

國學院大學學術情報リポジトリ

テキストアナリシスによる『むつぼしのひかり墨字
訳第1集～第4集』のことば

メタデータ	言語: Japanese 出版者: 公開日: 2024-03-12 キーワード (Ja): 近代語, 点字資料, 「むつぼしのひかり」, 墨字訳, テキストアナリシス キーワード (En): 作成者: イトウ, タカユキ メールアドレス: 所属:
URL	https://doi.org/10.57529/0002000142

テキストアナリシスによる 『むつぼしのひかり墨字訳第1集～ 第4集』のことば

伊 藤 孝 行

キーワード

近代語 点字資料 「むつぼしのひかり」 墨字訳 テキストアナリシス

要旨

「むつぼしのひかり」は、1903（明治36）年から1973（昭和48）年まで刊行されていた点字による雑誌である。「むつぼしのひかり」はその刊行期間の長さ（明治から昭和まで）、刊行頻度の高さ（戦前まではほぼ月刊、戦時中は休刊、戦後は年刊）、内容の豊富さ（論説・講演・雑録・和歌等）を兼ね備えている。「むつぼしのひかり」は明治から昭和にかけての盲教育史を映した資料でもあり、同時に明治から昭和にかけての点字による日本語の使用実態がどのようなものであったのかを記録した言語資料でもあり、いずれの面からみても資料価値はひじょうに高いと言える。2016年以降、『むつぼしのひかり』データ化・研究プロジェクトによって墨字訳が公開されている。2023年9月時点で第1号から第40号までの墨字訳が出版され、点字を読めない人々もその内容にアクセスできるようになっている。ただし、原文は主に表音式仮名遣いで記されているため、墨字訳ではこれがひらがな・カタカナ・漢字等に適宜変換されている。本稿では、特に「むつぼしのひかり」第1号から第40号までに掲載されていることばの使用実態に焦点を当てる。具体的には、『むつぼしのひかり墨字訳第1集～第4集』（第1号から第40号）を全文手入力でテキストデータ化を行い、Rとそのパッケージを適宜使用した。このテキストアナリシスにより

どのような言葉がどれほど使用されているのか（文字数、総語数、異語数、異語率）、品詞別頻度、品詞別割合、品詞別高頻度語、高頻度のことばの連なり（単語 2-gram、単語 3-gram）、高頻度語と共起することばについても調査をおこなった。また、TF-IDF（Term Frequency-Inverse Document Frequency）や潜在的ディリクレ配分法（LDA: Latent Dirichlet Allocation）による特徴語の抽出も行い、その中で「夢」が先行研究に於いてとりあげられることのなかった、隠れていたキーワードの1つであることも明らかにした。

1. はじめに

「むつぼしのひかり」は、1903（明治36）年から1973（昭和48）年まで刊行されていた点字による雑誌である。「むつぼしのひかり」はその刊行期間の長さ（明治から昭和まで）、刊行頻度の高さ（戦前まではほぼ月刊、戦時中は休刊、戦後は年刊）、内容の豊富さ（論説・講演・雑録・和歌等）を兼ね備えている。「むつぼしのひかり」は明治から昭和にかけての盲教育史を映した資料でもあり、同時に明治から昭和にかけての点字による日本語の使用実態がどのようなものであったのかを記録した言語資料でもあり、いずれの面からみても資料価値は非常に高いと言える。「むつぼしのひかり」刊行までの経緯や書誌については下田（1990）、青松・岩崎（2009）、土居（2016）、大塚（2016）に詳述されている。原文は一部を除き点字によって記されているため、点字を読むための知識とスキルが必須であった。2016年に点字雑誌『むつぼしのひかり』データ化・研究プロジェクトによる墨字訳『むつぼしのひかり墨字訳第1集』（「むつぼしのひかり」第1号から第10号）が社会福祉法人桜雲会点字出版部から刊行され、2023年9月時点で『むつぼしのひかり墨字訳第4集』（「むつぼしのひかり」第31号から第40号）まで刊行されている。そのおかげで「むつぼしのひかり」第1号から第40号まで、点字を解していない者にとってもその内容を理解することができるようになった。

ただし、原文は主として表音式仮名遣いであり、『むつぼしのひかり墨字訳第1集～第4集』では表音仮名遣いが適宜ひらがな・カタカナ・漢字等に変換されていることに留意する必要がある。参考までに諸星・伊藤・中野（2019）にある「むつぼしのひかり」第1号表紙を掲げた。この表紙は点字と墨字が併記してある。諸星・伊藤・中野（2019）より、点字をそのまま墨字にしたものを左側に、点字雑誌『むつぼしのひかり』データ化・研究プロジェクト（2016）より、点字を墨字訳にしたものを対照のために並記する。

〈表音仮名遣い〉	〈墨字訳〉
むつぼしの ひかり	『むつぼしのひかり』
だい 一ごー	第一号
めいち 三十六 ねん	明治三六年
六 がつ はつこー	六月発行
とーきよー もーあ がくこー	東京盲啞学校
もーせい どーそーかい	盲生同窓会

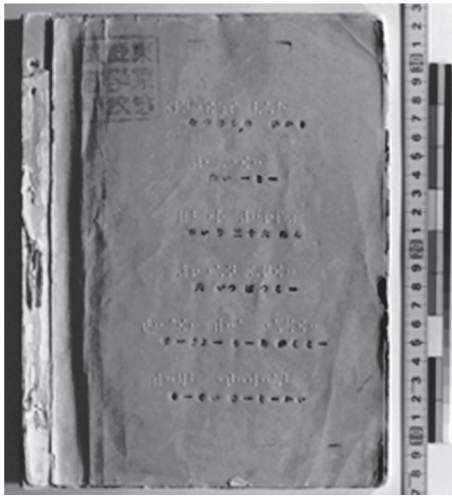


図1 「むつぼしのひかり」第1号 表紙

本稿では「むつぼしのひかり」の表音仮名遣いにはふれず、『むつぼしのひかり墨字訳第1集～第4集』に掲載されている「むつぼしのひかり」第1号から第40号までに掲載されていることばを対象とし、その使用実態を調査対象とする。具体的には「むつぼしのひかり」第1号から第40号にはどのようなことばがどれほど使用されているのか、どのようなことばの連なりがどれほど使用されているのか、どのようなことばが高頻度語と共起しているのか、「むつぼしのひかり」各号の特徴語はどのようなことばなのかを調査する。

2. テキストアナリシスによる『むつぼしのひかり 墨字訳第 1 集～第 4 集』 (第 1 号から第 40 号)

2. 1 テキストアナリシスのための環境設定

テキストアナリシスによる調査にあたり、『むつぼしのひかり 墨字訳第 1 集～第 4 集』(第 1 号から第 40 号)を全文手入力でテキストデータ化を行った。文字数は合計約 486500 字。第 1 号から第 40 号の間に 3 号欠号があるので、全 37 号である。テキストアナリシスにあたっては下記の環境下で行った。

macOS Ventura 13.4.1

R version 4.3.0 (2023-04-21)

RStudio 2023.06.0+421

MeCab 0.996

mecab-ipadic-NEologd : Neologism dictionary for MeCab

Documentation for package 'ca' version 0.71.1

Documentation for package 'ggplot2' version 3.4.3

Documentation for package 'readxl' version 1.4.3

Documentation for package 'RMeCab' version 1.11

Documentation for package 'tidyverse' version 2.0.0

Documentation for package 'tm' version 0.7-11

Documentation for package 'topicmodels' version 0.2-14

なお、以下に示す解析結果、具体的には品詞の分類・語の認定に関して、伝統的な文法体系の観点から見ると相違する結果があることについて触れておく。形態素解析は形態素解析器と形態素解析用辞書で解析される。本稿では上記の通り MeCab 0.996 (形態素解析器)、mecab-ipadic-NEologd (形態素解析用辞書)を使用した。mecab-ipadic-NEologd は、日本語の形態素解析用辞書としてこれまで自然言語処理分野で広く用いられてきた IPAdic⁽¹⁾に「MeCab の標準のシステム辞書では正しく分割できない固有表現などの語の表層(表記)とフリガナの組を約 319 万組(重複エントリを含む)採録⁽²⁾」したもので、言わば IPAdic の増補改訂版である。形態素解析の仕組みについて、小木曾(2014)によると「文字列で表現された文を入力として受け取り、これを単語に区切って品詞や読みなどの情報を付けて出力するものである。(中略)区切り方にはさまざまな組み合わせがあ

り得るから、形態素解析器は、その中で最適の組み合わせを出力する必要がある」とある。品詞の分類・語の認定伝統的な文法体系の観点から見ると相違する結果があるのはこのような事情ゆえのことである。また、形態素解析の弱点として未知語の問題・稀例の問題・表記と曖昧性の3点を挙げ、「漢字で書かれるのが普通である語が（中略）ひらがなで書かれている場合」に未知語として解析され誤解析が生じること、「かな書きによって別の語である可能性があるために解析を誤る場合も多い」とある。形態素解析はどの形態素解析用辞書を使用しても100%の解析精度ではないことも指摘している（IPAdicは新聞記事では98%近く、文学作品やブログでは95～96%程度）。そのような事情を踏まえ、『むつばしのひかり墨字訳第1集～第4集』（第1号～第40号）に含まれる語彙を定量的に分析し、適宜註を付す。

2. 2 文字数、総語数、異語数、異語率

『むつばしのひかり墨字訳第1集～第4集』（第1号から第40号）に掲載されている文字数、総語数、異語数、異語率は下記の通りである。

文字数については十の位を四捨五入した概数を掲げた。文字数がもっとも少ないのは第11号の約5900、文字数がもっとも多いのは第6号の20300である。総語数については、もっとも少ないのは第11号の3846、もっとも多いのは第6号の13128である。異語数については、もっとも少ないのは第11号の1132、もっとも多いのは第6号の3032である。異語率については単に異語数を総語数で除し、小数第三位を四捨五入したものであるが、もっとも小さいのは第19号の0.21、もっとも大きいのは第11号0.29である。前掲表においては第11号と第24号が共に0.29であるが、第11号が0.294、第24号が0.293であるので第11号が異語率のもっとも大きい号となる。「むつばしのひかり」の規模を概観する目安として平均についてもふれると、文字数は約13000、総語数は約8500、異語数は約2000、異語率は0.24となる。第1号から第40号の中では第10号と第39号が平均的な号といえる。

表1 『むつぼしのひかり墨字訳第1集～第4集』 文字数・総語数・異語数・異語率

号	文字数 (概数)	総語数	異語数	異語率 (TTR)
1	12600	8007	1896	0.24
2	12400	8057	2088	0.26
3	15700	9890	2384	0.24
4	15300	10102	2424	0.24
5	14800	9332	2142	0.23
6	20300	13128	3032	0.23
7	14800	9158	2227	0.24
8	14000	8967	2205	0.25
9	15200	9894	2318	0.23
10	13000	8441	2078	0.25
11	5900	3846	1132	0.29
12	9600	6135	1662	0.27
13	7600	4868	1321	0.27
14	9500	6140	1703	0.28
15	欠号			
16	14500	9332	2323	0.25
17	10000	6356	1673	0.26
18	欠号			
19	15000	9653	2010	0.21
20	10000	6422	1553	0.24
21	16200	10525	2411	0.23
22	16200	10492	2713	0.26
23	欠号			
24	6200	4127	1212	0.29
25	12300	8199	1996	0.24
26	13600	8809	2057	0.23
27	14400	9236	2283	0.25
28	15000	9855	2192	0.22
29	18600	12101	2766	0.23
30	15100	9519	2263	0.24
31	11900	7718	2030	0.26
32	12300	7956	1833	0.23
33	12000	7977	1984	0.25
34	16000	10562	2307	0.22
35	11500	7634	1725	0.23
36	10700	7194	1569	0.22
37	12300	7946	1934	0.24
38	14100	9416	2079	0.22
39	13100	8697	2104	0.24
40	14800	9661	2091	0.22
1-40	486500	332000	66500	

2. 3 品詞別頻度、品詞別割合

『むつぼしのひかり墨字訳第1集～第4集』（第1号から第40号）に掲載されている第1号から第40号の品詞別頻度は下記の通りである。なお、パーセンテージについては小数第三位を四捨五入した。

表2 『むつぼしのひかり墨字訳第1集～第4集』品詞別頻度

品詞	頻度	%
名詞	107485	34.08
助詞	79310	25.15
記号	48723	15.45
動詞	37610	11.93
助動詞	19840	6.29
副詞	7347	2.33
形容詞	4209	1.33
連体詞	3929	1.25
接続詞	3608	1.14
接頭詞	2781	0.88
感動詞	380	0.12
フィラー	138	0.04
その他	8	0.00

品詞別の頻度については名詞が約34%、動詞が約12%、副詞が約2%、形容詞が約1%となっている。名詞の頻度が高く、副詞・形容詞の頻度が低いことがわかる。編集方針に使用語彙の品詞は名詞を多めにし、副詞・形容詞を控えめにする等の記述はないが、「むつぼしのひかり」第1号「発刊の辞」に「平易な文章で」という記述がある。この調査結果からも「むつぼしのひかり」は「発刊の辞」の通り、読者にとって理解しやすい記述を心がけていた編集方針がうかがえる。下記、参考までに「発刊の辞」を引用する。

『東京盲啞学校盲生同窓会報告』と、鍼按学友会の発行していた『盲人世界』とがなくなって、その代わりに生まれ出でたのが、この『むつぼしのひかり』である。

二種類の雑誌を発行して力を分割さるるよりも、一つに集めて全力をそそいだ方が得策であることは、もっとも見やすい道理である。それゆえ、今度両会が合併してこの挙に及んだので、この雑誌の体裁は、上の二雑誌を折衷したのをもととして、平易

な文章で、興味と実益とのなるたけ多いように、また多少文学的趣味を含むようにするつもりで、その材料は広く全国に募る考えである。

あわれこの愛児よ、今は幼くして顧みる人少なきも、日を経るに従い、多くの同胞の歓迎を受けて、いよいよ健全に發育して、『むつぼしのひかり』なるその名に願わくはもとらざれ。

明治三十六年六月 代表者記す。

(第1号、発刊の辞、下線は伊藤による。以下同じ)

2. 4 品詞別高頻度語

『むつぼしのひかり墨字訳第1集～第4集』(第1号から第40号)に掲載されている品詞別高頻度を掲げ、適宜用例を掲げる。

2. 4. 1 動詞(動詞 - 自立⁽³⁾、2732種、上位20 = 頻度174以上)

『むつぼしのひかり墨字訳第1集～第4集』(第1号から第40号)において使用されている動詞およびその頻度は下記の通りである。

する	5522	もつ	449	おる	383	知る	297	終わる	221
ある ⁽⁴⁾	2125	あり	436	見る	348	せる ⁽⁸⁾	292	聞く	184
なる	1591	思う	405	なす	338	できる	225	向かう	182
いう ⁽⁵⁾	755	言う ⁽⁶⁾	387	しかる ⁽⁷⁾	297	至る	222	行う	174

2. 4. 2 形容詞(形容詞 - 自立⁽⁹⁾、430種、上位20 = 頻度32以上)

『むつぼしのひかり墨字訳第1集～第4集』(第1号から第40号)において使用されている形容詞およびその頻度は下記の通りである。量をあらわす「多い」「少ない」、大きさをあらわす「大きい」「小さい」、善悪をあらわす「良い」「悪い」の頻度が高い。

ない	857	少ない	100	広い	61	よろしい	48	良い	39
多い	177	小さい	65	強い	58	早い	47	嬉しい	35
よい	120	長い	65	高い	55	硬い	43	悪い	35
深い	102	無い	62	大きい	54	遠い	41	ゆるい	32

以下、『むつぼしのひかり墨字訳第1集～第4集』(第1号から第40号)における使用実態および資料紹介を兼ね、初出例を掲げる。

それゆえ、今度両会が合併してこの挙に及んだので、この雑誌の体裁は、上の二雑誌を折衷したのをもととして、平易な文章で、興味と実益とのなるだけ多いように、また多少文学的趣味を含むようにするつもりで、その材料は広く全国に募る考えである。

(第1号、発刊の辞)

第二 生物を作る元素には炭素、酸素、水素、窒素が多くありますが、無生物には炭素や窒素があるものが少ない。

(第1号、鍼按学講義)

もし色がひどくついたり、傷が大きいときには百倍の石炭酸水をもって湿した湿布を施さねばならぬ。

(第1号、鍼按学講義)

そこで、サルはいかにヒトと似ておるにもせよ、頭はそのからだに比して小さい。

(第1号、鍼按学講義)

が、本月書かないものは来月の投書と一緒にして、良いのを選んで出すつもりだから、ここへ出なくも失望しないで、どしどし投書してもらいたい。

(第16号、五行文)

この法をはじめてから、寒中でもすこぶる健康でどこが一つ悪いということがありません。

(第2号、空気浴)

2. 4. 3 名詞 (名詞 - 一般⁽¹⁰⁾、8136種、上位20 = 頻度157以上)

『むつぼしのひかり墨字訳第1集～第4集』(第1号から第40号)において使用されている名詞および使用数は下記の通りである。「盲人」「盲」「盲啞」が高頻度であることは「むつぼしのひかり」の特徴をよくあらわしている。「諸君」は「むつぼしのひかり」が読者への呼びかけを多用しており、前掲「発刊の辞」の内容と合致する。編集側から読者側へ向けて一方的に情報を提供する編集方針ではなく、読者からの声と共に編集している方針を明確にあらわしている証左であり、「諸君」が高頻度であることも「むつぼしのひかり」の特徴をよくあらわしている。「骨」「関節」「患者」「鍼按」は『むつぼしのひかり墨字訳第1集～第4集』(第1号から第40号)に按摩や鍼按に関する記事が多数あることからこのような結果となっている。

盲人	758	盲	260	手	220	患者	167	本校	162
人	616	骨	259	盲啞	206	下	166	世	161

諸君	353	吾	234	心	205	社会	163	鍼按	161
縁	272	関節	234	花	176	学校	162	雑誌	157

以下、初出例を掲げる。

こたび発行の『むつぼしのひかり』は、ひとり本会の報告会員の相互いの便宜を図るのみの目的ならで、広く世間の盲人諸氏のために、按摩、鍼ならびに音楽上の、学術技芸に関わることは言うもさらなり。

(第1号、『むつぼしのひかり』発行につきて)

幸いに小学校令には漢字を制限せられたるも数百年來の慣用、俗を成し、世人をして今にわかにこれのみに準拠せしめあたわざる事情もあり、徒勞と知りながら、これを教えねばならぬ教員の苦心と生徒の困難を思うばかりもなかなかの煩悶にして、国語の一定を一日も早く実行するに至ることを希望するもの本校教員生徒にしくものなかるべく、なお上は大学生より下は尋常小学校生徒に至るまでの徒勞困難にして、決して盲啞に限れるものとして等閑に付すべきものでないと固く信じています。

(第1号、卒業証書授与式における小西校長の告辞の概要)

しかし、相変わらずろくなことは書けぬのでありますから、読まるる諸君は、我が文の拙きと事柄の間違いとをこらえて読んで下さい。

(第1号、我が国盲人諸君に望む)

骨傷⁽¹¹⁾ これはからだのうちの骨が折れるので、よほど難しい。

(第1号、救急法)

小坂井氏の肩甲関節脱臼についての記事、また多少有益なり。

(第4号、岐阜と掛川の同窓会報告を読む)

まず患者を静かに寝せて、緩やかに衣服を脱がしめ、氷嚢をもってその部を冷やすぐらいに留めて、早く医者を招かなければならぬ。

(第1号、救急法)

このたび鍼按⁽¹²⁾学友会が、盲生同窓会の鍼按部となりし結果、『盲人世界』も『盲生同窓会報告』と合併して、『むつぼしのひかり』ちょう一の新雑誌となりて現わるることとなった。

(第1号、鍼按学講義)

2. 4. 4 副詞（副詞 - 一般⁽¹³⁾、683 種、上位 20 = 頻度 55 以上）

『むつぼしのひかり墨字訳第1集～第4集』（第1号から第40号）において使用されている副詞および使用数は下記の通りである。「しかして」「しかるに」が副詞として形態素解析されているが、これらは接続詞として捉えるのが一般的であるため、接続詞の項にも掲載する。

しかして ⁽¹⁴⁾	231	よく	128	常に	91	るる ⁽¹⁵⁾	75	しかるに ⁽¹⁶⁾	64
実に	152	ついに	103	決して	85	すでに	69	少しく	64
いかに	129	ことに	100	大いに	79	もし	69	もちろん	58
まず	128	もっとも	94	第一	78	ほとんど	67	互いに	55

以下、同様に初出例を掲げる。

生物は必ず同じ種類の親より生まれ、いくらかの間生きておって、その定まりたる時を経ればまた必ず死ぬものである。しかしてその生きておる間に、また同じ種類の子を産み残すものであります。

（第1号、鍼按学講義）

あまりと思ったゆえ下女に聞いてみると、この地の相場は二、三〇銭とのこと。実にひどいやつもあるものでないか。

（第1号、（二）汽車中のことば）

本年、盲生卒業生にして郷里に帰られしは、小松、小野、永堀の三氏なり。錦を着て帰るふるさと、いかに愉快であろうか。塵の都を去って、山青く水清き故山の地において、のこりなく発揮せよ、学びえし技術と学理。

（第1号、（三）卒業生を送る）

血が出るならば、二五倍ぐらいの石炭酸または清水でよく洗い、二五倍の石炭酸で湿した湿布を巻きつけ静かにしておいて、後は医者に譲るがよい。

（第1号、救急法）

しかるに、無生物は決してさようなことがないもので、親もなく、子もなく、ただ自然にその分子が寄り集まれば体をなし、別れ散れば体が消え失するというまでのことであります。

（第1号、鍼按学講義）

さて、それで諸君が、ものに生物と無生物という区別のあることがおわかりになった

ら、これからその生物と無生物とにいかなる違いがあるかということを、少しく述べてみようと思います。

(第 1 号、鍼按学講義)

2. 4. 5 接続詞 (接続詞⁽¹⁷⁾、78 種、上位 20 = 頻度 40 以上)

『むつぼしのひかり墨字訳第 1 集～第 4 集』(第 1 号から第 40 号)において使用されている接続詞および使用数は下記の通りである。

また	873	あるいは	227	かつ	98	されど	60	さて	43
すなわち	348	ただ	201	しかも	81	そうして	55	もしくは	42
および	256	なお	184	ただし	73	それから	49	それで	41
ゆえに	236	または	141	しかし	67	例えば	46	否 ⁽¹⁸⁾	40

以下、同様に地の文で使用されている初出例を掲げる。

かのバイ菌、アメイバの類がすなわちこの微生物体であります。

(第 1 号、鍼按学講義)

ゆえに冷水摩擦を毎朝一回行い、寝に就く時空気浴を行う順序に、双方を実行したなら冬の衛生には最良法であります。

(第 2 号、空気浴)

されど寒中裸体となるゆえ、ただじっとしていることは寒さに耐えられず、かえって風邪をひくようなことになるから、これには運動が伴わなければなりません。

(第 2 号、空気浴)

有眼者の鍼按業断じて許すべからず、否、進んで従来のを厳禁せよ。

(第 2 号、鍼按專業論)

2. 5 『むつぼしのひかり墨字訳第 1 集～第 4 集』の語の連なり - 単語 n-gram

(n = 2-3、動詞 (一般)・形容詞 (自立)・名詞 (一般)・副詞に限定)

『むつぼしのひかり墨字訳第 1 集～第 4 集』(第 1 号から第 40 号)において使用されていることばの連なりの実態について調査する。本稿では連なりの長さを 2 および 3 とし、対象とする品詞を動詞 (一般)・形容詞 (自立)・名詞 (一般)・副詞に限定し、単語 2-gram および単語 3-gram を調査する。

〈2-gram〉(上位5 = 頻度48以上)			〈3-gram〉(上位3 = 頻度9以上)		
盲啞 - 学校	名詞 - 名詞	74	化 - 骨 - する	名詞 - 名詞 - 動詞	11
訓 - 盲	名詞 - 名詞	69	我が国 - 盲人 - 社会	名詞 - 名詞 - 名詞	10
なる - おる	動詞 - 動詞	54	私立 - 訓 - 盲	名詞 - 名詞 - 名詞	9
する - おる	動詞 - 動詞	49			
技芸 - 欄	名詞 - 名詞	48			

単語 2-gram の頻度が高いものとして「盲啞 - 学校」「訓 - 盲」の連なりが多いことがわかる。「むつぼしのひかり」の内容からこれらの連なりが多いことはもっともなことではあるが、単語 2-gram による調査からもあらためてその特徴が明らかになった。「なる - おる」「する - おる」についても「なっておる」「しておる」という表現が多用されていることがわかる。「技芸 - 欄」は「むつぼしのひかり」に「技芸欄」があることから高頻度の連なりであることがわかる。

単語 3-gram の頻度が高いものとして「我が国 - 盲人 - 社会」が連なりの多いものとなっている。これもまた「むつぼしのひかり」において講演や告知等でよく使われる表現である。単語 2-gram でも「我が国 - 盲人」(21 例)と「わが国 - 盲人」(22 例)は高頻度の連なりである。

吾人は号を追うに従い、実質、体裁ともに精選改良して、我が国盲人社会の羅針盤たり、灯明台たることを確信す。嬉しさのあまり、妄評多謝。

(第2号、諸方より本誌の発行を祝して賜りたる詩歌および批評)
とにかく、困難なる雑誌の発行をば、両院において持続さるるは、我が国盲人社会のために、もっとも吾人が感謝するところなり。願わくは、ますます発達進歩せんことを。

(第4号、(三) 岐阜と掛川の同窓会報告を読む)
産まれてわずか七ヵ月の嬰兒なれども、我が国盲人社会の羅針盤をもって自ら任ずるは本誌なり。

(第5号、改善の辞)

2. 6 『むつぼしのひかり墨字訳第1集～第4集』高頻度語との共起語

前述した高頻度語から「盲人」「諸君」を検索語とし、検索語の近くで使用されていることば、つまり「盲人」「諸君」の共起語を調査する。本稿では検索語「盲人」「諸君」の前後5語を共起語の調査範囲とし、「盲人」「諸君」との共起が統計的に偶然ではないと

され得る共起語を明らかにする。前述の単語 2-gram および単語 3-gram との違いは、「盲人」「諸君」ということばと連なっている必要はなく、「盲人」「諸君」ということばと近い箇所で使用されていることばを知ることができる。共起語を調査することにより、「むつぼしのひかり」で使用されている「盲人」「諸君」がどのような文脈において使用されているのかを把握することができる。

2. 6. 1 「盲人」と共起度の高い語（上位 10）

『むつぼしのひかり墨字訳第1集～第4集』（第1号から第40号）において「盲人」と共起が統計的に偶然ではないとされ得ることばは下記の通りである。石田（2017）にある「統計学では T 値の絶対値が 2 を超えるかどうかを簡易的な目安にすることがある」「MI 値が 1.58 を超える場合、二つの単語の間に共起関係があると判断される」という基準を使用し、T 値が 2 以上かつ MI 値が 1.58 以上のことばを対象とした。

なお、Before は Term にあることばが検索語（ここでは「盲人」）の前に現れる頻度、After は Term にあることばが検索語（ここでは「盲人」）の後に現れる頻度、Span はそのことばがスパン（本稿は 5 とした）の中に現れる頻度、Total はテキスト全体の出現頻度、T 値は実測値から期待値を引いた値を実測値の平方根で除した値、MI 値は検索語が共起語の情報をどの程度持っているかを示す指標である。

Term	Before	After	Span	Total	T	MI
教育	10	56	66	278	7.301560178	3.304147667
界 ⁽¹⁹⁾	1	38	39	67	5.987132	4.598007649
我が国	27	1	28	69	4.978086266	4.077525086
職業	2	31	33	108	5.292687592	3.668201238
英国	18	3	21	41	4.367532457	4.413460039
協会	1	10	11	10	3.244155295	5.516198144
観	2	13	15	20	3.748865009	4.963657121
社会	3	34	37	164	5.434733342	3.230595981
わが国	24	0	24	64	4.58498217	3.963657121
団体	3	6	9	8	2.935905566	5.548619622

今より三三年前、英国衆議院議員医学博士アーミテージ氏、同国盲人教育視察に先立ち、各国盲人調査委員長としてヴィクトリア皇帝陛下の諮問に答申したところであります。

（第1号、卒業証書授与式における小西校長の告辞の概要）

しかして、我が国現時の盲人教育界は、よくこれらに向かつて不足なきか、吾人、疑いなきあたわざるなり。

(第7号、(一) 現今我が国盲人教育者の欠点)

吾人は号を追うに従い、実質、体裁ともに精選改良して、我が国盲人社会の羅針盤たり、灯明台たることを確信す。

(第2号、諸方より本誌の発行を祝して賜りたる詩歌および批評)

2. 6. 2 「諸君」と共起度の高い語（上位 10）

『むつぼしのひかり墨字訳第1集～第4集』（第1号から第40号）において「諸君」と共起が統計的に偶然ではないとされ得ることばは下記の通りである。こちらも前項同様に T 値が2以上かつ MI 値が1.58以上のことばを対象とした。

Term	Before	After	Span	Total	T	MI
盲人	776	18	794	758	27.87690195	6.548165444
教育	10	56	66	278	7.741010945	4.406677332
界	1	38	39	67	6.124910007	5.700537314
我が国	27	1	28	69	5.145544873	5.180054751
職業	2	31	33	108	5.534124791	4.770730903
社会	3	34	37	164	5.78097585	4.333125646
英国	18	3	21	41	4.482430229	5.515989704
協会	1	10	11	10	3.282875804	6.618727809
観	2	13	15	20	3.815181535	6.066186786
わが国	24	0	24	64	4.752751184	5.066186786

2. 7 『むつぼしのひかり墨字訳第1集～第4集』の特徴語

2. 7. 1 TF-IDF

『むつぼしのひかり墨字訳第1集～第4集』（第1号から第40号）における特徴語を別の観点から調査すべく TF-IDF（Term Frequency – Inverse Document Frequency）を調査する。TF-IDF は自然言語処理分野等で用いられている手法であり、文書に使用されていることばの重要度を数値化する方法の一つである。TF はあることばが文書の中においてどれほど使用されているかをあらわし、IDF はあることばが含まれている文書がどの程度少ない頻度であるかをあらわし、TF と IDF を乗じたものが TF-IDF である。つまり、

TF-IDF の値が大きいほどその文書において特徴的なことばといえる。

本稿では『むつぼしのひかり墨字訳第 1 集～第 4 集』（第 1 号から第 40 号）における TF-IDF の値が大きい名詞（一般）に限定し、特徴語上位 10 を明らかにする。これまで明らかにした高頻度語や単語 2-gram と併せ、『むつぼしのひかり墨字訳第 1 集～第 4 集』（第 1 号から第 40 号）における特徴語を調査する。

盲人	0.739154517
人	0.576420078
縁	0.419905298
関節	0.410906879
諸君	0.352801053
骨	0.339781682
吾	0.260514576
盲	0.255677929
盲啞	0.235996511
手	0.229556128

2. 7. 2 潜在的ディリクレ配分法 (LDA)

『むつぼしのひかり墨字訳第 1 集～第 4 集』（第 1 号から第 40 号）における特徴語をもう 1 つの方法で調査する。ここでは潜在的ディリクレ配分法 (LDA: Latent Dirichlet Allocation) を使用し、「むつぼしのひかり」各号を特徴づけているトピックは何なのか、そしてそれぞれのトピックを特徴づけていることばを探る。

下記に各トピックの特徴語上位 20 を掲げる。トピック 1 は「むつぼしのひかり」が同窓会誌以上の盲人のための会報誌としての一面が推定できる。トピック 2 は身体部位の名称や場所を表すことば（上、下、縁（へり））があることから按摩や鍼灸の情報誌としての一面が推定できる。トピック 3 は「夢」（156 件）と推定できる。「むつぼしのひかり」には「夢」ということばも高頻度語である。前述した高頻度語上位 20 にはないが 21 位が「夢」である。また、読者からの投稿にもよく使用されていることばである。トピック 4 は同窓会誌としての一面が推定できる。これらから『むつぼしのひかり墨字訳第 1 集～第 4 集』（第 1 号から第 40 号）は、発刊の辞にあった盲人にとって有用な会員のための情報誌であり、多くの盲人が生計を立てるうえで持ちあわせていた按摩や鍼灸についての具体的な情報誌であり、同窓会誌であり、編集側・読者側と共に各々の「夢」を育んでいた雑誌

なのではないか。殊に、トピック3の「夢」は、「むつぼしのひかり」を扱った先行研究においてとりあげられることのなかったことばであり、今後「むつぼしのひかり」におけるキーワードの1つとして留意しておくとなつたな発見や解釈の一助となるのではないだろうか。

Topic 1	Topic 2	Topic 3	Topic 4
盲人	縁	盲人	盲人
人	人	人	人
関節	盲人	盲	諸君
骨	諸君	吾	盲啞
諸君	孔	諸君	盲
吾	下	夢	骨
候	心	細胞	本校
鍼按	後ろ	手	職業
世	外	雑誌	韌帯
花	吾	社会	学校
盲	手	鍼按	教員
学校	上	先生	患者
社会	腹部	縁	下
先生	患者	本校	手
鍼治	花	点字	関節
会員	口蓋	本会	卒業生
心	次	心	世
本会	骨	郵税	腹部
他	溝	身	お呼び
月	三味線	盲啞	マッサージ

3. おわりに

以上、『むつぼしのひかり墨字訳第1集～第4集』（第1号から第40号）に掲載されている「むつぼしのひかり」第1号から第40号までに掲載されていることばを対象とし、「むつぼしのひかり」第1号から第40号の高頻度語、品詞別高頻度語、単語 2-gram・単語 3-gram、高頻度語「盲人」「諸君」の共起語、TF-IDF および LDA による特徴語を調査

した。『むつぼしのひかり墨字訳第 1 集～第 4 集』においては「夢」がキーワードの 1 つであることを指摘した。

注

- (1) IPAdic の品詞体系については、浅原・松本 (2008) および下記を参照されたい。
https://www.gavo.t.u-tokyo.ac.jp/~mine/japanese/nlp+slp/NAIST-JDIC_manual.pdf
<https://www.unixuser.org/~euske/doc/postag/#chasen>
<https://hayashibe.jp/tr/mecab/dictionary/ipadic>
- (2) <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>
- (3) 浅原・松本 (2008) によると IPAdic の品詞体系では動詞の下位分類として自立、接尾、非自立の 3 つが立項されている。接尾は「学校文法では助動詞にあたるもの」、非自立は「(て) くる」が例として挙げられている。
- (4) ラ行五段動詞「ある」・ラ変動詞「あり」については解析がうまくいかないという報告 (https://qiita.com/ensan_hcl/items/885588c7d2d99de85b44#fnref3) がある。
- (5) 動詞「いう」(例: 一般世の人は新年というと喜ばしがるが、歳末は少しも喜ばぬ。)
- (6) 動詞「言う」は 456 例
- (7) 動詞「叱る」は 0 例なので、誤解析と言える。
- (8) 動詞と形態素解析されているが、漢語サ変動詞の一部である (例: 附着せる、観察せる)。
- (9) 浅原・松本 (2008) によると「形容詞の活用型のうち、語幹の最後の母音がアウオのいずれかで終わるもの」・「形容詞の活用型のうち、語幹の最後の母音がイで終わるもの」・「いい」「ええ」・「形容詞の中で「基本形」の用法しかもたないもの」とある。
- (10) 浅原・松本 (2008) によると「普通名詞、あるいは、下位分類が未定の名詞」とある。
- (11) 「骨傷」は 2 例。「骨傷」は「骨」と「傷」に形態素解析される。
- (12) 「鍼按」は 1 語として形態素解析されている。「鍼按」が含まれる語は「鍼按科」20 例、「鍼按家」18 例、「鍼按学友会」7 例、「鍼按業」13 例、「鍼按部」27 例。
- (13) 浅原・松本 (2008) によると「必ず後ろで切れるもの、連体修飾が不可能なもの」とある。
- (14) 接続詞であるが、解析結果は副詞となっている。
- (15) 誤解析。
- (16) 接続詞であるが、解析結果は副詞となっている。
- (17) 浅原・松本 (2008) によると「独立に現れる接続詞」とある。
- (18) 感動詞であるが、解析結果は接続詞となっている。
- (19) 浅原・松本 (2008) によると「界」「観」は名詞・接尾に分類される。「名詞、あるいは他の品詞の

語幹〔ガル接続〕や〔連用タイ接続〕に接続して複合名詞を形成する語のうち、下位の分類にあてはまらないもの。一般に「接尾語」というよりも範囲が広く、複合名詞の後ろ要素として用いられることが普通なもの」とある。例として「観」「性」「学」「類」「面」「用」が挙げられている。

参考文献

- 青松利明、岩崎洋二（2009）「資料紹介／点字総合雑誌『むつぼしのひかり』第1号および『むつぼしのひかり』の墨訳・研究の現状」、『研究紀要』41、pp.37-48
- 浅原正幸、松本裕治（2008）『NAIST Japanese Dictionary version 0.4.0 ユーザーズマニュアル』、奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座
- 石田基広（2017）『Rによるテキストマイニング入門 第2版』、森北出版
- 大塚美紀（2016）「『むつぼしのひかり』刊行の歴史的意義と第10号までの概要」、『むつぼしのひかり 墨字訳 第1集』、pp.11-15、社会福祉法人桜雲会点字出版部
- 小木曾智信（2014）「第5章 形態素解析」、『講座日本語コーパス 2. 書き言葉コーパス—設計と構築—』、pp.89-115、朝倉書店
- 下田知江（1990）「むつぼしのひかり」、『障害者教育 福祉 リハビリテーション 目次総覧 別巻 第1期 第6巻・別巻1』、pp.14-16、大空社
- 点字雑誌『むつぼしのひかり』データ化・研究プロジェクト（2016）『むつぼしのひかり墨字訳第1集』、社会福祉法人桜雲会点字出版部
- 点字雑誌『むつぼしのひかり』データ化・研究プロジェクト（2017）『むつぼしのひかり墨字訳第2集』、社会福祉法人桜雲会点字出版部
- 点字雑誌『むつぼしのひかり』データ化・研究プロジェクト（2019）『むつぼしのひかり墨字訳第3集』、社会福祉法人桜雲会点字出版部
- 点字雑誌『むつぼしのひかり』データ化・研究プロジェクト（2020）『むつぼしのひかり墨字訳第4集』、社会福祉法人桜雲会点字出版部
- 中村満紀男（2018）『日本障害児教育史【戦前編】』、明石書店
- 諸星美智直、伊藤孝行、中野真樹（2019）「近代日本語資料としての点字雑誌『むつぼしのひかり』：福祉言語史の基礎資料としての近代日本語点字資料の調査と整備」、日本学術振興会科学研究費助成事業 基盤研究（C）研究成果報告書、JSPS 科研費 15K02574

謝辞

本研究は JSPS 科研費 JP19K00630 の助成を受けたものです。（研究種目：基盤研究（C）、研究課題：福祉言語史の基礎資料としての近代日本語点字資料の調査と整備）

