

日本語言い換えデータベースの構築と言語的個人性変換への応用

水上 雅博 Graham Neubig Sakriani Sakti 戸田 智基 中村 哲
 奈良先端科学技術大学院大学 情報科学研究科

{masahiro-mi, neubig, ssakti, tomoki, s-nakamura}@is.naist.jp

1 まえがき

意味が近似的に等価な言語表現の異形を言い換えという。言い換えは機械翻訳や質問応答、文章要約など様々な応用分野に貢献する横断的な技術であり、言い換えを生成する様々な手法が提案されている [13]。

特に、対訳コーパスを用いた言い換えの抽出 [2] は近年注目されている。この手法は機械翻訳に用いられるフレーズテーブルを利用したピボット方式の生成手法であり、中間意味表現として他言語のフレーズを利用するという特徴がある。これにより、本来問題となる中間意味表現の設計を行う必要がなく、また生成される言い換えは単語列の単純な置き換えに限定されるため、利用しやすい。更にフレーズテーブルに与えられた翻訳モデル確率を利用して、言い換えを行う際の尤もらしさを考慮できる。また、英語とスペイン語における言い換えデータベース [3] の構築手法にも利用されている。

本稿では、対訳コーパスを用いた言い換えの抽出を用いて日本語の言い換えデータベース（以下、PPDBと呼ぶ）を構築する。構築した日本語 PPDB に対する内的評価として、主観評価を用いて言い換えの妥当性を検証する。また、日本語 PPDB の外的評価として、日本語 PPDB が実際に自然言語処理の応用に利用できるか検証する。応用例として、我々の提案する話し言葉における言語的個人性の変換 [12]（以下、言語的個人性変換と呼ぶ）を対象とし、日本語 PPDB を適用する。言語的個人性変換では、今まで単語ベースの言い換えを用いて目標話者の個人性を再現を試みてきた。しかしながら、日本語 PPDB を利用する場合は単語ベースではなくフレーズベースでの言い換えが求められる。そこで、本稿ではフレーズベースの言い換えを可能とする言語的個人性変換のモデル化について述べる。また、日本語 PPDB を用いた言語的個人性変換の結果から、言語的個人性変換における日本語 PPDB の有効性を示す。

2 PPDB の構築

2.1 対訳コーパスからの言い換え抽出

先行研究 [3] では対訳コーパスから言い換えを抽出する手法を用いて英語とスペイン語の言い換えデータベースを構築した。対訳コーパスを用いた言い換え抽出では、言い換えを生成したい言語（英語）に対して他言語（フランス語）を中間意味表現として利用し、ピボット方式で言い換えを生成する。本稿では日本語の言い換えを抽出するため、中間意味表現として英語

を用いる。言い換え（日本語）と中間意味表現（英語）の例を図 1 に示す。例では、二つの対訳文から共通の中間意味表現（ここでは“translation”という単語）を経由し、「訳」と「翻訳」の言い換えを抽出する。

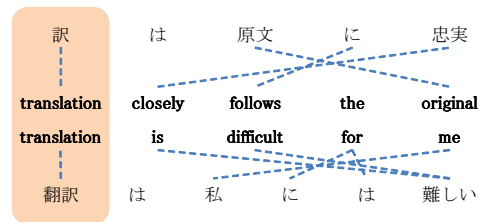


図 1: 英語をピボットとした言い換えの抽出

また、これらの言い換えが起きえる確率（以下、言い換え確率という）は、ある日本語の形態素列 j_1 が中間意味表現である英語の単語列 e に翻訳され、さらにそれが別の日本語の形態素列 j_2 に翻訳される確率であると解釈できる。ただし、全ての中間意味表現の可能性を考慮する必要がある。これは、以下のように、 e に対して周辺化をした形で定式化できる [2]。

$$P(j_2|j_1) = \sum_e P(j_2|e)P(e|j_1) \quad (1)$$

2.2 抽出された言い換え

言い換え抽出に用いた対訳コーパスの緒元を表 1 に示す。

表 1: 対訳コーパスの緒元

使用コーパス	Wikipedia, 講演, 新聞, 雑誌, 対話等を含む対訳コーパス
単語数	24.2M (en) 29.6M (ja)
対訳フレーズ数	67.1M
フレーズ最大長	7 単語
アライメント	Nile[9]
形態素解析	Kytea[8]

対訳コーパスからの言い換え抽出では、共通の中間意味表現を持つすべての言い換えを抽出する。この時、抽出された言い換えの確かさは、基本的に言い換え確率の高さと比例する。しかしながら、語尾やフィルター

といった、多様な言い換えが可能のために個々の言い換え確率が下がる言い換えも存在するため、一概に言い換え確率の低い言い換えが間違いと断定できない。そこで、本稿では言い換への抽出精度と言い換への多様性を両立するために、抽出された言い換を採用する条件を以下のいずれかを満たすとした。

- ある j_1 に対して言い換え確率 $P(j_2|j_1)$ の上位 20 組
- 言い換え確率 $P(j_2|j_1)$ が 1%以上

言い換え抽出によって 47.2M 組の言い換を含む、日本語 PPDB を構築することができた¹。表 2 に抽出された言い換の一部を挙げる。

表 2: “翻訳された”に対する言い換の一部 (j_1 = “翻訳された”)

j_2	$P(j_2 j_1)$	$P(j_1 j_2)$
翻訳された	0.083	0.083
に 翻訳された	0.034	0.074
翻訳	0.012	0.0003
共訳	0.011	0.026
訳される	0.011	0.012
と 訳された	0.002	0.004
他 20 件		

3 言語的個人性変換

3.1 変換のモデル化

言語的個人性変換 [12] とは、与えられたテキストを目的の話者の個人性を持ったテキストへと変換する手法である。基本となる変換手法は、機械翻訳の代表的なモデルである雑音のある通信路モデルを用いた話し言葉の整形 [7] を利用しており、変換の対象となる話し言葉のテキスト V から、変換先の目標となる個人性を持つテキスト W への変換を、両テキスト間の翻訳処理とみなして処理を行う。

しかしながら、個人性変換では、一般的な統計的機械翻訳と異なり、同じ意味で別の話者の個人性を持つ特殊な対訳コーパスを必要とする。そのような対訳コーパスを収集するのは容易ではなく、対訳コーパスから翻訳モデルを学習するのは困難である。この問題に対して、先行研究 [12] では、類義語を集めた辞書と n-gram 類似度から計算された擬似的な翻訳モデル確率を用いて、翻訳モデルを構築している。

個人性変換のモデル化では、十分量確保可能な目標話者の個人性を持つ言語モデル確率 $P(W)$ と、類義語とその類似度から構築した擬似的な翻訳モデル確率 $P(V|W)$ の二つを用いて、事後確率 $P(W|V)$ を以下のようにモデル化する。

$$P(W|V) = \frac{P(V|W)P(W)}{P(V)} \quad (2)$$

¹<http://ahclab.naist.jp/resource/jppdb>

与えられた V に対して $P(W|V)$ が最大となる \hat{W} を探索する。 $P(V)$ は W の選択によらず変動しないため、以下のように表せる。

$$\hat{W} = \underset{W}{\operatorname{argmax}} P(V|W)P(W) \quad (3)$$

先行研究 [12] で用いた類義語辞書と n-gram 類似度から計算された擬似的な翻訳モデル確率による翻訳モデルは、個人性を変換する上で十分効果があった。しかしながら、これらの擬似的な翻訳モデルでは、類義語辞書は高品質な人手による辞書か、言い換えよりも意味的な類義語を集めた既存の言語資源から構築した辞書、言い換えに対する裏付けの薄い自動構築された辞書しか利用できなかった。また、n-gram 類似度から計算された擬似的な翻訳モデル確率も表層的な単語の類似度を評価するのみで、意味的な言い換え可能性を評価していなかった。

そこで、本稿では翻訳モデルの代用として日本語 PPDB を利用する。2 章で述べたように、日本語 PPDB からは中間意味表現に則った高品質な言い換えと言い換え確率を得られるため、言い換への尤もらしさを考慮して変換を行うことができ、先行研究に比べてより高い変換精度を期待できる。

3.2 日本語 PPDB の利用

3.2.1 フレーズベース変換の導入

先行研究 [12] では、変換はある単語を別の単語に置き換える単語ベースで行われていた。しかしながら、日本語 PPDB は単語列の言い換えに対応しているため、変換ではある単語列を別の単語列に置き換えるフレーズベースの処理が必要となる。そこで、本研究では雑音のある通信路モデルによる変換をフレーズベースの処理に対応する。具体的には、ラティス構造の探索グラフを作成し、各単語列の言い換えに対して $P(W|V)$ が最大となる経路をビタビアルゴリズムを用いて探索する。今回はフレーズとして認識する単語列の最大長を 5 単語とした。

3.2.2 ワードペナルティの導入

フレーズベース変換のように異なる単語数を持つ変換候補を生成し得る場合、言語モデル確率を最大化するため変換前より小さい単語数の変換候補を選択し、不当に短い変換が行われることがしばしばある。これを防ぐために、機械翻訳では出力単語数に比例してスコアに加算されるワードペナルティを導入し、文章の単語数を制御している。本研究でも同様に変換モデルにワードペナルティを導入する。本来、文章の単語数は言語的特徴であり、個人性とは深い関係があると予測される [6]、しかしながら、現状では個人性に合わせて文章の単語数を制御するモデルは確立していない。そこで今回は、変換前後で文の平均単語数の変化が最小となるようなワードペナルティを算出する。ワードペナルティを変化させたときの変換前後の平均単語数と変化量のグラフを図 2 に示す。

図 2 より、ワードペナルティが 1.3 のときに変換前後の単語数の変化が最小となったため、この値を以降の実験に用いた。

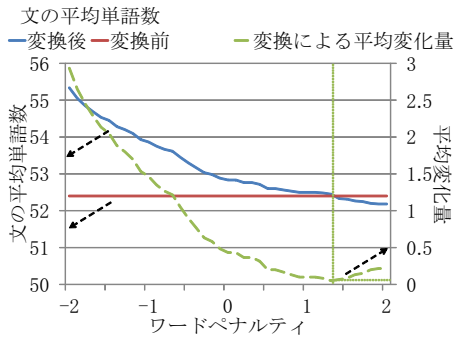


図 2: ワードペナルティと変換による単語数の変化

4 実験的評価

4.1 日本語 PPDB の内的評価

日本語 PPDB の品質を評価するために、日本語 PPDB の中からランダムに 90 組の言い換えを取り出し、それらを被験者 3 名に“5: 言い換えられる”～“1: 言い換えられない”の 5 段階で評価してもらう。図 3 に主観評価と言い換え確率の関係を示す。

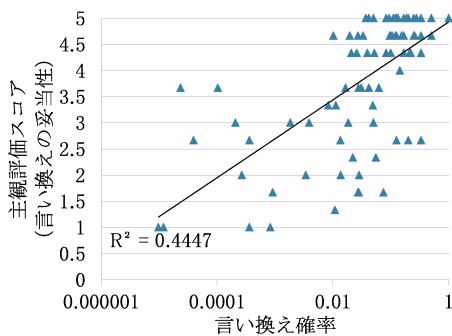


図 3: 言い換え確率と主観評価の関係

言い換え確率と主観評価の平均スコアに対して対数近似的な決定係数 R^2 を計算したところ、決定係数 $R^2 = 0.44$ となった。このことから、人間の直観的な言い換えの妥当性と言い換え確率は十分に相関があると言える。また、主観評価の平均スコアが 3 よりも大きい言い換えの数は 64、平均スコアが 4 よりも大きい言い換えの数は 53 であり、今回提示した言い換えにおいて 60 % 近くの言い換えが主観評価を通して言い換え可能であると判断された。

4.2 日本語 PPDB の外的評価

4.2.1 実験条件

先行研究と同様の条件で実験を行う [12]。実験の題材としてカメラ販売対話コーパス [4] を利用する。コーパスをそれぞれ話者ごとに分け、さらに学習データと評価データに分けて利用する。

実験的評価では、まず変換目標の話者の学習データを被験者に読んでもらい、次に変換前の文章とそれに対して個人変換を行った文章を読んでもらう。変

換結果に対して主観評価指標である WER と目標話者らしさを評価してもらう [12]。この実験は被験者 3 名に対して行い、各被験者には 3 話者、1 話者あたり 10 個の計 30 の変換結果に対して評価を行ってもらう。また、各評価指標の信頼区間を有意水準 $p < 0.05$ の Bootstrap Resampling を用いて求める [5]。

変換では、本稿で提案する日本語 PPDB を用いて個人変換を行い、その結果から有効性を検証する。また、先行研究より、助詞・助動詞・感動詞・フィラーを集めた辞書 (Particle) およびシソーラスを用いて構築した辞書 (Thesaurus) に対して n-gram 類似度を用いた翻訳モデル確率を付与した翻訳モデルを用意し、日本語 PPDB を翻訳モデルとして利用した場合 (PPDB) と変換前の文章 (Source) と比較した。

4.2.2 実験結果

個人変換を行った結果の客観評価指標として、カバレッジを図 4 に、エントロピーを図 5 に示す [12]。

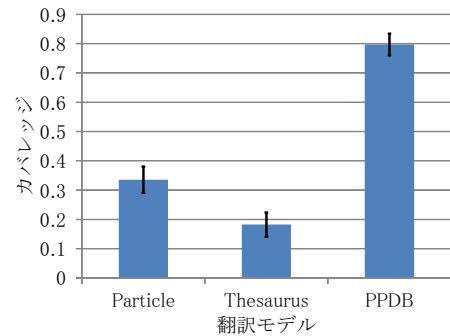


図 4: 各翻訳モデルにおけるカバレッジ

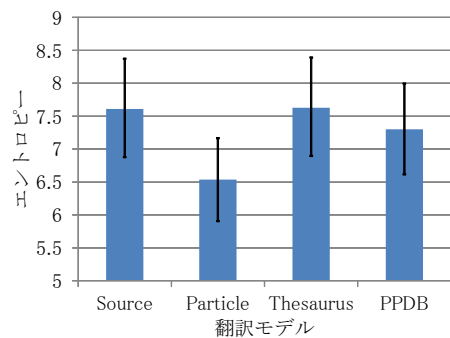


図 5: 各翻訳モデルにおけるエントロピー

先行研究で提案された Particle, Thesaurus に比べ、PPDB は 80 % 近くの高いカバレッジを得られ、多くの言い換えを網羅していることが分かる。また、エントロピーも Particle には劣るが、Source, Thesaurus に比べ低い傾向にある。

次に、主観評価の結果として、WER を図 6 に、主観評価スコアを図 7 に示す。

PPDB を用いた場合、高いカバレッジを持つにもかかわらず、Thesaurus に比べ WER が明らかに低い値となった。これは、PPDB に与えられた言い換え確

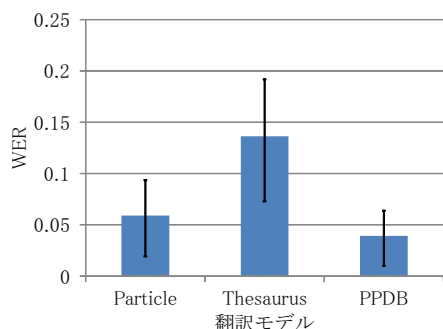


図 6: 各翻訳モデルにおける WER

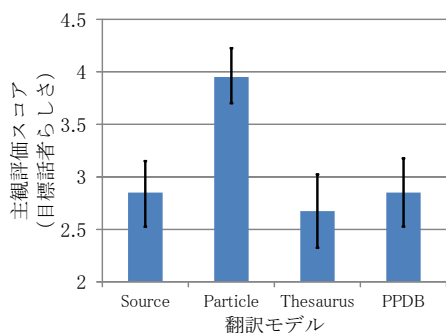


図 7: 各翻訳モデルにおける主観評価スコア

率が正確であることを示している。目標話者らしさを評価する主観評価スコアにおいては Thesaurus に比べ向上しているが、Source とほぼ同程度のスコアで、Particle に比べて低い値である。

客観評価指標であるエントロピーと主観評価スコアを比較すると、各翻訳モデルの優劣関係が一致していることが分かる。このことから、エントロピーによる客観評価が、人間の主観的な個人性評価とよく対応していると言える。

実験結果から、PPDB による個人性変換では、Thesaurus, Particle に比べて多くの言い換えを網羅し、言い換え確率による精度の高い言い換えが可能となっている。しかしながら、主観評価スコアは Particle に比べ劣る結果となった。Particle と PPDB の言い換えを比較すると、Particle の機能語の言い換え 1338 組のうち、PPDB がカバーできた機能語の言い換えは 144 組で、11%程度しかカバーできないことが分かる。PPDB は多数の言い換えを網羅している一方、対訳コーパスの前処理でフィルター削除が行われている点や、中間意味表現の関係から機能語同士の言い換え抽出が難しいことから、機能語の言い換えを網羅できなかった。また、言語学においては機能語は役割語として個人性を表現することが分かっている [11]。これらの理由から、PPDB では Particle ほど個人性を表現する言い換えを行えなかったと考えられる。

5 まとめ

本稿では、日本語における言い換え生成として、翻訳フレーズテーブルを利用して日本語 PPDB を構築

した。日本語 PPDB に対する主観評価によって、品質は十分示されている。

また、自然言語処理への応用として、著者らの提案する個人性変換に日本語 PPDB を適用した。日本語 PPDB を利用する効果は個人性変換におけるカバレッジと WER 評価で如実に表れた。しかしながら、個人性変換においては、終助詞の変換は上手く行われず、主観評価スコア自体には多くの影響を与えなかった。

今後は、言い換え抽出手法を改良し、終助詞に対して柔軟な日本語 PPDB を構築し、内容語のみならず、機能語の言い換えへの対応を進める。構築した日本語 PPDB は言語資源として公開する予定である。また、先行研究で利用した翻訳モデルと PPDB による翻訳モデルを混合して利用した場合の評価も進め、個人性変換の精度向上に努める。

参考文献

- [1] ALC. 英辞郎. <http://shop.alc.co.jp/cnt/eijiro/>. Accessed: 2013-07-26.
- [2] C. Bannard and C. Callison-Burch. Paraphrasing with bilingual parallel corpora. In *Proc. ACL*, pp. 597–604. Association for Computational Linguistics, 2005.
- [3] J. Ganitkevitch, B. Van Durme, and C. Callison-Burch. PPDB: The paraphrase database. In *Proc. HLT-NAACL*, pp. 758–764, Atlanta, Georgia, June 2013. Association for Computational Linguistics.
- [4] T. Hiraoka, G. Neubig, S. Sakti, T. Toda, and S. Nakamura. Construction and analysis of a persuasive dialogue corpus. In *IWSDS*, Napa, California, USA, 1 2014.
- [5] P. Koehn. Statistical significance tests for machine translation evaluation. In *Proc. EMNLP*, pp. 388–395, 2004.
- [6] F. Mairesse and M. A. Walker. Controlling user perceptions of linguistic style: Trainable generation of personality traits. *Computational Linguistics*, 37(3):455–488, 2011.
- [7] G. Neubig, Y. Akita, S. Mori, and T. Kawahara. A monotonic statistical machine translation approach to speaking style transformation. *Computer Speech and Language*, 26(5):349–370, 2012.
- [8] G. Neubig, Y. Nakata, and S. Mori. Pointwise prediction for robust, adaptable japanese morphological analysis. In *Proc. ACL, HLT '11*, pp. 529–533, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [9] J. Riesa, A. Irvine, and D. Marcu. Feature-rich language-independent syntax-based alignment for statistical machine translation. In *Proc. EMNLP*, pp. 497–507. Association for Computational Linguistics, 2011.
- [10] T. Takezawa, E. Sumita, F. Sugaya, H. Yamamoto, and S. Yamamoto. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world. In *Proc. LREC*, pp. 147–152, 2002.
- [11] M. Teshigawara and S. Kinsui. Modern Japanese ‘Role Language’ (Yakuwarigo): fictionalised orality in Japanese literature and popular culture. In *Sociolinguistic Studies Vol 5-1*. Sheffield: Equinox Publishing, 2012.
- [12] 水上, G. Neubig, S. Sakti, 戸田, 中村. 話し言葉における言語情報の個人性変換における変換辞書拡張. 電子情報通信学会音声研究会 (SP), 千葉, 9 2013.
- [13] 乾, 藤田. 言い換え技術に関する研究動向. 自然言語処理, 11(5):151–198, 2004.