

日本語・手話電子化辞書の構成法について

安達 久博, 下山 豪彦, 播磨 良洋, 松浦 威日

宇都宮大学 工学部 情報工学科

〒321 栃木県宇都宮市石井町2753番地

TEL:(0286)61-3401(702) FAX:(0286)89-0981

E-mail: adachi@guru.infor.utsunomiya-u.ac.jp

あらし

対訳辞書を構築する場合には、源言語側と目標言語側との間で単語の登録語数と対応する概念の差が最大の問題となる。本研究では、手話単語の造語成分に着目し、手話単語を基礎語と合成語に大別した。合成語は限られた基礎語を概念規定要素として構成されている。また、合成語の多くは、その最後の要素となる基礎語が単語全体の概念を規定する中心語となっている。これをキーにして手話単語を分類し、日本語のシソーラスとのマッピング操作により未定義の手話単語に対して類義概念の手話単語を提示し、登録語数のギャップを埋める方法を述べる。更に、基礎語をキーにして逆引きも可能な辞書構成法を提案する。

本方法により、手話の未定義語を新規登録する場合に、既に登録されている手話単語の構成要素と比較することで、容易に類義概念との弁別要素を半自動的に推論し、合成語として記述できる利点を持つ。

キーワード 電子化辞書, 手話, シソーラス

A Method to Construct A Japanese-Sign Electronic Dictionary

Hisahiro Adachi, Hidehiko Shimoyama, Yoshihiro Harima and Takeharu Matsuura

Department of Information Science

Faculty of Engineering

Utsunomiya University

2753 Ishii-machi, Utsunomiya, Tochigi 321, Japan

Abstract

The objective of the study is to establish a construction method of a Japanese-Sign electronic dictionary. The most important problem to construct a bilingual dictionary is a difference of vocabularies between two languages. While a sign language, which is used among hearing impaired people, has a few thousand words, Japanese language has more than 50 thousands words. In the process of translation from Japanese to a sign language, it is often necessary to put Japanese words into another way and to translate to corresponding sign words.

Sign words can be classified into types, i.e., basic words and composite word. We pay attention to the property that the composite words are expressed as a sequence of basic words and that the sign language has a high word-building capacity by changing basic wrds of the sequence. This suggests that it is desirable that the words in the dictionary must be represented a sequence of words. In this study, we construct a dictionary according to this observation and a function of mapping using Japanese thesaurus. The resultant dictionary can be used to inverted search.

Keywords Electronic Dictionary, Sign Language, Thesaurus

1 まえがき

近年、手話に対する関心が高まって来ている。特に、健聴者側から手話を理解しようという動きが各所で見られる。しかし、実際に手話を学び聴覚障害者と会話を始めると日本語と手話の多くのギャップに直面する。まず、現在市販されている手話辞書（冊子体）の登録語数（日本語見出し）は約4,000語程度であるのに対して、日本語辞書のそれは約6万語程度であり、この登録語数の差が問題となる[2, 3, 4, 5, 6, 7]。更に、全国的に浸透している手話単語数は約1,000程度との報告もある[1]。実際、日本語の放送原稿を毎日、手話通訳している現場ではこの語数のギャップが最大の問題となっている[1]。本稿では、この語数のギャップを吸収する枠組みを提案し、実際に日本語・手話電子化辞書を試作し、その有効性と問題点の検討を行う。

従来、手話に対する工学的なアプローチの多くは、画像処理（手話の認識、生成）をターゲットとした手指動作に力点を置いたものである[8, 10, 11]。また、手指動作に基礎を置いた単語のクラスタリングも一部で試みられている[9]。我々は、従来手話をノンバーバル（非言語）の一種として定義されていた考えを否定し、バーバル（言語）をノンボイカル（非音声言語）とボイカル（音声言語）とし、手話はバーバルの一種、すなわち自然言語として定義する考えを提唱する。本研究は、手話通訳システムの研究の一環として位置づけられ、手話に対する新しい工学的アプローチとして、手話の言語的特徴に着目し、その言語処理に力点を置いたものである。

本方法は、手話単語の造語成分に着目し、約1,000個の基礎語彙を抽出した。また、造語成分の最後の要素に主概念が制約されること、更に、既存の日本語シソーラスとのマッピングが容易に実現でき、この基礎語彙の組み合わせで手話を記述できることを述べる。これにより、今後辞書を大規模化する際に、この制限された基礎語彙の範囲で記述することは、語数のギャップを解消するのに有効である。更に、手話の発信・受信の双方で理解容易性を向上させる効果が期待できる。

以下、2章で、本稿で対象とした言語データの特徴について述べ、3章で、この言語データから抽出された手話単語の造語成分とその特徴について述べる。4章で、既存の日本語シソーラスと手話単語の分類結果とのマッピング方法を述べ、5章で、試作した電子化辞書の構成法とその検索機能、実験結果を示す。最後に、6章で考察と今後の課題について述べる。

2 源辞書の言語データの特徴

手話単語を電子化する際の言語データとしては、既存の冊子体の辞書が複数市販されている。一般に、そこに掲載されている記述項目は、日本語見出し、対応する手話イラスト（写真）、あるいは手指動作記述文（手続き文）、更には類義語などである。しかし、その編集方針として手話のタイプを規定している。日本の手話は日本語との関係で3つに大別され、より聴覚障害者側からより健聴者側にシフトした、伝統的手話、中間型手話、同時法的手話と呼ばれている。これら3つのタイプの違いは参考文献[8]に示し、本稿では省略する。このタイプの差は、手話単語の造語法と対応概念の弁別に影響を与える。

今回、対象とした源辞書は、イラスト手話辞典[2]である。この辞書は、収録語数が5,177見出しであり、手話のタイプとしては、中間型に分類できる。中間型は中途失聴者を含め広く普及しているタイプである。また、他の辞書との大きな違いは、手話単語が基準語（2,021語）と非基準語（3,156語）に分けられており、非基準語は索引にその構成要素成分を記載するのみとし、本文では基準語のみ掲載している点が挙げられる。

2.1 基礎語の定義

本稿で以後使用する用語を定義する。

- 基準語：源辞書で造語成分の明記されていない手話単語見出し
- 非基準語：源辞書で造語成分の明記されている手話単語見出し
- 基礎語：非基準語の造語成分として記述されている単語見出し

一般に、語彙論では、基本語彙を「ある範囲の言語資料体において、延べ語数の多くを記述できるように選定された比較的小数の見出し語」と定義される。更に、基礎語彙とは、「それだけで色々な意味を表わすことができるもの」と定義される。例えば、英語辞書の LDOCE[] は、約 56,000 語の見出しを 2,851 語の制限言語（基礎語彙）で語義を記述している。本稿の基準語、非基準語、基礎語の定義はこの考えに準拠している。

例 1. 対象データの基準語、非基準語、基礎語の例

- 基準語：愛、相交わず、挨拶、...
- 非基準語：相性（気持ち+一致）、愛情（愛+気持ち）、愛人（内緒+愛+人）、...
- 基礎語：気持ち、一致、愛、内緒、人、...

尚、今回は分析の対象としなかった基準語の中にも、その手指動作記述文を解析すると非基準語の範疇に分類される見出しが存在する。つまり、基準語が減少し、非基準語が増加するが、基礎語は逆に減少する傾向にある。

例 2. 基準語（閉会）の場合

- 閉会：動作 1 指の間を開けた両手を前後にして立て、両側の手を前方に出す。
- 動作 2 掌を前方に向けた両手を左右から中央に寄せていき、人差指側をつけ合わせる。
- 閉会 = 動作 1 + 動作 2 = 大会 + 閉まる ここで「+」はメタ言語（概念記述子）である。

例 3. 基準語（馬鹿、賢い）の場合

- 馬鹿：こめかみに親指をあてて、他の四指を勢いよくその親指につけ合わせる。
- 賢い：親指と人差指の指先を付け合わせてこめかみにあて、人差指を上にはね上げる。
- 馬鹿 = 反対（賢い） ここで「反対（*）」は、メタ言語（概念記述子）である。

2.2 合成語の定義

- 合成語：基礎語をその造語成分の要素として持つ非基準語の内で、
複数要素（2個以上）から成る単語見出し
- 単純語：基礎語をその造語成分の要素として持つ非基準語の内で、
単数要素（1個）から成る単語見出し

更に、合成語と非合成語ともに同一要素（順序保存）で記述された単語を同義語として扱う。また、日本語と手話単語の概念の弁別の差から日本語としては、類義（上位下位関係を含む）語と考えられるものも同義語として扱う。この差は、4章で述べる日本語シソーラスとのマッピング操作で吸収される。

例 4. 同義語列の例

- 演じる（基礎語）：劇、芝居、演劇（非合成語の同義語列）
- 暗い+過ぎる（基礎語列）：深夜、夜更け、夜中（合成語の同義語列）

3 合成語の造語成分の特徴

本章では、手話単語の造語成分である基礎語を構成要素数と要素順序からの2つの観点からその特徴を分析する。尚、単純語は、892語抽出され、その内同義語をマージした結果、565語に圧縮された。つまり、延べ327見出しが同義語として圧縮された。

3.1 造語成分の要素数から見た特徴

合成語(2,269語)の要素数の分布を表1に示す。この結果、合成語全体の約78%が2項関係で構成され、3項関係までで93%をカバーしている。参考のため、例5に6項関係と9項関係を示す。更に、3項関係から2項関係への置換率は39%、4項関係から2項関係への置換率は48%の結果を得た。例6に、多項関係から2項関係に置換された例を示す。

表1. 合成語の構成要素数の分布と2項関係への置換率

要素の個数	総数	マージ後の総数	同義総数	置換数	置換率
2項関係表現	1,772	1,619	153	—	—
3項関係表現	431	423	9	210	39%
4項関係表現	61	58	2	28	48%
5項関係表現	4	4	0	—	—
6項関係表現	1	1	0	—	—
9項関係表現	1	1	0	—	—
合計	2,269	2,106	164		

例5. 6項関係、9項関係の単語とその造語成分

6項関係表現：災害＝雨＋風＋地震＋いろいろ＋起こる＋混乱

9項関係表現：ゴールデンウィーク＝四＋～月＋～から＋五＋～月＋祝う＋休む＋休む＋間

例6. 多項関係から2項関係への還元例

4項関係から2項関係への置換：

外交官＝世界＋交渉＋責任＋人

＝外交＋担当者

(世界＋交渉＝外交，責任＋人＝担当者)

3項関係から2項関係への置換：

残業＝暗い＋過ぎる＋仕事

＝深夜 | 夜更け | 夜中＋仕事

(暗い＋過ぎる＝深夜 | 夜更け | 夜中)

上記の結果から、原則として、手話単語は前方から順にまとめられ、概念を構成していく要素合成法に基づいて造語され、受信者側でも効率良く概念を再構成し、理解を助けていると考える。

3.2 造語成分の要素順序から見た特徴

「最後の成分要素が共通している」という観点で合成語をマージした結果の一部を例7に示す。

例7. 合成語を最終要素でマージした結果の一部 (基礎語：場所)

51 場所 位置/居所/オフィス/会場A/会場B/家庭裁判所/金融機関/刑務所/
劇場/現場/公園A/公園B/皇居/興信所/拘置所/交番/国籍/国会議事堂/
盛り場/座席/事務所/地元/住所/塾B/職場/職業安定所A/職業安定所B/

食堂/閑所/戦場/そこ/そば/台所/待避所/中央/駐在所 A/
駐在所 B (会社の)/勤め先/場/福祉事務所/保育所/保健所 A/保健所 B/
本籍/本店/名所/役所/役場/遊園地/郵便局/留置所

この結果、最後の要素成分は936項目あり、基礎語全体が1158語中の81%が最後の要素として出現している。更には、全体の上位9%の基礎語22語が合成語全体の57%をカバーしている。

次に、最後の要素成分(基礎語)の中には、例えば「人、人々、男、女...」の様に概念が近似している項目があり、これらを角川書店の類語新辞典[7]をソーラスとして大中分類で単純マージすると98項目(10%)に圧縮できることを確認した。しかし、例8に示すように最後の要素のみで分類した結果を分析すると、整合性のないものが存在する。(「法人」、「保証」、など)

例8. 手話側の分類項目「人(基礎語)」に分類された単語群

47人 愛人/アナウンサーA/恩人/外交官/カウンセラー/学者/歌手A/
患者A/教師A/教授A/玄人/恋人A/講師A/故人/社会福祉法人/
社団法人/主人/証人/新人/スター/スポンサー/第三者/他人A/他人B/
担当者/知人/長老A/~人/犯人/被告人/美人A/評論家/二人A
/弁護士A/変人A/変人B/法人/保護者/保証/保証人/本人/
盲人/目撃者/喪主/有識者/容疑者/レポーター

この中で、例9に示す様に「法人」は「裁判+人」を要素として構成されている。一方、日本語ソーラス上では分類項目「団体(713)」に分類されている。また、「保証」は日本語ソーラス上では、分類項目「約束(448d)」に分類されている。しかし、「保証」の造語要素と「保証人」の造語要素を比較すると、それぞれ「保証人」に対する別の手話表現(同義関係)と考えられる。これと同様な例として、「漫画」と「劇画」の関係がある。「漫画=面白い+絵+本、劇画=演じる+面白い+絵」つまり、「漫画」を「漫画本」とし、「漫画」に対して、最後の要素「本」を削除することで整合性がとれる。更に、「二人」は「ふたり(カップル)」を手話では意味している。

この様に、手話の造語成分をキーにした意味分類と日本語ソーラスの分類は単純に、一対一に対応できず、それぞれの分類項目をマッピングする操作を必要とする。これは、手話側での未定義語に対して日本語の分類項目をキーにして類義語を検索する際の問題点の一つである。この解決方法は、次章で述べる。

例9. 「法人、二人、保証」の造語成分と関連語の造語成分

法人=裁判+人
二人=二B+人 (二人=カップル)
保証=認めるA | 認めるB+証明+人
保証人=副+証明+人
証人=証明+人

4 日本語と手話の分類項目の適合方法

ここで、手話単語の概念は、その造語成分(基礎語)に依存し、特に最後の成分要素で規定されると仮定し、最後の成分要素を「概念規定語」と呼ぶことにする。

4.1 適合方法

日本語の単語の多くは複数の意味を含んでいるため、日本語シソーラス中の各単語は複数の概念項目番号が記述してある。例えば、概念規定語「場所」を持つ手話単語は例7に示した様に51個ある。各単語を日本語シソーラスで見出しをキーに検索する。この結果、複数の概念項目を持つ単語は7個ある(台所(740, 941), 機関(733, 991)など)。「台所(740, 941)」は、それぞれ「経済」、「部屋」の意味に対応する。しかし、手話単語は「台所=料理+場所」であり、「経済(台所が火の車)」の意味は担っていない。そこで、「台所(740, 941)」の分類番号740を削除する。これらの作業はすべて人手で行った。これらの作業後、概念規定語「場所」は、「場所(04, 10, 70, 72, 77)」の様に概念規定語をキーとして纏められた各単語の分類番号の上位2桁の分類番号の集合で規定される。これを単純に「場所」を見出しとして日本語シソーラスを検索すると「場所(100, 702, 724a)」の3つの分類項目しか得られない。

しかし、この作業の過程で例8に示した様に、概念規定語「人」を持つ単語群の中に「法人」が出現している。これは「人」で概念を規定するのでは無く、「裁判+人」で規定していることを意味する。従って、概念規定語に派生語として「法人」を新たに加える必要がある。これは、先に述べた単語の構成要素の数を減らす方向に働き、かつ上記の仮定にも反しない。ただ、この派生語を概念規定語に加える作業は、注意を要する。

この様に概念規定語の分類番号付与作業が終了すると、約1000個の概念規定語は約100個に分類され上位1桁でマージすると約10個に収束し、階層化される。また、概念規定語以外の基礎語についても、同様に分類番号の上位2桁でマージする。

この分類結果の利用法は、次章で詳しく述べるが以下にその一例を示す。

4.2 分類結果の利用法

手話側での未定義語の推論機能について「喜劇」を例に説明する。

「喜劇」に対応する日本語側での分類番号は881である。概念規定語の中で分類番号88(881の上位2桁)を持つ概念規定語を検索する。その結果、対応する概念規定語として「演じる」が選択され、図1に示した様に「演劇(881), 劇(881), 公演(884a), 芝居(881), 独演(883), 悲劇A(芝居)(881), 余興(890)」の各単語を表示する。

ここでユーザが、一番意味の近い「悲劇A(芝居)」を選択すると、概念規定語の直前の基礎語と同じ分類番号を持つ他の基礎語で置換した推論結果を提示し、ユーザに判断を指示する。この結果、「喜劇」に対する手話単語「すごい+面白い+演じる」が選択される。

```
input data = 喜劇
...undefined, but, now reasoning

1. 演劇, 劇, 芝居 = 演じる
2. 公演 = 公+演じる
3. 独演 = 自分+演じる
4. 余興 = 楽しい+演じる
5. 悲劇A (芝居) = すごい+悲しい+演じる
input number = 5

1. 喜劇 = すごい+哀れ+演じる
2. 演劇 = すごい+楽しい+演じる
3. 演劇 = すごい+楽しい+演じる
4. 演劇 = すごい+面白い+演じる
5. 演劇 = すごい+楽しい+演じる
input number = 4
```

図1. 未定義語「喜劇」に対する手話単語の推論

5 日本語・手話電子化辞書の試作とその検索機能

5.1 辞書の構成

本電子化辞書は、既存の日本語シソーラスから抽出した日本語分類番号対応テーブルと、既存の手話辞書記載の日本語単語を概念基準語で分類した手話分類番号対応テーブルで構成される。まず、日本語単語が入力されると、日本語分類番号対応テーブルで分類番号を検索し、この分類番号をキーとして、手話分類番号対応テーブルを検索し、手話の概念基準語が選択される。辞書本体は、この概念基準語をインデックスとして持つ複数のブロックで構成されている。これは、新規登録によりデータ規模が拡大した際に、ディスクからメモリ領域に必要最小限のブロックのみをロードするために利用する。

また、辞書データの各レコードは、日本語見出し（読み）、表記、分類番号、語源、手指動作記述文、次のレコードのポインタの6項目で構成されている。内部的には、見出し等をキーにした検索に対応するための各種のインデックステーブル（対応するポインタ番号表）を持っている。

5.2 検索機能

本電子化辞書の特徴は、この検索機能にある。一般に、手話を学び始めた初心者は、ある単語を学習した際に、その手話に関連した他の手話にどのような単語があるのかに興味を覚える。また、新しい手話単語を登録する際に、既に同一要素の組み合わせで別の単語が登録されているかを調べたり、概念を弁別するためには何処の要素を変更するか、あるいは追加すべきかを検証する機能が必要不可欠である。本システムでは、各単語は基礎語のみで構成されているので、ある単語の構成語をキーにして逆引き検索が可能である。例えば、「演劇」を検索すると構成語として「演じる（基礎語）」が提示され、その手指動作記述文から手話を学習できる。更に、前節で述べた未定義語に対して、推論結果から新規辞書登録に有益な情報を提示する機能を持つ。これらの機能から類似概念間の不整合（前述の劇画と漫画の例）を解消できる。

次に、この「演じる」を例に、以下の検索機能を説明する（図2に実行例を示す）。

1. 同義語検索 （「演じる」のみで構成されている単純語の検索）
2. 後方検索 （「演じる」を先頭の構成語として持つ合成語の検索）
3. 前方検索 （「演じる」を最後の構成語として持つ合成語の検索）
4. 埋め込み検索 （「演じる」を中心として前後に別の基礎語を持つ合成語の検索）
5. 類義語検索 （「演じる」を構成要素として持つ全単語の検索）

```
input data = 演劇
1 演劇 [えんげき] = 演じる

input data = *演じる
1 演劇 { えんげき = 演じる
2 演劇 { げき = 演じる
3 芝居 { しばい = 演じる

input data = *演じる
1 公演 { こうえん = 公+演じる
2 独演 { どくえん = 自分+演じる
3 悲劇A (芝居) { ひげき = すこい+悲しい+演じる
4 余興 { よきょう = 楽しい+演じる

input data = *演じる
1 アトラクション { あとらくしよん = 楽しい+演じる+披露
2 演芸 { えんげい = 楽しい+演じる+いろいろ

input data = *演じる
1 演技 { えんぎ = 演じる+技術
2 演出 { えんしゅつ = 演じる+指示
```

図2. 「演じる」を例とした検索実行例

6 おわりに

本稿では、既存の手話辞書と日本語シソーラスを利用し、手話の未定義語に対しても情報を提示する電子化辞書の構成法とその試作について述べた。今回、対象とした言語データは基準語を除いた非基準語に範囲を限定したが、基準語の多くは、手指動作記述文を解析することで合成語に分割できる。従って、限定された基礎語の語彙で手話単語を記述する本方法は有効であり、一般性を保つと考える。同様に、本電子化辞書の最終出力は手話を画像で表示する予定であるが、その際にも各単語毎に異なるフレーム集合を一对一に対応させると記憶領域が大きくなりすぎる。しかし、本辞書構成なら少ない基礎語に対してのみフレーム集合を用意し、これらを組み合わせることで実現できる可能性が高い。また、辞書の性質上頻繁な修正や登録が発生するが、その際にも柔軟に対応できる構成と考える。

今後の課題としては、対訳辞書のため日本語側の概念に依存した類義関係を導出したが、手話の造語成分（動作成分でもある）に着目すると、上位下位、反義対義関係の導出が挙げられる。例えば、概念規定語の「いろいろ」は抽象概念を表現する際に、具象概念を前置することで合成語を構成している。同様に、反義対義関係は動作が逆転している関係の手話が多いため、上記の基準語数を更に減少させる際に有効である。また、概念規定語の単語の結束性にバラツキがあり、これが手話側の対応概念を持つ単語の不足に影響されているのか否かを調査し、概念規定語の見直しを行う必要がある。また、概念を弁別する際に単語の「音」を指文字で表現している手話単語もあり、指文字を基礎語のサブクラスに加える必要がある。

謝辞

本研究を進めるにあたり、貴重なご意見ならびに資料を御提供頂いたNHK報道局紫垣喜紀チーフプロデューサ、聴力障害者情報文化センタ手話試験部山城秀生副部長、手話ニュース担当者の皆様に深謝します。また、日頃ご指導頂いている鎌田一雄教授に感謝します。

参考文献

- [1] 紫垣喜紀：手話ニュース，NHK内部資料，1991.
- [2] 丸山浩路編：イラスト手話辞典，ダイナミックセラーズ，1984.
- [3] 赤松保羅監修，金田富美編著：日本手話辞典，光書房，1980.
- [4] 栃木県聾学校，栃木県聴覚障害者協会編著：手指法辞典，1978.
- [5] 手話コミュニケーション研究会編：新・手話辞典，中央法規，1992.
- [6] 全日本ろうあ連盟編：わたしたちの手話，1972.
- [7] 大野晋，浜西正人著：角川類語新辞典，角川書店，1981.
- [8] 安達久博：“手話通訳システムの研究動向—自然言語処理の立場から—”，電子情報通信学会技術研究報告，NLC92-5，pp. 33-40，1992.
- [9] 藤野他：“機械処理のための手話構成パラメータの検討”，電子情報通信学会技術研究報告，HC90-10，1990.
- [10] 田村他：“手話動画像認識システム”，情報処理学会，コンピュータビジョン研究会資料，441，1986.
- [11] 長嶋他：“日本語手話の形態素分析とその記号化に関する検討”，日本ソフトウェア科学会第8回全国大会，1991.