

V

産業日本語研究会ワークショップ講演

平成 29 年 12 月 19 日に開催した産業日本語研究会ワークショップでは、文章の品質や作成能力という点で産業日本語活動と共通する取り組みをしている外部講演者より、3 件の講演をいただいた。以下に当日の講演内容を掲載する。

1. 「いかに情報をシェアするか ～オントロジー構築と共通語彙構築の経験から～」

いかに情報をシェアするか ～オントロジー構築と共通語彙構築の経験から～

武田 英明
国立情報学研究所
takeda@nii.ac.jp

社会に役立つオントロジー・語彙

- 社会に役立つオントロジー・語彙を提供するにはどうしたら良いか。
 - オントロジー・語彙の**構築**の仕方
 - 農作業オントロジー構築の経験から
 - オントロジー・語彙を**維持**の仕方
 - 共通語彙基盤での経験から

農作業オントロジー構築の経験から

農作業名称の標準化

■ 現状

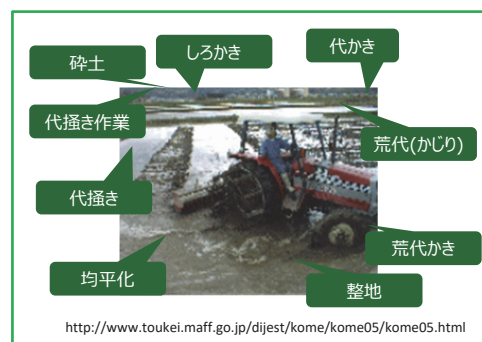
農業ITシステムの普及により農業現場では管理や記録など作業を効率的に行うことが可能となった。そして、膨大なデータが発生しており、そのデータは統計・分析により作業の最適化、予測に重要な資料として期待されている。

■ 問題点

基準となる情報が標準化されていないため、ベンダーは独自のデータ項目を定義して対応しており、これがデータの連携、統合に妨げとなる。

■ 本研究の目的

農作業名称における既存語彙体系について考察し、その考察に基づいて農作業名称の標準語彙となる農作業基本オントロジーを構築する。



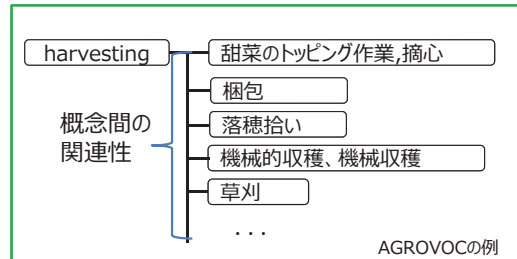
農作業オントロジーの構築 – シソーラスとオントロジー

■ シソーラス (Thesaurus)

単語を同義語や意味上の類似関係、包含関係などによって分類する体系。(AGROVOC)

効率的な関連語検索が可能。

上位/下位階層の関係性が明確に定義されていないため同じ階層の概念間,すなわち兄弟関係の概念の関連性も定義されていない。

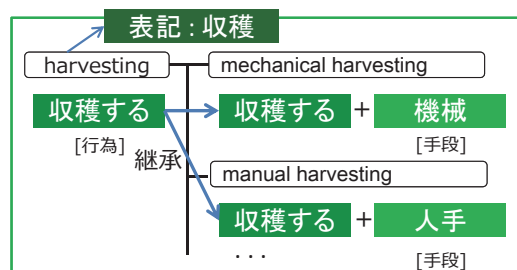


■ オントロジー (Ontology)

対象の概念と概念間の相互関係を定義する体系。

- 上位クラスを持つ属性を下位が継承
- 概念と表記を分離

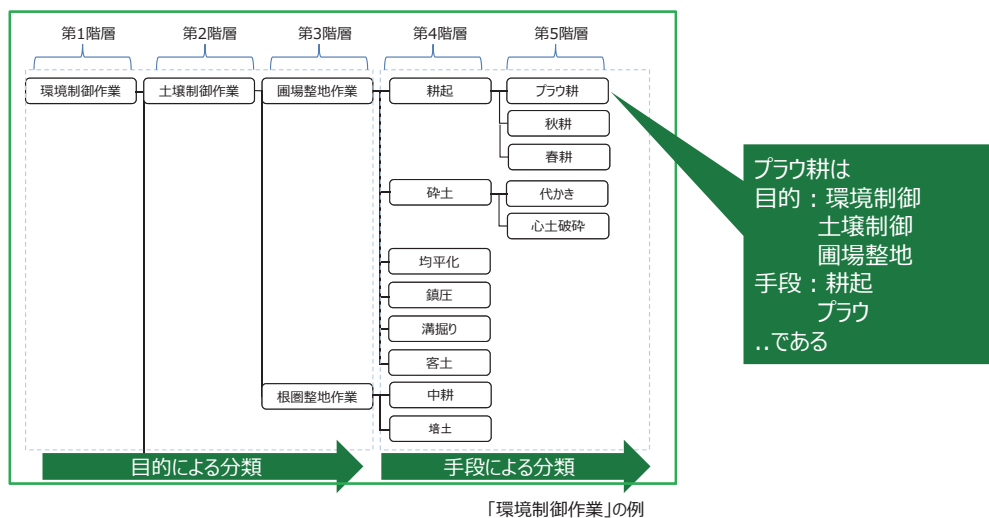
➡ 農業分野におけるドメインオントロジーを構築する



農作業オントロジーの構築 – 農作業の基本概念と構造化(1/2)

■ 農作業の構造概念

「農作業」を「作物生産作業」、「環境制御作業」、「管理作業」、「間接作業」に分類し上位概念とする。そして、属性として「目的」と「手段」を用意し2層構造とし、分類基準とする。

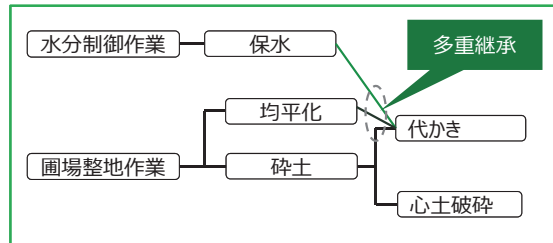


農作業オントロジーの構築 – 農作業の基本概念と構造化(2/2)

■ 概念間の多重継承

農作業の中では複数の目的で行われる作業が存在しており、それを概念の定義に適用させるために多重継承を考慮する必要がある。

例)「代かき」目的「砕土」
「保水」
「均平化」



■ 概念と表記

同じ作業に対して複数の表記が存在する場合に対応する必要がある。

- ひらがなと漢字による表記 例)「播種」と「は種」、「代かき」と「代掻き」など
- 英語と漢字による表記 例)「ハードニング」と「硬化」など

農作業オントロジーの構築 – 農作業基本オントロジー(1/2)

■ AAO : Agricultural Activity Ontology

農作業における基本概念185語を網羅したオントロジー

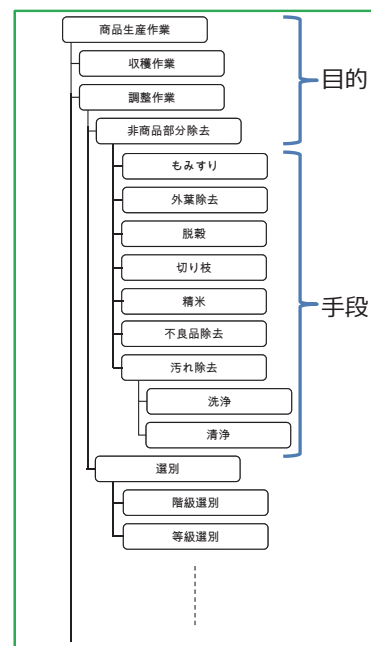
■ 特徴

「目的」と「手段」による2層構造
概念の多重継承と複数の表記に対応

■ 公開

“CAVOC : Common Agricultural VOCabulary”
<http://www.cavoc.org> でAAO ver.0.94を公開中

- 検索
- 名前空間

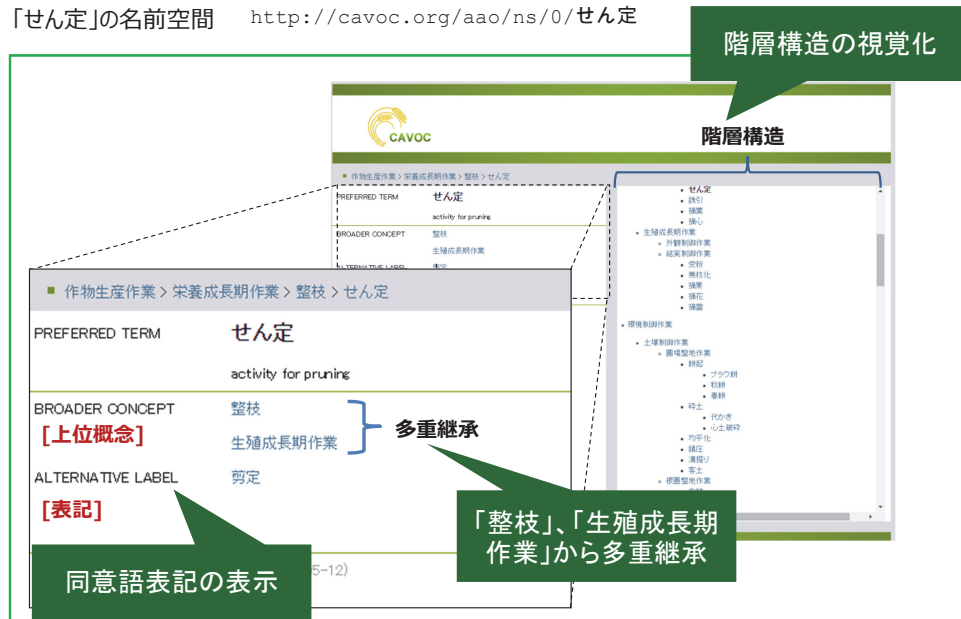


「商品生産作業」の例

農作業オントロジーの構築 – 農作業基本オントロジー(2/2)

■ www.cavoc.org

「せん定」の名前空間 <http://cavoc.org/aao/ns/0/せん定>

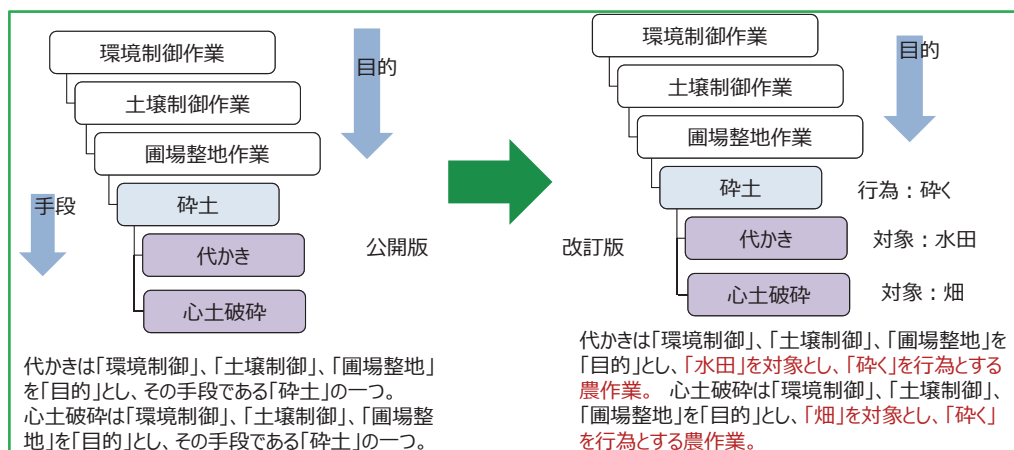


今後の課題 (1/5)

■ 属性の具体化

公開版(ver 0.94)では「目的」と「手段」による2層構造となっており、農作業の概念を定義するために属性の数が十分ではない。農作業名称の多様性と規模に対応するために属性を拡張する必要がある。

改訂版では属性として「作物」、「目的」、「行為」、「対象」、「手段(=道具)」、「時期」を用い、より明確な定義を行い、分類を行う。



農作業オントロジーの設計(3)

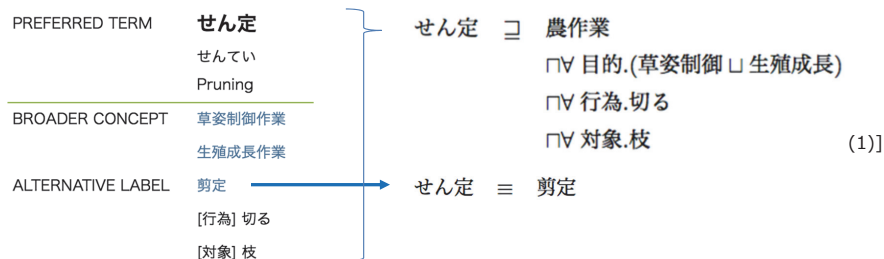
■ 別名の表記(同意語)

農作業名称は同じ作業なのに様々な別名が存在する。

- 1) 作業名に含まれる漢字が常用漢字ではない場合：は種 = 播種
- 2) 英語の作業名と日本語の作業名が混在する場合：ハードニング = 硬化
- 3) 最後に同じ作業でありながら判断の基準によって名称が異なる場合：秋耕 = 荒起こし
は種 = 種まき

■ 記述論理に基づく構築

記述論理を用いて表現



農作業基本オントロジー

農作業基本オントロジー(AAO, Agriculture Activity Ontology)の公開

Ver. 公開日 語数 属性数 階層の最大深さ

0.94	2015/05/12	185	2	6
1.00	2015/11/02	301	7	6
1.10	2016/02/12	330	7	7
1.31	2016/04/22	355	8	7



農作業基本オントロジー (Agriculture Activity Ontology)

AAO 農作業基本オントロジー (Agriculture Activity Ontology)

ID: [http://www.cavoc.org/aao/](#)

NAME: 農作業基本オントロジー

URL: [http://www.cavoc.org/aao/](#)

LATEST VERSION: 1.31

UPDATED: 2016-04-21

Download: [Turkish \(日本語\)](#) | [Turkish \(English\)](#)

AAO SPARQL Endpoint: [http://www.cavoc.org/aao/](#)

Version History

Version	Date	Concepts
1.31	2016-04-22	355
1.10	2016-02-12	330
1.00	2015-11-02	301
0.94	2015-05-12	185

[<http://www.cavoc.org/aao/>]

オントロジー構築手順

- 設計プロセス
 - 0th Step: プロジェクト形成
 - 1st Step: サーベイ
 - 2nd Step: データ分析
 - 3rd Step: 第1バージョン構築
 - 4th Step: 形式化(記述論理導入)
 - 5th Step: 専門家評価

設計プロセス

- 0th Step: プロジェクト形成-

- SIPプロジェクトの中のサブプロジェクト
 - メンバー
 - 農研機構2名
 - NII 2名



NII 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics



NARO 農研機構
国立研究開発法人 農業・食品産業技術総合研究機構

設計プロセス

- 1st Step: サーベイ -



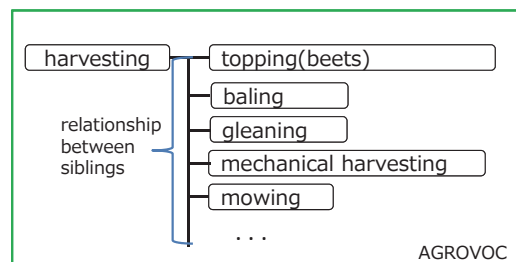
- 既存の語彙、タクソミー、オントロジーの調査
 - Agrovoc: FAO作成. この分野で最も著名な語彙
 - 国際的
 - 継続的メンテナンス
 - 機械可読 (LOD)
 - Agropedia
 - 日本語
 - 説明あり
 - 農水省ガイドライン
 - 公式
 - 農業統計

AGROVOC

AGROVOC はFood and Agriculture Organization(FAO)が用意するシソーラスで、約32000個の用語で、農業、漁業、食品、環境、関連分野をカバーする。

■ シソーラス Thesaurus

AGROVOC は単語を同義語synonym, 狭義/広義narrower/broader, 関連語 related relationshipで繋げている。



稲作等、日本の農業の特徴的な部分はそれほどサポートされていない。

- economic activities - agriculture - agricultural practices - agronomic practices - beating - bundling - handling - bulk handling - fish handling - food handling - loading - mechanical handling - pneumatic handling - pumping - stacking - unloading - fish unloading - harvesting - baling - gleaning - haymaking - manual harvesting - mechanical harvesting - mowing - picking - tapping - resin tapping - topping (beets) - windrowing - lifting - threshing - manual threshing - mechanical threshing - animal husbandry - animal husbandry methods	- economic activities - 農業 - agricultural practices - agronomic practices - beating - 結束 - 運搬 - 大量運搬、大量取り扱い、ばら荷扱い、バラ積取扱 - 水産物処理 - food handling - 載荷 - 機械運搬 - 空気による流動化操作 - ポンプ揚水 - 積み上げ - 荷おろし - 水産物水揚げ - 収穫 - 梱包 - 落穂拾い - 乾草作り - manual harvesting - 機械的収穫、機械収穫 - 草刈 - 葉ぬき - タッピング - 樹脂採取、乳液を採取する切付け - 甜菜のトップピング作業、摘心 - 地干し列、地干し刈り - 根菜堀上げ - 脱穀 - manual threshing - mechanical threshing - 畜産業 - 家畜飼養法
--	--

大項目	中項目	小項目（例示） 【ユーザーによって拡張（組み替え）可能】
1 種子等予措	1 選種 2 消毒 3 催芽 4 その他	種もみの選種 種子消毒 浸種 球根予措、種子コーティング
2 育苗	1 床土・苗床作り 2 は種 3 施肥 4 かん水 5 換気 6 間引き 7 防除 8 移植 9 除き木 10 その他	苗箱播種 等 苗床消毒 ずらし、間引き移植
3 耕起整地	1 耕起・砕土 2 整地 3 畝立て 4 温室、ビニールハウスの本ほ床土作り・床土入れ換え 5 その他	荒起し、秋田起し、代かき、ハロー等 畦塗り
4 基肥施肥	1 運搬 2 配合 3 施肥 4 養液栽培における定植時に行う養液作りや施肥 5 その他	緑肥すき込み、秋落ちを防ぐための客土、施肥濃度調整等
5 は種・定植 (本圃)	1 は種・覆土 2 苗取り・運搬 3 定植・移植 4 補植 5 その他	緑肥すき込み、秋落ちを防ぐための客土、施肥濃度調整等 直まき、種まき
6 追肥施肥	1 運搬 2 配合 3 施肥 4 養液栽培における定植後の養液の濃度調整 5 その他	1 圃場（けい野を含む）及び農道の整備・修復 2 施設の整備・修繕 3 かん排水 4 畝枝 5 生育調節（畝枝以外） 6 防風・防霜 7 排水・換気 8 見回り 9 収穫後の株（根）等の除去 10 その他
7 除草	1 中耕 2 除草剤散布 3 土入れ・土寄せ 4 草刈り・下刈り 5 マルチング 6 その他	畦畔除草、土壌改良材散布、圃場周回除草、防風林枝払い 等 ハウス修繕、機械設備、果樹樹の補修、マット交換・掃除 後追肥の補立て、取り壊し、資材器具の補立て、設置、取り壊し かん水、落水、水管理、排水調整等 せん定、支柱立て、誘引、摘心、摘果等 妻踏み、間引き、ホルモン剤処理、受粉、袋かけ等 こもかけ、防風ネット設置 等
8 病害虫防除 (雑草以外)	1 化学的防除（農薬散布） 2 化学的防除（土壌消毒） 3 物理的防除 4 生物防除 5 被害葉等の除去及び焼却 6 その他	生育調査、作柄見回り 等 鳥獣害対策、融雪剤散布 等 稲刈り、刈取り、摘取り、摘取り等 ラッピング 等 もみ運搬 等 稲架け 5 刈除 6 その他 稲わら処理、除袋、果実着色、貯蔵前予措等 乾燥、選別 等 もみ穀処理、もみ及び玄米の運搬 等
9 管理		1 乾燥・調製 2 もみすり 3 収穫物の水洗い 4 外葉除去・切断 5 切り枝・整形 6 その他
10 収穫		1 選別、包装、荷造り作業 2 出荷前の貯蔵作業 3 搬出、出荷作業 4 その他
11 調製		1 集金出庫（打合せ等） 2 出荷前の貯蔵作業 3 搬出、出荷作業 4 その他
12 出荷		1 集金出庫（打合せ等） 2 出荷前の貯蔵作業 3 搬出、出荷作業 4 その他
13 生産管理労働		1 ミニマライズ 等 2 農具修理、講習会受講 等 3 作業記録等の記録、パソコンデータ入力 等
14 間接労働		1 自給肥料の生産 2 運搬、自動車及び農機具の修繕 3 農具修理、講習会受講 等 4 水利用状況 5 その他

農業ITシステムで用いる農作業の名称に関する個別ガイドライン（試行版）

大項目	中項目	小項目（明示） 【ユーザーによって 異なる（仮）】	2014 y 農機、移植 機	2014 y 農機、乾田 直播	2014 y 農機、湛水 機	2014 y 農機（移植・ 湛水）	湛水機（移植・ 乾田直播・湛水機）	湛水機（移植・ 乾田直播・湛水機）	稲作アグリ（湛水・ 湛水機）	八幡（移植）
1 種子等移植	1 湛種 2 湛種（4 曜の含まれる） 3 湛種 4 湛種 5 その他	種もみの選種	湛種・湛水機							
2 育苗	1 湛土・湛種作り 2 は種 3 湛土 4 かん水 5 湛水 6 湛引き 7 防除（土壌消毒） 8 湛種 9 湛き木 10 その他	種もみの選種 ずらし、湛引き移植 古代理理、除草、 こまめかけ	育苗湛種機：その他 湛種機 古代理理（湛水の 中項目も含む） 湛種機：湛出し（の り）		湛種コーティング		湛種機 湛種機：コー ティング：ハ ウス湛種機	湛種機 湛種機：コー ティング：ハ ウス湛種機	湛種機 湛種機：コー ティング：ハ ウス湛種機	湛種機 湛種機：コー ティング：ハ ウス湛種機
3 移植地	1 湛種・湛土 2 湛土 3 湛土 4 湛土 5 湛土 6 湛土 7 湛土 8 湛土 9 湛土 10 湛土 11 湛土 12 湛土 13 湛土 14 湛土 15 湛土 16 湛土 17 湛土 18 湛土 19 湛土 20 湛土 21 湛土 22 湛土 23 湛土 24 湛土 25 湛土 26 湛土 27 湛土 28 湛土 29 湛土 30 湛土 31 湛土 32 湛土 33 湛土 34 湛土 35 湛土 36 湛土 37 湛土 38 湛土 39 湛土 40 湛土 41 湛土 42 湛土 43 湛土 44 湛土 45 湛土 46 湛土 47 湛土 48 湛土 49 湛土 50 湛土 51 湛土 52 湛土 53 湛土 54 湛土 55 湛土 56 湛土 57 湛土 58 湛土 59 湛土 60 湛土 61 湛土 62 湛土 63 湛土 64 湛土 65 湛土 66 湛土 67 湛土 68 湛土 69 湛土 70 湛土 71 湛土 72 湛土 73 湛土 74 湛土 75 湛土 76 湛土 77 湛土 78 湛土 79 湛土 80 湛土 81 湛土 82 湛土 83 湛土 84 湛土 85 湛土 86 湛土 87 湛土 88 湛土 89 湛土 90 湛土 91 湛土 92 湛土 93 湛土 94 湛土 95 湛土 96 湛土 97 湛土 98 湛土 99 湛土 100 湛土 101 湛土 102 湛土 103 湛土 104 湛土 105 湛土 106 湛土 107 湛土 108 湛土 109 湛土 110 湛土 111 湛土 112 湛土 113 湛土 114 湛土 115 湛土 116 湛土 117 湛土 118 湛土 119 湛土 120 湛土 121 湛土 122 湛土 123 湛土 124 湛土 125 湛土 126 湛土 127 湛土 128 湛土 129 湛土 130 湛土 131 湛土 132 湛土 133 湛土 134 湛土 135 湛土 136 湛土 137 湛土 138 湛土 139 湛土 140 湛土 141 湛土 142 湛土 143 湛土 144 湛土 145 湛土 146 湛土 147 湛土 148 湛土 149 湛土 150 湛土 151 湛土 152 湛土 153 湛土 154 湛土 155 湛土 156 湛土 157 湛土 158 湛土 159 湛土 160 湛土 161 湛土 162 湛土 163 湛土 164 湛土 165 湛土 166 湛土 167 湛土 168 湛土 169 湛土 170 湛土 171 湛土 172 湛土 173 湛土 174 湛土 175 湛土 176 湛土 177 湛土 178 湛土 179 湛土 180 湛土 181 湛土 182 湛土 183 湛土 184 湛土 185 湛土 186 湛土 187 湛土 188 湛土 189 湛土 190 湛土 191 湛土 192 湛土 193 湛土 194 湛土 195 湛土 196 湛土 197 湛土 198 湛土 199 湛土 200 湛土 201 湛土 202 湛土 203 湛土 204 湛土 205 湛土 206 湛土 207 湛土 208 湛土 209 湛土 210 湛土 211 湛土 212 湛土 213 湛土 214 湛土 215 湛土 216 湛土 217 湛土 218 湛土 219 湛土 220 湛土 221 湛土 222 湛土 223 湛土 224 湛土 225 湛土 226 湛土 227 湛土 228 湛土 229 湛土 230 湛土 231 湛土 232 湛土 233 湛土 234 湛土 235 湛土 236 湛土 237 湛土 238 湛土 239 湛土 240 湛土 241 湛土 242 湛土 243 湛土 244 湛土 245 湛土 246 湛土 247 湛土 248 湛土 249 湛土 250 湛土 251 湛土 252 湛土 253 湛土 254 湛土 255 湛土 256 湛土 257 湛土 258 湛土 259 湛土 260 湛土 261 湛土 262 湛土 263 湛土 264 湛土 265 湛土 266 湛土 267 湛土 268 湛土 269 湛土 270 湛土 271 湛土 272 湛土 273 湛土 274 湛土 275 湛土 276 湛土 277 湛土 278 湛土 279 湛土 280 湛土 281 湛土 282 湛土 283 湛土 284 湛土 285 湛土 286 湛土 287 湛土 288 湛土 289 湛土 290 湛土 291 湛土 292 湛土 293 湛土 294 湛土 295 湛土 296 湛土 297 湛土 298 湛土 299 湛土 300 湛土 301 湛土 302 湛土 303 湛土 304 湛土 305 湛土 306 湛土 307 湛土 308 湛土 309 湛土 310 湛土 311 湛土 312 湛土 313 湛土 314 湛土 315 湛土 316 湛土 317 湛土 318 湛土 319 湛土 320 湛土 321 湛土 322 湛土 323 湛土 324 湛土 325 湛土 326 湛土 327 湛土 328 湛土 329 湛土 330 湛土 331 湛土 332 湛土 333 湛土 334 湛土 335 湛土 336 湛土 337 湛土 338 湛土 339 湛土 340 湛土 341 湛土 342 湛土 343 湛土 344 湛土 345 湛土 346 湛土 347 湛土 348 湛土 349 湛土 350 湛土 351 湛土 352 湛土 353 湛土 354 湛土 355 湛土 356 湛土 357 湛土 358 湛土 359 湛土 360 湛土 361 湛土 362 湛土 363 湛土 364 湛土 365 湛土 366 湛土 367 湛土 368 湛土 369 湛土 370 湛土 371 湛土 372 湛土 373 湛土 374 湛土 375 湛土 376 湛土 377 湛土 378 湛土 379 湛土 380 湛土 381 湛土 382 湛土 383 湛土 384 湛土 385 湛土 386 湛土 387 湛土 388 湛土 389 湛土 390 湛土 391 湛土 392 湛土 393 湛土 394 湛土 395 湛土 396 湛土 397 湛土 398 湛土 399 湛土 400 湛土 401 湛土 402 湛土 403 湛土 404 湛土 405 湛土 406 湛土 407 湛土 408 湛土 409 湛土 410 湛土 411 湛土 412 湛土 413 湛土 414 湛土 415 湛土 416 湛土 417 湛土 418 湛土 419 湛土 420 湛土 421 湛土 422 湛土 423 湛土 424 湛土 425 湛土 426 湛土 427 湛土 428 湛土 429 湛土 430 湛土 431 湛土 432 湛土 433 湛土 434 湛土 435 湛土 436 湛土 437 湛土 438 湛土 439 湛土 440 湛土 441 湛土 442 湛土 443 湛土 444 湛土 445 湛土 446 湛土 447 湛土 448 湛土 									

大項目	中項目	小項目（指示） 【ユーザによって 拡張（組み替 え）可能】	2014y 農機、耕種	2014y 農機、乾田 整備	2014y 農機、湛水 農機	間刈農機（耕種、 整備）	湛野農機（耕種、 乾田整備、湛水農 機）	アグリタキスと （浮橋、乾田整 理・湛水農機）	耕畜アグリ（耕種、 整備）	八幡（耕種）
	4 追肥の配合									
	5 養液栽培における定植後の養液の濃度管理 や補充など									
	6 その他	施設溝張り等								
7 除草	1 中耕			出芽前除草剤、出 芽後除草剤、防草 剤散布		一発除草剤：除草 剤散布	本田除草剤	除草剤		除草剤散布（田 間）
	2 除草剤散布		除草剤散布		除草剤散布					
	3 土入れ、土寄せ									
	4 草刈り・下刈り		手取除草	手取除草	手取除草	草抜き				草刈（トラウ ク）：本田草刈 機
	5 マルチング	ゼニールマルチ、 苧わら、防草等								草引き：モア リ
	6 その他						アイガモ管理			
8 防除 （雑草以外）	1 化学的防除（農薬散布）								農薬散布	殺虫剤散布
	2 化学的防除（土壌消毒）									
	3 物理的防除	かんれいしや罫り 等								
	4 生物的防除	組皮ばさ、バント 捕殺、結核灯管理 等								
	5 被害箇所等の除去及び補却									
	6 その他		病害虫防除（中耕 により実施）	病害虫防除	病害虫防除	防除 1 回目：防除 2 回目		病害虫防除		播種時防除：防 1 回目、防除 2 回 目
9 管理	1 圃場（わいせを含む）及び農道の整備・修 復	いろいろな作業 がここに入る時 間が多いので、 4 か分 （具体的な例が必要 ではない）	間耕除草（除草の 4 か分）	間耕除草	間耕除草	間耕除草	整地・補修：保全 管理：春播：間耕 除草剤：間耕除草 剤散布	スライドモーター 路側刈り	間耕除草	
	2 施設の状態・修繕	ハウスの修繕、機械 整備、果樹園の補 修、マット交換・ 補修 暖房施設の結成 で、取りまき、暖 房設備の点検・修 繕、取りまき							農舎清掃	
	3 かん排水	かん水、排水、水 管理、排水溝張り 等	溝張り、水管理： 入水、中干し	溝張り、水管理： 入水、中干し、 排水、排水溝 整備	水管理：入水、中 干し、排水、明 水、排水溝整備	入水、中干し、 排水、排水溝 整備	溝張り、本田水 管理	排水作業、経路防 止：排水溝、排水 溝防除		
	4 栽培	せん定、支柱立 て、盛り、成込、 収穫等								
	5 生育調整（収成以外）	表踏み、間引き、 ホルモン処理等 こまかけ								
	6 防風・防霜									
	7 保水・湛水									
	8 臭い	生食調査、作柄見 回り								
	9 収穫後の（根）等の除去		結草摘取	結草摘取	結草摘取					
	10 その他	鳥獣害対策								
10 収穫	1 収穫	間刈り、刈取り、 播取り、掘取り等	間刈り	間刈り	間刈り	収穫：直播、刈取 乾燥（調製を含む む）：刈取乾燥機		コンバイン刈取り 機	稲刈（オペレ ーション）	収穫

設計プロセス

- 3rd Step: 第1版の提案 -

■ 階層関係の明確化

階層関係は人間が理解しやすい形式であるが、いろいろな関係性が混ざることが多い。

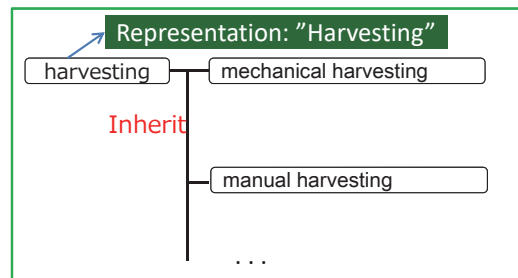
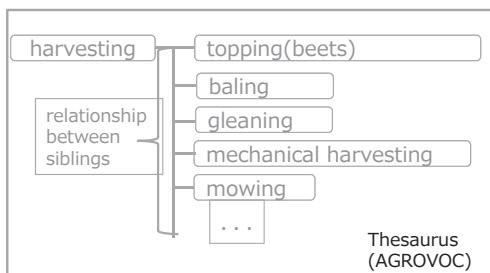
➡ ここでは、isa関係のみに限定する。



■ 同意語の許容

一つの概念がいろいろな用語で指されることがある（作物の違いなど）。

➡ 概念と表示される語の分離



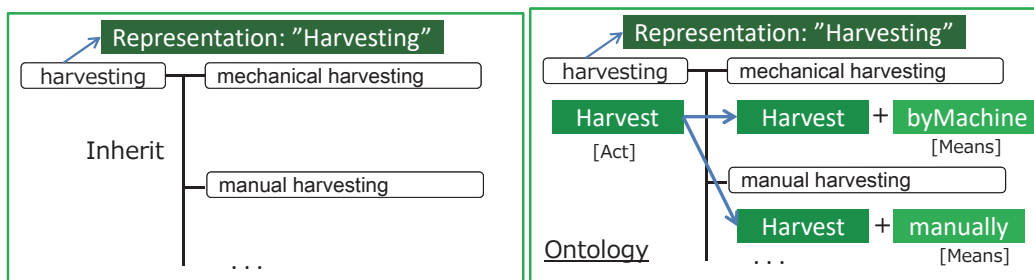
最上位階層	第二階層(概念)	第二階層(表記)	第三階層(概念)	第三階層(表記)	第四階層(概念)	第四階層(表記)	第五階層(概念)	第五階層(表記)	多重継承
作物生産作業	繁殖作業		種子繁殖作業		採種		選種		
					予措		催芽		
					は種				
			栄養繁殖作業		苗取り				
					挿し木				
					接ぎ木				
					取り木				
					株分け				
					分球				
	栄養成長調節作業		草丈制御作業						
			移植		定植		鉢上げ		
			間引き		補植				
			ハードニング(硬化)						
	生殖成長調節作業		外観制御				玉なおし(スイカ)		
			結実制御		受粉				
					無核化				
			開花制御						
			整枝		せん定				作物生産作業 > 栄養成長調節作業
	整枝 他の項目は「制御」これだけ性質が異なる。栄養成長調節が目的の場合もあり				誘引				
					摘果				
					摘葉				
					摘蕾				
					摘心				
環境制御作業	土壌の制御作業		圃場整地作業		耕起		ブラウ耕		
							秋耕		
							春耕		
					砕土		代かき		
					均平化		心土破碎		
					鎮圧				水分環境制御 > 排水
					溝掘り				環境整備作業 > 栄養の制御作業
					客土				
			根圏整地作業		中耕				
					耕+				

設計プロセス



- 4th Step: Introduction of Description Logics -

- Consideration of the structure
 - Discovery of logical structure
 - Reformation of the structure by Description Logics
 - Use of a property for each is-a relation
 - Introduction of a new property
 - Is-a hierarchy of a property value
 - Re-arrangement of classes



設計プロセス



- 5th Step: 専門家による評価と追加-

- 専門家への評価依頼
 - 作物ごとの専門家(農研機構)
 - 農業情報システム開発会社
- Feedback
 - 一部の概念の階層の変更
 - 多くの新規用語
 - 作物固有の用語
 - 地域固有の用語

Version History

Version	Date	Concepts
1.41	2017-01-05	410
1.33	2016-09-23	374
1.31	2016-04-22	367
1.10	2016-02-12	330
1.00	2015-11-02	301
0.94	2015-05-12	185

学んだこと

- 既存語彙のサーベイと批判
 - メリットとデメリットの理解
 - ターゲットを決める
- データ・ドリブンアプローチ
 - 抽象的な議論を避ける
- 両方からの少人数の専門家によるチーム編成
 - 建設的議論
- 核を作り、それを拡張する
 - その後、AIの専門家へ
 - その後、ドメインのより専門家へ
- コミュニケーション大事



CVO : 農作物語彙

農作物語彙体系(CVO)構築_背景・目的

【背景】

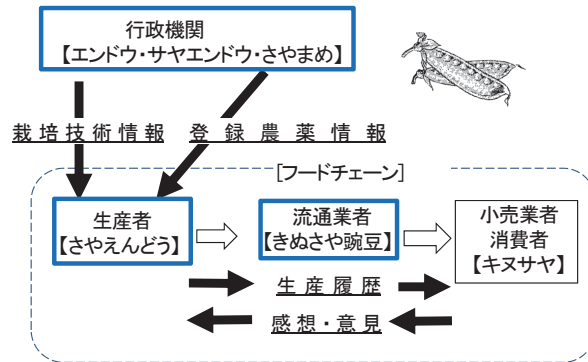
- ・フードチェーンでの食品の安全性確保の重要性の高まり。
=>フードチェーンの各段階で収集管理されたデータを相互利用できるシステムの要望。
- ・フードチェーンの多様なキープレーヤーは同一農作物のデータに異なる項目名を付けて管理
=>データの相互利用を阻害する要因の一つ。

【目的】

データ項目名が示す農作物概念の共通化を進めシステム間のデータ連携を促進するため**農作物語彙体系** (Crop Vocabulary; CVO)を構築・公開

【前提条件】

フードチェーンの上流
(生産者=>卸売業者)が対象
特に農薬情報の流通に着目して構築



農作物語彙体系(CVO)構築_特徴

1. 花き・樹木を除いた**農薬情報に含まれる農作物が主な対象**
2. 農作物を標準和名、学名、よみ、別名、英名等で表現
3. CVOの農作物名(標準和名)は
 - 1) 植物学的分類に基づいた**栽培植物**、
 - 2) 利用部位、用途、特性等により**栽培植物を分類した集合**を示す。
例1; 飼料用トウモロコシ、成熟トウモロコシ
例2; キュウリ(葉)、キュウリ(花)
4. CVOの農作物語彙は外部語彙リストとマッピングされる。
農薬・食品成分・学術・品種情報
=> CVOを基盤とした語彙リストの連携が実現
5. 農研機構OB・大学教授による精査(作物・果樹・野菜専門家)
現在も継続中

農作物語彙体系(CVO)構築_公開



<http://www.cavoc.org/cvo/ns/1/キュウリ>

農作物語彙体系 (Crop Vocabulary)

【農作物情報】

標準和名	キュウリ
読み	きゅうり
科名	ウリ科
英名	Cucumber
別名	キウリ 胡瓜 きうり
学名	<i>Cucumis sativus</i>
下位作物名	キュウリ(葉) キュウリ(花)

【CVO語彙リスト】

- * 科ごとに整理
- * 一部階層構造

－ キュウリ
 キュウリ(葉)
 キュウリ(花)

－ キンショウ
－ 金槌
－ 銀糸

**CVOは固有の名前空間を持ち
各概念はその上で定義**

語彙リスト	発行機関	対応語彙数
農薬登録における適用農作物名	農林水産省	972語
農産物等の食品分類表	厚生労働省	743語
食品の原料となる生物種の英名・学名	文部科学省	274語
Taxonomy Database	米NCBI	972語
ウィキペディア	ウィキメディア財団	768語
CAVOC品種リスト	CAVOC	69語

データ・ドリブンアプローチ

農薬登録における適用農作物名
(農林水産省)

農作物名称
ガイドライン
(農水省,
2017)

農作物名称
ガイドライン
(農水省,
2016)

家計調査
(総務省)

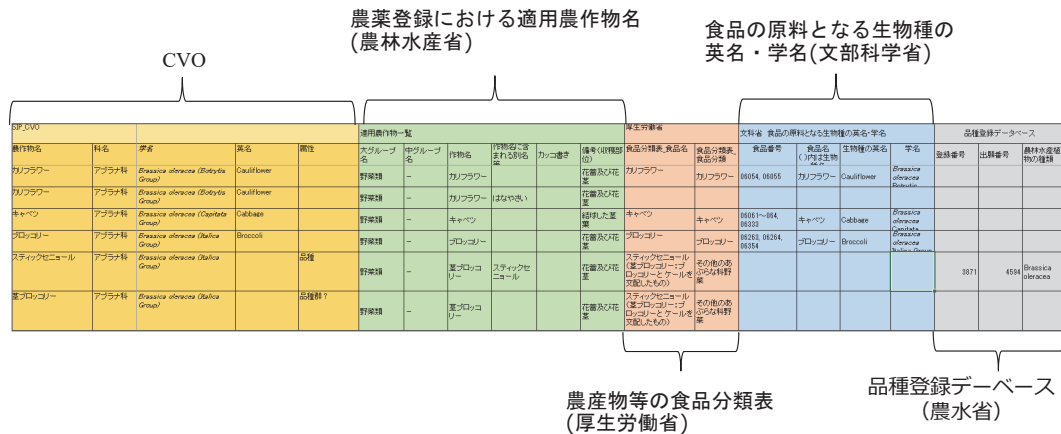
ベジブル
コード

食品の原料
となる生物
種の英名・
学名(文部科
学省)

[illegible]

CVO (Crop Vocabulary)

CVO has links to other standards in food industry, agricultural chemical industry.



31

CVO構築

- 作物 $\neq$ 種
 - 作物 = 種
 - 種 + 食用/非食用
 - + 利用部位
 - + 成熟/未成熟
 - + 栽培方法
 - + 品種
 - 最小限の関与
 - それは実際に使われているか? (市場で? 農家で?)
 - 適切な名前の付与 Give the suitable names
 - 種名 (総称)
 - 種名 (利用部位)
 - 専門家からのフィードバック
- 名前をどうする?

共通語彙基盤での経験から

共通語彙基盤(IMI)

- 電子行政分野におけるオープンな利用環境整備に向けた政府のアクションプランの一環
– 「世界最先端IT国家創造宣言・官民データ活用推進基本計画」[1]
- データに用いる文字や用語を共通化することで情報の共有や活用を円滑に行うための基盤

語彙の設計方針

- 日本語の語彙とする
- 意味や構造は明確になるようにする
- 業界を越えての情報交換を可能にする
- 用途に応じて適した物理形式によるデータ交換・共有が可能とするため、特定の物理形式に依存しないようにする
- 一部の項目を利用することや、項目を追加すること、他の語彙との併用が可能となるようにする
- 継承などにより、コア語彙など基本的な語彙を最大限活用できるようにする
- 既存システムの変更をすることなしに、データ交換の際に利用できるようにする
- 順次導入が可能なようにする

IMI共通語彙とは

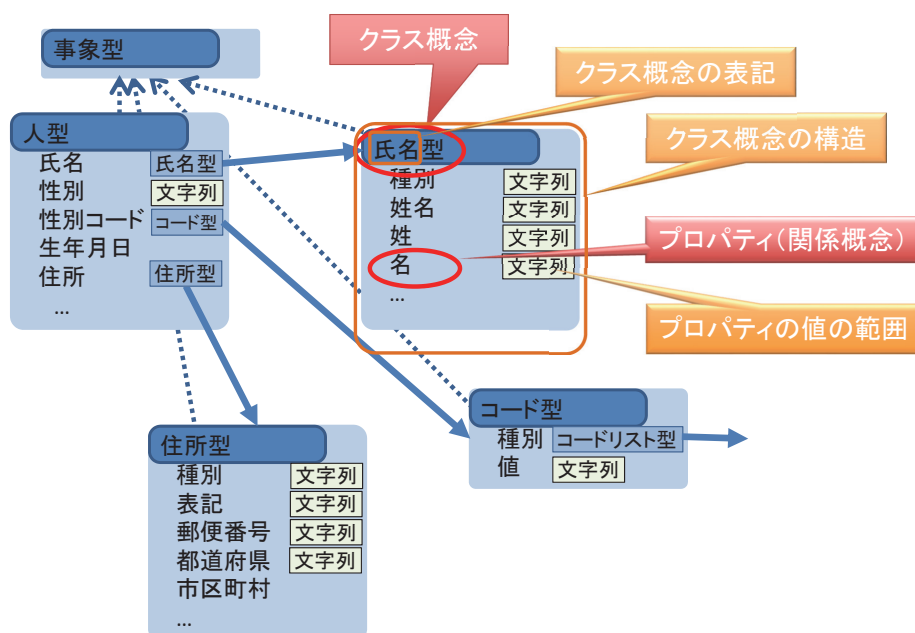
- **構造化概念辞書**
 - － 概念辞書
 - **概念の表記としての用語**
 - － 各項目は概念であって用語でない。
 - － 構造化辞書
 - 概念は相互につながっていて、**その組み合わせ(構造で意味を表現する**

用語の種類

用語は「クラス用語」と呼ばれる用語と「プロパティ用語」と呼ばれる用語の2種類からなる。どちらも言葉として意味をもっているので「用語」と呼ぶが、その構造は異なる。

- **クラス用語(クラス概念)**
 - ある事柄を表現する用語。自身の名前とプロパティ用語を構成要素として持つ。
- **プロパティ用語(関係概念)**
 - 事柄の性質や属性を表現する用語および事柄と事柄の関係を表現する用語。自身の名前と使われる事柄(クラス用語)、指し示すもの(クラス用語あるいは値の型)から構成される。

IMI共通語彙の構造



クラス用語

- 項目名、説明、継承、及び、プロパティ用語の集合で構成される
- 各プロパティに対して、カーディナリティ(出現数)とクラス用語における意味の説明をもつ ※ 「クラス用語」とは、項目名のことではないことに注意

人

識別子:

ic:人型

継承:

ic:実体型

説明:

人の情報を表現するためのクラス用語

プロパティ:

識別子	項目名	値型	回数	説明
ic:ID	ID	ic:ID型	0..n	人に割り振られたIDを記述するためのプロパティ用語
ic:氏名	氏名	ic:氏名型	0..n	氏名を記述するためのプロパティ用語
ic:性別	性別	xsd:string	0..1	性別の表記を記述するためのプロパティ用語
ic:性別コード	性別コード	ic:コード型	0..1	性別コードを記述するためのプロパティ用語
ic:生年月日	生年月日	ic:日付型	0..1	生年月日を記述するためのプロパティ用語
ic:死亡年月日	死亡年月日	ic:日付型	0..1	死亡年月日を記述するためのプロパティ用語
ic:住所	住所	ic:住所型	0..n	現住所を記述するためのプロパティ用語
ic:本籍	本籍	ic:住所型	0..1	本籍を記述するためのプロパティ用語
ic:連絡先	連絡先	ic:連絡先	0..n	連絡先を記述するためのプロパティ用語

プロパティ用語

- クラス用語を説明するためのクラス用語や値を表すための用語
- 項目名、データ型、及び説明から構成される
- データ型はクラス用語の型かXMLスキーマで定義される基本型をとる

項目名	データ型	説明	定義域
氏名	ic:氏名型	人の氏名	△

- クラス用語の子要素として使用される

人

識別子:

ic:人型

継承:

ic:実体型

説明:

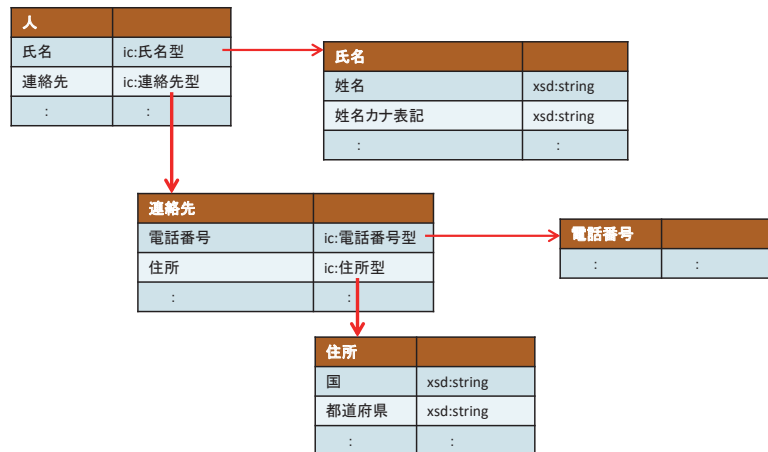
人の情報を表現するためのクラス用語

プロパティ:

識別子	項目名	値型	回数	説明
ic:ID	ID	ic:ID型	0..n	人に割り振られたIDを記述するためのプロパティ用語
ic:氏名	氏名	ic:氏名型	0..n	氏名を記述するためのプロパティ用語
ic:性別	性別	xsd:string	0..1	性別の表記を記述するためのプロパティ用語
ic:性別コード	性別コード	ic:コード	0..1	性別コードを記述するためのプロパティ用語

用語の構造

- プロパティ用語のデータ型としてクラス用語を指定することができる。その場合、データ型であるクラス用語がまたプロパティ用語をもつので、図のような概念的な構造をもつ



クラス用語の継承

- 他のクラス用語を継承して新たなクラス用語を作成することができる
- 継承したクラス用語は、継承元のすべてのプロパティをもつ
- 継承したクラス用語は継承元にはないプロパティを追加することができる
- 継承したクラス用語は継承元のクラス用語としても使用できる。例えば、データ型が「組織」となっているプロパティには、「業務組織」の値を指定することができる。



共通語彙基盤 コア語彙2.4.1

To the extent possible under law, Ministry of Economy, Trade and Industry and Information-technology Promotion Agency, Japan has waived all copyright and related or neighboring rights to Infrastructure for Multilayer Interoperability (IMI) Core Vocabulary 2.4.1. This work is published from: Japan .

赤い取り消し線が引かれた用語およびプロパティは将来の版では使用できなくなるため、使用を推奨しません。

=>プロパティ用語一覧

クラス用語一覧

人型
連絡先型
法人型
場所型
製品型
施設型
設備型
測定値型
重量型
価格型
日付型
定期スケジュール型
単位コード型
概念型
サービス型
コード制約型

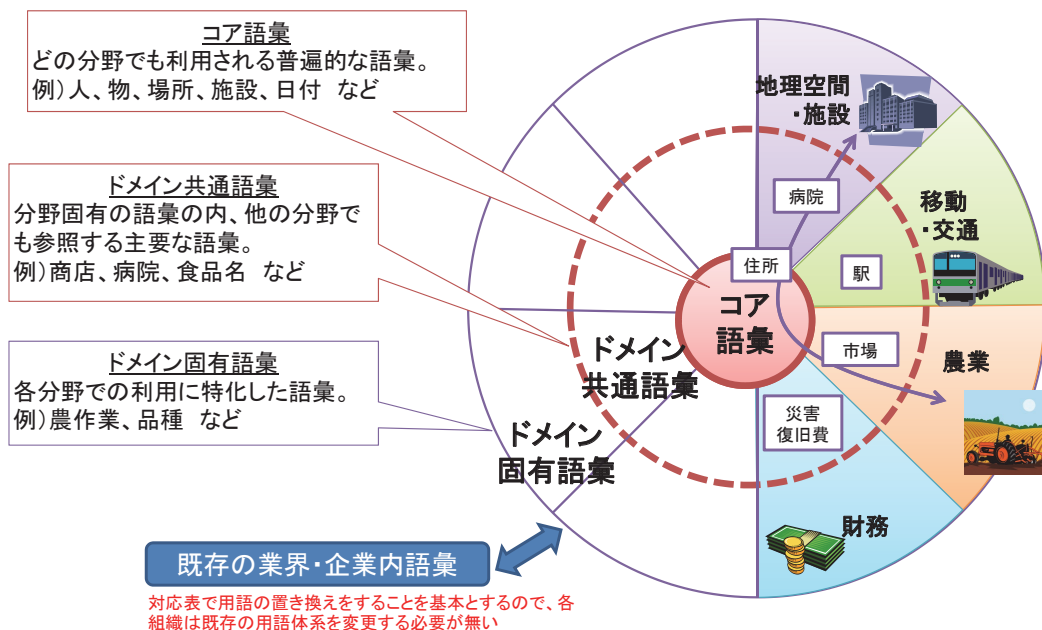
名称型
電話番号型
人数型
座標型
製品個品型
施設関連型
イベント型
数量型
長さ型
実体型
期間型
詳細スケジュール型
コードリスト型
事物型
文書型
範囲制約型

氏名型
組織型
構成員型
ID型
地物型
建物型
活動型
容量型
時間型
状況型
期間スケジュール型
詳細スケジュール規則型
アクセス型
参照型
対象型
期間制約型

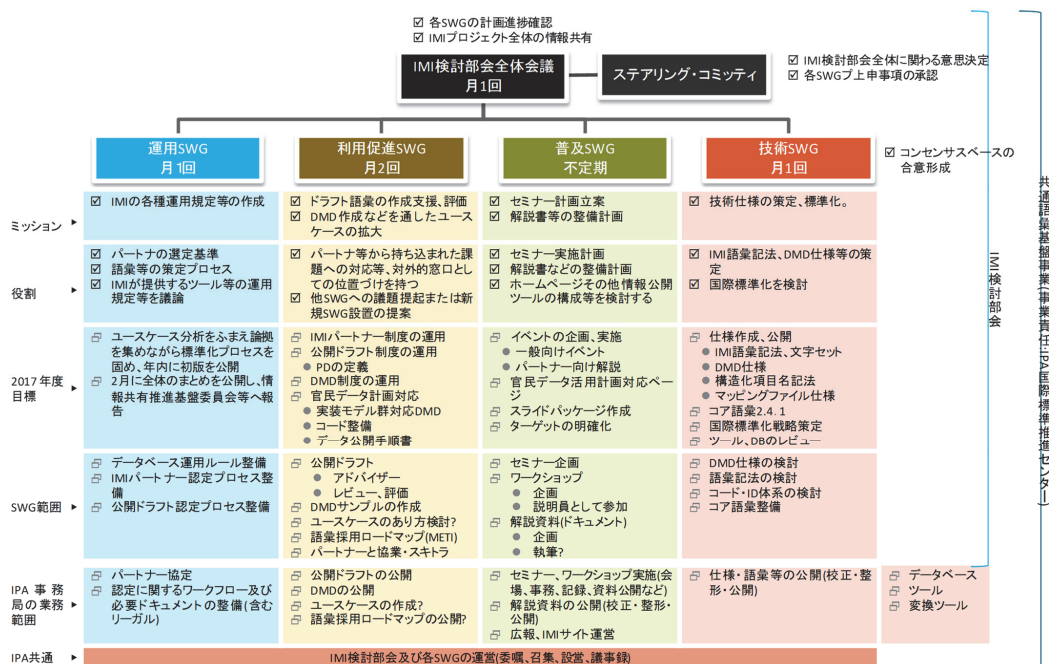
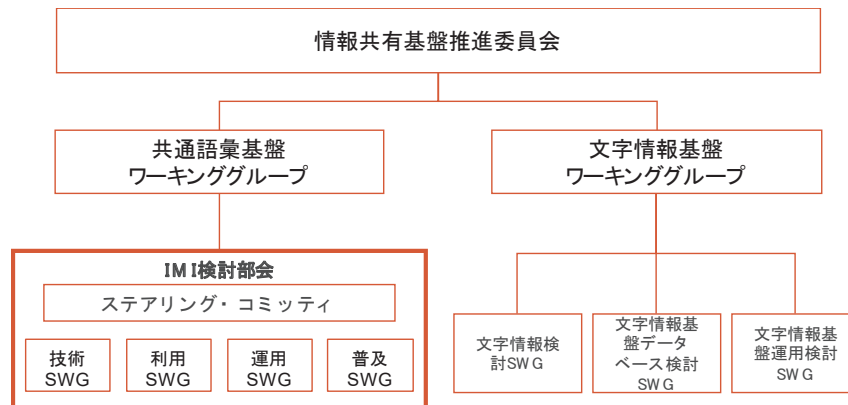
住所型
業務組織型
組織関連型
ID体系型
土地型
駐車場型
関与型
面積型
金額型
日時型
イベントスケジュール型
コード型
アクセス区間型
記述型
制約型

IMI共通語彙基盤の構造

- 情報交換のための用語連携インターフェースは、社会全体で使われる用語であるコア語彙、共通語彙及びドメイン固有語彙から構成される。



語彙基盤のマネージメント



いくつかのタスク

- プロジェクトの維持
- コア語彙のメンテナンス
- ドメイン語彙、応用語彙の構築

コア語彙のメンテナンス

- データの提供
 - 多様フォーマット
 - 共通語彙記法
 - RDF (JSON-LD, turtle)
 - XML
 - 構造化項目名
 - HTML
- バージョンアップ
 - メジャー／マイナー／マイナー・マイナー
 - バージョンアップ手順の制式化
 - 提案 – 意見募集 – 決定 – 周知
 - バージョンアップ手順のシステム化

ドメイン語彙、応用語彙の構築

- ドメイン語彙の構築
 - あるドメインの共通語彙。誰かがこれをコア語彙の同様に維持。
 - 今の所、未構築
- 応用語彙
 - 特定の用途のための語彙。
 - 誰でも提案できる。ただし、IMIパートナーとなることでIMIサイト上にPD(Public Draft)語彙として置くことができる。
 - パートナーとIMIとの協同作業で語彙を提案

IMIパートナー一覧

2017年5月現在のIMIパートナーです(五十音順)。

株式会社アスコエパートナーズ 



一般社団法人XBRL Japan 



一般社団法人オープン・コーポレイツ・ジャパン 



公益財団法人 九州ヒューマンメディア創造センター 



Code for Tokyo 



バイザー株式会社 



ビッグデータ&オープンデータ・イニシアティブ九州 



一般社団法人 ユニバーサルメニュー普及協会 



まとめ

オントロジー・語彙の構築と維持

- 構築する
 - オントロジー・語彙の専門家とそのドメインの専門家との協調プロセスをいかに作るか
- 維持する
 - 実は大変。
 - オントロジー・語彙は変化する。それをどう維持するか。
 - 体制づくり

2. 「辞書記述の精緻化と シソーラス・コーパスの活用」

2017/12/19

辞書記述の精緻化と シソーラス・コーパスの活用

柏野 和佳子



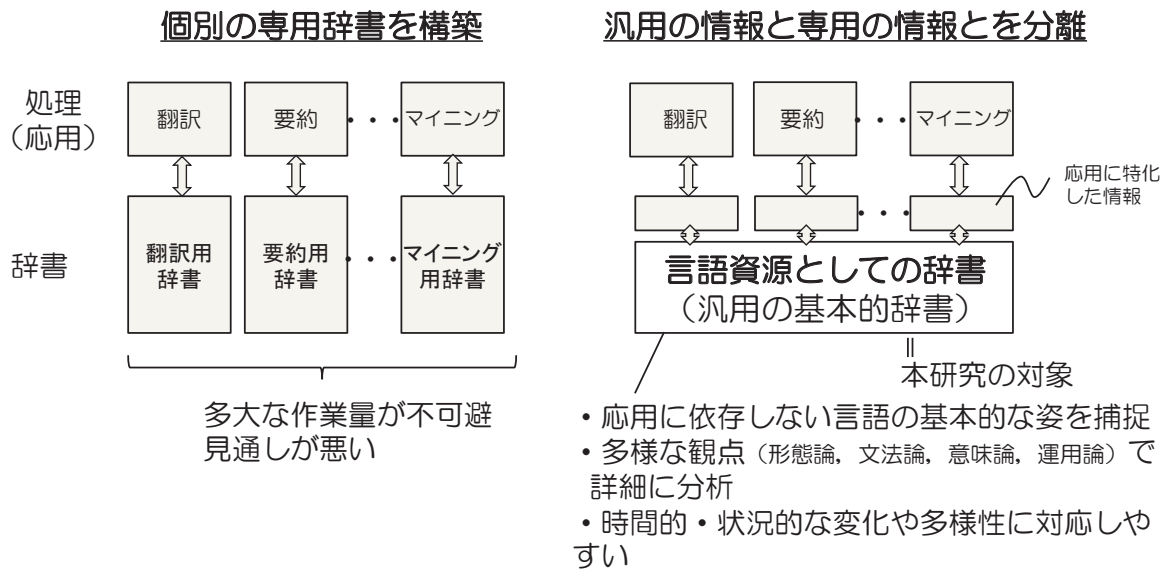
辞書記述の精緻化

柏野和佳子 (2016)コーパスに基づく辞書記述の精緻化の研究.東京工業大学総合理工学研究科博士論文.

背景

「言語資源としての辞書」の構築の必要性

汎用の基本的辞書の整備が応用上も言語学上も重要



3

これまでの問題点と本論文の目的

国語辞典 電子化したものを様々な言語処理に適用した例は多い
計算機用辞書 例：IPAL(1987～1996), EDR(1993), NTT日本語語彙体系(1997)

問題点

編集者の知識や内省への依存度が高い

- ① 国語辞典 → 母語話者利用を暗黙の前提としているため重要な情報が不足
- ② 計算機用辞書 → これまで以上に精度の高い解析・生成処理を実現するためにはまだ情報が不足



本論文の目的

- (1) 辞書記述の骨格となる見出し語の選定・語義区分・多義語の構造の記述を客観化・明示化して行う方法を提案する。
- (2) 解析・生成処理の精度向上のために、コーパスに基づく新たな辞書情報の記述方法を提案する。
- (3) 均質的・効率的な辞書構築を作業ベースで行えるように辞書記述を手順化する。

4

辞書に記載すべき情報と本論文で着目する情報

見出し語

見出し語レベルの《メタ情報》**頻度情報**, 《形態情報》表記,
《統語情報》語種, 品詞, 活用型など, 《意味情報》**多義の構造**

語義1

《意味情報》語義の記述1

語義1レベルの《メタ情報》
頻度情報など...

語義1レベルの《形態情報》
表記など...

語義1レベルの《統語情報》
コロケーションなど...

語義1レベルの《意味情報》
使用域など...

語義2 ...

《意味情報》語義の記述2

5

辞書に記載すべき情報と本論文で着目する情報

見出し語

見出し語レベルの《メタ情報》**頻度情報**, 《形態情報》表記,
《統語情報》語種, 品詞, 活用型など, 《意味情報》**多義の構造**

語義1

《意味情報》語義の記述1

語義1レベルの《メタ情報》
頻度情報など...

語義1レベルの《形態情報》
表記など...

語義1レベルの《統語情報》
コロケーションなど...

語義1レベルの《意味情報》
使用域など...

語義2 ...

《意味情報》語義の記述2

用例を収集し, 「見出し語」を選ぶ。

用例を分析し, 「多義の構造」を記述する。

用例を分析し, 「区分」する。

用例を分析し, 《各情報》を記述する。

6

本論文の着眼点と 提案する辞書記述の要件

(1) 辞書記述の骨格における客観化・明示化の提案

要件1-1 見出し語の選定基準

- 対象とする語の取舍選択の主観への依存性を低減

要件1-2 語義の区分

- 区分の主観への依存性を低減，文法・用法による多義性の解消が可能

要件1-3 多義語の意味関係の構造

- 多義性の仕組みに根差した多義性の解消が可能

(2) 辞書に記述すべき新規な情報の提案

要件2-1 多義性をもつコロケーション（名詞句＋格助詞＋述語）の意味の分離

- コロケーションへの修飾情報の記述により多義性の解消が可能

要件2-2 語の使用域と位相

- 文脈に相応しい語選択が可能

要件2-3 語の使用頻度の経年変化

- 語の流行り廃りなどの情報により文脈や目的に応じた語選択が可能

7

コーパスに基づくことによる問題解決の見通し

(1)客観的・明示的になる

頻度情報や実例に基づき，辞書情報の抽出と記述とが客観的・明示的に行えるようになる。

(2)網羅的になる

生きた用例を可能な限り網羅的に収集できるようになる。

(3)具体的になる

資料の幅が広がり，語の多様性を具体的に捉えることができる。ジャンルあるいはレジスタなどと呼ばれる，言葉の使用域が異なる場合，それを具体的に辞書記述に反映できる(例えば，話し言葉のか，書き言葉のか，フォーマルか，インフォーマルか)。

8

要件1-3 多義語の意味関係の構造①

用例より区分間の意味関係を認定し、明示化

本論文での適用対象：「名詞の下位区分間の意味関係」 （『IPAL名詞辞書』）

- 1 区分間の関係を、メタファー(隠喩)、メトニミー(換喩)、シネクドキー(提喩)のいずれで捉えられるか判定し、記号と区分番号とを用いて関係を記述する。
- 2 拡張の方向性は、歴史的な派生順序にできるだけ沿うように与える。（『日本国語大辞典』（小学館、1976年）を適宜参照）
- 3 可能な限り単純明瞭に記述するために、3タイプの比喩の融合と捉えられる場合も、いずれのタイプにできるだけ特定して記述する。
- 4 比喩によって捉えきれない多義については、できるだけパターン化した文章でその多義がどう捉えられるかを記述する。

Wakako Kuwahata, Minako Hasimoto, Fumihiro Aoyama (1996) Senses of Polysemous Nouns : Building a Computational Lexicon of Basic Japanese Nouns. *Proc. of The 16th International Conference on Computational Linguistics (COLING1996)*, pp.1082-1085.
 柏野和佳子, 本多啓 (1998) IPAL名詞辞書における多義構造の記述. 『情報処理学会論文誌』 39(9), pp.2603-2612.

9

比喩に基づく意味の拡張関係

- **シネクドキー（提喩）** カテゴリ全体とその一部という包摂関係によって意味を拡張する
 - ・ 着物：着物 (衣服) を揃える ⇨ 着物 (和服) を着付ける
 - ・ 酒：熱燗の酒 (日本酒) ⇨ 世界各国の酒 (アルコール飲料)
 - ・ スピード：スピードが遅くなる ⇨ スピードがある
 - ・ 広さ：部屋の広さに圧倒された ⇨ グラウンドの広さをはかる
- **メタファー（隠喩）** 類似性によって意味を拡張する
 - ・ 肩：人の肩 ⇨ 洋服の肩, ... ⇨ 山の肩, 文字の肩
 - ・ 皮：動物の皮 ⇨ みかんの皮, まんじゅうの皮, ... ⇨ 化粧の皮
- **メトニミー（換喩）** 密接な関係にあるものや属性などの隣接関係によって意味を拡張する
 - ・ 肩：肩がこる [身体部位] ⇨ [機能] いい肩をしている
 - ・ 鍋：鍋で煮る [容器] ⇨ [中身] 鍋を食べる
 - ・ 茶：茶を摘む [材料] ⇨ [加工品] 茶を入れる
 茶を入れる[手段] ⇨ [行為] お茶にする

（柏野和佳子, 本多啓(1998) 多義構造を辞典に書く. 『日本語学』明治書院, 17(14), pp.54-63.
 ほか, ウルマン1962, レイコフ・ジョンソン1986, 国広1982, 山梨1995）

10

要件1-3 多義語の意味関係の構造②

提案する記述の具体例1

「のど」

[区分01] 管の形をしている口の奥の部分。

例：のどに魚の骨がひっかかってしまった。

[区分02] 頭と肩をつなぐ身体部分の前面。

例：喉に湿布をする。

[区分03] 歌う声。

例：彼女の 自慢の喉を聞かせてもらいましょう。

[区分04] 本の綴じ目の部分。

例：本ののどがこわれる。

＜関係＞	01 [物] → [側面] 02。	: メトニミー
	01 [手段] → [行為] 03。	: メトニミー
	02 ... > 04。	: メタファー

11

提案する記述の具体例2

例：「あし【足，脚】」

[区分01] 動物の胴に付属し、歩行や体を支える機能を果たす身体部分。

例：彼女は足が長い。

[区分02] 「あし01」の一部で、くるぶしより下の部分。

例：彼は足が大きい。

[区分03] 歩いたり、走ったりすること。 例：彼は足が速い。

[区分04] 移動の手段となるもの。 例：台風で市民の足が乱れる。

[区分05] 椅子などの下部で、台の部分や上にある物を支える役割を果たす部分。

例：机の脚が折れる。

01 >> 02。 01 [身体部位] → [機能] 03 ... > 04。 01 ... > 05。

[区分02]は[区分01]と全体部分関係のメトニミー

[区分03]は[区分01]と[身体部位]と[機能]の関係のメトニミー

[区分04]は、その[区分03]からの機能の類似性によるメタファー

[区分05]は[区分01]からの形態的な類似性によるメタファー

12

多義構造の記述の試行

辞書に収録する多義語の意味的關係を記述 → 全体の約 86% の記述が可能

- 多義語560語の区分間の数 1,364
- そのうち比喩的關係として捉えて記述できた数 1,174(86%)

表4.14 各カテゴリにおける意味的關係の記述数

カテゴリ	メタファー		メトニミー			シネクドキー				(他)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
サ変	53	18	3		7	34				8
形容詞転成	67				1	10		4	3	
形容動詞	12	1			8	21		2		12
相対	75	42			38	24	1			49
副詞	19	2			5	8		1		3
述語用法2	23	27		1	12	21		15		14
被修飾2	24	13		1	7	52		1		19
動詞転成	5	5			5	8				
植物		22	11	2	6	4				
食べ物	6	4			16	6	2			4
身体	25	56	10	4	47	2				17
動物	1	14	3	2	9	9	2			11
場所	5	9	2	1	5	19	1			3
建物・家具	1	12	1	1	6	3				3
乗り物		2		2	3	1				
洋装品	4	11			4	3				3
仕事・役割	15	9			12	20				22
自然物	3	15	2	1	18	5				12
自然現象	5	8			2	5		1		6
組織・団体	2	2			4	7				4
小計	345	272	32	15	215	262	6	24	3	190
合計	617		262			295				190

合計 1,174

13

多義性をもつコロケーション

表5.1 重複記載の4パターン

a. 異区同素(異なる区分で同じ素性と重複)	3,608例
(例) [区分01] ACT (敵軍の)攻撃が始まる	
[区分02] ACT (議長への)攻撃が始まる	
b. 異区異素(異なる区分で異なる素性と重複)	2,260例
(例) [区分01] MEA (大豆の)収穫が多い	
[区分02] GRA (会議の)収穫が多い。	
c. 同区同素(同じ区分で同じ素性と重複)	821例
(例) [区分01] CON 自転車(のサドル)が高い	
区分01] CON 自転車(の価格)が高い	
d. 同区異素(同じ区分で異なる素性と重複)	934例
(例) [区分01] CON (シャワーデ)汗ヲ流す	
[区分01] LIQ (額二)汗ヲ流す	
合計	7,623例

14

コロケーション重複の要因

要因(A)-1 多義である名詞が同じ素性を持つため、素性に特有の同じ述語と結び付く。

例えば、〈LOC〉(場所)であれば「行く、戻る、広い、狭い」、〈ACT〉(行為)であれば、「始まる、終わる、止まる、早い、遅い」といったものが素性に特有の述語である。

要因(A)-2 述語の抽象度が高いため、同じ述語と結び付く。

例えば、「ある」「ない」「する」「なる」といった述語は抽象度が高いため、見出し語が多義であるとき、それぞれに結び付きやすく、その結果、重複する数も多くなる。

要因(B)-1 名詞と述語の組合せであるコロケーション自体が比喩的意味を持つため、同じように結び付く。

区分間にメタファーの関係があると、コロケーションが重複しやすい。

要因(C)-1 多義である述語が同じ名詞と結び付く。

述語の指す意味はまったく異なるのに、それが異なる述語ではなく、同じ1つの述語の多義によるものであるため、結果として重複が生じる。

15

重複記載を区別する違い

表5.2 重複記載を区別する違いの5種類

a. 他の名詞句が立つ格助詞が違う(他格)	2,421例
(例) 02:CON～ガ～二お茶ヲ入れる	
02:PRO～ガお茶ヲ入れる	
b. 他の名詞句が違う(他名)	1,145例
(例) 02:HUM(控え室)ニ家族ガいる	
02:ROL(彼女)ニ家族ガいる	
c. 見出し語の先行句が違う(先行)	4,888例
(例) 01:ACT～ガ(原野の)開拓ヲ行う	
02:ACT～ガ(技術の／新市場の)開拓ヲ行う	
d. 見出し語の後行句が違う(後行)	91例
(例) 01:POTのど(の調子)ガいい	
03:FORのどガいい(＝歌う声がいい)	
e. 見出し語が立つ格助詞が違う(格助)	1,794例
(例) 01:INT～ガ舞台ヲ／カウ出る	
01:SPA～ガ舞台ニ出る	

16

要件2-1 多義性をもつコロケーションの意味の分離①

コロケーションへの修飾情報を記述し、コロケーションのもつ多義性を解消

本論文での適用対象：「多義性をもつ名詞のコロケーション」（『IPAL名詞辞書』）

- 1 コーパスからコロケーション（名詞句＋格助詞＋述語）を抽出する。
- 2 「～の」の形で名詞句を連体修飾する用例を詳細に記述する。
- 3 形容詞や「～な」の形で名詞句を連体修飾する用例を詳細に記述する。
- 4 テイル形に関するアスペクト情報や副詞表現についての情報抽出が可能であれば抽出し、それも詳細に記述する。

例：「骨が折れている」→慣用句（苦勞という意味）の用法ではない。

桑畑(柏野)和佳子, 橋本三奈子, 青山文啓 (1998) IPAL名詞辞書による多義性解消のためのコロケーションの分析.
『情報処理学会論文誌』39(6), pp.1925-1934.

17

要件2-1 多義性をもつコロケーションの意味の分離②

提案する記述の具体例

「いろ【色】」

[区分01] GRA(口紅の／…)色が濃い

〈対象〉[CON | TIM | …]信号の一, 表紙の一, 服の一, 炎の一, 花の一, 醤油の一, 夕暮れの一, 緑の一, 葉の一。

[区分03] APPあせりの／…色が濃い

〈部位〉顔の一, 目の一。

〈具体化(心情を表す名詞)〉不安の一, 警戒の一, 反省の一, 焦慮の一, 憂慮の一, 苦悩の一, 傷心の一, いらだちの一。

[区分05] GRA敗戦の／…色が濃い

〈具体化(状況を表す名詞)〉秋の一, 自己防護の一, 敗北の一, 崩壊の一。

[区分07] GRAハト派の／…色が濃い

〈具体化(思想を持つ団体やその考え方・行動を表す名詞)〉タカ派の一, 保守の一, 過激なイスラム原理主義の一, 徒弟制度の一, 仏教国の一。

18

辞書情報による重複記載の多義性解消の確認

辞書に収録された多義性のある用例を記述 → 全体の約 84.4% の多義性が解消

- 重複するコロケーション(延べ10,415例, 異なり4,578例)の重複対 7,623例
- そのうち記述により多義性の解消ができた数 6,436例(84.4%)

表5.2重複記載を区別する違いの5種類

重複記載を区別する違い5種類						重複記載を区別する違い5種類					
タイプ	他格 他名	先行	後行	格助	出現数(%)	タイプ	他格 他名	先行	後行	格助	出現数(%)
1	-	-	-	-	1,187(15.6)	13	+	+	-	-	594(7.8)
2	-	-	-	+	77(1.0)	14	+	+	-	+	913(12.0)
3	-	-	+	-	20(0.3)	15	+	+	+	-	2(0.0)
4	-	-	+	+	2(0.0)	16	+	+	+	+	3(0.0)
5	-	+	-	-	2,594(34.0)	17	×	-	-	-	524(6.9)
6	-	+	-	+	162(2.1)	18	×	-	-	+	15(0.2)
7	-	+	+	-	13(0.2)	19	×	-	+	-	0(0.0)
8	-	+	+	+	2(0.0)	20	×	-	+	+	1(0.0)
9	+	-	-	-	330(4.3)	21	×	+	-	-	551(7.2)
10	+	-	-	+	533(7.0)	22	×	+	-	+	52(0.7)
11	+	-	+	-	12(0.2)	23	×	+	+	-	2(0.0)
12	+	-	+	+	34(0.4)	24	×	+	+	+	0(0.0)
計											7,623(100.0)

これ以外の合計 6,436

19

要件2-3 語の使用頻度の経年変化①

頻度情報により経年変化を明示化

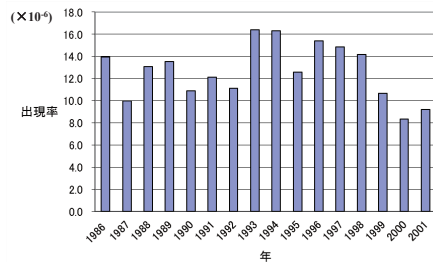
本論文での適用対象：「外来語」

- 経年のデータが蓄積している新聞データベース等を時間的使用推移の頻度情報を得て、使用推移のグラフを作成する。
- 1で作成したグラフを基に、使用推移が次のいずれかの型であるかを判断する。
 - 平行型 平ら：出現期間内をならすと平ら
 - 増加型 右上がり：出現期間内をならすと増加
 - 減少型 右下がり：出現期間内をならすと減少
 - 山 山：3年以上の増加後3年以上の減少
 - 山の後平ら 山の後平ら：増加，減少，平らが各3年以上で連続
 - 凹凸型 突出：特定の年のみ増加
 - 上記以外
- 2で判断した推移の型を辞書に明記するとともに、その判断の元となった頻度情報を記載する。頻度情報は、人が閲覧する用途では、閲覧用プログラムが必要に応じてグラフとして表示してもよい。

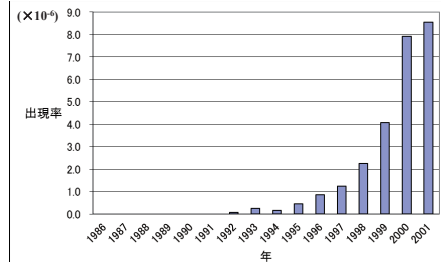
柏野和佳子, 山口昌也, 桐生りか(2005) 新聞記事データベースを利用した外来語の出現率の推移調査. 20
『自然言語処理』12(4) pp. 97-116

要件2-3 語の使用頻度の経年変化②

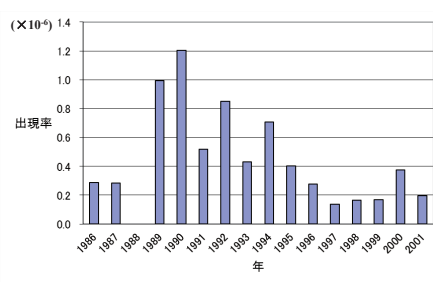
例1：シェア 平行型



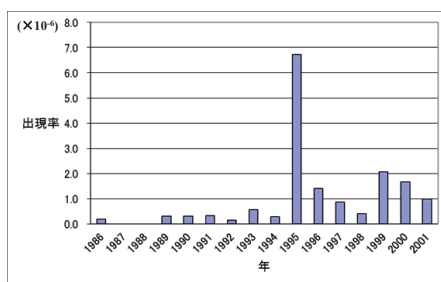
例2：バリアフリー 増加型



例3：フレックスタイム 減少型



例4：ライフライン 凹凸型



21

分類型の安定性の確認①

外来語156語について、新聞二紙間の頻度推移4分類の型の安定性を調査
→ 型分類で67.3%が一致、相関あり(相関関数0.5以上)も同じく67.3%

- 読売新聞、毎日新聞の1991年から2004年までの14年分の記事データを使用。
- 156語について、両紙の頻度間の相関を調査。平均0.57。0.5以上が105語(67.3%)。0.5以上を相関あり、とすれば、全体の67.3%に相当する語が相関あり。

表5.20 頻度推移類型の分類結果

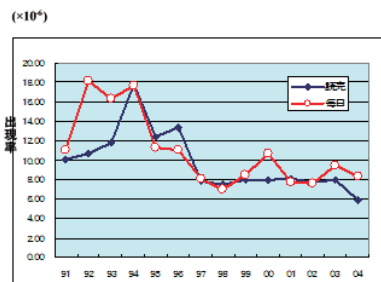
分類	両紙一致の語数		両紙不一致の語数	
	読売・毎日		読売	毎日
平行型	9		1	2
増加型	20		9	7
減少型	56		25	16
凹凸型	20		16	26
合計(語)	105		51	51

67.3%

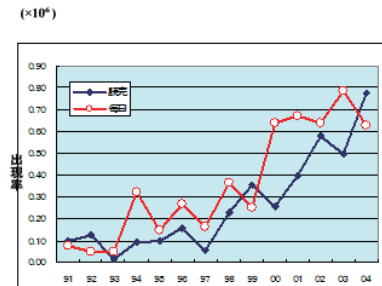
22

分類型の安定性の確認②

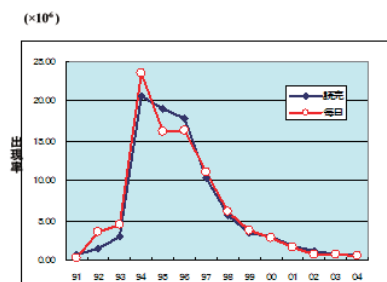
例1：ビジョン 平行型



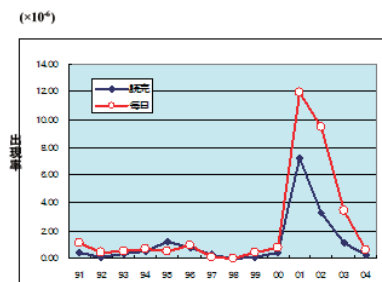
例2：スキル 増加型



例3：マルチメディア 減少型



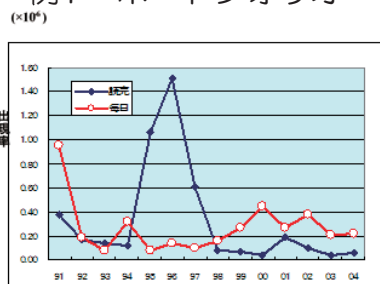
例4：セーフガード 凹凸型



23

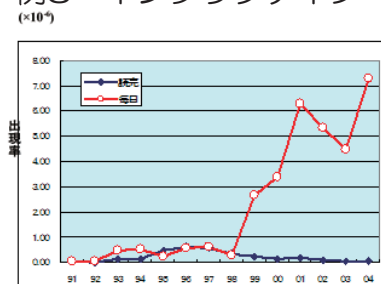
分類型の安定性の確認③

例7：ポートフォリオ 不一致



「ニッセイ投資顧問チーフポートフォリオマネージャー」の連載コラム

例8：インタラクティブ 不一致



毎日新聞の新しいホームページ「毎日インタラクティブ」

24

本論文の到達点

● 辞書記述の骨格における客観化・明示化の提案

- ・ コーパスの頻度情報に基づく見出し語の選定方法とその有効性を示した。
- ・ 名詞に着目し、名詞を下位区分する際の客観的・明示的な区分方法とその有効性を示した。
- ・ 名詞の多義語に着目し、その意味関係の記述方法とその有効性を示した。

● 辞書に記述すべき新規な情報の提案

- ・ 用例を網羅的に収集した場合に生じるコロケーションの多義性の問題に着目し、それを解決するための記述方法とその有効性を示した。
- ・ 「古風な語」に着目し、使用域・位相情報の記述方法とその有効性を示した。
- ・ 「外来語」に着目し、使用頻度の経年変化に基づく辞書記述方法とその有効性を示した。

期待される効果

- ・ 多義性解消の精度向上，語選択・生成処理の精度向上。
- ・ 多義の関係，豊富なコロケーション，文体・位相情報による人の参照性の向上。

25

残された課題

● 辞書記述の精緻化に関する課題

- － 低頻度語の扱い
- － コーパスの規模と見出し語・用例の採集効果の関係の統計的な解析
- － コロケーションの結び付きの強度の明示化
- － 文体・位相情報の体系的な記述方法の提案
- － 意味・用法の自動分類

● 大規模辞書の構築

- － 数万語以上の規模の実用的辞書の構築
 - ・ 本論文では、古風な語，外来語などといった限られた対象について議論した。これを足掛かりとし辞書を構築する。

26

国語辞典・シソーラス

27

技と工夫 その1

- 似たものを使って説明する。

パセリ

畑に作る二年草。葉はニンジンの葉のように細かく切れている。
かおりが強く、洋食のつまにする。オランダゼリ。

(新明解国語辞典)



28

技と工夫 その2

- 意味、用法を似たもの、近いものをまとめて説明する。

VS

- 意味、用法を細く分けて並べて説明する。

とる【取る・採る・執る・捕る】：

『岩波国語辞典』 8個

(①～⑧の下に、小項目が3個、8個、3個、2個、3個、3個、1個、1個、計24個あり)

『明鏡国語辞典』 1～59まで、59個

学習用では・・・

『新レインボー小学国語辞典』 1～14まで、14個

29

技と工夫 その3

- 意味を主観的に（自分の感じ方や考え方で）説明する。

VS

- 意味を客観的に（ものごとの見え方や考え方で）説明する。

こい【恋】

特定の異性に深い愛情を抱き、その存在が身近に感じられるときは、他のすべてを犠牲にしても惜しくないほどの満足感・充足感に酔って心が高揚する一方、破局を恐れての不安と焦躁(シヨウソウ)に駆られる心的状態。「一に落ちる／一は盲目／一は思案の外(ホカ) [=恋する人の気持や行動は常識や理性では割り切れないものだ] / 老いらくの一」 (新明解国語辞典)

こい【恋】

異性に愛情を寄せること、その心。恋愛。「一は盲目」(恋をすると無分別になる)▽本来は、(異性に限らず)その対象にどうしようもないほど引きつけられ、しかも、満たされず苦しくつらい気持を言う。「一に焼けて死ぬ虫になったって、思いはとげて見せるぞ」のような用例から見て取れるとおり、心の活動(やその内容)を言う「思い」とは区別される。「一は楽し、野辺の花よ」のような言い方は、1910年代ごろからのもの。→こう(恋) (岩波国語辞典)

30

技と工夫 その4

- 言葉を記述的に（あるがまま）説明する。
- 新語・新用法を積極的にのせる。

VS

- 言葉を規範（きはん）的に（お手本として）説明する。
- 新語・新用法は慎重にのせる。

さわり【触り】

④曲などの出だしの部分

〔参考〕④は最近の用法。

（新選国語辞典）

『三省堂国語辞典』は改版時に新語を中心に約4,000語が増補されている。

六版（2008年）「イケメン、がつつり、逆ギレ、激～、～じゃないですか、どん引き、パニックる、真逆、むずい、やさぐれる、ロン毛」など。

七版（2014年）「いらっと、上から目線、オソロ、空気感、恋バナ、ニーハイ、まるっと、スイッチが入る、パンダ目、ほうれい線、変顔」など。

31

技と工夫 その5

- 類義語だけで説明しないようにする。

つきあう【付き合う】

交際する。

こうさい【交際】

つきあうこと。

（多くの辞典）

つきあう【付き合う】

①よその人となかよく話したり、何かをいっしょにしたり、助け合ったりして、したい関係をつくる。

②恋人として交際する。

（以下、略）

こうさい【交際】

〔人間どうしの〕つきあい。つきあうこと。まじわり。

（三省堂国語辞典）

32

技と工夫 その6

- 語釈にあう用例を入れる。

岩国《進む》

(1) (I) 仕事はかどる。	「工事が着着とー (=完成に向かう)」
(3) (ウ) 勢いが増す。 ⇔衰える。	「食がー」 (食欲が盛んになる) 「病勢がー」 (病気が悪化する)

明鏡《進む》

③物事が（計画した通りに）進行する。	「仕事が予定通りにー」 「工事がー」 「筆がー」 「大いに食がー」
④物事の状態・程度がはなはだしくなる。進行する。	「過疎化〔相互理解・世代交代・活字離れ〕がー」 「病状〔近視〕がー」

33

技と工夫 その7

- 見出し語を一つにする、あるいは分ける。

岩波国語辞典 第七版新版 岩波書店

こ-える【**越える**】

- ①手前から向こうに、その上を過ぎて移る。「山をー」「国境をー」「2メートルのバーをー」
- ②ある時期、特に年を過ぎてその後になる。「暑い夏をー」「ー・えて天明9年の正月」
- ③ある基準より上になる。「分ぶんをー・えてしゃしゃり出る」
- ㊦...より大きな値になる。「100人をー」「30度をー暑さ」「予想をー」
- ④ぬきんでて、まさる。「人にー」（人よりまさる）。超越する。「現代をー」「党派をー・えて大局を見る」㊦㊧は「**超える**」とも書く。
- ㊨順序を追わず、先に出る。飛び越す。「兄をー・えて弟が家をつぐ」

34

技と工夫 その7

辞林21 三省堂

こ・える【**越える**】

(1)障害物や境界線の上を通り過ぎて、その向こう側へ行く。越す。
「峠を－・える」「－・え難い壁」(2)ある日時が過ぎる。「－・えて1990年の1月」

こ・える【**超える**】

(1)ある基準・数値を上まわる。「4万人を－・える大観衆」(2)自分の考え方や立場をすてて先へ進む。超越する。「怨讐を－・えた援助」
(3)他よりもすぐれる。まさる。「力量が人に－・える」

※大辞林（三省堂），広辞苑（岩波書店）は、

こ・える【**越える・超える**】

35

国語辞典の情報には多くの技・工夫がある

- 見出し語
- 表記
- 品詞
- 意味
- 用法
- 用例
- その他



どれをとっても，辞典間でさまざまな考え方や立場の違いによって書き方に差がある。

国語辞典それぞれに，それぞれの，技や工夫が多くある！！



36

たくさんの国語辞典



小型 現代生活に必要な語、使用頻度の高い語を収録。約6～9万語。『岩波国語辞典』（岩波書店），『三省堂国語辞典』（三省堂），『新選国語辞典』（小学館），『新明解国語辞典』（三省堂），『明鏡国語辞典』（大修館書店）など



中型 現代語と百科項目を収録。約20～25万語。『広辞苑』（岩波書店），『大辞林』（三省堂），『大辞泉』（小学館）など



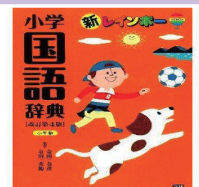
大型 国語項目と百科項目を収録。約50万語。『日本国語大辞典』（全13巻，小学館）。

37

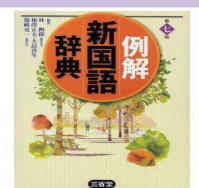
学習用国語辞典



ほとんどがひらがな。小学校入学前から小学校低学年向け。440～1,200語。『三省堂こどもこくごじてん』（三省堂），『こどもにほんごじてん』（講談社）など



漢字全てに読み仮名。小学校低学年から。2.5～3.7万語。『例解学習国語辞典』（小学館），『新レインボー小学国語辞典』（学研），『下村式小学国語学習辞書』（偕成社），『小学新国語辞典』（光村教育図書）『チャレンジ小学国語辞典』（ベネッセ），『くもんの学習 小学国語辞典』（くもん出版），『例解小学国語辞典』（三省堂），『旺文社小学国語新辞書』（旺文社）など



中学生から。4.5～5.7万語。『例解新国語辞典』（三省堂），『学研現代標準国語辞典』（学研），『旺文社標準国語辞典』（旺文社），『ベネッセ新修国語辞典』（ベネッセ），など

38

シソーラス・類語辞典とは・・・



「シソーラス」
「類語辞典」

語句が五十音順ではなく、
意味によって分類，配列
され，似た言葉どうしが
近くに並べられているもの。

39

どんなふうにご利用できるか

言語研究：

ことばの分布
作品の用語調査
ことばの結びつき



表現辞典：

適切な言葉

情報処理：

情報検索
類似度の計算

40

シソーラス・類語辞典の目的と体裁

目的 ①適切な言葉探し
②語彙の分布や偏りを見る
③使い分けを知る

シソーラス

類語辞典

体裁 A) 語釈や用例などがあるか
B) 分類体系は品詞別であるか
C) どのような語がどれくらいの数，収録対象か

41

『分類語彙表』とは

意味の世界を分類したもの



『分類語彙表』の「まえがき」より：

ここに分類語彙表というのは、一般に一つの言語体系の中で、その語彙を構成する一つ一つの単語が、それぞれどのような意味で用いられるかを一覧できるように、単語が表わし得る意味の世界を分類して、その分類の各項にそれぞれの単語を配当したものである。

42

『分類語彙表』の公開



- 国立国語研究所資料集6 『分類語彙表』 秀英出版 1964年
初版刊行（1997年 31版）
- 『分類語彙表 フロッピー版』 秀英出版 1994年 刊行
約3万7千語
- 『分類語彙表 増補改訂版』 大日本図書 2004年 報告書刊行,
電子化データベースの公開
約9万6千語

いずれも研究利用に無償で公開されている。

語彙調査データ

http://pj.ninjal.ac.jp/corpus_center/archive.html

43

『分類語彙表』の担う役割



- ①言葉や概念を手がかりに、適切な言葉を見つけるもの
- ②語彙の分布や偏りを見るための「物差し」となるもの

44

『分類語彙表』の紹介記事



日経新聞 プラス1
(朝刊別刷り)
2001年12月8日(土)

45

主なシソーラス・類語辞典①

(1) 『角川類語新辞典』 (1981)

- ・ 収録語数：約60,000語
- ・ 分類の特徴：大分類(10), 中分類(100), 小分類(1,000), 細分類(約3,000)の4階層。品詞の区別なし。大分類は, 「自然」「人事」「文化」という枠の中で設けられたもの。
- ・ その他特徴：語釈, 用例あり。位相が付与されている。

(2) 『日本語語彙体系』 (1997)

- ・ 収録語数：約300,000語
- ・ 分類の特徴：1階層目が品詞。2階層～12階層目に2,876個の意味属性を木構造で配置。
- ・ その他特徴：さらに6000語の用言には日英の対訳文型14,000パターンを付す。機械翻訳を主目的にして開発されたため, 固有名詞も多く収録されている。

46

主なシソーラス・類語辞典②

(3) 『類語大辞典』(2002)

- 収録語数：約79,000語
- 分類の特徴：カテゴリー(100), 小分類(916), 品詞(11), 小見出しの4階層。カテゴリーは用言に基づいて作成し, 人に近いところから遠いところへという原則にしたがって配置。
- その他特徴：語釈, 用例あり。位相が付与されている。

(4) 『日本語大シソーラス』(2003)

- 収録語数：延べ約320,000語, 異なり約200,000語
- 分類の特徴：カテゴリー(1,044), 小語群(14,000), セミコロン(任意)の3段階。品詞の区別なし。カテゴリー部分が先に2～4階層に分類されているため, あわせると一番深い場合で6階層。語釈なし。文法的カテゴリーなどの一部に用例あり。
- その他特徴：故事成句, 人名やオノマトペが多く収録されている。

47

主なシソーラス・類語辞典③

(5) 『類語例解辞典』新装版(2003)

- 収録語数：約25,000語
- 分類の特徴：大分類(10), 中分類(200), グループ(約6,000)の3段階。助詞・助動詞は別枠で78に下位分類。品詞の区別なし。
- その他特徴：語釈, 用例のほか, 使い分けの解説, 対比表などがある。

(6) 『三省堂類語新辞典』(2005)

- 収録語数：約50,000語
- 分類の特徴：柱(3), ジャンル(18), 分野, 領域の3段階。品詞の区別なし
- その他特徴：語釈, 用例のほか, 位相, 使い分けの解説, 図解などがある。また, 類語のニュアンスを解説するコラムが86ある。オノマトペも積極的に収録されている。

48

比較一分類体系は品詞別であるか

品詞別である	『分類語彙表』 『類語大辞典』 cf. 『ロジェのシソーラス』
品詞別ではない	『日本語大シソーラス』 『角川類語新辞典』 『三省堂類語新辞典』

<品詞別の利点>

- ・ 同じ品詞どうして、語の比較がしやすい
- ・ 品詞別にどのような意味分布があるかの見通しが良い

49

比較一人名はどう分類されるか

『分類語彙表』	<1.2390 人名> 01 佐藤 鈴木 田中 高橋 山田 スミス ジョンソン 09 紫式部 清少納言 鴨長明 滝沢馬琴 夏目漱石 森鷗外 芥川龍之介 17 浦島太郎 シンデレラ ピノキオ ロミオ ジュリエット
『日本語大シソーラス』	「シンデレラ」 <架空人物> <美女> <なさぬ中> (継子関係) <いじめる>

「人名は人物典型を表現する際の最初の手掛かりになるものである」 (『日本語大シソーラス』序)

50

比較—属性分類か主題分類か①

属性分類 語が本来もつ性質によって分類することを優先するもの

(例)『分類語彙表』 <1.5 自然現象> 「**発病**」
<1.3 人間活動> 「**治療**」

主題分類 ある主題のもとに関連する語を集めて分類することを優先するもの

(例)『角川類語新辞典』 073「**発病**」> C「**治療**」

51

コーパス

52

コーパスとは何か

目的

言語の実態を分析するための重要なデータとして、言語学、言語教育、辞書編集、言語情報処理などの分野で利用されている。

歴史

Brown Corpus(1964年、米、100万語)が最初。その後、British National Corpus(1994年、英、1億語)など、英語のコーパスを中心に発展してきた。

言語

英語、韓国語、中国語、スペイン語、ドイツ語、フランス語など、さまざまな言語でコーパスが作られている。

種類

書き言葉コーパス、話し言葉コーパス、通時コーパス、学習者コーパス、対訳コーパスなど、目的により多様なコーパスが作られている。

規模

書き言葉では、1億語以上のものが多い。話し言葉では、British National Corpusの一部(約1000万語)や『日本語話し言葉コーパス』の約700万語が最大級。

構築

コーパスの構築には、多額の費用が必要なため、大規模なコーパスのプロジェクトには国費が投入されている。

53

『現代日本語書き言葉均衡コーパス』

- 「コーパス」とは、言葉の実際の使われ方をしらべるために作られた、言語研究のための資料。体系的に言語の資料を収集し、研究用の情報（たとえば、品詞情報）を付加したもの。
- 目的： 現代日本語の書き言葉を代表するコーパス
- 対象： 1976年～2005年に出版された書き言葉
 - － 書籍，雑誌，新聞，白書，教科書，Web文書…
 - － 全部で1億語（文庫本1700冊分相当）
- 公開： Web上でオンライン検索『少納言』にて公開中

書き言葉の総体



コーパス

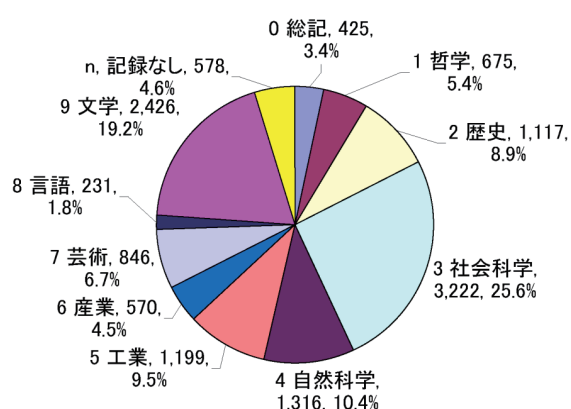
54

『現代日本語書き言葉均衡コーパス』の構成

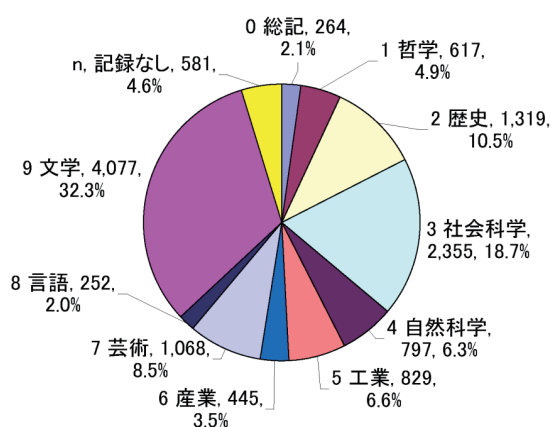
現代日本語書き言葉均衡コーパス

出版 サブコーパス	図書館 サブコーパス	特定目的 サブコーパス
書籍・雑誌・新聞	書籍	白書・教科書・ 広報紙・ ベストセラー・ Yahoo!知恵袋・ Yahoo!ブログ・ 韻文・法律・ 国会会議録
2001-2005年 約3,500万語	1986-2005年 約3,000万語	約3,500万語

書籍 出版・図書館サブコーパス



出版サブコーパスにおける
取得サンプル数のNDC別内訳
(サンプル数およびその割合
(%))



図書館サブコーパスにおける
取得サンプル数のNDC別内訳
(サンプル数およびその割合
(%))

実際と同じ配分になるように設計する

現代語の中の「さもありなん」

まさか和久さんが指導員として復帰してるなんて
思わなかったから。でも、**さもありなん**、という気もする。

君塚良一|脚本；丹後達臣|ノベライズ『踊る大捜査線スペシャル』フジテレビ出版；扶桑社
(発売)，1998

美也子たちの離婚を、夫の実家では、**さもありなん**、といった諦めの心境
で受けとめてくれた。

藤堂志津子|著『恋人よ』講談社，2003

「でもさ、霊ってお墓に住んでいるわけじゃないと思うんだけど…」あま
りにも「**さもありなん**」なオチだったので私は呆れて言った。

田口ランディ|著『ほつれとむすぼれ』角川書店，2004

「第十二深宇宙探検隊の隊長の秋篠氏と君は同じ苗字だけど、親戚な
の？」教室中に緊張が走った。**さもありなん**。さっき同じ質問をしたルイ
スは完璧に打ちのめされたんだ。

榎木洋子|著『Step out』集英社，1996

57

現代語の中の「なきにしもあらず」

また、さざ波のように押しよせる技術革新と対峙してきたブルーカラーに
してみれば、OA化に浮足だつホワイトカラーの姿は、いまさらの感も**無
きにしもあらず**で、滑稽にすら映るかもしれない。

加藤仁|著『日本の課長』PHP研究所，1988

その声には、質問せずにいられないだけでなく、思いあたる節が**なきにし
もあらず**という気持ちもあらわれていた。

クーパー|作；村山淳彦|訳『開拓者たち下』岩波書店，2002

「そのデータが中尾弥生に渡っていた可能性も**なきにしもあらず**ね」

「そうですね」

志茂田景樹|著『美雪総監と消えたクルーザー』光文社，1993

でないと、“ピアノ離れ”がドーンと出てしまった『子供のピアノ教室』
の二の舞になる恐れが**なきにしもあらず**だわ。

大村典子，大崎妙子|著『大人のピアノ長続きのコツ』ヤマハミュージックメディア，1997

58

現代語の中の「べし」

男は黙って「魂」を聴くべし。

鈴木敦史 | 著『クラシック批評こてんぱん』洋泉社, 2001

もしあなたが深夜に鍾路界限にいて、空腹を満たしたくなったらこのエリアに直行すべし。

松井聖一郎 | 著『韓国 長期滞在者のための最新情報55』三修社, 1996

◆賢い消費行動のための基本原則は「資産と負債の総合管理」にあり。

◆資産が増えないのであれば、負債を縮小させるべし。

小学館文庫編集部 | 編『デフレ時代を生きる知恵』小学館, 2001

これはこれでいいクルマ。いやはやイタリア恐るべし

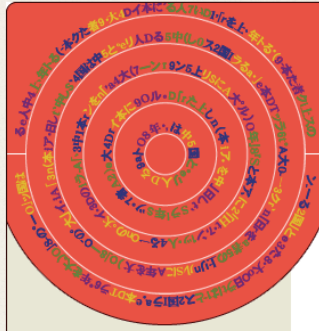
福野礼一郎 | 著『礼一郎式外車批評』双葉社, 2003

そんな尾崎さんの描く EME が、数量限定でクリアファイルと下敷きになりました。ゲットすべし。

瀧川武司 | 著『EME blue』富士見書房, 2002

59

コーパス：公開①



少納言

KOTONOHA「現代日本語書き言葉均衡コーパス」

検索にあたっての注意点

検索文字列の制約：
本サイトで検索できる文字列は最長で10文字までです。

検索方式：
本サイトでは全文検索をおこなっていますので、指定された文字列を含むすべてのテキストが検索されます。例えば検索文字列に「リズム」を指定すると、「リズム」だけでなく、「アルゴリズム」や「フォルマリズム」などが検索され、「国語」を検索

このサイトでは大学共同利用機関法人人間文化研究機構国立国語研究所と文部科学省科学研究費特定領域研究「日本語コーパス」プロジェクトが共同で開発した『現代日本語書き言葉均衡コーパス』（BCCWJ: Balanced Corpus of Contemporary Written Japanese）のデータを検索できます。BCCWJには、現代の日本語の書き言葉の全体像を把握できるように集められたサンプルが約1億語収録されています。

本サイトで可能な検索は文字列検索（全文検索）ですが、形態論情報を利用した検索サイト「中納言」も開設しています。詳しくは以下の URL をご覧下さい。なお、中納言の利用には申請が必要です。

<http://chunagon.ninjal.ac.jp/>

コーパスの構築にあたっては、全サンプルについて著作権処理を実施しており、その結果、権利者の方から利用を拒否されたサンプルは公開対象から除外してあります。

<http://www.kotonoha.gr.jp/shonagon/>

60

コーパス：公開②

NINJAL-LWP For BCCWJ (NLB)

国語研ホームページ

検索を開始する 操作説明書を開く

トップ NLBとは 使い方 2語比較機能 使用コーパス ご利用にあたって 更新履歴 お問い合わせ

■ 現代語自他対一覧表を公開
2015/11/25

『有対動詞の通言語的研究—日本語と諸言語の対照から見えてくるもの—』（くろしお出版）の付録『現代語自他対一覧表』のExcel版を公開しました。書籍版では収録できなかったローマ字表記、英訳、活用などのデータが含まれています。使役交替言語地図のサイトからダウンロードできます。頻度の計数には、BCCWJとNLBを利用しています。

■ 基本動詞ハンドブック、新たに8見出しを公開
2015/11/24

基本動詞ハンドブックのサイトから、新たに8見出し（合う、打つ、返す、切る、切れる、漬す、漬れる、抜く）が公開されました。ハンドブックの執筆・編集には、NLBやNLTが活用されています。

■ 筑波大学からNLT1.30版を公開
2015/3/11

お知らせ 公開先が以下に変更されました。
<http://nlt.tsukuba.legoinst.info>

■ NLB1.30版を公開
2014/8/25

<http://nlb.ninjal.ac.jp/>

61

名詞＋助詞

パターン	頻度	比率
...が現れる	5,537	
...は現れる	808	
...も現れる	355	
...の現れる	80	
...を現れる	21	
...に現れる	4,582	
...へ現れる	123	
...で現れる	586	
...と現れる	62	
...から現れる	437	
...まで現れる	60	
...より現れる	23	

...が現れる 1729種類

コロケーション	頻度	MI	LD
【人名】が現れる	265	2.03	2.75
効果が現れる	164	7.48	7.83
症状が現れる	160	9.05	8.84
人が現れる	149	3.25	3.96
者が現れる	111	3.43	4.12
【一般】が現れる	104	1.56	2.28
それが現れる	86	3.27	3.96
ものが現れる	82	2.78	3.48
男が現れる	75	5.25	5.79
変化が現れる	74	6.75	6.99
姿が現れる	67	6.04	6.42
たちが現れる	58	3.46	4.12
影響が現れる	56	6.19	6.47
画面が現れる	48	7.25	7.06
顔が現れる	47	4.65	5.18
さが現れる	46	3.58	4.22
ことが現れる	37	0.06	0.78
傾向が現れる	37	6.44	6.44
女が現れる	34	4.51	5.00
さんが現れる	32	2.47	3.14
差が現れる	32	6.26	6.25
成果が現れる	31	7.03	6.63
女性が現れる	30	4.24	4.74
気持ちが見える	29	4.78	5.17
現象が見える	28	6.66	6.38

変化が見える 全74件

■ 外見にも変化が見えます。
(『40代から、もっときれいになる本』, 2004,)

■ 彼女の唇にも変化があらわれた。
(西田道世著『Love History』, 2000, 913)

■ お子さんに何か変化が見えましたか？
(Yahoo!知恵袋, 2005, 知恵袋, 知恵袋, 知恵袋)

■ 大正時代末、仕込み容器に変化が表れます。
(佐藤伸雄著『道と路のはなし』, 2005, 588)

■ 変化—その結果、どういった変化があらわれるのか
(杉本修三著『記憶力30秒増強術』, 2002, 141)

■ 4 人間関係に変化があらわれることを知っておく
(マリアンヌ・レマクマナス著・林寿男、林由美訳・訳『災害ストレス』, 1995, 369)

■ 突然、大型スクリーンの画面に変化があらわれた。
(ジェイムズ・H.コップ著・白幡雄之訳『ステルス艦隊コンカム出版』, 1999, 933)

■ その結果、当然人格にも変化があらわれるであろう。
(アルビン・トフラー著・鈴木健太郎訳『第三の波』, 1980, 304)

■ 五歳半くらいから、速くんに一寸した変化が見えた。
(小久江伸子著『きんぐ』, 2005, 949)

■ もちろん早ければ6カ月くらいで変化が見えてきます。
(竹田隆一著『小さな会社がいかに儲かるか』, 2004, 673)

■ 輝夜の瞳には三年の間に著しい変化があらわれていた。
(高橋道雄著『龍の伝説』, 1994, 913)

■ 一行がじっと見つめるうちに、明確な変化があらわれた。
(ロバート・J.ソウワ著・内田昌之訳『古聖賢アサンの道徳論』, 1994, 933)

■ 昌幸の声にも、多少の変化があらわれてきたようであった。
(池田正太郎著『真田太平記』, 1984, 913)

■ 女性が独立して仕事をするとどんな体格の変化が表れますか？
(Yahoo!知恵袋, 2005, コスメ、美容)

■ さあ、再考にその後、どんな体の変化が見えたのでしょうか？
(小沢たけし著『たけし』, 2002, 498)

■ 昌幸の声にも、多少の変化があらわれてきたようであった。
(池田正太郎著『真田太平記』, 2005, 913)

■ 堤防おへて、社を頼りにウキウキの平にも、変化が見れる。
(Yahoo!ブログ, 2005, 乗り物)

■ しかし、暗黒を重ねるうちに、戸田の歌に大きな変化が表れる。
(池田大作著『新・人間革命』, 1999, 913)

■ このような潮の流れの変化が53年度中に現れたとみられる。
(経済白書, 1979,)

■ 吟詠の登場により、注ぐ縁づくりに変化が見えてきました。
(佐藤伸雄著『道と路のはなし』, 2005, 588)

■ ではこの年に至って何故このような変化があらわれたのだろうか。
(谷間隆夫著『近世史』, 1988, 791)

■ 一度うまくいけば、あとは弾みがついて次々に変化が表れるはずだ。
(堀江利典著『リー・モーション・デザイン』, 2005, 913)

『LWP』の2語比較

NINJAL-LWP for BCCWJ

ここに調べたい語を入力（読み可）

絞り込み 元に戻す

2語比較検索

NINJAL-LWP for BCCWJ

問題 課題

絞り込み 元に戻す

2語比較

名詞	動詞	形容詞・連体詞	副詞	オノマトペ
問題	見出し	読み	ローマ字表記	頻度
問題	モンダイ	mondai	68,064	
課題	カダイ	kadai	8,035	

63

動詞+動詞

問題	頻度	課題	頻度
問題が...	7738	課題が...	447
問題に...	7552	課題に...	771
問題を...	7268	課題を...	1063
問題と...	3189	課題と...	1378
問題は...	2338	課題は...	130
問題も...	1379	課題も...	79
問題で...	1018	課題で...	51
問題から...	203	課題から...	11
問題の...	185	課題の...	13
問題まで...	45		
問題へ...	43	課題へ...	3
問題より...	25	課題より...	1

問題を... 749 課題を... 247

問題	頻度	MI	LD	課題	頻度	MI	LD	LD差
問題を引き起こす	111	9.6	8.59					8.59
問題を起こす	163	8.32	8.48					8.48
問題をほらわ	49	10.39	7.68					7.68
問題を生じる	57	6.47	6.79					6.79
問題を生み出す	30	7.17	6.56					6.56
問題を回避する	20	8.59	6.35					6.35
問題を討議する	18	10.71	6.31					6.31
問題を先送りする	15	10.83	6.06					6.06
問題を持ち出す	16	7.58	5.94					5.94
問題を重視する	18	6.66	5.89					5.89
問題を生ずる	15	7.24	5.81					5.81
問題を追及する	13	8.38	5.76					5.76
問題を内包する	12	9.79	5.72					5.72
問題を説明する	12	8.16	5.64					5.64
問題を担当する	14	6.59	5.6					5.6
問題を含める	79	7.13	7.36	課題を含める	1	3.6	1.78	5.58
問題を扱う	95	8.23	8.02	課題を扱う	1	4.43	2.5	5.52

問題を... 749 課題を... 247

問題	頻度	MI	LD	課題	頻度	MI	LD	LD差
課題を遂行する	13	10.79	7.87					-7.87
課題を採択する	6	9.67	6.76					-6.76
課題を公算する	3	11.46	6.39					-6.39
課題をやり遂げる	3	10.14	6.2					-6.2
課題を達成する	7	8.26	6.07					-6.07
課題を取り扱う	2	9.87	5.68					-5.68
課題を握り起こす	2	9.85	5.67					-5.67
課題を実行する	5	7.4	5.32					-5.32
課題を内含する	1	14.06	4.94					-4.94
課題を担わず	1	12	4.91					-4.91
課題を感得する	1	10.74	4.87					-4.87
課題を羅列する	1	10.54	4.86					-4.86
課題を深化する	1	10.36	4.84					-4.84
課題をレポートする	1	10.05	4.82					-4.82
課題を提出する	1	9.53	4.77					-4.77
課題を上乗せする	1	9.27	4.74					-4.74

コーパス：公開③



<https://chunagon.ninjal.ac.jp/> (申請が必要)

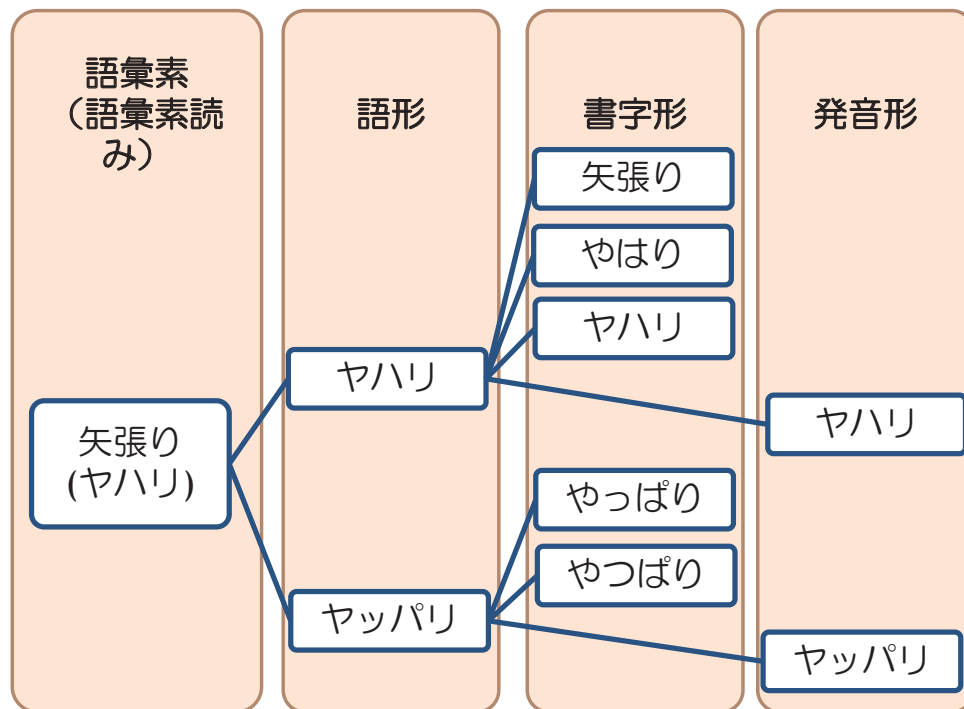
65

中納言の特徴

- 短単位と長単位とで検索できる
- 文字列検索だけではなく、見出し語検索ができる
- 品詞を指定した検索ができる
- 前にくるもの、後ろにくるもの、といった検索ができる

66

「矢張り(ヤハリ)」の階層構造



※「ヤッパ」「ヤッパシ」「ヤパ」「ヤパリ」「ヤーッパリ」もあり。67

現代日本語書き言葉均衡コーパス (NumTrans 版) BCCWJ-OT

短単位検索 長単位検索 文字列検索 位置検索

短単位検索

前方共起条件の追加

キー: 品詞 の 大分類 が 形容詞

後方共起: キーから 1 語

キーと結合して表示

書字形出現形 が 言葉

後方共起条件の追加

1,311 件の検索結果が見つかりました。そのうち 500 件を表示しています。
検索対象語数: 124,624,010 記号・補助記号・空白を除いた検索対象語数: 105,377,883

サンプル ID	開始位置	語	前文脈	キー	後文脈	語彙素読み	語彙素	語彙素分類	品詞	活用型	活用形	レジスター	執筆者	書名・出典	編纂者等	出版者	出版年
PB59_00435	38580	25910	晴雨、昼夜など。#おに地理がある。#遠近、陸奥、広域—もっ	生々しい	(音真)でいえば、死地と生地だ。#これらのものは、人力でもって	ナマナマシイ	生々しい		形容詞一般	形容詞一般	連体形一般	出版・書籍	伴野 朗(著)	佐子蘭	伴野朗著	徳間書店	2005
PM41_01283	87920	55180	一段下の階級であるわけだから、1対1リンクとしては編纂国家を	強い	(音真)を後ろ。#コアの標準漢語もどうだった。#彼は「コーポレーション」を攻撃	ツヨイ	強い		形容詞一般	形容詞一般	連体形一般	出版・雑誌	有賀 真紀(著)	朝日経理リポート 4R21		朝日新聞社	2004
OY04_05034	14100	9050	じゃあ、きがるさんのみに「音真」(音真)のCD/DVDをあげると	やさしい	(音真)をあげてくれました。#うひゃひゃひゃ!これからはきがるさん	ヤザシイ	やさしい		形容詞一般	形容詞一般	連体形一般	特定目的・ブログ		Yahoo!ブログ		Yahoo!	2008
PB37_00019	70000	50580	彼は、今までの人がけてて言わなかった。音真の人が聞いたたらう	ない	(音真)を口にしました。#それは「Part4」の結語の章で	ナイ	無い		形容詞非自立可能	形容詞一般	連体形一般	出版・書籍	山田 邦子(著)	邦子の「しあわせ」 音楽	山田邦子著	海電社	2003
LB01_00012	51090	32920	ているのは、例えば、小生業長が	強い	(音真)の前でウーンと知りつつ全話	スルドイ	強い		形容詞一般	形容詞一般	連体形一般	図書・書籍	大庭 秀(著)	権力とはどんな力か	大庭秀著	勁草書房	1991
PB59_00178	84940	41830	も、十年でも、百年でも待ってやろうか?」# などと	ざったらしい	(音真)を編纂する前にしていたりする。# 華はとたん滅亡になり、	アマツライ	ざったらしい		形容詞一般	形容詞一般	連体形一般	出版・書籍	みさき 志穂(著)	眠がしてみやがれ!!	みさき志穂(著)	ファンタム出版(ラジス書院(発売))	2005
OB3X_00225	34800	21570	情報をファイルしたらいい。#一冊は本人の手記で—お世辞ならた	いい	(音真)が並んでる—ほかには新聞の切り抜きがある。#ヤンキー—ター・ブートリス	ヨイ	良い		形容詞非自立可能	形容詞一般	連体形一般	特定目的・ベストセラー	フレデリック・フォーサイス(著) 篠原 慎(訳)	ネゴシエーター	フレデリック・フォーサイス(著) 篠原慎(訳)	角川書店	1989
LB49_00170	52930	34700	わたしはかかって言った。# 2年	しかつめらしい	(音真)を、無理して使わなくていいの。#なんと言おうと、ばばあは、	シカツメラシイ	しかつめらしい		形容詞一般	形容詞一般	連体形一般	図書・書籍	吉川 博子(著)	善徳忌	吉川博子著	実業之日本社	1990
OB4X_00133	79920	50580	んや!はしたるの中で進行した。# デビッドが顔を覗かせ、	優しい	(音真)を囁いてくれる。#いつの間にか、二人は並んでペダルに	ヤザシイ	優しい		形容詞一般	形容詞一般	連体形一般	特定目的・ベストセラー	ジドニヤ・シェルダン(著) 荒馬 竜行(訳) 岸記 希隆(訳)	私は別人	ジドニヤ・シェルダン(著) 荒馬竜行(訳) 岸記希隆(訳)	アカデミー出版(本語文庫)	1993

現代日本語書き言葉均衡コーパス(第 NumTrans 版) BCCWJ-OT

短単位検索

長単位検索

文字列検索

位置検索

短単位検索

▼ 前方一致条件の追加

① キー

1

語

■ キーの条件を指定しない

品詞

の

大分類

が

形状詞

■ 短単位の条件の追加

② 後方一致1

キーから

1

語

■ キーと結合を表示

音字形出現形

が

因

■ 短単位の条件の追加

③ 後方一致2

キーから

2

語

■ キーと結合を表示

音字形出現形

が

言葉

■ 短単位の条件の追加

▲ 後方一致条件の追加

1,120 件の検索結果が見つかりました。そのうち 500 件を表示しています。
 検索対象語数: 124,624,010 記号・補助記号・空白を除いた検索対象語数: 105,377,883

サンプル ID	国名 名称	通 番	対文版	キー	後文版	国 語 書 法 読み	国 語 書 法	国語書 法分類	品 目	活 用 型	活 用 形	レッ ス ター	執筆者	書名・出版	国語書等	出版者	出 版 年
OM21_00004	555550	348470	この二月一日の諸事録の 中で何共同研究開発という 言葉は希少な	いろいろ	(注)言葉は金でもごまか して、場所によらしてその 義が区々で	口ロ口 色々			形状詞 一般			特定の目的 ・国国会議 録		国会会議録			1982
OT02_00019	20330	13070	、外来語を転換することで 表裏を疊かにしてきた。#日 本語の	微妙	(注)言葉の使い分けを知る ことで、留学生などの言葉に 対する感覚も磨かれる	ミマウ	微妙		形状詞 一般			特定の目的 ・教科書		国語 1	宮地裕江が著	光村図書出版 株式会社	2005
OT02_00019	38860	24810	ないう。#費用物。ことわざ #「耳を振る」「毒とったま づがの	よう	(注)言葉は変わる。#日本語 の特徴#外国人によって、日 本語のどんな	ヨウ	様		形状詞 ・動詞 語幹			特定の目的 ・教科書		国語 1	宮地裕江が著	光村図書出版 株式会社	2005
OC02_00026	1330	730	で喜んでる途中で。#音 速スラングと言うのはその 仲間内でしか理解できない	特殊	(注)言葉で「ずかぢゃんね る」は毒でも「ずかに」に なりますか、#意味ない	トクシユ	特殊		形状詞 一般			特定の目的 ・知恵帳		Yahoo!知恵帳		Yahoo!	2005
OY14_13248	19480	13080	1「あ……オ、い」 「ん」何やま、いま	重要	(注)言葉はあつと受け流し たろう。1「あ」「ん」はあそ びか、	ジュウウ	重要		形状詞 一般			特定の目的 ・ブログ		Yahoo!ブログ		Yahoo!	2008
OV2X_00013	09050	6280	人の耳にあわれにひびいた #寒い地方 #暑い地方#痛 風をまわって来たその	わずかに	(注)言葉はその季節・季節の 風の変化にあわれにひびいて 消えたりした#痛風	ウスカ	僅か		形状詞 一般			特定の目的 ・論文		中野里治特集	中野里治著	思兼社	1988
OC14_10116	110	50	ありがとう。をもっと	柔軟	(注)言葉で「言えませんか」 #「君の真心、俺が受け止め ない」	ステキ	柔軟		形状詞 一般			特定の目的 ・知恵帳		Yahoo!知恵帳		Yahoo!	2005

69

現代日本語書き言葉均衡コーパス (非 NumTrans 版) BCCWJ-OT

短単位検索 長単位検索 文字列検索 位置検索

短単位検索

検索フォームで検索 検索条件式で検索 履歴で検索

1 キー 書字形出現形 1 語 キーの条件を指定しない

書字形出現形 が 短単位の条件の追加

2 接方共起1 キーから 1 語 キーと結合して表示 この条件をキーに この共起条件を削除

語彙 が 的 短単位の条件の追加

接方共起条件の追加

276,104 件の検索結果が見つかりました。そのうち 500 件を表示しています。
検索対象数: 124,624,010 記号・補助記号・空白を除いた検索対象数: 105,377,883

サンプルID	開始位置	演奏番号	前文脈	キー	後文脈	演奏楽器読み	演奏楽器	演奏楽器分類	品名	活用型	活用形	レジスター	執筆者	書名/曲名	編者等	出版者	出版年
LB17_00035	124780	79500	カウンターに立つことになる。おちうん。山は技術面の習得のりこに定まる。	アメリカ的	なサービス業のあり方をもとにつけ、オプミナルの外に出て	アメリカ力	アメリカ力	America	名詞・有名人・地名・国			図書館・書籍	東宮 清(著)	「戦車クラブから歌謡曲へ」	東宮清(著)	みすず書房	200
LB19_00175	55930	35340	いった。#「二足す二は結構!五にはならない!一度以上に	アメリカ的	癖は定めようとならぬ。私はまず手始めにそうしたい。#「二足す				未知語			図書館・書籍	マイク・ティッドウェル(著)/吉澤英美訳	アマゾン・の白い紙兵	マイク・ティッドウェル(著) 吉澤英美訳	精治社	199
LB2a_00047	85230	59490	ものはかりでない。一見ふとしたところの驚きのない施設や事業までも	イデオロギー的	意味深い!白色の減り方を試みた。光の点では、先の泉	イデオロギー	イデオロギー	Ideologie	名詞・香港名詞一般			図書館・書籍	中島 三男(著)	天皇の代官!と国民	中島三男(著)	香木書店	199
OC03_00410	490	240	あるクレジットカードは母が思い浮かぶますか?アメリカンエクスプレスかな。#中島は知ら	イメージ的	イメージ。にほ。	イメージ	イメージ	image	名詞・香港名詞・言葉可能			特定目的・知恵袋		Yahoo!知恵袋		Yahoo!	200
FB49_00275	41340	25680	きき複製を命令する。#それではみな研究力の意味において、あるいは	カント的	「言語一語」の意思にありて、あるいはイデオロギー的な「ひと」即ち	カント	カント	Kant	名詞・有名人・人名一般			出版・書籍	三木 清(著)	日本近代文学評論選	千葉俊二、坪内三編	岩波書店	200

70

71

– 136 –

73

コーパス：公開④

『国語研日本語ウェブコーパス (NWJC) 』

258億語の日本語コーパスをウェブで公開～国立国語研究所 -INTERNET Watch

<http://internet.watch.impress.co.jp/docs/news/1047913.html>

- NWJCは、2014年10～12月にインターネット上で8399万2556のURLから収集した258億3694万7421のコーパスを収録。文数では14億6314万2939、のベ文数は38億8588万9575に上るデータが格納されている。
- 梵天の一般公開版では、この258億のコーパスに対する文字列検索が行える。

「梵天」の検索結果

http://pj.ninjal.ac.jp/corpus_center/nwic/bonten-results.html

74

『梵天』

「梵天」の検索機能：文字列検索

http://pi.ninjal.ac.jp/corpus_center/nwjc/

「文字列検索」は最も手軽にNWJCのデータを調べることができる検索方法です。

検索の手順

- 1 「文字列検索」では、はじめに「文字列検索」タブをクリックする。

The screenshot shows a search interface with three tabs: '文字列検索' (Text Search), '品詞検索' (Part-of-speech Search), and '係り受け検索' (Dependency Search). The '文字列検索' tab is selected and highlighted in orange. Below the tabs, the text '文字列検索' is displayed in a larger font.

- 2 「検索文字列」の入力欄に調べたい文字列を入力する。

The screenshot shows a search form with two input fields. The first field is labeled '検索文字列' (Search Text) and is empty. The second field is labeled 'URLドメイン' (URL Domain) and contains the text '※ac.jp など末尾2パート'. Below the fields, there is a '検索' (Search) button and a link '検索条件を初期化する' (Reset search conditions).

- 3 「URLドメイン」を指定したい場合は、末尾2パート（第2レベルドメイン）を入力する。
- 4 「検索」ボタンを押す。
- 5 検索結果が「検索」ボタンの下部に表示される。
件数が50件以上ある場合の検索結果は、1～50件のみが表示されている。

<http://bonten.ninjal.ac.jp/>

3. 「ソフトウェア要求仕様書の品質 ～ IEEE Std. 830-1998 より～」

ソフトウェア要求仕様書の品質 ～IEEE Std. 830-1998より～

東京女子大学
白銀純子

平成29年度 産業日本語研究会・ワークショップ

1

講演者

- 白銀純子
 - 現職: 東京女子大学准教授
 - 専門: ソフトウェア工学
 - 主な研究テーマ: 要求工学およびユーザインタフェース(ソフトウェアの画面構成)の開発支援

平成29年度 産業日本語研究会・ワークショップ

2

本日の内容

- 要求仕様書とは
- IEEE Std. 830-1998～ソフトウェア要求仕様書～
- IEEE Std. 830-1998～品質特性～
- 要求仕様書とソフトウェアの品質の相関

平成29年度 産業日本語研究会・ワークショップ

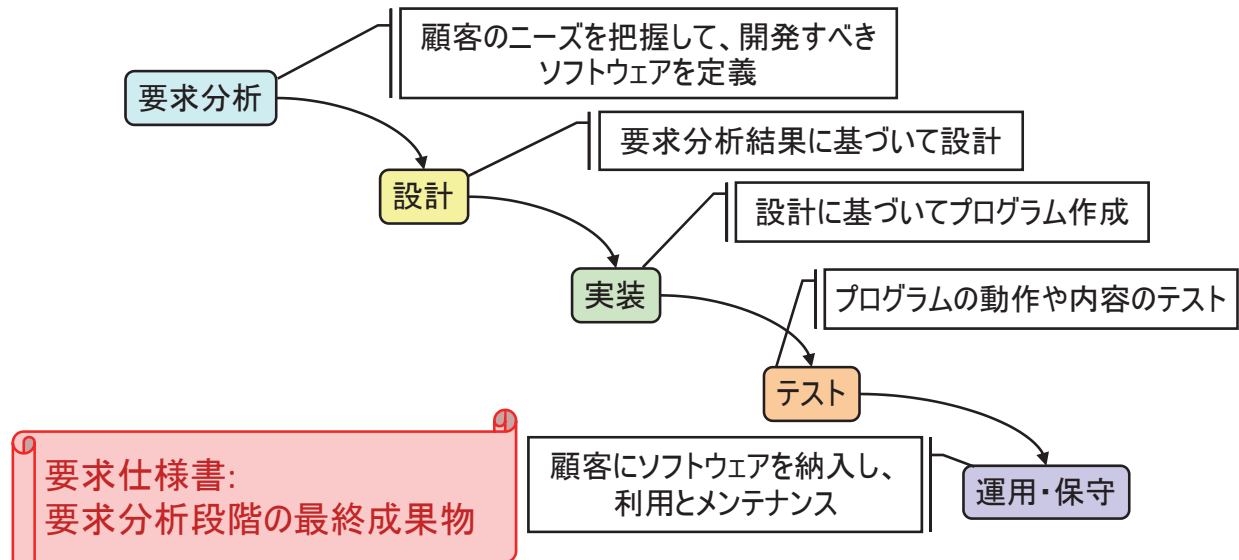
3

要求仕様書とは

平成29年度 産業日本語研究会・ワークショップ

4

ソフトウェア開発の流れ



平成29年度 産業日本語研究会・ワークショップ

5

要求仕様書とは?

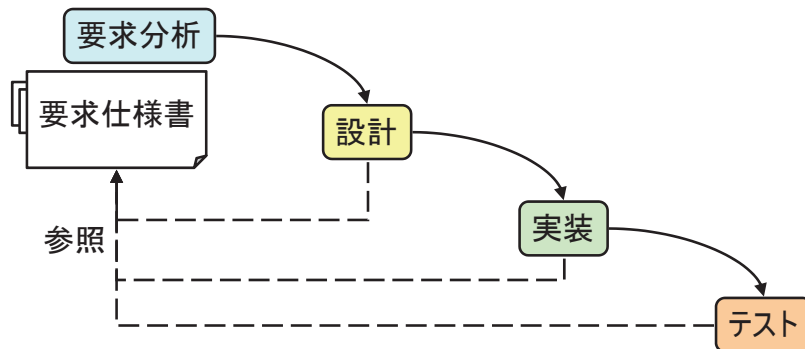
- どのようなソフトウェアを開発すべきかを文書化したもの
 - ソフトウェアに対して求めること: 要求
 - ソフトウェアが持つべき機能
 - ソフトウェアが持つべき特性
 - 開発時に考慮すべき条件, etc.
- 要求分析段階での最終成果物
 - 顧客からの聞き取りや業務の分析、資料の調査など、様々な作業の後、結果をまとめて作成
- 開発するソフトウェアに対する、顧客側と開発側との合意形成に利用
 - 顧客と開発者との一種の契約書のようなもの

平成29年度 産業日本語研究会・ワークショップ

6

要求仕様書の役割

- 設計、実装、テストといった後の開発の段階で常に参照される
 - どんな機能を実現すれば良いか?
 - どんな制約があるか?
 - ソフトウェアは要求通りに作成されているか?, etc.



平成29年度 産業日本語研究会・ワークショップ

7

要求仕様書の重要性

- 要求仕様書に不備があると...
 - 必要な機能や入出力内容
 - 必要な制約
 - ソフトウェアにとって重要な属性 (セキュリティ, 保守性, etc.)

抜けていたり、不明確だと..?

↓

 - 顧客に確認 → 手戻りの発生による時間のロス
 - 開発側で判断して開発続行
 - 顧客のニーズとは合わないソフトウェアができあがる可能性
 - 顧客に受け入れてもらえず、時間やコストのロスや顧客の不満

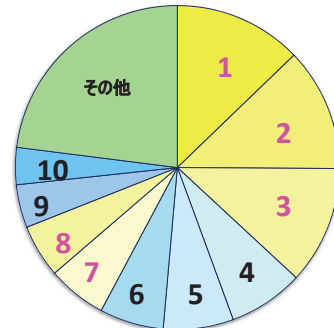
顧客のニーズを漏れなく正確に要求仕様書に記述することが重要

平成29年度 産業日本語研究会・ワークショップ

8

参考: ソフトウェア開発プロジェクトが失敗する原因^[1]

1. ユーザの参加不足
2. 要求および仕様が不完全
3. 要求および仕様がよく変更になる
4. 経営者側からの支援不足
5. 技術的能力の不足
6. リソースの不足
7. 現実的でない期待
8. 不明確な目的
9. 現実的でない納期
10. 全く実績のない新規技術



失敗する原因として、要求分析に関係するものは多い

平成29年度 産業日本語研究会・ワークショップ

9

IEEE Std. 830-1998～ソフトウェア要求仕様書～

平成29年度 産業日本語研究会・ワークショップ

10

IEEE Std. 830-1998[2]とは

- IEEEという国際的な学会が定めた、ソフトウェアの要求仕様に対して推奨される標準・規格
- 要求仕様書のテンプレートと品質特性を規定
 - この構成の内容を含むことで、少なくとも大きなミスや抜けがなく、ある程度の品質が確保された要求仕様書を作成可能
 - 要求仕様書の記述前に、記述内容を十分に収集することも極めて重要
 - 記述された要求仕様書を、品質特性の観点でチェックすることで、完成度の高いものに洗練

平成29年度 産業日本語研究会・ワークショップ

11

規定されている要求仕様書の構成

- 1. はじめに(Introduction)
 - 1.1 目的(Purpose)
 - 1.2 適用範囲(Scope)
 - 1.3 用語・略称の定義(Definitions, acronyms, and abbreviations)
 - 1.4 参考文献(References)
 - 1.5 概要(Overview)
- 2. 概要(Overall description)
 - 2.1 ソフトウェアの全体像(Product perspective)
 - 2.2 ソフトウェアの機能(Product functions)
 - 2.3 ユーザ特性(User characteristics)
 - 2.4 制約条件(Constraints)
 - 2.5 前提と依存関係(Assumptions and dependences)
 - 2.6 要求の割り当て(Appportioning of requirements)
- 3. 要求仕様の詳細(Specific requirements)
 - 3.1 外部インターフェース(External interfaces)
 - 3.2 機能(Functions)
 - 3.3 パフォーマンス要求(Performance requirements)
 - 3.4 論理データベース要求(Logical database requirements)
 - 3.5 設計上の制約(Design constraints)
 - 3.6 ソフトウェアシステムの属性(Software system attributes)
 - 3.7 個別の要求の構成(Organizing the specific requirements)
 - 3.8 追記(Additional comments)
- 4. 補足情報(Supporting information)
 - 4.1 目次と索引
 - 4.2 付録

平成29年度 産業日本語研究会・ワークショップ

12

IEEE Std. 830-1998～品質特性～

平成29年度 産業日本語研究会・ワークショップ

13

品質特性とは？

- 要求仕様書が「良い」ものであるために必要な要素
- 要求仕様書が品質特性を満たすことで
 - 開発側と顧客側の、開発するソフトウェアに対するイメージに違いがなくなる
 - 開発されたソフトウェアについての検証が容易になる
 - 要求の変更もしやすくなる
 - etc.

要求仕様書の第3章について、品質特性を満たすよう
記述することが重要

平成29年度 産業日本語研究会・ワークショップ

14

要求仕様書の品質特性[2][6]

- 妥当であること (Correct)
- あいまいでないこと (Unambiguous)
- 完全であること (Complete)
- 矛盾がないこと (Consistent)
- 重要度と安定度のランク付けがされていること (Ranked for importance and/or stability)
- 検証可能であること (Verifiable)
- 変更が容易であること (Modifiable)
- 追跡可能であること (Traceable)

平成29年度 産業日本語研究会・ワークショップ

15

妥当であること

- 要求仕様書に記述されている要求が、ソフトウェアが満たすべき事柄と一致していること
 - 要求仕様書に記述されている要求が正確に記述されていることを示す特性
 - 顧客のニーズを反映しているか
 - 関連する他のシステムとの整合性や、適用すべき標準・規格を満たしているか, etc.
 - この品質特性を満たしていない = 顧客のニーズが正確に記述されていない
 - 顧客のニーズとは合わないソフトウェアが完成する



- 顧客による確認
 - 関連する文書等との比較
- により、できるだけ正確に記述することを目指す

平成29年度 産業日本語研究会・ワークショップ

16

あいまいでないこと

- 要求仕様書に記述されている要求が、ただ一通りに解釈できること
 - － この品質特性が満たされていない要求が存在する
 - = その要求に対する解釈が複数通り存在する
 - そのまま開発が進行すると、顧客のニーズとは合わないソフトウェアが完成する

平成29年度 産業日本語研究会・ワークショップ

17

あいまいでないこと～自然言語の特性～

- 自然言語はあいまいなところを完全には排除できない
 - － 要求仕様書を読む個人によって、記述された要求の解釈が異なることがある
- 顧客側と開発側で、業務において利用している用語が異なることがある
 - － 記述された要求に対する誤解が生じることがある
 - － 開発側と顧客側で必ずしも同じバックグラウンドを持っていないため、開発側にとって理解しやすい要求仕様書は、顧客側にとっては理解しにくいこともある



- 様々な人によるレビュー
 - 用語集の作成
- により、できるだけあいまいでなくすることを目指す

平成29年度 産業日本語研究会・ワークショップ

18

完全であること

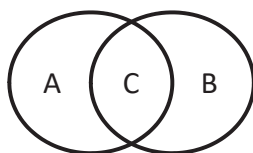
- 顧客のニーズが漏れなく要求仕様書に記述されているかを示す特性
 - この品質特性が満たされていない = 要求仕様書に記述されていない要求が存在する
 - 完成したソフトウェアで
 - 必要な機能が実装されていない
 - 実装された機能が必要な振る舞いをしない
 - 必要な性能を満たしていない
- ↓
- 機能や性能、設計上の制限、属性、他のシステムとのやりとり等の重要な要求を全て含んでいるか
 - 起こりうる全ての状況における入力に対する応答の定義を含んでいるか
 - 特に有効・無効な入力値の両方に対する応答が特定されているか
 - 要求仕様書中の全ての図表へのラベルと参照や、全ての用語の定義、単位の定義をしているか
- をチェックすることにより、できるだけ漏れをなくすることを目指す

平成29年度 産業日本語研究会・ワークショップ

19

「妥当であること」と「完全であること」の関係^[3]

- A: 開発するソフトウェアが満たさなければならない要求
- B: 要求仕様書に記述されている要求
- C: 開発するソフトウェアが満たさなければならない要求で、かつ要求仕様書に記述されているもの($A \cap B$)



- A-C: 満たさなければならない要求で、要求仕様書に記述されていないもの(要求の漏れ)
 - 品質特性「完全であること」が満たされていない
- B-C: 必ずしも満たさなくて良い要求で、要求仕様書に記述されているもの
 - 品質特性「妥当であること」が満たされていない
- A = B = Cとなったとき、要求仕様書は妥当かつ完全

平成29年度 産業日本語研究会・ワークショップ

20

矛盾がないこと

- 要求仕様書内部で矛盾がないこと
 - － 矛盾: ある要求を実現すると、別の要求が実現できない、という状況
 - 記述箇所により、利用するものやその状況が違う
 - 記述箇所により、ソフトウェアの動作が違う
 - 記述箇所により、使う用語が違う, etc.
 - － この品質特性を満たしていない = 最終的に達成できる要求と達成できない要求が存在する
 - 設計や実装の段階で、矛盾する要求をどのように扱うか、どちらを実現させるか、混乱が生じる

平成29年度 産業日本語研究会・ワークショップ

21

重要度と安定度のランク付けがされていること(1)

- 各要求について、重要度あるいは安定度を示す指標を明確につけておくこと
 - － 要求仕様書に記述された要求は、全てが実現すべき必要な要求とは限らない
 - 必ず実現すべき重要な要求
 - 必ずとは言わないまでもできるだけ実現してほしい要求
 - 余裕があれば実現してほしい要求、etc
 - － コストや納期などの面で要求仕様書に記述された要求全てを実現することが難しいこともある
 - ランク付けにより、優先すべき要求が明確になる

平成29年度 産業日本語研究会・ワークショップ

22

重要度と安定度のランク付けがされていること(2)

- 各要求について、重要度あるいは安定度を示す指標を明確につけておくこと
 - － この品質特性が満たされていない = 要求仕様書を読む個人によって、各要求の重要度の考え方が異なる
 - a. 顧客側と開発側でのコミュニケーションに支障
 - b. 設計・実装段階での混乱
 - c. 顧客側にとって重要な要求の実現がなし
 - － この品質特性が満たされている = 優先的に実現すべき要求が明確になる
 - a. ～ c. の問題の回避
 - コストや納期等の見積もりにも有効

平成29年度 産業日本語研究会・ワークショップ

23

重要度と安定度のランク付けがされていること(3)

- 要求の重要度のランク付け方法(例)
 - － 必要なもの (Essential)
 - これらの要求が満たされなければ、ソフトウェアが受け入れられない
 - － 条件つき (Conditional)
 - ソフトウェアの価値を高める要求であるが、満たされていなくても受け入れられないという要求ではない
 - － オプション (Optional)
 - なくても大きな問題はないが、要求仕様書を超える何かを提供する機会を開発側に与える

※この他、「重要度1～5」のような詳細な順位づけの方法もありうる

平成29年度 産業日本語研究会・ワークショップ

24

重要度と安定度のランク付けがされていること(4)

- 安定度
 - － 経験や知識に基づいた要求への将来の変更予測
 - 記述された要求が変更される可能性の有無
 - 予測される変更の回数
- などを明確にしておく

平成29年度 産業日本語研究会・ワークショップ

25

検証可能であること

- ソフトウェアが要求を満たしているかどうかを機械的に確認可能であること
 - － 妥当なコスト内、妥当な時間内で、人またはコンピュータがチェックできる手法が存在すること
 - － 一般的に、あいまいな要求は検証可能でない

平成29年度 産業日本語研究会・ワークショップ

26

変更が容易であること(1)

- 要求へのあらゆる変更が容易に、完全に、一貫して行えるような構成になっていること
 - － ソフトウェア開発の進行中であっても、要求は随時変更
 - ビジネス環境の変化やソフトウェアを稼働させる環境の変化のため
 - 要求の見落としや要求の記述ミスのため, etc.
- 要求の変更にあたっての重要事項
 - － 変更対象の要求がどこに書かれているかを容易に発見できること
 - － 変更対象の要求に依存する要求(併せて変更が必要となる要求)ができるだけ少ないこと

平成29年度 産業日本語研究会・ワークショップ

27

変更が容易であること(2)

- この品質特性を満たすための重要事項
 - － 目次や索引、明確なクロス・リファレンスが整備され、使いやすい構造であること
 - － 冗長でないこと
 - 例えば、同じ要求が要求仕様書内で複数箇所に記述されていないこと
 - － 他の要求と混ざらず、各要求を独立・分離して表現していること
 - 要求が互いに依存しないこと
 - 互いに依存する要求が多くなると
 - － 設計や実装、テスト部分の変更も多くなる
 - － 見落としやミスの原因に
 - － コスト超過や納期遅延などの原因の原因に

平成29年度 産業日本語研究会・ワークショップ

28

変更が容易であること(3)

- 冗長性
 - 冗長であること自体はエラーにはならないが、エラーに容易に結びつく可能性
 - 冗長であることは要求仕様書をより読みやすくすることも多いが、文書がアップデートされたときに問題が起こる可能性
 - 例えば、複数箇所に記述されている事柄のうちの1つだけを変更するなど(矛盾が存在することに)
 - 冗長性が必要な場合には、明確なクロス・リファレンスを含むこと

平成29年度 産業日本語研究会・ワークショップ

29

追跡可能であること(1)

- 個々の要求の起源が明確で、将来の開発で改良された文書等との対応付けがとれること
 - ソフトウェア開発は、要求仕様書をもとにして設計・実装・テストへと進行
 - 開発の段階ごとに様々な文書が作成
 - 設計・実装・テストの各段階で作成される文書も、必要に応じて随時変更
- この品質特性が満たされていると
 - 要求仕様書に記述された要求から各種文書の内容へたどることができる
 - 各種文書内の記述から要求仕様書に記述された要求にさかのぼることができる

以下を容易に確認することができる

 - 要求の変更が各種文書のどの部分に影響するか
 - 各種文書の変更によってどの要求が影響を受けるか

平成29年度 産業日本語研究会・ワークショップ

30

追跡可能であること(2)

- 2種類の追跡可能性が存在
 - － 後方追跡可能性 (Backward traceability)
 - 開発が始まる以前の段階への参照
(要求仕様書→要求のソースとなる文書への参照)
 - 各要求が、以前に作成された文書においてそのソースを明確に参照できること
 - － 前方追跡可能性 (Forward traceability)
 - この要求仕様書をもとにして作成された全ての文書への参照
(要求仕様書→各種文書への参照)
 - 要求仕様書内の各要求が、一意の名前や参照番号を持つこと

参考: IEEE Std 830-1998以外の品質特性^[4]

- 達成可能であること
 - － 要求仕様書に記述された全ての要求を満たすソフトウェアを構築できること
 - 技術的に不可能な要求が含まれていない
 - 矛盾する要求が存在しない, etc.
- 注釈がつけられていること
 - － 他の要求との関連を含めて要求の特色がわかりやすく記述されていること
 - 要求の起源や、誰がその要求の実現に責任を持つか、その要求の主要な顧客は誰か、など、要求に関係する様々な情報

要求仕様書とソフトウェアの品質の相関

平成29年度 産業日本語研究会・ワークショップ

33

ソフトウェアの品質

- ソフトウェア開発: ソフトウェアの品質を高める必要
 - 開発のコストやスケジュールを考えると、あらゆる品質を高めるのは難しい
 - 顧客の特性や利用状況、ソフトウェアの特性などにより、重要な品質は異なる

ソフトウェアの品質特性(ISO/IEC 9216 (JIS X 0129-1994))[5]*

- 機能性
- 信頼性
- 使用性
- 効率性
- 保守性
- 移植性

要求仕様書の品質特性

- 妥当であること
- あいまいでないこと
- 完全であること
- 矛盾がないこと
- 重要度と安定度のランク付けがされていること
- 検証可能であること
- 変更が容易であること
- 追跡可能であること

品質特性同士の関連は？

- 要求仕様書のどの品質特性を満たせば、ソフトウェアのどの品質特性を満たせるかが予測できれば、要求仕様書の記述時に、特定の品質特性に注力することができる[3]

平成29年度 産業日本語研究会・ワークショップ

*現在は、新しくISO/IEC 25010という規格に置き換わっている

34

相関関係の検討

- 情報処理学会ソフトウェア工学研究会要求工学ワーキンググループにより検討
 - ワーキンググループのメンバー全員が、相関表に◎(強い相関あり)、○(相関あり)、△(弱い相関あり)、なし(相関なし)を記入して持ち寄り
 - 相関表: ソフトウェアの品質特性を列、と要求仕様書の品質特性を行とする表
 - ◎・○・△・なしの理由を議論し、修正を加え、得点化して平均
 - 平均をもとに、◎・○・△・なしに再度変換

平成29年度 産業日本語研究会・ワークショップ

35

相関表の利用

- 重視するソフトウェアの品質特性を決定し、この相関表をカスタマイズ
 - 開発対象のソフトウェアにより相関表は異なる

	機能性				
	合目的性	正確性	相互運用性	標準適合性	セキュリティ
妥当であること	○	◎	△	△	○
あいまいでないこと	○	◎	△	△	○
完全であること	○	◎	△	△	○
矛盾がないこと	△	△	△	△	○
ランク付けがされていること	◎	なし	なし	なし	なし
検証可能であること	△	○	△	△	△
変更が容易であること	計測しやすい	計測しやすい	計測しやすい	計測しやすい	計測しやすい
追跡可能であること	計測しやすい	計測しやすい	計測しやすい	計測しやすい	計測しやすい

※条件によって相関が変わることもあるので、注釈がついている項目も多い

平成29年度 産業日本語研究会・ワークショップ

36

参考文献

- [1] Standish Group. CHAOS Report, 1994.
- [2] IEEE, Recommended Practice for Software Requirements Specifications, Std. 830-1998, 1998.
- [3] 要求工学ワーキンググループ: 要求定義で困ってませんか? 要求仕様書の品質に関する研究成果報告, <http://www.selab.is.ritsumei.ac.jp/~ohnishi/RE/rewg-tr1v2.pdf>, 2007.
- [4] Davis, A. M., Just Enough Requirements Management: Where Software Development Meets Marketing, Dorset House, 2005.
- [5] International Standard ISO/IEC 9126-1. Software engineering - Product quality - Part 1: Quality model, 2001.
- [6] 妻木, 白銀: 要求工学概論, 近代科学社, 2009

— 禁無断転載 —

平成 29 年度
産業日本語研究会 報告書
「産業日本語」
Technical Japanese

平成 30 年 3 月

一般財団法人 日本特許情報機構 特許情報研究所

東京都江東区東陽 4 丁目 1 番 7 号

TEL 03-3615-5511