

統計的機械翻訳におけるピボット翻訳の比較

内山 将夫 井佐原 均
独立行政法人情報通信研究機構 (NICT)

1 はじめに

コーパススペースの機械翻訳 [11, 2] の順調な発展は、大規模なパラレルコーパスにより可能になった。そのようなパラレルコーパスの例としては、Linguistic Data Consortium による中国語と英語およびアラビア語と英語のコーパスや、Europarl コーパス [7] がある。しかしながら、多くの言語対については、利用可能な大規模コーパスは存在しない。

大規模なパラレルコーパスがない状況においても、コーパススペースの機械翻訳を実現するために、英語をピボットとするコーパススペースの機械翻訳を本稿では提案する。すなわち、原言語と英語および英語と対象言語の二つのパラレルコーパスを利用することにより、原言語と対象言語の機械翻訳をする。なお、英語以外のものをピボットとすることも当然可能である。

本稿では、ピボット翻訳の2つの方法を比較する。1つ目はベースラインとしての単純な方法である。この方法は、まず、原言語の文を英語に翻訳し、次に、その翻訳された英語の文を、対象言語に翻訳する。この方法においては、原言語と英語のパラレルコーパスから得られた機械翻訳システムと英語と対象言語のパラレルコーパスから得られた機械翻訳システムをパイプラインとして利用することにより、原言語の文を対象言語に翻訳する。この方法を以下では SntTrans と呼ぶ。SntTrans では、2節のシステムを2つ続けて利用している。

ピボット翻訳の2つ目はフレーズテーブルの合成に基づくものである。本稿においてはコーパススペースの機械翻訳システムとして2節で述べるフレーズベースの統計的機械翻訳 (SMT) システムを利用する。フレーズベースの SMT においては、パラレルコーパスから作成されたフレーズテーブルを利用することにより、翻訳をする。そのため、このフレーズテーブルを、原言語と英語および英語と対象言語のフレーズテーブルから合成するこ

とにより、原言語と対象言語の翻訳をする。この方法を以下では PhraseTrans と呼ぶ。

以下では、まず、フレーズベースの SMT について説明したあとで、フレーズに基づくピボット翻訳について述べる。次に、実験によりこれら2つのピボット手法を比較する。なお、実験などの詳細については文献 [17] を参照のこと。

2 フレーズベースの SMT

本稿では、2006 NAACL/HLT で開かれた「統計的機械翻訳に関するワークショップ」で利用されたベースラインシステムを利用する [9]。このシステム (Pharaoh) は対数線型モデル [13] に基づくフレーズベースの SMT システムである [10, 6]。

Pharaoh では、次式に基づき、与えられた原言語の文 f を対象言語の文 \hat{e} に翻訳する。

$$\hat{e} = \arg \max_e \Pr(e|f) \quad (1)$$

$$= \arg \max_e \sum_{m=1}^M \lambda_m h_m(e, f) \quad (2)$$

ただし、 $h_m(e, f)$ は素性関数であり、 λ_m はその重みである。ベースラインシステムが利用する素性関数は次の8つである。(1) 対象言語の3-gram 言語モデル、(2,3) 双方向のフレーズ翻訳確率、(4,5) 双方向の語彙翻訳確率、(6) 単語ペナルティ、(7) フレーズペナルティ、(8) 並べ替えのペナルティ。これらについて詳しくは [10, 6, 8] を参照のこと。なお、 λ_m は、パラメタ調整用データにおける BLEU [14] が最大となるように設定される [12]。

3 フレーズに基づくピボット翻訳

フレーズベースの SMT では、言語モデルとフレーズテーブルが翻訳に必要である。このうち言語モデルは、対象言語のみから作ることができる

ので、フレーズテーブルを作成することができれば、フレーズベースのSMTをすることができる。

この節では、原言語であるフランス語 f について、英語 e をピボットとして、対象言語であるドイツ語 g に翻訳するものとして、フレーズテーブルの作成法を説明する。

まず、仏英および英独パラレルコーパスから、ベースラインシステムに付随する訓練プログラムを利用して、フレーズテーブル T_{FE} と T_{EG} を作成する。次に、これらを合成して、仏独のフレーズテーブル T_{FG} を作成する。ここで、 T_{FG} を作成するには、双方向のフレーズ翻訳確率 $\phi(\bar{f}|\bar{g})$ と $\phi(\bar{g}|\bar{f})$ 、および、双方向の語彙翻訳確率 $p_w(\bar{f}|\bar{g})$ と $p_w(\bar{g}|\bar{f})$ を推定する必要がある。ただし、 \bar{f} と \bar{g} は仏独のフレーズである。

これらは、 T_{FE} と T_{EG} におけるフレーズ翻訳確率 $\phi(\cdot)$ と語彙翻訳確率 $p_w(\cdot)$ を用いて、以下のように推定する。

$$\phi(\bar{f}|\bar{g}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} \phi(\bar{f}|\bar{e})\phi(\bar{e}|\bar{g}) \quad (3)$$

$$\phi(\bar{g}|\bar{f}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} \phi(\bar{g}|\bar{e})\phi(\bar{e}|\bar{f}) \quad (4)$$

$$p_w(\bar{f}|\bar{g}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} p_w(\bar{f}|\bar{e})p_w(\bar{e}|\bar{g}) \quad (5)$$

$$p_w(\bar{g}|\bar{f}) = \sum_{\bar{e} \in T_{FE} \cap T_{EG}} p_w(\bar{g}|\bar{e})p_w(\bar{e}|\bar{f}) \quad (6)$$

ただし、 $\bar{e} \in T_{FE} \cap T_{EG}$ は英語のフレーズ \bar{e} が T_{FE} と T_{EG} の双方に含まれていることを示す。

式3-6は、翻訳確率が、英語を条件として独立である(例えば $\phi(\bar{f}|\bar{g}, \bar{e}) = \phi(\bar{f}|\bar{e})$)と仮定している。なお、フレーズ翻訳確率と語彙翻訳確率の定義については、文献[17, 10]を参照のこと。

4 実験

ピボット翻訳の性能を、パラレルコーパスが直接利用できる場合と比較するために、Europarlコーパスを実験に利用した。Europarlコーパスでは、複数の欧州言語がパラレルになっているため、原言語 英語と英語 対象言語に加えて、原言語 対象言語のパラレルコーパスも得られる。そのため、原言語と対象言語のパラレルコーパスから直接訓練したSMTシステムをDirectと呼

言語対	Direct	PhraseTrans	SntTrans
西仏	35.78	32.90 (0.92)	29.16 (0.81)
仏西	34.16	31.49 (0.92)	27.99 (0.82)
独仏	23.37	22.47 (0.96)	21.64 (0.93)
仏独	15.27	14.51 (0.95)	14.21 (0.93)
独西	22.34	21.76 (0.97)	20.97 (0.94)
西独	15.50	15.11 (0.97)	14.61 (0.94)

表 1: BLEU スコアと相対的な性能

び、以下では、Direct, PhraseTrans, SntTrans の3つを比較する。

4.1 実験手順

実験には、前述のSMTワークショップにおけるデータ[9]を利用した。その訓練データは、仏英、西(スペイン)英、独英のパラレルコーパスからなる。これらから、各英語文に対応する仏独西の文を抽出することにより、英仏独西の4言語にパラレルなコーパスを作成した。その大きさは各言語につき585830文である。この4言語のパラレルコーパスから、各言語対のための訓練データを作成した。パラメタ調整は、パラメタ調整用データにおける最初の500文を利用して行ない、テストは、テスト用の3064文を利用した。翻訳の評価基準としては%BLEUスコアを利用した。

4.2 実験結果

表1に3種類の方法によるBLEUを示す。括弧内の数字はDirectと比べての相対的な性能である。たとえば、西仏では、DirectのBLEUは35.78, PhraseTransのBLEUは32.90, 相対的な性能は0.92=(32.90/35.78)である。

表1より、Direct > PhraseTrans > SntTransが全ての言語対について成立することがわかる。Directが最も性能が良いのは当然である。DirectはPhraseTransやSntTransの性能の上限といえる。一方、PhraseTransの方がSntTransよりも性能が良い理由は、PhraseTransでは、個々のフレーズを翻訳して、そのフレーズを組合せるという方法を利用しているため、SntTransに比べて、多くの種類の翻訳候補のなかから対象言語の翻訳を選ぶことができるためだと考える。PhraseTrans

	フレーズ数 (“M” は 10^6)			R	P
	Direct	PhraseTrans	共通		
西仏	18.2M	190.8M	6.3M	34.7	3.3
仏西	18.2M	186.8M	6.3M	34.7	3.4
独仏	7.3M	174.9M	3.1M	43.2	1.8
仏独	7.3M	168.2M	3.1M	43.2	1.9
独西	7.5M	179.6M	3.3M	44.1	1.9
西独	7.6M	176.6M	3.3M	44.1	1.9

表 2: フレーズテーブルの大きさ

の相対的な性能は 0.92 ~ 0.97 と高い性能を示している。

表 2 には、フレーズテーブル中の対訳フレーズの数を示す。これより Direct には 7.3 ~ 18.2 M のフレーズがあり、PhraseTrans には 168.2 ~ 190.8 M のフレーズがあることがわかる。したがって、フレーズテーブルの大きさは 10 倍以上と非常に大きくなる¹。

表 2 における「共通」の欄にある数字は、Direct と PhraseTrans で共有されるフレーズの数である。また、R と P は再現率と適合率であり、それぞれ、 $R = \frac{\text{共通のフレーズ数} \times 100}{\text{Direct でのフレーズ数}}$ 、 $P = \frac{\text{共通のフレーズ数} \times 100}{\text{PhraseTrans でのフレーズ数}}$ である。これより、再現率はある程度高いが、適合率は低いことがわかる。しかし、PhraseTrans の性能は、Direct と比べてもそれほど悪化していないので、適合率の低さは、性能には悪影響をそれほど与えていないと言える。

5 関連研究

ピボット翻訳は規則に基づく機械翻訳において利用されてきた。たとえば、文献 [1] は、ピボット翻訳の得失について論じているし、文献 [16] では、ピボット言語としては、自然言語を使うべきであると論じている。また、統計的な機械翻訳に

¹合成されたフレーズテーブルの大きさは、以下のように見積もることができる。まず、あるフレーズテーブルにおける英語のフレーズ i の数を f_i とする。すると、異なる英語のフレーズ数が n としたとき、フレーズテーブルの大きさは $s_1 = \sum_{i=1}^n f_i$ である。次に、合成したフレーズテーブルの大きさは、概算では、 $s_2 = \sum_{i=1}^n f_i^2$ である。したがって、両者の相対的な大きさは、 $E(f)$ と $E(f^2)$ を f_i と f_i^2 の平均としたとき、 $r = \frac{s_2}{s_1^2} = \frac{E(f^2)}{E(f)^2}$ である。具体例として、独英のフレーズテーブルについて、これらの値を調べてみると、 $E(f) = 1.5$ 、 $E(f^2) = 43.7$ 、 $r = 28.9$ である。これより、たとえ、 $E(f)$ が小さくても、相対的なサイズである r が大きくなることがわかる。

おける例としては、文献 [5, 3] がある。翻訳に関連する領域におけるピボット言語の利用例としては、文献 [15, 18, 4] などがある。

6 おわりに

本稿では、コーパススペースの機械翻訳におけるピボット翻訳の例として、文自体を翻訳する方法と、フレーズテーブルを合成する方法とを比較した。比較実験においては、Europarl コーパスを利用することにより、原言語と対象言語の平行コーパスから直接訓練した SMT システムと、ピボット翻訳とを比較した。実験の結果、フレーズテーブルを合成する方法が、文自体を翻訳する方法よりも BLEU 値により測定した性能が高いこと、および、その性能は直接訓練した SMT システムの 0.92 ~ 0.97 であることがわかった。

本稿での実験は、原言語と対象言語の平行コーパスが利用可能な状況における実験のため、本当にピボット翻訳が必要な状況、つまり、原言語と対象言語の平行コーパスが存在しない状況を正確にはシミュレートしていないが、PhraseTrans が SntTrans よりも性能が良いということ自体は、平行コーパスが存在しない状況においても成立すると考える。これについて確認するのは、今後の課題である。

参考文献

- [1] Christian Boitet. Pros and cons of the pivot and transfer approaches in multilingual machine translation. In Dan Maxwell, Klaus Schubert, and Toon Witkam, editors, *New Directions in Machine Translation*. Foris, 1988. (appeared in Sergei Nirenburg, Harold Somers and Yorick Wilks (eds.) *Readings in Machine Translation* published by the MIT Press in 2003).
- [2] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Com-*

- putational Linguistics*, Vol. 19, No. 2, pp. 263–311, 1993.
- [3] Adrià de Gispert and José B. Mari no. Catalan-English statistical machine translation without parallel corpus: Bridging through Spanish. In *Proc. of LREC 5th Workshop on Strategies for developing Machine Translation for Minority Languages*, 2006.
- [4] Tim Gollins and Mark Sanderson. Improving cross language information retrieval with triangulated translation. In *SIGIR*, 2001.
- [5] Manuel Kauers, Stephan Vogel, Christian Fügen, and Alex Waibel. Interlingua based statistical machine translation. In *ICSLP*, 2002.
- [6] Philipp Koehn. Pharaoh: a beam search decoder for phrase-based statistical machine translation models. In *AMTA*, 2004.
- [7] Philipp Koehn. Europarl: A parallel corpus for statistical machine translation. In *MT Summit*, 2005.
- [8] Philipp Koehn, Amittai Axelrod, Alexandra Birch Mayne, Chris Callison-Burch, Miles Osborne, and David Talbot. Edinburgh system description for the 2005 IWSLT speech translation evaluation. In *IWSLT*, 2005.
- [9] Philipp Koehn and Christof Monz. Manual and automatic evaluation of machine translation between european languages. In *Proceedings on the Workshop on Statistical Machine Translation*, pp. 102–121, New York City, June 2006. Association for Computational Linguistics.
- [10] Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical phrase-based translation. In *HLT-NAACL*, 2003.
- [11] Makoto Nagao. A framework of a mechanical translation between Japanese and English by analogy principle. In *the International NATO Symposium on Artificial and Human Intelligence*, 1981. (appeared in Sergei Nirenburg, Harold Somers and Yorick Wilks (eds.) *Readings in Machine Translation* published by the MIT Press in 2003).
- [12] Franz Josef Och. Minimum error rate training in statistical machine translation. In *ACL*, 2003.
- [13] Franz Josef Och and Hermann Ney. Discriminative training and maximum entropy models for statistical machine translation. In *ACL*, 2002.
- [14] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In *ACL*, 2002.
- [15] Charles Schafer and David Yarowsky. Inducing translation lexicons via diverse similarity measures and bridge languages. In *CoNLL*, 2002.
- [16] Klaus Schubert. Implicitness as a guiding principle in machine translation. In *COLING*, 1988.
- [17] Masao Utiyama and Hitoshi Isahara. A comparison of pivot methods for phrase-based statistical machine translation. In *HLT-NAACL*, 2007 (to appear).
- [18] Haifeng Wang, Hua Wu, and Zhanyi Liu. Word alignment for languages with scarce resources using bilingual corpora of other language pairs. In *COLING/ACL 2006 Main Conference Poster Sessions*, 2006.