

国立国語研究所学術情報リポジトリ

On the expansion of thesaurus entries

メタデータ	言語: jpn 出版者: 公開日: 2019-03-25 キーワード (Ja): キーワード (En): 作成者: 山崎, 誠, YAMAZAKI, Makoto メールアドレス: 所属:
URL	https://doi.org/10.15084/00002157

シソーラスの可能性

山崎 誠

(国立国語研究所)

キーワード

シソーラス, 『分類語彙表』, 類語辞典, 言語単位, 分類と配列

要 旨

本稿では、シソーラスの定義を具体的に検討することにより、シソーラスとはどのようなものであるかを明らかにし、それを踏まえてシソーラスの拡張の可能性を考察した。今までシソーラスが収録してきた範囲はいわゆる単語が中心であったが、単語は形態素から文にいたるさまざまな言語表現全体の中ではごく一部分にすぎず、組み合わせの自由度が高い連語や複合語などの生産的な表現にも類義的な対象を見出すことができる可能性を示唆した。

1. はじめに

近年、日本語を対象としたシソーラスや類語辞典が相次いで刊行され、日本でもシソーラスの認知度が高まってきた。この機会にシソーラスとは何か、どのような性質を持っているか振り返り、今後のシソーラスの発展や拡張の可能性を探りたい。シソーラスの拡張と言うと、新語の追加や語彙を分類する観点の多様化などが思い浮かぶが、ここでは別の立場からの提案を行う。シソーラスに収録されている語を言語単位や意味の性質からとらえたとき、そこで積極的に排除されている言語表現があることが分かる。本稿では、そのような今までシソーラスの収録対象と扱われてこなかった言語表現に着目し、その活用が有用であることを指摘する。

2. シソーラスとは何か

「シソーラス」を辞書で引いてみると、ほぼ同じような、しかし、完全には一致しない説明が見出される。以下にその幾つかを引用する。

- (1) 「(宝庫の意) 語を意味的類似により分類・配列したもの。分類語彙表。イギリスのロジェ (P. M. Roget 1779~1869) が1851年に刊行した辞典名に由来。」(岩波書店『広辞苑』第五版)
- (2) 「類義語辞典の一種。語を意味によって分類・配列したもの。分類語彙集。」(三省堂『三省堂類語新辞典』)
- (3) 「①類義語辞典の一種。意味の近い語句を集めて配列した語彙(ゴイ)集。②情報検索(ケンサク)に使うキーワードと関連語との関係をしるしたリスト」(三省堂『三省堂国語辞典』第五版)

- (4) 「①語を意味の上から分類・配列した辞書。②情報検索のために、術語の類義・対義、上位・下位などの関係を明確に定義した辞書。」(小学館『新選国語辞典』第八版)
- (5) 「①分類体辞典をさしている。中国の「爾雅」や日本の「和名類聚抄」がこれにあたる。②コンピュータなどの情報検索に用いる指標(インデックス)で、広く同義語、類義語を分類整理したもの。」(小学館『言泉』)
- (6) 「①単語を集め、意味などによって分類した本。類語辞典、分類語彙表などのたぐい。②コンピュータなどの情報検索に用いる指標(インデックス)で、広く同義語、類義語を分類整理したもの。」(小学館『日本語新辞典』)
- (7) 「①語句を同義・類義などによって分類・配列した語彙集。また、類義語集。②コンピュータで、同義語・類義語などからも検索できるようにした辞書機能。」(大修館書店『明鏡国語辞典』)
- (8) 「①語句を意味によって分類・配列したもの。分類語彙表。②情報検索において、キーワードの範囲、関連語との関係などを記したリスト。」(岩波書店『岩波国語辞典』第六版)
- (9) 「①知識の宝庫としての辞典。〔狭義では、△類語(関連語)辞典を指す〕②情報検索に使われるキーワードと関連語との関係を示した、一種の辞書。」(三省堂『新明解国語辞典』第六版)
- (10) 「①知識の宝庫。辞書・事典などをいう。とくに類義語辞典を指す。②コンピュータで記憶された情報の索引。」(新潮社『新潮現代国語辞典』第二版)

情報検索に使われるシソーラスを別にすれば、以上に共通する特徴は、およそ次の表1のようなものになるだろう。

表1 シソーラスを特徴付ける共通項

何を(対象)	: 語(句)を
どのように(基準)	: 意味的類似によって
どうする(方法)	: 分類・配列した
何(主体)	: 辞書(あるいは語彙集)

シソーラスを特徴付けるこのようなそれぞれの性質は、シソーラスとそうでないものとを区別する働きがある。上記の特徴群をシソーラスの定義と考えれば、それらから外れたものはシソーラスとは呼べないということになるだろう。例えば、次のようなものである。

- ・語でないもの(文や文章など)を意味的類似により分類・配列したもの
- ・意味的類似によらない分類を施したもの(品詞による分類、語の長さによる分類など)
- ・意味的類似によらない配列(J I Sコード順配列、語末の音による配列など)

「意味的類似」の考え方にもよるが、ある語から連想する語(「煙草」に対する「火」「禁煙席」「パイプ」など)の集合はシソーラスとは呼べないほうがよいだろう。これは、意味的類似では

なく語用論的な場面に基づいた分類である（石綿 1980）。ある語と全体・部分の関係を構成する語の集合（「車」と「タイヤ」「ハンドル」など）もシソーラスにおける収録状況を見ると、近接した項目に現れることもあるが、積極的に意味的類似ととらえられてはいないようだ。

意味を手がかりに配列することがシソーラスの特徴であるが、検索のためには、形による配列（＝索引）も付いていなければ、特定の語を探し出すような場合には、実用的でない。意味には、形と違って固定的な順列が定まっていないからである。実際、近年の英語シソーラスの中にはアルファベット順のものが多く見受けられる。

国語辞典における説明を見る限りでは、語の意味説明（語釈）のあるなしは、シソーラスの本質とは無関係のようであるが、語釈がある場合は「辞典」という性質が強くなり、実質的に類義語辞典に近いものになるだろう。日本では、シソーラスという語自体があまり普及していないためか、意味的類似による分類・配列をほどこしていても、書名としては類語辞典等の名称を付与し、語釈を伴うものが多いのが現状である。

シソーラスの性質を包括的にとらえるために、俯瞰的な立場からシソーラスをとらえなおしてみることにしよう。

3. シソーラスが取り扱っている言語単位はどのようなものか

一般の国語辞典もそうであるが、シソーラスもその目的によって収録する言語単位の大きさや言語単位の意味的特徴が限定されている。例えば、国立国語研究所（1964）『分類語彙表』の「まえがき」には、以下のようにある。

「この分類語彙表に収めた語は、およそ三万二千六百語である。これらの語は、国立国語研究所報告21『現代雑誌九十種の用語用字』第一分冊（1962）の語彙表に掲げる高使用率の語のうち、人名、会社名、球団名等の個別の名、および記号の類を除く約七千語を中心とし、それにつづく使用率をもつ約五千語を補い、」（国立国語研究所 1964：7）

ここで注意したいのは、次の2点である。

- ・固有名詞を対象外としていること
- ・語彙調査の結果である語彙表をデータとしていること

固有名詞の除外は、言語単位の意味的特徴に関する限定であり、語彙調査の結果である語彙表をデータとしていることは、言語単位の大きさに関する限定である。

この限定の内容は以下のとおりである。雑誌九十種の語彙調査では、使用度数7以上の7,217語のうち、約9%にあたる652語が固有名詞（人名・地名）である（国立国語研究所（1997）を使用して算出）が、一部の代表的地名を除いては『分類語彙表』（1964）には収録されていない。また、この語彙調査は、調査単位として β 単位（二次結合以上の長い複合語を認めない単位）を用いているため、長い複合語（「共同研究」「日常生活」など）、接辞や造語成分の付いた形（「面白さ」「都会的」「諸問題」など）、慣用句（「首をかしげる」など）は分割されている。その帰結として、『分類語彙表』（1964）にはこれらの語句はあまり収録されていない。

言語単位の大きさとその意味的特徴を元にシソーラスが扱っている範囲を示したものが図1で

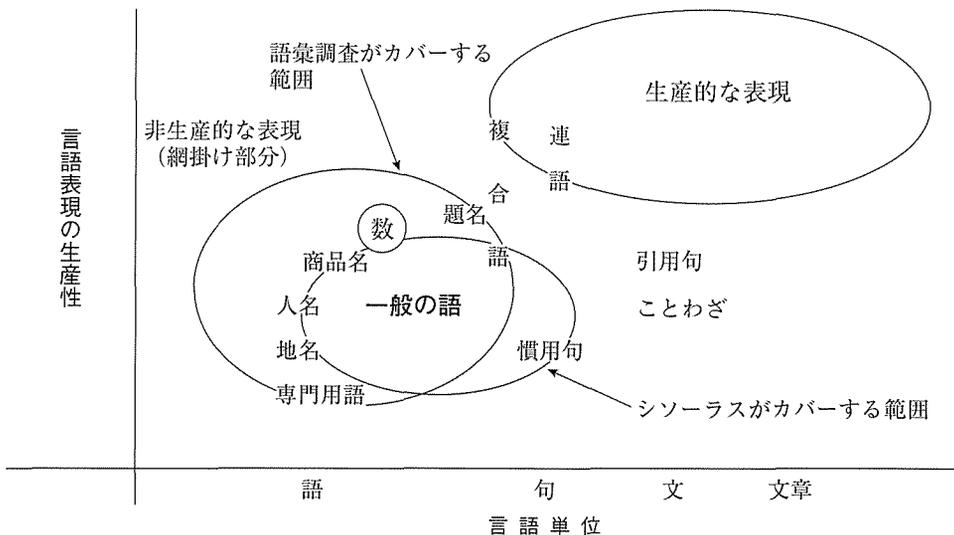


図1 シソーラスの範囲

ある。この図は特定のシソーラスを想定したものではなく、日常語を対象としたシソーラスの一般的な姿と考えられる範囲を示したものである。個別的には、この範囲からの逸脱は十分ありうる。また、図中に配置された語句の位置関係は、厳密なものではない。

図1には、語彙調査がカバーする範囲も併せて示した。一般的に語彙調査は特定の意味の語を排除することはせず、調査対象内に現れた言語表現を網羅的に扱う。ただし、言語単位の大きさは、せいぜい句（あるいは文節）までが通常である。

4. シソーラスがカバーしていない言語表現の特徴はどのようなものか

本節では、シソーラスが取り扱っていない言語表現の特徴を「生産性」という観点から考察する。本稿で言う生産性とは、言語単位の組み合わせにおける自由度と言い換えることもできる。この生産性という観点から言語表現を二分して考えてみよう。生産的な表現とは、文に代表されるような個別の出来事を表す言語表現で、語の組み合わせ方が比較的自由であるものである。例えば、新聞や雑誌、日常会話における個々の文などであり、個別の場面に依存する性質を持つ。非生産的な表現とは、語に代表されるような現実の一断片を表す言語表現である。複数の語から構成される言語単位は、その組み合わせ方が固定的なものは非生産的な表現とする。複数の語から構成される場合として、慣用句や複合語があるが、複合語には、生産的なものもあり、単純にどちらかに決めることはできない。なお、この区別は本稿での説明のため便宜的に行ったものである。

図1によれば、シソーラスがカバーしてないものとして、専門用語や人名、地名などの固有名詞、あるいは、引用句、ことわざなどがあるが、これらは、日常語を離れば、それぞれの独自の分野で用語集やシソーラスが編集されている。すなわち、シソーラスがカバーする最大の範囲

は、非生産的な表現と一致すると考えてよいだろう。

そうすると、従来のシソーラスがカバーしていない範囲とは、生産的な表現ということになる。したがって、シソーラスを拡張するとしたら、この生産的な表現の部分を取り込む方策が考えられる。生産的な表現と非生産的な表現（図1の網掛け部分）の特徴を言語資料、言語データとしての活用のされ方という観点で整理したのが表2である。

表2 生産性の違いによる言語データとしての活用

言語データとしての活用	非生産的な表現	生産的な表現
辞書・事典の見出しとして	多い	まれ
使用頻度を計る調査単位として	適切	不適切
意味記述をほどこす対象として	適切	不適切
言い換えや対訳をほどこす対象として	適切	適切
参照される用例として	不適切	適切

生産的な表現は、用例の提示、あるいは言い換えや対訳の対象として利用されることが多い。また、生産的な表現と非生産的な表現とは、言語データとしての活用のあり方に共通性が少ないことが分かる。

5. 生産的な表現をシソーラスでどのように利用するか

表2から、「言い換えや対訳をほどこす対象として」の取り扱いが、生産的な表現にも非生産的な表現にもどちらにも適用できるものであることが分かる。これを利用して両者の間の意味的な対応を考えてみる。

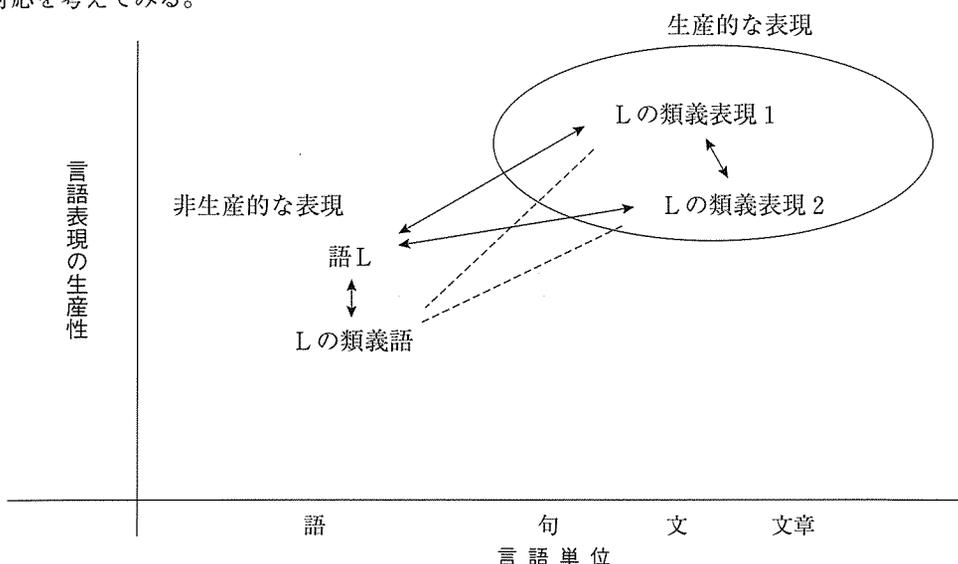


図2 生産的な表現と非生産的な表現の意味的対応

従来のシソーラスは、非生産的な表現の中での意味的な関係を整理したものと言えるだろう。この中に、生産的な表現からの意味的対応付けを行うことは容易である。もっとも基本的な対応付けは、意味的類似性である。図2を見てみよう。

類義関係が非生産的な表現内部への対応で閉じていたものが、生産的な表現を導入することで、もう一つの大きな言語表現の集合への対応に拡大することができる。生産的な表現は語の組み合わせが比較的自由であることから、非生産的な表現に比べて、表現可能な意味の領域が広く、きめの細かい意味の表示が可能である。このことを端的に証明しているのが辞書における見出し語と語釈との関係である。どのような見出し語に対しても、言葉を尽くせば、他の語と弁別可能な語釈を付けることができる。

非生産的な表現から生産的な表現への意味的類似による対応の例を、共同通信社編（2005）『記者ハンドブック 第10版』の用字用語集の部分の書き換え対象語とされたものの中からいくつか拾ってみた。

迂遠→实际的でない	交誼→親しい交わり，厚いよしみ
紆余曲折→複雑な経過	洒脱→あか抜けした，俗気のない
蘊蓄→深い知識	人口に膾炙する→広く世間に知れ渡る，有名になる
闊歩→大手を振って歩く，威張って歩く	逼塞→ひっそりと暮らす
杞憂→無用の心配	敬虔→信仰心のあつい
教鞭を執る→生徒を教える	平坦→起伏がない

また、生産的な表現と非生産的な表現の関係に関連するものとして、対象言語とメタ言語との関係がある。メタ言語とは、言語表現そのものについて記述する際の表現様式のこと、例えば、「あけましておめでとうございます」（対象言語）に対する「新年の挨拶」という語句がメタ言語となる。このような関係は言語行為に関する表現に多く見られる。以下に挙げるような三つの表現の関係がシソーラスにおいて記述されるためには、生産的な表現の導入が不可欠である。

「感謝する」（メタ言語，非生産的表現）

「お礼を言う」（メタ言語，生産的な表現）

「お心遣いありがとうございます」（対象言語，生産的な表現）

生産的な表現は、膨大に存在するため、どのようにして適切に最適な表現をデータから取り出して来たらよいかその方法が今後の検討課題となる。

6. 意味的類似による分類はどのような言語単位に対して有効か

シソーラスを特徴付ける性質のうち、分類・配列について考えてみよう。ここでは、形による配列（五十音順配列など）を「形」、意味的な類似性に基づく配列を「意味」として単純化して考える。

日本語のキーワード（見出し）を付け、そのキーワードの五十音順に配列したユニークな辞書である。

もともと意味には定まった順序がなく、基本的に始まりや終わりのある線条的な構造を持っていない。そのことを特徴的にとらえたのが林(1957:105)の「星図になぞらえた語彙表」(図3)である。

特定の語を参照しているときは、その語の周りに関連語が配置されていけばよいのである。この特性を考慮すれば、シソーラスの配列は、(端を持たない)球面上に配置するのが適切なものかもしれない。このような発想は他の分野にも見られる。久保田他(2004)では、様々な知的コンテンツを蓄積する「知球」というシステムを提案している。

7. シソーラスに欠けている情報はどのようなものか

Crystal(1987)によると、シソーラスには類義性の具体的な内容が記述されていないため、言語研究者は興味を持たないという指摘がされている。

しかしながら、言語研究者にとってはその価値は小さい。というのは、シソーラスには個々の語彙素のあいだの意味関係に関する情報が含まれず、また、地域・社会・職業による差異についても、異種の語彙素が注釈なく並列されているからである。意味の構造を研究するには、語彙素間の意味関係をつきとめるためのより正確な手段が必要とされる。(邦訳162)

このような位相に関する情報や体系を構成する個々の語同士の意味関係を構造的かつ統一的に把握するには、使用文脈から自動的に意味関係を解析する手法の導入が必要となるだろう。その一例として、岩城・蛭谷(2006)がテキストマイニングの方法を用いて、分類語彙表の部分体系を構造的に示した例が挙げられる。また、使用頻度や語種などの情報もシソーラスには欠かせないものとして今後記載されることになるだろう。とくに使用頻度については、コーパスの活用が期待される。

8. おわりに

本稿では、言語的特徴をもとにシソーラスを今後どのように発展させられるかについて述べた。シソーラスの拡張は、実際にどのような活用場面があるかと深く関連しているので、シソーラスを必要とする現実的な場面からのフィードバックも今後検討したい。

参考文献

- Crystal, D(1987) *The Cambridge Dictionary of Language*, Cambridge : Cambridge University Press.
(邦訳：風間喜代三他監訳(1992)『言語学百科事典』大修館書店)
石綿敏雄(1980)「シソーラス」国語学会編『国語学大辞典』, 466-467, 東京堂出版
岩城裕之・蛭谷安紀子(2006)「分類語彙表」3.1910-10語彙の部分体系—同一文中に出現する他語との関係を手がかりに一」『語彙研究』3, 1-11

共同通信社編(2005)『記者ハンドブック 第10版 新聞用字用語集』共同通信社
久保田秀和他(2004)「知球」：持続的に発展可能な時空間記憶の構築』『ヒューマンインタフェース』(情報処理学会研究報告), 2004(90), 1-8
国立国語研究所(1964)『分類語彙表』秀英出版
国立国語研究所(1997)『現代雑誌九十種の用語用字 全語彙・表記[FD版]』三省堂
三省堂編修所編(2002)『カタカナ類語辞典』三省堂
林大(1957)「語彙」『講座現代国語学Ⅱ ことばの体系』, 95-125, 筑摩書房
宮島達夫(1994)『語彙論研究』むぎ書房

付 記

本稿は、2005年12月17日に、国立国語研究所にて開催された公開研究発表会「シソーラスの^{へんさん}編纂と活用」の予稿集に掲載したものを改稿したものである。

山崎 誠 (やまざき まこと)

国立国語研究所 研究開発部門

190-8561 東京都立川市緑町3591-2

yamazaki@kokken.go.jp