

翻訳支援における訳語信頼度

富士 秀、大倉 清司、長瀬友樹

㈱富士通研究所

fuji.masaru@jp.fujitsu.com

1. 概要

翻訳支援のための訳語信頼度表示機能を開発し、実用場面での有用性を検証した。表示機能は、対訳コーパスから訳語信頼度情報を抽出して翻訳者に提示する。産業翻訳の現場において翻訳者は訳語選択に膨大な工数を費やしているが、翻訳対象分野の大量対訳コーパスが用意できれば、ここから抽出された信頼度情報を使うことによって作業の効率化が見込まれる。

今回開発した表示機能は、形態素解析や、対訳コーパスに対する単語自動アラインメント技術等の基礎的な言語処理技術を利用しているが、実際の利用場面を徹底的に分析しこれに即した情報提示を行う設計としたことで、効率化のための有用な機能となった。

2. 背景

産業翻訳は、高い品質を維持しながら大量の専門文書を翻訳することが求められる業界であり、大きな市場を形成している。産業翻訳では、訳語選択の作業に大きな工数がかかっているが、今のところ人手による作業が中心となっている。

訳語選択の問題

産業翻訳では、翻訳者が訳語として用いる用語は、当該分野において適切な用語である確証が必要となる。このため翻訳者は、自分の使おうとする訳語の「裏取り」作業を行ってからはじめて、自信を持ってその訳語を使える状態になる。

このことから、翻訳者が翻訳作業全体の中で訳語選択に費やす時間の割合は大きなものとなっている。分野や作業手順によってこの割合は変わってくるが、場合によっては全体工数のうちの3割以上をこの作業に費やしている場合もある。現在人手で行われている訳語の裏取り作業を効率化することは、翻訳作業全体の効率化につながる。

訳語選択に用いられる言語リソース

人手作業において、裏取り作業のための最も理想的な言語リソースは、業界で権威のある対訳用語集である。このような用語集に掲載された訳語は信頼性が高く、語釈や用例も掲載されているため、確信を持って使うことができる。しかしながら、このような用語集がカバーするのは実際に出現する専門用語のほんの一部である。

次に信頼できるのが対訳用例であるが、これも翻訳者が利用できる用例数は限られており、また従来型

の検索システムをそのまま使うと検索して例文を読む作業に大きな手間がかかる。

これらのリソースで見つけることのできない多くの用語に関しては、単言語の言語リソースを利用して裏取りを行わざるを得ない。よく用いられるのが単言語の検索エンジンである。例えば、日英翻訳を行う場合、翻訳対象の日本語の単語について英語の訳語候補を想起し、これらの英語訳語候補を適当に組み合わせることで英語検索エンジンを検索して、その英語表現が適切であるかの検証を行う。

実際には、この後者の作業を行うことが多いが、手間がかかるということと、分野を指定した検索ができる検索エンジンは現実的には存在しないため、例文を読んで分野における適正もチェックしながら作業を行う必要性もある。

3. 解決手段

訳語選択効率化の要件

訳語の裏取りを効率よく行うためには、以下の要件を備えている必要があるが、前項で述べたように、これらをすべて満たす言語リソースは存在しない。

- ・ 対訳用語の形態になっている：最初から用例を読んで用語を拾う作業には時間がかかるため
- ・ 訳語を決定するための手がかりが提供されている：頻度、単語属性、出展情報、など
- ・ 対象分野に即した情報が提示される

最初に、候補を絞り込むための手がかり情報を訳語候補に付与して翻訳者に提示し、候補をある程度絞り込んだところで、実際の例文を見に行かれるようにすると効率化につながると考えられる。

解決のためのアイデア

まず、対訳用語を豊富に持っているのが機械翻訳用対訳辞書であるのでこれを利用する。機械翻訳用辞書そのままでは訳語決定のための根拠に乏しいため、対象分野のコーパスから頻度を付与した形で提示する。また単語対からだけでは判断できない時のために、単語と候補訳語の両方を含んだ対訳用例を効率よく表示する機能を提供する。

例えば、化学分野の翻訳において、「粉末を混ぜる。」という原文に対して翻訳者が“The powder is mixed.”という訳文を考えたとする。ここで、“powder”という訳語がどの程度信頼できるかを調べる一つの有力な方法は、化学分野の対訳コーパス中の出現頻度を利用することである。英辞郎を検

索すると、「粉末」に対応する訳語は、「flour」、「powder」、「dust」が掲載されているが、ある化学分野対訳コーパスで「粉末＝flour」、「粉末＝powder」、「粉末＝dust」が出現する頻度を数えてみると、頻度はそれぞれ 0、242、2 であった。このことから、「powder」という訳語の信頼性は高いと考えることができる。

本研究では、このような頻度情報を自動的に取得して表示するシステムを構築する。たとえば、システムが「powder」という訳語の信頼度が高いと判定した場合には下線表示とし、「To provide powder.」等と表示することができる。翻訳者は、信頼度が高ければそのまま追加調査を行わずに訳語を採用するという判断をすることができ、作業の効率化につながる。さらにこの例では、「粉末＝dust」も 2 回出現しているが、翻訳者は念のためこの例も確認しておきたい場合も出てくる。このために、システムでは、単語対を指定して例文を提示する機能も用意した。

4. 実験環境の構築

以上のアイデアを実現する訳語信頼度表示機能を構築した。なお、ここでは日英翻訳支援を例にとるが、任意の言語対に適用可能である。

4.1. 構成

対訳コーパスの構築

訳語選定に有効な頻度を抽出するのに十分な量の対訳コーパスを手で構築した。分量と有効性の相関に関する一般に認められた指標はないが、経験則から判断して、「電気」、「化学」等の科学技術分野の大分類レベルで、数十万文対の対訳コーパスを用意した。

形態素解析と訳語候補付与

対訳コーパスの原文・訳文それぞれについて、機械翻訳システムの形態素解析を適用し、文単位の単語（形態素）リストを作成する。機械翻訳用形態素解析では、各単語に訳語候補が付与されて出力される。

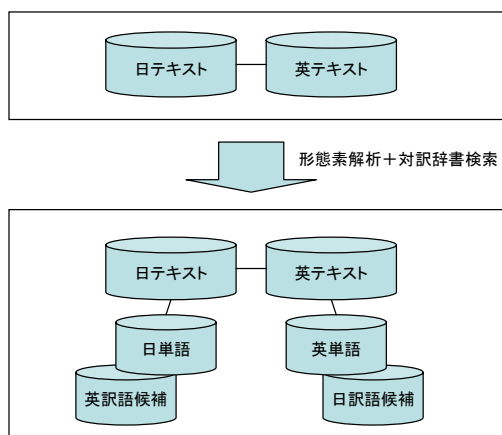


図 1. 対訳コーパスへの形態素解析 + 訳語候補付与

対訳コーパス中の対訳用語の推定

原文中の単語リストとその訳語候補群、訳文中の単語リストとその訳語群を、互いに照合し、対訳関係にある単語対を推定する。この処理は形態素解析までの処理のみを使っているため全体に軽い処理で済む。構文解析を行っていない分、誤った対応が見つかる可能性もあるが、対訳文を対象にしていることから、今回の用途には十分な推定精度が得られている。

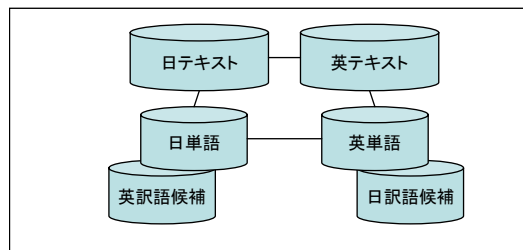


図 2. 対訳用語の推定

単語対応付き対訳コーパスの格納

前項で得られた単語対応付き対訳コーパスをデータベースに格納し、任意の入力単語対を含む対訳例文の出現頻度を返せるように、頻度情報を持った構成とする。さらに、入力単語対を入力としてそれを含む対訳例文を検索できるような構成とする。

訳語候補の信頼度計算と表示インターフェース

翻訳者が使うためのインターフェースを用意する。調査対象の入力単語に対して、まず対応する訳語候補を表示し、各訳語候補に頻度情報を付与して表示する。

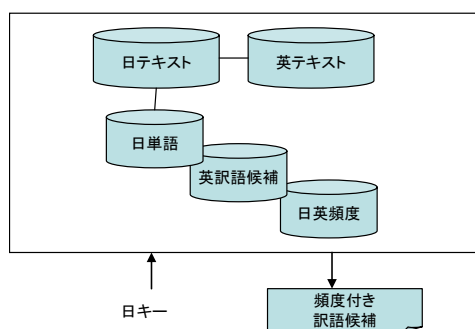


図 3. 頻度付き訳語候補の表示

対訳用語対を含む用例の表示

前述のデータベースでは、任意の入力単語対から、それら両方の単語を含む例文が検索できるようなインデックスを作成しているため、これを利用して、単語対を含む訳例の表示機能を構築した。

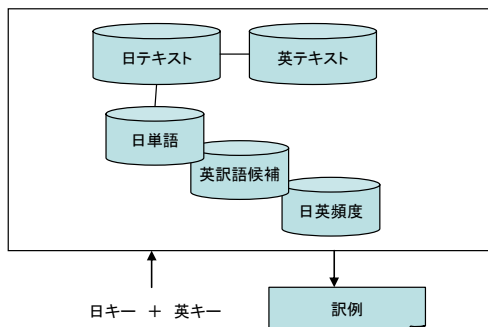


図 4. 入力単語対を含む用例の検索

4.2. 第 1 候補訳語への信頼度付与

訳語選択の効率化には、各単語対の出現頻度をもとにした、第 1 訳語候補への信頼度情報の付与がカギとなる。信頼度算出のもととなる頻度情報には以下のものがある。

絶対頻度

翻訳者にとってまず関心があるのが、対象分野における単語と訳語の対の出現頻度である。

相対頻度

ある単語において、単語⇔訳語候補 1 の共起頻度と、単語⇔訳語候補 2 の共起頻度に圧倒的な差があれば、翻訳者としては、頻度の高いほうを採用してよいと判断できる場合が多い。しかし、複数の訳語候補が接近した頻度を持っている場合には、それぞれの頻度が高かったとしても、それらいずれの訳語を採用すべきかを検討するのに手間がかかる。

そこで、訳語候補の頻度間の差の大きさを信頼度の指標とすることが求められる。

信頼度の設定

訳語候補の絶対頻度と訳語間の相対頻度をもとに、表 1.のような信頼度を設計した。実際の表示インタフェースでは、第 1 候補訳語に対するハイライト表示（色表示等）によって、信頼度の高さを表現する。

条件			第 1 候補の信頼度
訳語候補が一つのみ	第 1 候補頻度 > X	第 1 候補頻度 > Y * 第 2 項頻度	
○	○		高
×	○	○	高
×	○	×	中
×	×	○	低

(画面例では、X=20、Y=3、高：赤、中：青、低：黒)

表 1. 信頼度設定の例

4.3. 画面例

以下では「粉末を混ぜる」という日本語文が入力されたときの画面例を示す。英語に翻訳する場合の信頼度表示の例を示す。なお、本論文ではカラー印刷ができないため、色表示の説明は参考としていただきたい。

図 5 では、上部に原文・訳文が表示され、下部に指定した単語の訳語候補が表示される。

下部では、「粉末」が選択されてこの単語に対する訳語候補が表示されている。ここで、訳語「powder」の絶対頻度は規定値（ここでは 20）を超えており、第 2 候補「dust」との相対頻度も規定値（ここでは 3 倍）を超えていることから、「powder」という訳語の信頼度は「高」となる。このことから、上部の訳文中では、まず最高頻度の候補である「powder」が訳語として使われ、この訳語の信頼度（「高」）が単語の色（赤）で表示されている。

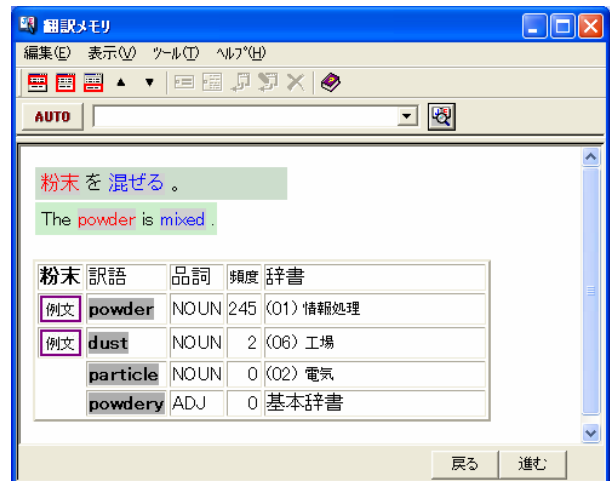


図 5. 「置換」に対する訳語候補と提示情報

図 5 で、「粉末」に対する「powder」という訳語は信頼度が「高」であって、それ以上調べなくてもそのまま使えると考えられる。しかし、「混ぜる」に対する訳語「mix」は信頼度が「中」であり、さらに時間をかけて調査する必要がある。

図 6 は、「混ぜる」に対して訳語「mix」を指定したときの例文検索結果である。ここで、元単語と訳語を両方含む例文が表示されている。単言語のみの検索と比べ、関連する例文が効率よく検索できる。

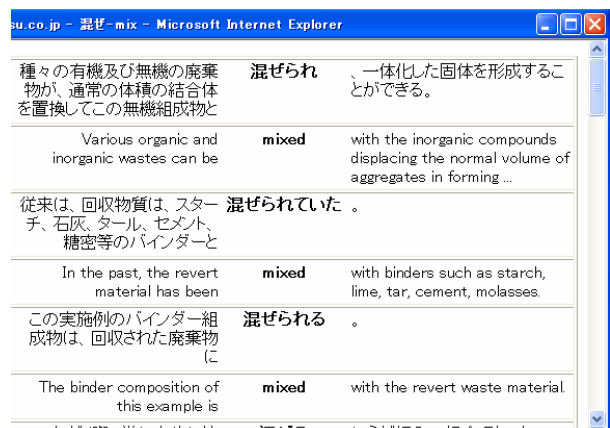


図 6. 単語対指定による訳例表示

5. 実験手順

これまでに開発してきた翻訳支援システム ([1],[2],[3]) に、今回開発した訳語信頼度表示機能を組み込み、実際の被験者を使った実験を行った。

実験環境

産業翻訳に従事する典型的な翻訳者と同等の英語力 (TOEIC 換算で 800 点台後半から 900 点中盤) を持つ 3 名の被験者に実験会場に通ってもらい、翻訳支援システムを使った実験に参加していただいた。3 名のうち 1 名はプロ翻訳者で、残り 2 名は非翻訳者。期間は 2 週間。最初はシステムの習得から始め、慣れてきた時点で、システムあり翻訳とシステムなし翻訳の比較を実施した。システムあり・なしいずれの場合も、通常使っている環境 (辞書やウェブ等) は使ってよいものとした。

対象文書と測定手順

対象作業は、日本語特許抄録の英語への翻訳作業。システムや環境に慣れるための練習用として 12 文書、本番用に別の 12 文書を用意した。12 文書は、人手翻訳の際の標準翻訳時間がほぼ等しくなるようなものをあらかじめ選んでおいた。システムあり・なしそれぞれの条件で、訳語選択作業にかかった時間を測定した。訳語選択作業のみでは現実的な測定にならないため、翻訳作業を一通り行ってもらう中で、翻訳作業全体の時間と訳語選択作業の時間をそれぞれ記録してもらうようにした。また、完成した翻訳文に対する訳質評価を行った。

6. 結果と分析

ここでは、翻訳作業全体のうちの、訳語選択作業のみの測定結果について述べる。実験に参加いただいた 3 名について結果を示す。

表 2. 文書毎平均の訳語選択所要時間

	システムなし	システムあり	比率
翻訳者 A	22.1	18.1	0.82
翻訳者 B	13.2	12.1	0.91
翻訳者 C	25.1	10.9	0.43
平均	20.1	13.7	0.68

別途訳質を評価したところ、全翻訳者とも、システムあり・なしで有意差が見られなかったため、所用時間の差がそのまま効率の差とみなすことができる。翻訳者によってばらつきがあるものの、システムを使うことによって翻訳者全員の作業効率が向上した。

あわせて記述式のアンケート調査を行ったところ、全翻訳者が、システム利用による効率化について実感を持っていることもわかった。

7. 考察

実験から、信頼度表示システムを使うことによる、訳語選択作業の効率化が確認できた。ただし、翻訳者による「比率」のバラつきが出ていることから、実験方法自体に改善の余地があると思われる。バラつきが出た原因としては以下が考えられる。

信頼度自体の信頼性の情報を与えなかった

訳語に対して、システムのほうで「高信頼度」の表示がされていても、翻訳者としてはどこまで信用してよいかかわからない、という問題があった。純粋に信用した人は提示された候補をそのまま使う傾向にあり、そうでない人はさらに確認作業に時間をかけるということになり、測定が安定しなかった。実験後、別途、信頼度自体の評価を行い、業界において適当と判断される訳語に実際に高い信頼度が与えられているかどうかの調査を行った。この結果、本システムで高信頼度を得ている訳語はほとんどの場合にそのまま使ってよいことがわかり、今後の実験では、その部分の確認作業を省略することを手順に含めることによって測定が安定すると考えられる。

訳語選択作業時間の切り分け手順の問題

今回の測定は、実際に翻訳文を最後まで完成させるという作業の中においての、訳語選択時間の測定なので、そういう意味では現実味のある測定となった。しかし、訳語選択作業のみの時間を測るということ自体に不自然さが出てくる場合がある。例えば、訳語だけを最初に選択してしまってその時間を測定しても、後工程の訳文作成中にさらに訳語の調査が必要になるような場合も出てくる。このような切り分けの手順の明確化が不十分であったため、測定の不安定さが生じた可能性もある。

8. まとめと今後

本訳作業における訳語選択を効率化するための枠組みとして訳語信頼度表示機能を提案した。機能を実装し被験者実験を行ったところ、機能の有効性が確認された。今後は、機能の改良と並行して測定手順の改良も進めて、より利用価値の高いシステムを目指す。

参考文献

- [1] 大倉清司、富士秀、長瀬友樹. オートコンプリートによる翻訳支援. 言語処理学会第 13 回年次大会予稿集, 2007.
- [2] 大倉清司、富士秀、徐国偉、長瀬友樹、潮田明. Cliché: さらになる翻訳効率化のための翻訳支援インタフェース. 言語処理学会第 12 回年次大会予稿集, 2006.
- [3] 潮田明、富士秀、大倉清司、山下達雄. 機械翻訳と訳例検索を統合した翻訳支援システム. 言語処理学会第 9 回年次大会予稿集, 2003.