

# クラウドソーシングによる対訳コーパス構築における 音声入力インターフェースの検討

小村 和輝 塩田 嶺明 中澤 敏明 黒橋 禎夫

京都大学

{komura, shioda}@nlp.ist.i.kyoto-u.ac.jp

{t\_nakazawa, kuro}@i.kyoto-u.ac.jp

## 1 序論

近年、インターネットの普及等の理由から、他言語での日常会話に触れる機会は増加しており、一般的な文の機械翻訳に対する需要は高まっている。現在主流であるコーパスベース機械翻訳は、高品質な翻訳を実現するために大規模な対訳コーパスが必須であるが、誰もが自由に使え、且つドメインが限られていない大規模なものは現状では存在しない。例えば、自由に入手できる日本語の意味辞書開発を目的とした「日本語 WordNet<sup>1</sup>」は約 18 万対訳文を収録しているが、この程度の規模では高精度な翻訳を行う事は難しい。またこの対訳文は辞書の定義文や例文で構成されているため、一般的な文の翻訳にはやや不適切である。京都大学格フレームをベースに日本語の基本的な文を自動抽出し、人手で修正を行った「日英中基本文データ<sup>2</sup>」は日常会話に近い形ではあるが、約 5000 対訳文のみの収録にとどまるため有用な規模ではない。

ユーザーが語学学習を目的として取り組んでいる翻訳作業が、クリエイティブ・コモンズの文書翻訳作業になるという「Duolingo<sup>3</sup>」や、興味のある Web コンテンツを投稿し、様々な人に無償で翻訳してもらうという「みんなの翻訳<sup>4</sup>」など、クラウドソーシングの対訳コーパス構築サイトは既にいくつか存在するが、どれもコーパスデータは非公開である。このような状況から、我々は新たにクラウドソーシングによる対訳コーパス構築フレームワークを考案し、対訳文の収集、精練、公開を行うことで、広く機械翻訳技術の発展に貢献することを目指す。

クラウドソーシングによる対訳コーパス構築に関する研究としては、Mechanical Turk を利用したものが

提案されている [1, 2]。先行研究では、まず様々な言語と英語との一文翻訳当たりのコストについて調査し、それに付随してクラウドへのタスクの示し方を検討し、最終的にはその翻訳の正確さを測定することで、クラウドソーシングでも十分な正確性をもつ対訳文の収集を行うことが可能であると結論付けている。しかし Mechanical Turk を利用しているため、作業への賃金の支払いが必要である。我々の研究では、金銭的なインセンティブによる構築ではなく、人々の自主的な貢献を促進するような枠組みを目指しており、この点で先行研究とは異なる。

クラウドソーシングの利用にあたっては、作業者の負担を減らすことが重要な問題であり、我々は効率的な入力インターフェースを用意することが負担軽減につながると考えた。そこで本研究では、比較的新しい技術である音声入力を入力インターフェースとして用いることを検討する。具体的には、被験者による「音声入力とタッピング入力による訂正」「タッピング入力のみ」「ハードウェアキーボードによる入力」の 3 種類の入力インターフェースを用いた英日翻訳実験を行い、入力インターフェースの違いによる作業者の負担や作業効率の違いを明らかにする。また、被験者へのアンケートやインタビューを通じて得た、入力インターフェースの改善に関する知見を報告する。

## 2 入力文選択

クラウドソーシングによる対訳コーパス構築の際には、少数の文で効率よく翻訳精度の向上が見込めるよう、作業者に翻訳してもらう文の選択も考慮すべき問題の一つである。これには能動学習が主に用いられている。Eck ら [4] は、翻訳済みコーパスに出現しない単語 n-gram の数を、次に翻訳すべき文のスコア付けに用いている。Ambati ら [3] はこれに加え、未翻訳

<sup>1</sup><http://nlpwww.nict.go.jp/wn-ja/>

<sup>2</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?日英中基本文データ>

<sup>3</sup><http://duolingo.com/>

<sup>4</sup><http://trans-aid.jp/>

順位	単語	出現回数
1	the	703,922,784
2	be	566,553,782
3	and	362,379,193
4	to	349,559,172
...	...	...
1997	palm	557,250
1998	ft	556,823
1999	adjust	556,446
2000	severe	556,328

表 1: Web での頻度上位 2000 語の例

コーパスでの n-gram の出現頻度も考慮したスコア指標を提案している。Haffari ら [5] は単語 n-gram だけでなく、文を翻訳した精度のスコアや、目的言語から原言語に再翻訳した文と原文との類似度など、様々な要素を用いて能動学習を行っている。

我々もこれら先行研究に習って文選択について考慮するべきであるが、今回は入力インターフェースの違いによる作業への負担を明らかにすることが目的であるため、実験で用いる入力文は以下に述べる簡単な方法で、半自動的に選択した。まず英語 Web コーパス 15 億文より、各単語の原形の出現回数を計数し、その上位 2000 語を抽出した。表 1 は上位 2000 語の例である。参考までに、上位 2000 語に入らなかった単語の例としては、“insight”, “intentional”, “schizophrenic”, “impressive”, “dozen”, “hook”, “biotechnology”, “currency” などがある。次に Web コーパスより、上位 2000 語のみで構成される文を抽出し、さらに 15 単語以下の短い文のみに絞り、その中から文構造の比較的平易なものを人手で選び、本インターフェースで用いる原文データとした。以下に原文の例を示す。

- They were the toughest team we have played so far.
- A new kind of war requires a new kind of peace.
- Let's look at the history of this.
- This drink is the only way to save yourself from hit of summer sun.
- Some large houses were built on these lots in the early 1900's.

### 3 インターフェース概要

今回の実験では、

1. 音声入力とタッピング入力による訂正
2. タッピング入力のみ
3. ハードウェアキーボードによる入力



図 1: インターフェース メイン画面 (左) と音声入力時の画面 (右)

の 3 種類の入力方法を用意している。音声入力は、Google 社が提供している音声入力 API を用いている。これを Android アプリケーション上に組み込み、Android 端末に実装することで、入力インターフェースを開発した。音声入力によって、まず大まかな訳文を素早く入力し、その後タッピング入力により部分修正を加えるという入力方法を想定している。「タッピング入力」とは Android 端末上のソフトウェアキーボードに触れて入力を行う方法である。「ハードウェアキーボードによる入力」は一般的なパソコンを用いて、Android アプリケーションと同様の機能を実装した CGI 上で入力を行う方法である。

補助機能として、Google 翻訳 API を用いて原文の機械翻訳の結果を同時に表示し、翻訳者の語彙力などをサポートするよう工夫した。また、訳文の質を保つため、翻訳が困難な文を飛ばす機能も実装した。実験で使用した入力インターフェースのスクリーンショットを図 1 に示す。

## 4 実験・考察

音声入力による訳文作成におけるヒューマンファクタへの影響を明らかにし、入力インターフェースの検討とさらなる改善に役立てるため、翻訳実験を行った。

### 4.1 実験設定

原文データは先の章で述べたものから 20 文を 3 セット選び、計 60 文を用いた。これを被験者 10 名に翻訳してもらったが、その際 20 文セット毎に 3 種類の入力インターフェースをそれぞれ使用してもらい、比較

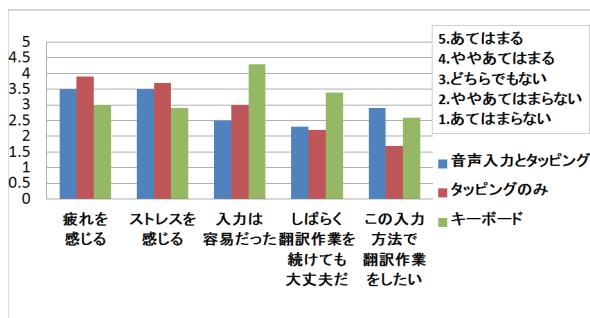


図 2: 入力インターフェースごとのアンケート結果

入力インターフェース	人数
音声入力とタッピング	6 人
ハードウェアキーボード	3 人
タッピングのみ	1 人

表 2: 選好アンケート (どのインターフェースを一番使いたいと感じたか) の結果

実験を行った。各場合において、それぞれ「翻訳時間」「訳文」「アンケート」に関するデータを測定、および収集した。翻訳時間はアプリケーション上で測定し、訳文データと共にサーバーへ送信した。また、「音声入力とタッピングによる訂正」の場合のみ、音声入力ミスの「部分訂正回数」と「全文訂正回数」を、翻訳者にアプリケーション上のカウンターボタンをタップしてもらう方法で測定した (図 1 参照)。

## 4.2 結果・考察

図 2 に各インターフェースに関するアンケートの結果を、表 2 に「どのインターフェースを一番使いたいと感じたか」という選好アンケートの結果を示す。このうち「疲れを感じる」「ストレスを感じる」「入力は容易だった」「しばらく翻訳作業を続けても大丈夫だ」の四つの項目の結果から、被験者らは音声入力を難度の高いものと感じ、それが心身への負担につながっており、長時間の作業は困難であると考えていることが分かる。これに反して「この方法で翻訳作業をしたい」という項目での音声入力の点数は高く、選好アンケートの結果を見ても、音声入力を使いたいと考える被験者は多いことが分かる。これは表 3 の記述式アンケートの質問と回答から、音声が入力として認識された際の楽しさなど感覚的理由に起因していると推察できる。

図 3 に翻訳速度測定の結果を、図 4 に音声入力時の訂正回数を、選好アンケートにおける各インターフェースの選択者ごとに示す。これによれば、「音声入

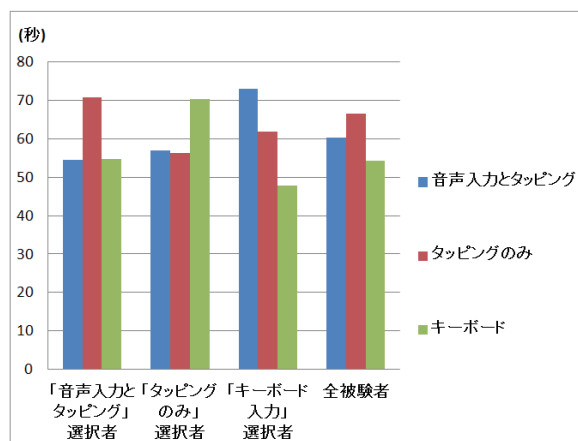


図 3: 各選好ごとの一文当たりの翻訳速度

力とタッピング入力による訂正」を好む被験者らは、音声入力による翻訳速度が他の被験者や他のインターフェースと比較して速く、また音声入力の訂正回数も少ない。したがって、音声入力を好むかどうかは翻訳速度と関係しており、かつ訂正回数に大きく影響を受けていることが分かる。また、訂正回数が翻訳者の心身への負担に深く関係していることは、表 3 において、音声入力の訂正時に時間がかかることを全被験者が指摘しており、他の項目に比べ不満も多いことなどから分かる。実際に、全文訂正・部分訂正ともに平均約 10 秒を要するとした場合、全被験者平均で一文あたり約 13 秒のロスが発生していることとなる。「音声入力とタッピング」および「キーボード」の全被験者平均の翻訳速度の差が約 6 秒であることから、インターフェース開発において訂正方法を重点的に改善すれば、翻訳速度の優劣関係が逆転し、さらには心身への負担も軽減されることで、音声入力を使いたいと考える翻訳者は大きく増加すると予測される。

また、図 3 より、キーボード入力を好む被験者は、比較的キーボード入力の熟練度が高いことが分かる。したがって、キーボード入力に慣れていない翻訳者は、より音声入力を好むと考えられる。また、先に述べたように、感覚的理由により音声入力を好む翻訳者もいることから、こういった音声入力特有の娯楽性を利用した入力インターフェースを考案することも必要であると考えられる。

## 5 結論

本研究では、クラウドソーシングによる対訳コーパス構築の実現に向けた入力インターフェースについて検討した。実験により、一般の人々に幅広く、効率的

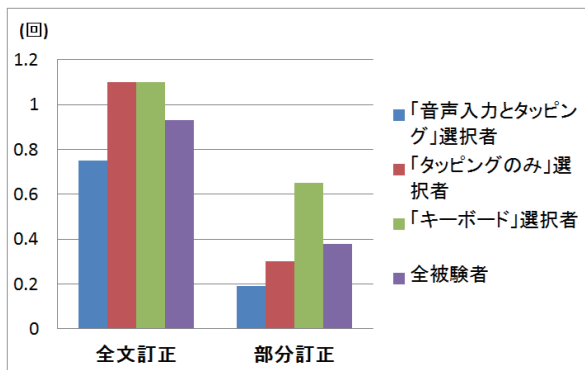


図 4: 音声入力時の訂正回数

に翻訳をしてもらう上では、音声入力が適していることが示され、目標達成に前向きな結果となった。今後は、先の章で述べたような改善点や表 3 の他の意見を踏まえたさらなるインターフェースの改良や、訳文の質の確保、文選択における能動学習の応用などの検討課題を解決していき、クラウドソーシングによる対訳コーパス構築の実現を目指す。

## 参考文献

- [1] Vamshi Ambati and Stephan Vogel. Can crowds build parallel corpora for machine translation systems? In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 62–65, Los Angeles, June 2010. Association for Computational Linguistics.
- [2] Vamshi Ambati, Stephan Vogel, and Jaime Carbonell. Active learning and crowd-sourcing for machine translation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may 2010.
- [3] Vamshi Ambati, Stephan Vogel, and Jaime Carbonell. Multi-strategy approaches to active learning for statistical machine translation. In *Proceedings of MT-Summit XIII*, 2011.
- [4] Matthias Eck, Stephan Vogel, and Alex Waibel. Low cost portability for statistical machine translation based on n-gram coverage. In *Proceedings of MT-Summit X*, 2005.
- [5] Gholamreza Haffari, Maxim Roy, and Anoop Sarkar. Active learning for statistical phrase-based machine translation. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 415–423, Boulder, Colorado, June 2009. Association for Computational Linguistics.

1. どのような場合に時間がかかりましたか
  - 音声入力の訂正時 ... 10 人
    - 音声認識の大部分が間違っており、音声入力し直す時
    - 漢字の訂正が多かった
    - 余計な言葉を認識してしまうため、その削除に時間がかかった
    - 二か所以上あると面倒に感じる
    - 記号の場所が分かりづらい
  - 音声入力の前にある程度の長さの日本語を頭の中で作るとき ... 3 人
  - タッピングミス時 ... 2 人
  - 上手い表現がでてこないとき
  - 単語が分からない時
  - 漢字変換時
  - 疲れてきたとき
  - 数字入力時
  - タイピングミス時 (キーボード入力)
  - タイピングが苦手なため、長い文の時 (キーボード入力)
2. 入力方法に関して、改善点等をご自由にお答えください
  - 間違えて「スキップ」や「送信」を押してしまう ... 4 人
  - もう少し画面を大きくしてほしい ... 3 人
  - タッチミスをもう少し防いでほしい ... 3 人
  - 音声認識が遅い ... 3 人
  - フリック入力の方が良いのでは ... 2 人
  - 音声認識のコツをつかむのに、与えられた時間以上に練習が必要 ... 2 人
  - 機械翻訳を修正するという形式が良いのでは
  - 音声入力の精度を上げてほしい
  - 訂正時、再び音声で入力するかタッピングで訂正するか判断しがたい場合がある
  - キーボード入力は単調な作業のため苦痛
3. 翻訳をスムーズに進めるため、改善点等をご自由にお答えください
  - 音声入力後に選択肢が出てほしい ... 4 人
  - 文節ごとに分けて入力するとうまくいった
4. 他に何か御意見があれば、ご自由にお答えください
  - 音声を入力する際、恥ずかしさがある
  - 短い文は認識しやすい
  - 音声認識は速いが、誤認識が多いため疲れる
  - より良い收音デバイスがあればもう少し正確になるかもしれない
  - 音声入力だと、視覚をあまり使わなくてよく、老人等には楽かもしれない
  - 音声がかきちんと認識されたときに嬉しくて続ける気がおきる
5. 表示される英文に関して改善点等をご自由にお答えください
  - ちょうどよかった ... 4 人
  - 長すぎて言いづらいものもあった (音声入力時) ... 2 人
  - 訳し易いものを最初の方にすると良い
  - 短い文章を最初の方にすると良い
  - 英文に SVOC などの構造的情報があると良い
  - 抽象的な文章の方が認識がうまくいくと感じた
  - 文脈に意味が左右されるものは悩んでしまい、時間がかかる
  - 文構造がもう少し難しくても大丈夫な気がした
6. 「参考訳」などの補助機能に関して、改善点等をご自由にお答えください
  - 必要ない・無い方が速くなる ... 4 人
  - 単語単体で辞書的な意味を出してほしい ... 4 人
  - 精度を上げてほしい ... 3 人
  - 参考訳が正しい時には、それを見ながら言えて楽だった
  - 英文と比較できる形で縦に表示されているのが良かった
  - 語彙を知りたい単語が訳されていない時があった
  - 英語をそのままカタカナで表示するのはやめてほしい
  - 分からない単語の意味を知ることのみ役立った

表 3: 記述式アンケートの質問と回答