句の翻訳順序パタンを考慮した統計的機械翻訳モデル

大橋一輝 [1]　山本和英 [1]　齋藤邦子 [2]　永田昌明 [2]

長岡技術科学大学 [1]　NTT サイバースペース研究所 [2]
{ohashi, ykaz}@nlp.nagaokaut.ac.jp　{saito.kuniko, nagata.masaaki}@lab.ntt.co.jp

1　はじめに

近年、統計的機械翻訳の分野において、句に基づく翻訳モデルが研究の主流になっている。その理由として、訳語の選択能力および局所的な語の並べ替え能力が語に基づく翻訳モデルに比べて高いことが挙げられる。しかし、従来の句に基づく翻訳モデル [1, 2] は、語順の違いを表現する歪みモデルが単純なため、大局的な句の並べ替え能力が低い。この歪みモデルは、翻訳先言語と翻訳元言語の句の並び方が同じでない場合にペナルティを与えるものである。語順が近い言語間の翻訳であれば有効に働くが、日本語と英語のように語順が離れた言語間の翻訳は難しい。

本稿では、句に基づく翻訳モデルにおける新たな歪みモデルを提案する。この歪みモデルでは、直前に翻訳した翻訳元言語句に対して次に翻訳する翻訳元言語句がどの位置にあるかという翻訳順序のパタンを 4 つに分類する。そして、これらのパタンが発生する確率は、直前に翻訳した翻訳元言語句とそれに対応する翻訳先言語句、および次に翻訳する翻訳元言語句とそれに対応する翻訳先言語句、の 4 つの句に依存するというものである。

この歪みモデルを構築するために、対訳文の対応付けを行う。対応付けが得られていれば翻訳順序パタンが決定されるため、パタンの相対頻度によって確率を算出する。さらに、パタンの頻度が極端に少なくならないように、パタンの発生に影響する 4 つの句をクラスタリングおよび句の品詞によって分類する。

対訳文の対応付けにはデコーダを用いる。デコーダはワードグラフを構築し、前向きビーム探索および後向き A* 探索によって上位 N 位 (N-best) の解を出力する [3]。

2　従来の翻訳モデル

句に基づく翻訳モデルは次式で表される。

$$p(\bar{f}_1^I|\bar{e}_1^I) = \prod_{i=1}^{I} \phi(\bar{f}_i|\bar{e}_i) d(a_i - b_{i-1}) \tag{1}$$

ここで、$\phi(\bar{f}_i|\bar{e}_i)$ を翻訳確率、$d(a_i - b_{i-1})$ を歪み確率と呼ぶ。翻訳モデルはこれら二つの確率を考慮する。式中の I は翻訳元言語 f の形態素の連なりの数、$\bar{f}_1^I$ はこれを句に分割したもの、$\bar{f}_i$ は分割したそれぞれの句、$\bar{e}_i$ は $\bar{f}_i$ に対応した句、a_i は新たに翻訳する句の左端の位置、b_{i-1} は直前に翻訳した句の右端の位置である。

翻訳確率は、次式による相対確率で算出する。

$$\phi(\bar{f}|\bar{e}) = \frac{count(\bar{f}, \bar{e})}{\sum_f count(\bar{f}, \bar{e})} \tag{2}$$

歪み確率は、次式によって算出する。

$$d(a_i - b_{i-1}) = \alpha^{|a_i - b_{i-1} - 1|}$$

これは、翻訳する翻訳元言語の句の位置のずれに依存するモデルである。すなわち、直前に翻訳した句の右端の位置と、次に翻訳する句の左端の位置の差の絶対値を考慮する。この例を図 1 に示す。例は英日翻訳であり、“help” の次に “disposed to” を翻訳しようとしている。

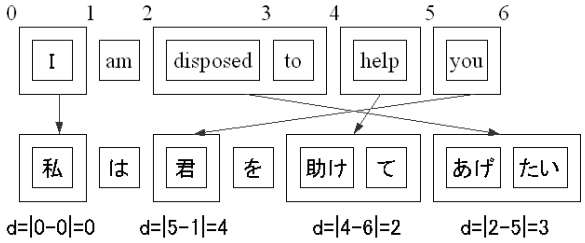

図 1　歪み確率モデル

ここで、a_i は新たに翻訳する句の左端の位置であり、“disposed to” の左端は 2 である。そして、b_{i-1} は直前に翻訳した句の右端の位置であり、“help” の右端は 5 である。よって、$d = |2 - 5| = 3$ となる。

この歪みモデルでは句の相対的な位置のみを考慮している。このため、英語の動詞は文頭近くにあるが日本語の動詞は文末にある、というような語順の違いを表現できない。語順が大きく異なる言語間では、句についての情報をもっと考慮した歪みモデルが必要だと考えられる。

3　提案モデル

本稿では、2 節で述べた句に基づく翻訳モデルの中の歪み確率について新しいモデルを提案する。従来の歪みモデルは、直前に翻訳した翻訳元言語句と新たに翻訳する翻訳元言語句との相対的な距離を考慮している。提案モデルは、この相対的な距離を絶対的な距離の 4 つのパタンに拡張し、これを翻訳順序パタンと呼ぶ。さらに、この翻訳順序パタンの発生に影響する因子として 4 つの句を考慮する。

3.1　翻訳順序パタン

図 2 を用いて翻訳元言語を翻訳する順序のパタンを考える。

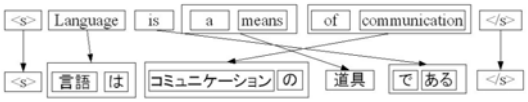

図 2　翻訳する順序のパタン

図では英日翻訳における対訳文の句対応付けを示している。<s> は文頭記号、</s> は文末記号である。ここで、デコーダの計算量削減のため、英語文を翻訳して日本語文

を文頭から生成していくことを考える。このとき、言語間の語順の違いがあるため、英語文も文頭から翻訳していけばいいというわけではなく、どの英語句から翻訳していくかという順序を考える必要がある。これを、直前に翻訳した翻訳元言語句に対して次に翻訳する翻訳元言語句がどの位置にあるかという次の 4 つのパタンとして考える。

(1) [正順] 直前に翻訳した翻訳元言語の句と次に翻訳する翻訳元言語の句が 文頭から文末方向に 連なっている。

(2) [逆順] 直前に翻訳した翻訳元言語の句と次に翻訳する翻訳元言語の句が 文末から文頭方向に 連なっている。

(3) [正順間隙あり] 直前に翻訳した翻訳元言語の句と次に翻訳する翻訳元言語の句が 文頭から文末方向に あり、かつ 間に別の句 がある。

(4) [逆順間隙あり] 直前に翻訳した翻訳元言語の句と今回翻訳した翻訳元言語の句が 文末から文頭方向に あり、かつ 間に別の句 がある。

図 2 でこれらのパタンの例を示す。最初に翻訳される “Language” は、直前に文頭記号を翻訳したと考えると文頭記号に対して文頭から文末方向へ連なっているので「正順」である。次の “of communication” は、直前に翻訳した “Language” とは正順でかつ間に別の句があるため「正順間隙あり」である。“a means” は “of communication” とは文末から文頭方向へ連なっているため「逆順」である。“is” も “a means” とは「逆順」である。もし、“is” と “a means” の間に別の句があるときは「逆順間隙あり」となる。

以上の 4 つの翻訳順序パタンを考える。これらのパタンがある条件下で発生する確率を考慮することにより、翻訳元言語をより適切な順序で翻訳できると考えられる。

3.2 翻訳順序パタンに影響する因子

次に、これらの翻訳順序パタンの発生確率に影響する因子として $\bar{e}_{i-1}$、$\bar{e}_i$、$\bar{f}_{i-1}$、$\bar{f}_i$ の 4 つを考える。$\bar{e}_{i-1}$ と $\bar{e}_i$ は隣り合っている翻訳先言語句である。$\bar{f}_{i-1}$ と $\bar{f}_i$ はそれぞれ $\bar{e}_{i-1}$ と $\bar{e}_i$ に対応する翻訳元言語句である。これを数式で表すと次のようになる。

$$p(d|\bar{e}_{i-1}, \bar{e}_i, \bar{f}_{i-1}, \bar{f}_i) \quad (3)$$

ここで、d は翻訳順序パタンを表している。すなわち翻訳順序パタンは、直前に翻訳した翻訳元言語の句とそれに対応する翻訳先言語の句、及び次に翻訳する翻訳元言語の句とそれに対応する翻訳先言語の句の 4 つの因子によって決定される。図 2 に示した例では、直前に翻訳した翻訳元言語句 $\bar{f}_{i-1}$ を “Language”、次に翻訳する翻訳元言語句 $\bar{f}_i$ を “of communication” だとすると、前者に対応する翻訳先言語句 $\bar{e}_{i-1}$ は「言語 は」、後者に対応する翻訳先言語句 $\bar{e}_i$ は「コミュニケーションの」となり、翻訳順序パタンは正順間隙ありとなる。

3.3 歪みモデルの種類

ここで、考慮する因子の数を変化させた 5 つの歪みモデルを考える。まず、翻訳順序パタンのみを考慮したモデルが考えられる。次に、$\bar{e}_i$ を決定するときに影響が強いのは $\bar{f}_i$, $\bar{e}_{i-1}$, $\bar{f}_{i-1}$ の順であるとすると、影響力が強い順に因子を考慮することで式 4 から式 8 までの歪みモデルが考えられる。

$$p(d) \quad (4)$$
$$p(d|class(\bar{f}_i)) \quad (5)$$
$$p(d|class(\bar{e}_{i-1}), class(\bar{f}_i)) \quad (6)$$
$$p(d|class(\bar{e}_{i-1}), class(\bar{f}_{i-1}), class(\bar{f}_i)) \quad (7)$$
$$p(d|class(\bar{e}_{i-1}), class(\bar{e}_i), class(\bar{f}_{i-1}), class(\bar{f}_i)) \quad (8)$$

以降では、これらの歪みモデルを先頭から順に歪みモデルタイプ 1、2、3、4、5 と呼ぶ。

ここで考えた影響力の強さは、IBM Model 4 の歪みモデルとのアナロジーに基づいている。この歪みモデルでは、次に翻訳する翻訳元言語に対応する翻訳先言語の単語のクラス、および直前に翻訳した翻訳元言語の単語のクラスを考慮している。これは、$\bar{e}_{i-1}$ と $\bar{f}_{i-1}$、$\bar{e}_i$ と $\bar{f}_i$ があるとき、それぞれの組は互いに翻訳になっているので強い関係があると言えるからである。

3.4 歪みモデルの汎化

```
-1 から are you using it|1|1
-1 から can i order this|1|2
```

図 3 句の表記による歪みモデル

図 3 はタイプ 3 の歪みモデルのモデルの一部である。一行はスペース及びパイプによって区切られており、行の先頭から翻訳順序パタン ID、$\bar{f}_i$、$\bar{e}_{i-1}$、確率、頻度である。翻訳順序パタン ID は、-1 が逆順、-2 が逆順間隙あり、1 が正順、2 が正順間隙ありを表している。図 4 および図 5 で示すモデルも同様の書式である。例えば、直前に “are you using it” を翻訳していて、次に翻訳する句の翻訳先言語側が「から」のとき、翻訳順序パタンが -1、すなわち逆順である確率は 1 である。このモデルでは句の表記をそのままモデルにしているため、頻度が低くて有効な確率を得られない場合がある。そこで「句の品詞」と「クラスタリング」を導入することでモデルを汎化する。

3.4.1 句の品詞

言語間における語順の違いは、英語の動詞が文頭付近にあるのに対して日本語の動詞は文末にあるなど、ある程度品詞で説明することが出来る。しかし、単語には品詞が存在するが、句の品詞というものは存在しない。

そこで、句の先頭もしくは末尾の単語の品詞を句の品詞として扱う。日本語の場合、係り受けが文頭から文末方向であるため、句の末尾の単語の品詞を用いる。例えば「私 は 日本 の」という句であれば、末尾の単語「の」の品詞「助詞」を句の品詞とする。逆に、英語や中国語の係り受けは文末から文頭方向であるため、句の先頭の単語を用いる。句の品詞の有効性については、これを用いて作成した歪みモデルで実験を行い、翻訳精度が向上したことを以前に報告した [5]。ただし、以上の方法で句の品詞を決定すると言語に依存してしまうが、句の先頭及び末尾の単語の品詞を両方用いれば言語非依存となる。句の品詞を用いた歪みモデルタイプ 3 の例を図 4 に 2 行だけ示す。

```
-1 動詞-自立 NNP|0.625|10
-1 動詞-自立 NNP|0.01428571428571143|2
```

図 4 句の品詞による歪みモデル

3.4.2 句のクラスタリング

IBM Model 4 では歪みモデルに単語クラスタリングを用いる。そこで、句の歪みモデルでも句のクラスタリングを使用する。クラスタリングのツールとして、GIZA++ で IBM Model を推定する際の単語クラスタリングに用いる mkcls[1] を使う。これは、bigram を元に最尤推定を行い、単語を指定された数のクラスに分類する。入力として一行一文かつスペースにより区切られた単語群を受け取る。

単語のクラスタリングツールで句のクラスタリングをするため、句を単語として扱う。句の対応付けが得られているとき、対応付けの中でひとつの句に属している単語をすべてアンダーバーで結ぶ。その他の部分、すなわち句の境界となっている部分はスペースを残す。このように句を構成する単語をすべて結ぶことにより、mkcls に句を単語として扱わせる。句のクラスタリングによって作成した歪みモデルタイプ 4 の例を図 5 に 2 行だけ示す。

```
-1 1 2 5|0.0128462464873545|32
-1 1 3 1|0.8304020100502 51|661
```

図 5　クラスタリングによる歪みモデル

3.5　歪みモデルの学習方法

本稿で提案する歪みモデルを学習する手順を述べる。まず句翻訳モデルを作成し、このモデルを用いて対訳コーパスの対応付けを行う。対応付けが得られれば、直前に翻訳した句に対して次に翻訳する句がどこにあるか、という翻訳順序パタンがわかる。この頻度から歪みモデルを作成する。

対訳コーパスの対応付けにはデコーダを用いる。デコーダについては 4 節で述べる。デコーダが翻訳元言語の文に応じた翻訳先言語の文を生成するという問題において、与えられた翻訳先言語の文だけを生成するという制約をかけると、それは対訳文の対応付けを求める問題と同じになる。

例えば、“Language is a means of communication” の翻訳文を探索するのがデコーダの機能である。ここで、デコーダに「言語/は/コミュニケーション/の/道具/です」という翻訳文しか生成しないという制約をかける。すると、デコーダの出力結果は対訳文の翻訳先言語側と常に同じになる。このとき、翻訳元言語のどの部分をどの翻訳先言語の句に翻訳したのかを見れば対訳文の対応付けが得られる。

ただし、デコーダが探索しても与えられた翻訳先言語の文にならない可能性がある。この場合、句の対応付けが得られないので、その対訳文はモデル構築のための学習に用いることはできない。ここで、翻訳先言語の文が与えられるため探索範囲は通常の翻訳より狭くなっていることから、通常のデコーダよりも枝狩りの条件を緩くする。これにより、対応付けが求められない対訳文を少なくできる。

デコーダは N-best の対応付けを求められるため、どれだけの解を学習に使うかによって翻訳精度が変化する。

4　デコーダ

デコーダは前向きビーム探索でワードグラフを構築し、このグラフに対して後ろ向き A* 探索を行うことにより N-best の解を探索する [3]。ワードグラフを構築した時点でスタートノードから各ノードへの確率は算出できているため、これがそのまま予測確率となる。すなわち、予測確率は常に正しい。予測確率が最大の経路をたどっていくことで、簡単に確率最大の解を求めることができる。2 番目に確率の高い解以降は、確率最大の解からの分岐を比較して最も確率の高い経路をたどることで N-best の解を求められる。ビーム探索であるためモデルの最適解が求まる保証は無いが、構築したワードグラフにおける最適解を正しく求められる。

デコーダでは、翻訳確率及び歪み確率のほかに、句を構成する単語により計算する句の翻訳確率 Phrase Translation Probability[4] を使用している。

5　実験

5.1　使用するコーパス及びツール

句の品詞による歪みモデルでは品詞を使用している。品詞のタグ付けには、日本語は ChaSen[2]、英語は MXPOST[3]を使用した。MXPOST の品詞体系は Penn Treebank Tagset であり、その品詞タグ数は 45 である。ChaSen の品詞体系は深さ 4 まで階層化されており、本実験では第 2 階層までの品詞を使用した。その品詞タグ数は 50 である。

句翻訳モデルの構築には Pharaoh Training[4]を用いた。句の抽出方法はデフォルトの grow-diag-final である。言語モデルは、Palmkit[5]により作成した backoff 3-gram を用いた。翻訳モデルの学習に用いる対訳コーパスの英語側のみを言語モデルの学習コーパスとして使用した。ただし、英語はすべてを小文字に変換している。

コーパスとして、IWSLT(International Workshop on Spoken Language Translation) 2005 で使用されたコーパスを用いる。IWSLT は多言語話し言葉翻訳技術の評価に関する国際ワークショップである。翻訳入力文として、普通のテキストと音声認識結果のテキストの 2 種類が用意されている。本稿では普通のテキストを使用し日英翻訳を扱う。コーパスのドメインは旅行会話であり、学習コーパス 20000 対、開発セット 500 文およびテストセット 500 文を用いた。

デコーダの確率の重みはすべて 1 とした。Minimum Error Rate Training は行っていない。

5.2　結果

提案したモデルによる翻訳の結果を図 6 に示す。横軸は歪みモデルの種類で、縦軸は BLEU のスコアである。歪みモデルの学習には上位 100 位の対応付けを用いており、クラスタリングのクラス数は 5 である。4 種の歪みモデルすべてにおいてタイプ 2 まではベースラインを上回っているが、タイプ 3 以降はスコアが急激に落ちている。これは頻度の不足が原因だと考えられる。歪みモデルタイプの数が大きくなればなるほど多くの情報を考慮するため、頻度は逆に低下する。句の表記を使ったモデルは翻訳に用いる句の異なり数だけバリエーションがある。句の品詞は 50 種類あるため、ひとつの品詞を用いた歪みモデルタイプ 3 では 2500、タイプ 4 では 125000 と非常に多様になる。

クラス数 5 でクラスタリングした結果は、タイプ 2 以降のスコアにほとんど差が無い。これはクラス数 5 であるため頻度の不足とは考えにくい。タイプ 3 以降で考慮している直前に翻訳した句の情報は、翻訳順序パタンの決定に寄与していないと考えられる。

句の表記とクラスタリングの歪みモデルではタイプ 5 でスコアが上昇しており、特に句の表記の上昇幅は大きい。これは、タイプ 3 および 4 で考慮している直前に翻訳した句の情報よりも、タイプ 5 で考慮している次に翻訳する句自

[1] http://www.fjoch.com/mkcls.html

[2] http://chasen.aist-nara.ac.jp/

[3] http://www.cis.upenn.edu/~adwait/statnlp.html

[4] http://www.iccs.informatics.ed.ac.uk/

[5] http://palmkit.sourceforge.net/

身のクラスが重要であることを示唆していると言える。歪みモデルのタイプには改善の余地がある。

次に、クラスタリングによる歪みモデルにおいて、クラスタリングのクラス数を変化させたときの結果を図 7 に示す。歪みモデルの学習には上位 100 位の対応付けを用いている。クラス数 20 のときのタイプ 2 が一番良く、クラス数 10 や 30 でも同程度のスコアである。句の品詞で扱っている品詞の種類は 50 程度あるので、これを減らすことで句の品詞を用いた歪みモデルの精度が向上する可能性がある。

最後に、歪みモデルの学習に用いる N-best の数を変化させたときのクラスタリングによる歪みモデルの結果を図 8 に示す。クラス数は 20 である。5-best のスコアはベースラインより悪いが、10-best 以上を学習すればベースラインを上回る。学習量を 100-best まで増やしても精度に大きな変化はないため、未知の句が少なくなるよう多くの N-best を学習させる方が良い。

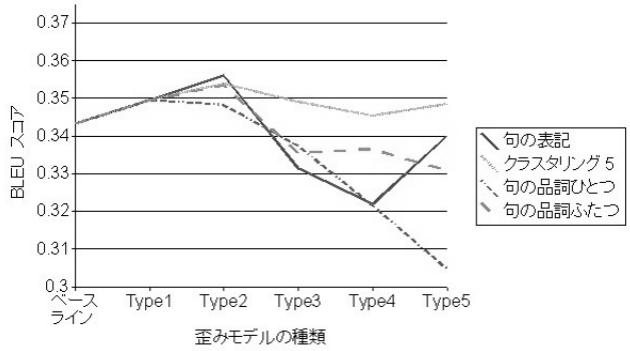

図 6　提案する歪みモデルによるスコアの改善

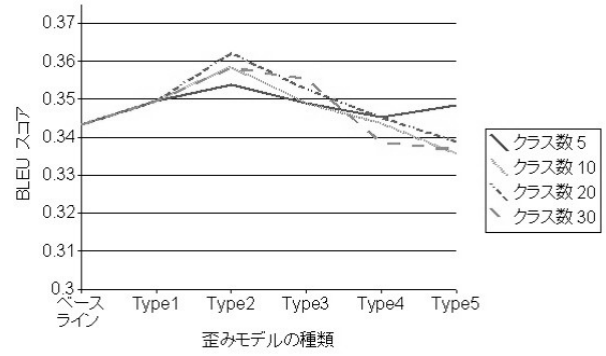

図 7　クラスタリングのクラス数を変化させる

6　従来研究との比較

翻訳順序パタンを Left(逆順), Right(正順) および Neutral(正順間隙ありおよび逆順間隙あり) の 3 つのパタンとし、直前の翻訳順序パタンおよび句によって次の翻訳順序パタンおよび句が決まるというモデルが提案されている [6]。本研究のモデルは、翻訳順序パタンを 4 つに分け、直前及び次の句によって次の翻訳順序パタンが決まるというものであり、さらにこれを汎化している。

それから、ふたつの句が正順もしくは逆順でつながっている状態を文全体で保つという制約をかけることで句の並びを改善する ITG 制約がある [7]。本研究は翻訳順序を分類するが制約としては使用せず、パタンの出現をコーパスから学習する。

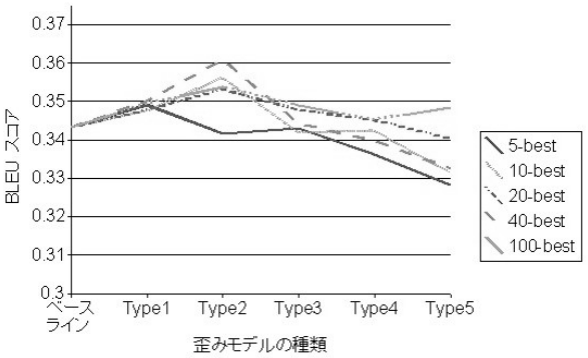

図 8　歪みモデル学習に用いる N-best を変化させる

7　おわりに

本稿では、句に基づく翻訳モデルにおける新たな歪みモデルの提案を行った。モデルは、直前に翻訳した翻訳元言語句に対して次に翻訳する翻訳元言語句がどの位置にあるかという翻訳順序のパタンを 4 つに分類し、さらにこれらの翻訳順序パタンの発生に影響する因子を 4 つ考慮する。さらに、これらの因子をクラスタリングおよび句の品詞によって汎化した。提案する歪みモデルを実装し実験をしたところ、自動翻訳評価指標 BLEU において従来の歪みモデルよりも高い精度を得ることが出来た。

参考文献

[1] Philipp Koehn, Franz J. Och, and Daniel Marcu. Statistical phrase-based translation. In *Proceedings of HLT-NAACL 2003*, pp. 127–133, 2003.

[2] Franz Josef Och and Hermann Ney. The alignment template approach to statistical machine translation. *Computational Linguistics*, Vol. 30, pp. 417–449, 2004.

[3] Nicola Ueffing, Franz Josef Och, and Hermann Ney. Generation of word graphs in statistical machine translation. In *Proceedings of EMNLP 2002*, pp. 156–163, 2002.

[4] Stephan Vogel, Ying Zhang, Fei Huang, Alicia Tribble, Ashish Venugopal, Bing Zhao, and Alex Waibel. The CMU statistical machine translation system. In *Proceedings of MT Summit IX*, pp. 23–27, 2003.

[5] Kazuteru Ohashi, Kazuhide Yamamoto, Kuniko Saito, and Masaaki Nagata. NUT-NTT Statistical Machine Translation System for IWSLT 2005. In *Proceedings of International Workshop on Spoken Language Translation*, pp. 128–133, 2005.

[6] Christoph Tillmann and Tong Zhang. A Localized Prediction Model for Statistical Machine Translation. In *Proceedings of ACL'05*, pp. 557–564, 2005.

[7] Richard Zens, Hermann Ney, Taro Watanabe, and Eiichiro Sumita. Reordering Constraints for Phrase-Based Statistical Machine Translation. In *Proceedings of Coling 2004*, pp. 205–211, 2004.