

SignWriting を用いたアメリカ手話-日本手話対訳辞書の検討

杉山 真也[†] 松本 忠博[†] 加藤 三保子[‡]

[†] 岐阜大学大学院工学研究科

[‡] 豊橋技術科学大学総合教育院

1 はじめに

アメリカ手話 (American Sign Language, ASL) は主に北米英語圏で使用されている手話であるが、母語を共有しない話者間の会話で共通語としても利用され、国際手話 (International Signs) にも大きな影響を与えているという。外国語の学習や機械翻訳には辞書が欠かせないが、手話には標準的な文字がないことから、市販の手話辞典の中で、手話からそれに相当する日本語 (英語) を引くものはまれであり¹、異なる手話言語間の対訳辞書は見られない。本研究では、手話の文字として提案されている SignWriting[1] (以下, SW) と既存の辞書を利用して、ASL の単語をそれに相当する日本手話 (JSL) の単語に変換する ASL-JSL 辞書について検討し、試作を行った。

SignPuddle Online[2] には、SW コミュニティによって構築された ASL 辞書 (単語の SW 表現とそれに対応する英語の語句のデータ) と ASL 文書が公開されている (図 1)。本研究における辞書では、検索対象の単語をこの ASL 辞書から高瀬ら [3] の SW 単語間類似度算出アルゴリズムを用いて見つけ出し、その英訳を取得する。次に、英和辞書から得られた日本語を介して、対応する日本手話の単語を求める。日本手話の辞書掲載語数は数千語と、日本語に比べてかなり少ないため、日本語の語句に対応する手話が辞書に掲載されていないことも多い。そこで単語が手話辞典に掲載されていない場合は、シソーラスを併用して類義語も検索するようにした。

2 SignWriting による単語表現

ここでは ASL 辞書および検索対象の単語の表記に用いる SW について述べる。言語を活用する上で文字は重要な役割をもつが、手話には文字がないため、これまでにいくつかの表記法が提案されている。SW の特徴は、日常生活における利用を想定している点にある。手話単語を構成する手の形、動き、位置、お

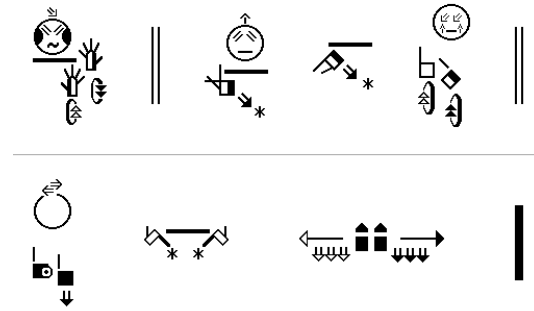


図 1: SignWriting による ASL 文の例 (SignPuddle², “Sleeping Beauty” より)

よび、顔の表情などの非手指要素を表す図像的な記号 (International SignWriting Alphabet, ISWA) を 2 次元的に配置して単語を表現することで、人にとって比較的分かりやすい表現になっている。

内部的には、SW における手話単語は、単語を構成する ISWA 記号の集合 $\{s_1, s_2, \dots, s_n\}$ であり、各記号 s_i は、記号コードと単語内での位置の情報を持つ。記号コードは、図 2 に示すような記号のカテゴリ (手の形、動き、頭・顔、体など)、カテゴリ内グループ、基本記号、手のひらの向き・視点、指先の向きなどを表す 6 つの値からなる。

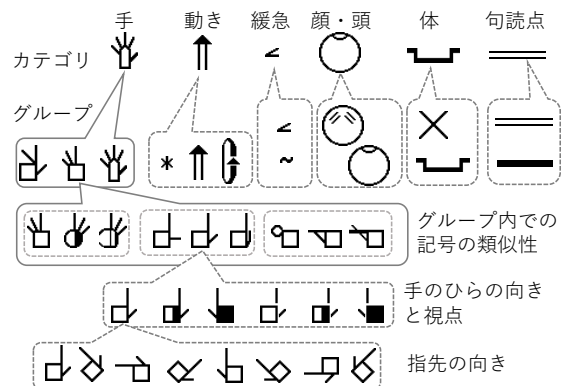


図 2: SW の表記に用いる記号 (文字) の種類

¹手の形から単語を検索できるものが市販されている。

3 ASL-JSL 辞書機能

ASL 単語の SW 表記からそれに相当する JSL 単語を求める処理の流れを図 3 に示す。処理は、SW 形式の ASL 単語の同定と英訳の取得 (図中①)、英語の語句から日本語ラベルで表された JSL 単語の取得 (②、③)、日本語ラベル形式から SW 形式の JSL への変換 (④) の各ステップに分けられる。ここでは、①から③について述べる。④については手話文書エディタ JSPad[4] 用に構築中の SW 辞書を使用する。

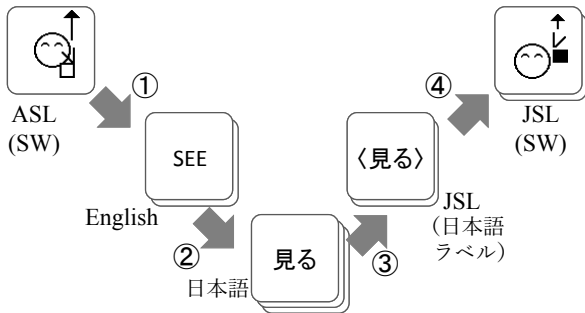


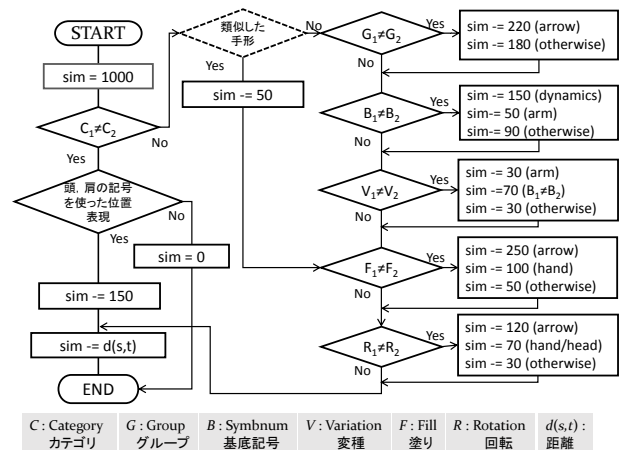
図 3: ASL-JSL 変換処理の流れ

3.1 ASL 単語の同定による英訳の取得

SW による手話単語表記は視覚的に分かりやすい反面、自由度が高く、書き手による表記のゆれが大きいため、単語の同定が機械的にに行いにくい。そこで、文献 [3] の JSL の単語間類似度算出アルゴリズムを用いて、目的の単語を ASL 辞書から検索する。ASL 辞書は SignPuddle から取得した XML 形式のものを SQLite データベースに変換して用いる。ただし、文献 [3] で使用した JSL 辞書の登録語数が約 680 (現在は約 1,130) であったのに対し、SignPuddle から取得した ASL 辞書は 9,600 以上と大幅に増えたことから、単語を構成する記号の数が一定以上になると実用的な時間で検索できないという処理時間の問題が表面化したため、高速化を含めたアルゴリズムの改良を行った。

単語間類似度の算出にはまず、2 つの単語を構成する各記号間の類似度を求める。記号の類似度は、記号コードから得られる記号のカテゴリ、グループ、基本記号番号などの値と、各単語内での記号の位置のずれを元に算出する (図 4)。これにより、図 5 のような記号間類似度の表を作成する。なお、同じグループ番号を持つ記号でも類似していないものも存在するため、今回は個別の記号の類似性も考慮するようにした。

単語間の類似度は、単語を構成する記号の数の差が閾値 k より大きいなら 0 であり、そうでなければ、2



C: Category カテゴリ G: Group グループ B: Symbnum 基底記号 V: Variation 変種 F: Fill 塗り R: Rotation 回転 d(s,t): 距離

図 4: 記号間類似度の算出

	☞	☞	☞☞	☞☞	☞
☞	972	876	0	0	0
☞	884	970	0	0	0
☞☞	0	0	880	750	0
☞☞	0	0	630	1000	0
☞	0	0	0	0	934

図 5: 2 つの単語を構成する記号間の類似度の表 (類似度の最大値は 1,000)

2 つの単語を構成する各記号の対応関係のうち、最も記号間類似度の平均値が高くなるような組み合わせにおける記号間類似度の平均値と定義している。この組み合わせの探索に時間がかかっていたため、記号間類似度が 0 のものがある (高い類似度が得られる可能性が低い) 組み合わせであることが判った時点で、その組み合わせの計算を取り止めるようアルゴリズムを変更した。この改良により、例えば改良前に 5 分 20 秒かかった手話単語 TEND (記号数 7) の検索時間は、0.09 秒に短縮できた。

3.2 英語の語句から JSL 単語名の取得

(英語から日本語へ) ASL 辞書から得られた手話単語の英訳 (英語の語句) から、JSL 単語を表す日本語の語句 (日本語ラベル) を取得するため、まず、英和辞書を参照して英語の語句の日本語訳を取得する。

(日本語から JSL 単語名へ) 日本語に対応する手話単語を検索するため、『わたしたちの手話 学習辞典』(全日本ろうあ連盟, 全国手話検定試験出題単語), 『日本語-手話辞典』(全日本ろうあ連盟) など複数の手話辞典を参考にして, 手話単語名(単語の日本語ラベル)とその日本語訳の対応テーブル(SQLite データベース)を作成した。登録手話単語の日本語ラベルによる異なり数は 8,690 語, 訳語の異なり数は 8,882 語である。ただし, 同じ手話単語が辞典によって異なる日本語ラベルで記載されていることも多いが, それによる単語の重複については考慮してないため, 実際の単語数はこれより少ない。

(シソーラスの利用) JSL の語彙は(辞書に掲載されたものを比較すると)日本語の語彙よりかなり少ないため, ASL 単語の日本語訳が上記のテーブル上に存在しない可能性がある。その場合は, 日本語大シソーラスを参照して日本語訳の類義語を取得し, 再度 JSL 単語を検索する。

4 ASL-JSL 辞書の試作と評価

前節で述べた方法により ASL-JSL 対訳単語辞書を試作した。ただし, 現状では SW 形式の JSL 辞書の登録単語数が約 1,130 語程度とまだ少ないため, Sign-Puddle の ASL 辞書に記載された単語の英訳から, JSL の単語(日本語ラベル)を取得するステップのみを実施した。

当該 ASL 辞書に登録された単語数は, 原稿執筆時点で 9,666, 単語名の異なり数は 5,839 (同名の単語が複数登録されている), 単語の英訳語句の異なり数は 7,447 である。この 7,447 の英語の語句に対応する JSL の単語名を検索し, ASL 単語とその英訳語句, JSL 単語名の対応テーブルを作成した。

ASL 単語の英訳語句から JSL 単語に変換できた数と割合を表 1 に示す。ASL 単語のうち, 何らかの訳が得られたものは全体の約 55%程度であった。

「ASL 単語」列は ASL 単語ごとの統計, 「英訳語句」列は ASL 単語の英訳語句ごとの統計を表す(1つの単語に複数の英訳が付与されている場合がある)。

「JSL 訳あり」の行は, 当該英語語句の和訳が JSL 辞書の日本語訳の中に 1つ以上見つかったものの数, 「JSL 訳あり(類)」は, 和訳は JSL の日本語訳と一致するものがなかったが, 和訳の類義語とは一致した数である。いずれも, 訳語の妥当性については考慮していない。「JSL 訳なし」は JSL 訳が得られなかった

表 1: ASL から JSL に変換できた語句の数と割合

	ASL 単語	英訳語句
JSL 訳あり	2,479 (42.5 %)	2,942 (39.5 %)
JSL 訳あり(類)	769 (13.2 %)	950 (12.8 %)
JSL 訳なし	663 (11.4 %)	897 (12.0 %)
和訳なし	1,928 (33.0 %)	2,658 (35.7 %)
合計(異なり数)	5,839 (100 %)	7,447 (100 %)

もの, 「和訳なし」は英語の和訳が得られなかったものを表す。

変換結果の例を表 2 に示す。英語の和訳が得られなかったものには人名などの固有名詞のほか, 句や文が多く含まれていた。手話では動詞に道具や対象, 動作主体などの情報が埋め込まれるため, 手話単語が音声言語では複数の単語で表されることになり, 辞書の見出しになりにくいためと考えられる。なお, 和訳を得るために参照した英和辞書の見出し語句の数(異なり数)は 973,710 であった。

和訳が得られて JSL 訳が得られなかったものには, 外国の地名や, 日本で見かけない動物(シマウマ, 狼)などが含まれていたほか, sneeze など比較的簡単な単語が JSL 辞書に存在しない例が見られた。

シソーラスを介して JSL 訳が得られたものは, 妥当と思われるものも含まれるものの, 訳語としてふさわしくないものが多く混入した。類語を利用する場合は, 人手による取捨選択が不可欠と考えられる。

表 2: ASL の英訳から JSL への変換結果の例

JSL 訳あり	name → 「名前」, noon → 「正午」, where → 「どこ」, store → 「店, 商業, 商店, 保存, 覚える」, ring → 「指輪, ボクシング, 組織, 取り囲む, 飾る」,
JSL 訳あり(類)	highway → 「高速道路, インターチェンジ, ドライブ, レストラン」, many → 「多い, らしい, 十分」, geography → 「地図, 工事, 作る」, zebra → 「島, 線, 採点」
JSL 訳なし	Arizona, Asian, otherwise, your, cannot, “one more”, only, themselves, June, “back and forth”, sneeze, mammal, wolf
和訳なし	1/2, “10th grade”, “Valerie Sutton”, “We two”, “What’s wrong?”, abdo-man, “address (location)”, “bring (right)”, call-to-me

5 むすび

手話の書記体系 SW と既存の辞書を利用して、ASL の単語をそれに相当する JSL の単語に変換する ASL-JSL 辞書について検討し、試作を行った。SW で書かれた ASL 単語の同定は、アルゴリズムの高速化によりごく短時間で実行可能になった。精度に関してはまだ十分検証できていないため、今後検証し、精度の向上を図りたい。ASL の英語訳から JSL 単語への変換では、約 55% の ASL 単語から何らかの JSL 単語が取得できた。手話と音声言語の構造の違いもあり、実用的な辞書を構築するには人手による修正・補完が必要である。

参考文献

- [1] Sutton, V. (2002). *Lessons in SignWriting, Textbook and Workbook*, 3rd ed., The deaf action committee for SignWriting.
- [2] Slevinski, S., Sutton, V., *SignPuddle Online v2.0*, <http://www.signbank.org/signpuddle/>.
- [3] 高瀬友宏, 松本忠博, 加藤三保子 (2012). *SignWriting* を利用した手話文書エディタ JSPad における手話-日本語辞書について, 言語処理学会第 18 回年次大会発表論文集, pp.1220–1223.
- [4] Matsumoto, T., Kato, M. and Ikeda, T. (2009). “JSPad—a sign language writing tool using SignWriting,” *Proc. 3rd Int’l. Universal Comm. Sym.*, pp.363–367. (<http://www.mat.info.gifu-u.ac.jp/jspad/>)