

統計翻訳指標を導入した構文トランスファに基づく用例翻訳

今村 賢治, 大熊 英男, 渡辺 太郎, 隅田 英一郎

ATR 音声言語コミュニケーション研究所

{kenji.imamura,hideo.okuma,taro.watanabe,eiichiro.sumita}@atr.jp

本稿では、統計翻訳のモデルを利用して最適訳選択を行う、構文トランスファ方式の用例翻訳器を提案する。従来用例翻訳は、原言語の類似度だけを利用して入力文と似た用例を選択していたため、翻訳文が誤ることがあった。本稿で提案する方式は、統計翻訳と同様に、翻訳モデルと言語モデルを利用して、入力文に最も適合する出力単語列を選択するため、用例翻訳単体に比べ、翻訳品質を向上させることができる。本方式の特徴の一つは、構文トランスファで語順変更を行った後にモデルを適用するため、語彙モデルだけでも翻訳品質を確保することが可能なことである。さらに、構文トランスファが出力する木構造を利用してボトムアップにモデルを適用することにより、翻訳速度を向上させた。

Example-based Machine Translation Based on Syntactic Transfer with Statistical Models

Kenji Imamura, Hideo Okuma, Taro Watanabe, and Eiichiro Sumita

ATR Spoken Language Translation Research Laboratories

{kenji.imamura,hideo.okuma,taro.watanabe,eiichiro.sumita}@atr.jp

This paper presents example-based machine translation (MT) based on syntactic transfer, which selects the best translation by using models of statistical machine translation. Example-based MT sometimes generates invalid translations because it selects similar examples to the input sentence based only on source language similarity. The method proposed in this paper selects the best translation by using a language model and a translation model in the same manner as statistical MT, and it can improve MT quality over that of ‘pure’ example-based MT. A feature of this method is that the statistical models are applied after word re-ordering is achieved by syntactic transfer. This implies that MT quality is maintained even when we only apply a lexicon model as the translation model. In addition, translation speed is improved by bottom-up generation, which utilizes the tree structure that is output from the syntactic transfer.

1 はじめに

対訳コーパス充実の流れに伴い、コーパスから知識、モデルを自動獲得する翻訳方式が提案されてきている。現在、その方式として、大きく用例翻訳と統計翻訳の2種類がある。

用例翻訳 (Nagao, 1984) は、対訳コーパスを一種のデータベースと見なし、入力文と似た用例を用例

ベースから検索する。そして、用例の目的言語側を対訳辞書等を用いて修正・更新することにより翻訳を行う。用例には句や文を単位とした対訳が用いられることが多く、格関係や慣用表現を考慮した翻訳ができることが特徴である。しかし用例翻訳は、検索時に用例同士が競合したとき、原言語の情報だけを使ってスコアリングしているため、最終的に得られた翻訳文が入力文の翻訳として正しいかどうか、

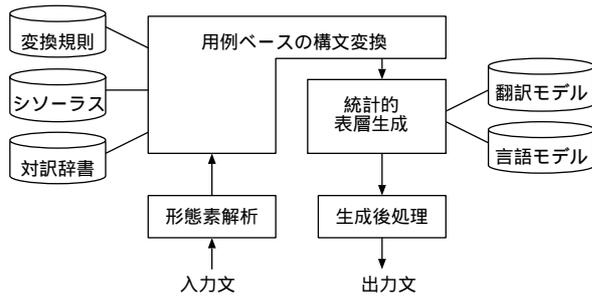


図 1: 本方式の構成

わからないという問題点がある。

一方、IBM モデルを用いた統計翻訳 (Brown et al., 1993) は、単語翻訳と語順調整を組み合わせる翻訳を行っている。そのため、日本語と英語のように、語順が著しく異なる言語対に適用した場合、探索空間が非常に大きくなり、大域最適解を見つけることが困難になる (Watanabe and Sumita, 2003)。

しかし、統計翻訳は、大域最適解の探索に成功した場合、良質な訳を出力することが可能である。つまり、統計翻訳で用いられているモデルは、機械翻訳の訳質を判断する指標としても優れている。この特徴を利用し、Akiba et al. (2002) は、複数の機械翻訳システムが出力した翻訳文を、統計翻訳の指標にしたがってランキングし、最適訳の選択に成功している。

本稿では、用例翻訳と統計翻訳の弱点をカバーする方式として、統計翻訳のモデルを用いて最適訳選択を行う、構文トランスファ方式用例翻訳を提案する。本方式は、大きく 2 つのモジュールから成り立っている (図 1)。一つは、用例ベースの構文変換部である。ここでは変換規則を参照しながら構文解析やノードのマッピングを行い、目的言語の木構造を作成する。もう一つは、統計的表層生成部である。ここでは目的言語の木構造から生成される単語列のうち、最適な列を統計翻訳と同様の方式を用いて決定する。つまり、本方式は用例翻訳と統計翻訳を直列につなげた構成となっている。

本方式は、以下の特徴を持つ。

- 用例翻訳からみた場合、入力文と用例の原言語側の類似性だけでなく、目的言語の正しさ、および単語訳の正しさを判断して翻訳結果を選択するため、訳質が向上する。
- 統計翻訳からみた場合、構文トランスファにより語順変更が行われた後に探索を行うため、比較的単純なモデルでも適切な翻訳文を得ることができる。また、用例翻訳である程度正しい訳文候補を絞り込んだ上でモデルを適用するため、探索空間が狭められる。

以下、第 2 章では、用例ベースの構文変換部について説明し、第 3 章では、統計的表層生成部について説明する。第 4 章では、実験システムについて評価を行い、第 5 章では他の用例翻訳・統計翻訳のハイブリッド方式との違いについて述べる。

2 用例ベース構文変換

本稿で用いる用例ベースの構文変換は、階層的句アライメントベース翻訳システム (HPAT; (Imamura, 2002) 参照) を改良したものである。以下、その概要を、日英翻訳を例に取り説明する。

2.1 変換規則

変換規則は、階層的句アライメント (今村, 2002) を用いて、対訳コーパスから自動獲得することにより作成される。階層的句アライメントでは、対訳文の構文解析を行い、構文木のノードレベルで原言語・目的言語間の対応づけを行う。そのノードレベルの対応を規則化したものが変換規則である。その例を図 2 に示す。図 2 中の X, Y 等は、原言語と目的言語間で対応する非終端記号を表す。変換規則は、同期文脈自由文法と見なすことができる。

同期文脈自由文法と変換規則の相違点は、変換規則には用例が付与されている点である。用例は、訓練文に現れた、各規則の非終端記号の主辞の実例である。たとえば、図 2 の規則 1 は、日本語動詞句 “フライト (を) 予約する” が、英語動詞句 “make a reservation (予約する) for the flight (フライトを)” に翻訳されたことを意味する。

2.2 構文トランスファ処理概要

実際の変換時には、上記変換規則を利用して、以下の 3 ステップで目的言語の構文木を作成する。

1. まず、変換規則の原言語文法を用いて、入力文を構文解析する。

番号	原言語文法	⇒	目的言語文法	用例
1	VP → X _{PP} Y _{VP}		VP → Y _{VP} X _{PP}	((フライト, 予約する) ..)
2			VP → Y _{VP} X _{ADVP}	((そこ, 行く) ..)
3			VP → Y _{BEVP} X _{NP}	((橋, ある) ..)
4	S → X _{NP} は Y _{VP} ます	⇒	S → X _{NP} Y _{VP}	((彼, 演奏する) ..)
5			S → X _{NP} will Y _{VP}	((バス, 停まる) ..)

図 2: 変換規則の例

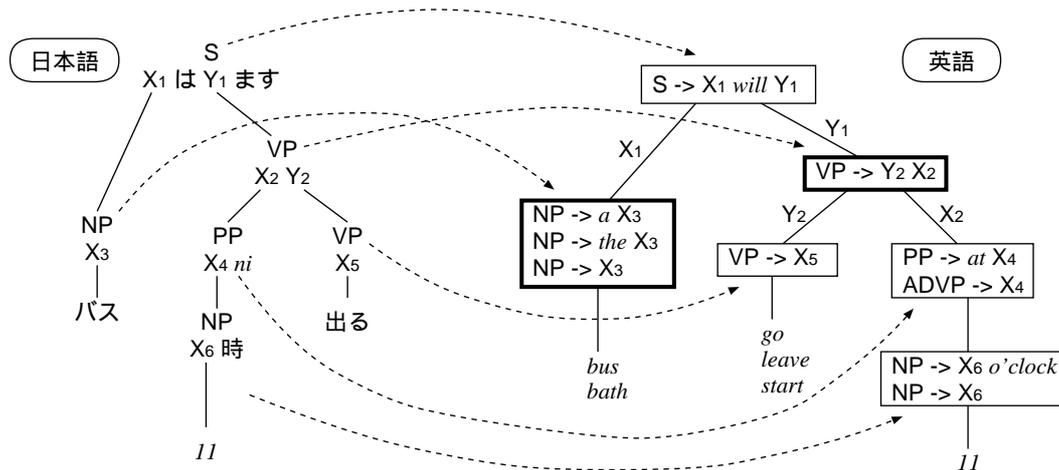


図 3: 構文トランスファによる変換過程
(太枠は、本文から参照されている構文ノードを表す。)

- 次に、原言語側の構文木の各ノードを、目的言語にマッピングする。
- 目的言語の木構造の葉に非終端記号が残る場合は、対訳辞書を参照して、訳語候補を埋め込む。

入力文「バスは11時に出る」が与えられたときの変換処理の例を図3に示す。ここで注意したい点が2つある。まず、原言語・目的言語間で語順が異なる場合、構文トランスファによって語順が逆転した候補を生成している点である(太枠のVPノード参照)。つまり、語順調整は、構文トランスファによって行われている。もう一つは、原言語・目的言語間で対応しない単語(*a, the*等)は、目的言語文法によって自動的に埋め込み、または消去されている点である(太枠のNPノード参照)。つまり、構文トランスファ用の変換規則は、IBMモデルの *distortion, NULL, fertility* モデルと同等の機能を果たしている。

2.3 用例の利用

用例ベースの構文変換部では、変換時、および構文解析時の曖昧性解消に、図2の用例を利用する。具体的には、入力文の主辞と用例を比較し、最も意味距離 (Sumita and Iida, 1991) が近い規則だけを利用して目的言語の構文木を作成する。単語間の意味距離は、シソーラス (大野・浜西, 1984) 上におけるノード間の距離として定義され、Most Specific Common Abstraction (MSCA) から算出される。

たとえば、入力の句が、「家に帰る」である場合、図2の規則1から3が適用され、変換後、3種類のノードが作成される。しかし、Xの主辞='家'、Yの主辞='帰る'と用例を比較した場合、シソーラスを参照すると、X='そこ'、Y='行く'との意味距離が最も近くなるため、規則2のノードだけが作成される。なお、現在は同距離の規則がある場合はどちらも目的言語にマップするようにしている。

このように、用例ベースの構文変換部では、用例

との意味距離を使い、格関係や慣用表現を考慮した翻訳を行っている。

3 統計的表層生成、最適訳選択

3.1 翻訳モデル、言語モデル

統計的表層生成部では、用例ベースの構文変換部が出力する木構造から生成される目的言語単語列のうち、最適な列を探索する。「最適」な組み合わせは、統計翻訳と同様に、言語モデルと翻訳モデルから決定する。すなわち、入力単語列を F 、出力単語列を E としたとき、以下の式を満足する \hat{E} を探索する。

$$\begin{aligned}\hat{E} &= \operatorname{argmax}_E P(E|F) \\ &= \operatorname{argmax}_E P(E)P(F|E)\end{aligned}\quad (1)$$

本稿では、Vogel et al. (2003) と同様に、翻訳モデルとして語彙モデル (Lexicon Model) だけを用いる。すなわち、 f, e をそれぞれ入力、出力単語としたとき、翻訳モデルの確率値は以下の式で計算される。

$$P(F|E) = \prod_j \sum_i t(f_j|e_i)\quad (2)$$

IBM モデルの場合、他にも fertility, NULL, distortion モデルを用いているが、用例ベースの構文変換と組み合わせる場合、変換部でこれらモデルに相当する変換を行っているため、最も基本的な語彙モデルを適用するだけでも翻訳品質を確保できると考えられる。

また、言語モデルには、標準的な単語 n-gram を用いる。

3.2 ボトムアップ型表層生成

用例ベースの構文変換部が出力する木構造は、直線化すると単語グラフを作ることができる。しかし、元の木構造は同じ原言語を指し示すノードは既に共有されており、直線化しながら確率値を算出することも可能である。これを本稿ではボトムアップ型表層生成と呼ぶ。

図 4 は、図 3 の構文木が得られたときのボトムアップ型表層生成の例 (一部) である。それぞれのノードでは、下位のノードから得られた単語列から、まずラティス構造を作成する。そして、ラティスのすべての組み合わせの中から、n-best の単語列を選択する。その際、下位ノードで既に計算された翻訳モデル、言語モデルの確率値を利用し、その部分列に対しては再計算しない。このように計算された新しい単語列と、翻訳・言語モデル確率値を上位ノードに引き継ぐ。なお、翻訳モデル計算時には、トランスファのマッピングを逆にたどって入力の単語列を取得し、その範囲内で確率値を計算する。つまり、翻訳モデルは入力文全体に対して計算するのではなく、対応する句の範囲内だけに局所的に適用される。

セット名	項目	英語	日本語
訓練セット	文数	152,170	
	総形態素数	886,708	1,007,484
テストセット	文数	510	
	総形態素数	2,973	3,340

表 1: コーパスサイズ

生成が最上位ノードに達すると、言語モデル確率は文頭・文末記号と共に再計算され、ソートされる。結果的に、図 3 の木構造から、“*The bus will leave at 11 o'clock*” が得られる。

このようにボトムアップにデコードすることにより、共有されたノードの確率値は一度計算するだけでよく、木構造の情報を効果的に利用することができる。

このようにボトムアップにデコードすることにより、共有されたノードの確率値は一度計算するだけでよく、木構造の情報を効果的に利用することができる。

4 実験

用例翻訳に対する統計翻訳モデル導入の効果を測るため、ここでは日英翻訳を対象に、いくつかの方式を比較する。

4.1 実験条件

対訳コーパス 本稿で用いたコーパスは旅行会話基本表現集 (Takezawa et al., 2002; Kikui et al., 2003) で、旅行会話に頻出する表現を集めたものである。このコーパスを、表 1 に示すように、訓練セット、テストセットに分割して実験した。

変換規則 変換規則は、階層的句アライメントを用い、上記訓練コーパスから獲得した。なお、訓練

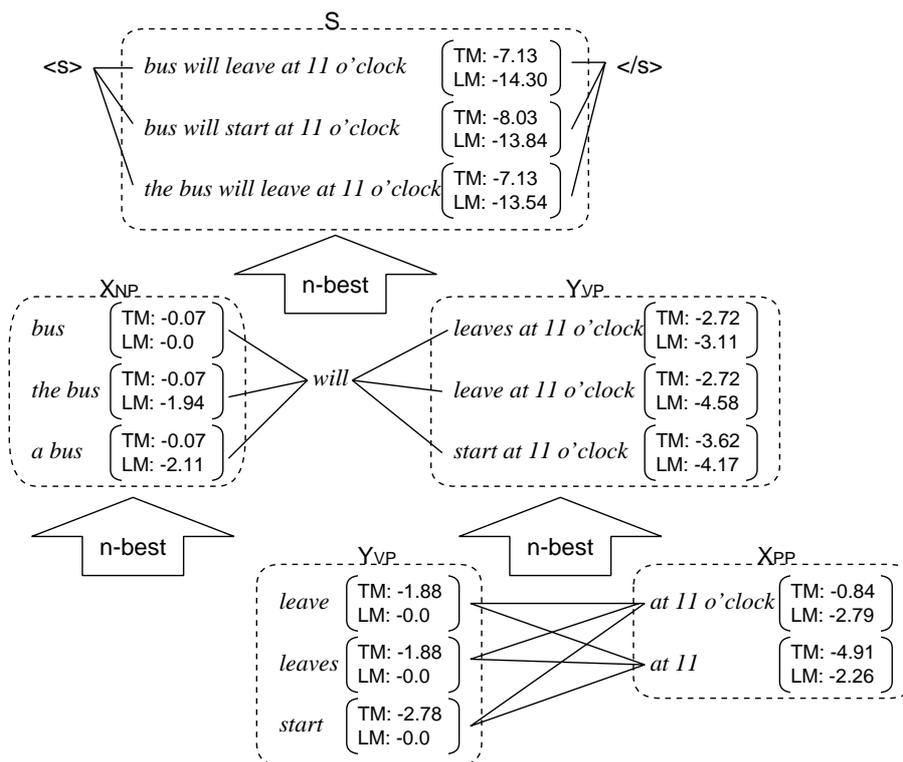


図 4: ボトムアップ型表層生成の例

(TM は翻訳モデルの log 確率値、LM は言語モデルの log 確率値を表す)

セット上で頻度 1 以下の規則を除外して使用した。変換規則の総数は、24,310 規則である。

翻訳モデル、言語モデル 翻訳モデルは、GIZA++(Och and Ney, 2003) を用いて学習した IBM モデル 4 の語彙モデルを使用した。また、言語モデルは、CMU-Cambridge Statistical Language Modeling Toolkit (Clarkson and Rosenfeld, 1997) を用いて、単語 bigram および trigram を学習した。

比較方式 今回、以下の 4 方式について比較を行った。

- ベースライン (用例変換のみ)

用例ベース変換部が出力した木構造から、意味距離が最も小さい候補を一つランダムに選択した。ただし、対訳辞書については、あらかじめ最も頻度の高い訳語を一つに絞り込んで使用した。これは、用例ベースの構文変換だけで翻訳を行ったときのベースラインである。

- ボトムアップ型表層生成

ボトムアップ型表層生成を用いて、最も確率値の高い候補を選択した場合。この実験では、100-best を選択した。

- 全探索型表層生成

用例ベースの構文変換部の出力から作成された単語グラフを全探索し、最も確率値の高い候補を選択した場合。つまり単語グラフを探索空間としたときの全域最適解。

- 言語モデルのみ

用例ベースの構文変換部の出力から単語列を全探索し、言語モデルの確率値の最大候補を選択した場合。

評価方法 表 1 のテストセット 510 文について、以下の評価を行った。なお、自動評価に用いた参照訳は、1 文あたり 16 参照訳である。

BLEU: BLEU スコア (Papineni et al., 2002)。値が高いほど翻訳品質はよい。

NIST: NIST スコア (Doddington, 2002)。値が高いほど翻訳品質はよい。

mWER: すべての参照訳との Word-error-rate のうち、最も誤り率の低いもの。値が低いほど翻訳品質はよい。

主観評価: 英語ネイティブ話者 1 名による主観評価。(A: 完全訳, B: 部分訳, C: 理解可能訳, D: 不可訳)。

4.2 実験結果

各方式における翻訳品質と、翻訳速度を表 2 に示す。

まず、ボトムアップ型と用例ベースの構文変換単体(ベースライン)を比較すると、どの評価方法に対しても、精度が向上している。このように、統計翻訳のモデルを用例翻訳に導入することは有効である。

また、ボトムアップ型と全展開型を比較すると、ボトムアップ型は、若干翻訳品質が低い。ボトムアップ型表層生成は、翻訳モデルを局所的に適用するため、対応する部分木に入らない単語アライメントが欠落した状態で確率を計算されることがある。そのため、大域最適解とは一致しない場合がある。

言語モデルだけを使用して最適訳選択を行った場合、ベースラインに比べ翻訳品質は向上したが、翻訳モデルを併用した場合(ボトムアップ型、全展開型)に比べると、かなり低下した。これは、本稿で使用した n-gram 言語モデルが出力長を考慮していないため、短すぎる訳が選択されたことが原因である。言語モデルは、最適訳選択に対してある程度までは有効であるが、単語訳の正しさまでは評価できないため、訳質向上には翻訳モデルは必要である。

最後に各方式の翻訳速度について述べると、ボトムアップ型では、全展開型に比べ、最大翻訳時間が著しく向上した。ボトムアップ型では木構造の共有ノードを効率的に利用しているため、速度を向上させることが可能である。そのため、音声対話翻訳のような実時間性が要求される翻訳タスクに適合すると考えられる。

5 議論

本稿では、用例翻訳に対して、統計翻訳のモデルを導入したが、統計翻訳の初期解を用例翻

訳を用いて与える方法は提案されている。たとえば、Marcu (2001) は、あらかじめ句単位に分解しておいた対訳を組み合わせて初期翻訳文を生成し、そこから統計翻訳のモデルを適用して初期翻訳文を書き換える方式を提案している。Watanabe and Sumita (2003) は、文単位に入力文と似た対訳を取得し、そこから書き換えるデコーダを提案している。

これらの方式との違いは、書き換え処理を行うか否かという点である。本稿で提案した方式は、用例翻訳が出力する単語列のうち、最もモデルに適合する組み合わせを選択している。用例翻訳は、従来から研究されているように、単体でもある程度の精度で翻訳することが可能であるため、Marcu (2001) や Watanabe and Sumita (2003) の初期翻訳文に比べ、品質は高いと推測される。しかし、本方式は書き換え処理を行わないため、用例翻訳が出力する結果中に、解が存在するという前提に立っている。つまり、本方式の性能の上限は、用例翻訳の性能によって制限されている。用例翻訳の精度向上には、翻訳品質の自動評価を用いた最適化法 (Imamura et al., 2003) が提案されており、これらを利用することによって、本方式の精度を更に向上させることが可能であると考えられる。

6 まとめ

本稿では、用例翻訳に対して、統計翻訳のモデルを適用することにより、精度を向上させることができることを示した。本稿で用いた用例翻訳は、構文トランスファを基にしているため、語順変更は変換部が行う。この特徴を利用し、語彙モデルと n-gram の言語モデルだけを使って最適単語列を選択し、用例翻訳単体に比べ高品質の翻訳を行った。また、出力の木構造を利用して、ボトムアップにモデルを適用することにより、翻訳速度を向上させた。

謝辞

類語新辞典体系の研究利用を許可してくださった、(株)角川書店に感謝いたします。

また、本研究は情報通信研究機構の研究委託「大規模コーパス音声対話翻訳技術の研究開発」により実施したものです。

方式	自動評価			主観評価			翻訳速度	
	BLEU	NIST	mWER	A	A+B	A+B+C	平均 (sec./文)	最大 (sec.)
ベースライン	0.410	9.06	0.423	51.6%	64.3%	70.4%	0.180	10.82
ボトムアップ	0.491	9.99	0.366	62.2%	72.5%	80.4%	0.211	5.03
全探索	0.498	10.04	0.353	62.9%	73.1%	80.8%	1.23	171.31
言語モデルのみ	0.491	9.11	0.385	57.6%	66.9%	72.0%	1.624	220.69

表 2: 生成方式別翻訳品質と翻訳速度

参考文献

- Yasuhiro Akiba, Taro Watanabe, and Eiichiro Sumita. 2002. Using language and translation models to select the best among outputs from multiple MT systems. In *Proceedings of COLING-2002*, pages 8–14.
- Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.
- Philip Clarkson and Ronald Rosenfeld. 1997. Statistical language modeling using the CMU-Cambridge toolkit. In *Proceedings of EuroSpeech 97*, pages 2707–2710.
- George Doddington. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the HLT Conference*, San Diego, California.
- Kenji Imamura, Eiichiro Sumita, and Yuji Matsumoto. 2003. Feedback cleaning of machine translation rules using automatic evaluation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003)*, pages 447–454.
- Kenji Imamura. 2002. Application of translation knowledge acquired by hierarchical phrase alignment for pattern-based MT. In *Proceedings of the 9th Conference on Theoretical and Methodological Issues in Machine Translation (TMI-2002)*, pages 74–84.
- 今村賢治. 2002. 構文解析と融合した階層的句アライメント. *自然言語処理*, 9(5):23–42, 10月.
- Genichiro Kikui, Eiichiro Sumita, Toshiyuki Takezawa, and Seiichi Yamamoto. 2003. Creating corpora for speech-to-speech translation. In *Proceedings of EuroSpeech 2003*, pages 381–384.
- Daniel Marcu. 2001. Towards a unified approach to memory- and statistical-based machine translation. In *Proceedings of 39th Annual Meeting of the Association for Computational Linguistics*, pages 386–393.
- Makoto Nagao. 1984. A framework of mechanical translation between Japanese and English by analogy principle. In *Artificial and Human Intelligence*, pages 173–180, Amsterdam: North-Holland.
- Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.
- 大野晋、浜西正人. 1984. 類語新辞典. 角川書店.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–318.
- Eiichiro Sumita and Hitoshi Iida. 1991. Experiments and prospects of example-based machine translation. In *Proceedings of the 29th ACL*, pages 185–192.
- Toshiyuki Takezawa, Eiichiro Sumita, Fumiaki Sugaya, Hirofumi Yamamoto, and Seiichi Yamamoto. 2002. Toward a broad-coverage bilingual corpus for speech translation of travel conversations in the real world. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, pages 147–152.
- Stephan Vogel, Ying Zhang, Fei Huang, Alicia Tribble, Ashish Venugopal, Bing Zhao, and Alex Waibel. 2003. The CMU statistical machine translation system. In *Proceedings of the 9th Machine Translation Summit (MT Summit IX)*, pages 402–409.
- Taro Watanabe and Eiichiro Sumita. 2003. Example-based decoding for statistical machine translation. In *Proceedings of Machine Translation Summit IX*, pages 410–417.