第 5 章

システム辞書仕様

本章では，EDR から提供されるテキスト形式の辞書 (以下，テキスト辞書と呼ぶ) と，辞書トランスレータにより作成される EDBroW 用の辞書 (以下，システム辞書と呼ぶ) について，その対応関係を示す．

5.1 各辞書に共通な事柄

概念識別子

テキスト辞書では，概念識別子はそのほとんどが16進数6桁の文字列で記述されているが，一部の概念識別子 (数字を表す概念) では，概念識別子が16進数で6桁より大きくなることがある．

また，システム辞書では，概念識別子は unsigned long で格納している．このため，テキスト辞書中で16進数6桁を越える概念識別子については，別途用意するテーブルにより変換した後に格納する．

変換の対象となっているのは，以下の 6 つの概念識別子である．

テキスト辞書中の 概念識別子	システム辞書中の 概念識別子
000103e8	a00001
00012710	a00002
00010f4240	a00003
000105f5e100	a00004
0001e8d4a51000	a00005
00012386f26fc10000	a00006

なお，上記以外で，テキスト辞書中で6桁を越える概念識別子を含むレコードは，エラーとしてシステム辞書には格納されない．

文法コード

品詞コード，左右連接属性コードについては，テキスト辞書中のコード文字列を，本章末に示すテーブルにより，数値としてシステム辞書に格納する．

また，日本語単語辞書，日英対訳辞書において，"JN1;JVE" という品詞は，サ変名詞を表わすものであり，システム辞書では，"JSA" (品詞番号 38) として格納する．

品詞，連接以外の文法コードについては，文字列をそのまま格納する．

以下に，システム辞書に格納されるデータの名称と，対応するテキスト辞書の名称との対応を示す．

5.2 単語辞書

システム辞書名称	テキスト辞書名称	備考
単語見出し *1	不変化部分	
読み *2	カナ表記 (日本語) 音節区切り (英語:慣用句以外) 見出し表記 (英語:慣用句)	
発音情報 *3	発音情報 (慣用句以外)	
構成語情報 *3	単語見出し (慣用句のみ)	
概念識別子	概念識別子	
左連接情報	左連接情報	慣用句の場合には
右連接情報	右連接情報	格納しない
品詞	品詞	
言語識別 *4	–	
レコード種別 *5	–	
慣用句フラグ *6	–	
文法情報	品詞 / 構文木以外の 文法情報 + 用法	各文法情報コードを ';' で連結して格納
構文木	構文木	慣用句以外は空

*1 慣用句の場合には，先頭の 1 単語のみを格納する．

*2 英語単語の慣用句の場合，テキスト辞書では音節区切りのフィールドが空であるため，かわりに見出し表記を格納する．

*3 発音情報 / 構成語情報 は，慣用句とそれ以外のレコードでフィールドが重複しないため，システム辞書中では同フィールドとして格納する．

*4 レコードIDの別により判定．"JWD" を含む場合には 0 を，"EWD" を含む場合には 1 を格納

*5 レコードIDの別により判定．先頭が 'T' で始まる場合には 1 を，それ以外は 0 を格納

*6 主に管理のための情報．レコードが構文木を持つ場合に 1 を，そうでない場合に 0 を格納

※ただし，テキスト辞書の頻度情報は，システム辞書には格納しない

5.3 対訳辞書

システム辞書名称	テキスト辞書名称	備考
単語見出し	単語見出し	
読み	カナ表記 (日英)	英日の場合は空
品詞	品詞	
概念識別子	概念識別子	
言語対識別 *1	–	
レコード種別 *2	–	
対訳情報	対訳情報	

*1 レコードIDの別により判定．"JEB" を含む場合には 0 を，"EJB" を含む場合には 1 を格納

*2 レコードIDの別により判定．先頭が 'T' で始まる場合には 1 を，それ以外は 0 を格納

5.4 共起辞書

システム辞書名称	テキスト辞書名称	備考
受け側単語見出し	受け側単語表記	
受け側表記情報	表記情報 *1	カナ表記もしくは原形
受け側品詞	品詞 *1	
受け側概念識別子	概念識別子 *2	
受け側慣用句フラグ	慣用句フラグ *1	
共起関係子	共起関係子	
関係単語表記	関係単語表記	
係り側単語見出し	係り側単語表記	
係り側表記情報	表記情報 *1	カナ表記もしくは原形
係り側品詞	品詞 *1	
係り側概念識別子	概念識別子 *2	
係り側慣用句フラグ	慣用句フラグ *1	
概念関係子	概念関係子 *3	
関係子方向 *4	–	
表記順 *4	–	
言語識別 *5	–	
表層共起頻度	表層共起頻度	
共起項目頻度	共起項目頻度	
受け側共起要素頻度	受け側共起要素頻度	
係り側共起要素頻度	係り側共起要素頻度	
例文	例文	情報を省略して格納 *6

*1 受け側 / 係り側の各情報は，共起句要素情報の中から，受け側 / 係り側に対応するもののみを格納

*2 概念識別子が「補足付き概念説明」である場合には，他の概念識別子と重ならない識別子を新設して格納することが可能．ただし，新設した概念識別子については，「補足付き概念説明」文字列を「日本語概念説明」のフィールドに格納した形式の概念見出し補助ファイルを作成し，他の概念見出し辞書ファイルとマージする (辞書トランスレータにより，格納方法を選択可能)

*3 概念関係子が "" である場合には，関係子 "nil"(関係子番号 80) を格納

*4 要素番号をすべて格納しないかわりに，意味情報における始点 / 終点側と，受け側 / 係り側との対応をとるために関係子方向を利用．始点 = 受け側の場合に 0 を，始点 = 係り側の場合に 1 を格納．また，句見出しを格納しないかわりに，句見出しにおける受け側 / 係り側単語の出現順を表記順とし

て格納する．受け側が先行する場合に 1 を，係り側が先行する場合に 0 を格納．

*5 レコードIDの別により判定．"JCC" を含む場合には 0 を，"ECC" を含む場合には 1 を格納

*6 例文は，1 例文につき文IDが複数個ある場合には，先頭の文IDのみを格納し，最大 5 例文を格納する．

※句見出しは格納しないが，受け側 / 係り側単語見出し，関係単語表記，表記順により，再構成が可能である．

5.5 概念見出し辞書

システム辞書名称	テキスト辞書名称	備考
概念識別子	概念識別子	
日本語概念見出し	日本語概念見出し	
英語概念見出し	英語概念見出し	
日本語概念説明	日本語概念説明	
英語概念説明	英語概念説明	

5.6 概念体系辞書

システム辞書名称	テキスト辞書名称	備考
上位概念	上位概念	
下位概念	下位概念	

5.7 概念記述辞書

システム辞書名称	テキスト辞書名称	備考
概念識別子 1	概念識別子 1	
概念識別子 2	概念識別子 2	
関係子	関係子	
真偽値	真偽値	
記述区分 *1	記述区分	

*1 J記述を 1，E記述を 0 として格納

5.8 コード一覧

次ページ以降に，左右連接情報，品詞，概念関係子のコード番号と，対応するコードを示す．

5.8.1 日本語左連接情報コード

番号	コード
0	–
1	JLN1
2	JLN2
3	JLN3
4	JLN4
5	JLN5
6	JLN6
7	JLN7
8	JLN8
9	JLV1
10	JLA1
11	JLM1
12	JLD2
13	JLR1
14	JLC1
15	JLC3
16	JLT1
17	JSF1
18	JLJ6
19	JSF4
20	JSF5
21	JSF7
22	JSF8
23	JSF9
24	JSFA
25	JLS1
26	JLS2
27	JLJ1
28	JLJ2
29	JLJ3
30	JLJ4
31	JLJ5
32	JLJ7
33	JLJ8
34	JLJ9
35	JLJA
36	JLJB
37	JLWA
38	JLJD
39	JLJE
40	JLJF
41	JLJG
42	JLJH
43	JLJI
44	JLJJ
45	JLJK
46	JLJL
47	JLJM
48	JLJN
49	JSD1
50	JSD2
51	JSD4
52	JLW1
53	JLW2
54	JLW3
55	JLW4
56	JLW5
57	JLW6
58	JLW7
59	JLW8
60	JLW9
61	JLWA
62	JLWB
63	JLWC
64	JLWD
65	JLWE
66	JLWF
67	JLWG
68	JLWH
69	JLWI
70	JLWJ
71	JLWK
72	JSVK
73	JSVG
74	JSVS
75	JSVT
76	JSVN
77	JSVB
78	JSVM
79	JSVR
80	JSVW
81	JSV1
82	JSV2
83	JSV3
84	JSV4
85	JSV5
86	JSV6
87	JSV7
88	JSA0
89	JSA1
90	JSA2
91	JSA3
92	JSM0
93	JSM1
94	JSM2
95	JSM3
96	JSM4
97	JSMA
98	JSM5
99	JSM7
100	JSM8

5.8.2 日本語右連接情報コード

番号	コード
0	–
1	JRN1
2	JRN2
3	JRN3
4	JRN4
5	JRN5
6	JRN6
7	JRN7
8	JRN8
9	JRN9
10	JRNA
11	JRNB
12	JRND
13	JRNE
14	JRNF
15	JRVK
16	JRVG
17	JRVS
18	JRVT
19	JRVN
20	JRVB
21	JRVM
22	JRVR
23	JRVW
24	JRV1
25	JRV2
26	JRV3
27	JRV4
28	JRV5
29	JRV6
30	JRV7
31	JRA1
32	JRA2
33	JRA3
34	JRM1
35	JRM2
36	JRM3
37	JRM4
38	JRMA
39	JRM5
40	JRM9
41	JRM7
42	JRM8
43	JRD2
44	JRD3
45	JRD4
46	JRD5
47	JRD6

48	JRD7
49	JRD8
50	JRD9
51	JRDA
52	JRDB
53	JRDC
54	JRDD
55	JRDE
56	JRDF
57	JRDG
58	JRDH
59	JRR1
60	JRC1
61	JRC2
62	JRT1
63	JRP1
64	JRP2
65	JRP3
66	JRP4
67	JRP5
68	JRP6
69	JRP7
70	JRS1
71	JRS2
72	JRJ1
73	JRJ2
74	JRJ3
75	JRJ4
76	JRJ5
77	JRJ6
78	JRJ7
79	JRJ8
80	JRJ9
81	JRJA
82	JEM1
83	JEM2
84	JEM3
85	JEM4
86	JEM5
87	JEM6
88	JEM7
89	JEM8
90	JED1
91	JED2
92	JED3
93	JED4
94	JED5
95	JED6
96	JED7
97	JED8
98	JED9
99	JEDA
100	JEE1
101	JEE2
102	JEE3
103	JEE4
104	JEE5
105	JEE6
106	JEA1
107	JEA2
108	JEC1
109	JEC2
110	JEI1
111	JEI2

5.8.3 日本語品詞コード

番号	コード
0	–
1	JN1
2	JN2
3	JN3
4	JN4
5	JN7
6	JVE
7	JAJ
8	JAM
9	JD1
10	JD2
11	JNM
12	JC1
13	JC3
14	JT1
15	JT2
16	JT3
17	JT4
18	JN5
19	JB1
20	JUN
21	JN6
22	JEV
23	JEA
24	JEM
25	JNP
26	JPR
27	JAP
28	JMP
29	JIP
30	JSE
31	JJO
32	JJ1
33	JJD
34	JJP
35	JAX
36	JIT
37	JSY
38	JSA

5.8.4　英語左連接情報コード

番号	コード
0	–
1	ELB1
2	ELN1
3	ELN2
4	ELP1
5	ELP2
6	ELA1
7	ELA3
8	ELD1
9	ELD3
10	ELV1
11	ELV2
12	ELT1
13	ELT2
14	ELW1
15	ESN1
16	ESN2
17	ESN3
18	ESN4
19	ESN5
20	ESN6
21	ESN7
22	ESN8
23	ESF1
24	ESF2
25	ESF3
26	ESF8
27	ESF4
28	ESF5
29	ESF6
30	ESF7
31	ESV1
32	ESV2
33	ESV3
34	ESV4
35	ESV5
36	ESV6
37	ESV7
38	ESV8
39	ESV9
40	ESVA
41	ESVB
42	ESPB
43	ESPD
44	ESPF
45	ESPG
46	ESPK
47	ESPL
48	ESPM
49	ESPN
50	ESPP
51	ESPR
52	ESPT
53	ESPV
54	ESA1
55	ESA2
56	ESA3
57	ESAB
58	ESAD
59	ESAF
60	ESAG
61	ESAK
62	ESAL
63	ESAM
64	ESAN
65	ESAP
66	ESAR
67	ESAS
68	ESAT
69	ESAV
70	ESAZ
71	ESD1
72	ESD2
73	ESD3
74	ESDB
75	ESDD
76	ESDF
77	ESDG
78	ESDK
79	ESDL
80	ESDM
81	ESDN
82	ESDP
83	ESDR
84	ESDS
85	ESDT
86	ESDV
87	ESDZ
88	ELS1
89	ELS2
90	ELS3
91	ELS4
92	ELS5
93	ELS6
94	ELS7
95	ELS8
96	ELS9
97	ELSA
98	ELSB
99	ELSC

5.8.5　英語右連接情報コード

番号	コード
0	–
1	ERB1
2	ERR1
3	ERR2
4	ERR3
5	ECF1
6	ECN1
7	ECN2
8	ECN3
9	ECN4
10	ECN5
11	ECN6
12	ECN7
13	ECN8
14	ERN1
15	ERN2
16	ERN3
17	ERN4
18	ERN5
19	ERN6
20	ERN7
21	ERP1
22	ERP2
23	ERP3
24	ERP4
25	ERP5
26	ERP7
27	ERP8
28	ERP9
29	ERPA
30	ERPB
31	ERPC
32	ERPD
33	ERPE
34	ERL1
35	ERL2
36	ERL3
37	ERL4
38	ERL5
39	ERQ1
40	ERQ2
41	ERQ3
42	ERQ4
43	ERQ6
44	ERM1
45	ERM2
46	ERE1
47	ERE2

48	ERE3
49	ERE4
50	ERE5
51	ERE6
52	ERE7
53	ERE8
54	ERO1
55	ERO2
56	ERO3
57	ERH1
58	ERH2
59	ERH3
60	ERH4
61	ERX1
62	ERC1
63	ERC2
64	ECV1
65	ECV2
66	ECV3
67	ECV4
68	ECV5
69	ECV6
70	ECV7
71	ECV8
72	ECV9
73	ECVA
74	ECVB
75	EPVB
76	EPVD
77	EPVF
78	EPVG
79	EPVK
80	EPVL
81	EPVM
82	EPVN
83	EPVP
84	EPVR
85	EPVT
86	EPVV
87	ERV1
88	ERV2
89	ERV3
90	ERV4
91	ERV5
92	ECA1
93	ECA2
94	ECA3
95	ECAB
96	ECAD
97	ECAF
98	ECAG
99	ECAK
100	ECAL
101	ECAM
102	ECAN
103	ECAP
104	ECAR
105	ECAS
106	ECAT
107	ECAV
108	ECAZ
109	ERA1
110	ERA2
111	ERA3
112	ECD1
113	ECD2
114	ECD3
115	ECDB
116	ECDD
117	ECDF
118	ECDG
119	ECDK
120	ECDL
121	ECDM
122	ECDN
123	ECDP
124	ECDR
125	ECDS
126	ECDT
127	ECDV
128	ECDZ
129	ERD1
130	ERD2
131	ERD3
132	ERD5
133	ERIT
134	ERPP
135	ERS1
136	ERS2
137	ERS3
138	ERS4
139	ERS5
140	ERS6
141	ERS7
142	ERS8
143	ERS9
144	ERSA
145	ERSB
146	ERPF
147	ERTO
148	ERSC

5.8.6 英語品詞コード

番号	コード
0	–
1	EN1
2	EN2
3	EN3
4	EN4
5	EN5
6	EP1
7	EP2
8	EP3
9	EP4
10	EP5
11	EVE
12	EBE
13	EAJ
14	ED1
15	ED2
16	ED3
17	ED4
18	ED5
19	EPR
20	EPR2
21	ET1
22	ET2
23	EAR
24	EAV
25	EAV2
26	EIT
27	EC1
28	EC12
29	EC2
30	EC22
31	EPF
32	EEN
33	EEV
34	EEA
35	EED
36	ENP
37	EVP
38	EAP
39	EDP
40	EPP
41	EIP
42	ESE
43	EUN
44	ESY

5.8.7 共起辞書品詞コード

番号	文字列
0	–
1	名詞
2	動詞
3	形容詞
4	形容動詞
5	副詞
6	接尾語
7	NOUN
8	PRON
9	DEMO
10	INDEF
11	VERB
12	ADJ
13	ADV
14	UNIT
15	PTCL

※ 1 ~ 6 までは日本語共起辞書用，8 以降は英語共起辞書用

5.8.8 概念関係子コード

番号	文字列
0	–
1	agent
2	object
3	manner
4	implement
5	material
6	time
7	time-from
8	time-to
9	duration
10	location
11	place
12	source
13	goal
14	scene
15	beneficiary
16	a-object
21	condition
22	cooccurrence
23	sequence
31	part-of
32	element-of
42	unit
43	quantity
44	degree
45	number
46	modifier
51	possessor
52	purpose
53	cause
54	basis
55	and
56	or
57	from-to
60	kind-of
61	equivalent
62	similar
71	statement
80	nil
101	all

※ nil は共起辞書用