

国立国語研究所学術情報リポジトリ

Homophonic words and homographic words in modern newspapers

メタデータ	言語: jpn 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 田中, 章夫, TANAKA, Akio メールアドレス: 所属:
URL	https://doi.org/10.15084/00001008

新聞語彙調査の同音語と同形語

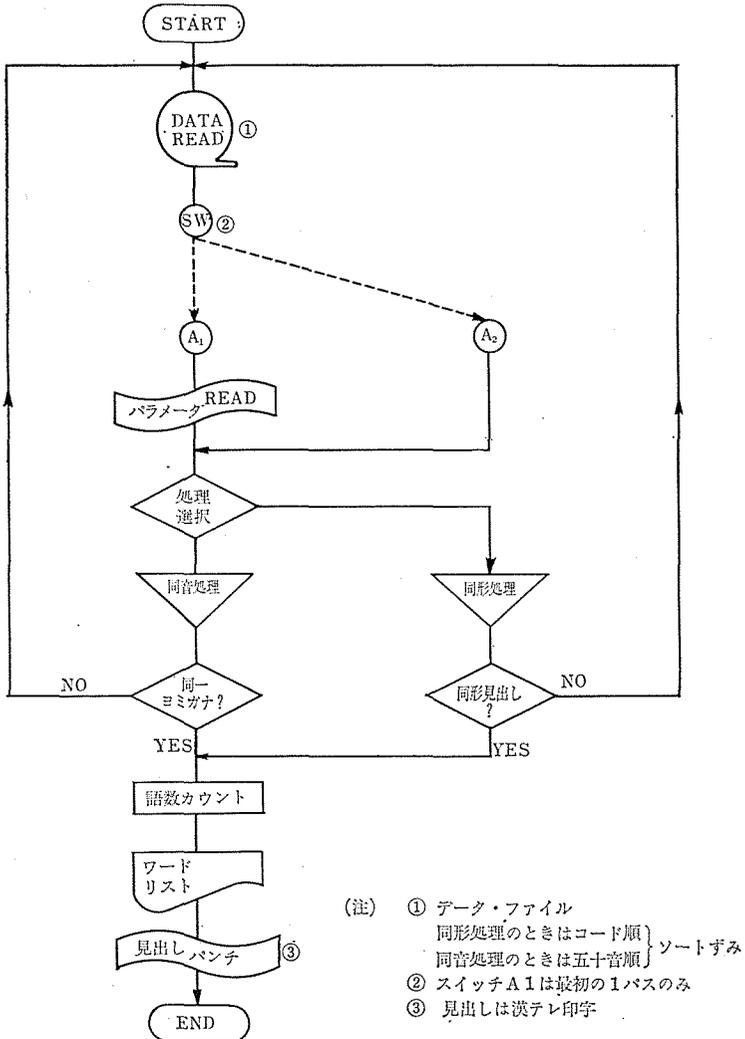
田 中 章 夫

0. はじめに

国立国語研究所報告38「電子計算機による新聞の語彙調査(II)」には、「同音短単位表」と「同形短単位表」が収められている。これらは、いずれも、現在までに集計が終了した一紙一年分相当量の新聞語彙調査のデータから、コンピュータ処理によって作成したものである。処理システムは、図1の流れ図の通りである。

図1の流れ図からわかるように「同音短単位表」は、「よく／良く／欲／翼」など、ヨミが同じで語形(表記形式)の異なるものを集めた表であり、「同形短単位表」は、逆に語形が同じでヨミの異なる「工夫(クフウ／コウフ)」の類を集めたものである。ここで使用した「ヨミ」はデータ(磁気テープ)の「ヨミガナ欄」に記入されているもので、原文でかな書きの場合は、ヨミガナ欄には、もとの表記がそのまま転写されている。原文が漢字書きのものについては、短単位処理作業において作業者が与えたヨミガナが記入されている。これは、短大卒程度の作業者が、新聞を読んでいくさいに最も常識的に採用するヨミガナを記入することを原則としたが、「三五円」「一九七〇年」などの形で用いられた場合の漢数字にかぎり「三(さん)五(ご)円」「一(いち)九(く)七(しち)〇(れい)年」のように、特別な規定を設けた。その理由は、数字は、「三／五／円」「一／九／七／〇／年」のように、一字一字が、一短単位として区切られて処理されるためである。ここで「短単位」というのは、語の長さのうえでは、「雑誌九十種の語彙調査(国立国語研究所報告21)」、「総合雑誌の語彙調査(国立国研究所報告12)」で採用した「β単位」に、ほぼ一致する言語単位(単語)である。くわしくは、電子計算機による新

図1 <同音・同形短単位表作成>
フロー・チャート



聞の語彙調査(Ⅱ)」の第Ⅰ章の「2. 単位の説明」を参照されたい。なお、「同音単位表」「同形短単位表」の作成作業には、研究補助員・堀江久美子の助力を得た。

1. 同音語の分析

1.0. 同音短単位表の性格

現在進行中の新聞の語彙調査の中間報告「電子計算機による新聞の語彙調査(国立国語研究所報告37)」では、短単位については度数5以上の13234の見出しについて語彙表を作成してある。これらの見出しは、いずれも表記別に集計してあるので、同一単語であっても、表記の異なるものには、それぞれ別の見出しが立っている。たとえば「後ろ(度数10)／うしろ(度数12)」「一斉に(度数5)／いっせいに(度数8)」となっている。これらは、表記別の見出しが立っても、いずれの表記も度数5以上なので、語彙表の上から、その単語の度数が、「ウシロ(22)」「イッセイニ(13)」のように、おさえられる。

しかし、「晴間(5)／晴れ間(3)」「電話(346)／でんわ(2)」のように、度数5に満たない表記をもつ語の場合には、「晴れ間」「でんわ」は、語彙表に掲載されないので、「晴間」「電話」についての集計結果しか知ることができない。また、「木綿(4)／もめん(2)」「若葉(4)／わかば(1)」などの類は、単語としては度数5以上でありながら、表記別集計のために、短単位語彙表には掲載されない結果となる。

表記別集計は、上に述べたように、語の表記の面での実態がわかる利点はあるが、語彙表を、ある度数で切った場合には、語としての度数が明らかにならない欠点を持つ。

「同音短単位表」を作成した第1の目的は、国立国語研究所報告37「電子計算機による語彙調査」の「五十音順短単位表」と「度数順短単位表」とがもっている上記の欠点を補うことにある。

しかし、ヨミガナ情報の一致する短単位を抽出していくと、「後ろ／うしろ、晴間／晴れ間、木綿／もめん」のような表記の異なる同音短単位ばかりで

なく、「科学／化学」「市立／私立」などの、いわゆる同音語も、すべて抽出されてくる。コンピューター処理としては、きわめて単純に、ヨミガナ情報の一致する短単位データを集計したにすぎないが、その結果、作成された「同音短単位表」には、つぎのような短単位セットが、それぞれの度数とともに、すべて集められている。

①表記にゆれのある短単位

(例) 歓び(6)／喜び(17)／よろこび(6) 道連れ(1)／道づれ(5)／みちづれ(1) 見事(9)／見ごと(1)／みごと(6) 日ざし(4)／陽ざし(1) 張り紙(2)／貼り紙(2) 肌(2)／膚(1) 函(2)／箱(9) 年々(11)／年年(1)

②送りがなにゆれのある短単位

(例) 異なる(7)／異なる(5) 晴着(9)／晴れ着(4) 日付(5)／日付け(1) 引受(5)／引受け(12)／引き受け(6) 少い(1)／少ない(36) 係り官(1)／係官(3) 付添(3)／付き添い(1)／付添い(1) 半ば(4)／半(7)

③いわゆる同音異義語のセット

(例) 市立(20)／私立(17) 科学(117)／化学(116)／価額(2) 学会(22)／学界(9) 国家(79)／国歌(1) 機関(129)／期間(76) 対照(8)／対象(38) 体制(68)／態勢(11) 福利(12)／複利(4) 要員(17)／要員(64)

④同音類義語のセット

(例) 機具(1)／器具(13) 追求(15)／追及(38)／追究(2) 辞典(19)／事典(13) 平行(2)／並行(7) 泣く(7)／鳴く(2) 暑い(5)／熱い(9) 塩(18)／潮(4) 玉(1)／珠(9)／球(8) 形(13)／型(98)
()の数值は、出現度数を示す。

ただし、上記(1)の場合、実際には、「タバコ／たばこ／煙草」のようなセ

ットが成立していたとしても、カタカナ書きの「タバコ」は、この表には出てこない。その理由は、原文で漢字書きのものとひらがな書きのものについては、ヨミガナ情報が、ひらがなで与えられているのに対して、原文でカタカナ書きのデータのヨミガナ情報は、カタカナで与えられているからである。上の例でいえば「タバコ(タバコ) / たばこ(たばこ) / 煙草(たばこ)」のように、()内の形でヨミガナ情報が与えられているため「同音短単位表」には、ヨミガナ情報の一致する「たばこ/煙草」のみが抽出され「タバコ」は出てこない。同じ理由で「ガラス(ガラス) 硝子(がらす)」のような、同音の短単位が存在したとしても、「同音短単位表」には抽出されてこない。この点については、今後プログラムの一部を修正する予定である。

この稿では、上記の3種の同音短単位の中から、③の「いわゆる同音異義セット」だけをとりあげて、考察を試みることにする。

1. 1. 「同音短単位表」の同音異義語

いままでに作成された各種の同音語表にくらべて、「同音短単位表」の同音語の、著しい特徴は、各語に出現度数がついていることである。従来の同音語表においては、同音語が、「歌学/化学/価額/課学/科学/家学/科学/下学/下顎」「布陣/夫人/不仁/婦人/不尽」あるいは、「摂政/殺生/折衝」のように列挙されているだけであるが、「同音短単位表」の場合は、出現度数は、これらの同音語のまぎれやすさを推定する一つの目安になり、出現度数の上から、ある程度のウェイトづけを試みることもできる。

上の例でいえば、「カガク」については、「科学(117) / 化学(116) / 価額(2)」となっていることから、「科学」と「化学」の間の混同が、他に比べて、格段に多く起こる可能性があるろうとか、「フジン」については、「婦人(155) / 夫人(50) / 布陣(3) / 不尽(1)」からみて、やはり「婦人/夫人」がまぎれやすそうだというようなことを推定することができる。

また、「セッション」の場合には「折衝(18) / 殺生(1) / 摂政(1)」の結果からみて、そう問題にはなるまいといった見当がつけられよう。同様な

例に「生産(134)／清算(5)／制酸(2)」のようなものがあるが、これも「生産」が他にくらべて圧倒的に頻度が高いことからみて、まぎれやすい同音語として問題になるようなものではなさそうだといえよう。

まぎれやすい同音語として有名な「市立(20)／私立(17)」は、やはり、度数が伯仲している。同様なものに「現象(21)／減少(20)」「地震(52)／自身(51)／自信(35)」「性格(34)／正確(23)」などがある。その逆に、「実験(68)／実権(2)」「調停(14)／朝廷(1)」「企画(56)／規格(2)」のように、度数がアンバランスなものも出てくる。もちろん、度数だけから、同音語のまぎれやすさを推定することはできないが、国立国語研究所報告20「同音語の研究」においても、同音語のまぎれやすさを生み出す条件の一つとして、使用頻度をとりあげ、「総合雑誌の用語・前編(国立国語研究所報告12)」の語彙表にもとづいて、まぎれやすさの段階づけを行なっている。

以上、述べたように、「同音短単位表」の各同音語セットに出現度数がついていることは、同音語のまぎれやすさを考えるさいの一つの手がかりを与えることになる。

つぎに、「同音短単位表」はコンピューターの出力結果を、そのまま掲げたため、従来の同音語表に見られない、つぎのような同音語のセットを見つけることができる。

①活用語の変化形を含む同音語セット

(例) 長さ(20)／流き(2) 反し(7)／半紙(1) 死な(7)
品(17) 良く(6)／欲(5)／翼(7)／翌(13)／浴(172)
来い(1)／濃い(10)／恋(45) 買っ(34)／飼っ(3)
／勝っ(13)／駆っ(1)

②現代かなづかい表記で同一よみがなとなる同音語セット

(例) 小売り(16)／公吏(1) 王(80)／追う(8)／負う(5)
乞う(21)／港(47)／甲(4) 問う(3)／父(10)／
党(522)／塔(4)

③固有名詞を含む同音語セット

(例) 佐藤(276)／砂糖(17) 義雄(13)／好男(1)／由夫(1)

／嘉夫(1)／芳郎(2)／良雄(5) 応酬(4)／欧州(33)
／奥州(1)／押収(6) 長野(68)／永野(7) 日光(22)
／日航(25)

④ 助詞・助動詞・接辞などを含む同音語セット

(例) まい(65)／毎(13)／枚(124) たく(29)／宅(44)／卓
(3) 頭(1)／かしら(6) 曹達(17)／そうだ(77)
其の(3)／園(18)

従来同音語集は、一般に発音式で見出しを立て、主に自立語のみを採集している。また、活用語は、終止形の形で採録するのが普通である。固有名詞については、国名・都道府県名あるいは大都市、山河、海の名程度を収容したものはあるが、固有名詞を広く採録しているものは、ほとんどない。

「同音短単位表」の同音語セットは、もちろん系統的ないしは体系的に選定したものではないが、現実の資料に出現したものを忠実に採録した点に、一つの意味があるといえよう。

1. 2. セットとなる語の種類が多い同音語

「同音短単位表」において、もっとも見出しの種類(数)が多い同音セットは、「こう」の39種類である。以下「かん(29)」「しょう(29)」「し(27)」と続いている。表1のようにして、種類の多い順に、同音短単位を示したのが、表1である。このように、種類の多い同音セットを集めてみると、目立つのは、一字漢語と姓名の名とである。前者の一字漢語は、その多くが、接頭・接尾的につかわれ、「光(化学)／高(価額)」「(試験)官／(試験)官」のような同音複合語を生み出す要素となるものである。後者は、いわゆる同姓同名を作り出す要素である。

表1 同音短単位表 (10種以上の語数順表)

() 内の数値は合計度数

こう	39種 (2036)	氏文 450	紙資 173	件具 47	けん 3	けんじ 19種 (64)	
興後口工公広行交光向好校考江高功紅鉸鋼坑	90: 47 3: 1 6: 10 639: 3 30: 193 3: 4 8: 9 24: 4 2: 10 18: 14 17: 49 167: 1 1: 2 2: 2 264: 23 4: 1 1: 1 110: 80 171: 21 1: 1	死系司私詞使試始指史 4 604 77 49 1 10 1 4 4 1 1 1 64	四し銅詩賜師誌視刺 5 1244 6634 11 44 3 96 43 13 2	間研權犬兼絹堅券 392 6 5 237 7 21 7 1 50	けん 3 鍵健 6 健健 14 健健 9 健健 4 軒謙 19 圈賢 22 堅拳 4 券 5	けんじ 1 研治 1 持二 3 堅次 1 度次 1 けんじ 8 司次 2 健次 2 健次 2 健治 2 健治 1	検事 9 謙治 1 謙司 10 憲治 2 賢治 2 賢治 1 賢治 1 拳健 14
かん	29種 (1727)	漢刊 6 2: 11 27: 2 60: 1 1: 17 416: 2 80: 441 18: 1 2: 1 122: 12 1: 69 2: 4 2: 5 129: 1 217: 1	規規 7 規規 2 喜喜 15 喜喜 11 喜喜 60 喜喜 11	世奇 23 奇製 1 製製 3 生聖 246 聖性 5 性整 217 整聲 6 聲制 1 制西 247 3	せい 20種 (1130)	か 18種 (3575)	
管汗感鑑簡間卷関貫環劬歛伍官館	75: 6 2: 11 27: 2 60: 1 1: 17 416: 2 80: 441 18: 1 2: 1 122: 12 1: 69 2: 4 2: 5 129: 1 217: 1	規規 7 規規 2 喜喜 15 喜喜 11 喜喜 60 喜喜 11	孝二 1 浩司 3 浩二 9 浩二 9 浩次 1 浩治 1 浩二 2 浩二 2 虹二 2 魏次 1 孝次 1	政勢 1 製製 2 製製 186 せい 21 正清 14 青盛 37 盛請 6 精 5 103	かい 19種 (2664)	きょう 18種 (390)	
しょう	29種 (1133)	捷 1 蔣 3 しょう 5 礁障 2 13: 1 209: 2 1: 1 39: 11 1: 32 27: 3 123: 124 4: 1 120: 2 329: 1 1: 1	工事 95 公示 1 広広 1 光光 1 光光 1 光光 2 光光 2 虹二 2 魏次 1 孝次 1	とう 20種 (1164)	きょう 18種 (390)		
性正相昇章小少昭消証商傷勝省昌	2: 1 8: 3 55: 5 12: 2 13: 1 209: 2 1: 1 39: 11 1: 32 27: 3 123: 124 4: 1 120: 2 329: 1 1: 1	新夫 1 壯夫 1 義男 3 義男 2 義男 7 義男 13 吉男 3 吉男 3 喜男 1 美男 1 由男 1 良男 1	投東 136 党盜 184 等島 522 3 24 102	とう 20種 (1164)	きょう 18種 (390)		
し	27種 (9682)	けん 22種 (1777)	よしお 23種 (55)	とう 20種 (1164)	きょう 18種 (390)		
士	136: 子 7	建 744	見 170	とう 20種 (1164)	きょう 18種 (390)		

一雄 9	嘉寿男 2	球究 32	求邱 73	報 17	邦 1	だん 10種 (377)	
かずお 2	賀寿老 1	4	1	1		男弾 1: 譚 2	
こうき 12種 (48)		304	76	かいそう 10種 (39)		122: だん 1	
後期 8	紅旗 7	う 3	817	廻送 1: 改葬 1		110: 談 36	
工期 1	鋼機 13	急級 138	125	回想 7: 改装 2		1: 断 8	
機器 2	香気 2	けいこ 11種 (76)		回路 14: 海藻 4		1: 暖 4	
光機 1	衛機 7	経子 1	けいこ 24	階会 7: 快走 1		ち 10種 (335)	
好機 3	網紀 2	景子 5	佳子 1	かんこう 10種 (152)		遅智 2: 置血 2	
好季 1	幸輝 1	子子 1	子子 17	管工 1: 肝硬 4		1: ち 28	
しげお 12種 (31)		慶子 6	啓子 2	管公 1: 完工 1		3: 致 7	
栄夫 1	茂穂 1	圭子 11	桂子 7	管寒 1: 慎行 7		12: 政 3	
成夫 1	茂雄 2	けい子 1		官公 6: 寛與 1		265: 数值 12	
重男 7	しげお 1	けい子 1		刊行 9: 観光 121		と 10種 (10752)	
重男 11	繁雄 1	けい子 1		きかん 10種 (234)		都土 520: 杜 2	
茂男 1	繁雄 1	計次 1	慶二 2	気管 4: 機関 129		5: 鍍 1	
茂夫 1	滋夫 1	京滋 2	治持 1	帰艦 1: 旗艦 1		7: 戸 10173	
はん 12種 (442)		敬治 1	携啓 1	帰還 4: 器官 7		23: 塗 6	
犯 17	判はん 3	刑事 40	掲示 4	基幹 1: 貴官 1		6: 9	
反版 66	はん 20	慶治 1		期間 76: 既刊 10		ぼう 10種 (201)	
伴 1	範畔 1	しんじ 11種 (47)		こうえん 10種 (220)		亡防 8: 某 5	
半落 275	畔班 6	信治 2	治二 1	後援 22: 好演 2		7: 棒 10	
さ 3	30	信二 10	事二 2	口演 1: 高遠 1		110: 房 12	
ひでお 12種 (83)		信二 26	慎二 1	公演 73: こうえん 5		11: 質 1	
栄男 1	日出男 2	新二 1	真二 1	公演 1: 講演 35		1: 質 36	
栄夫 3	豪雄 1	審心 1		公園 79: 幸延 1		みつお 10種 (37)	
英男 2	秀生 2	たん 11種 (131)		こうよう 10種 (40)		光生 1: 満雄 2	
英男 20	秀男 4	短炭 12	担単 1	公用 2: 紅葉 6		光男 4: 充夫 1	
英郎 1	秀夫 13	炭淡 56	探探 33	広葉 1: 昂揚 2		光夫 14: 三雄 8	
英雄 16	秀雄 18	反譚 1	探嘆 7	光洋 10: 黄葉 1		光雄 3: 2	
りょう 12種 (604)		1	嘆鍛 1	光陽 1: 甲陽 12		美津男 1: みつお 1	
漁 2	亮慶 4	たん 8		高揚 1: 効用 4		やすお 10種 (36)	
両 84	梁慶 6	ふ 11種 (634)		せいじ 10種 (267)		保男 1: 泰夫 3	
良 60	量領 1	不府 253	夫婦 10	誠司 5: 清二 1		保夫 11: 泰雄 1	
料 102	量領 71	43	符 62	清司 1: 政治 252		1: 康夫 2	
寮 268	涼 3	33	譜 4	清志 2: 精二 1		靖男 8: 康雄 4	
了 1	2	33	賦 4	清次 2: 静司 1		4: 康朗 1	
きこう 11種 (85)		198		清治 2: 静児 1		り 10種 (96)	
気工 11	奇香 1	普 17		ただし 10種 (90)		里利 2: り 3	
気候 4	寄港 4	ほう 11種 (886)		正 33: 但し 12		7: 季 3	
起紀 2	寄稿 4	法砲 232	方彭 565	正し 3: 質し 1		58: 吏 1	
機工 6	寄紀 9	7	ほう 9	3: 忠 7		1: 履 3	
機構 44	貴公 1	12	ほう 40	9: 蔽 1		1: 李 17	
機甲 1		1	豊 1				
きゅう 11種 (1768)							

1. 3. 高頻度語を含む同音語セット

「同音短単位表」において、同音セット内の合計度数の最も高いのは、「の」の32604であり、以下「に (21484)」「は (13494)」の順になる。合計度数3000以上について、度数順の表を作成すると、表2の通りである。この表からわか

るとおり、合計度数の高い同音セットは、助詞あるいは数字を含むものである。この二つの度数が高いために、同音セット全体の度数が高くなっていくのである。しかし、いかに頻度が高くても、助詞の場合には、文脈（話線）において出現する位置が、他と異なるため、たとえ同音語が存在しても、ほとんど問題にならない。それに対して、数字の場合には、「三列／参列」「五本／御本」のように、同音語を作り出していく可能性がある。

表2 同音短単位表（合計度数順の表）

() 内の数値は合計度数

の (32604)	戸時 6	市 41	架 2
之野 6	時杜 6	菅 1	荷 2
の 5	鏡 2	一 5576	禍 1
乃 32590	塗 1	いち 21	
	10173		れい (3279)
	9	さん (4906)	例 82
に (21484)	し (9682)	山 46	令 15
式 1	士 136	参 1	〇 3078
二 4435	子 7	産 439	麗 1
に似 17040	氏 450	産 34	零 29
煮 1	支 4	算 2	冷 21
	市 604	三 3538	礼 50
は (13494)	死 77	さん 846	靈 3
刃 1	糸 49		
派 135	司 1	ご (3982)	まん (3247)
葉 9	私 10	後 156	万 3237
は 13324	詞 1	語 125	満 10
波 1	使 1	御 96	
鹵 24	試 4	呉 5	
	始 1	五 3118	
た (11629)	指 1	ご 473	
他 1026	史 64	基 7	
多 29	紙 173	伍 1	
田 2	資 4	誤 1	
誰 2	至 5		
た 10570	四 1244	か (3575)	
	し 6634	下 35	
が (11492)	飼 11	科 160	
画 10	詩 44	化 636	
が 11482	賜 3	課 217	
	師 96	加 2	
て (11435)	誌 43	可 756	
手 230	視 13	家 233	
て 11205	刺 2	火 88	
		価 12	
と (10752)	で (8564)	過 7	
都 520	出 213	歌 12	
土 5	で 8351	日 78	
途 7		蚊 3	
徒 23	いち (5661)	か 1321	
	位置 22	菓 10	

1. 4. 報告20「同音語集」との比較

同音語については、国立国語研究所第1資料研究室が、昭和33年から3年間にわたって調査研究を実施し、その研究成果は、筆者も加わって、昭和36年に、国立国語研究所報告20「同音語の研究」として、まとめた。そのさい、この研究の資料とした、約10000セット（48000語）の中から、現代でも使われる可能性のあるもの7803セット（約25000語）を選び出して、「同音語表」を作成し、巻末に掲げた。

この国研報告20の「同音語表」において、15種以上の語をもつセットを、とりあげ、今回の「同音短単位表」の同音セットと比べてみると、表3の通りである。両者に共通のものは別として、「同音語表」だけに見られるものは、現代語としては、やや古めかしいものが多い。また、「同音短単位表」にのみ現われるものの中には、固有名詞や略称・略語の類が目立つ。

なお、表3は、国研報告20「同音語表」について、セットとなる同音語の種類（数）の多いものから少ないものに、順次示したものである。

表3 「同音短単位表」と「同音語集（報告20）」との比較表

<同>は同音語集（報告20）の見出し語数。
 <新>は同音短単位表（報告3）の見出し短単位数。
 () で囲んだ見出しは、同音語集のみに出ているもの。
 * 印の見出しは、同音短単位表のみに出ているもの。

コウショウ	(降将)	紀行	9	(紀綱)	(好士)
<同> 28種	(工廠)	寄稿	2	(希覲)	(厚志)
<新> 7種	(工匠)	寄稿	4	(亀甲)	(公司)
公称	(工商)	貴公	1	(氣孔)	(考試)
交渉	(公娼)	機構	44	(氣硬)	(口試)
考証	(好尚)	機甲	1	* 氣工	11 (高師)
高商	(高承)	機工	6	* 奇肴	1 (嚙矢)
厚相	(咬傷)	(寄航)			
(公傷)	(哄笑)	(婦港)		コウシ	(皓齒)
(公証)	(高翔)	(婦航)		<同> 27種	(公示)
(行賞)	(鉞床)	(婦校)		<新> 8種	(甲子)
(後章)	(硬焼)	(騎行)		公使	4 (鉞滓)
(後証)	* 江商	(奇行)	2	公子	1 (後視)
(口証)	* 鴻省	(奇効)	1	講師	42 (後し)
(高小)		(奇功)		公私	4 (口し)
(口誦)	キコウ	(起稿)		行使	12 (厚糸)
(口承)	<同> 27種	(季候)		(後嗣)	(光子)
(高唱)	<新> 11種	(貴校)		(皇嗣)	(格子)
(交唱)	寄港	(婦降)	4	(孝子)	* 交支
	起工	(婦耕)	2	(高士)	* 抗し
					1
					2

同形語の問題は、徒来は、おもに漢字のよみ方の違いとして、たとえば「ニホンかニッポンか」「ハクシかハカセか」というようにとらえられてきた。「工場 コウバ/コウジョウ」「初産 ウイザン/ジョザン」「間食 カンシヨク/アイダグイ」あるいは「女性 ジョセイ/ニョジョウ」なども同様な例であろう。しかし、「大家 オオヤ/タイカ/タイケ」「人気 ニンキ/シンキ/ヒトケ」「心中 シンチュウ/シンジュウ」「工夫 クフウ/コウフ」「最中 モナカ/サイチュウ」などとすると、意味上のへだたりが大きいので、単に漢字のよみ方の違いとして扱うわけにはいかなくなる。こうした類の同形語は、電子計算機を使って、漢字表記のままの形で日本語データを処理するさい、大きな障害となってくる。これは、なにも名詞ばかりでなく、「得る ウル/エル」「通った トオッタ/カヨッタ」「行って イッテ/オコナッテ」あるいは「米 ベイ/コメ/マイ/メートル」「上京 カミギョウ/ジョウキョウ」のようなものも出てくる。しかし、現在のところでは、同形語については、見るべき資料もなく、研究も、ほとんど行なわれていないようである。

そこで、現在、進行中の新聞語彙調査のデータを使って、同形語の資料を得るとともに、若干の調査を試みた。

なお、言語情報処理関係の論文では、「同表記語」と呼んでいるものが見られるが、「同音語」に対応して、この問題をとらえる意味から、この稿では「同形語」と呼ぶことにした。

2. 1. 同形短単位表の性格

「電子計算機による語彙調査(Ⅱ)」の「同形短単位表」は、はじめにも述べたように、一紙一年分相当量の新聞語彙調査のデータから、見出しが同一で、ヨミガナ欄の表記の異なるものを、すべて抜き出して作成したものである。したがって、「工夫 こうふ(7)/くふう(2)」「入会 にゅうかい(9)/いりあい(8)」のような、同形で異義の語ばかりでなく、「牧場 まきば(4)/ぼくじょう(17)」のような同形類義語とでも呼ぶべきものや、「各国 かくこく(42)/かっこく(23)」「勇者 ゆうしゃ(7)/ゆうじゃ(1)」など、音の交替や連濁によって生じたものも、すべて掲げられてい

る。また、ヨミガナつけの作業は、長単位の処理段階で行ない、同形短単位表は、短単位に分割したものを見出しとしているために、「時計 とけい (34) /どけい (9)」「舟 ふね (2) /ふな (2)」のようなセットも現われている。

この表にあがっている同形短単位を、分類してみると、¹ほぼ、つぎのようなことになるろう。

①同形異義語とみられるもの

(例) 末期 まっき (4) /まつご (2) 声明 せいめい (52) /しようみょう (1) 取得 とりえ (3) /しゅとく (2) 質 しつ (22) /しち (1) 根本 ねもと (1) /こんぽん (26) 瓦 かわら (3) /ぐらむ (7) 立 りつ (24) /りっとる (3) 角 かく (24) /かど (4) /すみ (1) /つの (71)

②活用形において同形となるもの

(例) 入り いり (38) /はいり (27) 出入り でいり (3) /ではいり (1) 守り まもり (3) /もり (1) 下り くだり (4) /おり (9) 来 き (6) /こ (1)

③音韻上の変化現象によって同形となるもの

(例) 三つ みつ (1) /みっつ (39) 胸 むね (36) /むな (1) 共存 きょうそん (5) /きょうぞん (7) 各界 かくかい (2) /かっかい (10) 価額 かかく (3) /かがく (2) 愛想 あいそ (1) /あいそう (1) 文字 もじ (32) /もんじ (3) 茨城 いばらき (1) /いばらぎ (44) 精進 しょうじん (4) /しょうじ (1)

④同音類義のセット

(例) 零 れい (29) /ゼロ (3) 得る うる (6) /える (7) 世論 せろん (25) /よろん (1) 昨日 きのう (3) /きくじつ (16) 月末 つきずえ (2) /げつまつ (4)

これらのうち③④の類には、徒来「語形のユレ」としてとらえられてきたよ

うなものが、数多くみられる。

2. 2. 同形短単位表の同形語セット

同形短単位表を見わたしてみても、もっとも例が多いのは、「商売 しょうばい (13) / あきない (1)」のように、音読み語と訓読み語との組合せである。「身体 からだ (1) / しんたい (27)」のような熟字訓・あて字の類まで含めると、この種のものが、大半を占めている。

音読み同士のセットでは、やはり「変化 へんか (52) / へんげ (7)」のように、漢音と呉音の対立・対応が目立つ。

同形語としては、やや特殊なものかもしれないが、固有名詞を含むものが、数の上では、きわめて多い。「博士 ひろし (2) / はくし (24) / はかせ (16)」「上京 かみぎょう (1) / じょうきょう (13)」などが、その例である。

そのほかでは、活用語の活用形を含む同形セットが、みられる点も注目すべきであろう。同形セットを生み出す活用形としては、連用形が、やはり圧倒的に多い。

このような観点から、同形語の特徴的な例を示すと、つぎの通りである。

①音読みと訓読みとのセット

(例) 市場 いちば (2) / しじょう (108) 上手 うわて (7) / じょうず (7) 高潮 たかしお (1) / こうちょう (1) 工場 こうば (7) / こうじょう (166) 水浴 すいよく (1) / みずあび (1) 尼寺 あまでら (1) / にじ (1) 翼 よく (7) / つばさ (1) 訳 やく (32) / わけ (49) 大人 おとな (6) / たいじん (7) 気質 きしつ (1) / かたぎ (1) 白髪 はくはつ (1) / しらが (1) 梅雨 つゆ (2) / ばいう (3) 夫婦 ふうふ (50) / めおと (3) 今日 こんにち (90) / きょう (5) 永遠 えいえん (5) / とわ (2)

②音読み同士のセット

(例) 言語 げんご (8) / ごと (2) 女体 じょたい (1) / に

よたい (1) 来日 らいじつ (1) / らいにち (56) 施工
しこう (18) / せこう (1) 工夫 こうふ (7) / くふう (2)
便 べん (29) / びん (29) 歩 ふ (33) / ほ (3) 性 せい
い (217) / しょう (2)

③訓読み同士の同形セット

(例) 紅 べに (5) / くない (8) 行く いく (59) / ゆく (1)
柄 え (6) / がら (3) 灯 ともしび (3) / ひ (1) 何
処 どこ (3) / いずこ (1) 生き (2) / いき (1) / なま
(6) 末 すえ (46) / まつ (32) 私 わたくし (371) /
わたし (12)

④固有名詞を含むセット

(例) 国立 くにたち (6) / こくりつ (59) 田畑 たばた (2) /
たはた (3) 小豆 あずき (10) / しょうど (1) 大山 お
おやま (32) / だいせん (33) 大木 おおき (19) / たいぼく
(4) 平家 へいけ (8) / ひらや (1) 利子 りし (14)
/ としこ (1)

2. 3. セットとなる語の種類が多い同形語

「同音短単位表」を使って、見出しの種類(数)の多い順に、順位表を作ってみると、表4のようになる。

この表の上位の方で、もっとも目につくのは、一字漢字からなる同形セットである。一字漢語について成立する同形セットにおいては、「本ほん(200) / もと(1)」「角かく(24) / かど(4) / すみ(1)」のように一語単位のものも、もちろんあるが、接頭・接尾的な用法や助数詞的な用法をもつものが、大部分である。接頭・接尾的な用法や助数詞的なものでは、たとえば「立りつ(24) / りっ(19) / りっとる(3)」となっていたとしても、その用法は、「(東京都)立 / 立(候補) / (1. 8)立」のように、かなり、はっきり分離してくる。また「歩ふ(33) / ぶ(4) / ほ(903) / ぼ(29)」においても、将棋の「歩」, 「五反歩」の「歩」, 「一步二歩」の「歩」というよ

家	4種(424)	こじろ	3	ぶん	153	楽	(45)
いえ	144	上	(921)	ぶん	1172	がく	25
かけ	233	う	547	平	(18)	たのしみ	6
か	21	え	1	たいら	9	たらく	14
や	26	う	4	たいらの	1	掛	(3)
角	(100)	か	369	だいら	7	かけ	1
か	24	じょう		ひら	1	がかり	1
か	4	真	(33)	歩	(969)	がけ	1
か	1	しん	22	ふ	33	笠	(5)
か	1	ま	8	ぶ	4	かさ	1
つ	71	ま	2	ほ	903	かり	3
後	(407)	ま	1	ば	29	ゆう	5
あ	247	生	(255)	方	(980)	球	(41)
と	3	い	1	か	263	き	32
こ	156	き	2	か	20	たま	8
の	1	せ	246	た	565	だ	1
ち		な	6	ほう	132	求	(75)
広	(84)	石	(125)	ほう		き	73
こ	3	い	32	本	(307)	ゆ	1
ひろ	2	く	1	ほん	200	と	1
い	75	こ	1	ばん	41	と	1
ひろ	4	せ	91	ぼ	65	強	(12)
し		き		も	1	き	8
行	(33)	船	(137)	と		つ	2
い	1	せん	117	無	(346)	つ	2
き	11	ふ	1	む	125	よ	
ょう	8	な	18	ない	1	し	
き	13	ふ	1	なし	218	橋	(270)
ゆ		ね		な	2	は	42
好	(22)	中	(986)	ぶ		し	224
こ	17	じ	38	衣	3種(25)	ば	4
す	2	ゅう	533	い	14	き	
ず	1	ち	1	こ	9	眺	(8)
よ	2	ゅう		ろ	2	あ	4
し		な	414	も		あ	2
幸	(13)	直	(23)	も		き	2
こ	1	ただ	3	位	(319)	ょう	
う	3	し	16	い	91	玉	(24)
ち	4	く	2	く	220	ぎ	18
あ	5	よ	2	ら	8	た	1
わ		っ	2	い		だ	5
せ		し		一人	(152)	ま	
み		な		い	7	勤	(41)
ゆ		お		ち	1	き	39
き		し		に	144	ん	1
際	(79)	ら		ん		と	1
き	58	頭	(264)	と		め	1
わ	15	あ	233	り		め	1
い	5	た	1	印	(73)	つ	1
い	1	か	1	い	36	近	(83)
時	(2130)	が	29	ん	36	き	81
じ	1965	と	2	し	8	ん	1
と	6	とう	5	る	29	か	1
と	156	等	(194)	し		か	1
と	3	ひと	5	る		し	1
き		と	102	二	(8)	九	(1261)
七十	(83)	ら	5	な	3	き	304
し	24	など	82	わ	1	ゆう	956
じ	2	二	(4463)	わ	4	く	1
じ	2	じ	4	音	(41)	こ	1
じ	54	に	4435	お	29	君	(105)
な	3	に	6	ん	10	き	34
な		い	6	ね	2	み	3
所	(579)	ふ	18	河内	(6)	ぎ	68
し	291	入	(40)	か	1	くん	
じ	270	い	10	わ	3	形	(107)
と	12	り	4	か	2	か	84
と	6	れ	25	わ	2	た	13
と		に	1	ち		け	10
小	(8)	は		学	(130)	い	
お	1	い		が	124	型	(121)
き	2	り	(2030)	が	3	か	21
し	2	ふ	634	っ	3	た	98
こ	2	ん	71	ま	3	が	
じ		ぶ		な		た	
ょう		ぶ		ぶ		た	

けい	2	じ	79	にん	1034	からだ	20
係	(268)	てら	8	心	(107)	たい	28
かかり	37	でら	5	ごころ	1	てい	1
がかり	230	四	(2253)	こころ	92	男	(297)
けい	1	し	1244	しん	14	だん	1
建	(780)	しよ	252	神	(35)	なん	12
けん	744	よん	752	かみ	31	おとこ	284
たて	1	四十	(396)	がみ	1	値	(39)
だて	35	しじゅう	18	しん	3	あたい	5
限	(62)	よんじゅう	345	神野	(13)	ち	12
かぎり	52	よんじゅう	33	かみの	1	ね	22
かぎる	1	実	(43)	かみや	11	着	(68)
げん	9	じつ	23	こうの	1	き	11
湖	(79)	じっ	1	星	(43)	ぎ	12
うみ	2	みのる	19	せい	3	ちゃく	45
こ	71	守	(13)	ほし	38	鳥	(33)
みずうみ	6	まもり	3	ほし	2	ちよう	1
口	(43)	りる	9	正	(55)	とどり	27
くち	56	もり	1	しょう	8	殿	(12)
ちち	52	取	(53)	せい	14	でんの	3
こう	6	おさむ	1	ただし	33	とどの	6
向	(43)	おさめる	51	声	(104)	どの	3
こう	18	しゅう	1	こえ	102	渡	(9)
むき	22	重	(93)	ごえ	1	わたし	5
むけ	3	おも	2	せい	1	わたる	3
高	(360)	しげ	3	政治	(254)	わたる	1
こう	264	じゅう	88	せいじ	252	当	(27)
たか	4	出	(216)	まじ	1	あたり	5
だか	92	しゅう	2	まきはる	1	あてう	1
紅	(14)	で	213	盛り	(4)	とう	21
くれない	8	でる	1	さかり	2	島	(103)
こう	1	小	(226)	ざかり	1	しま	44
べに	5	おこ	1	もり	1	じまう	35
国	(371)	しょう	16	切	(37)	とう	24
くに	205	昭	(44)	きれ	34	灯	(32)
こく	151	あき	3	せつ	2	ともしび	28
ごく	15	しょう	2	素	(3)	ひ	3
山	(140)	あしょう	39	す	1	刀	(9)
さん	46	譲	(27)	そもと	1	かたな	3
ざん	15	じょう	1	組	(183)	がたな	1
やま	79	ゆずり	25	くみ	38	とう	5
三	(3550)	ゆずる	1	ぐみ	143	年	(2286)
さん	3538	色	(141)	そ	2	とし	717
みみ	7	いろ	60	蔵	(5)	し	1
みつ	5	しき	5	くら	3	どし	1568
私	(393)	しよく	76	ぐらう	1	ねん	(27)
し	10	信	(67)	ぞう	1	波	(27)
わたくし	371	のぶ	1	相	(66)	なみ	22
わたくし	12	まこと	2	あい	5	は	1
紙	(245)	しん	64	しょう	55	ば	4
かみ	36	進	(22)	そう	6	杯	(42)
がみ	36	しん	5	造	(48)	はい	21
し	173	すすみ	1	いたる	6	はい	4
事	(53)	すすむ	16	ぞう	37	ばい	17
こと	24	人	(1616)	づくり	5	敗	(102)
ごと	21	ひと	371	体	(49)	ばい	40
じ	8	じん	211			はい	61
寺	(92)						

ばい	1	品	(201)	きた	213	ば	1
博士	(42)	しな	17	きたの	1	よう	5
はしく	24	ひん	178	はく	1	来	(124)
ひろし	2	びん	6	本町	(42)	き	6
ひかせ	16	仏	(38)	ほんちよう	39	こ	1
八幡	(89)	ふつ	13	ほんまち	1	らい	117
はちまん	9	ほとけ	23	もとまち	2	力	(276)
やはた	4	ぼとけ	2	満	(23)	ちから	92
やわた	76	文	(30)	まん	10	つとむ	1
版	(70)	おみ	2	みちる	10	りょく	183
はん	10	もん	27	みつる	3	林	(103)
ばん	54	粉	(73)	夜	(303)	はやし	74
ばん	6	こな	17	や	35	ばやし	4
美	(49)	こ	8	よる	1	りん	25
うつくし	1	ふん	48	薬	(170)	話	(215)
び	47	編	(55)	くすり	17	はなし	205
よし	1	あみ	1	ぐすり	1	ばなし	5
匹	(31)	へん	48	やく	152	わ	9
ひき	6	べん	6	裕	(5)	囊	(4)
びき	8	かた	2	ひろ	1	じょう	2
びき	17	片	(4)	ひろし	3	のぼる	1
百	(387)	へん	1	ゆたか	1	ゆずる	1
ひゃく	377	べん	1	有	(568)		
ひゃく	8	かた	2	あり	5		
びゃく	2	豊	(17)	ある	546		
票	(62)	とよ	1	ゆう	17		
ひょう	51	ほう	1	余	(64)		
びょう	1	ゆたか	15	あま	12		
びょう	10	邦	(6)	まり	9		
俵	(11)	くに	2	あま	43		
たわら	1	ほう	1	よ	15		
びょう	2	ほう	3	葉	(15)		
びょう	8	北	(215)	は	9		

2. 4. 高頻度語を含む同形語セット

「同形短単位表」において、同形セット内の合計度数の最も高いのは「一」の6637であり、以下「二(4463)」「日(3910)」の順になる。合計度数500以上のセットについて度数順表を作ると表5の通りである。

この表からわかる通り、合計度数の高い同形セットは、漢数字と、やはり一字漢字の語が多い。一字漢字のものについては、前節で述べたが、表4と同様に、その大部分は、漢字のヨミの違うもので占められ、同形語として興味をひかれるようなセットは、ほとんど出てこない。このことは漢数字についても同様である。

同音語の場合には、種類が多いセット(表1)や、頻度の高いセット(表2)において、すでに、同音異義語として、しばしば取りあげられるようなも

のが、現われて来たが、同形語では、そうした類のものは、表4・表5のいずれにも、ほとんど現われない。このことは、同形異義語として、その判別に苦しむようなセットは、ほとんど二語の組み合わせのものに限られ、頻度も、それほど高くないことを示している。

表5 同形短単位表 (合計度数505以上の同形群)

() 内の数値は合計度数

一 (6637)	いち	5576	じゅっ	183	九 (1261)	きゅう	304	うえ	547	
	いっ	945		分 (2030)			く	956	うわ	1
	いつ	1	ぶん	634		このつ	1	かみ	4	
	かず	13	ぶ	71	他 (1044)			建 (780)		
	はじめ	12	ぶん	153		た	1026	けん	744	
	ひと	90	ぶん	1172	ほか	18	たて	1		
二 (4463)	じ	4	六 (1827)	ろく	1692	町 (1025)	ちょう	736	大 (741)	
	に	4435		ろっ	135		まち	289	だい	680
	にい	6	月 (1766)	がつ	1	中 (986)	う	38	お	54
	ふた	18		がつ	1588		じゅう	38	お	2
日 (3910)	か	78	げつ	2	ちゅう	533	お	1	お	4
	じつ	7	つき	174	ちゅう	1	お	1	お	4
	にち	3026	づき	1	なか	414	お	1	お	4
	にっ	136	者 (1713)	しゃ	1617	二十 (982)	にじゅう	912	付 (709)	
	び	356		もの	96		にじゅっ	70	ずき	5
	び	305					つき	491	つき	2
三 (3550)	さん	3538	人 (1616)	ひと	371	方 (980)	かた	263	つき	13
	み	7		じん	211		がた	20	ほう	2
	みつ	5	にん	1034	ほう	565	ほう	132	つけ	198
年 (2286)	とし	717	日本 (1610)	にっぽん	626	前 (975)	ぜん	101	新 (663)	
	どし	1		にほん	984		まえ	874	あたらし	6
	ねん	1568						しん	657	
四 (2253)	し	1244	七 (1497)	しち	1311	歩 (969)	ふ	33	工 (654)	
	しよ	252		なな	186		ふ	4	こう	639
	よん	757				ぼ	903	たくみ	15	
時 (2130)	じ	1965	八 (1489)	はち	1411	名 (939)	ない	47	市 (645)	
	と	6		はっ	78		めい	892	いち	41
	とき	156	的 (1299)	てき	1291	上 (921)			し	604
	どき	3		まと	8				長 (619)	
十 (2063)	じゅう	1880						ちょう	618	
								なが	1	
								会社 (618)		
								かいしゃ	162	
								がいしゃ	456	

所 (579)		三十 (565)		車 (551)		都 (537)	
しよ	291	さんじゅう	487	くるま	187	と	520
じよ	270	さんじゅう	78	しゃ	364	みやこ	17
と	12						
ところ	6						
有 (568)		店 (564)		面 (542)		通 (513)	
あり	5	てん	387	づら	1	つう	442
あ	546	みせ	177	めん	541	どおり	71
ゆ	17						