

確率的用例ベース翻訳の実現

荒牧 英治 †† 黒橋 禎夫 †† 柏岡 秀紀 † 加藤 直人 †††

† 東京大学大学院情報理工学系研究科

‡ ATR 音声言語コミュニケーション研究所

††† NHK 放送技術研究所

{aramaki, kuro}@kc.t.u-tokyo.ac.jp

hideki.kashioka@atr.jp, naoto.kato@atr.jp

1 はじめに

近年、統計ベース翻訳 [5] や用例ベース翻訳 [10] など大量のテキストを用いた翻訳手法 (コーパスベース翻訳) が注目されている。我々は、用例ベース翻訳に焦点を当て研究を行っている。

用例ベース翻訳の基本的なアイデアは、入力文の各部分に対して類似している用例を選択し、それらを組み合わせて翻訳を行うことである。ここでいう類似とは、通常、入力文とできるかぎり大きく一致していればいるほどよいと考えられてきた。なぜならば、用例のサイズが大きくなればなるほど、より大きなコンテキストを扱うことになり、正確な訳につながるからである。したがって、大きな用例が利用可能なドメイン、すなわち、特許翻訳のような類似文が多いドメインにおいて、用例ベース翻訳の可能性が注目されている。しかし、これまでの用例ベース翻訳システムは、用例のサイズ/類似度などを経験則による指標で計算してきたため、統計ベース翻訳システムに比べて、そのアルゴリズムが不透明でアドホックであった。

提案手法は、翻訳確率という尺度のみを用いて用例選択を行う。提案する翻訳確率は、統計ベースのそれとは異なり、語や句単位の小さな単位から、文全体まで、あらゆるサイズをカバーして構築される。この枠組みの上では、大きなサイズの用例は安定した翻訳先を伴うため、高い翻訳確率を持つと考えられる。したがって、翻訳確率が高い用例を選ぶことで、自然と用例のサイズを考慮した用例の選択が可能となる。この翻訳確率の考え方は簡潔で自然であり、実験の結果、従来手法と同等の精度が得ることができたので報告する。提案手法は言語ペアを特定しないが、本稿は日英翻訳方向で説明し、実験を行う。

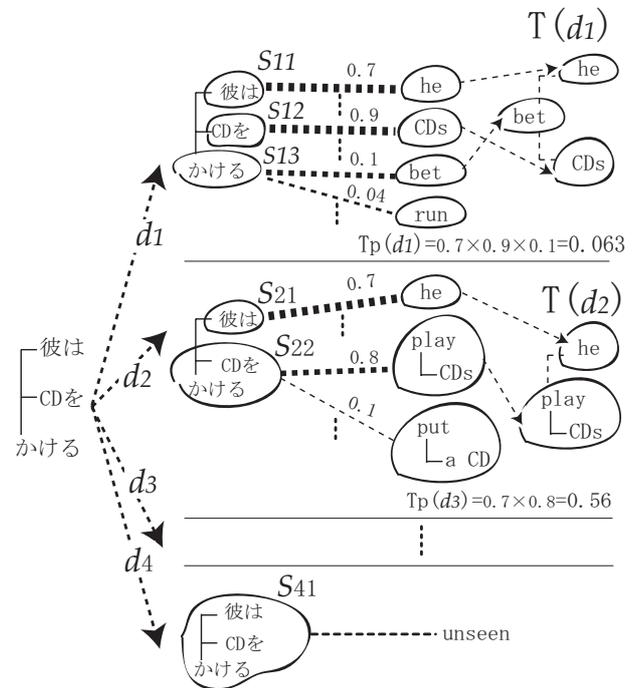


図 1: 翻訳のながれ。

本稿では、依存構造木は左を親とし、子を右に描く。また、本研究では、ノードの単位は日本語は文節、英語は base-NP, VP とする。

2 提案手法

用例ベース翻訳の基本的な原則はできるだけ大きなサイズの用例を用いて翻訳を生成することである。例えば、“彼は CD をかける”を翻訳する場合、“かける”単独で用例を検索した場合、“bet”や“run”など不適切な訳語を選ぶ可能性がある。そこで、“CD をかける”など大きなサイズで用例が存在する場合は、可能な限り大きな用例を用い、正確な訳語 (“play”) を用い

たい．提案手法は，この用例の大きさに対する選好を確率的に定式化して実現する．

まず，提案手法は入力文を可能なかぎりの部分木の組合せに分解する：

$$D = \{d_1, \dots, d_N\}. \quad (1)$$

ここで， d_i は入力文の分解のパターン， D は d_i の集合とする．例えば，図 1 左の入力文の場合， d_1, \dots, d_4 の 4 通りの部分木の組合せで表現できる．

次に， d_i は入力文を M_i 個の部分木に分解しているとする：

$$d_i = \{s_{i1}, s_{i2}, \dots, s_{iM_i}\}, \quad (2)$$

ここで， s_{ij} は入力文の部分木である．例えば，図 1 では， d_1 は入力文を 3 つの部分木 s_{11}, s_{12}, s_{13} に分解している．

次に，各部分木 s_{ij} について，もっとも翻訳確率 $P(t_{ij} | s_{ij})$ (この確率の計算方法は次節にて述べる) の高い用例を選び，それらの積を翻訳文の翻訳確率 $P(d_i)$ とする：

$$P(d_i) = \prod_{s_{ij} \in d_i} \max_{t_{ij}} P(t_{ij} | s_{ij}). \quad (3)$$

ここで， t_{i1}, \dots, t_{iM_i} を d_i の翻訳とみなし， $T(d_i)$ と表記する．

最後に，もっとも高い翻訳確率を持つ d_m を以下の式によって探索し，最終的な翻訳 $T(d_m)$ を得る：

$$d_m = \arg \max_{d_i \in D} P(d_i). \quad (4)$$

例えば，図 1 の $T(d_1)$ のように，入力文を小さな部分木に分解した場合は，曖昧性のある日本語“かける”に対して，“bet”，“run”や“play”など様々な英語表現が考えられる．この場合，適切な訳である $P(\text{play} | \text{かける})$ の翻訳確率は低く，適切な翻訳は行われない．

一方， $T(d_2)$ では，より大きな用例“CD をかける”を用いている．この用例の英語表現としては，ほとんどが“play”となり，用例の翻訳確率は高くなる．その結果，用例群の翻訳確率の積である $P(d_2)$ も高くなり，この結果が翻訳として採用される．

また，図 1 の $T(d_4)$ のように，大きすぎる用例を検索した場合は，コーパス中に存在せず，確率が定義されない場合がある．

2.1 パラメータの推定

本節では，用例の翻訳確率の推定方法を述べる．まず，英語部分木 t と日本語部分木 s からなる用例があ

用例		context_sim
英語側	日本語側(とコンテキスト)	
play	(CDを)かける	0.8
play	(テープを)かける	0.8
put	(MDを)かける	0.8
⋮	⋮	⋮
set	(目覚ましを)かける	0.6
⋮	⋮	⋮
bet	(お金を)かける	0.7
bet	(財産を)かける	0.7
bet	(100ドルを)かける	0.3

* 実際は用例は木構造の形で扱われているが，表記を簡潔にするため，ここでは用例の構造は記していない．

図 2: “かける”を含んだ用例とそのコンテキスト (コンテキストは括弧で示されている)．

るとする．この翻訳確率 $P(t | s)$ は，アライメントされたコーパス中での対応 (t, s) の出現頻度を直接数えて求める：

$$P(t | s) = \frac{\text{count}(t, s)}{\text{count}(*, s)}, \quad (5)$$

ここで， $\text{count}(t, s)$ は，アライメントされたコーパスにおける対応 (t, s) の出現頻度， $\text{count}(*, s)$ は日本語部分木 (s) の出現頻度である*．

ただし，この頻度の計算にあたっては，次節のフィルタリングを行う．

2.2 context_sim による用例のフィルタリング

用例の選択にあたって重要な手がかりは用例のサイズであり，それはすでに提案された翻訳確率の枠組みで実現されている．しかし，サイズに加えて，コンテキストの類似も重要な手がかりである．

例えば，コーパスに存在しない句“レコードをかける”を図 2 の用例を用いて翻訳することを考える．大きなサイズで一致するものがない場合，“かける”単独で翻訳確率を計算することになり，前述したように“bet”など不適切な訳語が選ばれる可能性がある．このような場合においては，“レコード”と“CD”は似ているため，用例“CD をかける”を用いて，“レコードをかける”を翻訳したい．そこで，翻訳に使う部分の周辺の句 (“CD”) をコンテキストとみなし，入力文

*後述する実験では，データスペースへの対処として s と t は内容語に汎化して確率の計算を行った．

と類似したコンテキストを持つ用例の翻訳確率を高くすることを考える。

まず、提案手法を説明する前に、用例 A と入力文のコンテキストの類似度を次の式で定める：

$$\text{context_sim}(A) = \sum_{i \in N} \text{sim}(i, j), \quad (6)$$

ここで、 i は用例 A の日本語側で翻訳に使う部分の周辺の句、 j は i と対応する入力文の句、 N は i の集合、 $\text{sim}(i, j)$ はシソーラス [15] を用いて計算する i と j の類似度 ($\max=1$) である。

次に、式 5 にて用例 A の翻訳確率を計算する際には、 $\text{context_sim}(A)$ 以上の類似度を持つ用例だけを集計して翻訳確率を計算し、 $\text{context_sim}(A)$ 未満の類似度の用例は、用例 A の翻訳確率の計算には用いない。この操作を用例のフィルタリングと呼ぶ。このフィルタリングによって、高い context_sim を持つ用例の翻訳確率は、それよりも低い context_sim を持つ用例の影響を受けず、正確な訳語が選好される。

例えば、用例 “CD をかける” と入力文 “レコードをかける” の context_sim が 0.8 であるとすると、同じく 0.8 以上の context_sim を持つ用例 (図 2 では点線で示されている) だけを用いて翻訳確率を計算する。この場合、用例の数は 3 つだけとなるが、その英語表現は安定しており、 $P(\text{play} | \text{かける}) = \frac{2}{3}$ 、 $P(\text{put} | \text{かける}) = \frac{1}{3}$ となる。このように類似したコンテキストを持つ用例の翻訳確率は自然と高くなる。

3 実験

提案手法の妥当性を検証するため、用例ベース翻訳システム [2] の用例選択部分を提案手法に置き換えて実験を行った。

3.1 実験設定

実験は、(1) 提案システム (PROPOSED)、および、(2) 経験則によるメジャーにより用例を選択するシステム [2] (BASIC) の 2 つの翻訳システムを表 1 の自動評価法を用いて比較することで行った。

コーパスは IWSLT04[1] にて配布されたコーパス (トレーニングとテストセット) を用いた。トレーニングセットは旅行対話ドメインの 20k の日英対訳文からなる。これらに対して、翻訳辞書を用いた手法 [4] でア

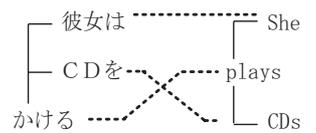


図 3: アライメントされた対訳文の例。

ライメントを行った。アライメントされた結果例を図 3 に示す。この対応関係から用例の翻訳確率は計算した。利用した辞書は英辞郎[†]などいくつか辞書をマージしたもので、延べ百万語を含む。テストセットは日本語文 (500 文) とそれらの 16 通りの英語翻訳 (500 × 16 文) からなる。

3.2 実験結果

各手法の精度を表 2 に示す。表 2 に示されるように、提案手法 PROPOSED は、経験則による BASIC と比べて僅かに高い精度を持ち、提案する確率による選択が妥当であることを示している。

3.3 用例のフィルタリングの効果

次にコンテキストの類似度を考慮した効果を調べてみた。これは、提案手法 (PROPOSED) とシソーラスを用いない手法 (WITHOUT_SIM) の精度を比較することで行った (表 2)。

実験結果は、NIST においては PROPOSED が高く、BLEU においては WITHOUT_SIM が僅かに高かった。NIST は訳語選択により鋭敏に反応するため、コンテキストの類似度は訳語選択に貢献すると考えられる。

例えば、以下のように、“使えますか” の翻訳において、WITHOUT_SIM では、用例 “ビザカードは使えますか” を用いて不適切な訳であったが、PROPOSED においては、用例 “スペアは使えますか” を用いて改善された例が観察された。

入力文:	このカートは使えますか。
PROPOSED:	Can I use this cart ?
WITHOUT_SIM:	Do you accept this cart ?

4 関連研究

これまで様々な用例ベース翻訳システムが提案されてきたが、それらは経験則に基づいて用例を選択して

[†]<http://www.eijiro.jp/>

表 1: 自動評価手法.

BLEU	正解との n-gram の適合率の相乗 (幾何) 平均 [12].
NIST	正解との n-gram の適合率の相加 (算術) 平均 [6].
WER	Word Error Rate. 正解との編集距離 [11].
PER	Position Independent Word Error Rate. 語順を用いない正解との編集距離 [7].
GTM	general text matcher. 正解との一致した最長語列の適合率, 再現率の調和平均 [14].

* BLEU, NIST, GTM については大きな値ほど精度がよい. WER, PER については小さな値ほど精度がよい.

表 2: 実験結果.

	bleu	nist	wer	per	gtm
PROPOSED	0.41	8.04	0.52	0.44	0.67
BASIC	0.39	7.92	0.52	0.44	0.67
WITHOUT_SIM	0.42	7.67	0.49	0.42	0.68

* WITHOUT_SIM については 3.3 章で述べる.

おり, 提案手法のような確率的な尺度に注意を払っていない.

例えば, [13] はマニュアルドメインの用例ベース翻訳システムを提案した. 彼らのシステムは, 用例と入力文の間で一致する部分のサイズのみを用いて用例を選択している. これは, マニュアルドメインの翻訳においてコンテキストの必要性が少ないからだと考えられるが, 一般的な翻訳ドメインを扱うためには, 彼らの手法は不十分だと考えられる.

[8, 9] は, 一致サイズとコンテキストの類似度の両方を用いて用例を選択している. [3] は, それらに加え, さらにアライメントのもっともらしさを用いて用例を選択している. これらの手法では, 複数の尺度をどのような重みで考慮するか, という重み付けの問題が存在する.

5 おわりに

本稿では, 大きな用例ほど翻訳確率が高くなるという考えに基づき, 翻訳確率だけを用いて用例を選択する用例ベース翻訳手法を提案した.

実験の結果は, 従来の経験則による用例選択を行うシステムと同程度の精度を得ることができ, 提案手法の妥当性を示している.

参考文献

- [1] Yasuhiro Akiba, Marcello Federico, Noriko Kando, Hiromi Nakaiwa, Michael Paul, and Jun'ichi Tsujii. Overview of the IWSLT04 evaluation campaign. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, pp. 1–12, 2004.
- [2] Eiji Aramaki and Sadao Kurohashi. Example-based machine translation using structural translation examples. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, pp. 91–94, 2004.
- [3] Eiji Aramaki, Sadao Kurohashi, Hideki Kashioka, and Hideki Tanaka. Word selection for ebmt based on monolingual similarity and translation confidence. In *Proceedings of the HLT-NAACL 2003 Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, pp. 57–64, 2003.
- [4] Eiji Aramaki, Sadao Kurohashi, Satoshi Sato, and Hideo Watanabe. Finding translation correspondences from parallel parsed corpus for example-based translation. In *Proceedings of MT Summit VIII*, pp. 27–32, 2001.
- [5] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, Vol. 19, No. 2, 1993.
- [6] G. Doddington. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of HLT*, pp. 257–258, 2002.
- [7] Nicola Uerger Franz Josef Och and Hermann Ney. An efficient a* search algorithm for statistical machine translation. In *Proceedings of ACL 2001 Workshop on Data-Driven Machine Translation*, pp. 55–62, 2001.
- [8] Osamu Furuse and Hitoshi Iida. Constituent boundary parsing for example-based machine translation. In *Proceedings of the 15th COLING*, pp. 105–111, 1994.
- [9] Kenji Imamura. Application of translation knowledge acquired by hierarchical phrase alignment for pattern-based mt. In *Proceedings of TMI-2002*, pp. 74–84, 2002.
- [10] Makoto Nagao. A framework of a mechanical translation between Japanese and English by analogy principle. In *Elithorn, A. and Banerji, R. (eds.): Artificial and Human Intelligence*, pp. 173–180, 1984.
- [11] S. Niessen, G. Leusch F. J. Och, and H. Ney. An evaluation tool for machine translation: Fast evaluation for machine translation research. In *Proceedings of the Second Int. Conf. on Language Resources and Evaluation (LREC)*, pp. 39–45, 2000.
- [12] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of ACL 2002*, pp. 311–318, 2002.
- [13] Stephen D. Richardson, William B. Dolan, Arul Menezes, and Monica Corston-Oliver. Overcoming the customization bottleneck using example-based mt. In *Proceedings of the ACL 2001 Workshop on Data-Driven Methods in Machine Translation*, pp. 9–16, 2001.
- [14] J. P. Turian, L. Shen, and I. D. Melamed. Evaluation of machine translation and its evaluation. In *Proceedings of MT Summit IX*, pp. 386–393, 2003.
- [15] NTT コミュニケーション科学研究所. 日本語語彙大系. 岩波書店, 1997.