付録資料:「テキストマイニングのための1・0データファイルの作成手順」

# テキストマイニングのための1・0データファイルの作成手順

東北農業研究センター 総合研究部

# 磯島昭代

# 目次

1.はし	じめに	••	•	•	•	1
2.本書	書の解説内容					
(1)	) 本書の解説範囲と取り扱うデータの種類	••	•	•	•	2
(2)	)1・0データファイル作成手順の概要	••	•	•	•	4
(3)	)マクロの内容	• •	•	•	•	5
3 . デ-	- タファイル作成手順					
(1)	) 文章を形態素解析ツールで分解する	••	•	•	•	7
(2)	) 形態素に文章番号を付与する	••	•	•	•	10
(3)	) 品詞情報を利用してキーワード候補の抽出を行う	••	•	•	•	12
(4)	) 基本形の「ひらがな」を「カタカナ」に変換する	••	•	•	•	15
(5)	) 形態素番号と出現数を割り当てる	••	•	•	•	18
(6)	) 出現数からキーワードを絞り込む	••	•	•	•	21
(7)	)「1・0 データ化」の準備	••	•	•	•	26
(8)	)マクロの実行	••	•	•	•	30
(9)	)1.0データファイルの完成	• •	•	•	•	34
4.補足	2説明					
(1)	)2つ以上のキーワードをまとめる場合	••	•	•	•	38
(2)	) マクロを使わずに1・0 データファイルを作成する方法	• •	•	•	•	42
5.む?	すび	••	•	•	•	50
付録	マクロのフローチャート		•	•	•	51
付録	マクロのコード	••	•	•	•	52
付録	「茶坊主くん.txt」からVBファイルを作成する	••	•	•	•	56

1.はじめに

アンケート調査などで得られる自由記述回答文やクレームデータなどに含まれる文章データは、 非定型の定性的データであるためこれまでは計量的な分析手段がなく、活用されることもほとん どなかった。近年、こうした大量の文章データを電子化して計量的に分析するための手法として テキストマイニング<sup>注(1)</sup>が注目されるようになったが、同手法は未だ開発途上にある上、市販さ れている専用のソフトウエアは非常に高額であるため、誰もが利用できる状況にあるとはいえな い。

こうした現状を踏まえ、磯島(2002)はフリーソフトの形態素解析システム「茶筌(ちゃせん)」 <sup>注(2)</sup>と表計算ソフト「Excel2002」および統計処理ソフトを用いて、低コストで簡便にできるテキ ストマイニングの手法(以下、本手法)を紹介している。

本書では、このテキストマイニングの過程で必要なデータファイル作成の手順を説明する。上 記の通り、本手法はフリーソフトと一般的に使用されている表計算ソフトおよび統計処理ソフト を利用しているため、新たな投資を必要とせず、低コストで分析を行うことができる。一方、高 度な専門知識を要さないシンプルな手法であるため、簡便ではあるがその作業手順は若干繁雑で ある。基本的には Excel のコピー&貼り付け機能、オートフィルタ機能、データの並べ替え機能、 簡単な関数などでほとんどの作業を行い(一部 Word を利用)、単純作業の繰り返しが多くなる

「1・0データ化」の部分だけは簡単なマクロプログラム(以下、マクロ)を組んでいる<sup>注(3)</sup>。

本手法は試行段階であるため、最小限の作業しかマクロを組んでおらず、汎用性に乏しい。従 って、その前段階であるキーワードの抽出において、データの位置やシートの名前などは既定の ものとしている。そのため、「とりあえずやってみよう」という方は、これからお示しする手順に 沿って作業をしていただきたい。一方、マクロの知識をお持ちの方は、ご自由に使い勝手の良い ように書き換えていただければと考えている。

なお、本マクロにおいて不具合を発見された方は、ご一報いただければ幸いであるが、筆者は プログラミングの初心者であり、必ずしも対処できるとは限らないのでその点はご了承いただき たい。また、本マクロを使用して発生したいかなる損害に対しても責任を負いかねることをここ に明記しておく。

注(1) テキストマイニングの概要については、市村・長谷川・渡部・佐藤(2001)を参照のこと。 注(2)奈良先端科学技術大学院大学自然言語処理学講座が提供しているフリーソフトウエア。詳細は

http://chasen.aist-nara.ac.jp/index.html.ja を参照のこと。

注(3)筆者が使用したパソコンの環境は以下の通りである。OS:Microsoft Windows XP SP1、メモリ: 1 GB、Excel2002、Word2002。 2.本書の解説内容

(1)本書の解説範囲と取り扱うデータの種類

筆者が想定しているテキストマイニングのプロセスと、本書において解説する範囲について説 明する(図1)。テキストマイニングのプロセスは、大きく、 文章データの入手、 1・0デー タファイルの作成、 データの分析、の3段階に分けることができる。このうち、本書で扱う範 囲は、 の「1・0データファイルの作成」を中心とする部分である。

文章データの入手および分析については、本書では解説の対象外とするが<sup>注(4)</sup>、取り扱う文章 データの種類について若干触れておく。

文章データについては、アンケート調査の自由記述回答文、クレームデータの他、Web 上の掲 示板など、様々な入手方法が考えられる。また、データの種類も、「 についてご自由にお書き 下さい」という非定型の自由記述文から、「 は、( )なので、( )だと思う」など のように、文章の一部を空白にしたフォーマットを用意し、その空白を埋めてもらう形の定型自 由文など様々である<sup>注(5)</sup>。

本書で取り上げるテキストマイニング手法では、分析する文章データとして、

文章データの1単位(例えば、1人の発言、1クレームなど)が Excel の1つのセルに入力 できる程度の分量であること、

上記単位の文章が、統計的処理を行うに相当する分量だけあること、

を原則としている。そのため、論文1本、議事録の1議題などの長文は想定外としているが、比較的短い文章もしくは単語が大量に存在しているような文章データであれば、定型・非定型によらず基本的に適用可能であると考えている(図2)。

注(4)データの分析については、磯島(2002) 磯島(2004) 磯島・野中・清野(2004)を参照のこと。 注(5)林(2002)は定型自由文へのテキストマイニングの適用事例を紹介している。





M M	icros	oft Exce	el - <del>an -</del>				1	
	ファイル	,(F) 編集	<sub>≢(E)</sub> 1	人の回	答または		ータ(D) ウィンドウ(W) ヘルプ(H)	
読ん	,de!!]]	](K)	1	件のク	レームなどが	r		_ 8 ×
D I	<u> -</u>	. 🔒 🖷				7117		100% - »
	A1		• • •	フのセ		1(115		•
	A	В	С	D			E	<b></b>
1	No	code1	code2	code3			記述文	
2	1	2	1	2	同じ物でも、とて	も高く感じる	る時は手頃な青菜にしていま	₫. <b>&gt;</b>
3	2	3	4	2	安くて、新鮮です	安全なもの		
4	3	3	6	2	旬の時期になる	べく購入し	旬でない時は売っていても他	の野菜で
5	4	3	3	1	新しい種類が売	っていても、	、利用のしかたが分からなくで	て買わない
6	5	3	3	2	生協にある青菜	ならば(野)	菜全般にも思っていることだた	が)安くても
7	6	2	4	2	青菜類は100円	程度の価格	8の時に購入するので、冬はる	あまり購入
8	<u></u>	۔ حج حد ر		2	新鮮でなるべく!	県産品を購	入	
9	因	い又草	または	2	家族が少ないと	:一束が多す	「ぎてすぐ葉が萎びてしまうの	で、少量
10	畄	諈が	大量に	1	ほぼ毎日、必ず	食べる様に	している。	
11			八主に	4	袋に入っている	けれど、葉の	Dほうは空気にふれているの	<u>で、傷み</u>
12	入	力され	ている	2	時期にあったも	のを購入し1	ています。その方が美味しく、	また、安い
13				2	行者にんにくを「	訂において	ましい	
14	13	4	3	2	新鮮で無農薬で	であること		
15	14	3	5	2	安さ 安全 地物	勿		
16	15	2	4	2	栄養価のことを:	気にして、意	意識して購入しています。(カリ	1ウム、カ/
17	16	3	4	2	特にありません			
18	17	2	4	₽	やはり、気になる	るのは、農薬	<u> 築。もっと、はっぎりと、分かり</u>	やすい、 🛃 🚽
H 4	► ►	(調査結					· · · · · · · · · · · · · · · · · · ·	•
עדב	۲						NUM	

図2 基本となるデータファイル

(2)1・0データファイル作成手順の概要

1 ・ 0 データファイルの作成手順の詳細については次章で述べることにし、ここでは簡単に概要を説明する(図3)。

- (A)まず、文章データを形態素に分解する。これには、フリーソフトウエアの形態素解析ツ ール「茶筌」を用いる。
- (B)次に、分解された形態素が、どの文章に含まれていたかを明らかにするために、各形態 素に文章番号を付与する。
- (C) 続いて、各形態素の品詞情報から、キーワードの候補となる形態素を抽出する。
- (D)抽出した形態素は、基本形が同じである語をまとめて、形態素番号と出現数を割り当て る。
- (E) この出現数を基準にして、キーワードの絞り込みを行う。
- (F)抽出したキーワードと各文章番号とを関連づけるために、マクロを使用して「1・0デ ータ化」を行う。
- (G)最後に、元の文章データと完成した1・0データを結合する。



図3 1・0 データファイルの作成手順

(3)マクロの内容

本書で紹介する1・0データファイルの作成手順では、抽出したキーワードが各文章に出現す るか否かを示す「1・0データ化」の部分にのみマクロを使用している。ここでは、そのマクロ の内容について簡単に説明する。なお、巻末にマクロのフローチャートとコードを掲載している ので、詳細についてはそちらをご覧いただきたい。

前述のように、マクロを実行する前段階として、各形態素にはどの文章に出現した語かを示す 文章番号を付与しており、さらに同じ基本形をもつ形態素には、同じ形態素番号を割り当ててい る。従って、それぞれの文章に含まれる形態素の番号と、キーワードとして抽出した形態素番号 (以下、「キーワード番号」)とを照合することにより、各文章における各キーワードの出現の有 無を調べることができる。

本書で用いるマクロでは、まず文章番号の小さい順に1つの文章を選び、その文章に出現する 形態素番号を小さい順に1つずつキーワード番号と照合していく(図4)。キーワード番号と出現 形態素番号が等しい場合には1を出力し、次のキーワード番号および出現形態素番号に移動する。 キーワード番号が出現形態素番号よりも大きい場合には、次の出現形態素番号と照合する。出現 形態素番号がキーワード番号と等しくならずにキーワード番号よりも大きい数値を示したら、そ のキーワード番号に対応するセルには0を出力し、次のキーワード番号に移る。

これを、もう少し具体的に説明する。例えば文章1に出現した形態素番号が「1、2、3、5、

7、8、12、抽出したキーワード番号が「2、4、8、10」であったとする(図5)。 まず、形態素番号「1」とキーワード番号「2」を照合する。 「形態素番号」<「キーワード番号」なので、形態素番号「2」に移る。 形態素番号「2」とキーワード番号「2」を照合する。 「形態素番号」=「キーワード番号」なので「1」を出力する。 形態素番号「3」、キーワード番号「4」にそれぞれ移る。 形態素番号「3」とキーワード番号「4」を照合する。 「形態素番号」<「キーワード番号「4」を照合する。 「形態素番号」>「キーワード番号「4」を照合する。 「形態素番号」>「キーワード番号「4」を照合する。 「形態素番号」>「キーワード番号」なので、「0」を出力する。 キーワード番号「8」に移る。 形態素番号「5」とキーワード番号「8」を照合する (以下略)。

🛛 Microsoft Excel - 形創	整索.xls		
😰 ファイル(E) 編集(E) 表	示(⊻) 挿入Φ 書式(Q) ツ	ール(II) データ(II) ウィンドウ(W)	ヘルプ(王)
			_ 8 ×
🗈 🖆 🔛 🔒 🖏 🎒 🕻	à 🖤 🐰 🖻 🛍 • 🚿 🗠	• 🖙 + 🍓 Σ • f <sub>×</sub> 🛔 🖁	🗼 🛍 🚜 100% 🔹 😤
E2 🔻	fx 1		
A B	C D	E F í	Э Н 🔺
1 表層語 ▼基本形	▼カタカナ ▼品詞 🛛 ▼	記述Now 形態素 出現	文章に出現した
2 高く 高い	高イ 形容詞-自	1 12	又半に ユ 光 さ に
<u>3しする</u>	スル	1 29	形態系の番号
4 感じる 感じる			1 1
0 百米 百米 6 壬頃 壬頃	- 「日米 - 石町一版 - 毛頃 - 夕詞形図		1 10
7 安く 安い			1 12
8 安全 安全	安全名詞一形容	2 229	1 5
9 新鮮 新鮮	新鮮 名詞-形容	2 236	1 17
10 し する	スル 動詞-自立	3 →音1にけ	9
		3	<b>`</b>
12 なる なる 13 高	- アル - 朝詞 目立 高山 - 静語 白立	<u> </u>	、189、235 🖁 🔤
13 90つ 元回	一 一 助った日立 一 助った日立 一 助った日立	3 の5つの形	態素が ち
15 代用 代用	代用 名詞-サ変	3 + + +	7 1
<u>16</u> 旬 旬	旬 名詞-一般	3 宮まれてい	ବ 3
17 旬 旬	旬 名詞-一般	3 174	2 2
18 他 他	他名詞一一般	3 199	1 2 -
◀ ▶ ▶   \形態素 / 形態	素②/抽出語) <u>data</u> /出現	枚/3以上/ <i>1</i> ·0/	
אעדב א			UM
Microsoft Excel - 形態	≩≢,.xls		
■ Microsoft Excel - 形態 ■ ファイル(E) 編集(E) 表 読んde!!ココ(K)	<mark>終了:xls</mark> 示① 挿入印 書式の ツ・	-ルロ データロ ウィンドウミン	
Microsoft Excel - 形態 ■ ファイル(E) 編集(E) 表 読んde!!ココ(K) □ 22 日 高 戦 6 20 0	示① 挿入① 書式② ツ い い ♥ よ 12 ♂ い	-ルロ データロ ウィンドウW	
<ul> <li>Microsoft Excel - 形態</li> <li>アァイル(E) 編集(E) 表 読んde!!ココ(K)</li> <li>ご</li></ul>	k w 挿入印書式の ッ・ ★ ♥	-ルロ データロ ウィンドウW ・ 29 ・ 49 - 49	
<ul> <li>Microsoft Excel - 形態</li> <li>ファイル(E) 編集(E) 表 読んde!!ココ(K)</li> <li>ご ご 日 ご (K)</li> <li>MS Pゴシック</li> <li>11</li> <li>A94</li> </ul>	k → x ls 示① 挿入① 書式の ツー 、 **	-ルロ データロ ウィンドウW ・ つ 29 ・ 49 國 写 % ぷ ぷ 年	
<ul> <li>Microsoft Excel - 形線</li> <li>ファイル(E) 編集(E) 表 読んde!!ココ(E)</li> <li>ロ ご (D) ご (D) (D) (D) (D) (D) (D) (D) (D) (D) (D)</li></ul>	k7,x1s 示① 挿入① 書式の ツ ・	-ルロデータロウインドウW 29 - 49 國 9 % かった	
<ul> <li>Microsoft Excel - 形線</li> <li>アァイル(E) 編集(E) 表 読んde!!ココ(K)</li> <li>ロ 谷 岡 谷 〇</li> <li>MS Pゴシック</li> <li>A94</li> </ul>	A ☆	-ル① デーカ② ウィンドウW - 29 - 49 國 愛 % ぷ ? ? 年 12 13	
X Microsoft Excel - 形態 ③ ファイル(E) 編集(E) 表 読んde!!ココ(K) D 2 日 合 包 の 0 MS Pゴシック 11 A94 A 4 1 2 1	A → ×Is 示(火) 挿入(2) 書式(2) ッ ・ B J U ・ B J U ・ B J U ・ B J U ・ 12 ・ D 10 0 0 0	-ル① データ ウィンドウ W 29 49 図 9 % 2010 12 13 1 0	
<ul> <li>■ Drfル(E) 編集(E) 表 読んde!!コ(K)</li> <li>□ ご 目前の目前の目前の目前の目前の目前の目前の目前の目前の目前の目前の目前の目前の目</li></ul>	A → XIS 示① 挿入① 書式② ツ B I U F 0 0 0 0	-ルロ データロ ウィンドウW 29 - 49 図 9 % ぷ ぷ モ 12 13 1 0	
<ul> <li>Microsoft Excel - 形線</li> <li>アァイル(E) 編集(E) 表 読んde!!ココ(K)</li> <li>ロ ご 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日</li></ul>	xx1s 示 ☆ 挿入 @ 書式 の ッ ・ B / U ・ C - D ・ C - D	-ルロデータD ウィンドウW 29 - 49 国 9 % からう 12 13 1 0 - ド番号と、	
<ul> <li>Microsoft Excel - 形線</li> <li>ファイル(E) 編集(E) 表 読んde!!ココ(K)</li> <li>ご ご 日 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一</li></ul>	A ☆ #入 @ 書式 Ø ッ ★ ↓ 12 ↓ ↓ B J U ↓ ↓ 12 ↓ ↓ 7 9 10 0 0 0 0 +-ワ- タ ☆ デ	-ル① データ ウィンドウ W 29 · 49 図 9 % * * * * * 12 13 1 0 ・ * 番号と、	
<ul> <li>Microsoft Excel - 形線</li> <li>アァイル(E) 編集(E) 表 読んde!!コ(K)</li> <li>ご ご 記 記 記 記 記</li> <li>MS Pゴシック</li> <li>1</li> <li>A94</li> <li>A94</li> <li>4</li> <li>2</li> <li>1</li> <li>2</li> <li>4</li> <li>3</li> <li>5</li> <li>4</li> <li>6</li> <li>5</li> <li>7</li> <li>文 た 6</li> </ul>	k → x ls h () 挿入() 書式() ツ h () 挿入() 書式() ツ h () 目 () 日 h () H ()	-ル① データ ウィンドウ W 29 49 図 % ページ 症 12 13 1 0 - ド番号と、 ーに出現した キ	ヘルズビ 189 0 235 189 0 235 15 18 0 15 18 0 -ワード番号
<ul> <li>■ Dr(1)(E) 編集(E) 表 読んde!!コ(K)</li> <li>□ ご 目前 目前 目前 目前</li> <li>▲ A94</li> <li>▲</li></ul>		-ルロ データD ウィンドウW 29 49 国 9 % 20 5 12 13 1 0 - ド番号と、 ーに出現した 等号を照合し、	ヘルズゼ 189 00 235 ・ ▲ ・ 、 15 18 0 - ワード番号
<ul> <li>■ ファイル(E) 編集(E) 表 読んde!!ココ(K)</li> <li>□ ご □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □</li></ul>	xx1s fxW 挿入Q 書式の ッ x ↓ 12 ↓ 12 x ↓ 12 ↓ 12 x ↓ 12 ↓ 12 x ↓ 12 ↓ 10 0 0 0 0 x ↓ 12 ↓ 10 0 0 0 0 x ↓ 10 x ↓	-ルロ データD ウィンドウW 29 49 国 9 % 20 10 年 12 13 1 0 - ド番号と、 - に出現した 雪号を照合し、 - の場合は 1	ヘルスHP 189 0 235 
<ul> <li>■ 7r1µ(E) 編集(E) 表 読んde!!ココ(K)</li> <li>□ 2 日 日 2 日 1</li> <li>▲ 84</li> <li>▲ 94</li> <li>▲ 95</li> <li>▲ 10</li> <li>➡ 9</li> <li>■ 4</li> <li>■ 5</li> <li>■ 4</li> <li>■ 5</li> <li>■ 7</li> <li>■ 7</li> <li>■ 8</li> <li>■ 10</li> <li>■ 9</li> </ul>	xx1s f W 挿入 @ 書式 Ø ッ x ♥ x 12 0 x ♥ x 10 0 x ♥ x 0 0 x 0 0 0 x ♥ x 0 0 x 0 0	-ル① データ ウィンドウ W 29 49 図 % 20 1 12 13 1 0 -ド番号と、 ーに出現した 音号を照合し、 音の場合は1、	<ul> <li>ヘルンH<sup>0</sup></li> <li>189 0 235</li> <li>・ ふ ・ ふ</li> <li>・ ふ ・ ふ</li> <li>・ ふ ・ ふ</li> <li>・ ふ ・ ふ</li> </ul>
<ul> <li>Microsoft Excel - 形線</li> <li>アァイル(E) 編集(E) 表 読んdel!ココ(K)</li> <li>ご ご 副 副 図 の (C)</li> <li>MS Pゴシック</li> <li>A94</li> <li>A94</li> <li>A94</li> <li>A94</li> <li>A94</li> <li>(C)</li> <li></li></ul>	<ul> <li>★</li> <li>★<td>-ル① データD ウィンドウW 29 49 図 % % で 12 13 1 0 ・ド番号と、 ーに出現した 音号を照合し、 の場合は1、 トは0とする。</td><td>ヘルズ世 189 0 235 ・ ▲ ・ ** 15 18 0 - ワード番号</td></li></ul>	-ル① データD ウィンドウW 29 49 図 % % で 12 13 1 0 ・ド番号と、 ーに出現した 音号を照合し、 の場合は1、 トは0とする。	ヘルズ世 189 0 235 ・ ▲ ・ ** 15 18 0 - ワード番号
<ul> <li>■ 7rイル(E) 編集(E) 表 読んde!!コ(K)</li> <li>□ 2 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日</li></ul>	<ul> <li>★</li> <li>★<td>-ル① データD ウィンドウW 29 49 国 9 % 200 年 12 13 1 0 - ド番号と、 ーに出現した 号を照合し、 号の場合は1、 トは0とする。</td><td>ヘルズH<sup>0</sup> 189 00 235 ・ ▲ ・ 、 15 18 0 - ワード番号</td></li></ul>	-ル① データD ウィンドウW 29 49 国 9 % 200 年 12 13 1 0 - ド番号と、 ーに出現した 号を照合し、 号の場合は1、 トは0とする。	ヘルズH <sup>0</sup> 189 00 235 ・ ▲ ・ 、 15 18 0 - ワード番号
<ul> <li>■ 7rイル(E) 編集(E) 表 読んde!!コー(K)</li> <li>□ 2 日 2 日 2 日 2 日</li> <li>▲ 94</li> <li>▲ 100</li> <li>■ 9</li> <li>■ 11</li> <li>■ 10</li> <li>■ 12</li> <li>■ 12</li> </ul>	<ul> <li>★ x ls</li> <li>★ ★ 12</li> <li>★ ↓ 12<td>-ル① データD ウィンドウW 29 49 国 9 % 2 10 年 12 13 1 0 - ド番号と、 - に出現した 等号を照合し、 その場合は1、 トは0とする。</td><td><ul> <li>189 0 235</li> <li>189 0 235</li> <li>15 18</li> <li>0</li> </ul></td></li></ul>	-ル① データD ウィンドウW 29 49 国 9 % 2 10 年 12 13 1 0 - ド番号と、 - に出現した 等号を照合し、 その場合は1、 トは0とする。	<ul> <li>189 0 235</li> <li>189 0 235</li> <li>15 18</li> <li>0</li> </ul>
<ul> <li>■ 7r1µ(E) 編集(E) 表 読んde!!コー(K)</li> <li>■ ○ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □ □ ○ □</li></ul>	A     F     XIS       F     (M)     挿入(0)     書式(0)       F     (M)     (M)       F     (M)	-ル① データD ウィンドウW 29 49 國 9 % 20 10 年 12 13 12 13 1 0 - ド番号と、 つに出現した 音号を照合し、 50場合は1、 トは0とする。	
<ul> <li>■ 7rイル(E) 編集(E) 表 読んde!!コ(K)</li> <li>□ ご □ ご □ ご □ ご □ ご □ ご □ ご □ ご □ ご □</li></ul>	A     F     X     Ja       F     M     M     M     M	-ル① データD ウィンドウW 29 49 国 9 % 200 年 12 13 1 0 -ド番号と、 中に出現した 音号を照合し、 音の場合は1、 トは0とする。 - た - た - た - た - に出現した - た - た - た - に出現した - た - た - た - た - の - に出現した - た - た - 一 - 一 - 一 - 一 - 一 - 一 - 一 - 一	<ul> <li>ヘルズビー</li> <li>189 00 235</li> <li>・ シーム・ **</li> <li>15 18</li> <li>0</li> </ul>
<ul> <li>Microsoft Excel - 形線</li> <li>アァイル(E) 編集(E) 表 読んdel!ココ(K)</li> <li>ロ ご 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日 日</li></ul>		-ルロ データD ウィンドウW 29 49 図 9 % 20 50 50 12 13 12 13 1 0 - ド番号と、 ーに出現した 音号を照合し、 合の場合は1、 トは0とする。 を照合する な/3以上、1.0	

図4 マクロの内容



図5 形態素番号とキーワード番号の照合手順

3.データファイル作成手順

(1) 文章を形態素解析ツールで分解する

本手法は、形態素解析ツール「茶筌」を利用している。「茶筌」は奈良先端科学技術大学院大学 自然言語処理学講座で開発された日本語形態素解析システムのフリーソフトであり、以下のサイ ト(画面1)で管理、配布されている。

#### http://chasen.aist-nara.ac.jp/

なお、本書では茶筌の説明は割愛する。詳しい内容や使用方法については、上記サイトおよび 林(2002)を参照すること。

分析する文章データは、あらかじめ Excel ファイルに入力しておく。ここでは、アンケート調査の結果(回答数92件)を事例としているが、1つの行につき1人分の文章番号(回答者番号、記述番号)、回答者属性、文章データを入力している(画面2)。クレームデータなども同様に、 1行に文章データとそれに関連するデータを入力する。このとき、文章データは原則として1つのセルに収める。

また、この文章データでは、「半角カタカナ」を使わないように注意したい。なぜならば、「茶 筌」で形態素解析を行う際に「半角カタカナ」があるとバグが発生し、一見すると普通に解析し ているようであっても、以降の手順に大きな障害が残るからである。既に入力されているデータ に半角カタカナが含まれている場合には、Word の「文字種の変換」機能(後述)などを用いて、 全角カタカナに変換しておくことが大切である。

まず、「茶筌」の初期画面を立ち上げる(画面3)。「クリア」ボタンをクリックして「文エリア」 を空にし、そこに画面2の文章データを一括してコピー&貼り付けする。「文エリア」の下にある 「表層語」「基本形」「品詞」にチェックを入れ、最後に「全文解析」をクリックする。

ここで、通常は画面下部の「形態素エリア」に解析結果が表示されるが、テキストマイニング を行うほどの文章量になると「解析結果が大きすぎて全体を表示できません。今すぐ結果を保存 しますか?」と聞かれるので、「OK」をクリックしてテキストファイルで保存する。

次に、保存したテキストファイル(事例では「形態素.txt」)を Excel で読み込む。形態素解析 の結果はテキスト保存してあるので、Excel でファイルを開く場合は「ファイルの種類(T)」を 「すべてのファイル(\*.\*)」にしてファイル名を選択する。さらに、「データファイルの形式を選 択して下さい」では、「カンマやタブなどの区切り文字によってフィールドごとに区切られたデー タ(D)」にチェックを入れる。

読み込んだデータは、A列が文章を形態素に分解した「表層語」(これを縦につなげると元の文章と等しくなる)を、B列は「基本形」を、C列はその「品詞」を示している。各列の名前を付けるために一番上に1行挿入し、それぞれの列に「表層語」「基本形」「品詞」と入力する(画面4)。

🗿 GhaSen Home Page - Microsoft Internet Explorer	
ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)	A.
🔾 戻る 🔹 💿 · 💌 😰 🏠 🔎 検索 🥎 お気に入り 📢 メディア 🍕	😢 🔗 🍓 🔜 🛄
アドレス(1)) 🍘 http://chasen.aist-nara.ac.jp/index.html.ja	Note: N
[English]	
形態素解析システム 茶筌	
<ul> <li> <u>新着情報</u> <u>Widows 版 chasen-233</u> リリース(2003/8/18)         chasen-233 リリース(2003/8/15)         ipadic-263 リリース(2003/8/15)         <u>茶筌均ま?         茶筌均ま? 茶筌均ま? <u>茶筌均配布         </u> <u>Manual Maling List <u>FAQ         </u> <u>LINK 茶筌に関する文献         </u> <u>関係者一覧         </u> </u></u></li> </ul>	
<u>Wiki版 ChaSen Web</u>	<b>~</b>
🧃 ページが表示されました	🥥 インターネット

画面1 「茶筌」ホームページ

🔀 м	icros	oft Exce								
<b>1</b>	ווליקר	. <u>(F)</u> 編集	€ その	つ他の	□ <sub>書式(0)</sub> 又草テータのセルを <sub>プ(H)</sub>					
ÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌÌ	音番·	묵.	数值	∎data	全て選択・コピーする					
= └┙ .		◨ ▣;ਾ ™	-7							
	AI		•/							
	A No.		U							
		codel	codez	codes	<u>にためん</u> 同い物でも、とても直く感じる時は手順が表意にしています。					
2	2	2	1	2	190初にも,としも同く感じる時は子頃な月米にしています。 年/て 新鮮で完全たもの					
4	- 2	3		2	Qへて、新鮮で Q 生き000 旬の時期になるべく購入し 旬でない時は売っていても他の野菜で					
5	4	3	3	1	新しい種類が売っていても、利用のしかたが分からなくて買わな					
6	5	3	3	2	生協にある青菜ならば(野菜全般にも思っていることだが)安くて					
76				2	青菜類は100円程度の価格の時に購入するので、冬はあまり購入					
8	回答	≦者1.	人につき	<b>x</b> 2	新鮮でなるべく県産品を購入					
9		ан · / /- ·»		2	家族が少ないと一束が多すぎてすぐ葉が萎びてしまうので、少量					
10	1	「行すう	つ入力	1	ほぼ毎日、必ず食べる様にしている。					
_14⊑				4	袋に入っているけれど、葉のほうは空気にふれているので、傷み・					
12	11	2	4	2	時期にあったものを購入しています。その方が美味しく、また、安、					
13	12	4	4	2	行者にんにくを店においてほしい					
14	13	4	3	2	新鮮で無農薬であること					
15	14	3	5	2	女き女生 地物 労業(からちまたに) オ 奇渉( オ購 )」 オレキキ (カレウエーキ)					
10	15	2	4	2	木食油のことで気にして、息減して購入しています。(カリソム、力) 時にありませす					
18	17	3	4	2	すにのシュヒル やすり 気になるのけ 農薬 もっと けっきりと 分かけやすい ヲ					
10		∠ ±≣‱±	·甲/	-						
עדב ו	• •  K		达/		NUM					

画面2 文章データのファイル



画面3 「茶筌」による解析手順

🔀 M	icrosoft Ex	cel - 形態	≅.txt	
<b>图</b> 読/	ファイル(E) 新 vde!!ココ( <u>K</u> )	扁集(E) 表示	(⊻) 挿入⊈ 書式(2) ツール(1)	データ(D) ウィンドウ(W) ヘルプ(H) - <b> </b>
	🖻 🖪 🔒	B 🖨 🖪	🖓 👘 🛍 🕶 🖉	$\bullet \bigotimes \Sigma \bullet f_{\mathbb{X}} \stackrel{\circ}{=} \left\{ \begin{array}{c} \Sigma \\ \downarrow \end{array} \right\} \stackrel{\circ}{=} \left\{ \begin{array}{c} \Sigma \\ \bullet \end{array} \right\} {=} \left\{ \begin{array}{c} \Sigma \\ \bullet \end{array} \right\}$
	H1	-	fx	
	Â			F G H
1	表層語	基本形	品詞	
2			建14词 夕詞_非白☆	
 	で	1/0 1/-	お前方	─ それぞれの列の頭に
5	t.	<u>,</u> こ	助詞-係助詞	
6	,	,	記号-読点	
7	とても	とても	副詞─助詞類接続	と入力する
8	高く	高い	形容詞-自立	
9	感じる	感じる	動詞一自立	
10	時	時	名詞-非自立-副詞可能	
11	lt	lt	助詞-係助詞	
12	手頃	手頃	名詞-形容動詞語幹	
13	な	だ	助動詞	
14	青菜	青菜	名詞一一般	
15	15	15	助詞-格助詞-一般	
16	し し	する	動詞-自立	
17	7	7	助詞-接続助詞	
18	เา	いる	動詞-非自立	
14 4	▶▶∖形態	氢/		
עדב	۴			NUM

画面4 形態素解析結果

(2)形態素に文章番号を付与する

ここで作成しようとしている1・0データファイルは、「どの文章にどのキーワードが含まれて いるか」を明示するためのものである。そのため、いったん分解された形態素がどの文章におい て出現した語であるかを明確にしておく必要がある。ここでは、各々の形態素に元の文章と対応 する文章番号(回答者番号、クレーム番号など。以降、説明画面では「記述 No」と表示)を付与 することにする。

「茶筌」で分解された形態素は、表層語として縦1列に連なっており、これを上から読んでい くと元の文章と等しくなる。そして、1つの文章が終了すると、その末尾に「EOS」と表記される。 従って、文章の区切りは、この「EOS」を用いることで判断することができる(画面5)。

文章番号の付与は「IF 関数」を利用する(画面6)。まず、1行目のセル(D1)には、文章番号を示す列であることがわかる名前をつける(事例では「記述No」)。次の行、すなわち最初の形態素の行のセル(D2)には、数値の「1」を入力する。そして次の行のセル(D3)には、「もし上の行の表層語が EOS ならば上の行の数値に1を加える。EOS でなければ上の行の数値と等しい」という条件式を入れる。[D3 の場合: IF(A2="EOS"、D2+1、D2)]

この条件式を最後の行までコピーする。ここで、最後の数値と文章の総数とが一致しているか を必ず確認する。もし、総文章数よりも大きい数値になっている場合には、「半角カタカナを使っ たことによるバグ」などで、途中に不要な「EOS」が紛れている可能性がある。

🔀 M	icrosoft Ex	cel - 形態詞	툕.txt					
₽ 読ん	ファイル( <u>F</u> ) ∦ ₀de‼コ⊒( <u>K</u> )	扁集( <u>E</u> ) 表示	(⊻) 挿入Ф 書式(型)	ツール(①	データ( <u>D</u> ) 「	→ → → → → → → → → → → → → → → → → → →	ルプ(出)	- <b>-</b>
	🖻 📕 🚔 I	1 de la	🂱 🐰 🖻 🛍 • 🚿	<b>N -</b> CI	- 🍓 Σ	列の名前で	をつける	*
	A	B	C C	D	E	F	G	
1	表層語	基本形	品詞	記述No	<u> </u>			
20		•	記号-句点					
21	EOS							
22		安い	形容詞 自立					
23	7	7	助司法税助司					
24	、 ≠ <b>F</b> &¥	、	記方-読点   夕詞  形の動詞語於		「FOS」と	r FOS L D	間が	
25	村川 開手    777		11111111111111111111111111111111111111			00] 00  		
27	安全	安全	名詞-形容動詞語幹		1つの文:	章を表す		
28	な	だ	助動詞					
29	200	もの	名詞-非自立-一般					
30	EOS			_				
31		旬	名詞一一般					
32	の n+ #n	の nt #n	助詞-連体化					
33	時期	時期	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~					
34	たる	たる	助命。哈奶前一板 新铜-白立					+
36	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	-a-2 -<	助动詞					<u>+</u>
	■ ■ ▼ 1	,					+	╧
עדב	F F (12)55	212/				NUM	1	

画面 5 文章番号の付与

<b>X</b> M	icrosoft Ex	cel - 形態	≅.xls			
8	ファイル( <u>E</u> ) 新	■集(E) 表示	── 挿入仰 書式(○)	ツール① データ② ウィ	ンドウW) ヘルプ(出)	_
:元A   []]	,de‼_]( <u>k</u> ) ⊯≟ □] _⊇	R A R	ッ 🕅 📾 📾 🗸 🛷 🛛		最初のセルは	• ×
	D19	<b>↓</b>	f <sub>★</sub> =IF(A18="EOS",	D18+1,D18)	数値の「1」	•
	А	В	С	DE	E G	<b>•</b>
1	表層語	基本形	品詞	記述No	2 行目以降は	
2	同じ	同じ	連体詞			
3	物	物	名詞-非自立-一般		IF 関数の条件式	
18	lγ	いる	動詞─非自立			
19	ます	ます	助動詞			7
20	$\frown$		記号-句点	1	EOS の有無で	_
21	EOS	)		1		_
22	$\overline{\mathcal{T}}$	安い	形容詞-自立	2		_
_23	7	7	助詞-接続助詞	2	判断する	_
_24	×	×	記号-読点	2 🕇	-	_
25	新鮮	新鮮	名詞-形容動詞語幹	2		
26	で	だ	助動詞	2		
27	安全	安全	名詞-形容動詞語幹	2		
28	な	だ	助動詞	2		
29	もの	もの	名詞-非自立-一般	2		
30	EOS			2		
31	旬	旬	名詞一一般	3		
32	Ø	0	助詞-連体化	3		
14 4	▶▶\				•	
עדב	۲				NUM	

画面6 文章番号の付与

(3)品詞情報を利用してキーワード候補の抽出を行う

文章を形態素に分解すると、膨大な数の形態素が検出される。これらの中には、句読点や括弧 などの記号や、「が」「を」などの助詞、「です」「ます」などの助動詞など、キーワードとして意 味を成さない語も多く含まれている。キーワードの抽出に際してこれらの語を1つ1つ検証する には大変な労力を要するので、不必要な語はある程度機械的に切り捨ててしまった方が効率的で ある。そこで、形態素の品詞情報を利用してキーワードの候補になりそうな語を抽出することに する。

まず、条件式で付与した文章番号が作業中に変化しないように、「数式」から「値」に変換して おく。画面7では、シートをコピーして新たに設け<sup>注(6)</sup>、「記述 No」の列を選択してコピー、「形 式を選択して貼り付け(S)」の「値(V)」にチェックして上書きしている。

続いて、オートフィルタ機能を利用する(画面8)。データのある適当なセルをクリックし、「デ ータ(D)」-「フィルタ(F)」-「オートフィルタ(F)」コマンドを選択する。1行目の各セルに「フ ィルタ矢印」がつくので、「品詞」の列の「フィルタ矢印」をクリックする。

リストには様々な品詞名が表示されるが、その中からキーワードになりそうな語を多く含んで いる品詞を選択する。ここでは、「形容詞 - 自立」「動詞 - 自立」「名詞 - サ変接続」「名詞 - 一般」 「名詞 - 形容動詞語幹」を選んでいるが、データの内容や分析の目的によっては他の品詞が必要 になることもあるので、個別に判断する<sup>注(7)</sup>。未知語の利用も考慮する必要がある。

ここで新たに「抽出語」シートを設ける。オートフィルタでキーワードとして使いたい品詞を 表示し、全ての行を選択してコピーし、「抽出語」シートに貼り付ける(画面9、画面10)。その 際、1つ目の品詞(例では「形容詞 - 自立」)は1行目(「表層語」「基本形」などの見出しの行) も含めてコピー&貼り付けをし、2番目の品詞からは、上の品詞に続けて形態素の行を貼り付け ていく。

注(6)以降、作業途中もしくはファイル完成後にファイルを保存する際には、以下の点に注意されたい。 テキストファイルを読み込み、新たなシートを設けた後でそのまま保存しようとすると「選択した ファイルの種類は複数のシートを含むブックをサポートしていません」というエラーメッセージが 出される。その場合、「ファイルの種類(T)」において「Microsoft Excel ブック(\*.xls)」を選択し、 Excel 形式で保存すること。

注(7)林(2002)は、「品詞ごとにキーワードとして適切かどうか」を一覧表にして示している。



画面7 キーワード候補の抽出



画面8 キーワード候補の抽出

🔀 Microso	ft Exce	- 形態素.	txt	+	7-	ドにかりる	そうか品部	ച്ച	. 🗆 🛛
🔊 ファイル(	E) 編集	€(E) 表示(V	) 挿入① 書式(_)	/- + \	ノ 88 + ロ		- ノ゚みぃぃぃ - <b>ナ</b> っ	-102	
読んde‼ココ	ß			行を1	<b>茜</b> //		-95		-8×
🗋 🗅 🚔 🔛	🔒 🖏	i 🖨 🗋 🚏	*   👗 🗈 🔁 • 🝼	50 + 0	i v	🍓 Σ 🕶 📠	Ż↓ X↓ 🛄	- 🔥 100	)% • *
A1		<b>•</b> ;	& 夫酒語					_	
	A	В	С		)	E	F	G	
<u>1</u> 表層	語 👤	基本形 💌	品詞	✓記述N	Jo 🔽	•			
<u>1931</u> 強い	1	強い	形容詞-自立		71				
2001 高い	1	高い	形容詞-自立		80	)			
<u>2034</u> 安い	1	安い	形容詞-自立		*	切り取り(工)			
<u>2067</u> 美明	記	美味しい	形容詞-自立		Ba	⊐ピ–(C)			
<u>2140</u> ない	1	ない	形容詞-自立			Etn(#(+/p)			
2154 長く		長い	形容詞-自立		ue i	8601910 \ <u>P</u> /			
<u>2159</u> 美明	紀けれ	美味しい	形容詞-自立			形式を選択して	貼り付け( <u>S</u> )		
2164 よく		よい	形容詞-自立			行の挿入の			
<u>2173</u> ない	1	ない	形容詞-自立			(二の背)服金(の)			
2219 多く		多い	形容詞-自立			1100H100A			
2233 良く		良い	形容詞-自立			数式と値のクリフ	7( <u>N</u> )		
<u>2243</u> 少な	th	少ない	形容詞-自立		P	セルの書式設定	E( <u>F</u> )		
<u>2254 &lt;</u> 61	.) С	くらい	形容詞-自立		_	行の高さ(R)			
<u>2255</u> 高く		高い	形容詞-自立						
2260 <mark>美</mark> 呀	むい	美味しい	形容詞-自立			表示しない(円)			
2269						<u> </u>			
2270									
H A P H	形態素	). 形態素 (2)	)/		1	ė –			
2267 レコード	中 69 個加	が見つかりました	io l	合計=32	79		NUM		

画面9 キーワード候補の抽出

<b>X</b> M	licrosoft Ex	cel - 形態	索.xls				
图 読/	ファイル(E) 新 {vde!!ココ( <u>K</u> )	扁集( <u>E</u> ) 表示	₩2 挿入Φ	書式(①) ツ	'−ルŒ	データ(型) ウィンド 新しいシート	<
	🖻 🖪 🔒	🔁 🎒 🗋	💱 🐰 🖻	🛍 • 🚿 🔺		$\neg \bigotimes \Sigma \cdot f_{\ast} \stackrel{\land}{\simeq} \downarrow \stackrel{?}{\simeq} \downarrow \stackrel{?}{\simeq} \downarrow \stackrel{?}{\simeq} \downarrow \stackrel{?}{\simeq} \downarrow \stackrel{?}{\simeq} \downarrow \stackrel{?}{\simeq} $	» ₹
	E71	-	fx				_
	Δ	В	C		E	F G H	•
1	表層語	基本形	品詞	記述No			
66	良く	良い	形容詞-自	91		1 行目(見出し行)を含め、	
67	少ない	少ない	形容詞一自	91		形容詞の行を貼り付ける	
<u>6</u> B	くらい	くらい	形容詞一目	92			
69	高く	高い	形容詞目	92	· · ·		
70	美味しい	美味しい	形谷詞日	92	<b></b>		
4	<u>感しる</u> し	感じる	動詞自立	1			
12	なる	サる	動詞 白立	3			
14	L	する	動詞-白立	3			
5	 売っ	売る	動詞-白立	3			
16	する	する	動詞-自立	3		すぐ下に続けて	
17	売っ	売る	動詞-自立	4		動詞の行を貼り付ける	
18	分から	分かる	動詞-自立	4			
19	買わ	買う	動詞-自立	4			
80	ある	ある	動詞-自立	5			
<u>s</u>	思っ	思う	動詞-自立	5	/		
82		5-3	動記白主	5			•
14 4	▶ ▶ \ 形態	素(形態素	() 抽出語/				
עדב	<i>י</i> ۲					NUM	

画面 10 キーワード候補の抽出

(4)基本形の「ひらがな」を「カタカナ」に変換する

ここで、基本形のひらがな表記をいったん全てカタカナ表記に変換する。これは、形態素の基 本形が何種類あるかを数える過程で、同じ語でもひらがな表記の語とカタカナ表記の語は、別の 語としてカウントしてしまうからである。従って、もとの文章データの段階でそうした表記の揺 れを修正している、または「ひらがな語」と「カタカナ語」は区別してカウントしたいという場 合には、この過程は無視してよい。ただし、後ほど行う「1・0データ化」の過程で、「記述 No」 は5列目、「形態素 No」は6列目と定義しているので、マクロを実行する際には、1列挿入する か、もしくはマクロの一部を書き換える必要がある。

まず、B列とC列の間に1列挿入し、表頭に「カタカナ」と入力する(画面11)。「基本形」の 列の語を全て選択してコピーし(画面12)、Wordを立ち上げ、「形式を選択して貼り付け(S)」-「テキスト」で貼り付ける(画面13)。

Word 上で、「編集(E)」 - 「すべて選択(L)」で文字をすべて選択し、「書式(0)」 - 「文字種の 変換(E)」 - 「カタカナ(K)」にチェックを入れて「OK」をクリックする(画面 14)。カタカナに 変換した語を再びコピーして Excel の「カタカナ」列に「テキスト」で貼り付けをする(画面 15)。

🔀 м	icrosoft Ex	cel - 形態計	ē.xls						
❷〕 読ん	ファイル( <u>E</u> ) 編 ude!!ココ( <u>K</u> )	<b>諜(E) 表</b> 示	☑ 挿入ጧ	書式(0) ツ	ハール① デー	タ( <u>D</u> ) ウイント	♡ウ₩) ヘルブ	и <u>н</u> ) –	8×
	🖻 🖬 🔒 🛉	8 🖨 🖪	🌮 🐰 🖻	🛍 • 🚿 🔺	n 🕶 🕬 👻 🍕	🛔 Σ 🔹 f 🛪	<b>≜</b> ↓ <b>Z</b> ↓ <b>∭</b>	📣 100%	• »
	C2	-	f <sub>x</sub>						
	A	В	С	D	E	F	G	Н	
1	表層語	基本形	<u>カタカナ</u>	品詞	記述No				
2	高く	高い		形容詞-自	1				
3	安く	安い		形容詞-自	2				
4	新しい	新しい		形容詞-自	4				
5	多い	多い		形容詞-自	4				
6	安<	安い		形容詞-自	5				
7	少ない	少ない		形容詞-自	8				
8	多	多い		形容詞一自	8				
9	ありがたい	ありがたい		形容詞-自	8				
10	美味しく	美味しい		形容詞-自	11				
11	安い	安い		形容詞-自	11				
12	安	安い		形容詞-自	14				
13	少なく	少ない		形容詞-自	17				
14	安けれ	安い		形容詞-自	21				
15	多い	多い		形容詞-自	22				
16	ない	ない		形容詞-自	22				
17	美味しい	美味しい		形容詞-自	23				
18	ປາປາ	ເນເນ		形容詞-自	24				-
14 4	▶ ▶ \/形態	素、形態素(	2)、抽出語。			•			
ייסר (	ĸ	,(	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,				NUM		

画面 11 「基本形」のカタカナ変換

🔀 M	icrosoft Ex	cel - 形t	<b>熊素</b> :	xls							
<b>剄</b> 読ん	ファイル(E) 縦 ,de‼ココ( <u>K</u> )	■集( <u>E</u> ) 表	£⊼( <u>∨</u>	) 挿入①	書式(0) い	y−ル(	D デー!	¢( <u>0</u> ) לאלי	*ウ( <u>W)</u> へル:	¢Щ) –	θ×
D 1	🛎 🖪 🔒 I	🗟 🎒 🖸	ð. 💞	/ 🐰 🖻	🛍 • 🚿   •	n • )	CH 👻 🎑	$\Sigma \bullet f_{s}$		- 🚯 100%	• »
	B699	-	ţ	≨ 新鮮							
	A	В		С	D		E	F	G	Н	
1	表層語	基本形		1タカナ	品詞	記述	No				
2	高く	高い			形容詞-自		1				
3	安<	安い			形容詞-自		2				
4	新しい	新しい		1	形		4				
5	多い	多い	Ъ	切り取り①			4				
6	安<	安い	<b>B</b>	⊐ピー( <u>C</u> )			5				
7	少ない	少ない	Ê.	貼り付け(P)	)		8				
8	多	多い			al <del>78</del> ⊦n/+/+/(		8				
9	ありがたい	ありがた		ルシェンを選び	COCREDITIO @	<u>v</u>	8				
10	美味しく	美味しい		挿入①…			11				
11	安い	安い		削除( <u>D</u> )			11				
12	安	安い		一	5U 77 (N)		14				
13	少なく	少ない		9XINCIE/0	577 <u>u</u> v		17				
14	安けれ	安い	<b>Č</b> a	コメントの挿	入( <u>M</u> )		21				
15	多い	多い	P	セルの書式	設定(F)		22				
16	ない	ない		ロフトかと認む	1R(K)		22				
17	美味しい	美味しい	-	リストから凄	1/(\ <u>K</u> /		23				
18	ເນເນ	ເນເນ		ハイパーリン	ク( <u>H</u> )		24				
14 4	▶ ▶ \ 形態	素(形態表	表(2)	入抽出語,	/			•			
עדב	۲								NUM		

画面 12 「基本形」のカタカナ変換



図形 13 「基本形」のカタカナ変換



画面 14 「基本形」のカタカナ変換

<b>X</b> M	icrosoft Ex	cel-形態	素.xls						
■ 読ん	ファイル(E) 新 ude!!ココ( <u>K</u> )	扁集(E) 表示	⊼Ϣ 挿入Φ	書式(2) ツ	νール( <u>Τ</u> ) デー	タ( <u>D</u> ) ウィント	♡☆(₩) ヘルブ	- (H)	đΧ
D	🖻 📕 🗃	🔁 🎒 🖪	💞 🐰 🖻	🛍 • 🚿 🕒	n + 🖂 + 📢	$\sum \cdot f_{\mathcal{H}}$	<b>≜</b> ↓ <b>Z</b> ↓ <b>∭</b>	뤻 100%	• *
	C2	-	fx ⊼↑						
	A	В	C C	D	E	F	G	Н	<b></b>
1	表層語	基本形	ガタカナ	品詞	記述No				
2	高く	高い	高イ		1				
3	安く	安い	安イ	形容詞一自	2	+ -	b ++ /−		
4	新しい	新しい	新シイ		4	132	リノに		
5	多い	- 多い - 安い	· 多1 中 7		4	変換し	した文字を		
	女気	女い	女1	形态前日	0			-	
6	ション	ション	シノコ	形在前门日		' フー	+ 7 ト 」 (	-	
a a	シ ありがたい	ありがた	マリガタイ	形物的石	8	貼りイ	すける		
10	ま味しく	美味しい	・ ノ ノノコント	形容詞-白	11				
11	安い	安い	安イ	形容詞-白	11				
12	安	安い	安イ	形容詞-自	14				
13	少なく	少ない	少ナイ	形審詞-自	17				
14	安けれ	安い	安イ	形容詞-自	21				
15	多い	多い	多イ	形容詞-自	22				
16	ない	ない	<b>7</b> 1	形容詞-自	22				
17	美味しい	美味しい	集味シイ	形容詞-自	23				
18	ເນເນ	ປາປາ	1	形容詞-自	24				
14 4	▶ ▶ \形態	素(形態素	<ul><li>②、抽出語</li></ul>			•			
עדב	۲						NUM		

画面 15 「基本形」のカタカナ変換

(5)形態素番号と出現数を割り当てる

品詞情報を用いて抽出した語には、基本形の等しいものが多数存在する。ここでは、この中に 何種類の形態素(基本形)が存在するかを明らかにし、後の「1・0データ化」の過程で必要と なる形態素番号(以降、説明画面では「形態素 No」と表示)の割り当てを行う。また、それらの 出現数の算出も行う。

まず、適当なセルを選択して、「データ(D)」 - 「並び替え(S)」を表示する(画面 16)。「最優 先されるキー」に「品詞」を、「2番目に優先されるキー」に「カタカナ」を選択する。また、「オ プション(0)」をクリックして、「方法」 - 「ふりがなを使わない(R)」がチェックされていること を確認してから、並び替えを実行する。

次に、F列、G列、H列の1行目に、それぞれ「形態素 No」「出現数」「出現数2」と入力する (画面17)。

「形態素 No」の入力は、最初の形態素の行のセル(F2)には数値の「1」を入力する(画面 18)。 次の行のセル(F3)には、「もし『カタカナ列』のセル(C3)の文字が、上のセル(C2)の文字と 同じであれば、上のセル(F2)と同じ数値にする。異なる場合には上のセルの数値に1を加える」 という条件式を入れる[F3の場合:IF(C3=C2、F2、F2+1)]。残りの行にはこの関数をコピー&貼 り付けする。

次に、「出現数」の入力を行う(画面 19)。最初の形態素の行のセル(G2)は、同様に「1」を 入力する。次の行のセル(G3)には、「もし『形態素 No』列のセル(F3)の数値が、上のセル(F2) の数値と同じであれば、上のセル(G2)の数値に1を加える。異なる場合には、数値の1を入力 する」という条件式を入れる[G3 の場合: IF(F3=F2、G2+1、1)]。残りの行にはこの関数をコピ ー&貼り付けする。

続いて、「出現数2」であるが(画面 20)、これは、「出現数」の入力と同じ作業を下(最後の 行)から行っている。これにより、「出現数」列で数値が1の行は、「出現数2」列でその形態素 の総出現数(以下、出現頻度)を示すことになるからである。まず、最後の行のセル(セル番号 はデータにより異なる。事例ではH699)に数値の「1」を入力する。続いて、下から2行目のセ ル(事例ではH698)に、「もし『形態素 No』列のセル(F698)の数値が、下のセル(F699)の数 値と同じであれば、下のセル(H699)の数値に1を加える。異なる場合には、数値の1を入力す る」という条件式を入れる[H698 の場合: IF(F698=F699、H699+1、1)]、そして、それより上の 行にはこの関数をコピー&貼り付けする。

■ Microsoft Excel - 形態素.xls												
図 ファイル(E) 編集(E) 表示(V) 挿入												
□ 最優先 - 「品詞」												
2番目 - 「カタカナ」												
を選択												
1 <u>気福田 本本ル パンパナ</u> 2 高く 高い 高イ												
3 安く 安い 安イ	形容詞-自 3番目に優先されるキー											
4 新しい 新しい 新シイ 5 多い 多い 多イ												
百安く 安い 安イ	並べ替えオブション ? 🔀											
_ 「ふりがなを使わない」	並べ替え順序の指定(U) C データ(W)											
にチェックを入れる	標準OKキャンセル											
	「大文字と小文字を区別する(Q)											
12 安 安) 安1	7回 (7) (7) (7) (7) (7) (7) (7) (7) (7) (7)											
<u>13</u> 少なく 少ない 少ナイ 14 安けれ 安い 安イ	○ 列単位(L)											
15 多い 多い 多イ	方法											
16 ない ない ナイ 17 美味しい 美味しい 美味シイ	<ul> <li>ふりがなを使わない(B)</li> </ul>											
18 111 111 11	OK ##/////											
▲ ◆ ▶   \ 形態素 ( 形態素 ② ) 抽出												
אעדב	NUM											

画面 16 形態素番号と出現数入力

🔀 м	icrosoft	Excel - 形態	索.xls						
₽〕 読ん	ファイル( <u>E)</u> ,de!!ココ( <u>K</u> )	並び替	- <u>^^ / ~</u> えて		ツール(11) デ	ータ( <u>D</u> ) ウ	心やりШ /	√ルプ(出) -	. 8 ×
D ı	🛩 🖪 🕯	同じ基	本形の語を	E 💅		🍓 Σ 🔹	f≈ Ž↓ Z↓	🛍 🚯 100%	• »
	F2	揃えて	おく			-			
	A			<u> </u>	E	F	G	H	
1	表層語	( 画面 16	の実行結	果)	<u>記述No</u>	形態素N	<u>lo 出現数</u>	出現数2	
2	ありがた	いめりかたい	רעתייני	形容詞一	3		_		
3	ເນເນ	<u></u>	11	形容詞一	24	, 			
4	ເງເງ ≂atari	()() 2.4-1-1	11	形容詞日	3 40	)			
5	วกบบ		フレシイ	形容詞目	∃ 50 5	)			
6	くらい	くらい	251	形容詞目	3 92	2	別の目出	ーた入力	74
7	くわしく	くわしい	<u> </u>	形容詞目	30	j	уј₩л		<u> </u>
8	ਗੁ <u>ਟ</u> <	すこい		形容詞目	3 29	)			
9	7ない オン	7211	テイ	形容詞目	3 22				
10	なし	7211	テイ	形容詞目	3 25	5			
11	/よし ナマレ	/al)	テイ		<u>∃</u> 33	}			
12	/JU +>L:	7261 #21.5	テイ	形容詞を	<u>∃</u> 42	<u>'</u>			
13	/JU) +>/	7263	<u> </u>	形容詞的	∃ 43 5 •	5			
14	/d.< +>1.5	7263	テ1	形容詞的	3 48	5 \			
15	/みし) ナマレッ	7261	<u> デ1</u> エノ	形谷詞	∃ 58 5 aa	j			
10	/みし) ナントッ	7201	71	形谷詞	3 bl	,			
17	(みし) ナミレッ	- 7ati 7ati	$r_1$	形谷詞		·			
18	(3.6) 1-71 - 1		11	形合词下	= 9( 				-+- <b>-</b> -
<b>H</b> 4	▶ ▶ \\刑!	態素 ( 形態素	(2)、 披出語	/			•		
コマン	۲						NUN	4	

画面 17 形態素番号と出現数の入力

MICTOSOTT EXCEL - 15 284	⊊.xls					×
<ul> <li>図 ファイル(E) 編集(E) 表示</li> <li>読んde!!□□(K)</li> </ul>	⊻ 挿入Φ	書式(0) ツ	νール( <u>Π</u> ) データ	1( <u>D</u> ) ウィンドウ((	最初のセルは	k
🗅 🛎 🖫 🔒 🖏 🎒 🔍	ABC U U	E * 👏 🗖	🔍 CH 🗸 🍓	$\Sigma \cdot f_{*}$	数値の1	>> *
F3 🔻	f≈ =IF(C3	=C2,F2,F2+	1)			_
A B	G	D	E		G H	
1   表層語   基本形	<u></u>	品詞	記述No	杉態素No 出	現数 出現数2	
2 ありがたい ありがたい	アリガタイ	形容詞-自	8			
	11	形容詞-自	24	2		
2 行目以降は	14	形容詞-自	40	2		
IF 関数を入力	ウレシイ	形容詞-自	50	3		
	クライ	形容詞-自	92	4		
7 くわしく くわしい	クワシイ	形容詞-自	36	5		
<u>8</u> すごく すごい	スゴイ	形容詞一自	29			
9 ない ない	ナイ	形容詞一自	22	_/		
10 なし ない	ナイ	形容詞一目	25	7	に条件式を	
11 なし ない	ナイ	形容詞一目	33	7		
12 なし ない	ナイ	形容詞一目	42	7	コピー&貼り	
13 /3() /3()	ナイ	形容調一目	43	7	付けすると	
14 /a< /al>	テイ	形容詞-目	-48	7		
15 /all) /all)	<b>アイ</b>	形容詞 自	58		同じ語は	
16 /2(1) /2(1)	<u> </u>	形容詞 白	60		日小水能主	
	<u> </u>	形容詞 白	8/		回し形態系	
	71		90		番号になる	-
● ● ▶ ▶  \ <u>形態茶 ( 形態茶  </u>   マンド	四人間出語				L,,	-

画面 18 形態素番号と出現数の入力

■ Microsoft Excel - 形態素.xls		
図 ファイル(E) 編集(E) 表示(V) 挿入(D) は(++++>>>(A)	書式(Q) ツール(T) データ(D)	ウイン 最初のセルは
		★ 数値の1 ・ ※
G3 ▼	=F2,G2+1,1)	*
A B	E	F G H A
1 表層語 基本形 カタカナ	品詞 記述No 形態 形容詞	
2 005% /2000 05% /2007 05% /2000 05% /2000 05% /2007 05% /2000 0000 000000000000000000000000000	形容詞-自 <u>24</u> 形容詞-自 <u>24</u> 形容詞-自 <u>40</u>	
5 IF 関数を入力 ウレシイ 6 マン・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・・	1回しか出現	
7 くわしく くわしい クワシイ 8 すごく すごい スゴイ	しない語では1 🗕	5 1
9 1 セル G4 以降は 1	が表示される	7 1 7 2
1 条件式をコピー	複数回出現する	7 3 7
1 &貼り付けする	語は数値が1つ	7 5
15 ない ない ナイ	ずつ大きくなる	7 7
16 ない ない ナイ		7 8
$17 \text{ all } 70 \text{ all } 71 \text$		7 9
H ↓ ▶ ▶ 下熊素 / 形熊素 (2) 抽出語		
אעקב איזאנאנע אויזענאני אויזענאני אויזענען איזענען אויזענען אויזענען אויזענען אויזענען אויזענען אויזענען אויזע		NUM

画面 19 形態素番号と出現数の入力

Microsoft Excel - 形態素.xls												
	ファイル( <u>E</u> ) 編 オーリココ(K)	<b>諜(E) 表示</b>	☑ 挿入璽	書式(0) じ	/-ル①	一番下の	セルに					
:≂≂n []](	≥ 🔲 🕾	6 4 6	ABC Y P		<b>.</b> . Ci ·	数値の1	を入力す	る <sub>100%</sub>	-			
	H698	- (	£ =IF(F6	98=F699,H6	99+1,1)							
	A	В		D	E	F	G	Н				
1	表層語	基本形	カタナナ	品詞	記述No	形態素No	出現数	出現数2				
687	大好ぎ	大好ぎ	大好キ	名詞-形容	46	239	1	1				
688	大切	大切	术切	名詞-形容	15	240	1	1				
689	大変	大変	大変__	名詞-形容	31	241	1	1				
690	大量	大量	大量	「出現数	1=1の時	242	+	1				
691	不安	不安	不安			243	1	4				
692	不安	不安	不安	「出現数 2	」のセル	t 243	2	3				
693	不安	不安	不安	你山田米	*==	243	3	2				
694	不安	不安	不安	総山現安	とを不9	243	4	1				
695	便利	便利	便利	名詞-形容	68	244	1	1				
696	無害	無害	無害	名詞-形容	75	245	1	1				
6	<u> </u>	) (二日	馬太	名詞-形容	51	246	1	1				
68	Г <i>Л</i> 9 2	-1J日	倒	名詞-形容	54	247		1				
68	より上に	こは	-Č	名詞一形容	30	248		1	لرز			
70	10 月月米ケス	★ \ + ≠ 7	-						_			
70		で八八りる	S						L			
702									L			
703												
14 4	▶ ▶ \ 形態	素/形態素(	2) <u>油出語</u>	/		4			•			
עדב	۲						NUM					

画面 20 形態素番号と出現数の入力

(6)出現数からキーワードを絞り込む

「形態素 No」「出現数」「出現数2」の入力が全て終了したら、並び替えなどで数値が変化しな いように、新しいシート(シート名「data」)にコピー&「形式を選択して貼り付け(S)」-「値 (V)」で貼り付けをする(画面21、画面22)。なお、ここで指定したシート名「data」は、後で用 いるマクロに組み込まれているので、他の名前にする場合にはマクロの書き換えが必要である。

抽出語データを「data」シートに貼り付けたら、「データ(D)」 - 「フィルタ(F)」 - 「オートフィルタ(F)」を選択して、1行目にフィルタ矢印をつける。次に、「出現数」列のフィルタ矢印を クリックして、1を選択する(画面 23)。

出現数が1の行だけを表示させたら、新たに設けた「出現数」シートにコピーして貼り付ける (画面 24)。前述したように、「出現数」列の数値が1の行は、「出現数2」列にその抽出語の出 現頻度が示されているので、「出現数」シートには各形態素の出現頻度を示した行だけが抽出され たことになる。

ここで、「データ(D)」 - 「並べ替え(S)」を選択し(画面 25)、「出現数2」列の値が大きい順 に並べ替える。「出現数2」列は各抽出語の出現頻度を示しているので、より多く出現する語から 順番に示される(画面 26)。

ここから、一定以上の出現頻度がある語をキーワードとして抽出する。これは、例えば出現頻 度が1というのは、たった1回しか記述がなかった言葉であり、全体の集計をする上でこれを取 り上げる意義は少ないと考えるからである。そこで、「最低何回出現したか」という基準でキーワ ードを絞り込むことにする。

この最低ラインについては、データのサイズや求める分析の精度によって異なる。データの量 が少ない場合に出現頻度の下限を高く設定すると、一般的な語(例えば、青果物に関する記述で 「野菜」と「購入」だけ)しか拾えない場合がある。また、なるべく多くキーワードを用意した 方が、類義語をまとめるなど分析に幅を持たせることができる。一方、大量のデータがあるのに 下限を低く設定すると、膨大な量のキーワードの扱いに苦慮することになる。また、出現頻度の 低い語を統計的に扱うことの意義についても問われることになるであろう。

ただし、本手法では「1・0データ化」の過程で、Excel の各列にキーワードを割り振る。す なわち、1回の作業で扱えるキーワード数は「Excel の列数の上限(256)」-「文章番号を示す 列(1)」=255 個である。これに回答者属性などのデータを加えることを考えると、扱いやすい キーワード数の上限は250 個程度であるといえる。もちろん、システム上の上限であるため、250 が適当なキーワード数であるという根拠はない。また、データを分割していくつかシートを作成 すれば、255 を上回るキーワードを抽出することも可能である。しかし、適当な出現頻度の下限 を設定できない場合には、1つの目安として、キーワード数が250 を超えない程度に抑えられる 出現頻度を挙げることができる。

ここで示す事例は説明用の簡易データなので、出現頻度3という低い数値で抽出している(画 27)。「出現数」シートの「出現数2」列の数値が3以上の行をコピーし、新たに設けたシート (事例ではシート名「3以上」)に貼り付ける。このシートにある語が、1・0データ化を行うキ ーワードとなる。

🔀 Mi	☑ Microsoft Excel - 形態素.xls												
<b>剄</b> 読ん	ファイル(E) .de!!ココ( <u>K</u> )	編	潗(Ľ) 表示	'⊻ 挿入Φ	書式( <u>O</u> )	y	ール① デ	-タ( <u>D</u> ) ウィン	ドウW) へル:	パ( <u>H)</u> ー ピ	۶×		
D	差 🔲 🖁	3 4	à 🖨 À	🕸 🎖 🌾	🛍 • 🚿	ĸ	<b>) →</b> Cil →	🍓 Σ 🗕 f 🕷	<b>≜</b> ↓ Z↓ 🛍	- 🐻 100%	• *		
	A1		-	★ 表層語	5. 								
	A		В	С	D		Е	F	G	Н			
1	表層語		基本形	カタカナ	品詞		記述No	形態素No	出現数	出現数2			
2	ありがた	<u>-</u> l1	ありがたい	アリガタイ	形容詞	-自	8	3 1	1	1			
3	ເນເນ	*	切り取り①		容詞·	-自	24	1 2	1	2			
4	เงเง ≂≠ระเม	6	⊐ピー( <u>C</u> )		~詞	-自	40	) 2	2	1			
5	วกมา	A	貼り付け(P)		学同	-目	50	) 3	1	1	_		
07	くちい			78Fn/#/4/c)	장물리	-日	92	2 4	1	1	_		
/	くわしく ナディ		パンゴンを注意すべい	20800100 @/	··· 국·문희	-日 占	30		1	1	_		
0	ターへ オペレン		挿入仰		(165) (전) (151)	-日 -白	28	ט ע ד ע	1	10			
10	'a∿' 721.		削除( <u>D</u> )		11日 家舗司-	白	22 2P	· /	2	10 Q			
11	なし		数式と値のク	リア(N)	~ 司	-白	33	, , ,	3	8			
12	なし				~ 高	-白	42	27	4	7			
13	ない	₩.	ビルの香い話	ΩÆ\ <u>Γ</u> /	字詞- 容詞-	-自	43	3 7	5	6			
14	なく		行の高さ( <u>R</u> ).		容詞·	-自	48	3 7	6	5			
15	ない		表示しない(	Ð	李詞·	-自	58	3 7	7	4			
16	ない		再表示(U)		容詞·	-自	60	) 7	8	3			
17	ない	_	/JUN	T1	形容詞	-自	87	7 7	9	2			
18	ない		ない	ナイ	形容詞	-自	90	) 7	10	1	-		
<b>H</b>	▶ N\无	態	人形態素	2) 抽出語	/data/	~		1			•		
עדב	*					合	計=117846		NUM				

画面 21 キーワードの絞り込み



# 画面 22 キーワードの絞り込み

N 12	licrosoft Ex	cel - 形!	艦素.xls						X
8	ファイル(E) 新	扁集( <u>E</u> ) → 表	転⊙ 挿入ወ	書式(2) ツ	ール① デー	タ( <u>D</u> ) ウィンド	לעא ישיל	î( <u>H</u> )	
記れ	Vde!!]]( <u>K</u> )								ч×
	🖻 📙 🔒 I	🔁   🖨 🛛	à. 🏹   X 🖻	🛍 • 🚿   🗉	א 🖓 די ד	$\Sigma \bullet f_{\pi}$	2 Z I 🗍	46 100%	• *
	G2	•	<i>∱</i> × 1						
	A	В	С	D	E	F	G	Н	-
1	表層語 💌	基本形	👤 カタカナ 👤	品詞 🔽	記述No 🔽	形態素N▼	出現数 🔽	出現数2▼	
2	ありがたい	ありがた			7 8	(すべて)	🔺	1	
3	ເນເນ	ປາປາ	'出現鉯	」の列か	24	(197) J		2	
4	ບບ	ປາປາ	1の行を	抽出する	40			1	
5	うれしい	うれしい	1 0313 G		50	3		1	
6	くらい	くらい	クライ	形容詞-自	92	4		1	
7	くわしく	くわしい	クワシイ	形容詞-自	36	ő		1	
8	すごく	すごい	スゴイ	形容詞-自	29	7		1	
9	ない	ない	ナイ	形容詞-自	22	9		10	
10	なし	ない	ナイ	形容詞-自	25	10		9	
11	なし	ない	ナイ	形容詞-自	33	12		8	
12	なし	ない	ナイ	形容詞-自	42	13		7	
13	ない	ない	ナイ	形容詞-自	43	15		6	
14	なく	ない	ナイ	形容詞-自	48	16	-	5	
15	ない	ない	ナイ	形容詞-自	58		7	4	
16	ない	ない	ナイ	形容詞-自	60	7	8	3	
17	ない	ない	ナイ	形容詞-自	87	7	9	2	
18	ない	ない	ナイ	形容詞-自	90	7	10	1	
14 4	▶ ▶ \ 形態	素/形態	素②/抽出語)	data /		I I			
נדב	۲×.		,,,====				NUM		

画面 23 キーワードの絞り込み

Microsoft Excel - 形態素.xls												
■ ファイル(E) 編 読んdeUココ(K)	[集( <u>E</u> ) 表示	R₩	挿入仰	書式(2) ツ	-μ(Ώ	データ	( <u>D</u> ) ウィンドウ	₩) ヘルプ( <u>H</u> )	- 6	×		
	B 🖨 🖪		出現数	が1の行れ	だける	. 🚷	$\Sigma \bullet f_{\pi} \stackrel{\text{A}}{\simeq}$	i zi 🛍 🥡	100%	•		
A1	-			- い - い - い - い	 <del>-</del>							
A	B	L	迭折し		9 0		F	G	H	_ <b>_</b>		
1 表層語 ▼	基本形   ▼	<u>・力:</u> エ	タカナ <u>▼</u> *2		記述N	<u>○</u> <u>▼</u> 用	2.11.11.11.11.11.11.11.11.11.11.11.11.11	<u>出現数 ▼</u> 出	<u>現数2</u> ▼	_		
<u>662</u> 手腔	于輇 壬/佰	于	粋 ⊬百	名詞 形谷		35	234		1			
663 新鮮	新鮮	V	9 <u>स</u> ∤ຫກ≣ຫກ/	ים אין וויים סיי ד)		2	236		17			
680 生	生	00 195		D.		5	237	1	6			
686 多目	多目	1	コピー©) 貼り付け(P)			69	238	1	1			
<u>687</u> 大好き	大好ぎ				46	239	1	1				
<u>688</u> 大切	大切		形式を選	択して貼り付け(	<u>S</u> )	15	240	1	1			
689 大変	大変		行の挿入	Φ		31	241	1	1			
690 八重 691 不安	不安		行の削除	( <u>D</u> )		40	242	1	4			
🥫 前もって			数式と値	のクリア( <u>N</u> )		68	244	1	1			
			わしの妻			75	245	1	1			
「 貼り111万	元の		にの書い			51	246	1	1			
肾 新しいシ	ートを			<u></u>		54	247	1	1			
	*		जित्तन्त् 	, v <u>H</u> /	-	30	248		1	-		
	5		冉表示([									
	表 / 形能表	(2)	/抽史語》	udata (北田)	数 /			V				
フィルタ モード	no A novet <del>no</del>	27	, <u>המרחמר</u> י		RA/ STE4104	1		NEM				
					01-410-			1011				

画面 24 キーワードの絞り込み



画面 25 キーワードの絞り込み

■ Microsoft Excel - 形態素.xls												
図 読ん	ファイル(E) ,de!!ココ( <u>K</u> )	編集( <u>E</u> ) 表:	示── 挿入Φ	書式(2) ツ	ν−ル① デー	タ <u>(D)</u> ウインド	やし ヘルブ	(H) _₽×				
D 1	🖻 🔲 🚔	) 🔁 🎒 🗋	💞 🐰 🖻	🛍 • 🚿 🛛 🕷	ວ 🕶 🖂 👻 📢	🛓 Σ 🝷 f 🛪		🚯 100% 🔹 🦓				
	H2	$\sim$	<b>∱</b> ∗ 70					$ \land $				
	A	В	С	D	E	F	G	Н				
1	表層語	基本形	カタカナ	品詞	記述No	形態素No	出現数	<u>出現数2</u>				
2	し	する	<u></u>	動詞-自立	1	29	1	70				
3	購入	購入		名詞-サ変	3	96	1	25				
4	思っ	思う	思ウ	動詞自立	5	62	1	21				
5	買わ	買う	買り	動詞一自立	4	79	1	21				
6	ある	ある		動詞目立	5	21		18				
7	農楽	農業	農業	名詞一般	5	214		18				
8	新鮮	新鮮	新評	名詞一形容	2	236		17				
9	食べる	食べる	食べル	動計目立	9	69	1	16				
10	育采	青米	育米	谷詞-一般	1	89	1	15				
11	/43 57/	143 111		朝記一日立	3	35	1	13				
12	女く #>1 >	女い オントン		形容詞目		10	1	12				
13	ない	(みい)		形谷祠一日			1	10				
14	ガルウ	アルる		各抽出部	音の	83	1	10				
10	ゴボ 羊中レノ	ゴボーム	ゴボー	出租粮度	きがわかス	223	4					
17	天中しく	一天中しい	<u> 長いい1</u>	山/元//只/5	211.1717.2	204	4	9				
10	ゴー 三/	희			1	10		7				
10						12						
	▶ ▶\ <u>/形</u> !	出来,北非来	いん細出語,	(data)出現	<u> #X</u> /	•						
עדב	۴						NUM					

画面 26 キーワードの絞り込み

🔀 Microsoft E	xcel - 形態	索.xl	S							
嚠 ファイル(E) 読んde‼ココ( <u>K</u> )	編集( <u>E</u> ) 表示	Ś	挿入邸 書式(2) ツール(1)	) データ( <u>D</u>	♪ ウィンドウ <u>№</u>	୬ ∧ルプ( <u>H</u> )	) (	ē ×		
🗋 🗁 🔚 🔒	🔁 🎒 📐	ABC	👗 🗈 🙉 + 🝼 ၊ 🗠 + 🔿	a 🗸 🎑 🗄	$\Sigma - f_{\pi} \stackrel{A}{\geq} \downarrow$	7 🕅 🕷	100%	• »		
A1	<b>•</b>		出現頻度がいくつ以	上の語	を					
A	В			きょう		G	Н			
47 知り	知る		キーワートとりるか	決めて	76	1	3			
48 入っ	入る	77	1	3						
49 売っ 売る 元/// 動画 日立 81 1										
<u>50</u> 生産	生産	*	切り取り(工)	56	102	1	3			
51 調理	調理	83	⊐ピー( <u>©</u> )	20	105	1	3			
52 派	成	A	貼り付け(P)	27	154	1	3			
<u>53</u> <u></u> 于快	子供		形式を避快して貼り付け(の)	15	164	1	3			
54 于	于		15148781/(OCHODINO/00/~~	31	109	1	3			
50 日 56 休	山		挿入仰	3	195	1	3			
57 <u>绛</u>	岱		削除( <u>D</u> )	10	197	1	3			
58 店頭	店頭		数式と値のクリア(№)	39	貼り付け	· 先の	3			
59 日持ち	日持ち	r R		45			3	/		
60 INN	ເນເນ			24	新しいシ	/ートを	2			
61 新しい	新しい		17の両で12	4	田音して	**	2			
62 良い	良い		表示しない( <u>H</u> )	70			2			
63 冷たく	冷たい		再表示(U)	31	20	1	2			
64 かかる	かかる	תנק	コル 動詞 目立	22	25	1	2			
■●▶■∖形態	素、形態素	Ôړ	抽出語 (data) <u>出現数 (3</u> .	UE/	I I					
אַלאַב			合計=861	6		NUM				

画面 27 キーワードの絞り込み

まず、1・0データ用のシートを新たに設ける(画面28)。シート名は「1・0」(全て全角) である。このシート名は「data」シートと同様、マクロに組み込まれているので、他の名前にす る場合には、マクロの書き換えが必要である。

次に、(6)で絞り込んだキーワードのシート(この事例ではシート名「3以上」)に戻り、「形 態素 No」の数値が小さい順に並べ替えをする(画面 29)。これが、キーワードとして扱う形態素 の番号である。この番号を選択し(画面 30)、「1・0」シートの1行目B列(セル「B1」)から 横に貼り付けをする(画面 31、32)。さらに、セル「A2」を選択し(画面 33)、縦に1から全文章 数(事例では文章数 92)までの文章番号を入力する。

ここで、「data」シートに戻る(画面 34)。出現数によるキーワード絞り込みの過程でオートフィルタ機能を使用したままの状態にあるので、「出現数」列のフィルタ矢印をクリックし、「すべて」を選択して全てのデータを表示させる。

次に、「記述 No」と「形態素 No」で並べ替えをする(画面 35)。「データ(D)」-「並べ替え(S)」 を選択し、「最優先されるキー」として「記述 No」を、「2番目に優先されるキー」として「形態 素 No」を選択し、どちらも「昇順」にチェックを入れる。これにより、文章番号の小さい順に並 び、同じ文章番号の場合には形態素番号の小さい順に並ぶことになる。



画面 28 「1・0 データ化」の準備



画面 29 「1・0 データ化」の準備

<b>X</b> M	icrosoft	Exc	el - 形如	鸄索.xls									$\mathbf{X}$
<b>B</b> )	ファイル(圧)	編	集(E) 表	示\∑) 挿入	た 単の	( <u>0</u> ) ))	ν−ル(T)	データ([	シ ウィンド	ウ(W) ^	いプ(日)		
_ 読ん	,de‼(K)	ж	切り取り	(T)	Ctrl+X							_ é	з×
B	2 🗖 🖻	Ba.	コピー(C)		Otrl+C	3	0 + CI	- 0.	$\Sigma - f_{\infty}$	AT ZT	60 🔊	100%	• »
	 D="2b			/m\	Owhor				* 0 00		··· A.		*
INIS.	PJ299		REDITIO	\ <u>C</u> /	Otri+v		194   19	•%,	00 + 0		i 🕈 थ	• 🗛 •	•
	F59		形式を運	観して貼り付け	t( <u>S</u> )				$\frown$				
	A		シートの削	训除①			E		F	G		Н	_
1	表層語		シートの利	多動またはコピー	-(M)		記述No	シ 正	態素No	出現数	出現	見数2	
2	ない			9900-C10-LC	<u></u>	1-自		2/	7		1	10	
3	よい	<b>6</b> 9	検索(上)。		Ctrl+F	1-自		<u> 1</u> 8	9		1	3	_
4	安く		置換( <u>E</u> ).		Ctrl+H	1-自		2	10	1	1	12	
5	高く	_		*		旧自		1	12		1	7	
6	少ない		少ない	シナイ	— 形容	詞自		8	13		1	4	L
7	33U)	-	<u></u> 3 ዓር ነ	- 391	形容	詞自		4	15		1	7	<u> </u>
8	美味しく	-	美味しい	実味シイ		詞一日		11	18		1	9	<u> </u>
9	<b>あ</b> る	- 2	<u>ති</u> ති	昇順にす	£べ替え	えた		5	21		1	18	<u> </u>
10	ເນລ ່		いる + 7	<b>取能表表</b>	日元			15	23		1	5	_
11			95 -/	加密杂目	850			1	29	_	1	/0	-
12	つい	_	) 57	コピーゔ	する			26	31	_	1	4	-
13	/J/S 14-75		ವ⊲ಾ   ದ-ಇಾ 7	· · ·	<b>314 617</b>			3	35	-	1	13	-
14	rμ.c. nas	, r	ሦር <u></u> ኤኋ	エテル	動詞	- 티 꼬		X-	43		1	3	-
15	- 咽つ /まね	L	<u> </u> 扱つ (また		動詞	- 변 곳		X	51	<b>/</b>	1	/	-
10	受り		ぼつ <u></u> 田に	- 127	動詞	- 티 꼬		'X	58		1	04	-
14 4	▶ ▶ \形	態素		気の/抽出	语/data	/出現	数∖3比	ĿN	0/ 1	•		<b>1</b>	
עדב	۲					4	計=7016			NUM	1		

画面 30 「1・0 データ化」の準備



画面 31 「1・0 データ化」の準備



画面 32 「1・0 データ化」の準備



画面 33 「1・0 データ化」の準備



画面 34 「1・0 データ化」の準備



画面 35 「1・0 データ化」の準備

(8)マクロの実行

まず、プログラムをインポートする<sup>注(8)</sup>。「ツール(T)」 - 「マクロ(M)」 - 「Visual Basic Editor(V)」を選択し、VBEを起動する(画面 36)。VBEの「ファイル(F)」メニューから「フ ァイルのインポート(I)」を選択し、「1・0データ化」のマクロ「茶坊主くん」(仮称)をインポ ートする(画面 37)。

Excelの画面に戻り、「1・0」シートが表示されていることを確認する(画面 38)。「ツール(T)」 -「マクロ(M)」-「マクロ(M)」を選択すると、マクロ名「茶坊主くん」が表示されるので、こ れを実行する。すると、「抽出したキーワードの数」と「1・0データ化する文章の数」がそれぞ れ表示されるので(画面 39、40)数に間違いがないことを確認して「OK」をクリックする。

「1・0データ化」は、文章番号1から形態素番号の小さい順に行われる。この事例のように 文章数 92×キーワード数 58 という少ないデータ数の場合には数秒で終了するが、もっとデータ の規模が大きい場合には、終了まで数分を要することがある<sup>注(9)</sup>。そのため、どの程度まで処理 が進行しているか、現在処理中の文章番号がセル「A1」に表示されるようになっている(画面 41)。 このセルの値が総文章数(事例の場合 92)と等しくなったら、「1・0データ化」は終了である。

注(8)ここでは、マクロのファイルが作成済みであることを前提としている。マクロの作成については 本書でコードを公表しているので、専用の解説書などを参照の上、各自で対応していただきたい。 注(9)データ規模が大きい場合には、処理速度を上げるために、「data」シートおよび「1・0」シー トのみを他のブックに移してマクロを実行することをお勧めする。



画面 36 マクロの実行



画面 37 マクロの実行



画面 38 マクロの実行



画面 39 マクロの実行



画面 40 マクロの実行

	ic 現在処刊		kls					
劉	┘ 流住之う	ェ , ッ (⊻) ⊒ が	) 挿入①	書式(Q) ツール	( <u>[</u> ) データ( <u>[</u> )	) ワインドワ№	0 ~N7(H)	_ 8 ×
	<u>へ</u> = =	nz 🕅	× 🖻 🕻	l - ≪   ∽ -	a - 🍓 🗴	E + f≈ A	ZI 🛍 🛷	100% - »
MS	₽ᠴᡔᢦᡔ		B 7 II	三三三陸	<b>9%</b> ,	<b>+.0</b> .00 <b>€</b>	E 🖂 • 🖄	• <u>A</u> • »
-		- (f	<b>9</b> 2	$\mathbf{)}$	Γ <sub>Π</sub>	《熊素番号	+	
	A	в		D	E			H
	92	7	9	10	12	13	15	18
2		0	0	0	1	0	0	
3	2	0	0	1	0	0	0	
4	3	0	0	0	0	0	0	
5	4	0	0	0	0	0	1	0
6	5	0	0	1	0	0	0	0
7	6	0	0	0	0	0	0	0
8	7	0	0	0	0	0	0	0
9	8	0	0	0	0	1	1	0
10	9	0	0	0	0	0	0	0
11	10	0	0			<u> </u>	0	0
12	11	0	0	_ 1 • 01Ł	これたテ	-~	0	1
13	12	0	0		0		0	0
14	13	0	0	0	0	0	0	0
 	<u><u></u> ≅<del>▼</del>□</u>		0	1	0	0	0	
	早留石 5		0	0	0	0	0	
17	▶ ▶ \\形態素	まん形態素 ②	/抽出語/@	ata/出現数/	(3以上) <u>1·</u> (	2/		
コマン	۲						NUM	

画面 41 マクロの実行

(9)1・0データファイルの完成

「1・0データ化」が終了したら、最後に元の文章データや属性コードのあるデータファイル と結合させる。

まず、1・0データの1行目にある形態素番号のセルを、数値から形態素(基本形)に書き換 えて、どのキーワードに関する出現状況を示しているかわかるようにする(画面 42)。先にキー ワードを絞り込んだシート(事例ではシート名「3以上」)に戻り(画面 43)「基本形」列にあ るキーワードを選択してコピーする。「1・0」シートに戻ってセル「B2」を選択、「形式を選択 して貼り付け(S)」で「行列を入れ替える(E)」にチェックを入れて貼り付ける(画面 44)。

最後に、キーワードの出現状況を示した列を全て選択し(画面 45)、文章データが入力してあ る元のファイルに貼り付け(画面 46)、1・0データファイルを完成させる(画面 47)。

なお、キーワードから元の文章を検索する場合には(画面 48)「データ(D)」-「フィルタ(F)」 -「オートフィルタ(F)」を選択して1行目にフィルタ矢印をつけ、検索したいキーワードのフィ ルタ矢印をクリックして1を選択すると、求めている元の文章を抽出することができる。

■ Microsoft Excel - 形態素.xls										
<b>圏</b> : 読ん	ファイル(E) 編集 jde‼ココ( <u>K</u> )	≹(E) 表示(⊻)	挿入①	書式(0) ツ	ール(工) データ	x( <u>D</u> ) ウィンド	ウѠ ヘルプ	(H) _ & ×		
	2 🖬 🔒 🐿	a 🛯 🖉	1 K 🖻 1	🔁 • 🚿 🗠	) + Ci + 🍓	$\Sigma \bullet f_{s}$		🚜 100% 🔹 🐥		
MST	Pゴシック	• 11 • F	JUI	E = =	œ <b>9 %</b>	00. 0.+		<u>A - A</u> - »		
-		<b>▼</b> fs	. 92				<del>-</del> ,   <u></u>			
	A		<del>.</del>	Ð	Ē	F	G			
1	92	7	9	10	12	13	15	18		
2	1			0	1	0	0			
3	2	0	0		0	0	0	0		
4	J 4	0	0		0	0	1	0		
6	5	0	0		<u> </u>	0	0	0		
7	6		ŏ	この部方	き	Ő	Ő	0		
8	7	0	0	キーワー	・ドに 🔎	0	0	0		
9	8	0	0	·····································		1	1	0		
10	9	0	0		<b>`</b>	0	0	0		
11	10	0	0	L		0	0	0		
12	11		0	1	0	0	0	1		
13	12	0	0	0	0	0	0	0		
14	13	0	0	0	0	0	0	0		
15	14	0			0	0	0	0		
10	15	0	0	0	0	0	0	0		
14 4	▶ ▶ \ 形態素	(形態素 ②	/抽出語/	data/出現	发(3以上)	<u>1•0/</u>				
עדב	r.						NUM			

画面 42 1・0 データファイルの完成

■ Microsoft Excel - 形態素.xls												
1 Xm 1	ta ta 128		<b>b</b> 🔬 🐝	防閉結果不疲惫	• 1 ≣((	・0」シ	ートの					
: <u>1</u>	い能	きまの「甘っ	k IV to		1行	i目と同じ	数値が					
111 (***)	,心思	ふり 幸2	<u>א</u> נידא	18式型 り			+> + P					
記元れ   一	" 選打	<b>そし、コピ-</b>	-する		同し	/順雷(と	ちちも升	順)で「『^				
L (			V   00 HE	🖪 + 🚿   K	? 並ん	でいるこ	とを確認	する 📑 🤻				
MS	Pゴシック	• 1 •	B <i>I</i> <u>U</u>	≣≣≣	B 9 %	, .00 + 0		× · · · · ·				
	B59	-	f∡ 不安	1								
	A	В	С	D	E	F	G	Н				
1	表層語	基本形	カタカナ	品詞	記述No	形態素No	出現数	出現数2				
2	ない	<b>ぶ</b> い	ナイ	形容詞-自	22	7	1	10				
3	よい	よい	<u> 1</u>	形容詞-自	48	9	1	3				
4	安<	安い	<b>安</b> イ	形容詞-自	2	10	1	12				
5	高く	高い	高イ	形容詞一自	1	12	1	7				
6	少ない	少ない	チナイ	形容詞一自	8	13	1	4				
7	-39() -39()	-33()	31	形容詞一自	4	15	1	7				
8	美味しく	美味しい	美味シイ	形容詞一自	11	18	1	9				
9	ある	ある	アル	動詞目立	5	21	1	18				
10	いる	いる	111	「3以上」	シート	23	1	5				
11	し	する	<u> オル</u> .			29	1	70				
12	つい	24	12 6	こ戻る	Ì	31	1	4				
13	なる	なる		期间日义	3	35	1	13				
14	ゆで	ゆでる	ユデル	動詞-自立	22	43	1	3				
15	喰っ	喰う	喰ウ	動詞-自立	27	51	1	7 🗸				
H 4	▶▶∖形	《素》形態素	② (抽出語)	(data/出現	<u>数、3以上</u> /	$10/\mathbf{V}$	•					
עדב	۲						NUM					





画面 44 1・0 データファイルの完成

🔀 Mi	icrosoft Excel -	形態	索.xls		<b>A</b>	]		
1	ta ta 🗷 🔁 ta	ı 🛛	🖪 🙆 🔻	●校開給	0 データ	の列を	終了 ( <u>N</u> ) 🖕	
<b>8</b> ) :	ファイル(E) 編集(E)	表	示⊠ 挿入頌	) <sub>書ゴ</sub> 全て	選択して	21	ドウ(W) ヘルプ	Έ
読ん	de!!( <u>K</u> )			¬ ~	゚ーする			_ 8 ×
Πı	😤 🔲 🙈 🗞 🖌	5 R	** X 🖻		. 9 0		AT ZT Mu	🚜 100% 🔍 👻
MS.		11 .				• • • .0 .00		Δ_ »
	- 12772			!  = = =		> <b>;</b> .00 <b>+</b> .0		<u>~</u> • <u>~</u> ••
	BG1 🔻	$\sim$	<u>た</u> 小丑			_		
1	A	В			上	ト	G All	
2	92 95 1		0		 	21211	-3261	
3	2	X	切り取り(T)			0	0	0 -
4	3				) ů	0 0	0	0
5	4				- 0	0	1	0
6	5		貼り付け(円)		0	0	0	0
7	6		形式を選択し	元貼り付け(S)	0	0	0	0
8	7		挿入仰		0	0	0	0
9	8	_	肖		0	1	1	0
10	9	_	新士や病の方	1200	0	0	0	0
11	10	_	東江へに回のシ	77° <u>11</u> 7	0	0	0	0
12	11		セルの書式設	定(E)	0	0	0	1
13	12	-	列の幅( <u>C</u> )		0	0	0	0
14	13	-	表示しない(上	Ð	0	0	0	
			用表示印刷			1.0 (		
	▶▶□\\兀恐来(井	ST	1190100		(3以上) (3以上)	(1.0)		
ולאב				Ê	GET=423		NUM	

画面 45 1・0 データファイルの完成



画面 46 1・0 データファイルの完成

🔀 М	icros	oft Ex	cel - 調	査結果い	S		
8	ファイル	レ( <u>F</u> ) 編	編集(E) ⅔	表示♡	■入① 書式(Q) ツール(T) データ(D) ウィ	ンドウW ヘルプ	(H) <b>- 8 ×</b>
	E1 1	79	- ×	J fx			
	Α	В	С	D	E F	G	Н
1	No	code1	code2	code3	記述文ない	よい	安い 高
2	1	2	1	2	司じ物でも,とても高く感じる問	0 0	0
3	2	3	4	2	安くて、新鮮で安全なもの	0 0	1
4	3	3	6	2	旬の時期になるべく購入し旬	0 0	0
5	4	3	3	1	新しい種類が売っていても、利	0 0	0
6	5	3	3	2	主協にある青菜ならば(野菜≦	0 0	1
7	6	2	4	2	青菜類は100円程度の価格の	0 0	0
8	7	3	3	2	新鮮でなるべく県産品を購入	0 0	0
9	8	3	3	2	家族が少ないと一束が多すぎ	0 0	0
10	9	2	1	1	まぼ毎日、必ず食べる様にし	0 0	0
11	10	3	3	4	淺に入っているけれど、葉のI	0 0	0
12	11	2	4	2	時期にあったものを購入して!	0 0	1
13	12	4	4	2	う者にんにくを店においてほし	0 0	0
14	13	4	3	2	新鮮で無農薬であること	0 0	0
15	14	3	5	2	安さ 安全 地物	0 0	1
16	15	2	4	2	栄養価のことを気にして、意識	0 0	0
17	16	3	4	2	特にありません	0 0	0
18	17	2	4	2	やはり、気になるのは、農薬。	0 0	0
19	18	2	3	2	断鮮であること。農薬をあまり	0 0	0
20	19	2	5	2	更いまわしがきくものを特に購	0 0	0
21	20	\ ■本	結重 / 2	2	個理学れなっと分かれ ビート	0 0	
入力	<b>P P</b> 1					NUM	

画面 47 1・0 データファイルの完成



画面 48 1・0 データファイルの完成

4. 補足説明

(1)2つ以上のキーワードをまとめる場合

1・0データファイルの作成を終了し、文章の検索や分析を進めていくうちに、同じような意味で使われている語など、いくつかのキーワードをまとめて1つの語として扱う必要が生じてくることがある<sup>注(10)</sup>。その場合、それらのキーワードの出現状況をあわせて1・0で示す列を新たに加えることになる。以下、その手順を説明する。

まず、合成したい語(ここでは「調理」と「料理」をあわせる事例を示す)の列を、別のシートにコピー&貼り付けする(画面49、50)。新たに、「合計」列をつくり、2列の数値をそれぞれの行ごとに合計する(画面51)。これにより、「合計」列では、どちらか一方の語が出現した文章には1、両方の語が出現している場合には2、どちらの語も出現していない場合には0が表示される<sup>注(11)</sup>。

次に、「合計」列の隣に、2つの語を合成した項目であることを示す名前を付ける(事例では、「調理・料理」)(画面 52)。そして、2行目以降には、「『合計』列が0の場合は0、それ以外は 1」という IF 関数を入力する。[E2の場合: IF(D2=0、0、1)]

最後に、合成項目の列を選択・コピーし(画面 53) 元のデータの最後尾に貼り付ける。この際、「数式」のまま貼り付けると隣の列と同じ数値になってしまうので、必ず「値」で貼り付ける よう注意が必要である(画面 54、55)。

- 注(10)同じ意味で使われる語の他、全く反対の意味を持つ語も1つにまとめることによって分析が容 易になる場合がある。例えば、「美味しい」と書かれた文章と、「美味しくない」と書かれた文章は、 全く反対の内容であるにも関わらず、形態素で抽出すると「美味しい」という語になる。一方、「不 味い」と書かれた文章は、「美味しくない」とほぼ同じ意味であるにもかかわらず、別の語として 抽出される。そこで、「美味しい」も「不味い」も1つにまとめて、「味」に関する記述と捉えるの である。
- 注(11)事例のように合成する語が2つの場合は、行ごとの合計を出さずに、直接「2列とも0の場合 は0、そうでなければ1」という条件式を使うことも可能である。ただし、合成する語が3語以上 の場合には、いったん合計した方が容易である。



画面 49 語の合成

M 🔊	icros	oft Excel	- Book	1								X
8	ファイル	レ(E) 編集(	E) 表示	ξŴ	挿入仰	書式	t( <u>O) ッー</u> ル(	I) データ( <u>D</u> )	ウィンドウW	ヘルプ(田)	- 6	۶×
	产 🖡	a 🔒 🗞	i 🕹	ABC	X 🖻	٦.	🚿 🗠 –	🖙 - 🝓 Σ	: • f* 👌	KI 🛍 🚜	100%	• *
MS	Pゴシ	ック •	11 -	в	ΙU	E	= = <b>=</b>	9%,	+.0 .00 €	- 🖓 🗸 🔕	• <u>A</u> •	»
-												•
	D	1 ·	•	fx.								
	Α	В		)	D		Е	F	G	Н	I	
1	No	調理	料理									_
2	1		)	0								_
3	2			0								
4	3		)	0								
5	4			0			新たに	シートを	田音」			_
6	5			0					_			
7	6	(		0			合成し	たい語の	列を			_
8	7			0			並べて	目上 いんナルナ	z			_
9	8			0					ວ			_
10	9		0	0								_
11	10			0								_
12	11	(	)	0								_
13	12	(		0								_
14	13		)	0								_
15	14		<u>ן</u>	0								-
16	15		) 	0								-
14 4	► N	<u>\Sheet1</u>	Sheet2	/Shee	et3/			•			•	
עדב	۲									NUM		

画面 50 語の合成



画面 51 語の合成

🔀 Mi	icros	oft Excel -	Book1					[	
8	ファイル	/E) 編集(E)	) 表示(⊻)	挿入① 書式	t@) ツール(T)	) データ( <u>D</u> )	ウィンドウ(型	) ヘルプ(円)	_ 8 ×
D	2	1 🔒 🔁 🖌	🗟 🖪 💱 🗍	X 🖻 🖻 -	🛷 🗠 - C	x - 🎑 🏾 🗴	- f* A	Z I 🛍 🚜 1	00% - 2
MS	ר P	ין ביין ביין ביין ביין ביין ביין ביין ב	11 <b>.</b> R	7 п =		9 9/	+.0 .00 <i>z</i> ≡	· · · · · ·	A _ >
				<u>x</u> <u>∩</u> ⊨		ay ∕o j	.00 +.0   ===	·   😐 ' 🚾 '	<u> </u>
	ES	5 -	Æ						
		B		-IR(D55-0,0	5,17	E C	스러 피구	× ۵ م ا± ۱+	<u> </u>
1	No	調理	料理	合計	調理・料理		百百丁」 外ル	りりの守は	0.
53	52	0	0	0	0	1	以上の時は	は1となる	Ī
54	53	0	0	0	0		問粉た)・	$h \neq z$	Γ
55	54	0	1	1	1		ぼ奴で八	/]93	
56	55	0	0	0	0				
57	56	0	0	0	0				
58	57	0	0	0	0				
59	58	0	0	0	0				
60	59	0	0	0	0				
61	60	0	0	0	0				
62	61	0	0	0	0				
63	62	0	0	0	0		「스計」	がっの時=	E -
64	63	0	0	0	0	<b>`</b>		Ŋ. Z ₩J#₫ -	
65	64	1	1		1		1になる	>	
00	05	Ü	1	1	1				
07	00	0	0	0	0				
14 4	► •	<u>\Sheet1/S</u>	heet2/Shee	et3/		•		· · ·	
ועדב	۲							NUM	

画面 52 語の合成

🔀 M	icros	oft	Exc	el – Book1									×
	ファイル	,( <u>F</u> )	編	兼(E) 表示(	⊻ 挿	た書 ゆん	(○) ツール	(T) データ(D	) ウィンドウ(W	) ヘルプ(円)		- 8	×
D	产 🛯	1.4	Ж	切り取り( <u>T</u> )		Ctrl+X	1 5-	a - 🎑 :	$\Sigma - f_{s} \stackrel{A}{=} \downarrow$	Z. I 🛍 🚜	1009	6 -	, »
MS	ー Pゴシ	יראי איל	Ê)	⊐ピ–©)		Ctrl+C		<b>9</b> %	+.0 .00 ₹	·	- A		»
			a	貼り付け(P)		Ctrl+V		~~ /o j	.00 +.0 ==	·  ··· · ···			•
	E			形式を選択	して貼り付	(†(S)							
	A			シートの買用金	÷(1.)	_	Е	F	G	Н		Ι	
1	No	調理		シートの設計	、 <u>⊑∕</u> ,≠ <i>t</i> -/+⊐µ	-00	]理·料理					_	
53	52					- <u>(191</u> 7	C				<u> </u>		
54	53		<i>6</i> 9	検索( <u>F</u> )		Ctrl+F	C		冒成し	に列を			
55	54			置換( <u>E</u> )		Ctrl+H	1		↓ 選択、	コピーす	3		_
56	55		_	0	*	h							-
57	56			0	0		0						-
50	57			0	0	- KI					-		-
60	59			0	0	- XI	C				-		-
61	60			0	ŏ	- Xi	0				-		-
62	61			0	Ō	d	C						-
63	62			0	0	d	C						
64	63			0	0	0	C						
65	64			1	1	2	1						
66	65			0	1	1	1						
67	66			0	0	0	9						
68 M 4	I 67   ▶ ₩	\Sh	eet1	/Sheet2/	Sheet3,	/			1	1		•	Ē
עדב	۲			.,,,			合計=(	3		NUM			

画面 53 語の合成



画面 54 語の合成

<b>M</b>	icro	soft Excel	- 調査	結果.xls														×
■ 読ん	ファイ) ,de!!ニ	レ(E) 編集( □(K)	(E) 表示	(⊻) 挿入	Ψ,	書式( <u>○</u> ) (□) た書	ッ	/-/J/(	D :	デー	ጵ( <u>D</u> )	ウィンド	უ₩)	ヘルプ	(H)	-	8	×
D I	<b>2</b> [	- 🔒 🖏	i 🖓	Ф <sup>р</sup> 🖌 [	<b>ì</b> (	3 - 🚿	ĸ	- C	Cil +	9	δ	• f <sub>*</sub>		l 🛍	-	100%	•	» •
MS	Pゴシ	ック	- 11 -	BZ	U		≣		9	%	, .	00. 00 00 + 00	<b>€</b> ≣	•	٩	• <u>A</u> •		»
	BL	55	•	fx 1														
	Α		E	$\searrow$	/	BI			ВJ		В	К	В	IL		BM		•
1	No		記述文		#	新鮮		生			<u>不安</u>		調理·	料理				
52	51	少人数世	<u>帯なの1</u>	で一把でフ	<u>a&lt;</u>		0			0		0		0				
53	52	新鮮なも(	<u>D。不足</u>	しないよう	574		1			0		0		0				
54	53	やはり新聞	鮮さか購	人の決め	커		1			0		0						
55	54	ホワレンン	クロは、	ומזמו			0			0		-2						
56	55	スーバー	マーケッ				0			0		0		0				
57	56	甲国産な	とで輸入	「野来の食	畏 <u>课</u>		0			0		0		0				
58	57	黒農楽や	産地かり	かから物	を員		0		_			0		0				
59	58	-∞-4 ⊾ *			<u>은 281</u>					0		1		0			+	
60	59	ਿੱਠ 'ਤ	又式」で	どはなく	•		0			0		0		0				
61	60	育ガ 「伯	す,でオ	5ろこと	- をる	隺訍				1		0		0				
62	61	/a/a "			+	CE HUS	H					0		0				
03	62	ᆘᅎᄅᅕ	小小朋友		7055		10					0		0				
64	03	八日座の	シンガオ	さる 秋開中 !	0.32 35		+			0		0						
60	04	のサ白田	1日見つし 1日1日年6日	いまり。	ບ-ງ ≐n±		4			-		0		1				
67	00	「有米はや	はツ利した話をは	いカかま	= শক		1					0		- 1			+	•
	N	(調査結)	2/	ALL					[								Þ	
עדב	۲												N	UM				

画面 55 語の合成

(2)マクロを使わずに1・0データファイルを作成する方法

最後に、マクロを使わずに1・0データファイルを作成する方法を紹介する。本手法において マクロは、単純ではあるが膨大な作業量を要する部分でのみ用いている。従って、この部分を手 作業で行うには、相当の労力を要する。しかしながら、マクロの実行において不具合が生じた場 合など、何らかの事情でマクロを利用できない事態を想定して、手作業で「1・0データ化」を 行う場合の手順を述べておく。

まず、キーワードの絞り込み(画面 27)までは、前章で述べた手順と同様である。ここから、 各キーワードがどの文章において書かれたものであるかをリストアップする。

抽出したキーワードを昇順に並べ替え(画面 56)、キーワードを示す番号の列(「形態素 No」列) とその基本形の列(「基本形」列)をそれぞれコピーし、新たに設けたシート(シート名「出現語」) の1行目、2行目に、「形式を選択して貼り付け(S)」-「行列を入れ替える(E)」で貼り付ける(画 面 57)。

ここで、「data」シートに戻る(画面 58)。「出現数」列のオートフィルタがかかった状態にあ る場合には、そのフィルタ矢印の「すべて」を選択して全データを表示させておく。そして、「形 態素 No」列のフィルタ矢印をクリックし、先に挙げたキーワード番号の1つを選択する。ここに 示された文章番号(「記述 No」列)が、そのキーワードが出現した文章であることを示すので、 これをコピーし、先に作成した「抽出語」シートに貼り付ける(画面 59)。この手順を、抽出し た全てのキーワード番号について行う。

次に、1・0データ用のシートを用意し(シート名「1・0 data」) 1列目2行目(セル A2)

から下に向けて文章番号を入力する(画面 60)。

さらに、「新規作成」で、新たに作業用のファイル(以下、「作業ファイル」)を用意する。ここ にも、1列目2行目(セルA2)から下に向けて文章番号を入力しておく(画面 61)。

「出現語」シートに戻り、1つめのキーワードついて、「基本形」とその下に入力されている文 章番号をコピーし(画面 62)、「作業ファイル」の1行目に「形式を選択して貼り付け(S)」-「行 列を入れ替える(E)」で横に貼り付ける(画面 63)。

これにより、A列縦方向に全体の文章番号、1行目横方向にキーワードの出現した文章番号が 入力されたことになる。そこで、各セルについて、表側の番号と表頭の番号が等しい場合には1、 異なる場合には0とするIF 関数を入力する(画面 64、65)。なお、IF 関数のセルをコピー&貼り 付けする場合には、表側と表頭の参照がずれないように、絶対参照「\$」をつける必要がある。 [B2 の場合:IF(\$A2=B\$1、1、0)]

全てのセルに条件式を入力したら、最後の列にカーソルを移動させ、新たに「合計」列を作る (画面 66)。ここに、各行の1・0の合計を入力する。もし、1つの文章に2回以上同じキーワ ードが出現している場合、この合計は2以上になるため、更に1・0への変換を行う(画面 67)。 「合計」列の隣に、現在1・0への変換を行っているキーワード名を入力し(画面では「ない」) 「『合計』列の数値が0の時は0、それ以外は1」という IF 関数を入力する。[M2の場合: IF(L2=0、 0、1)]

最後に、この列(「ない」列)をコピーし、「1・0 data」シートに「形式を選択して貼り付け (S)」の「値(V)」にチェックを入れて貼り付ける(画面 68)。この手順を、全てのキーワードに ついて行う。

🔀 Microsoft	Excel - 形態	索.xls						$\mathbf{X}$
🖳 ファイル(E)	編集(E) 表示	R <u>∭ tar</u> an	) 事式(() ツール	<u>m ≓-q</u> (	<u>D) ウィンドウ(M</u>	0 ヘルブ	(H)	
_ 読んde‼⊐⊐( <u>K</u> )		形	態素番号と					٦×
i 🗅 🚅 🔚 🔒	🗟 🛃 🛃 🖾	₩C ~7	れに対応した		Σ- 形	悲素霍	号で	» •
MS Pゴシック	- 11 -			±π/ \	, 號 昇	順に並	べ替える	»
F1	~	f +·	ーリート(基	4市 ( 14 日本)		1		
A	B	をj	選択しコピー <sup>.</sup>	する	F	G	Н	
1 表層語	基本形	カタカナ	品詞	t∛No /∄	態素No	見数	出現数2	
2 ない	ない	V.	形容詞一自	22	7	1	10	
3 よい	よい	1	形容詞-自	48	9	1	3	
4 安<	安い	371	形容詞-自	2	10	1	12	
5 高く	高い	高イ	形容詞-自	1	12	1	7	_
6 少ない	少ない	少ナイ	形容詞-自	8	13	1	4	
7 多い	多い	多イ	形容詞-自	4	15	1	7	
8 美味しく	美味しい	美味シイ	形容詞-自	1	18	1	9	
9 ある	ある	アレ	動詞-自立	5	21	1	18	
10 いる	いる	イレ	動詞-自立	5	23	1	5	
<u>11</u> し	する	スル	動詞-自立	1	29	1	70	
12 つい	2<	ック	動詞-自立	26	31	1	4	
13 なる	なる	1 JU	動詞-自立	<b>k</b>	35	1	13	
14 ゆで	ゆでる	ユデル	動詞-自立	22	43	+	ーワードな	を
15 喰っ	喰う	喰ウ	動詞-自立	37	51		1222 / +*	
16 使わ	使う	使り	動詞-自立	-17	58	殺	リ込んに	
17 思っ		思り	<b>動詞-白石</b>	SNL <sup>5</sup>		シ	<b>-</b> ト	_
<b>4 4 € €</b>  \ <u>∏</u> 2 	泥条人形態茶	☑∬1曲山詰	人uata人正说教》	ONE/				
1421						NOM		

画面 56 マクロを使わず「1・0 データ化」







<mark>⊠</mark> Microsoft Excel - 形態素.xls	
図 ファイル(E) 読んde!!ココ(K) キーワードの番号	書式(Q) ツール(T) データ(Q) ウィンドウ(W) ヘルプ(H) - ₽ ×
0 🛩 🖬 🔒 월 🎒 🗛 🖤 👗 🖻	🛍 • 🚿 🗠 • 🖙 🍓 Σ • f* 🛃 📶 🛷 100% 🔹 🤻
MSPゴシック • 11 • B / U	≣ ≣ ≣ ፼ % , ‰ , ☆ • <u>≫</u> • <u>▲</u> • →
B21 <b>▼ f</b> *	
A B C	DEFGH
1 形態素No 7 9	
2 基本形 ない よい 3 2 22 -	女い 高い 少ない 多い 美味しい め
4 25	
5 33	全てのキーワード番号について
6 42	
7 43	同様に又草番号をリストアッフする
8 48	
10 60	
11 87	
12 90	
14	出現した
16	→音の番号
◀ ◀ ▶ ▶ \北態素(北態素 (2)(抽出語)   ¬¬、⊮	<u> </u>
ACAE	NUM

画面 59 マクロを使わず「1・0データ化」



画面 60 マクロを使わず「1・0データ化」

🔀 Mir	rosoft Ex	cel - Book	2						
87	7ァイナ(E) 新	扁集( <u>E</u> ) 表示	500 挿入O	書式(0) ツ	ソール(T) デー	·タ( <u>D</u> ) ウィント	やし へいけ	γ( <u>H</u> ) _ I	θ×
	2 <b>D</b> A		ABC X BA	🙉 - 🛷 🗖	0 - 01 - 1	$\Sigma - f_{x}$	AL ZI III	<i></i> 100%	<b>,</b> »
									*
MIS F	-1790	+ 11 +	R I D		作業用	に別のフ	ァイルを	🖉 • 🗛 •	÷
	0.4		c .		田音才	z			
	A1	•	<i>†x</i>		用息 9	ວ 	_		
			C	D	E	F	G	H	
2	0								-
5	4								-
6	5								
7	6								
8	7								
9	8			文書番号	号を入力す	5			
10	9								
11	10								
12	11								
13	12								
14	13								
15	14								
16	15								
17		t1 /Sheet2	/Sheet3/		1				Ì
אַעדב			<u>, ,</u>		1		NUM		





画面 62 マクロを使わず「1・0データ化」







画面 64 マクロを使わず「1・0 データ化」

🔀 м	icrosoft Exce	l – Book2							X
	ファイル(圧) 編集	(E) 表示(⊻)	挿入① 書式	t( <u>O</u> ) ツール	<ul><li>(T) データ(</li></ul>	① ウィンドウ·	₩) ヘルプ(円)	_ 8	×
D	🛩 🖪 🔒 🐿	a 🐧 💖	X 🖻 🖻 -	🚿 🔊 +	CH + 🥘	$\Sigma - f_{x} = \frac{A}{Z}$	ZI 🛍 🚜	100% -	, »
MS	ー <u>-</u> Pゴシック	• 11 • B	/ Ⅱ Ξ	三三兩	9 %	+.0 .00		- A -	»
-	11	- £				3 .00 +.0		_	•
_		B	0	D -	F	F	G	н	
1	ない	- 02	25	33	42	43	48	58	
22	21			- Series	0	0	0	0	-
23	(22)		0	0	0	全てのヤノ	レに	0	
24	23		0	0	0			. 0	_
25	24	0		0	0	IF 関数を	貼り付ける	っと、p	-
26	(25)			0	0	表頭と表	則の数値が		-
27	20	0 -		<u> </u>					_
28	27	0	0			等しい時間	ま1、		-
29	28	0	0	0		それ以外に	はのとなる		-
31	30	0	0	0	— č				-
32	31	0	ŏ	- ŏ	ŏ	O	0	0	
33	32	0	0		0	0	0	0	
34	(33)	0	0	(1)	0	0	0	0	
35	34	o	o	- 🥁	0	0	0	0	
36	35	0	0	0	0	0	0	0	
37	36	0	0	0	0	0	0	0	
38	37	n /Shot2 /Sho	0	n		n	n		
עדב יין	κ ση <u>coneetr</u> γ Κ	(oneetzi/one	elay				NUM		



🔀 м	icrosoft Ex	cel – Book2				-			<
i 🔊	ファイル(E) 編 🚔 🔲 🖂 🏾	諜(E) 表示(⊻ B) ▲ B ഈ	)挿入① 書	900 「合計	宅 n と入力	( する 🛵 🎄		B列から	× »
[MS	ー EP 1 Pゴシック	- 11 - I	B Z <u>U</u>		₩ \$	+.0 .00 +.0	¥ A¥	行の合計を 計質する	• >> •
	IF	_ <del>- X V</del> )	🗣 =SUM(B2	::K2)		<u> </u>		미开 9 0	L
	A	Н	I	J	К 🚺		M	N L	*
1	ない	58	60	87	90 -	合計	-		
2	1	0	0	0	ý.	-SUM( <u>B2:K2</u>	$\searrow$		
3	2	0	0	0	0				
4	3	0	0	0	0				
5	4	0	0	0	0				
6	5	0	0	0	0				
7	6	0	0	0	0		670	いたち	
8	7	0	0	0	0		±ζυ		
9	8	0	0	0	0		合計を	を計算する	
10	9	0	0	0	0				
11	10	0	0	0	0				
12	11	0	0	0	0				
13	12	0	0	0	0				
14	13	0	0	0	0				
15	14	0	0	0	0				
16	15	0	0	0	0				
1/	16	0	0	0	0	•			-
18	► N \ Sheet	t1 /Sheet2 /S	heet3 /					  ▶[	-
参照	, (==		,		1-1		NUM		

画面 66 マクロを使わず「1・0 データ化」

<b>X</b> M	icrosoft Exce	l – Book2					
	ファイル(E) 編集	[(E) 表示(⊻) 挿入(₽)	書式(2) ツー	キーワー	ド名を	ヘルプ(円)	_ 8 ×
1D r	<b>₽ 0 0</b>	1.4 R V X B	B 9 0	列の名前	にする 🔒	1009	(
MS	ー Pゴシック	+ 11 + P I U	臣王王臣	88%	.0 .00 tE	- A - A	• »
	IF	X V 1/2 =IF(L2	=0,0,1)	- President - Frank			
	A	I J	ĸ	L	A N	0	-
1	ない	60 87	90 合	計 titu	ノ		
2	1	0 0	0	0 2=0,0	(1)		
3	2		0	0			
4	3	関数の引数				? 🛛	
5	4	IF					
6	5	論理式 L2=0	)		🔣 = TRUE		
7	6	真の場合 0			₹. = 0		<u></u>
8	/	偽の場合 1			<b>T</b> = 1		
9	8			12			
11	10	論理式の結果に広いて	と完された値を返し	. <b>#</b>	= 0		
12	11		BECHWCIE 2X50	JGK 970			
13	12						
14	13	偽の場合には、	〈論埋式〉の結果が	い FALST であったね	易合に返される値	を指定します。	
15	14						
16	15	数式の結果 =	0		'台計」:	列の値を	
17	16	<u>この関数のヘルプ(H)</u>			1 • 0 17	変換する	
18	17	/Shart? /Shart? /					
<b>行生</b>	M <u>Sneetl</u>	Koneers Voueers /				IM	
重来					NU	INIC	

画面 67 マクロを使わず「1・0 データ化」

🔀 Mi	icrosoft Ex	cel - 形態素	≅.xls						
<b>劉</b> : 読ん	ファイル( <u>E</u> ) 編 ide!!ココ(K)	<b>諜(E) 表</b> 示	⊻ 挿入Φ	書式(0) ツ	ール(工) デー	タ(D) ウイント 	でか(W) ヘルブ	<u>(H)</u>	la ×
	≥ 🛯 🔒 🤅	6 <i>6</i> 6	🌮 🐰 🗈 I	🔁 - 🝼 🔽	) + Ci +   <b>(</b>	Σ • f <sub>*</sub>	必ず「値	「」で	- - »
MSI	ー Pゴシック	• 11 •	B Z U		8 8 %	+.0	貼り付け	られて	»
-	E2	•	(fx 1)				いること	を確認	
	A	В	C	D	E	F	G	Π	
1		ない	よい	安い	高	$\checkmark$			
2	1	0	0	0		/			
 	3		0	I	0				+
5	4		0			→			+
6	5	0	0	1	0				
7	6	ſ	^ <b>- - - -</b>						
8	7	4	全てのキ	-9-61	こういて				
9	8		1.0~	の変換を征	行う				
11	9		0	0	0				+
12	11	0 O	0	1	0				+
13	12	0	0	0	0				
14	13	0	0	0	0				
15	14	0	0	1	0				
16	15	0	0	0	0				
17	▶ ▶ //形態	素②/抽出	語,/data,/出	現数/3以	L ∠(出現語)、	<u>1•Odata</u> /	•		
עדב	۴						NUM		

画面 68 マクロを使わず「1・0データ化」

5.むすび

本書では、低コストで簡便にテキストマイニングを行うために必要な1・0データファイルの 作成手順を解説した。主な作業の流れは、文章を形態素に分解する、品詞情報と出現数でキーワ ードの絞り込みを行う、各文章におけるキーワード出現の有無を1・0で示す、というものであ る。

本書で示した手順は未だ試用段階であり、マクロによる作業の自動化はごく一部でしか行って いない。市販のテキストマイニング用ソフトウエアと比較すると、繁雑な作業を要するため、膨 大な量の文章を日々の業務として分析する場合には作業効率の点で不向きであると言わざるを得 ない。

しかし、本手法は新たな投資を必要としないため、気軽に取り組むことができる。また、ほと んどの工程を手作業で行っているため、市販の専用ソフトウエアに見られるようなブラックボッ クスとなる部分がない、というメリットもある。そのため、「テキストマイニングがどのようなも のか、試しにやってみたい」という方、「たくさんの文章データを持っているがどう整理したらよ いかわからない」という方にとっては、テキストマイニングに触れる良いきっかけになるのでは ないかと考える。

本書の末尾には、本手法で使用したマクロのコードを紹介している。また、本書では、マクロ を使用しない場合の1・0データファイル作成手順についても補足説明を行っているので、必要 に応じて参考にしていただきたい。

本手法の適用により、これまで放置されていた大量の文章データから、有益な情報が抽出され るようになれば幸いである。

#### 参考文献

- 1)市村由美・長谷川隆明・渡部勇・佐藤光弘(2001): テキストマイニング 事例紹介、人工知 能学会誌、16巻2号、192-200
- 2)磯島昭代(2002):米に関する自由記述回答文の分析、農業経営通信214、26-29.
- 3)磯島昭代(2004): 定性情報分析の方法 テキストマイニング 、東北農研総合研究(A)第 15号、10-14.
- 4)磯島昭代・野中章久・清野誠喜(2004):テキストマイニングによるクレームデータの分析、
   農業経営研究、第42巻第1号(通巻120号) 148-152
- 5) 林俊克 (2002): Excel で学ぶテキストマイニング入門、オーム社



```
付録Ⅱ マクロのコード
Sub 茶坊主くん()
   Dim gyo As Integer, retu As Integer
   Dim goku As Integer, keygyo As Integer
   Dim bun As Integer, bunretu As Integer
   Dim cnt As Integer, cnt2 As Integer
   Dim kaz As Integer, bunkazu As Integer
   Dim i As Integer, j As Integer
   Dim motodata As String, ichizerodata As String
   retu = 5 'motodata の文章番号がある列
   goku = 6 'motodata の形態素番号がある列
   bunretu = 1 'ichizerodata の文章番号がある列
   keygyo = 1 'ichizerodata のキーワード番号がある行
   motodata = "data" '元のデータのあるシート
   ichizerodata = "1 • 0" '1 • 0データを作成するシート
   gyo = 2 'motodata の文章番号の行:初期値
   bun = 2 'ichizerodata の文章番号の行:初期値
   kaz = 0 'キーワード数:初期値
   bunkazu = 0'全文章数:初期值
   Do While Worksheets(ichizerodata).Cells(keygyo, kaz + bunretu + 1) <> ""
      kaz = kaz + 1
      'キーワードの数をカウント
   Loop
   MsgBox "抽出したキーワードの数は" & kaz
   Do While Worksheets(ichizerodata).Cells(bunkazu + keygyo + 1, bunretu) <> ""
      bunkazu = bunkazu + 1
      '全文章数をカウント
   Loop
   MsgBox "1・0データ化する文章の数は" & bunkazu
   Cells(keygyo, bunretu).Select
   'セル(1,1)をアクティブにする
   Cells(keygyo, bunretu).Interior.ColorIndex = 6
   'セル(1,1)に色を付ける
   Do While bun <= bunkazu + keygyo
   ichizerodataの文章番号の行を全て参照するまで行う。
       Worksheets(ichizerodata).Cells(keygyo, bunretu).Value = bun - 1
      'icizerodata のセル(1,1)に、現在操作中の文章番号を記入
      If Worksheets(motodata).Cells(gyo, retu) <> "" Then
```

```
- 52 -
```

'motodataの文章番号のセルが空白でなければ

```
Select Case Worksheets(ichizerodata).Cells(bun, bunretu)

'ichizerodataの文章番号が

Case Is < Worksheets(motodata).Cells(gyo, retu)

'motodataの文章番号より小さいとき

For i = 2 To kaz + 1

Worksheets(ichizerodata).Cells(bun, i).Value = 0

'値は一行全て0

Next i
```

```
Case Is = Worksheets(motodata).Cells(gyo, retu)
' (ichizerodataの文章番号が) motodataの文章番号と等しいときは
   cnt = bunretu + 1
   'ichizerodataの2列目(キーワードのある列)から
   Do While cnt <= kaz + bunretu
   'ichizerodataの全てのキーワード番号の列を参照するまで繰り返す
      Select Case Worksheets (ichizerodata). Cells (keygyo, cnt)
      'ichizerodata のキーワード番号が
      Case Is < Worksheets(motodata).Cells(gyo, goku)
      'motodataの形態素番号より小さいときは
          Worksheets(ichizerodata).Cells(bun, cnt).Value = 0
          '値を0にして
          cnt = cnt + 1
          '次のキーワード番号の列へ
          If cnt > kaz + bunretu Then
          'もし, ichizerodataの全てのキーワード番号を参照してしまったら
             Do While Worksheets(motodata).Cells(gyo, retu) = _
                 Worksheets(motodata).Cells(gyo + 1, retu)
             'motodataの文章番号が変わるまで行を進める
                gyo = gyo + 1
             Loop
          End If
      Case Is = Worksheets(motodata).Cells(gyo, goku)
       'motodataの形態素番号と等しいときは
          Worksheets(ichizerodata).Cells(bun, cnt).Value = 1
          '値を1にする
          If Worksheets(motodata).Cells(gyo, retu) = _
             Worksheets(motodata).Cells(gyo + 1, retu) Then
          'motodataの次の文章番号が変わらない場合
```

```
gyo = gyo + 1
      '同じ文章番号の次の形態素番号の行に移り
      cnt = cnt + 1
      'ichizerodataの次のキーワード番号に移る
      If cnt > kaz + bunretu Then
      'もし, ichizerodataの全てのキーワード番号を参照してしまったら
         Do While Worksheets(motodata).Cells(gyo, retu) = _
             Worksheets(motodata).Cells(gyo + 1, retu)
          'motodataの文章番号が変わるまで行を進める
             gyo = gyo + 1
         Loop
      End If
   Else
   'motodataの文章番号が変わった場合
      For i = cnt + 1 To kaz + 1
         Worksheets(ichizerodata).Cells(bun, i).Value = 0
         '残りの値を0にして
      Next i
      Exit Do
      ,ループを抜ける
   End If
Case Is > Worksheets(motodata).Cells(gyo, goku)
'motodataの形態素番号より大きいとき
   If Worksheets(motodata).Cells(gyo, retu) = _
      Worksheets(motodata).Cells(gyo + 1, retu) Then
   'motodataの次の文章番号が同じ場合
      gyo = gyo + 1
      ,同じ文章番号の次の形態素番号の行に移り
   Else
   'motodataの文章番号が変わった場合
      For i = cnt To kaz + 1
         Worksheets(ichizerodata).Cells(bun, i).Value = 0
         '残りの値を0にして
      Next i
      Exit Do
      'ループを抜ける
   End If
End Select
```

```
Loop

'ichizerodataの最後のキーワード番号の列を参照するまでループ

gyo = gyo + 1

'motodataの次の文章番号の行に移る

Case Is > Worksheets(motodata).Cells(gyo, retu)

'(ichizerodataの文章番号が) motodataの文章番号より大きいときは
```

End Select

### Else

```
'motodataの文章番号が空白になったら
```

```
For i = 2 To kaz + 1
Worksheets(ichizerodata).Cells(bun, i).Value = 0
,値は全て0にする
Next i
End If
```

bun = bun + 1 'ichizerodataの次の文章番号について行う Loop

End Sub

付録Ⅲ 「茶坊主くん.txt」からVBファイルを作成する

本文中では、「茶坊主くん」のマクロは作成済みという前提で解説しており、公開しているマク ロのコードを参考に各自で対応していただくことにしている。しかし、実際にコードを入力する 作業は面倒と感じられる方も多いと思われるので、テキスト形式でマクロのコードを配布するこ とにした。テキスト形式としたのは、セキュリティ上の問題に配慮したためであるが、テキスト 形式のファイルでは、そのままマクロとして実行することはできない。

そこで、以下では、テキスト形式のファイル「茶坊主くん.txt」から、マクロが実行可能なV Bファイル「茶坊主くん.bas」を作成する手順を紹介する。なお、ここではVBファイルの作成 について述べるが、もちろん、1・0データファイル作成の過程で、テキストのコードをVBE (Visual Basic Editor)のコードウィンドウに直接コピー&ペーストして実行していただいても 一向に構わない。

1. Excel の「ツール(T)」-「マクロ(M)」-「Visual Basic Editor(V)」を選択してVBEを 起動し、「挿入(I)」-「標準モジュール(M)」をクリックする。



2.「茶坊主くん.txt」のファイルを開き、コードをコピーする。

◎ 茶坊主くん.txt - メモ帳	X
ファイル(E) 編集(E) 書式(Q) 表示(V) ヘルプ(H)	
Sub 茶坊主くん()	^
Dim gyo As Integer, retu As Integer	
Dim goku As Integer, keygyo As Integer Dim burn As Integer, humantha As Integer	=
Dim bun As Integer, bunretu As Integer Dim opt ås Integer, opt2 ås Integer	-
Dim kaz As Integer, kunkazu As Integer	
Dim i As Integer, j As Integer	
Dim motodata As String, ichizerodata As String	
retu - 5 motodata())))文卓省与かめる列 golu = 6 motodataの形能表悉早がある列	
bunretu = 1 'ichizerodataの文音番号がある列	
keygyo = 1 'ichizerodataのキーワード番号がある行	
motodata = "data" '元のデータのあるシート	
ichizerodata = ´1・O´ `1・Oデータを作成するシート	
evo = 2 'motodataの文音悉号の行・初期値	
bun = 2 ichizerodataの文章番号の行:初期値	
kaz = 0 'キーワード数: 初期値	
bunkazu = 0 '全文章数:初期値	
	~

3. VBEのコードウィンドウにコードを貼り付ける。

着 Microsoft Visual B	asic - Book1 - [Module1 (]- f)]	
■ 🦂 ファイル(E) 編集(E) ウィンドウ(W) ヘルブ(H)	表示(V) 挿入(P) 書式(Q) デバッグ(D) 実行(R) ツール(T) アドイン(A) -	₽×
🛛 🐮 - 🔚 🕺 🖻	ÊM ∽ ∽ → II ■ M 🗞 🖻 🥞 🛠 🛛	» *
プロジェクト - VBAProje	(General) ▼ 茶坊主くん	•
Image: Second	Sub 茶坊主くん() Dim gyo As Integer, retu As Integer Dim goku As Integer, keygyo As Integer Dim bun As Integer, bunretu As Integer Dim cnt As Integer, cnt2 As Integer Dim kaz As Integer, bunkazu As Integer Dim i As Integer, j As Integer Dim i As Integer, j As Integer Dim motodata As String, ichizerodata As String	
全体 項目別   付ブジェクト名 Module1	retu = 5 'motodataの文章番号がある列 goku = 6 'motodataの形態素番号がある列 bunretu = 1 'ichizerodataの文章番号がある列 keygyo = 1 'ichizerodataのキーワード番号がある行 motodata = "data" '元のデータのあるシート ichizerodata = "1・0" '1・0データを作成するシート	
	gyo = 2 'motodataの文章番号の行:初期値 bun = 2 'ichizerodataの文章番号の行:初期値 kaz = 0 'キーワード数:初期値 bunkazu = 0 '全文章数:初期値	
	Do While Worksheets(ichizerodata).Cells(keygyo, kaz + bur kaz = kaz + 1 'キーワードの数をカウント	ırı
Gサブジェクト名 Module1	retu = 5 'motodataの文章番号がある列 goku = 6 'motodataの形態素番号がある列 bunretu = 1 'ichizerodataの文章番号がある列 keygyo = 1 'ichizerodataの大章番号がある行 motodata = "data" '元のデータのあるシート ichizerodata = "1・0"'1・0データを作成するシート gyo = 2 'motodataの文章番号の行:初期値 bun = 2 'ichizerodataの文章番号の行:初期値 kaz = 0 'キーワード数:初期値 bunkazu = 0 '全文章数:初期値 Do While Worksheets(ichizerodata).Cells(keygyo, kaz + bur kaz = kaz + 1 'キーワードの数をカウント	וזרי •

4.「ファイル(F)」-「ファイルのエクスポート(E)」を選択し、ファイル名を「茶坊主くん」として保存する。

者 Microsof	ft Visual Basic – Book1 – [Module1 (コード)]	
🤻 ファイル(! ウィンドウ(!!)	E) 編集(E) 表示(V) 挿入(D) 書式(D) デバッグ(D) 実行(B) ツール(T) アドイン(A) ) ヘルブ(H)	_ 8 ×
X 13 - C	- ½ = C + + + + + + ≥ ≥ C + 2	» •
ブロジェクトーヽ	ファイルのエクスポート 🔹 🛛 💽	•
	(保存する場所型: 🔜 3.5 インチ FD (A:) 🗸 🗢 🗈 📸 🖽	-
400 MU		
ブロパティー Mo		
Module1 Mod		
全体 項目		
はブジェクト名	·	
	ファイル名(N): 茶坊主(ん.bas 保存(S)	
	ファイルの種類(I): 標準モジュール (*.bas) <	
	bun = 2 'ichizerodataの文章番号の行:初期値	
	kaz = 0 キーワー F 数:初期値 bunkazu = 0 '全文章数:初期値	
	Do While Worksheets(ichizerodata).Cells(keygyo, kaz +	bunr
	kaz = kaz + 1 'キーワードの数をカウント	