

# 陸上無線技術士国家試験問題のテキストマイニング — ユーザ辞書の作成 —

松田豊稔\*

## Text Mining on Questions of the National Examination for Technical Radio Operator for On-the-Ground Services - Creating a User-defined Dictionary -

Toyonori Matsuda\*

KH Coder, which is a powerful free software for text mining, has a feature of a user-defined dictionary to adopt particular needs of users. In the report, a user-defined dictionary is created for content analysis of questions of the national examination for Technical Radio Operator for On-the-Ground Services in Japan: Technical terms necessary for learning radio engineering are extracted from all the questions in the national examinations that were ever carried out 36 times and then the technical terms are registered into the user-defined dictionary. Two examples of the content analysis with the user-defined dictionary are included.

キーワード：テキストマイニング、KH Coder、ユーザ辞書、陸上無線技術士

Keywords：Text Mining, KH Coder, User-defined Dictionary, Technical Radio Operator for On-the-Ground Services

### 1. はじめに

テキストマイニング<sup>(1)</sup>は、大量のテキストデータの中から自動的に語句を抽出し、その抽出した語句に対して検索・集計そして種々の統計手法やグラフ理論を用いた計量的な分析を行い、テキストデータが持つ特徴的なパターンや一定のルールなど有用な情報を取出す技術である。近年、様々なテキストマイニングのソフトウェアが開発され、その有用性が認められるようになり、テキストマイニングが企業等での商業目的から学術研究や教育現場など幅広い分野で利用されている<sup>(2)</sup>。

筆者は、テキストマイニング用のフリーソフトウェアの KH Coder 2<sup>(1)・(2)</sup>（以下、KH Coder）を用いて、第一級陸上無線技術士国家試験<sup>(3)</sup>（以下、一陸技）の試験科目「無線工学 B」の内容分析を行っている<sup>(4)</sup>。一陸技は、無線通信に用いる設備の技術操作を行うための最上位の国家資格で大学卒業レベルの内容で、技術範囲も電波の基礎理論から電波伝搬、アンテナ系等の理論、構造及び機能の詳細から測定装置まで幅広い。このように「無線工学 B」の国家試験は、出題範囲が広く、その内容は専門

的知識に加えて、電気・電子・通信の予備知識を要するものであり、受験者にとって試験の全体構成を理解するのは難しい。そこで、筆者は、「無線工学 B」の国家試験問題にテキストマイニングを行い、出題される設問や内容を分析し、その関連性や共通性を系統的にまた視覚的に把握できる学習支援用教材の制作を目指している。

KH Coder では、ユーザ辞書（利用者が指定する語句）を登録すれば、その語句は“タグ”という品詞で認識され、集計や分析の単位として指定でき、内容分析の基礎データとなる<sup>(1)</sup>。テキストマイニングで有用な情報を得るには、内容分析の目的に応じたユーザ辞書を用意する必要がある。ユーザ辞書の作成には、多くのテキストデータから語句を抽出し、その統計解析を行い、内容分析に反映される抽出語を選択しなければならない。

本報告では、一陸技「無線工学 B」の試験問題の内容分析に用いるユーザ辞書の作成を目的として、平成 19 年 7 月期から令和 2 年 1 月期までに実施された 36 回の試験問題から専門用語（以下、学習項目）を抽出し、それを集計・分析する。その分析結果をもとに、ユーザ辞書に登録する学習項目を選定する。さらに、作成したユーザ辞書を用いて全 36 回の試験問題の KH Coder による内容分析を行い、各期の試験で特徴のある学習項目を抽出し、その学習項目と各期の試験との関連性を視覚的に示す共起ネットワーク<sup>(1)</sup>と対応分析<sup>(1)</sup>の結果を示す。

電子情報システム工学系  
〒861-1102 熊本県合志市須屋 2659-2  
Faculty of Electronics and Information Systems Engineering,  
2659-2 Suya, Koshi-shi, Kumamoto, Japan 861-1102

\* Corresponding author:  
E-mail address: tmatsu@kumamoto-nct.ac.jp (T. Matsuda).

## 2. 分析結果及び考察

### 2.1 テキストデータファイル

KH Coder では、テキストデータを「章」・「節」・「文」といった階層構造として構成でき、これらを集計や分析の単位として指定することができる<sup>(4)</sup>。本調査では、表1に示すように、平成19年7月期から令和2年1月期までの36回の各試験を「章」に、その中の各設問を「節」に対応させ、「文」は句点により区別される。例えば、平成14年7月期の試験は、第1章に“H14-07”という見出しが付けられ、第1章中に、(設問数が25問なので)25の節と183個の文があることを示している。

表1 テキストデータの章の構成

試験	章	見出し	節	文	学習項目出現回数
平成14年7月期	1	H14-07	25	183	444
平成15年1月期	2	H15-01	25	203	548
平成15年7月期	3	H15-07	25	197	526
平成16年1月期	4	H16-01	25	178	420
平成16年7月期	5	H16-07	25	184	421
平成17年1月期	6	H17-01	25	171	421
平成17年7月期	7	H17-07	25	172	415
平成18年1月期	8	H18-01	25	173	363
平成18年7月期	9	H18-07	25	167	381
平成19年1月期	10	H19-01	25	186	448
平成19年7月期	11	H19-07	25	179	481
平成20年1月期	12	H20-01	25	188	468
平成20年7月期	13	H20-07	25	187	514
平成21年1月期	14	H21-01	25	188	509
平成21年7月期	15	H21-07	25	182	418
平成22年1月期	16	H21-01	25	184	420
平成22年7月期	17	H21-07	25	182	408
平成23年1月期	18	H23-01	25	184	476
平成23年7月期	19	H23-07	25	170	455
平成24年1月期	20	H24-01	25	177	421
平成24年7月期	21	H24-07	25	210	473
平成25年1月期	22	H25-01	25	211	551
平成25年7月期	23	H25-07	25	205	449
平成26年1月期	24	H26-01	25	216	418
平成26年7月期	25	H26-07	25	211	515
平成27年1月期	26	H27-01	25	237	514
平成27年7月期	27	H27-07	25	230	537
平成28年1月期	28	H28-01	25	229	463
平成28年7月期	29	H28-07	25	201	467
平成29年1月期	30	H29-01	25	200	444
平成29年7月期	31	H29-07	25	207	475
平成30年1月期	32	H30-01	25	188	484
平成30年7月期	33	H30-07	25	183	447
令和元年1月期	34	R01-01	25	181	541
令和元年7月期	35	R01-07	25	185	506
令和2年1月期	36	R02-01	25	184	452
合計	全36章		900	6913	16693

### 2.2 形態素解析とユーザ辞書

KH Coder では、前処理としてテキストデータ中の文を品詞単位(名詞、動詞など)の単語で区切り、その語を品詞別に抽出する形態素解析が行われる。形態素解析には辞書が必要であるが、特に指定しないとシステム辞書「茶釜(ちやせん)」<sup>(5)</sup>が用いられる。しかし、「茶釜」は一般的な辞書であり、ユーザが指定する語句を抽出するにはユーザ辞書が必要である。例えば、「送信アンテナ」は“送信”と“ア

ンテナ”と区切るのではなく、一語の複合語とする。このように、内容分析の対象となる語句や複合語を指定するのがユーザ辞書であり、その作成手順を次に述べる。

表1のテキストデータをシステム辞書「茶釜」を用いて形態素解析を行うと、品詞毎に語句が抽出される。この抽出された語句・複合語の中から「無線工学 B」の学習に必要な語句と複合語を学習項目として選定や作成を行う。ユーザ辞書は、学習項目の集合体で、実体は学習項目がセルに格納された csv ファイルである。以上の作業により、学習項目の総数が 559 個のユーザ辞書が作成された。表1の列「学習項目出現回数」は、ユーザ辞書に登録されている学習項目が各期の試験での出現回数を表し、合計 16693 は全 36 回の試験での学習項目の総出現回数である。

### 2.3 頻出語検索の結果

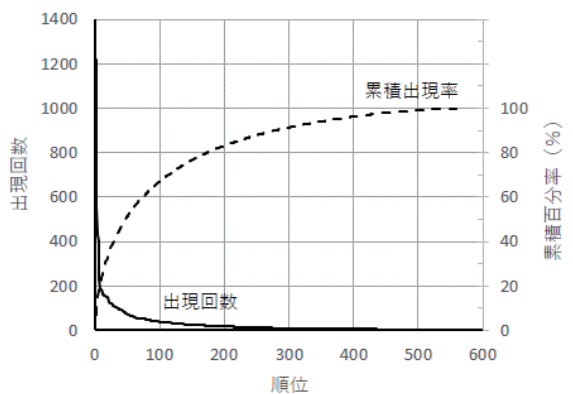
表1のテキストデータを対象として、ユーザ辞書の学習

表2 頻出語検索結果(上位50位)

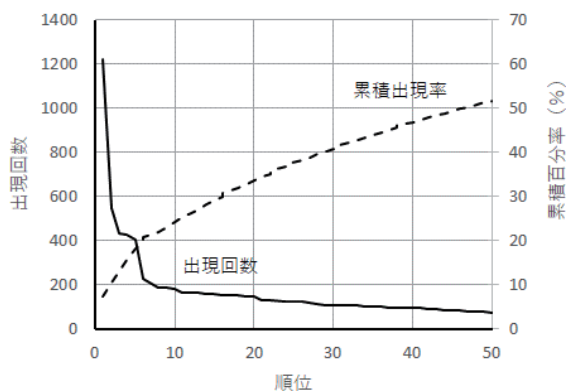
順位	学習項目	出現回数	出現率 %	累積出現率 %
1	アンテナ	1221	7.3	7.3
2	電波	546	3.3	10.6
3	周波数	431	2.6	13.2
4	インピーダンス	428	2.6	15.7
5	波長	402	2.4	18.1
6	利得	229	1.4	19.5
6	放射	229	1.4	20.9
8	入力	187	1.1	22.0
9	電界強度	185	1.1	23.1
10	指向性	183	1.1	24.2
11	給電線	166	1.0	25.2
12	受信アンテナ	164	1.0	26.2
13	電流	163	1.0	27.2
14	電界	158	0.9	28.1
15	伝搬	156	0.9	29.0
16	半波長ダイポールアンテナ	154	0.9	30.0
16	自由空間	154	0.9	30.9
18	導波管	150	0.9	31.8
19	電力	149	0.9	32.7
20	真数	148	0.9	33.6
21	整合	130	0.8	34.3
22	電磁波	129	0.8	35.1
22	dB	129	0.8	35.9
24	反射	124	0.7	36.6
24	絶対利得	124	0.7	37.4
26	送信アンテナ	123	0.7	38.1
27	偏波	119	0.7	38.8
28	受信点	112	0.7	39.5
29	送信	109	0.7	40.1
30	平面反射板	107	0.6	40.8
30	磁界	107	0.6	41.4
32	受信	106	0.6	42.1
33	開口面	105	0.6	42.7
34	モード	102	0.6	43.3
35	誘電体	101	0.6	43.9
36	減衰	99	0.6	44.5
37	電離層	95	0.6	45.1
38	出力	94	0.6	45.6
38	給電点	94	0.6	46.2
40	受信機	93	0.6	46.8
41	フェージング	92	0.6	47.3
42	導体	89	0.5	47.8
43	伝送	87	0.5	48.4
44	位相	85	0.5	48.9
45	雑音	83	0.5	49.4
46	反射波	81	0.5	49.9
47	平面波	78	0.5	50.3
48	実効面積	77	0.5	50.8
49	位相差	76	0.5	51.2
50	損失	73	0.4	51.7

項目の頻出語検索（出現回数を調べ、出現回数順に語句を並べる）を行った。表2は、出現回数上位50位までの学習項目を示している。第1位の“アンテナ”は、36期分の全データで1221回出現し、出現率7.3%は学習項目の総出現回数に対する割合(=1221/16693)である。累積出現率(%)は、第1位からその順位までの出現率を加えたもので、例えば10位“指向性”の累積出現率24.2%は、10位までに総出現回数の24.2%が出現することを表している。表2の結果から、学習項目の出現は上位に集中し、14位以下では出現率1%未満であることが分かる。

図1は、横軸に出現順位を縦軸に出現回数及び累積出現率をとった学習項目の出現回数の分布である。図1の(a)は全学習項目(559位まで)、(b)は上位50位までの出現回数の分布である。出現回数の分布は、第1位の“アンテナ”が突出し、第10位ぐらいまで急に減少し、それ以降は緩やかな減衰が続く。出現回数が多い学習項目は、各期の試験に複数回出現するもので、「無線工学B」の出題範囲全体に共通する基礎的内容の学習項目である。一方、「無線工学B」の専門性を表す学習項目は、出現回数が少なく、緩やかな減衰が続く分布の中に含まれる傾向がある。このことは、「無線工学B」の専門分野の出題範囲が広いことを示唆している。このように、学習項目の出現回数の分析から「無線工学B」の試験問題の内容が調べられる。



(a) 全学習項目に対する分布



(b) 上位10位までの分布

図1 学習項目の出現回数分布

## 2.4 内容分析の実施例

本報告で作成したユーザ辞書を用いて、KH Coderによる「無線工学B」の36回の試験問題の内容分析を実施した例を示す。本報告では、頻出語検索で求められた出現回数上位の学習項目(表2参照)と各期の試験との関連性を共起ネットワーク<sup>(1)</sup>と対応分析<sup>(1)</sup>により調べた。

図2(次頁)は、学習項目(図の丸印)と各期の試験(図の四角で見出しで表記)の共起ネットワークの結果であり、各期の試験で特徴のある学習項目(つまりその学習項目を含む文が多いこと)を視覚的に把握できる。図の円の大きさはその学習項目の出現回数に対応し、円の色は各期の試験との結びつきの多さを表している。学習項目と各期の試験を結ぶ線は、両者の結びつきを定量化したもので、線の太さが結びつきの強さを表し、結びつきが一定値以下では線は省略される。

図2の共起ネットワークは学習項目と各期の試験との関連性及びその程度を視覚的に把握できるが、その位置関係は情報を持っていない。図3(次頁)に示す学習項目と各期の試験の対応分析は、学習項目と各期の試験を座標で表し、その位置関係から学習項目と各期の試験の関連性を定量的に示したものである。

## 3. まとめ

KH Coderによる第一級陸上無線技術士国家試験の「無線工学B」の試験問題の内容分析に用いるユーザ辞書を作成した。ユーザ辞書は「無線工学B」の学習に必要な学習項目で構成されている。ユーザ辞書の中から内容分析の対象となる学習項目を選ぶことで、目的とする内容分析が可能となる。例えば、本報告では、頻出度上位の学習項目と各期の試験の関連性を調べる共起ネットワークと対応分析の例を示した。今後、作成した辞書を用いて「無線工学B」の試験問題の内容分析を行い、学習支援教材のためのデータを収集していく予定である。

謝辞 本研究は、JSPS 科学研究費助成事業 18K02887の助成によるものである。また、本稿の調査で資料作成に協力してくれた本校情報通信エレクトロニクス工学科5年の中村亘希君に深謝します。

(令和2年9月25日受付)

(令和2年12月7日受理)

## 参考文献

- (1) 樋口耕一:「社会調査のための計量テキスト分析」, ナカニシヤ出版, pp. 1-30, (2014), 東京.
- (2) <https://khcoder.net/>, 「2020.9.25閲覧」
- (3) 電波法 第40条第1項第4号イからロ
- (4) 松田豊稔:「テキストマイニングによる陸上無線技術士国家試験問題の内容分析」, 熊本高等専門学校研究紀要, vol. 11, pp. 94-97 (2020).
- (5) <http://chasen-legacy.osdn.jp>, 「2019.10.25閲覧」

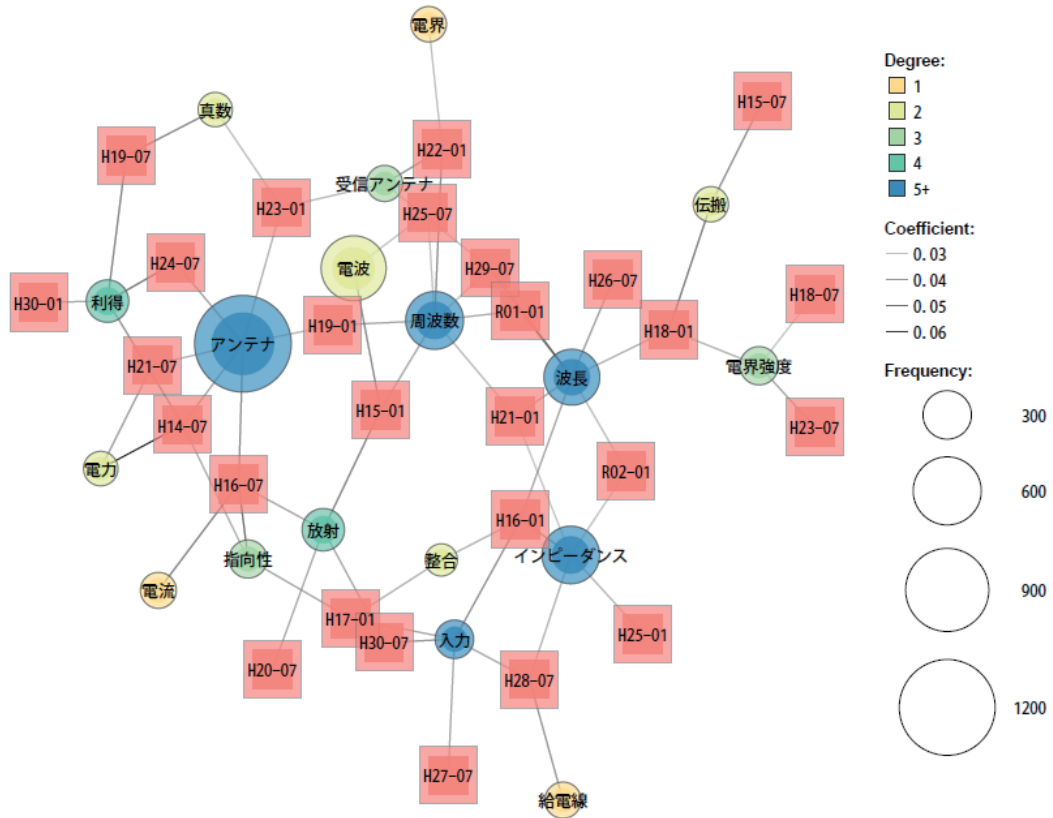


図2 内容分析の実施例：出現数上位の学習項目と各期の試験との共起ネットワーク

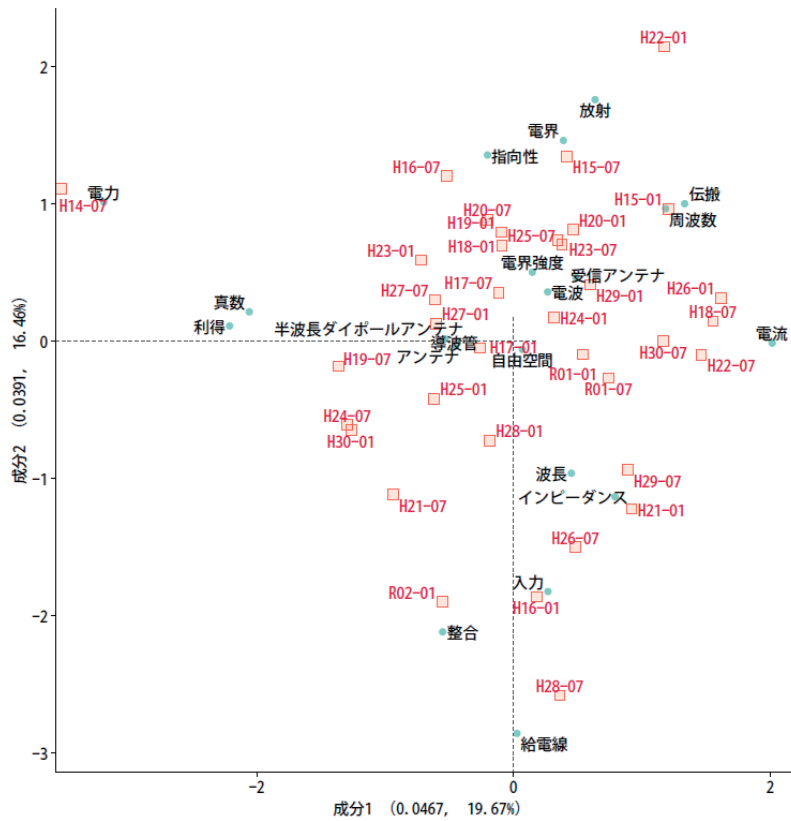


図3 内容分析の実施例：出現数上位の学習項目と各期の試験との対応分析