

多言語用例対訳共有システムのための用例間評価手法の評価

福島 拓†

吉野 孝†

重野 亜久里‡

† 和歌山大学

‡ 多文化共生センターきょうと

1 はじめに

現在、在日外国人や訪日外国人数は増加傾向にある [1]。しかし、母語以外の言語によるコミュニケーションは困難であるため [2]、日本人と外国人とが十分なコミュニケーションを行うことは難しいと考えられる。

正確なコミュニケーションが求められる分野の 1 つに医療がある。医療分野では、正確性の確保された用例対訳を用いた支援が行われている [3]。用例対訳とは用例を多言語に翻訳した多言語コーパスのことを指す。用例対訳は様々な団体が構築を行っているが、正確性の確保が十分に行われているとは言えない [4]。

そこで我々は Web 上で正確な多言語用例対訳の収集、共有、提供を目的とする多言語用例対訳共有システム TackPad (タックパッド) の開発を行い、評価機能によって用例対訳の正確性の確保を目指している [5]。しかし、用例対訳は量が多いため、評価者の負担が大きい。このため、評価対象となる用例対訳を絞って提示する必要があると考えられるが、不正確な用例対訳の抽出に関する研究は行われていない。

本稿では、用例対訳の評価を円滑に行うために、形態素解析器と Web 検索を用いて、不正確な用例対訳の抽出を行う用例間評価手法を提案し、その評価を行う。

2 システム設計

2.1 システム概要

本システムの画面例を図 1 に示す。収集言語は、日、英、中、韓、葡、西、越、泰、尼の 9 言語である。

本システムでは、医療従事者や患者などが他の言語に翻訳してほしい用例を提案し、その提案用例を翻訳者が翻訳している。実際に用例対訳を使用する利用者が用例を提案することで、医療現場で求められている用例対訳の収集を可能としている。

2.2 用例間評価手法

本システムでは用例対訳の正確性を確保するために、複数の評価軸を持った 5 段階評価を用いて、利用者相互で用例対訳の評価を行っている (図 1 右下)。しかし、評価対象用例対訳が多く存在するために、評価者の負担が大きくなっている。

そこで、本稿では形態素解析器と Web 検索を用いた用例間評価手法を提案する。本手法では、形態素解析器は言語グリッド¹⁾を、Web 検索は Google 検索²⁾をそれぞれ利用した。

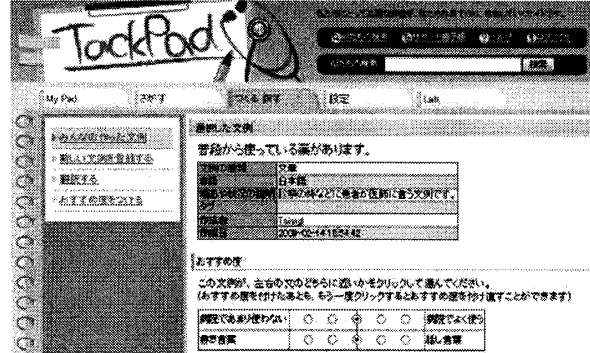


図 1: TackPad の画面例

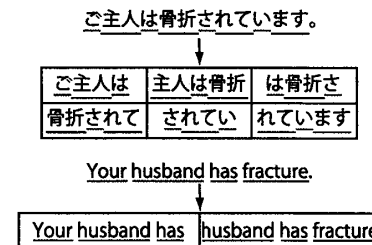


図 2: 形態素ごとの N-gram 分割用例の例

・用例対訳の 3-gram 分割例である。図中の下線は形態素を示す。

本手法は下記の順で不正確な用例対訳候補を抽出する。

1. 本システムで収集している用例を、形態素解析器で形態素ごとの N-gram データに分割する。その際、カンマやコロン、括弧でくくられた文字列を除去している。なお、N-gram に分割できない用例対訳は分析対象外としている。本稿では、分割したデータを「N-gram 分割用例」とする。図 2 に N-gram 分割用例の例を示す。
2. 用例対訳の N-gram 分割用例を組み合わせて Web 検索で検索し、検索ヒット件数を記録する。その際、N-gram 分割用例をダブルクォーテーションでくくり、完全一致で検索している。また、すべての N-gram 分割用例の組み合わせを検索している。N-gram 分割用例数が 2 個と 3 個の用例対訳の場合、組み合わせを変えて計 6 回検索を行う。表 1 に Web 検索の検索文字列の一例を示す。
3. 検索ヒット件数が 1 件以上の割合を用例対訳ごとに調べ、その割合を用例対訳の「Web ヒット割合」として記録する。
4. Web ヒット割合が 0% の用例対訳は Web 上に存在しないため、不正確な用例対訳候補とする。

Evaluation of Parallel-text Relation Evaluation Method of Multilingual Parallel-text Sharing System TackPad

†Taku Fukushima †Takashi Yoshino ‡Aguri Shigeno

†Wakayama University ‡Center for Multicultural Society Kyoto

¹⁾ <http://langrid.nict.go.jp/>

²⁾ <http://www.google.co.jp/>

表 1: 検索ワード例

| 検索ワード | ヒット数 |
|-------------------------------|------|
| “ご主人は” “Your husband has” | 5 件 |
| “ご主人は” “husband has fracture” | 0 件 |
| “主人は骨折” “Your husband has” | 0 件 |

- ・「ご主人は骨折されています。」と「Your husband has fracture.」の 3-gram 分割用例を用いた検索ワードの一例である。
- ・この用例対訳には、合計 12 対 (日本語 6 組×英語 2 組) の検索ワードがある。

表 2: 評価基準

| 値 | 評価基準 | 意味 |
|---|--------|----------------------------|
| 5 | All | 同じ意味 |
| 4 | Most | 文法などに多少問題があるが、 だいたい同じ意味 |
| 3 | Much | 意味はだいたいつかめる |
| 2 | Little | 雰囲気は残っているが、 もとの意味は分からない |
| 1 | None | 全く違う意味 |

3 実験

用例間評価手法の有用性確認のための実験を行った。本実験の目的は、2.2 節の手法を使用して不正確な用例対訳の抽出が可能かどうかの確認である。

3.1 用例対訳の評価

実験に使用する用例対訳を本システムから抽出した。その際、日英、日中、日韓、日葡、日西の用例対訳を各 200 対抽出して使用している。なお、日英の 1 対の用例対訳に問題があったため、実験結果では 999 対を使用している。また、抽出には不正確な用例対訳が含まれている可能性が高い、一般ユーザが作成した用例対訳を中心に抽出している。このため、不正確な用例対訳の割合が実際よりも大きくなっている可能性がある。

抽出した用例対訳の正確性を判別するために、通訳者 15 名 (各言語 3 名) に正確性の評価を依頼した。評価基準を表 2 に示す。

用例対訳の正確性評価の結果を表 3 に示す。本実験では、評価者による評価の平均値を用いて手法の有用性確認を行う。また、評価者の平均が 4 以下の用例対訳を不正確な用例対訳としている。表 3 より、不正確な用例対訳は 73 対存在したことが分かる。

3.2 実験結果と考察

3.1 節の調査結果を用いて、本手法の不正確な用例対訳の抽出性能を評価した。なお、今回は 2-gram, 3-gram, 4-gram のデータを分析に利用している。

本手法の適用前と適用後の用例対訳総数と不正確な用例対訳数を表 4 に示す。表 4 から、4-gram で不正確な用例対訳候補の抽出を行った場合、再現率が 98.11% であり、適合率が 0.77 ポイントあがったことが分かる。

不正確な用例対訳の抽出の際には、不正確な用例対訳を極力除外しない手法が必要なため、本手法は一定の効果があったと考えられる。

なお、4-gram に満たない用例対訳 (373 対) に含まれていた不正確な用例対訳 (20 対) には、用例そのものは正

表 3: 用例対訳の評価結果

| | 評価値の平均の区間 | 用例対訳数 | |
|-----|-------------------|-------|----|
| 不正確 | $1 \leq n < 2$ | 3 | 73 |
| | $2 \leq n < 3$ | 8 | |
| | $3 \leq n \leq 4$ | 62 | |
| 正確 | $4 < n \leq 5$ | 926 | |

表 4: 提案手法の適用結果

| | | 2-gram | 3-gram | 4-gram |
|-----|-------------------|--------------------------|---------------------------|---------------------------|
| 適用前 | 総数 | 830 | 710 | 616 |
| | 不正確 | 59 | 56 | 53 |
| | 適合率 F 値 | 7.11% 13.27% | 7.89% 14.62% | 8.60% 15.84% |
| 適用後 | 総数 | 111 | 485 | 555 |
| | 不正確 | 6 | 37 | 52 |
| | 適合率 再現率 F 値 | 5.41% 10.17% 7.06% | 7.63% 66.07% 13.68% | 9.37% 98.11% 17.11% |

- ・表中の総数は評価対象用例対訳数を、不正確は評価対象用例対訳のうち不正確な用例対訳の数を示す。
- ・適用前の F 値は、適用前の再現率を 100% として計算している。
- ・適用後の総数は、抽出した不適切な用例対訳候補数を示す。

確だが、用例対訳としては不正確なものが多く含まれていた。例として日本語で「尿管結石」と記述された用例が、中国語で「腎臓結石」と記述されていたものや、日本語で「スペイン語」と記述された用例が、ポルトガル語で「ポルトガル語」と記述されていたものがあった。このような用例対訳は本手法では検出できないため、別手法の適用が必要であると考えられる。

4 おわりに

本稿では、不正確な用例対訳の抽出を目的とした用例対訳抽出手法を提案し、その評価を行った。実験結果より、4-gram 分割用例を使用することで、不適切な用例対訳の抽出に一定の効果が見られた。今後はより精度の高い不正確な用例対訳の抽出手法の確立を目指す。

謝辞

本研究の一部は総務省の戦略的情報通信研究開発推進制度 (SCOPE) の平成 20 年度採択課題「多言語共生社会における医療対話支援のための多言語対話用例プラットフォームの構築」による。

参考文献

- [1] 法務省: <http://www.moj.go.jp/PRESS/>
- [2] Takano, Y., et al: A temporary decline of thinking ability during foreign language processing, *Journal of Cross-Cultural Psychology*, 24, pp.445-462(1993).
- [3] 宮部真衣ほか: 外国人患者のための用例対訳を用いた多言語医療受付支援システムの構築, *信学論*, Vol.J92-D, No.6, pp.708-718(2009).
- [4] 福島拓ほか: 多言語用例対訳を用いたコミュニケーションのための応答用例対作成システムの開発, *情処, DI-COMO2009 シンポジウム*, pp.1612-1618(2009).
- [5] 福島拓ほか: 多言語用例対訳共有システム TackPad における用例評価手法の比較, *情処, グループウェアとネットワークサービスワークショップ 2009*, pp.87-92(2009).