

家電製品マニュアルコーパスの構築と その利用

佐野 洋[†] 桑野 裕康^{††}

[†] 東京外国語大学 大学院総合国差額研究院

^{††} パナソニック株式会社 コーポレート R & D 戦略室

筆者等は、製品取扱説明書のことばの表現特徴の調査を進めている。調査では、製品の機能や操作手順説明を行うための表現としての正しさ、説明の正確さや分かりやすさを定量的に計測する手段を探っている。そして、その手段に基づき計量されたポイントを手続的に改善する方法も併せて探っている。本稿では、調査のために構築した製品取扱説明書コーパスについて概説する。このコーパスを使って、製品取扱説明書の文末表現を調査した結果について報告する。

Developing Corpus of Home Electric Appliance Manuals and its Application

SANO, Hiroshi[†] Hiroyasu Kuwano^{††}

[†] Tokyo University of Foreign Studies, Institute of Global Studies

^{††} Panasonic Corporation

This paper introduces the process of corpus development on home electric appliance manuals (in Japanese) and reports some results of the investigation using the corpus. The authors seek how manual writings are quantitatively-measured in order to clarify the correctness and appropriation of the explanation. Focusing on the types of the end of the sentences in the corpus, an experimental quantitative method and improvement technique of manual writings are suggested here.

1. はじめに

筆者等は、製品取扱説明書の言語表現を、日本語マニュアル表現の複雑さの観点から調査している。調査では、説明表現としての日本語の正しさや、内容説明の正確さや分かりやすさを定量的に計測する手段を見いだすこと、並びにその手段に基づき計量されたポイントを手続的に改善する方法を探っている。

本稿では、調査のために構築した製品取扱説明書コーパスについて概説する。さらに構築したコーパスを使って製品取扱説明書の文末述部表現を分析した結果について報告する。

日本語の特徴は、(1) 述部に文法要素が集中し、それらは形態的に顕れる、(2) 主述部は文末に置かれる、ことから本報告の調査では、製品取扱説明書の語彙分布に加えて、文末述部表現も調査した。仮定もしくは仮説として、述部表現の形態特徴を使った定量分析によって、複雑さの指標が得られると考える。もしこうした指標が得られれば、その指標によって文章特徴を客観的に比較することができ、客観評価が可能となる（複雑さの自動計算や、複雑な表現の自動修正の可能性が判断できるだろう）。

2. 製品取扱説明書と文体

工業技術を中心にした理屈によって作られている機器・装置である家電製品は、広く我々の社会に浸透し、家電製品を想定しない生活はあり得ない。さらに携帯電話やパソコンのように高度情報化社会のインフラとして機能している製品もある。

製品開発における重要な課題のひとつは、相応のコストで利用者のニーズに応える製品を届けることである。このコストには、技術開発コスト、製造コストや販売コストだけでなく、利用者が負担する製品機能の学習コストも含まれる。

とくに利用者が負担する学習コストは、製品のユーザインタフェースの実現の方法に強く依存しているほか、製品取扱説明書の可読性に依存している。利用者の認識行動を、意味論と構文論のモデルにした一般的な理論も存在し、ユーザインタフェースの開発に利用されている。マニュアルの可読性の向上のため、インデックスの工夫（必要機能の特定が早いこと）や、ビジュアル化などメディア複合（読むことに加え、見ることによって機能が理解でき、操作技能に結びつけられること）の取り組みが行われてきた。

家電製品は、すべての利用者に等価機能を提供し、反復して同等機能を提供するよう合理的に作られている。理想では、機能実現の仕組みやメカニズムを、筋道を違わず説明すれば、利用者は分かるはずである。イデオロギーや主義主張、評論ではないのであって、理屈や論理を伝える書き方や表現で説明すればよい。そして、理屈や論理を伝える書き方の表現特徴が定量化のポイントになる。筆者等は、[1]で「ことばの公共性」という基準を提案している。

3. コーパス構築

研究目的は、家電製品のマニュアル文章の複雑さの定量評価である。その調査のため家電製品マニュアルコーパスを構築した。

3.1 マニュアルの種類と数

表 1 に、コーパス化した製品取扱説明書の種類と数を示す。コーパス化対象のマニュアルは、HTML ファイルと PDF ファイルである。総計 116 ファイルで、35HTML ファイル、81PDF ファイルである。なお、製品分類（大小カテゴリ）名は、製造元の製品紹介ホームページに記載の用語を利用した。前者は、簡単にテキスト抽出が可能だが、後者は、基本的に手作業のため手間を要した。テキスト変換ソフトウェアを一部利用したが、図表を多用するデザインのため、人手によるテキスト部分の構造回復作業を行った。

表 1 製品取扱説明書の種類と数

ファイル番号	製品分類（大カテゴリ）	製品分類（小カテゴリ）
1, 2	調理	IH クッキングヒーター
3, 4	通信・ネットワーク	ネットワークカメラ
5~8	カーナビ・カーAV	カーAV
9	ワープロ	ワープロ
10, 11	カーナビ・カーAV	ETC
12	水回り	温水洗浄機
13, 14	カメラ	デジタルカメラ
15~17	BD・DVD・ビデオ	DVD レコーダー
18, 19	美容	ヘアケア・フェイスクア
20	空調・季節	加湿器
21	空調・季節	空気清浄機
22, 23	ワープロ	ワープロ
24	空調・季節	除湿器
25, 26, 45	パソコン	ノート PC
27~44	カーナビ・カーAV	カーナビ
46	通信・ネットワーク	ネットワークカメラ
47~51	パソコン	ノート PC
52	通信・ネットワーク	I/O コネクタ（ネットワークカメラ

		用)
53	電話機・FAX・テレビドアホン	ファクス
54	調理	IH 調理器
55, 56	パソコン	内蔵モデムコマンド一覧（ノート PC）
57	家事	洗濯乾燥機
58	調理	レンジ
59, 60	家事	食器洗い乾燥機
61	保冷・保温	冷凍冷蔵庫
62	カメラ	デジタルビデオカメラ
63	パソコン	ノート PC
64, 71, 81	通信・ネットワーク	ネットワークカメラ
65~70	パソコン	ノート PC
72~76	パソコン	ノート PC
77	水回り	温水洗浄便座
78	健康	マッサージチェア
79	健康	フィットネス機器
80	健康	血圧計
82	通信・ネットワーク	ネットワークカメラ
83~93	パソコン	ノート PC
94, 95	オーディオ	IC レコーダー
96~98	オーディオ	コンボ
99	通信・ネットワーク	SD カード（ネットワークカメラ用）
100	調理	ホームベーカリー
101	調理	炊飯器
102~107	テレビ	液晶プロジェクタ
108, 109	オフィス	電子黒板
110	カメラ	デジタルビデオカメラ
111	電話機・FAX・テレビドアホン	電話機
112	空調・季節	ルームエアコン
113	水回り	家庭用ヒートポンプ給湯器
114~116	調理	IH クッキングヒーター

3.2 タグ付きテキスト

マニュアルでは、メディア複合が進んでいる。タイトルや図表などの文章構造の情報もできるだけ保存したいことから文書構造タグを設定した上で、文章部分のテキスト化を行った。図 1 にタグ付けを施したテキスト化例を示す。

```
<manual type="PC" name="FW_U1CSD500_600.pdf" page="" pic="">
<head>最初にお読みください</head>
<head>パーソナルワープロ</head>
<head>取扱説明書</head>
<head>品番 FW-U1CSD500</head>
FW-U1CSD600
<head>入門ガイド</head>
<head>説明書の構成</head>
<fig></fig><head>取扱説明書</head>
<head>入門ガイド（本書）</head>
本体の取り扱い，ワープロの基本操作を知りたいときや，本体の異常で困ったときに。
また，文書作りの応用操作や住所録のはがき宛名印刷など代表的な機能を実例で説明しています。
<head>活用ガイド</head>
日本語ワープロや住所録を詳しく知りたい，スキャナーを利用したい，辞典を調べたい，表を作って計算したいなど，各機能を使いこなしたいときに。
<head>通信ガイド</head>
電子メールをやり取りしたり，パソコン通信やファクス通信などを行いたいときに。
ラインスキャナー拡張セット <symbol>*</symbol>取扱説明書
本体のスキャナーを取りはずして使用したいときに。
<symbol>* CSD500 : 別売 品番 FW-RCU1</symbol>
<head>まるごと CD（専用の CD と呼びます）</head>
（以下省略）
```

図 1 タグ付けされたテキスト

表題，図，表や文章中で使われているシンボル（機能アイコンやボタン，キートップなど）を特定するタグ（head, fig, table, symbol）を設けた。

3.3 テキスト解析

テキスト化したデータに対して，日本語形態素解析システム・茶筌(chasen)を使って形態素解析を実施した。形態素解析の結果例を図 2 に示す。

最初	サイショ	最初	名詞-一般
に	ニ	に	助詞-格助詞-一般
お	オ	お	接頭詞-名詞接続
読み	ヨミ	読む	動詞-自立 五段・マ行 連用形
ください	クダサイ	くださる	動詞-非自立 五段・ラ行特殊 命令
EOS			
パーソナル	パーソナル	パーソナル	名詞-一般
ワープロ	ワープロ	ワープロ	名詞-一般
EOS			
取扱	トリアツカイ	取扱	名詞-一般
説明	セツメイ	説明	名詞-サ変接続
書	ショ	書	名詞-接尾-一般
EOS			
品番	ヒンバン	品番	名詞-一般
FW	エフダブリュー	FW	名詞-一般
—	—	—	記号-一般
U	ユー	U	記号-アルファベット
1	イチ	1	名詞-数
CSD			未知語
500	ゴゼロゼロ	500	名詞-数
EOS			
FW	エフダブリュー	FW	名詞-一般
—	—	—	記号-一般
U	ユー	U	記号-アルファベット
1	イチ	1	名詞-数
CSD			未知語
600	ロクゼロゼロ	600	名詞-数
EOS			

図 2 テキストの形態素結果

4. 分析

4.1 語彙分析（自立語，付属語，分野別）

表 2 には，カテゴリ毎の総語数，異なり語数及び総語数比を示す。

表 2..語彙分析結果

カテゴリ	総語数	異なり語数	異なり語数比 (%)	総語数比	累積総語数比
パソコン	907192	3668	0.40	36.49	36.49
ワープロ	265668	4548	1.71	10.69	47.17
カーナビ・カーAV	185382	6856	3.70	7.46	54.63
調理	167357	4842	2.89	6.73	61.36
カメラ	160558	4519	2.81	6.46	67.82
テレビ	119509	3011	2.52	4.81	72.62
電話機・FAX・テレビドアホン	114072	2670	2.34	4.59	77.21
通信・ネットワーク	113195	2845	2.51	4.55	81.77
BD・DVD・ビデオ	111606	3197	2.86	4.49	86.25
オーディオ	102085	4335	4.25	4.11	90.36
水回り	50492	2423	4.80	2.03	92.39
空調・季節	47103	2410	5.12	1.89	94.29
健康	39056	2377	6.09	1.57	95.86
家事	32287	2351	7.28	1.30	97.16
オフィス	31573	1827	5.79	1.27	98.43
美容	25745	1880	7.30	1.04	99.46
保冷・保温	13410	1532	11.42	0.54	100.00
合計	2486290	55291		100.00	

図 3 は、カテゴリ別の語数分布を示す。図 4 は、異なり語数比をカテゴリ毎に示した。パソコンは、同種の文章が多いことから異なり語数比が極端に少ない。一方、オフィス製品や保冷・保温カテゴリのマニュアル文章は、テキスト化した文章量が少ないことから語数比が高くなっている。一般的には、異なり語数比は、4%~6%であると考えられる。

表 3 は、語種別の語数を示す。付属語は、助詞や助動詞である。自立語は、動詞や名詞、形容詞、副詞、形容動詞などである。マニュアルには、番号を含む製品記号や専門用語が多いことから、未知語や記号、活用語尾などを区分した。

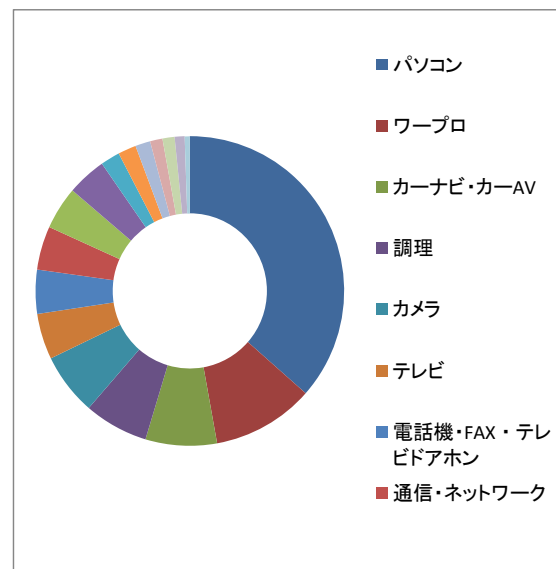


図 3 カテゴリ別語数

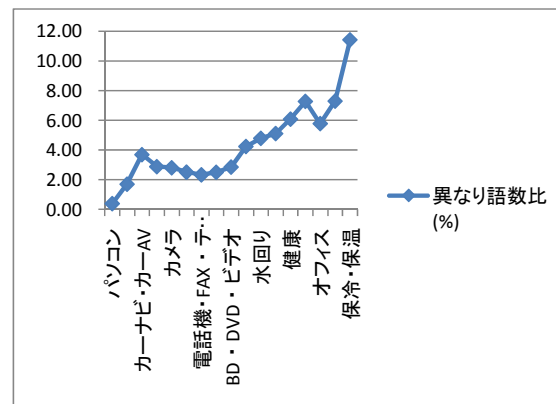


図 4 異なり語数比

表 3 語種別の語数

	付属語	自立語	活用語尾, 記号他	合計
総語数	402226	1139588	944476	2486290
異なり語数	101	14917	4086	19104

4.2 使用語彙

表 4 は、語彙の出現頻度順位である。ただし、カテゴリ別に語彙数の偏りがあるので（パソコン関連製品のマニュアルが多く、美容や保冷・保温などの製品マニュアルが少ない）、出現数はその影響を受ける。

カテゴリ毎に語彙順位をみると、付属語には違いがあまりない。自立語に違いがあり、それぞれのカテゴリで製品特徴に関係する語彙が顕れる。例えば、調理カテゴリでは「加熱(する)」「焼く」「使う」等、空調・季節カテゴリでは、「運転(する)」「ページ」「フィルター」等である。

表 4 語彙の頻度順位

	付属語	自立語	活用語尾, 記号他
1	を(助詞)	する(動詞)	。(記号)
2	の(助詞)	くださる(動詞)	, (記号)
3	に(助詞)	設定(名詞)	る(活用助辞)
4	が(助詞)	表示(名詞)	ず(活用助辞)
5	は(助詞)	いる(動詞)	ます(接辞)
6	で(助詞)	場合(名詞)	い(活用助辞)
7	と(助詞)	できる(動詞)) (記号)
8	や(助詞)	クリック(名詞)	((記号)
9	など(助詞)	画面(名詞)	て(活用助辞)
10	から(助詞)	ページ(名詞)	。(記号)
11	も(助詞)	こと(名詞)	[(記号)
12	か(助詞)	ある(動詞)] (記号)
13	について(助詞)	押す(動詞)	… (記号)
14	へ(助詞)	とき(名詞)	● (記号)
15	まで(助詞)	なる(動詞)	れる(接辞)
16	によって(助詞)	1 (名詞)	「 (記号)
17	のみ(助詞)	使用(名詞)	」 (記号)

18	ので(助詞)	電源(名詞)	り(活用助辞)
19	として(助詞)	接続(名詞)	ない(接辞)
20	ながら(助詞)	2 (名詞)	/ (記号)
21	だけ(助詞)	選ぶ(動詞)	— (記号)
22	による(助詞)	機(名詞)	た(活用助辞)
23	に従って(助詞)	使う(動詞)	く(活用助辞)
24	により(助詞)	入力(名詞)	φ(活用助辞)
25	という(助詞)	時(名詞)	: (記号)
26	より(助詞)	機能(名詞)	だ(活用助辞)
27	でも(助詞)	再生(名詞)	せん(活用助辞)
28	に関する(助詞)	戻る(動詞)	って(活用助辞)
29	ナ(助詞)	状態(名詞)	この(連体詞)
30	な(助詞)	メニュー(名詞)	お(接頭詞)

大小カテゴリごとの語彙の重複の程度を調査した。出現頻度順位 100 までの語彙について重複率(%)を求めた。自立語では、BD・DVD・ビデオ、オーディオ、カメラ、通信・ネットワークと電話機・FAX のカテゴリが語彙的に重複している(40%~50%)。他のカテゴリ間では、語彙の重複率は高くない(30%以下)。

付属語では、各カテゴリとも 60%~70%程度の語彙の重複が見られる。ただし、テレビ、パソコンカテゴリと保冷・保温カテゴリは 50%である。調理カテゴリも他のカテゴリとの重複率が低い(50%~60%)。

4.3 教科書データとの比較

4.3.1 日本語の述語と形態素

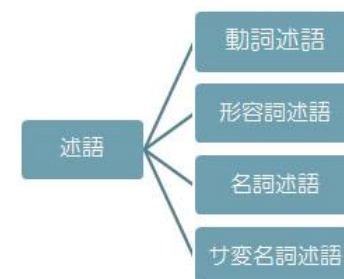


図 5 述語の分類

文末述語表現を図 5 のように区分し、製品取扱説明書の文末述部表現を分析した結果を示す。なお、例えば、動詞述語は「してください、あります、されます」など、形容詞述語は「よいだろう、多いだろう、いいだろう」など、名詞述語は「ことです、ものです、必要です」など、サ変名詞述語は「成立した、発展した、発達した」などである。また、述語を構成する形態素数 N によって、N 語述語という区分を設けた。同時に、頻度の高い語彙については、教科書データ[2]と比較して示す。

4.3.2 動詞述語

図 6 と図 7 には、動詞述語の出現数と異なり数を示す。表 5 には、1 語述語の出現数を教科書データと比較した結果を示す。

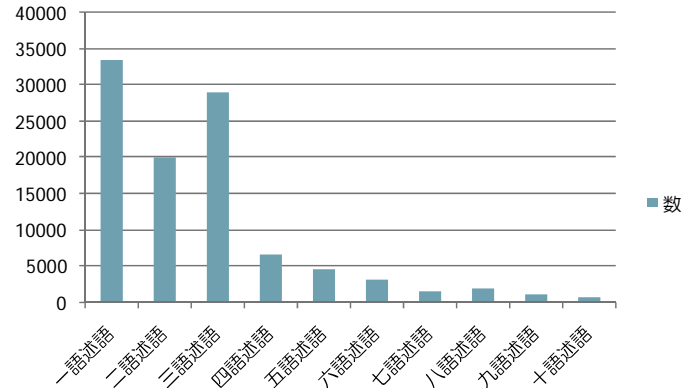


図 6 動詞述語の出現数

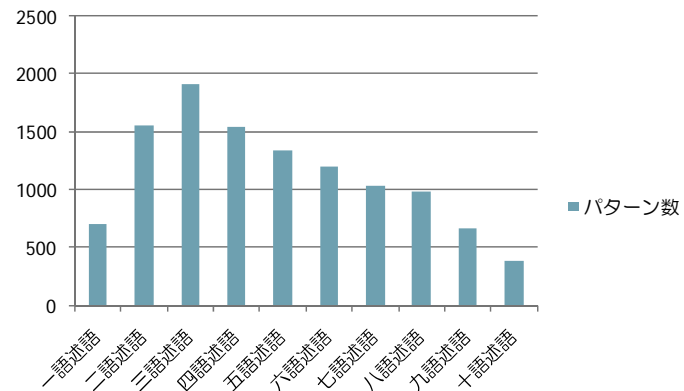


図 7 動詞述語の異なり数

表 5 1 語述語の種類と出現数

	マニュアル	語数	教科書	語数
1	する	13133	した.	29694
2	戻る	3514	する.	14042
3	押す	3320	なった.	12877
4	選ぶ	2201	ある.	9859
5	ください	1206	いう.	7350
6	入れる	619	なる.	6938
7	見る	602	できる.	5291
8	使う	484	あった.	1971
9	切る	231	おこった.	946
10	抜く	230	わかる.	902

4.3.3 名詞述語

図 8 と図 9 には、名詞述語の出現数と異なり数を示す。表 6 には、2 語述語の出現数を教科書データと比較した結果を示す。

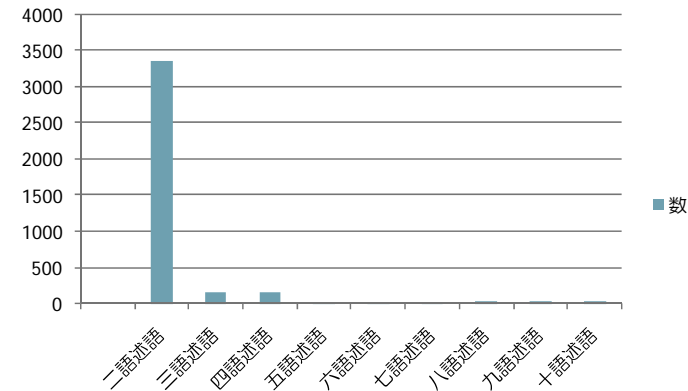


図 8 名詞述語の出現数

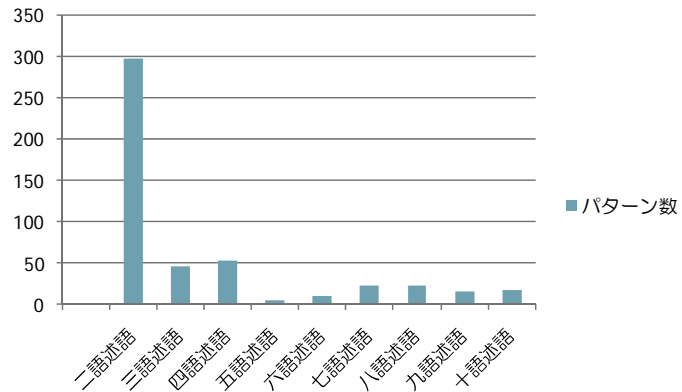


図 9 名詞述語の異なり数

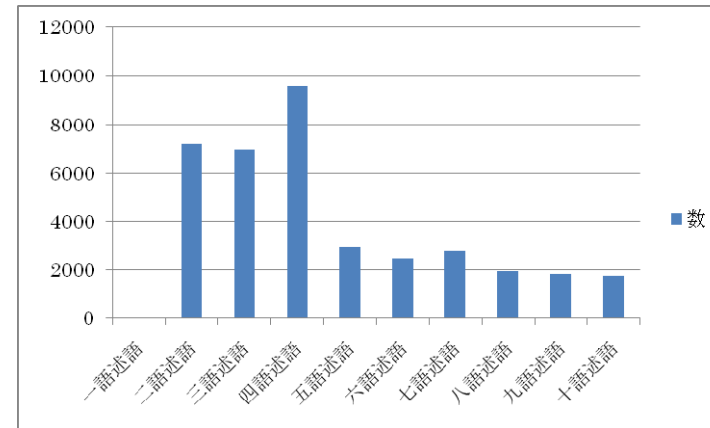


図 10 変名詞述語の出現数

表 6 2語述語の種類と出現数

	マニュアル		教科書	
1	必要です	380	の です.	2035
2	商標です	258	こと です.	303
3	完了です	235	もの です.	252
4	内容です	178	必要 です.	177
5	便利です	129	わけ です.	139
6	時間です	108	ん です.	139
7	ことです	94	大切 です.	101
8	機能です	92	ちがいない.	101
9	費用です	81	の でしょう.	91
10	ものです	75	そうです.	87

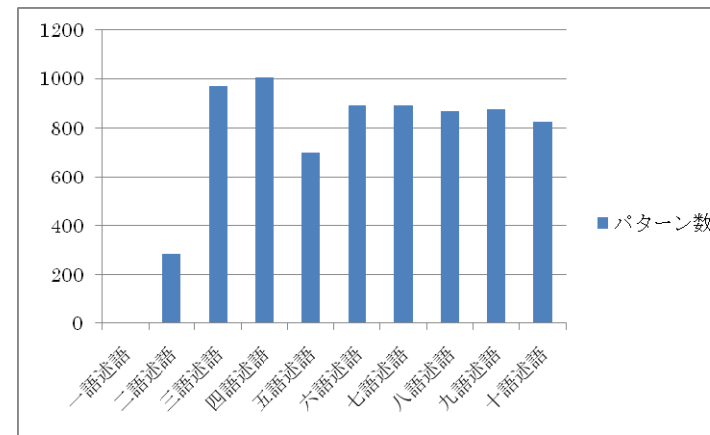


図 11 変名詞述語の異なり数

マニュアル文書に形容詞述語はなかった。

4.3.4 変名詞述語

図 10 と図 11 には、名詞述語の出現数と異なり数を示す。表 7 には、2語述語の出現数を教科書データと比較した結果を示す。

表 7 2 語述語の種類と出現数

	マニュアル		教科書	
1	設定する	779	成立した。	1105
2	確認する	455	発展した。	535
3	接続する	366	発達した。	439
4	変更する	312	主張した。	404
5	表示する	273	拡大した。	363
6	入力する	273	確立した。	359
7	調整する	212	実現した。	333
8	選択する	203	展開した。	329
9	決定する	193	観察する。	273
10	起動する	192	開始した。	268

4.4 結果

取扱説明書文章の文末表現は複雑ではなかった。形態特徴の比較では、教科書テキストとあまり差がない。文章特徴を挙げる。

- 文末に助動詞表現が出ない（表現が簡素）
 - 分野（製品）依存の語彙頻度の高さ（一般的傾向）
 - 操作手順に頻出する動作表現（「押す」「表示される」など）が多い（内容依存）
- 考えられる原因や要因としては、(1) 手続きや手順を示す表現に話し手の意図はないこと、(2) 事実説明が主な教科書テキストも同じ傾向を示すことが推察できる。

5. おわりに

本稿では、構築した製品取扱説明書コーパスについて概説した。さらに構築したコーパスを使って製品取扱説明書の文末述部表現を分析した結果について報告した。

コンピュータは、電子技術を中心とした理屈によって作られている電子機器・装置の一つで、数学や物理など合理的な思考を基盤とした電子技術やソフトウェア技術が理屈を支えている。そのため適切な指示を与えれば、複雑な機能を実現するのだが、その「適切な指示」の理解、選択や実施の仕方の学習に要する負担コストが大きくなっている。

製品取扱説明書は、多機能説明のために紙面が増えることのコスト増の過大を抱えている。文章説明を減らし視覚情報を増やして、メディア複合による紙面の簡便化を図る取り組みや、近年ではウェブ検索を前提に説明書を提供するサービスが活発であ

る[3]。

いずれにしても機能解説のための説明の工夫が増えていることには間違いはなく、如何に説明書を読んでもらうのかがポイントになっている。その下位の難事に、説明のための文章の書き方—日本語表現の工夫がある。「製品が実現する機能を叙述する」分かりやすい日本語表現とは何か議論されてきた。正しい日本語を特定する試みや、説明書表現が表す意味を分析することで、製品取扱説明書が、何をどのように伝えようとしているのかを探る試みもある[4,5]。今後も、製品説明書コーパスを利用することでマニュアル文章の特徴調査を複文に拡げてゆく予定である。

参考文献

- 1) 佐野洋, 桑野裕康: 日本語書き言葉の公共性について—取扱説明書による事例報告—, テクニカルコミュニケーションシンポジウム 2010 論文集, pp23-pp27, テクニカルコミュニケーションシンポジウム 2010, 2010.
- 2) 佐野洋, 于壮飛, 芝野耕司: 日本語教科書コーパスの構築, 2009 日本語教育国際研究大会(ポスター発表), ニューサウスウェールズ大学, 2009.
- 3) 久保達明, 細川尚嗣: 取説の近未来を考える, テクニカルコミュニケーションシンポジウム 2009 論文集, pp15-pp18, テクニカルコミュニケーションシンポジウム 2009, 2009.
- 4) 恒田杏子, 三波千穂美, 石井啓豊: マニュアルに表現された知識の構成, テクニカルコミュニケーションシンポジウム 2005 論文集, pp41-pp45, テクニカルコミュニケーションシンポジウム 2005, 2005.
- 5) 森口稔: 正しさは理解に影響するのか?—実験的アプローチ—, テクニカルコミュニケーションシンポジウム 2005 論文集, pp5-pp9, テクニカルコミュニケーションシンポジウム 2005, 2005.