collins First Dictionary* の登録語彙数が、2,000語強に対し、「新教育基本語彙」は20,000語を超えている)も大きいことから、英語翻訳テキストに出現する語彙も日本語原文テキスト上の日本語語彙と同程度の難易度と言える。つまり、日本語原文テキスト・英語翻訳テキストに出現する語は第一

¹ 以下で言及する語数・割合は、特に記載がない限り全て延語ではなく異なり語とする。

言語習得的観点から見て基本語であると言えると考えられる。

ちなみに、日英対訳絵本テキストに出現した用言のうち、『新教育基本語彙』/*Collins First Dictionary* に掲載のなかった語を見てみると、日本語では「うごかせる」「とべる」などの、可能動詞が多くみられた。英語では“guarantee,” “vanquish,” “deceive”などのように難易度が高いために掲載されていないと思われる語の他にも、“meaningless,” “unlucky,” “no-win,” “overexcited”といった語形成が複雑なために掲載がないと推察される語も見られた。

3. JFNとFNの語彙カバー率・意味フレームカバー率

日本語原文テキストと英語翻訳テキストに出現する用言の、語彙カバー率と意味フレームカバー率との調査結果は表1の通りであった。

	JFN (%)	FN (%)
語彙カバー率	38.4	74.9
意味フレームカバー率	86.2	89.6

表1 JFN・FNの語彙カバー率・意味フレームカバー率 (用言)

表1の語彙カバー率と意味フレームカバー率を、形容詞と動詞に分けて記載したのが表2である。日英の比較がしやすいように、日本語の形容詞には、形状詞 (例: だいじょうぶ、たいへん)、名詞 (形状詞可能) (例: せいけつ、ひま)、形状詞としての使われ方であったため名詞 (サ変・形状詞可能) (例: あんしん、じゃま) も含めた。動詞には名詞 (サ変可能) (例: しゅじゅつ、せんたく) も含めた。

	JFN/ FNに 登録さ れている 語数	JFN/ FNに 意味フ レーム は定義 済みだ が、語 彙は未 登録の 語数	異なり 語数	語彙カ バー率	意味フ レーム カバ ー率
形容詞 (日)	23	37	68	33.8%	88.2%

形容詞 (英)	64	19	94	68.1%	88.3%
動詞 (日)	66	74	164	40.2%	85.4%
動詞 (英)	109	15	137	79.6%	90.5%

表 2 JFN・FN の語彙カバー率・意味フレームカバー率 (用言・品詞別)

意味フレームカバー率については、やや JFN の方が低い傾向にはあるものの、押しなべて高い。ただし、FN において “mysterious,” “play” などの基本語が喚起する意味フレームが未だ定義されていないということも今回の調査で判明した。

意味フレームカバー率が高いのに対し、語彙カバー率は、FN においても 80% を超える数値は出なかった。特に形容詞の語彙カバー率は 70% を割り込んでいる。例えば、Aesthetic フレームに “cute” が登録されていない、などである。FN の語彙カバー率と意味フレームカバー率を調査した研究は本稿が初めてであるが、今後は FN の語彙カバー率が低い理由とその改善策について検討する必要があることが明らかとなった。

4. 日本語原文テキストと英語翻訳テキストにおける意味フレームの一致・不一致

日本語原文テキストと英語翻訳テキスト上で対応する語句が喚起するフレームが完全に一致した意味フレームは、226 意味フレームのうち 27 意味フレームにとどまった。どの対応する文のペアにおいても意味フレームが一致した例としては、dark & 「まっくら」と “dark” がいずれも Level_of_light フレーム、「こわれる」と “break” が Render_no_functional フレーム、「すむ」、「くらす」と “live” が Residence フレームなどがあつた。

完全に意味フレームが一致した件数が低かった要因としては、日本語 1 語に対応する英語の後の豊富さが考えられる。例えば、日本語原文テキストの「きれいなみずうみ」の「きれいな」が英語翻訳テキストでは “What a clean, pretty lake!” の “clean” と “pretty” の 2 語に対応する、などである。

5. 終わりに

本稿では、日英対訳絵本を題材に、JFN と FN 上で認知的観点から見た基本語・基本概念が網羅されているかを見るために両者の語彙カバー率と

意味フレームカバー率を調査した。また、対応する日本語と英語の語が喚起する意味フレームの一致・不一致について考察を行った。

JFN、FN 共に意味フレームカバー率が高いのに対し、語彙カバー率が低いことが判明したため、今後は新たな意味フレームを定義するよりも今回の調査で明らかとなった未登録語を該当意味フレームに登録するべきである。

また、1000 程度定義済みの意味フレームのうち、今回のテキストに出現する語が関与していた意味フレーム数は日本語・JFN で 218、英語・FN で 282 であつた。これらは JFN と FN の規模を拡大するために集中すべき意味フレームということもできる。

JFN 構築に関しては、今回完全でなくとも日英で一致した意味フレームに関与する用言のアノテーションはアノテータやクラウドワーカーにとって比較的難易度が低いと言えるのではと考えられる。

主要参考文献

- 小原京子 (2011). 「BCCWJ への日本語フレームネットの意味アノテーション」, 『現代日本語書き言葉均衡コーパス』完成記念講演会, pp. 371-376.
- 小原京子・大久保佳子 (2018). 「日英対訳絵本の語彙から見た日本語フレームネットの評価」, 『言語処理学会第 24 回年次大会予稿集』 pp. 1112-1114.
- 樺島忠夫 (1955). 「類別した品詞の比率に見られる規則性」『国語国文』24 巻 6 号, pp.385-387.
- 樺島忠夫 (1963). 『表現論——ことばと言語行動』綜芸舎.
- 国立国語研究所 (2001). 「教育基本語彙の基本的研究-教育基本語彙データベースの作成-」
- 阪本一郎 (1984) 「新教育基本語彙」
- 山崎誠(2014). 「言語単位と文の長さが品詞比率に与える影響」 第五回コーパス日本語学ワークショップ予稿集, pp.233-242.
- やなせたかし・たまきゆりこ訳(1991). 『英語対訳つき とべ! アンパンマン 1』フレーベル館.
- Collins First Dictionary (2014). Harpercollins Pub Ltd.

Flocabulary kindergarden ~ grade6
<https://www.flocabulary.com/>

Readingrockets Basic Spelling List
<http://www.readingrockets.org/article/basic-spelling-vocabulary-list>

Vocabulary List (For preschool or school age)
<https://cehs.unl.edu/documents/secd/aac/vocablists/VLN1.pdf>