

# 琉球大学学術リポジトリ

## インターネット上の出会い系掲示板を対象とした基礎的語彙調査

メタデータ	言語: ja 出版者: 琉球大学法文学部 公開日: 2010-06-08 キーワード (Ja): キーワード (En): 作成者: 金城, 克哉, Kinjo, Katsuya メールアドレス: 所属:
URL	<a href="https://doi.org/10.24564/0002005542">https://doi.org/10.24564/0002005542</a>

## インターネット上の出会い系掲示板を対象とした 基礎的語彙調査

金城 克哉

### 1. はじめに

これまで国立国語研究所によって大規模な語彙調査が行われてきた。2009年10月現在、同研究所は言語データベース整備計画（KOTONOHHA 計画）の一環として『現代日本語書き言葉均衡コーパス』構築を進めている（同プロジェクトの詳細については『国文学：解釈と鑑賞』2009年1月号所収の諸論文を参照）。しかしながら荻野（2002）が指摘するようにコンピューターの普及に伴い、従来ではなし得なかった大規模コーパスを用いた調査が可能になったとは言うものの、個人研究者のレベルでの様々な分野の「語彙調査」はそれほどなされているとは言えない状況にある（もちろん、杉本（2009）のように大規模コーパスを用いた意味分析や語の用法の分析などの研究は盛んに行われている）。

その一方で、インターネットの普及により、個人が（編集者などの校閲を経ることなく）自ら思うところをブログで公表したり日記という形で仲間内で共有したり（mixiなど）、また Yahoo 知恵袋（<http://chiebukuro.yahoo.co.jp/>）などのオープンな掲示板へ書き込んだりすることも可能となった。この後者の「掲示板」には様々な形態のものがあり、上記 Yahoo 知恵袋のように誰もが自由に携帯電話やパソコンなどを用いてアクセスできるものもあれば、特定の関心事を持つ人々のみがアクセスするようなもの（オープンではあるが、一般の人々にはあまり知られていない）も存在する。これらを広義の「言語活動」と捉え、そこでどのような言葉が用いられているのか、そこにはどのような語彙が現われるのかを調査することは、共時的な言語分析という視点から意義のあることであると思われる。

本稿ではそのような特定の関心を持つ人々がアクセスし書き込みをするという点に着目し、MSM を対象とした出会い系掲示板の語彙調査を実施した。なお、本稿では基礎的な「語彙の量的性格(特性)」の記述を中心とし、「語彙の量的構造」(伊藤 2002)の解明にまで言及することができなかつた。以下調査概要、調査結果、分析・考察の順に記述していく。(注1、注2)

## 2. 調査概要

### 2.1. 調査目的

コーパスを用いた特定分野 (MSM を対象とした掲示板) の語彙を調査することによって高頻度語および特徴的な語彙を明らかにすること。

### 2.2. 調査方法

今回調査対象としたインターネット上の掲示板はA～Fの計6つである。2008年11月1日～2009年4月30日までの半年間の全投稿のうち、過去ログとしてインターネット上で参照できるものを利用し、1回の投稿を1件としてカウントした。総投稿件数は42,560であった。それぞれの掲示板の月別の投稿件数の内訳は以下の通り：

表1 各掲示板の月別投稿件数と合計数

	11月	12月	1月	2月	3月	4月	合計
掲示板A	3,753	3,335	3,528	3,333	3,887	3,639	21,475
掲示板B	2,153	2,132	2,020	2,150	2,578	2,822	13,855
掲示板C	195	154	138	140	123	121	871
掲示板D	241	227	234	166	197	204	1,269
掲示板E	243	236	264	291	344	328	1,706
掲示板F	596	605	497	506	556	624	3,384

サンプリングのために投稿件数の最も少ない「掲示板C」を基準とし、それぞれの掲示板の投稿件数との比を計算した (小数点1位は四捨五入)。その結果、表2に示す割合が得られた。この比率が反映されるよう、それぞれの掲示

板について11月から4月までの投稿の中から総標本数が2,400（母集団の約5.64%）となるよう標本を採取した。標本抽出にあたってはコンピュータ上で発生させた乱数を用い、無作為抽出法をとった。

表2 掲示板Cを基準とした投稿数の割合と標本採取数

	A	B	C	D	E	F
Cを1とした割合	25	15	1	1	2	4
標本採取数	1,250	750	50	50	100	200

今回の投稿文の解析にあたってはKH Coderを用いた（解析器は茶筌、解析器システム用電子化辞書はUniDic）。（注3）

データ収集から解析までの手順は以下のとおりである。まず、各掲示板の過去ログを参照し、投稿1,000件毎にエディタに記録し基本ファイルを作成。それをもとに乱数にしたがって1件ごとにExcelに入力した。投稿文は氏名、タイトル、本文、プロフィール（身長・体重・年齢）から構成されているが、今回はタイトルと本文のみを扱った。また1件ごとに20代や30代というタグを付した。これをKH Coderにかけ、UniDicに登録されておらず解析がうまくいかない語（未知語）については「強制抽出語」として登録した（樋口 2009）。この作業を繰り返し、ある程度目処がついたところで一つ一つの見出し語について誤解析がなされていないかを検討し、最終的にはExcelの機能などを活用しながら手作業で数え上げた。

### 2.3. 調査単位の認定

単位（語）認定の基準を統一するため、UniDicに登録された単位を基準とした。動詞では可能形を1つの見出し語として立てた（例：「会う」と「会える」は1つの見出し語「会う」として統一するのではなく、別々の見出し語とした）。UniDicに登録されていない単位の認定については以下の点に留意した。

- ① 同一語と認定できる場合でも、漢字・ひらがな・カタカナ・記号の有無で表記が異なるものについては異なる見出し語とした（例：「良い」の意味での「いい」は「いい」の他にデータ上では「いい」、「いいー」、「イイ」、

- 「イィ」の表記があるが、それぞれを別の見出し語とした。また、頻出の「暇」、「ヒマ」、「ひま」も別の見出し語とした。
- ② 略語については1つの見出し語とした(例:「格好いい」は「格好」と「いい」という別々の見出し語を立てたが、今回のデータでは出現形として「かっけー」、「カッケー」、「かっけえ」の略語表記があり、それぞれを別の見出し語とした)
  - ③ 頻出語である「場所あり」の「あり」は他にも「アリ」、「有り」、「有」の表記があるが、それぞれを別の見出し語とした。またこれは動詞の連用形ではなく、名詞とした。
  - ④ 時の名詞「今、夜、明日」などは副詞的な用法であると考えられるものも、品詞分類上は名詞とした (cf. 『日本語教育事典』 p.140)。
  - ⑤ 今回の調査で特徴的な語であると考えられる「ガチムチ」、「セクフレ」、「ガン堀り」などの複合語については、例えば「ガチムチ」を「ガッチリ」と「ムッチリ」に分けて見出しを立てるのではなく、1つの見出し語とし、名詞に分類した。
  - ⑥ 伏字として記号の「○」が使用されている場合でも、それが前後の関係から語の一部であると認められる場合には、1つの見出し語とした(例:「中○生」は「中」「○」「生」に分解するのではなく、「中○生」を1つの見出し語とする)
  - ⑦ アルファベットについては、「H」(全角)と「H」(半角)が性交為の意味での「エッチ」の異表記である場合は1つの見出し語とした。また、「tough」などの外国語(アルファベット表記)については、登録語のみを採用し、その他はそれぞれのアルファベットを1つの見出し語とした。
  - ⑧ 顔文字と呼ばれる記号の組み合わせで作るエモーティコン(例:(^\_^)v)などは分解して分析し、全て独立した記号としてそれぞれを1つの見出し語とした。(注4)
  - ⑨ 音引き(長音)の記号「ー」の代替表現としてマイナス記号「-」やハイフン「-」が用いられているものについては、まとめて1つの見出し語とした(例:「メール」、「メール」、「メ-ル」は「メール」にまとめた。

ただし、省略形の「メル」については別の見出し語とした)。

### 3. 調査結果

#### 3.1. 延べ語数と異なり語数

調査の結果、延べ語数は77,843、異なり語数は3,548であった。以下の表3で各品詞ごとの述べ語数と異なり語数を示す(注5)：

表3 各品詞ごとの延べ語数と異なり語数

	形(非)	固	地	人	名	動	副	形	形動	接続
延べ語数	1,048	197	1,074	46	20,969	9,823	1,460	1,150	1,189	168
異なり語数	16	68	109	30	1,786	521	168	133	144	16

	接頭・尾	連体	助	助動	代	記	感	否	不明	合計
延べ語数	2,655	196	15,610	9,429	1,145	10,060	46	1,546	32	77,843
異なり語数	178	13	85	37	45	156	22	3	18	3,548

固有名詞・地名・人名を除いた名詞、動詞、形容詞(形(非)を含める)などの主要品詞構成の延べ語数と比率、また同じ主要品詞の異なり語数の度数と比率は以下の通りであった(小数点2位以下切り捨て)：

表4 主要品詞構成比率(延べ語数)

	名	動	形	形動	副	連体	接続	感動	合計
度数	20,969	9,823	2,198	1,189	1,460	196	168	46	36,049
比率(%)	58.17	27.24	6.09	3.29	4.05	0.54	0.46	0.12	99.95

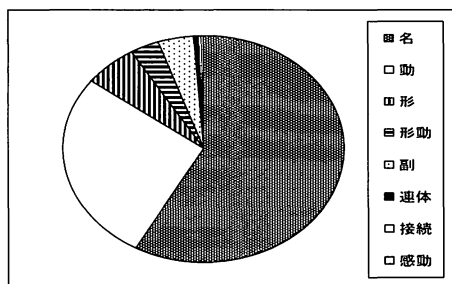


図1 延べ語数の品詞構成比率

表5 主要品詞構成比率（異なり語数）

	名	動	形	形動	副	連体	接続	感動	合計
度数	1,786	521	149	144	168	13	16	22	2,819
比率 (%)	63.35	18.48	5.28	5.1	5.95	0.46	0.56	0.78	99.96

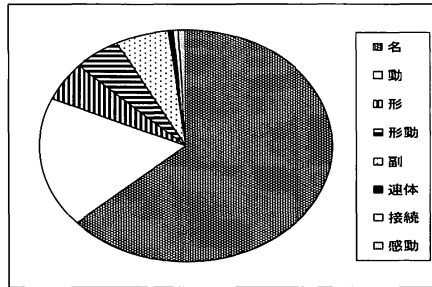


図2 異なり語数の品詞構成比率

これら主要品詞の度数順語彙については付表1～5を参照のこと（紙面の都合上150位まで。なお連体詞、接続詞、感動詞は省略）。

### 3.2. 母比率の推定

名詞・動詞・形容詞の95%信頼区間での母比率の推定値を次の公式によって求める（伊藤 2001）：

$$p \pm Z\alpha \sqrt{\frac{p(1-p)}{n}}$$

（ $p$ =使用率、 $n$ =標本数、 $Z\alpha$ =正規分布表の係数1.96）

- (1) 名詞（使用率  $p=20,969$ （名詞の延べ語数） $\div 77,843$ （総延べ語数） $= 0.269$ ）

母集団の名詞の比率は、信頼度95%で25.40%から28.39%の間であると推定できる

- (2) 動詞（使用率  $p=9,823 \div 77,843 = 0.126$ ）

母集団の動詞の比率は、信頼度95%で11.48%から13.71%の間であると推定できる

(3) 形容詞 (使用率 $p=2,198 \div 77,843 = 0.028$ )

母集団の形容詞の比率は、信頼度95%で2.24%から3.35%の間であると推定できる

## 4. 分析と考察

### 4.1. 樺島の法則 (1955)

樺島は「条件を一定にした文において、名詞の百分率が分かれば、他の類別された品詞の百分率が算出でき」(1955:386)、「名詞の百分率をもって、文章の特性を計る尺度となし得る」(同上)とし、次の3つの式を提案する(自立語のみ)

$$(1) Ad = 45.67 - 0.60N$$

$$(2) \log_{10} I = 11.57 - 6.56 \log_{10} N$$

$$(3) V = 100 - (N + Ad + I) \quad (\text{注6})$$

(但し  $N$  は名詞の構成比百分率、 $V$  は動詞の構成比百分率、 $Ad$  は形容詞・形容動詞・連体詞・副詞の合併類の構成比百分率、 $I$  は接続詞・感動詞の構成比百分率とする)

今回の調査では  $N = 58.17$  であったが、これを上記式に代入すると、

$$Ad = 45.67 - 0.60 \times 58.17 = 10.768$$

$$\log_{10} I = 11.57 - 6.56 \log_{10} 58.17 = -0.00643$$

という値が得られる。これは今回の調査で得られた値  $Ad$  (13.97)、 $\log_{10} I$  (-0.23657) とは異なり(水谷 (1977) の式でもほぼ同じ結果となる)、また水谷 (1977:66 図5) で示された『日本文学大辞典』の語類の割合に近い値となっている。

樺島は「Nの増加は話しことば的なものから書きことばへと向かって」おり(1955:387)、また「感情の表現をなすものから関係の表現をなすものへと、Nが増す」(同上)と指摘しているが、これは樺島が最も名詞の比率が高いもの



として新聞記事を念頭に置いているためである。今回の調査では掲示板への投稿者の意向を表わす助動詞「たい（てえ）」（例：「盛りたい」（データNo.2））の延べ語数は1,138、また投稿文の読み手を意識した丁寧体「です」（延べ語数1,108）・「ます」（延べ語数2,036）も用いられており、名詞の多用がすなわち新聞調の書き言葉に近づくとは一概には言えないであろう（佐竹（2009:106）では言語的特徴を抽出するための指標として文末形式の比率について触れている）。次のセクションで高頻度語について検討するが、樺島（1954）が指摘するように、「場面文脈依存度の高低が、文長の制約と共に品詞の割合を変化させる要因の一つ」であり（1954:18）、「文が言語表現以外の通達方法あるいは文脈を利用し得ず、自足的自立的であるためには、主要な概念項である名詞を省略し得ず、名詞の割合が大となる」（同上）という点は傾聴に値する。今回の調査で名詞の割合が大きかった要因は、「いつ」「誰が」「何を」（「どんな人」を）望んでいるのかという点に重点が置かれた記述が多かったためではないかと思われる。この点については次の高頻度語の分析で検討する（文末形式「つす」・「す」については金城（2009）を参照のこと）。

#### 4.2. 高頻度語

高頻度語は「無特徴語」（基本的な名詞（例：「こと、もの、人、時」）や基本的な動詞（例：「する、いる、ある、なる」））とテキストを特徴づける「特徴語」とからなるとされる（伊藤2001）。表6に本調査で得られた高頻度語50位までをあげる。これら50の語のうち、本稿が分析する出会い系掲示板の投稿文を特徴付ける（であろう）と考えられるものを以下のカテゴリーに分類した。

- ①時：「今」
- ②場所：「場所」「公園」
- ③役割：「タチ」「ウケ」（注7）
- ④身体：「短髪」「体型」「ケツ」「プロフ」「太め」「スリム」
- ⑤動作：「やる」「できる」「出来る」「掘る」「しゃぶる」「会える」「くれる」
- ⑥その他：「メール」「年下」「軽い」「募集」「足」「あり」「なし」

「人」や「方」があることから「人物」というカテゴリーも考えられたが、これらは無特徴語とみなし、ここでは「タチ」・「ウケ」という語が見られたため性行為の際に能動的もしくは受動的な役割をするかという役割の区分を示すものが必要であると考え、「役割」というカテゴリーを立てた。身体には身体特徴「短髪」・「太め」・「スリム」と身体部位「ケツ」、身長・体重・年齢をまとめた「プロフ（イール）」を含めた。「足」はその他に含めたが、これは「足」が「移動手段」の意味で用いられているためである。また、動作を表す語のうち、「やる」と「できる（出来る）」は性行為の隠語として上位概念を、「掘る」と「しゃぶる」はその下位区分として具体的な動作を表している。(注8)

表6 高頻度語（50位まで 数詞を除く）

順位	語	度数	順位	語	度数
1	いる	1,517	26	よろしく	203
2	する	1,511	27	好き	201
3	人	1,234	28	あり（有り）	192
4	メール	520	29	下さる	190
5	今	475	30	なし（無し）	188
6	場所	441	31	くださる	178
7	くれる	376	32	見る	178
8	掘る	329	33	出来る	177
9	いい	315	34	ケツ	175
10	タチ	310	35	会える	174
11	方（かた）	308	36	プロフ	173
12	普通	302	37	年下	164
13	やる	295	38	来る	164
14	行く	295	39	軽い	164
15	ない	292	40	太め	157
16	できる	289	41	公園	153
17	しゃぶる	286	42	居る	150
18	短髪	277	43	感じ	145
19	ある	269	44	よい	143
20	体型	251	45	良い	142
21	募集	243	46	スリム	141
22	ウケ	241	47	市	139
23	足	238	48	奴（やつ）	139
24	タイプ	218	49	希望	130
25	願い	203	50	今日	130

次に、特徴語には含めなかったが、高頻度語の1位である「いる」(42位の「居る」を含めると頻度は1,667となる)の他の語との共起関係を見てみよう。表7は「いる」と共起関係にある上位3語(「ます」・「人」・「ない」)をあげたものである。ここでは「いる」を中心に、両側の何番目に「ます」・「人」・「ない」が用いられているかがわかる。表が示すように、「いる」は助動詞「ます」と結びついて「います」の形で用いられている例が647あることがわかる。次に高頻度語である「人」も「いる」と結びつき、「人いる」のように用いられる例が510あり、また「いない」という否定の形で645例用いられている。

表7 「いる」のコロケーション統計(左右)

順位	語	品詞	左3	左2	左1	右1	右2	右3
1	ます	助動詞	7	2	1	<b>647</b>	0	10
2	人	名詞	11	24	<b>510</b>	22	8	3
3	ない	助動詞	4	10	0	<b>465</b>	1	12

また、「いる」の左側に焦点をあわせ、左1に現れる語を頻度の高い順に見ると、表8のような結果となる：

表8 「いる」のコロケーション統計(左側のみ)

順位	語	品詞	左5	左4	左3	左2	左1
1	人	名詞	12	7	11	24	<b>510</b>
2	て	助動詞	48	54	123	2	<b>126</b>
3	方	名詞	0	1	0	9	<b>88</b>
4	か	助詞	24	13	30	10	<b>84</b>
5	に	助詞	19	12	45	1	<b>79</b>
6	たち	名詞	2	2	9	4	<b>71</b>

「人いる」(例：「起きてる人いますか」データ No.21)、「方いる」(例：「欲しい方いましたらレスよろしくです」データ No.27)、「たちいる」(例：「会ってできるたちいるかな」データ No.176)など「いる」が他の頻出語とも共起関係にあることがわかる(第5位の「か」は「誰かいる」の例である)。

上記コロケーションは隣接する語にどのようなものがあるのかを示すものであるが、もう少し広い意味で「共起」を捉え、同一投稿文内(センテンスでは



## 5. まとめ

今回は個人で行う調査ということで標本数も2,400（母集団の約5.64%）とせざるを得なかったが、MSMを対象とした掲示板で用いられる語彙の一端でも示すことができたのではないかと思う。当初、データを収集する段階では高頻度語の中に性行為に関する語が多く含まれるであろうと予想していたが、分析結果はその予想に反する結果となった（各品詞ごとの高頻度語については付表を参照のこと）。

また、語彙という範疇を越えるため、分析・考察の部分では詳しく触れることができなかったが、「～人いる」という語と語の「共起」から次の段階として「表現」へ視点を移すと、肯定の「いる」ではなく、否定形の「いない」や「いません」を用いて否定疑問文にする例（「～人いない?」「～人いないかな?」「～人いませんか?」）が多用されていることも今回の調査で明らかとなった。こういった表現の特徴の分析については今後の課題としたい。また、年代別の語彙についても特徴を明らかにすることができなかったため、この点もさらに研究を進めたい。

最後になったが、今回の調査ではデータ提供を許可してくださった掲示板の管理・運営者の方と KH Coder の開発者である樋口耕一氏にこの場を借りて謝意を表したい。

### 注

1. 本研究は平成21年度厚生労働科学研究費による研究(研究代表者：加藤慶、課題番号：H20-エイズ-若手-012)の一部である。MSMはmen who have sex with menの略。
2. 今回の調査は、個人情報保護の観点から掲示板への投稿者の氏名やメールアドレス（これは基本的には掲示板では表示されず、またデータ収集の際にも調査者は知ることができない）、個人のプロフィールなどの情報を開示しないことなどを条件に管理・運営者の同意を得て実施した。また本稿が公表されることにより当該掲示板へ何らかの不利益が及ぶことを考え、今回は掲

示板の URL 開示を控える。

3. KH Coder は立命館大学の樋口耕一氏が開発したフリー・ソフトウェアである。本調査では Version 1.beta21 を用いた。これについては以下の URL より最新版が入手できる。

<http://khc.sourceforge.net/index.html>

この KH Coder は形態素解析器として茶筌を利用しているが、茶筌は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（松本研究室）によって開発されたものである。詳しくは次の URL を参照のこと。

<http://chasen-legacy.sourceforge.jp/>

また以下の URL より最新版のダウンロードが可能である。

[http://sourceforge.jp/projects/chasen-legacy/releases/?package\\_id=5864](http://sourceforge.jp/projects/chasen-legacy/releases/?package_id=5864)

今回の分析では解析器システム用電子化辞書として、IPADIC ではなく、千葉大学の伝康晴氏らが開発した UniDic を用いた（今回は unidic-chasen1312\_sjis.zip を利用）。UniDic については以下の URL を参照のこと。

[http://www.tokuteicorpus.jp/dist/modules/system/modules/menu/main.php?page\\_id=1&op=change\\_page](http://www.tokuteicorpus.jp/dist/modules/system/modules/menu/main.php?page_id=1&op=change_page)

4. UniDic にはこれら一部の顔文字が登録されているが、今回分析対象としたデータ中には登録外の顔文字も含まれており、統一を期するために構成要素となっている記号を分解して見出し語とした。
5. この品詞分類は KH Coder にもとづくものであるが（樋口 2009 : 9）、電子化辞書として上記 UniDic を用いたため一部改変してある。UniDic の品詞分類については伝他（2009 : 15）を参照のこと。
6. 水谷（1977）はこの樺島の法則を第一式と第二式を次のように修正し、近似度を高めることができるとする（括弧内に示した桁は有効数字ではないとする）：

$$(1) Ad = 44. (16) - 55. (75) N$$

$$(2) \log_{10} I = 12. (59) - 7.2 (36) \log_{10} N$$

7. 『分類語彙表』（1.1510）には「動き」の下位区分（04）として「受け太刀」という語が見られる。

8. 参考までに上位24位までの語を対象に階層的クラスター分析を行った結果を示す (Ward 法、クラスター数=5)。本文中の分類と比較されたい。

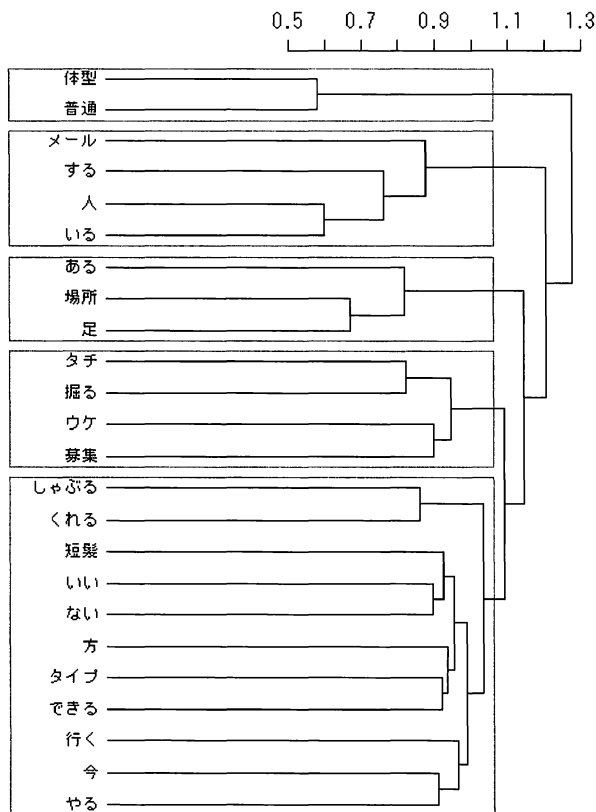


図 高頻度語24位までのクラスター分析

9. 例えば本稿の「いる」と「人」の Jaccard 係数は以下のように求められる (KH Coder のフォーラム ([http://koichi.nihon.to/cgi-bin/bbs\\_khn/khcf.cgi?no=473&mode=allread](http://koichi.nihon.to/cgi-bin/bbs_khn/khcf.cgi?no=473&mode=allread)) を参照)。Jaccard 係数については (金2009: 162) も参照のこと。

共起した投稿件数 (639) ÷ (「いる」が現れた投稿数 (1290) + 「人」が現れた投稿数 (934) - 共起した投稿数 (639)) = 0.4032

## 参考文献

- 伊藤雅光 2001 「推定と検定」『日本語学』Vol.20 4月臨時増刊号, pp.84-98
- 伊藤雅光 2002a 『計量言語学入門』大修館書店
- 伊藤雅光 2002b 「語彙の量的性格」北原保男（監修）斉藤倫明（編）『朝倉日本語講座4 語彙・意味』, pp. 29-53
- 小川芳男・林大ほか（編）1987『日本語教育辞典 縮刷版』大修館書店
- 荻野綱男 2002 「計量言語学の観点から見た語彙研究」『国語学』第53巻1号, pp.97-115
- 樺島忠夫 1954 「現代文における品詞の比率とその増減の要因について」『国語学』18号, pp.15-20
- 樺島忠夫 1955 「類別した品詞の比率に見られる規則性」『国語国文』24巻6号, pp.385-387
- 金城克哉 2009 「MSM出会い系掲示板における文末表現『っす』の用法について」『Southern Review』No. 24, pp.85-89
- 金明哲 2009 『テキストデータの統計科学入門』岩波書店
- 国立国語研究所 2004 『分類語彙表 増補改訂版』大日本図書
- 佐竹秀雄 2009 「新聞における投書の文章スタイルの計量的分析」三宅和子他（編）『メディアとことば』ひつじ書房 pp.96-127
- 杉本武 2009 「コーパスからみた類義語動詞：『ねじる』と『ひねる』」『文藝言語研究 言語篇』筑波大学文藝・言語学系 第55巻, pp.109-122
- 伝康晴 他 2009 「UniDic version 1.3.12 ユーザーズマニュアル」
- 樋口耕一 2009 「KH Coder 2.x リファレンス・マニュアル」
- 水谷修 1977 「語彙の量的構造」『岩波講座日本語9 語彙と意味』岩波書店 pp.43-86



付表1 名詞の度数順語彙表(150位まで、数詞を除く)

順位	語	度数	順位	語	度数	順位	語	度数
1	人	1234	51	年上	68	101	上	37
2	メール	520	52	抜きあい	66	102	返事	37
3	今	475	53	リバ	65	103	マッサージ	36
4	場所	441	54	秘密	65	104	以外	36
5	たち	310	55	暇	64	105	年	36
6	方(かた)	308	56	やつ(奴)	63	106	有り	36
7	普通	302	57	気持ち	63	107	夕方	36
8	短髪	277	58	無し	62	108	あたり	36
9	体型	251	59	一緒	61	109	しゃぶり合い	35
10	募集	243	60	デブ	59	110	人気(ひとけ)	35
11	ウケ	241	61	近く	59	111	あと	35
12	足	238	62	髭	58	112	経験	34
13	タイプ	218	63	辺り	58	113	前	34
14	願い	203	64	エロ	57	114	体系	34
15	あり(有り)	192	65	厳守	57	115	野外	34
16	なし(無し)	188	66	ドライブ	55	116	友達	34
17	ケツ	175	67	顔	54	117	アナル	33
18	プロフィール	173	68	画像	53	118	スポーツ	33
19	年下	164	69	頃	53	119	こと	32
20	公園	153	70	仕事	51	120	兄貴	32
21	感じ	145	71	前半	51	121	H	31
22	市	139	72	夜	50	122	ヤツ	31
23	奴(やつ)	139	73	マラ	49	123	車	31
24	希望	130	74	後半	49	124	投稿	31
25	今日	130	75	複数	48	125	オナニー	30
26	筋肉	128	76	ごめん	47	126	最近	30
27	ガチムチ	123	77	岩盤	47	127	親父	30
28	デカマラ	108	78	国際	47	128	間	29
29	笑	105	79	興味	46	129	ガン堀り	28
30	抜き合い	105	80	交換	46	130	スジ筋	28
31	自分	104	81	ノンケ	45	131	チンポ	28
32	坊主	104	82	体	45	132	ラブホ	28
33	目	104	83	野郎	44	133	帰り	28
34	一方	103	84	いま	43	134	近辺	28
35	互い	99	85	ノリ	43	135	部屋	28
36	気	94	86	見せ合い	42	136	ヒゲ	27
37	フェラ	88	87	観光	42	137	既婚	27
38	運動	87	88	湯	42	138	口	27
39	トイレ	86	89	明日	42	139	女	27
40	時間	85	90	ガタイ	41	140	全裸	27
41	今夜	82	91	サブ	41	141	年齢	27
42	アド	81	92	しゃぶりあい	40	142	勃起	27
43	ホテル	79	93	乳首	40	143	ガチポ	26
44	セーフ	78	94	ぬきあい	39	144	ちなみ	26
45	エッチ	77	95	パイ	39	145	リーマン	26
46	男	77	96	子	39	146	事	26
47	以上	76	97	触り合い	38	147	ネコ	25
48	同	74	98	後	38	148	やり取り	25
49	バリタチ	70	99	通り	38	149	休み	25
50	年代	70	100	髪	38	150	セフレ	24

付表2 動詞の度数順語彙表 (150位まで)

順位	語	度数	順位	語	度数	順位	語	度数
1	いる	1517	51	触る	24	101	来れる	8
2	する	1511	52	盛る	24	102	ぶっ放す	8
3	くれる	376	53	入れる(いれる)	23	103	集まる	8
4	掘る	329	54	責める	20	104	締まる	8
5	やる	295	55	攻める	19	105	いける	7
6	行く	295	56	あう(合う)	18	106	いじめる	7
7	できる	289	57	イケる	18	107	うづく	7
8	しゃぶる	286	58	もらう	18	108	おる	7
9	ある	269	59	しごく	17	109	こだわる	7
10	下さる	190	60	盛れる	17	110	だす	7
11	くださる	178	61	行ける	16	111	ぶっかける	7
12	見る	178	62	知る	16	112	飲める	7
13	出来る	177	63	有る	16	113	過ごす	7
14	会える	174	64	起きる	15	114	甘える	7
15	来る	164	65	離れる	15	115	向かう	7
16	居る	150	66	たまる(堪る)	14	116	構う	7
17	なる	128	67	見せる	14	117	考える	7
18	抜く	116	68	楽しむ	13	118	出る	7
19	会う	107	69	寝る	13	119	寝れる	7
20	思う	105	70	話す	13	120	酔う	7
21	待つ	104	71	いじる	12	121	分かる	7
22	みる	93	72	楽しめる	12	122	たまる(溜まる)	6
23	書く	69	73	帰る	12	123	あえる(合える)	6
24	遊ぶ	67	74	終わる	12	124	つかう	6
25	溜まる	67	75	疼く	12	125	なれる	6
26	やれる	66	76	かける	11	126	はじめる	6
27	言う	65	77	ぬく	11	127	もらえる	6
28	いく	62	78	まつ	11	128	逢う	6
29	太る	58	79	抜く	11	129	過ぎる	6
30	すむ	49	80	抱く	11	130	虐める	6
31	つける	46	81	つく(付く)	10	131	仕込む	6
32	飲む	45	82	あげる	10	132	受ける	6
33	感じる	44	83	しまう	10	133	助かる	6
34	くる(来る)	42	84	掘り込む	10	134	付き合う	6
35	鍛える	41	85	持つ	10	135	変える	6
36	付ける	41	86	住む	10	136	連れる	6
37	合う	38	87	助ける	10	137	イく	5
38	入る	38	88	送る	10	138	オナる	5
39	犯す	38	89	問う	10	139	くわえる	5
40	使う	37	90	あう(会う)	9	140	とまる	5
41	舐める	37	91	ほる	9	141	はめる	5
42	かく(書く)	34	92	まくる	9	142	ムラる	5
43	消す	33	93	逢える	9	143	もつ	5
44	泊まる	30	94	見せ合う	9	144	わかる	5
45	なさる	29	95	困る	9	145	廻す	5
46	載せる	29	96	出会う	9	146	掘り上げる	5
47	出す	27	97	眠れる	9	147	決める	5
48	教える	26	98	遊べる	9	148	見える	5
49	探す	25	99	絡む	9	149	合える	5
50	いう(言う)	24	100	あえる(会える)	9	150	似る	5

付表3 形容詞の度数順位語彙表

順位	語	度数	順位	語	度数	順位	語	度数
1	軽い	164	51	めんどい	4	101	ややこしい	1
2	悪い	109	52	高い	4	102	わるい	1
3	エロい	106	53	短い	4	103	汚い	1
4	遅い	51	54	ありがたい	3	104	危ない	1
5	若い	45	55	うまい	3	105	狭い	1
6	久しい	38	56	カッケー	3	106	堅い	1
7	毛深い	38	57	でかい	3	107	厳つい	1
8	近い	35	58	ヤバイ	3	108	固い	1
9	嬉しい	29	59	よろしい	3	109	厚い	1
10	優しい	29	60	強い	3	110	広い	1
11	激しい	27	61	仕方ない	3	111	黒い	1
12	太い	25	62	臭い	3	112	細かい	1
13	寒い	22	63	暖かい	3	113	若々しい	1
14	少ない	20	64	きつい	2	114	渋い	1
15	寂しい	18	65	キモイ	2	115	女々しい	1
16	上手い	16	66	スゲー	2	116	少ない	1
17	うれしい	15	67	すげー	2	117	情けない	1
18	詳しい	15	68	ヤバい	2	118	深い	1
19	デカイ	14	69	やばい	2	119	辛い	1
20	楽しい	12	70	暗い	2	120	大人しい	1
21	細い	12	71	古い	2	121	男くさい	1
22	でかい	11	72	酒臭い	2	122	低い	1
23	可愛い	11	73	男臭い	2	123	程よい	1
24	薄い	11	74	恥ずかしい	2	124	難しい	1
25	かわいい	10	75	白い	2	125	薄暗い	1
26	大きい	10	76	肌寒い	2	126	美味しい	1
27	早い	8	77	怖い	2	127	幅ひろい	1
28	多い	8	78	エロイ	1	128	幅広い	1
29	デカイ	7	79	おかしい	1	129	忙しい	1
30	緩い	7	80	おっきい	1	130	面白い	1
31	小さい	7	81	かっけえ	1	131	有り難い	1
32	かるい	6	82	かる〜く	1	132	頼りない	1
33	やさしい	6	83	キショい	1	133	淋しい	1
34	明るい	6	84	キツイ	1			
35	恋しい	6	85	きもい	1			
36	いやらしい	5	86	ゴツイ	1			
37	かっけー	5	87	ごっつい	1			
38	カワイイ	5	88	こわい	1			
39	ゆるい	5	89	さびしい	1			
40	安い	5	90	しつこい	1			
41	新しい	5	91	すくない	1			
42	凄い	5	92	すげ〜	1			
43	長い	5	93	スゲー	1			
44	熱い	5	94	ちっちゃい	1			
45	濃い	5	95	でっかい	1			
46	イカツイ	4	96	ねちっこい	1			
47	おそい	4	97	ひさしい	1			
48	キモい	4	98	ひどい	1			
49	すげえ	4	99	ものすごい	1			
50	すごい	4	100	やべえ	1			

付表4 形容動詞の度数順位語彙表

順位	語	度数	順位	語	度数	順位	語	度数
1	好き	201	51	不安	3	101	マメ	1
2	太め	157	52	不潔	3	102	ムリ	1
3	スリム	141	53	有名	3	103	むり	1
4	暇	88	54	やんちゃ	2	104	めんど	1
5	淫乱	60	55	ヤンチャ	2	105	リアル	1
6	ガチ	38	56	淫ら	2	106	ワイルド	1
7	太目	32	57	下手	2	107	安易	1
8	短め	29	58	完全	2	108	暗め	1
9	面倒	29	59	嫌い	2	109	可哀想	1
10	無理	18	60	自然	2	110	格安	1
11	苦手	17	61	純粹	2	111	楽	1
12	マッチョ	16	62	真っ白	2	112	気楽	1
13	気軽	14	63	退屈	2	113	健康	1
14	極端	13	64	大柄	2	114	嫌	1
15	真面目	13	65	大変	2	115	懸命	1
16	爽やか	12	66	適度	2	116	元気	1
17	ヒマ	11	67	半端	2	117	控えめ	1
18	清潔	11	68	卑怯	2	118	豪華	1
19	大丈夫	10	69	敏感	2	119	骨太	1
20	可能	9	70	不満	2	120	自由	1
21	スケベ	8	71	変	2	121	手軽	1
22	小柄	8	72	豊富	2	122	上手	1
23	タフ	7	73	余計	2	123	上等	1
24	ピンピン	7	74	あっさりめ	1	124	色々	1
25	細め	7	75	あんな	1	125	真剣	1
26	色黒	7	76	いたずら好き	1	126	親切	1
27	エッチ	6	77	いろいろ	1	127	正直	1
28	ひま	6	78	うすめ	1	128	素朴	1
29	マニアック	6	79	エロチック	1	129	駄目	1
30	めちゃくちゃ	6	80	おだやか	1	130	大すぎ	1
31	場違い	6	81	おっけ～	1	131	大事	1
32	大好き	6	82	がち	1	132	沢山	1
33	でかめ	5	83	きれい	1	133	短かめ	1
34	マジ	5	84	グチャグチャ	1	134	遅め	1
35	軽め	5	85	こまめ	1	135	丁重	1
36	綺麗	5	86	こなな	1	136	低め	1
37	ダメ	4	87	ずき	1	137	当たり前	1
38	簡単	4	88	スマート	1	138	得意	1
39	気長	4	89	そんな	1	139	濃厚	1
40	小さめ	4	90	てきとー	1	140	薄め	1
41	少なめ	4	91	どんな	1	141	微妙	1
42	早め	4	92	バカ	1	142	別々	1
43	大胆	4	93	バラバラ	1	143	無駄	1
44	長め	4	94	バルキー	1	144	愉快	1
45	すき	3	95	ひんぱん	1			
46	真っ赤	3	96	ブサイク	1			
47	大きめ	3	97	フリー	1			
48	同じ	3	98	フレンドリー	1			
49	派手	3	99	まじ	1			
50	必要	3	100	まじめ	1			

付表5 副詞の度数順位語彙表 (150位まで)

順位	語	度数	順位	語	度数	順位	語	度数
1	よろしく	203	51	よく	6	101	多分	2
2	ガッチリ	106	52	結構	6	102	アッサリ	1
3	サクッと	63	53	いろいろ	5	103	アンアン	1
4	ムラムラ	63	54	とりあえず	5	104	いちゃいちゃ	1
5	宜しく	59	55	なるべく	5	105	ウジウジ	1
6	まったく	54	56	プラブラ	5	106	うじゃうじゃ	1
7	あまり	39	57	ゆっくり	5	107	うずうず	1
8	どう	37	58	よろ	5	108	ウズウズ	1
9	かなり	35	59	グチョグチョ	4	109	うろろう	1
10	がつちり	29	60	そう	4	110	ガッシリ	1
11	ヨロシク	28	61	たまに	4	111	ガバガバ	1
12	少し	27	62	トロトロ	4	112	がんがん	1
13	ガツツリ	25	63	ひたすら	4	113	ギュッと	1
14	ガンガン	25	64	まあ	4	114	きらきら	1
15	まだ	25	65	まだまだ	4	115	きりっと	1
16	ちよっと	22	66	もち	4	116	グッチョグッチョ	1
17	もちろん	21	67	モロ	4	117	くっと	1
18	マッターリ	20	68	沢山	4	118	こう	1
19	即	20	69	ウロウロ	3	119	さっぱり	1
20	あっさり	18	70	オラオラ	3	120	シコシコ	1
21	必ず	18	71	サクっと	3	121	しっかり	1
22	ぼちゃ	17	72	じっくり	3	122	スグ	1
23	すぐ	15	73	そこそこ	3	123	すこし	1
24	ぼつちャリ	15	74	どうぞ	3	124	ずっと	1
25	ポチャ	14	75	どー	3	125	すでに	1
26	がつつり	13	76	とても	3	126	せつかく	1
27	しばらく	12	77	はっきり	3	127	ぜひ	1
28	たっぷり	12	78	いきなり	2	128	そろそろ	1
29	なかなか	12	79	いっばい	2	129	だいぶ	1
30	まず	12	80	ギンギン	2	130	タツプリ	1
31	もう	12	81	さすが	2	131	ちよっくら	1
32	色々	12	82	さらに	2	132	どうせ	1
33	是非	12	83	ズボズボ	2	133	ときどき	1
34	同じく	12	84	たくさん	2	134	とくに	1
35	あんまり	10	85	ただ	2	135	とつても	1
36	やや	10	86	ちゃんと	2	136	トバーツ	1
37	初めて	10	87	ドキドキ	2	137	なおさら	1
38	あんま	9	88	なお	2	138	ネチネチ	1
39	イチャイチャ	9	89	はじめて	2	139	ノンビリ	1
40	めっちゃ	9	90	ぶらぶら	2	140	のんびり	1
41	めっちゃ	9	91	ポッキリ	2	141	ハッキリ	1
42	もし	9	92	ポツチャリ	2	142	パツと	1
43	とにかく	8	93	まあまあ	2	143	バリバリ	1
44	仲良く	8	94	また	2	144	ヒューヒュー	1
45	ちょい	7	95	ムチムチ	2	145	ぶよぶよ	1
46	全然	7	96	ムツチリ	2	146	ふらふら	1
47	さくっと	6	97	めつき	2	147	ぶらぶら	1
48	すつきり	6	98	より	2	148	ペロペロ	1
49	むっちり	6	99	わりと	2	149	ぼおー	1
50	もっと	6	100	一番	2	150	ボキッ	1

# A basic vocabulary survey of local bulletin boards for MSM

Katsuya Kinjo

This study investigates vocabulary across 6 local bulletin boards for MSM (men who have sex with men). During the period of November 1 (2008) to April 30 (2009), there appeared 42,560 cases (target population) in the 6 bulletin boards. 2,400 units were selected using random numbers generated by Excel. Selected data were analyzed by KH Coder (UniDic as the electric dictionary and ChaSen as the analyzer). The result shows that the number of token words is 77,843 and the number of type words is 3,548.

Among the high-frequent top 50 words, there appeared some unique nouns and verbs: those relating to the sex position, *tachi* (top) and *uke* (bottom), and those action verbs such as *yaru* (do), *dekiru* (can do), and *horu* (dig). It was also found that those high-frequent words are related each other.