

厚生労働科学研究費補助金（食品の安全確保推進研究事業）

総括研究報告書

## 基盤の違いを補足するための伝達方法の検討に関する研究

研究代表者 種村 菜奈枝 慶應義塾大学 薬学部・助教

### 研究要旨

ミス・コミュニケーションが生じる原因として、一般消費者と専門家との間における共通基盤の違いがある。この違いを平準化するため、難解語を特定し、平易化することが重要である。しかし、機能性表示食品の一般消費者向け情報に含まれる難解語を特定するための形態素解析用辞書はない。そこで、形態素解析用辞書の作成およびその精度評価を目的とした。医学または臨床試験用語を収録する10種類の用語集（取得用語数973,895語）を用い、203,095通りの辞書を作成した。その後、機能性表示食品の一般消費者向け情報1,310件からランダムに66件（4.65%）抽出し、作成した全辞書の精度評価を行った（データ取得日2018年7月）。その結果、最も再現率が高い辞書（検出数：MeCab 1,725, マニュアル 529, 精度：適合率 0.283, 再現率 0.924, F値 0.434）で、F値は最大0.961となると推定された。この形態素解析用辞書を用いて、一般消費者向け情報1,310件全ての難解語候補の抽出を行った結果、総計3,691語のうち出現回数10回以上の1,428語において、明らかに医学または臨床試験用語である難解語候補とはならない用語が887語（例：が、ます）含まれていた。最終的に、摂取（15,232回）、評価（7,315回）、健康（4,885回）といった用語の他、計541語の難解語候補が抽出された。今後、さらに実証研究の実施が望まれる。

- |       |        |   |
|-------|--------|---|
| 研究分担者 | 山本 景一  | 和歌山県立医科大学医学部附属病院・部長・准教授                           |
| 研究分担者 | 長 雄一郎  | 東京医科歯科大学 大学院医歯学総合研究科・助教                           |
| 研究分担者 | 中谷 英仁  | 大阪大学 大学院医学研究科・寄付講座助教                              |
| 研究分担者 | 小野寺 理恵 | 札幌医科大学 医学部・講師                                     |
| 研究分担者 | 柿崎 真沙子 | 名古屋市立大学 医学部・特任講師                                  |
| 研究協力者 | 矢澤 一良  | 早稲田大学 早稲田大学ナノ理工学研究機構<br>規範科学総合研究所ヘルスフード科学部門・研究院教授 |
| 研究協力者 | 漆原 尚巳  | 慶應義塾大学 薬学部・教授                                     |
| 研究協力者 | 木村 毅   | 健康食品産業協議会・会長                                      |
| 研究協力者 | 長村 洋一  | 日本食品安全協会・理事長                                      |

研究協力者 阿南 久 消費者市民社会をつくる会・代表理事  
研究協力者 森田 満樹 Food Communication Compass・組織代表

## A. 研究目的

ミス・コミュニケーションが生じる原因として、一般消費者と専門家との間における共通基盤の違いがある（添付1）。この違いを平準化するため、難解語を特定し、平易化することが重要である。そこで我々は、一般消費者向け情報に含まれる医学または臨床試験用語の専門用語を難解語の候補として抽出するための形態素解析用辞書を作成することを目的とした。

## B. 研究方法

### (ア) 方法

#### i. 形態素解析用辞書の作成

##### ① 使用した用語集

本研究では、次の計10種類の用語集「用語集#A)から#J)」(取得用語数973,895語)を用いた(2018年7月時点)。

また、各用語集から除外した用語は次の3つ場合である。

1) 空白入りの英文字で構成された用語を除外した。例えば、「abdominal abscess」「femoral artery」である。

2) 用語の一部に「( かっこ)」が含まれる場合は「( かっこ)以降の文字を除外した。例えば、「習慣性(精神生理学)」は「習慣性」へ変換した。

3) 用語集#G)「医学用語シソーラス 第8版」については、辞書の属性が「情報科学」、「人文科学」、「地理的位置」または「人名・地名」である場合を除外した。例えば、「MP3プレイヤー(情報科学)」、

キリスト教(人文科学)、アイスランド(地理的位置)、あきる野市(人名・地名)である。

さらに、次の場合については、変換した用語も新たに用語集へ追加した。用語の一部に「・ 中黒」や「、句点」が含まれる場合は、「・ 中黒」または「、句点」を除去した残りの用語を追加した。例えば、「アーリー・エクスポージャー」の場合、「アーリーエクスポージャー」を追加した。

よって、最終的に取得用語数973,895語から313,466語を除外、13,194語を追加した結果、延べ673,623語、重複なし472,977語となった(添付5)。それぞれの用語集の取得用語数、除外用語数、および追加用語数は以下の通りであった。

- 用語集#A)「CDISC 標準推進プロジェクト 臨床試験用語集」

CDISC (Clinical Data Interchange Standards Consortium) は国際的な臨床研究データ交換基準を提供している団体である。CDISCから臨床研究に関連する用語集が公開されている。取得用語592語、除外用語17語、追加用語121語、最終的に696語を採用した。

- 用語集#B)「ATOK用 テキスト形式 医学変換辞書」

医療辞書を無償で提供する電子カルテ用標準病名マスターを土台にした医学用語変換辞書である。“みんなで作る医学用語変換

辞書”として一般公開されている。取得用語 98,027 語，除外用語 38 語，追加用語 634 語，最終的に 98,623 語を採用した。

- 用語集#C)「MS-IME 医学用語変換辞書 (1)」

Microsoft の日本語入力システムである IME 用の医学用語変換辞書であり，上述した辞書と同様，医療辞書を無償で提供する電子カルテ用標準病名マスターを土台にした医学用語変換辞書である。“みんなで作る医学用語変換辞書”として一般公開されている。取得用語 45,311 語，除外用語 22 語，追加用語 563 語，最終的に 45,852 語を採用した。

- 用語集#D)「MS-IME 医学用語変換辞書 (2)」

「MS-IME 医学用語変換辞書 (1)」とは異なる作成者により無償配布されている Microsoft の日本語入力システムである IME 用の医学用語変換辞書である。取得用語 44,228 語，除外用語 20 語，追加用語 174 語，最終的に 44,382 語を採用した。

- 用語集#E)「DMiME 医学用語変換辞書」

オープンライセンスの医学・医療用語変換辞書である。Google 日本語入力用のユーザ辞書ファイルとして開発された辞書である。取得用語 41,640 語，除外用語 11 語，追加用語 531 語，最終的に 42,160 語を採用した。

- 用語集#F)「日本医学会医学用語辞典」

この辞書は、『医学・医療関係者が教育，研究，診療，医療行政などの場において，論文や教科書の執筆，診療記録の記載，行政文書の作成などをする際に必要な医学用語を選定し，収載したものである。』とされ，日本医学会により作成された辞書である。取得用語 141,229 語，除外用語 51,045 語，追加用語 877 語，最終的に 91,061 語を採用した。なお，当該辞書は，ユーザ登録が必要であるため，本研究に使用するにあたり事前に使用許諾を日本医学会より得た。

- 用語集#G)「医学用語シソーラス 第 8 版」

医学中央雑誌刊行会が，医学，薬学，公衆衛生学等の分野で使われている用語を関連付けたキーワード集であり，無償で公開されている。このシソーラスに含まれる用語のうち，冒頭の削除用語に関する取り決め事項の他，用語属性が情報科学，人文科学，地理的位置，人名であった用語をさらに削除した。取得用語 587,594 語，除外用語 262,312 語，追加用語 10,171 語，最終的に 335,453 語を採用した。

- 用語集#H)「音訳の部屋 読み方辞典」

音訳者・点訳者，読みを必要としている方のために制作された読み方辞典である。医学用語 1 (病名)，医学用語 2 (部位・症状等)，医学用語 3 (臨床で使う単位) の部分を使用した。取得用語 13,475 語，除外用語 1 語，追加用語 31 語，最終的に 13,505 語を採用した。

- 用語集#I)「国立国語研究所「病院の

言葉」データベース」

「病院の言葉」の分かりにくさには、いくつかの類型があり、詳しく取り上げた57語に、複合語・関連語を加えた152語が、国立国語研究所のホームページにて公開されている。取得用語152語、除外用語0語、追加用語0語、最終的に152語を採用した。

- 用語集#J)「健康食品の素材情報データベース」

国立研究開発法人 医薬基盤・健康・栄養研究所が公開する「健康食品」の安全性・有効性情報のデータベースであり、消費者が適切に商品を選択できるようにするために、「健康食品」に添加されている素材の安全性・有効性の情報を集めたものである。取得用語1,647語、除外用語0語、追加用語92語、最終的に1,739語を採用した。

## ② 形態素解析用辞書の作成

本研究における難解語候補（以下、難解語）の定義は、「医学または臨床試験用語から健康食品の素材情報データベースに含まれる機能性関与成分の用語を除外した用語」とした。そこで本研究では、上述の操作化定義に基づき、「医学または臨床試験用語」を抽出するための用語集として「用語集#A)から#I)」を使用し、さらに「機能性関与成分の用語を除外」するための用語集として「用語集#J)」を使用した（添付2）。

ただし、一般消費者向け情報から医学または臨床試験用語を高精度に抽出するための用語集は存在しないため、表1で示した

「用語集#A)から#I)」のそれぞれの用語集に対し、次に定義する条件0, 1, 2および3の4種類の条件のいずれか1つを適応し、あらゆる用語集の条件の組み合わせから、全203,095通りの形態素解析用辞書を作成した。それぞれの用語集で適応した条件に関する定義は、次の通りとした。「条件0：当該用語集に存在しない」、「条件1：当該用語集に存在する」、「条件2：無条件」、「条件3：“条件3”と指定した用語集のうち、いずれかの用語集に存在する」。

最後に、本研究で作成した全203,095通りの辞書から英単語、数値、記号といった用語を除外した。

## ii. 形態素解析用辞書の精度評価

### ① 解析対象情報

一般消費者向け情報1,408件のうち旧届出様式を除いた1,310件を精度評価のための解析対象情報とした（データ取得日2018年7月）(21)。その後、1,310件からランダム抽出した66件(4.65%)の情報を対象に、精度評価を行った。これら一連の精度評価を「調査No.1」とした。

### ② 正答用語集合の作成

疫学、医学および薬学分野に精通した2名が独立して、それぞれ解析対象情報となった66件の一般消費者向け情報から難解語をマニュアル抽出した。これらの難解語は、表2で示した正答用語集合のうち難解語“あり”である「TP (True Positive) とFN (False Negative)」で構成される（表2）。その後、意見が分かれた用語に対して2名で討議を行い、疑義がなくなる

まで繰り返し検討を重ねた(添付3).

### ③ 形態素解析の実行

本研究で作成した形態素解析用辞書203,095種類を用いて、解析対象情報となった66件の一般消費者向け情報の形態素解析を行い、難解語を抽出した(添付3). 形態素解析エンジンは、オープンソースの形態素解析器 MeCab 0.996 を使用した. 形態素解析の結果、難解語“あり”と判定されたものは、「TP (True Positive) と FP (False Positive)」で示される(添付6).

### ④ 精度評価

本研究では、作成した形態素解析用辞書203,095種類のうち、どの辞書が最も高精度に難解語を抽出することができるかを特定するため、次の3つの指標を用いて抽出精度を評価した. 一般的に形態素解析の精度評価の指標として再現率、適合率、F値が用いられるため、本研究においても同様の評価指標を採用した(添付3,6).

#### ● 適合率:

本研究で作成した形態素解析用辞書を用いて形態素解析を実行することにより難解語と予測した用語「TP (True Positive) と FP (False Positive)」のうち、実際に正答用語集合の難解語「TP (True Positive)」であった割合.

#### ● 再現率

正答用語集合に含まれた全ての難解語のうち「TP (True Positive) と FN (False Negative)」のうち、形態素解析を実行す

ることにより難解語候補と予測した用語「TP (True Positive)」の割合.

#### ● F値

再現率と適合率の調和平均

### ⑤ 形態素解析用辞書の採用

上述の「調査No.1」の精度評価の結果、再現率またはF値が最も高い辞書を一般消費者向け情報に含まれる難解語を抽出するための候補辞書とした. なお、再現率が最も高い辞書を用いた検討を「条件1」、F値が最も高い辞書を用いた検討を「条件2」とした. その後、「条件1」または「条件2」のどちらの辞書を選択した場合において、最終的に難解語の抽出精度がより向上するのかを予測するため試行実験を行った.

### iii. 形態素解析用辞書の抽出精度の向上予測のための試行実験

#### ① 次回除外係数の算出

上述した「条件1」または「条件2」の2つの候補辞書を用いて、「調査No.1」の解析対象情報66件の一般消費者向け情報を対象に形態素解析器で抽出した難解語集合を「MeCab (TP+FP)」とし、ここからマニュアル抽出した正答用語集合の難解語(TP)を除外した集合「MeCab (FP)」を特定した.

次に、「調査No.1」で解析対象情報とならなかった残りの1,244件の一般消費者向け情報を「試行No.2」から「試行No.20」に19分割した. これら19個の試行グループそれぞれにおいて、「条件1」または「条件2」の2つの候補辞書を用いて、形

形態素解析器 MeCab で難解語を抽出した。それらを集合「MeCab (TP+FP)」のうちに、「調査 No. 1」で特定した難解語ではない用語集合「MeCab (FP)」が何語含まれたか算出した。

最後に、「試行 No. 2」から「試行 No. 20」の試行グループの範囲で、「MeCab (FP)」の平均値である除外平均用語数を求め、この値を「調査 No. 1」の「MeCab (FP)」数で除したものを「次回除外係数 (%)」と定義した。

## ② 抽出精度の向上予測

### (1) 難解語抽出用語数および除外用語数の算出方法

「条件 1」または「条件 2」の 2 つの形態素解析用辞書の難解語抽出精度の向上予測をした。詳細の手順は以下の通りである。

まず、「調査 No. 1」で得られた難解語抽出用語数およびその精度の結果を基準とした。「調査 No. 1」で得た“正答用語集合 (TP+FN)”に該当する用語を「(a) マニュアル (TP+FN) 【調査 No. 1】」とし、さらに「調査 No. 1」で得た TP に該当する用語を「(c) 共通 (TP)」として、これら 2 つの値を「調査 No. 1」, 「試行 No. 2」から「試行 No. 20」まで固定した。

続いて、「(d) MeCab (FP)」は、「(b) MeCab (TP+FP)」から「(c) 共通 (TP)」を引いた値、また「(e) 次回除外用語数」は、「(d) MeCab (FP)」に「次回除外係数 (%)」を乗じたものとした。ただし、「試行 No. 2」以降の「(b) MeCab (TP+FP)」は、前の調査または試行の「(b) MeCab (TP+FP)」から前調査または試行の「(e)

次回除外用語数」を引いたものと定義した。

### (2) 精度評価

その後、試行ごとに「(a) マニュアル (TP+FN) 【調査 No. 1】」, 「(b) MeCab (TP+FP)」および「(c) 共通 (TP)」を用いて、「条件 1」または「条件 2」の形態素解析用辞書の精度 (適合率, 再現率, および F 値) を算出した。

### (3) 形態素解析用辞書の選択

最後に、「条件 1」または「条件 2」の 2 つの形態素解析用辞書の難解語抽出精度の向上予測曲線を「調査 No. 1」, 「試行 No. 2」から「試行 No. 20」それぞれの F 値を用いて作成し、F 値が最も 1 に近くなる高精度の辞書を最終的に形態素解析用辞書として選択した。

## iv. 難解語候補の抽出

最終的に選択した形態素解析用辞書を用いて、一般消費者向け情報 1,310 件を対象に形態素解析器 MeCab で難解語を抽出した。

### (倫理面への配慮)

本研究の調査対象は、消費者庁のホームページや一般公開されている既存資料であるため、該当事項なし。

## C. 研究結果

### (ア) 形態素解析用辞書の精度評価

「調査 No. 1」の精度評価の結果を示す。形態素解析用辞書 203,095 件の辞書ごとに適合率, 再現率, F 値を算出し、再現

率およびF値のそれぞれ上位15種類の精度評価の結果を添付7および添付8に示した。最も再現率が高い形態素解析用辞書を使用した場合は、「用語集#A), #B), #D)から#G)」は条件3, かつ「用語集#C), #H), #I)」は条件2, とした場合であった(検出数: MeCab 1725, マニュアル 529, 精度: 適合率 0.283, 再現率 0.924, F値 0.434) (添付7)。また, 最もF値が高い形態素解析用辞書を使用した場合は、「用語集#A), #D), #E), #G), #I)」は条件3, かつ「用語集#B), #C), #F), #H)」は条件2, とした場合であった(検出数: MeCab 1116, マニュアル 529, 精度: 適合率 0.358, 再現率 0.754, F値 0.485) (添付8)。

以上より, これら2つの辞書を一般消費者向け情報に含まれる難解語を抽出するための候補辞書とした。

## (イ) 抽出精度の向上予測の試行実験

### i. 次回除外係数

「条件1」の辞書を用いた場合, 除外平均用語791語を「調査No.1」の難解語ではない用語「MeCab (FP)」1,236語で除した結果, 次回除外係数は64.0%となった(添付9)。続いて, 「条件2」の辞書を用いた場合, 除外平均用語433語を「調査No.1」の難解語ではない用語「MeCab (FP)」717語で除した結果, 次回除外係数は60.4%となった(添付10)。

### ii. 抽出精度の向上予測

「条件1」または「条件2」の2つの形態素解析用辞書の難解語抽出精度の向上予測曲線を「調査No.1」, 「試行No.2」から

「試行No.20」それぞれのF値を用いて, 難解語抽出精度向上の予測曲線を作成した(添付4)。

「条件1」の場合, 「試行No.9」以降は「(e) 次回除外用語数」が0, F値は最大0.961となり, それ以降は変化が見られなかった(添付11)。一方, 「条件2」の場合, 「試行No.9」以降は「(e) 次回除外用語数」が0, F値は最大0.860となり, それ以降は変化が見られなかった(添付12)。

これらの結果を踏まえ, 再現率が最も高い辞書「条件1」を選択した場合, F値が最も高い辞書「条件2」を選択した場合と比べて, 高精度に一般消費者向け情報から難解語を抽出することができると判断した。よって, 「条件1」の辞書を形態素解析用辞書として最終的に選択した。

## iii. 難解語候補の抽出

その後, 「条件1」の形態素解析用辞書を用いて一般消費者向け情報1,310件を対象に, 形態素解析器 MeCab で難解語の抽出を行い, うち総計3,691語のうち出現回数10回以上の1,428語を添付13に示した。

明らかに医学または臨床試験用語である難解語候補とはならない用語が887語(例: が, ます)含まれていた。最終的に, 摂取(15,232回), 評価(7,315回), 健康(4,885回)といった用語の他, 計541語の難解語が抽出された。(添付13)。

## D. 考察

我々は, 一般消費者向け情報に含まれる医学または臨床試験用語を抽出するために203,095通りの形態素解析用辞書から最も再

現率の高い辞書を選択した。

しかし、我々が選択した辞書の場合、現状では、再現率が 0.924 と高い一方、適合率 0.283 かつ F 値 0.434 といったように、やや精度は低値であった。これは、既存の医学または臨床試験の用語集の中には、これらの分野に無関係な用語が一定数以上含まれていることが原因として考えられた。

今後、本研究で作成した形態解析用辞書の実運用上での利用可能性を考慮した際、我々が作成した辞書に含まれる用語集データから医学または臨床試験とは無関係な用語をさらに継続的に精査して除外（辞書クリーニング）することにより、さらに形態素解析用辞書の抽出精度の向上予測のための試行実験を通して辞書の精度指標である F 値が向上する見込みを推定できたことは大きな意義があったと考えた。

本研究で作成した形態解析用辞書を実際の社会で適応しつつ、医学または臨床試験とは無関係な用語を本研究で作成した形態解析用辞書から取り除く辞書クリーニングの工程プロセスを繰り返し重ねていくことで、さらに高精度に機能性表示食品の機能性および安全性に関する一般消費者向け情報から医学または臨床試験用語といった難解語を抽出することが可能になると考えられた。

## E. 結論

我々は、機能性表示食品の機能性および安全性に関する一般消費者向け情報に含まれる医学または臨床試験用語を抽出するための新たな形態解析用辞書を作成した。今後、食のリスクコミュニケーション推進のための試みや臨床現場等での利活用に向け、本研究で作成した形態素解析用辞書の実証研究の実施が望まれる。

## F. 健康危険情報

該当なし

## G. 研究発表

### 1. 論文発表

該当なし

### 2. 学会発表

該当なし

## H. 知的財産権の出願・登録状況

### 1. 特許取得

該当なし

### 2. 実用新案登録

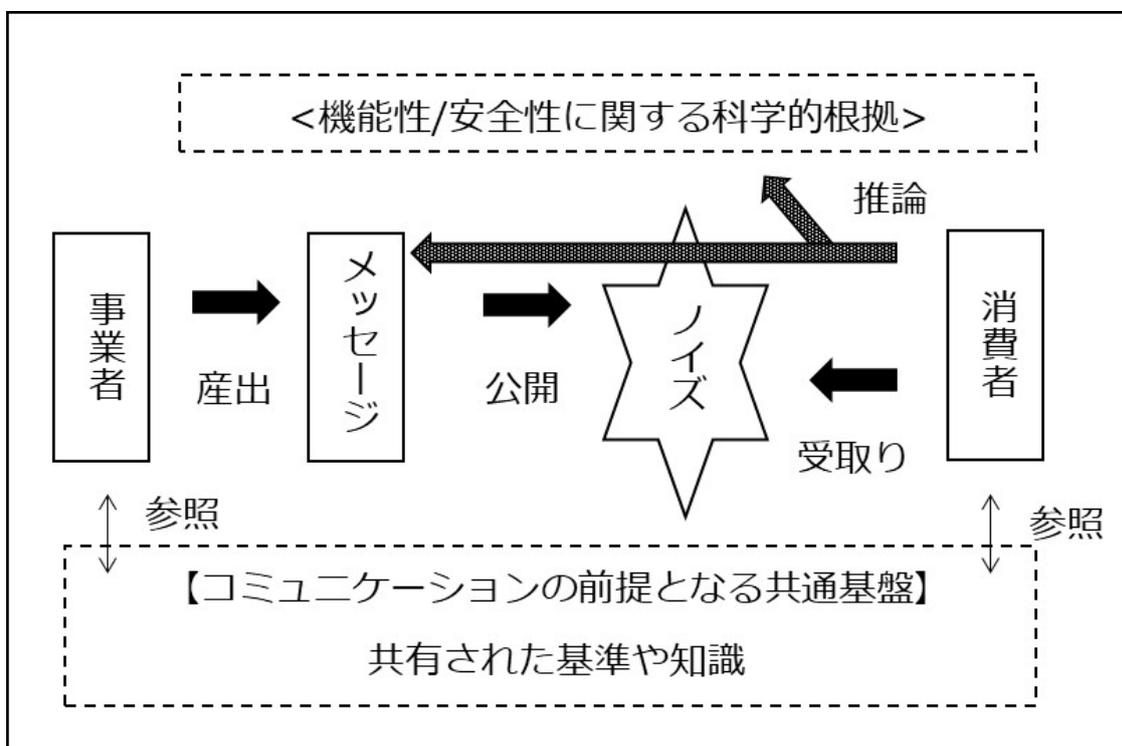
該当なし

### 3. その他

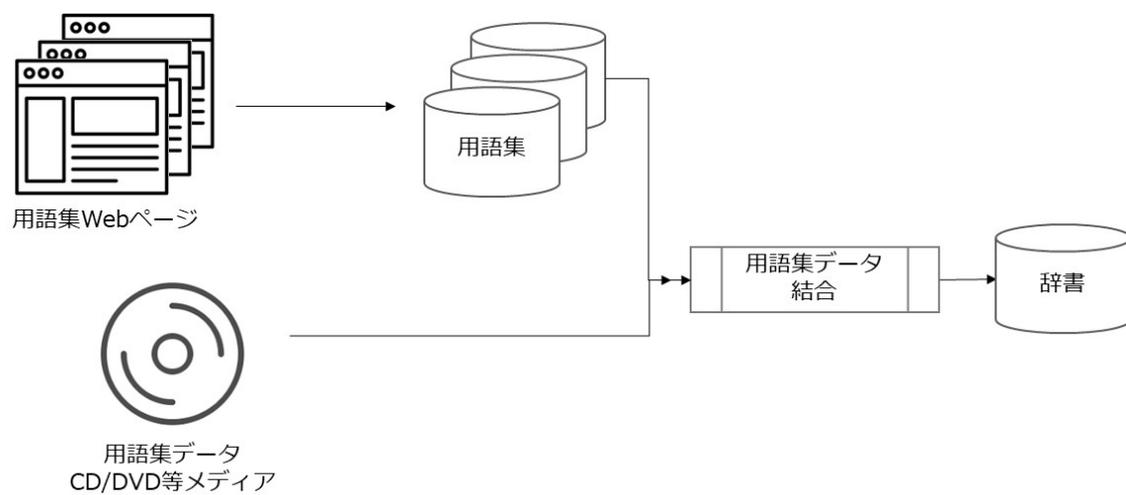
該当なし

# 資 料

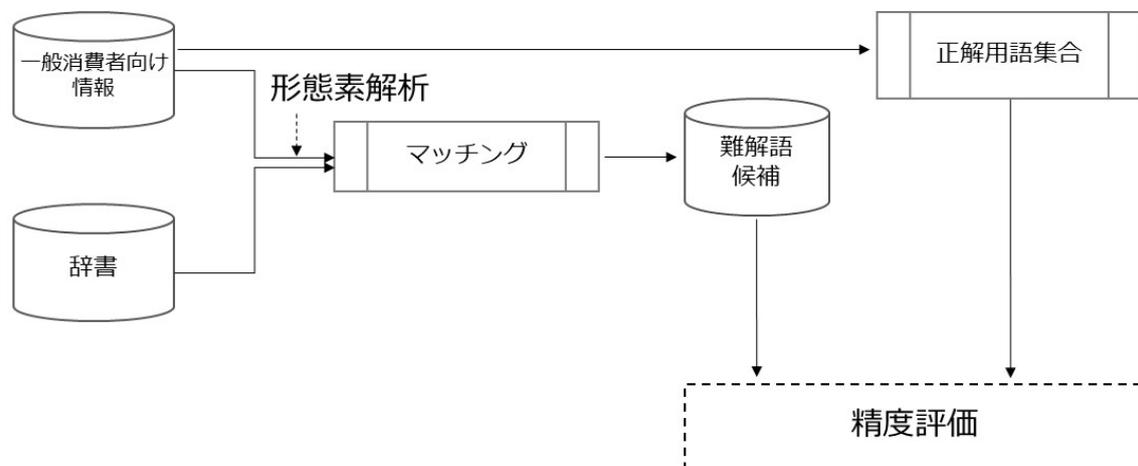
添付1. コミュニケーションモデル



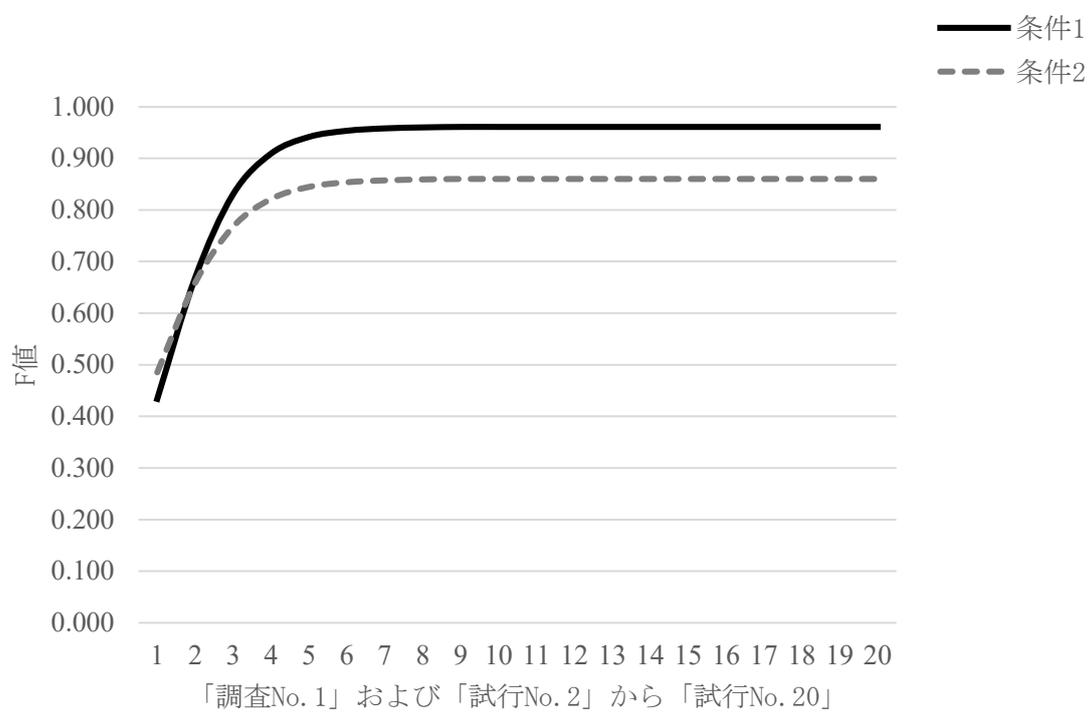
## 添付 2. 難解語抽出のための形態素解析用辞書作成の流れ



### 添付 3. 難解語抽出のための形態素解析用辞書の精度評価



#### 添付 4. 形態素解析用辞書の抽出精度の向上予測



添付 5. 形態素解析用辞書作成に用いた用語集の概要

#	用語集	取得用語数	除外用語数	追加用語数	採用用語数
A	CDISC 標準推進プロジェクト 臨床試験用語集	592	-17	121	696
B	ATOK 用 テキスト形式 医学変換辞書	98,027	-38	634	98,623
C	MS-IME 医学用語変換辞書 (1)	45,311	-22	563	45,852
D	MS-IME 医学用語変換辞書 (2)	44,228	-20	174	44,382
E	DMiME 医学用語変換辞書	41,640	-11	531	42,160
F	日本医学会医学用語辞典	141,229	-51,045	877	91,061
G	医学用語シソーラス 第8版	587,594	-262,312	10,171	335,453
H	音訳の部屋 読み方辞典	13,475	-1	31	13,505
I	国立国語研究所「病院の言葉」データベース	152	0	0	152
J	健康食品の素材情報データベース	1,647	0	92	1,739
				延べ用語数	673,623
				重複なし合計用語数	472,977

添付 6. 精度評価のための指標

		正答用語集合	
		難解語の有無	
		あり	なし
形態素解析	難解語の有無	あり	なし
		あり	なし
	あり	TP	FP
	なし	FN	TN

TP: true positive, TN: true negative, FP: false positive, FN: false negative

添付 7. 形態素解析用辞書の精度：適合率・再現率・F 値（再現率 上位 15 件）

#用語集とそれぞれの条件										難解語抽出用語数			精度		
A	B	C	D	E	F	G	H	I	マニュアル (TP+FN)	MeCab (TP+FP)	共通 (TP)	適合率	再現率	F 値	
3	3	2	3	3	3	3	2	2	529	1,725	489	0.283	0.924	0.434	
3	3	2	3	3	3	3	2	3	529	1,725	489	0.283	0.924	0.434	
3	3	3	3	3	3	3	2	2	529	1,725	489	0.283	0.924	0.434	
3	3	3	3	3	3	3	2	3	529	1,725	489	0.283	0.924	0.434	
3	3	2	3	3	3	3	3	2	529	1,728	489	0.283	0.924	0.433	
3	3	2	3	3	3	3	3	3	529	1,728	489	0.283	0.924	0.433	
3	3	3	3	3	3	3	3	2	529	1,728	489	0.283	0.924	0.433	
3	3	3	3	3	3	3	3	3	529	1,728	489	0.283	0.924	0.433	
3	3	2	3	2	3	3	2	2	529	1,709	486	0.284	0.919	0.434	
3	3	2	3	2	3	3	2	3	529	1,709	486	0.284	0.919	0.434	
3	3	3	3	2	3	3	2	2	529	1,709	486	0.284	0.919	0.434	
3	3	3	3	2	3	3	2	3	529	1,709	486	0.284	0.919	0.434	
3	3	2	3	2	3	3	3	2	529	1,712	486	0.284	0.919	0.434	
3	3	2	3	2	3	3	3	3	529	1,712	486	0.284	0.919	0.434	
3	3	3	3	2	3	3	3	2	529	1,712	486	0.284	0.919	0.434	

TP : True Positive, FP : False Positive, FN : False Negative

用語集#A) CDISC 標準推進プロジェクト 臨床試験用語集, 用語集#B) ATOK 用 テキスト形式 医学変換辞書, 用語集#C) MS-IME 医学用語変換辞書 (1), 用語集#D) MS-IME 医学用語変換辞書 (2), 用語集#E) DMiME 医学用語変換辞書, 用語集#F) 日本医学会医学用語辞典, 用語集#G) 医学用語シソーラス 第8版, 用語集#H) 音訳の部屋 読み方辞典, 用語集#I) 国立国語研究所「病院の言葉」データベース

一般消費者向け届出情報 1,310 件のうち, 61 件 (4.65%) を「調査 No.1」とした場合の精度評価結果である.  
各用語集の条件の定義は次の通りとした. 「条件 0: 当該用語集に存在しない」, 「条件 1: 当該用語集に存在する」, 「条件 2: 無条件」, 「条件 3: “条件 3”と指定した全ての用語集のうち, いずれかの用語集に存在する」

添付 8. 形態素解析用辞書の精度：適合率・再現率・F 値 (F 値 上位 15 件)

#用語集とそれぞれの条件										難解語抽出用語数			精度		
A	B	C	D	E	F	G	H	I	マニュアル (TP+FN)	MeCab (TP+FP)	共通 (TP)	適合率	再現率	F 値	
3	2	2	3	3	2	3	2	3	529	1,116	399	0.358	0.754	0.485	
3	2	2	3	3	2	3	2	2	529	1,114	398	0.357	0.752	0.484	
3	2	2	3	2	2	3	2	3	529	1,073	388	0.362	0.733	0.484	
3	2	2	3	2	2	3	2	2	529	1,071	387	0.361	0.732	0.484	
3	2	2	3	3	2	3	3	3	529	1,149	404	0.352	0.764	0.482	
2	2	2	3	3	2	3	2	3	529	1,080	387	0.358	0.732	0.481	
3	2	2	3	3	2	3	3	2	529	1,147	403	0.351	0.762	0.481	
3	2	2	3	2	2	3	3	3	529	1,114	395	0.355	0.747	0.481	
2	2	2	3	2	2	3	2	3	529	1,037	376	0.363	0.711	0.480	
3	2	2	3	2	2	3	3	2	529	1,112	394	0.354	0.745	0.480	
2	2	2	3	3	2	3	2	2	529	1,076	384	0.357	0.726	0.479	
2	2	2	3	3	2	3	3	3	529	1,114	393	0.353	0.743	0.478	
2	2	2	3	2	2	3	3	3	529	1,079	384	0.356	0.726	0.478	
2	2	2	3	2	2	3	2	2	529	1,033	373	0.361	0.705	0.478	
3	3	2	3	3	3	1	2	2	529	578	264	0.457	0.499	0.477	

TP : True Positive, FP : False Positive, FN : False Negative

用語集#A) CDISC 標準推進プロジェクト 臨床試験用語集, 用語集#B) ATOK 用 テキスト形式 医学変換辞書, 用語集#C) MS-IME 医学用語変換辞書 (1), 用語集#D) MS-IME 医学用語変換辞書 (2), 用語集#E) DMiME 医学用語変換辞書, 用語集#F) 日本医学会医学用語辞典, 用語集#G) 医学用語シソーラス 第8版, 用語集#H) 音訳の部屋 読み方辞典, 用語集#I) 国立国語研究所「病院の言葉」データベース

一般消費者向け届出情報 1,310 件のうち, 61 件 (4.65%) を「調査 No.1」とした場合の精度評価結果である.  
各用語集の条件の定義は次の通りとした. 「条件 0: 当該用語集に存在しない」, 「条件 1: 当該用語集に存在する」, 「条件 2: 無条件」, 「条件 3: “条件 3”と指定した全ての用語集のうち, いずれかの用語集に存在する」

添付 9. 次回除外係数：「条件 1（再現率が最も高い条件の辞書）」を使用した場合

難解語抽出用語数		
試行 No	MeCab (TP+FP)	MeCab (FP) *
2	1,707	820
3	1,807	830
4	1,766	801
5	1,706	786
6	1,759	779
7	1,881	856
8	1,721	783
9	1,557	749
10	1,889	810
11	1,667	808
12	1,705	812
13	1,676	786
14	1,600	733
15	1,778	789
16	1,624	756
17	1,760	786
18	1,692	784
19	1,698	748
20	1,757	817
平均	1,724	791

次回除外係数\*\* 64.0%

TP : True Positive, FP : False Positive

\* 精度評価（調査 No.1）において採用した辞書を用いて形態素解析を行った場合、MeCab 抽出用語である「MeCab (TP+FP)」のうちマニュアル抽出した正解用語集合に含まれなかった 1,236 語と合致していた用語数「MeCab (FP)」。

\*\* 除外平均用語数を「調査 No.1」で特定した難解語ではない用語「MeCab (FP)」1,236 語で除した値。

添付 10. 次回除外係数：「条件 2（F 値が最も高い条件の辞書）」を使用した場合

難解語抽出用語数		
試行 No	MeCab (TP+FP)	MeCab (FP) *
2	1,101	452
3	1,158	455
4	1,135	436
5	1,118	440
6	1,132	424
7	1,230	477
8	1,094	424
9	1,024	407
10	1,220	436
11	1,084	445
12	1,093	446
13	1,097	439
14	1,009	391
15	1,144	436
16	1,036	408
17	1,119	431
18	1,073	422
19	1,100	404
20	1,107	446
平均	1,109	433

次回除外係数\*\*                      60.4%

TP : True Positive, FP : False Positive

\* 精度評価（調査 No.1）において採用した辞書を用いて形態素解析を行った場合，MeCab 抽出用語である「MeCab (TP+FP)」のうちマニュアル抽出した正解用語集合に含まれなかった 717 語と合致していた用語数「MeCab (FP)」。

\*\* 除外平均用語数を「調査 No.1」で特定した難解語ではない用語「MeCab (FP)」717 語で除した値。

添付 11. 抽出精度の向上予測：「条件 1（再現率が最も高い条件の辞書）」を使用した場合

調査 または 試行 No	難解語抽出用語数				除外用語数	精度		
	(a) マニュアル (TP+FN)	(b) MeCab (TP+FP)	(c) 共通 (TP)	(d) MeCab (FP)	(e) 次回除外用語数	適合率	再現率	F 値
	【調査 No.1】 固定値	【試行 No.2 以降】 前回(b)-前回(e)	【調査 No.1】 固定値	(b)-(c)	(d)×次回除外係数			
1*	529	1,725	489	1,236	791	0.283	0.924	0.434
2**	529	934	489	445	285	0.524	0.924	0.668
3	529	649	489	160	102	0.753	0.924	0.830
4	529	547	489	58	37	0.894	0.924	0.909
5	529	510	489	21	13	0.959	0.924	0.941
6	529	497	489	8	5	0.984	0.924	0.953
7	529	492	489	3	2	0.994	0.924	0.958
8	529	490	489	1	1	0.998	0.924	0.960
9	529	489	489	0	0	1.000	0.924	0.961
10	529	489	489	0	0	1.000	0.924	0.961
11	529	489	489	0	0	1.000	0.924	0.961
12	529	489	489	0	0	1.000	0.924	0.961
13	529	489	489	0	0	1.000	0.924	0.961
14	529	489	489	0	0	1.000	0.924	0.961
15	529	489	489	0	0	1.000	0.924	0.961

調査 または 試行 No	難解語抽出用語数				除外用語数	精度		
	(a) マニュアル (TP+FN)	(b) MeCab (TP+FP)	(c) 共通 (TP)	(d) MeCab (FP)	(e) 次回除外用語数	適合率	再現率	F 値
	【調査 No.1】 固定値	【試行 No.2 以降】 前回(b)-前回(e)	【調査 No.1】 固定値	(b)-(c)	(d)×次回除外係数			
16	529	489	489	0	0	1.000	0.924	0.961
17	529	489	489	0	0	1.000	0.924	0.961
18	529	489	489	0	0	1.000	0.924	0.961
19	529	489	489	0	0	1.000	0.924	0.961
20	529	489	489	0	0	1.000	0.924	0.961

TP : True Positive, FP : False Positive, FN : False Negative

\* 「調査 No.1」の解析対象情報 66 件および「調査 No.1」で得られた難解語抽出用語数および精度結果を基準とした。

\*\* 「試行 No.2」以降は試行実験の結果である。

添付 12. 抽出精度の向上予測：「条件 2 (F 値が最も高い条件の辞書)」を使用した場合

調査 または 試行 No	難解語抽出用語数				除外用語数	精度		
	(a) マニュアル (TP+FN)	(b) MeCab (TP+FP)	(c) 共通 (TP)	(d) MeCab (FP)	(e) 次回除外用語数	適合率	再現率	F 値
	【調査 No.1】 固定値	----- 【試行 No.2 以降】 前回(b)-前回(e)	【調査 No.1】 固定値	(b)-(c)	(d)×次回除外係数			
1*	529	1,116	399	717	433	0.358	0.754	0.485
2**	529	683	399	284	172	0.584	0.754	0.658
3	529	511	399	112	68	0.781	0.754	0.767
4	529	443	399	44	27	0.901	0.754	0.821
5	529	416	399	17	10	0.959	0.754	0.844
6	529	406	399	7	4	0.983	0.754	0.853
7	529	402	399	3	2	0.993	0.754	0.857
8	529	400	399	1	1	0.998	0.754	0.859
9	529	399	399	0	0	1.000	0.754	0.860
10	529	399	399	0	0	1.000	0.754	0.860
11	529	399	399	0	0	1.000	0.754	0.860
12	529	399	399	0	0	1.000	0.754	0.860
13	529	399	399	0	0	1.000	0.754	0.860
14	529	399	399	0	0	1.000	0.754	0.860
15	529	399	399	0	0	1.000	0.754	0.860

調査 または 試行 No	難解語抽出用語数				除外用語数	精度		
	(a) マニュアル (TP+FN)	(b) MeCab (TP+FP)	(c) 共通 (TP)	(d) MeCab (FP)	(e) 次回除外用語数	適合率	再現率	F 値
	【調査 No.1】 固定値	----- 【試行 No.2 以降】 前回(b)-前回(e)	【調査 No.1】 固定値	(b)-(c)	(d)×次回除外係数			
16	529	399	399	0	0	1.000	0.754	0.860
17	529	399	399	0	0	1.000	0.754	0.860
18	529	399	399	0	0	1.000	0.754	0.860
19	529	399	399	0	0	1.000	0.754	0.860
20	529	399	399	0	0	1.000	0.754	0.860

TP : True Positive, FP : False Positive, FN : False Negative

\* 「調査 No.1」の解析対象情報 66 件および「調査 No.1」で得られた難解語抽出用語数および精度結果を基準とした.

\*\* 「試行 No.2」以降は試行実験の結果である.

添付 13. 難解語候補  
(出現回数 10 回以上)

Rank	用語	出現回数
1	が	41,065
2	摂取	15,232
3	ます	8,870
4	食品	7,798
5	評価	7,315
6	報告	6,083
7	あり	6,042
8	から	5,878
9	健康	4,885
10	的	4,549
11	試験	4,537
12	成分	4,409
13	G A B A	4,357
14	文献	3,921
15	性	3,196
16	以上	3,041
17	食後	2,548
18	目的	2,519
19	確認	2,517
20	判断	2,491
21	か	2,250
22	科学的根拠	2,247
23	工場	2,215
24	上昇	2,201
25	検証	2,085
26	血圧	2,013
27	配合	1,868
28	G M P	1,800
29	加工	1,756
30	抑制	1,733
31	主	1,638
32	相互作用	1,617

33	特性	1,528
34	背景	1,497
35	認証	1,478
36	罹患	1,452
37	特定保健用食品	1,403
38	花	1,402
39	維持	1,394
40	葛	1,392
41	プラセボ	1,265
42	ビフィズス菌	1,261
43	調査	1,217
44	G	1,155
45	注意	1,140
46	摂取量	1,138
47	人	1,127
48	健常成人	1,122
49	安全	1,030
50	バイアス	1,017
51	気	957
52	g	926
53	L-テアニン	914
54	品質管理	892
55	栄養	867
56	基準	862
57	飲料	847
58	有効性	840
59	水分	817
60	数	808
61	ヒアルロン酸N a	798
62	して	790
63	B	783
64	I S O	773
65	有意	754
66	肌	749
67	認知機能	737
68	衛生管理	725

69	被験者	710	105	化	522
70	体重	706	106	II	516
71	否定	693	107	内臓脂肪	510
72	一般	692	108	株	508
73	フラボノイド	691	109	製造工場	508
74	ラクトン	680	110	限界	506
75	生産	677	111	米	488
76	テクトリゲニン	666	112	妊産婦	485
77	継続	654	113	英語	482
78	大豆	649	114	健康食品	472
79	体制	643	115	4週間	464
80	相談	641	116	グルコサミン塩酸塩	463
81	吸収	637	117	環境	463
82	B B	623	118	構造化抄録	459
83	という	615	119	HM	453
84	規格	611	120	対照	449
85	R C T	608	121	飲用	445
86	データ	596	122	システマティックレビュー	444
87	睡眠	596	123	記憶力	444
88	有意な	587	124	非変性	443
89	期待	584	125	得	428
90	ストレス	574	126	錠剤	427
91	便通	572	127	還元型コエンザイムQ10	422
92	1日当たり	568	128	偏り	417
93	各	560	129	研究報告	415
94	過剰摂取	559	130	糖	405
95	作業	558	131	力	404
96	リスク	553	132	併用	403
97	協会	546	133	素材	401
98	調子	543	134	粒	400
99	認定	542	135	5g	400
100	型	540	136	研究論文	394
101	利益相反	537	137	ウ	390
102	ランダム化比較試験	535	138	許可	388
103	面積	534	139	血圧低下	387
104	上記	523	140	示唆	386

141	メタアナリシス	381	177	過程	288
142	発生	372	178	う	285
143	利用	371	179	分解	281
144	カプセル	367	180	傾向	280
145	体質	366	181	開始	280
146	パッケージ	365	182	届	279
147	食品添加物	364	183	エキス	276
148	一時	363	184	統合	273
149	J	362	185	パイナップル	270
150	種	354	186	範囲	270
151	健常人	346	187	評価項目	263
152	疲労	342	188	精神的ストレス	262
153	薬剤師	340	189	施設	261
154	C	340	190	p	258
155	比較試験	337	191	野菜	255
156	モノグルコシルヘスペリジン	336	192	脂肪の	253
157	調節	336	193	有害	250
158	乾燥	335	194	未成年者	248
159	カ	333	195	粉末	241
160	2週間	331	196	食生活	240
161	記憶	330	197	黄斑色素	239
162	検索対象	321	198	注意事項	237
163	計画	312	199	蒸散	235
164	研究デザイン	309	200	有効	234
165	食品安全	309	201	l	234
166	査読	308	202	測定	233
167	原因	307	203	s	229
168	排便	301	204	筋力	227
169	コントラスト感度	299	205	品質保証	225
170	抑制効果	297	206	適合	224
171	乾燥肌	295	207	定性的	221
172	機能性表示食品	294	208	コレステロール	220
173	胴囲	294	209	妊娠中	220
174	活動	290	210	R A S	219
175	限定	290	211	サラシノール	217
176	腹部脂肪	289	212	β-グルカン	215

213	色	215	249	文献調査	175
214	体脂肪	213	250	関節の	175
215	B i f i X	212	251	会社	174
216	日本全国	205	252	集団	173
217	色素	205	253	5 m g	173
218	水溶性	204	254	ランダム化	172
219	性別	201	255	起床時	170
220	年間	198	256	保護	169
221	厚生労働省	197	257	魚介	169
222	強	197	258	中高年	166
223	清涼飲料水	197	259	精製	166
224	授乳中	196	260	身体的な	165
225	果実	196	261	ジュース	164
226	規格外	196	262	20歳以上	164
227	o	195	263	緑茶	163
228	衛生	194	264	ポリフェノール	159
229	分析	193	265	動物	159
230	収縮期血圧	193	266	統計解析	156
231	長期摂取	193	267	歩行	154
232	ワルファリン	192	268	肌荒れ	153
233	無作為化	190	269	関連性	153
234	筋肉量	189	270	P	152
235	100 m g	188	271	回復	151
236	光	187	272	天然	151
237	排便回数	187	273	負荷	151
238	能力	187	274	除外	149
239	倍	186	275	無作為	148
240	専門家	186	276	支持	147
241	介入	185	277	ウエスト周囲径	146
242	包装	184	278	統計	146
243	H A C C P	182	279	食事の	146
244	サケ	181	280	デザイン	144
245	例えば	180	281	眠気	144
246	摂取期間	180	282	視覚	143
247	ガセリ菌	179	283	アンモニア	142
248	遺伝毒性	177	284	服用中	142

285	L a c t o b a c i l l u s	142	321	F D	120
286	便中	141	322	食事療法	119
287	I 度	140	323	1 0 0 g	119
288	変動	139	324	付	118
289	弱い	139	325	疲れ	118
290	添加物	139	326	曲線下面積	117
291	鼻の	139	327	果物	117
292	発酵乳	137	328	魚	116
293	C P	137	329	とき	115
294	なし	136	330	言葉	115
295	代謝	135	331	運動不足	115
296	4つ	135	332	L D L	115
297	医中誌	134	333	L D L - C	115
298	柔軟	134	334	抗	114
299	アミノ酸	133	335	元	113
300	精神的負荷	133	336	試験食	113
301	3つ	132	337	ターミナリアベリリカ	112
302	添加	131	338	身体的	112
303	V D T	130	339	1 5 0 m g	112
304	I	129	340	ホルモン	111
305	図形	129	341	参考文献	111
306	主要な	127	342	手法	111
307	ライト	126	343	牛乳	111
308	食品の安全性	126	344	e	111
309	ティロロサイド	125	345	刺激	109
310	ゼリー	124	346	食後に	109
311	東洋	124	347	r	109
312	記録	124	348	アウトカム	108
313	f	124	349	動物試験	108
314	変異原性試験	123	350	工業	108
315	消化	123	351	正常血圧	108
316	前後	122	352	知見	108
317	A U C	122	353	血中濃度曲線下面積	107
318	H D L	122	354	アグリコン	106
319	視機能	121	355	頻度	106
320	検索サイト	120			

356	3-ヒドロキシ-3-メチル ブチレート	106	391	60分	92
357	時期	105	392	毒性試験	91
358	健康状態	104	393	パン	90
359	無作為化対照試験	104	394	酸化ストレス	90
360	製造所	104	395	骨の	90
361	I	104	396	2004年	90
362	空腹時血糖値	103	397	ハンド	88
363	承認	101	398	精神的疲労	88
364	皮膚の	101	399	面	88
365	高用量	101	400	VP	88
366	TC	101	401	手順書	87
367	各指標	100	402	蓄積	87
368	植物由来	99	403	アリ	86
369	直接	99	404	メント	86
370	精確	99	405	欧州食品安全機関	86
371	精神的	99	406	食事摂取基準	86
372	臨床研究	99	407	コール酸	85
373	トウモロコシ	98	408	摂取後	85
374	作用機序	98	409	正常域	85
375	妊娠の	98	410	ADI	85
376	グアーガム	97	411	BB-1	85
377	因果関係	97	412	ヨーグルト	84
378	意味	97	413	急性毒性試験	84
379	WHO	97	414	文献数	84
380	投与量	96	415	エラグ酸	83
381	統計学	96	416	グラブリン	83
382	健康影響評価	95	417	チョコレート	83
383	利害関係	95	418	よらず	82
384	ナス	94	419	同等性	82
385	中止	94	420	米飯	82
386	穀物	94	421	h	82
387	感じ	93	422	層別	81
388	タイプ	92	423	独立行政法人	81
389	バラ	92	424	2週間後	81
390	観察	92	425	No.	81
			426	低用量	80

427	単一の	80	463	尚	71
428	妥当性	80	464	未報告	71
429	安定	80	465	監査	71
430	月	80	466	眠り	71
431	製剤	80	467	脳波	71
432	1月5日	80	468	III	70
433	X	80	469	反復投与	70
434	サンプルサイズ	79	470	感度	70
435	約2	78	471	未成年	70
436	自立	78	472	みそ汁	69
437	合成	77	473	予防	69
438	C T X	77	474	最大	69
439	D r e a m	77	475	5年以上	69
440	d	77	476	カロテノイド	68
441	ロンガム	76	477	パッケージング	68
442	健康な	76	478	判断力	68
443	3倍	76	479	培養細胞	68
444	I P	76	480	定量的	68
445	生活の質	75	481	手順	68
446	腐敗	75	482	豆	68
447	選択	75	483	豆腐	68
448	心理的ストレス	74	484	5つ	68
449	申請	74	485	i n	68
450	血流量	74	486	ウンシュウミカン	67
451	N A G	74	487	クロモグラニンA	67
452	信頼	73	488	中心	67
453	就寝前	73	489	光学密度	67
454	研究計画	73	490	変質	67
455	終了	73	491	推定	67
456	重篤な有害事象	73	492	清掃	67
457	鼠	73	493	無毒性量	67
458	ソーセージ	72	494	18歳以上	67
459	果	72	495	VMA	67
460	消費者	72	496	こんにゃく	66
461	除外基準	72	497	割合	66
462	大学	71	498	手洗い	66

499	無作為化比較試験	66
500	調味料	66
501	人種	65
502	例数	65
503	参考	65
504	抗凝固薬	65
505	雄	65
506	げっ歯類	64
507	スープ	64
508	ライン	64
509	亜慢性毒性	64
510	促進	64
511	含量	64
512	通院中	64
513	G C	64
514	N T T	64
515	唾	63
516	商品名	63
517	急性毒性	63
518	抗血栓薬	63
519	服装	63
520	液中	63
521	環境の	63
522	自覚	63
523	興奮剤	63
524	認識力	63
525	1 9 7 5 年	63
526	I I	63
527	L 2	63
528	ヒハツ	62
529	入退室	62
530	共同	62
531	妥当	62
532	容器	62
533	血液凝固剤	62
534	診断	62
535	k	62
536	コルチゾール	61
537	バランス	61
538	動き	61
539	日間	61
540	異物	61
541	皮膚疾患	61
542	給水	61
543	視覚化	61
544	言語	61
545	1 0 年以上	61
546	スマートフォン	60
547	メタ分析	60
548	構造	60
549	標準	60
550	T G	60
551	イソクエルシトリン	59
552	ヒペロシド	59
553	ワーファリン	59
554	味	59
555	睡眠中	59
556	計算	59
557	雌	59
558	高純度	59
559	切り口	58
560	推測	58
561	清涼飲料	58
562	生理学的指標	58
563	産物	58
564	集団の	58
565	1 9 8 0 年代	58
566	H	58
567	P U F A	58
568	ハーブ	57
569	亜慢性毒性試験	57
570	前述	57

571	国内の	57	607	生殖発生毒性試験	52
572	活性化	57	608	部分	52
573	睡眠時間	57	609	E F S A	52
574	腹部内臓脂肪	57	610	L	52
575	認知	57	611	下痢症状	51
576	30分	57	612	塩分	51
577	便性	56	613	抗凝血作用	51
578	単回投与	56	614	消費者庁	51
579	要素	56	615	食塩	51
580	りんご	55	616	骨代謝	51
581	厚生省	55	617	HD	51
582	容量	55	618	ヶ月	50
583	意欲	55	619	延長	50
584	網膜の	55	620	段階	50
585	LDLコレステロール	55	621	認識	50
586	P1	55	622	ST	50
587	VAS	55	623	厳密	49
588	仕事	54	624	品種	49
589	役割	54	625	最大血圧	49
590	痛み	54	626	最小血圧	49
591	誘発	54	627	植物	49
592	グリコ	53	628	研究報告書	49
593	トクホ	53	629	表	49
594	ヘマトコッカス藻	53	630	高年者	49
595	収載	53	631	1週間	49
596	層の	53	632	1950年代	49
597	文献情報	53	633	1970年代	49
598	昭和	53	634	ECF	49
599	孢子	53	635	平成8年度	48
600	調査研究	53	636	2週間以上	48
601	LD50	53	637	DBP	48
602	OS	53	638	F	48
603	Oligonol	53	639	SBP	48
604	収縮	52	640	スポーツ選手	47
605	外国語	52	641	一定の	47
606	微生物	52	642	交絡因子	47

643	外観	47	678	トレーニング	42
644	抗原性試験	47	679	柑橘類	42
645	着目	47	680	臨床栄養	42
646	血液凝固	47	681	試験期間	42
647	身体的活動	47	682	食品化学	42
648	Bacillus	47	683	120分	42
649	Bifidobacterium	47	684	3ヶ月間	42
650	グレア	46	685	SBT	42
651	リスクの評価	46	686	とうもろこし	41
652	一次予防	46	687	会議	41
653	野菜ジュース	46	688	協力	41
654	リスト	45	689	抗血小板薬	41
655	受け入れ	45	690	糞便中	41
656	必須アミノ酸	45	691	菌数	41
657	抗酸化物質	45	692	衛生局	41
658	毒性学	45	693	費用	41
659	米国食品医薬品局	45	694	鶏	41
660	約1	45	695	毒性	40
661	脱落	45	696	非ランダム	40
662	AI	45	697	食品安全管理	40
663	Health	45	698	Food	40
664	プロシアニジンB1	44	699	MPO	40
665	乳	44	700	アレルギー反応	39
666	別途	44	701	中高年女性	39
667	効果量	44	702	制限	39
668	医者	44	703	治療薬	39
669	技術	44	704	混合	39
670	認可	44	705	炭酸飲料	39
671	再生	43	706	胚芽	39
672	厚生科学	43	707	分布	38
673	吸収量	43	708	品質規格	38
674	層	43	709	流	38
675	総コレステロール	43	710	眼疲労	38
676	UMIN	43	711	調査対象	38
677	ことば	42	712	骨密度	38
			713	クロロゲン酸	37

714	ビデンス・ピローサ	37	750	健康増進	34
715	保健	37	751	定性	34
716	原データ	37	752	思い込み	34
717	発症リスク	37	753	気温	34
718	腹部皮下脂肪	37	754	血圧値	34
719	菌株	37	755	識別	34
720	調査報告	37	756	2000年	34
721	C T R	37	757	C O	34
722	E	37	758	N u t	34
723	R e	37	759	カプセル剤	33
724	中間	36	760	バリルチロシン	33
725	乾燥剤	36	761	収縮期圧	33
726	分離	36	762	甲殻類	33
727	増殖	36	763	糖尿病患者数	33
728	栄養調査	36	764	週間後	33
729	毒性作用	36	765	遺伝毒性試験	33
730	活力	36	766	1日摂取許容量	33
731	胃腸症状	36	767	6つ	33
732	資料	36	768	m i n	33
733	遅延	36	769	むくみ	32
734	運動後	36	770	エストロゲン受容体 $\alpha$	32
735	D e r m a t o l o g y	36	771	ドリンク	32
736	T M	36	772	ヒドロクロロチアジド	32
737	グルコシル	35	773	ビルベリー抽出物	32
738	プラズマローゲン	35	774	便	32
739	レモン	35	775	冷房	32
740	事務所	35	776	化学工業	32
741	作業負荷	35	777	危険性	32
742	抑制薬	35	778	卵黄	32
743	科学技術	35	779	味噌汁	32
744	経過	35	780	排出	32
745	血液凝固薬	35	781	沈香	32
746	評価方法	35	782	源	32
747	リンゴ	34	783	超	32
748	ルテインエステル	34	784	酸	32
749	位置	34	785	6-ショウガオール	32

786	6ージンゲロール	32	822	平均値	29
787	アメリカ食品医薬品局	31	823	推移	29
788	不足	31	824	栄養補助食品	29
789	定義	31	825	種子	29
790	新規	31	826	表面	29
791	真の	31	827	おそれ	28
792	紙	31	828	その間	28
793	美	31	829	みかん	28
794	20名	31	830	アムロジピン	28
795	3年間	31	831	エナラプリル	28
796	8つ	31	832	バルサルタン	28
797	ヘスペレチン	30	833	ロサルタン	28
798	充実	30	834	冷凍食品	28
799	反復	30	835	単盲検	28
800	反応	30	836	多岐	28
801	尿中デオキシピリジノリン	30	837	大豆食品	28
802	尿検査	30	838	寒天	28
803	抗酸化作用	30	839	漬物	28
804	摂取不足	30	840	管理方法	28
805	血液検査	30	841	胃腸障害	28
806	軽症高血圧	30	842	脱落例	28
807	除去	30	843	遵守	28
808	食品衛生法	30	844	関節機能	28
809	高年齢	30	845	食後高脂血症	28
810	html	30	846	75g	28
811	nih	30	847	LDLーコレステロール	28
812	ガ	29	848	エネルギー代謝	27
813	クチナシ	29	849	ジエチルプロピオン	27
814	一致	29	850	デンブン	27
815	中年期	29	851	ピーク値	27
816	反復投与毒性試験	29	852	二重盲検試験	27
817	回分	29	853	処方	27
818	増幅	29	854	抗凝固作用	27
819	大切	29	855	種々の	27
820	対照試験	29	856	HP	27
821	導入	29	857	SR	27

858	さけ	26	894	2種類	25
859	はじめとする	26	895	9年間	25
860	アロエ	26	896	B C A A	25
861	ビフィドバクテリウム	26	897	S	25
862	フェンテルミン	26	898	プソイドエフェドリン	24
863	リラックス	26	899	リスク因子	24
864	包括	26	900	ロイテリ菌	24
865	強化	26	901	厚生労働大臣	24
866	成人女性	26	902	実験	24
867	歯科治療	26	903	慢性疲労	24
868	溶出	26	904	摂取基準	24
869	理解	26	905	渡り	24
870	眼の	26	906	週目	24
871	累積	26	907	陰性	24
872	良品	26	908	食品規格	24
873	調査票	26	909	1990年	24
874	魚肉ソーセージ	26	910	5段階	24
875	1例	26	911	D I	24
876	1日1回	26	912	S c i e n c e	24
877	10年間	26	913	W	24
878	2例	26	914	w	24
879	5, 7-ジメトキシフラボン	26	915	たんぱく質	23
880	7つ	26	916	エピガロカテキン-3-O- (3-O-メチル) ガレート	23
881	D p d	26	917	バリア	23
882	I g A	26	918	作業従事者	23
883	N u t r i t i o n	26	919	危険	23
884	カプトプリル	25	920	室	23
885	クロスオーバー試験	25	921	心の健康	23
886	ミルク	25	922	目付け	23
887	体脂肪低減作用	25	923	標準化	23
888	同一性	25	924	結晶	23
889	同意	25	925	肉体	23
890	寒冷	25	926	蒟蒻	23
891	料理	25	927	農林水産省	23
892	温州みかん	25	928	食前	23
893	食品医薬品局 (F D A)	25			

929	C o Q 1 0	23	965	最小	21
930	か否か	22	966	果汁	21
931	だるさ	22	967	燕	21
932	の発作	22	968	用語	21
933	アジア人	22	969	皮膚表面温度	21
934	カフェイン	22	970	移行	21
935	コンセンサス	22	971	競輪	21
936	プロバイオティクス	22	972	缶詰	21
937	不安	22	973	自己	21
938	妥当性と信頼性	22	974	血圧降下作用	21
939	手足	22	975	表題	21
940	抗痙攣薬	22	976	都道府県	21
941	環境改善	22	977	関係者	21
942	血糖降下剤	22	978	1年間	21
943	賦形剤	22	979	2年間	21
944	農業	22	980	20年	21
945	酸化	22	981	3種類	21
946	食品医薬品局	22	982	60代	21
947	1日目	22	983	ARA	21
948	3例	22	984	AS	21
949	BL	22	985	PI	21
950	$\alpha$ -グルコシダーゼ	21	986	でんぷん	20
951	うんし	21	987	はっきり	20
952	まず	21	988	ウエストサイズ	20
953	ギムネマ酸	21	989	エピガロカテキンガレート	20
954	シクロ	21	990	キリン	20
955	システマティック・レビュー	21	991	サッカー	20
956	ドラッグストア	21	992	人種差	20
957	ヒット	21	993	体脂肪率	20
958	リスク評価	21	994	処置	20
959	各群	21	995	小核試験	20
960	培養	21	996	心臓疾患	20
961	安全性調査	21	997	必須	20
962	安全管理	21	998	意見	20
963	悪化	21	999	感覚	20
964	抑制の	21	1000	植物化学	20

1001	母乳栄養児	20	1037	R i s k	19
1002	注視	20	1038	におい	18
1003	相関	20	1039	アイ	18
1004	糖尿病薬	20	1040	エネルギー不足	18
1005	記憶の	20	1041	コロソリン酸	18
1006	試験管内	20	1042	データベース	18
1007	酵素の	20	1043	ラクトバチルス・ガセリ	18
1008	食あたり	20	1044	乳製品	18
1009	食品分析	20	1045	体重60kg	18
1010	高脂肪食	20	1046	倫理	18
1011	5年間	20	1047	対象部位	18
1012	C l i n i c	20	1048	推奨量	18
1013	HDLコレステロール	20	1049	構成割合	18
1014	u	20	1050	機構	18
1015	インゲン	19	1051	歩行時	18
1016	エビ	19	1052	比率	18
1017	不適格	19	1053	溶解	18
1018	体調不良	19	1054	濃縮	18
1019	加熱	19	1055	第1号	18
1020	単回投与毒性試験	19	1056	糖尿病患者	18
1021	外国人	19	1057	腰の	18
1022	栄養成分	19	1058	腹部膨満感	18
1023	桑	19	1059	色調	18
1024	構成成分	19	1060	茶系飲料	18
1025	用量依存性	19	1061	褐色脂肪	18
1026	発芽	19	1062	質の改善	18
1027	細胞の	19	1063	足の筋	18
1028	腸内フローラ	19	1064	軽減した	18
1029	臨床効果	19	1065	酢	18
1030	血圧降下剤	19	1066	酸乳	18
1031	行動	19	1067	防除	18
1032	軽微	19	1068	3月4日	18
1033	食品性状	19	1069	C P I	18
1034	高齢化	19	1070	O r g a n i z a t i o n	18
1035	高齢社会	19	1071	S p o r t s	18
1036	I n g e s t i o n	19	1072	T B C	18

1073	a m p	18	1109	カゼイン	16
1074	b l i n d	18	1110	ジャガイモ	16
1075	s k i n	18	1111	バラ科	16
1076	えび	17	1112	モニター	16
1077	かに	17	1113	中央	16
1078	せき	17	1114	冬期	16
1079	ほうれん草	17	1115	妊婦	16
1080	ジケトピペラジン	17	1116	平成5年	16
1081	一般化可能性	17	1117	心身機能	16
1082	両方	17	1118	所見	16
1083	乳児の健康	17	1119	搾汁	16
1084	低下傾向	17	1120	日本語論文	16
1085	医薬局	17	1121	気分障害	16
1086	吸収性	17	1122	牛	16
1087	大学生	17	1123	知能	16
1088	数の	17	1124	糖尿病治療薬	16
1089	最高	17	1125	規制	16
1090	歩数	17	1126	負荷食	16
1091	死因	17	1127	赤	16
1092	熟成	17	1128	連用	16
1093	発がん	17	1129	30歳	16
1094	短縮	17	1130	40名	16
1095	筋肉の	17	1131	90名	16
1096	経口免疫寛容	17	1132	G o o g l e	16
1097	脳血管疾患	17	1133	p p	16
1098	薬用植物	17	1134	かゆみ	15
1099	距離	17	1135	つま先	15
1100	醤油	17	1136	アスナロ	15
1101	4以上	17	1137	ガラクタン	15
1102	B. l o n g u m	17	1138	セサミン	15
1103	G S	17	1139	ランダム割り付け	15
1104	H y a l u r o n a n	17	1140	ランダム化試験	15
1105	K a w a	17	1141	下肢筋力	15
1106	h a	17	1142	体重50kg	15
1107	あん	16	1143	均質性	15
1108	アスピリン	16	1144	抗血小板剤	15

1145	拡張期圧	15	1181	外科手術	14
1146	改善傾向	15	1182	季節	14
1147	有限	15	1183	復帰突然変異	14
1148	水準	15	1184	急速	14
1149	発酵	15	1185	情報収集	14
1150	盲検化	15	1186	成人男性	14
1151	糖尿病の	15	1187	氷	14
1152	糖類	15	1188	汚染	14
1153	緊張	15	1189	生活環境	14
1154	臨床検査値	15	1190	生理学的	14
1155	臭い	15	1191	生鮮食品	14
1156	血糖降下薬	15	1192	異常の	14
1157	見解	15	1193	発酵食品	14
1158	陽	15	1194	皮膚アレルギー	14
1159	食事直後	15	1195	社製	14
1160	高温	15	1196	系統	14
1161	1ヶ月間	15	1197	能	14
1162	12月末	15	1198	菓子	14
1163	2回目	15	1199	藻類	14
1164	2月末	15	1200	血中コレステロール値	14
1165	FDA	15	1201	補完代替医療	14
1166	L. g a s s e r i	15	1202	視界	14
1167	L i p i d s	15	1203	視覚の	14
1168	R e g i s t r y	15	1204	観察期間	14
1169	U	15	1205	認知症患者	14
1170	f o o d	15	1206	身体活動	14
1171	クレモリス菌	14	1207	閉経後の	14
1172	グレード	14	1208	降圧作用	14
1173	ブレンド茶	14	1209	非直線性	14
1174	モデル	14	1210	食品工業	14
1175	世代	14	1211	食品衛生	14
1176	付着	14	1212	200万	14
1177	体積	14	1213	T h	14
1178	保健所	14	1214	l i p i d s	14
1179	冷水	14	1215	イミノ	13
1180	塩酸塩	14	1216	イワシ	13

1217	カニ	13	1253	30名	13
1218	ステージ	13	1254	4例	13
1219	タンパク質	13	1255	4週間後	13
1220	ネオコタラノール	13	1256	500個	13
1221	ノンレム睡眠	13	1257	65歳未満	13
1222	ファセオラミン	13	1258	7年間	13
1223	中途覚醒	13	1259	ALA	13
1224	予定	13	1260	C1	13
1225	二重盲検ランダム化比較試験	13	1261	Cr	13
1226	付加	13	1262	S-	13
1227	体脂肪量	13	1263	SAFETY	13
1228	保健の	13	1264	$\alpha$ 波	12
1229	優位	13	1265	たもぎ茸	12
1230	内1	13	1266	の予防	12
1231	凝血	13	1267	アルミ	12
1232	外の	13	1268	カロテン血症	12
1233	大量	13	1269	キウイ	12
1234	完了	13	1270	グリコサミノグリカン	12
1235	形成	13	1271	コントラスト	12
1236	抗血小板	13	1272	サラシア属	12
1237	拡大	13	1273	ダイエット	12
1238	流出	13	1274	ファゴミン	12
1239	破損	13	1275	一日許容摂取量	12
1240	美肌効果	13	1276	予想	12
1241	舞茸	13	1277	分化	12
1242	薬事	13	1278	制御	12
1243	薬物代謝酵素	13	1279	労働者	12
1244	調理	13	1280	十分量	12
1245	起源	13	1281	単語	12
1246	農産物	13	1282	唾液の	12
1247	運動習慣	13	1283	崩壊性	12
1248	酸味	13	1284	打錠	12
1249	錠	13	1285	抗酸化能	12
1250	1か月	13	1286	提示	12
1251	12か月	13	1287	更年期の	12
1252	1960年代	13	1288	最低血圧	12

1289	最高血圧	12	1325	にんにく	11
1290	標準体重	12	1326	アレルギー	11
1291	機能性食品	12	1327	ビスクロン	11
1292	歳未満	12	1328	一定時間後	11
1293	気分	12	1329	予備	11
1294	清潔	12	1330	二重盲検法	11
1295	発酵乳製品	12	1331	人参	11
1296	相同性	12	1332	冷凍	11
1297	短期記憶	12	1333	前提	11
1298	研究も	12	1334	加糖	11
1299	置き換え	12	1335	効果指標	11
1300	腹部肥満	12	1336	受け付け	11
1301	薬物	12	1337	同等以上	11
1302	血液凝固阻止薬	12	1338	学識経験者	11
1303	行政	12	1339	想起	11
1304	製品の回収	12	1340	把握	11
1305	評価基準	12	1341	有機	11
1306	輸出	12	1342	残留農薬	11
1307	連合	12	1343	治療方法	11
1308	適応	12	1344	測定データ	11
1309	選択バイアス	12	1345	甘味料	11
1310	遺伝子突然変異	12	1346	相関性	11
1311	重量	12	1347	研究内容	11
1312	間隔	12	1348	移植	11
1313	青年	12	1349	習慣	11
1314	高齢化社会	12	1350	脳内の	11
1315	鮭	12	1351	薬局	11
1316	1回目	12	1352	血液凝固抑制剤	11
1317	10週間	12	1353	製薬	11
1318	8例	12	1354	資金提供	11
1319	GE	12	1355	連携	11
1320	III	12	1356	連結	11
1321	LC	12	1357	運動機能	11
1322	Scale	12	1358	関心	11
1323	UK	12	1359	香料	11
1324	method	12	1360	2月3日	11

1361	20日	11	1397	樹皮	10
1362	Joint	11	1398	正常範囲内	10
1363	MMS E	11	1399	注釈	10
1364	QS	11	1400	漢方薬	10
1365	しわ	10	1401	炭酸水	10
1366	たるみ	10	1402	特許	10
1367	アセトアミノフェン	10	1403	生存	10
1368	アメリカ人	10	1404	監視	10
1369	ガム	10	1405	研究レベル	10
1370	キャップ	10	1406	肥満指数	10
1371	クラス	10	1407	脂質の過酸化	10
1372	クリプトキサンチン	10	1408	腸内の	10
1373	タカ	10	1409	腸内ガス	10
1374	チューインガム	10	1410	花粉	10
1375	ナシ	10	1411	若年	10
1376	フィルム	10	1412	財団	10
1377	フラバノール	10	1413	通過時間	10
1378	ミカン	10	1414	食糧農業機関	10
1379	ラベル	10	1415	1人当たり	10
1380	リスクファクター	10	1416	2錠	10
1381	一日摂取許容量	10	1417	20歳	10
1382	乳酸	10	1418	30分後	10
1383	代替	10	1419	4週間以上	10
1384	体の	10	1420	5例	10
1385	分包	10	1421	50歳以上	10
1386	化成	10	1422	C2	10
1387	取り上げ	10	1423	HANDBOOK	10
1388	同濃度	10	1424	I-1	10
1389	唾液αアミラーゼ	10	1425	II型コラーゲン	10
1390	国籍	10	1426	MS	10
1391	増加傾向	10	1427	STAI	10
1392	室温	10	1428	stress	10
1393	小豆	10			
1394	心理ストレス	10			
1395	抗肥満薬	10			
1396	教育	10			

灰色：難解語候補から除外した用語