

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

小原京子(慶應大・理研AIP)・大久保佳子(日本システムアプリケーション)



日本語フレームネット

- **日本語フレームネット (Japanese FrameNet, JFN)** (Ohara 2014)
- **意味フレーム** (特定の状況・事物・事象などを、各々に関わる**意味的要素**とともに記述した概念構造)で語彙を記述した日本語語彙情報資源。
- **英語フレームネット (FN)**と共同研究: データベースを移植。構造・手法・ツール共通。
- **目的**
BCCWJ書籍・白書コアデータへの全文テキストアノテーション結果の評価
- 現代日本語の基本語をカバーできているか? → JFNデータベースの拡張へ
- **方法**
日英対訳絵本中の自立語(用言)に関するJFNの**意味フレームカバー率**(当該語彙に相当する意味フレームが定義済みか)・**語彙カバー率**(当該意味フレーム(語彙)に語が登録済みか)の調査 → **FN**と比較
- **仮説**
 - 幼児対象の絵本は認知的な意味で基本的語彙のみで構成
 - 第一言語習得の観点から基本的語彙・意味が抽出可能

調査結果

	JFN 上意味フレームに登録済み語数	該当意味フレームは存在するが未登録の語数	語彙カバー率	意味フレームカバー率
形状詞	3	7	23.1	76.9
形容詞	15	19	40.5	91.9
動詞	55	60	41	85.8
形状詞可能	4	9	26.7	86.7
サ変可能	11	14	36.7	83.3
サ変形状詞可能	1	2	33.3	100
合計	89	111	38.4	86.2

JFN 意味フレームへの品詞別登録状況・語彙カバー率・意味フレームカバー率
語数: 異なり語数

考察1

● **意味フレームカバー率**
基本的語彙において、
小原(2011)の結果よりも**向上**。

● **語彙カバー率**
基本的語彙において、
極端に**低い**。
語彙数を増やすことがJFNの課題。

● **品詞別評価**
形状詞 (例: だいじょうぶ、たいへん、とうめい)
意味フレームカバー率・語彙カバー率ともに 低め。
拡張必要。

調査方法

- 使用テキスト: 『英語対訳つき とべ! アンパンマン1』(やなせ1991)



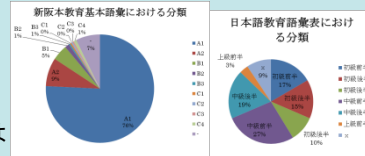
- ストーリー: 23
- 会話数: 652
- 日本語形態素数: 延べ2998
- 1会話あたり形態素数: 約5
- **調査手順**
 - 1) テキスト化
 - 2) 形態素解析(人手見直し含む)
 - 3) 意味フレーム付与 (JFNのアノテーション未経験者。全自立語に対して)
 - 4) 意味フレーム見直し (JFNのアノテーション経験者。全自立語に対して)
 - 5) 意味フレーム/語彙カバー率集計 (用言のみ抽出し、集計)

	延べ語数	異なり語数
形状詞	27	13
形容詞	111	37
動詞	456	134
形状詞可能	28	15
サ変可能	48	30
サ変形状詞可能	6	3
合計	676	232

調査対象用言の内訳と数

- **出現した自立語の例**
(原文)いきているパンができますように
⇒「いきる」「パン」「できる」

- **文章の難易度**
小学校1年~3年までに
習得する用語がほとんど
⇒ 基本的な語彙で構成
★ 日本語教育的観点からは
初級とは言えない



考察2

- **英語FNとの比較**
3話(79文)での比較
- **語彙カバー率**
FN英語 **88.4%**
JFN日本語 **44.8%**

* ただし、延べ語数。用言以外も含む比較。

	英語・語数	JFN上 英語割合%	日本語・語数	JFN上 日本語割合%
語彙カバー率	115	88.5	56	44.8
意味フレームカバー率	121	93.1	123	98.4
意味フレーム非カバー率	9	6.9	2	1.6
合計	130		125	

日英対訳絵本3話における比較

語数: 延べ語数

- **JFNアノテーションデータベースの統合**
BCCWJ書籍・白書コア(unidicベース)へのアノテーション
+ 青空文庫・京大コーパス(ipadicベース)へのアノテーション 要検討

今後の展望

- JFNデータベースの拡張方法
- 意味フレームカバー率が高いため、意味フレームを増やさずとも今回判明した未登録語をJFNの該当意味フレームに登録することで語彙カバー率向上可能
- クラウドソーシングによる、意味フレーム名付与済み例文増加(河原他、NLP2018)
- JFNアノテーションの難易度調査
- 調査手順3)における誤り分析 → JFNアノテーション(意味フレーム名付与)の難易度を明らかにする