

感覚・知覚領域を起源に持つ英語形容詞の意味拡張の調査・分析

進藤三佳 内元清貴 井佐原均

独立行政法人 情報通信研究機構

{mshindo, uchimoto, isahara}@nict.go.jp

ある語彙の定まった意味が、時の経過とともに変化していく意味変化・意味拡張の現象は、その表面的な言語現象の根底にある人間の認知メカニズムを反映している。本研究では、形容詞の意味をその修飾・叙述している名詞に注目することによって自動的に分類する方法を提案し、そのシステムを紹介する。このシステムによる検証の一例として、英語の感覚形容詞の大規模コーパスにおける意味分布を調査し、感覚形容詞が系統だった意味拡張をしているか否かを分析する。

1. 先行研究・理論的枠組み

言葉の意味のありようには、人間の認知的活動の諸相が強く影響している。殊に抽象的事物を解釈する際に、人間は自らの身体的経験をを用いて概念化しているとされ、感覚・知覚語彙が本来の意味以外に抽象の意味を獲得していく意味変化・意味拡張の現象は、その認知活動を映し出すものとして Ullmann (1951)の頃から注目されてきた。これらの研究は、主に動詞を中心にメタファー的観点から行われ、感覚知覚語彙は認知的抽象概念と規則的關係(たとえば、視覚 知性、聴覚 従順、味覚 個人的好悪、触覚 感情)があるとされてきた。しかしこのメタファー的方法論は、ある感覚領域が拡張しやすいと思われる抽象概念(例えば、視覚の場合は知性)に関連する語彙の内、その感覚領域に起源を持つ語がいかに多いかを列挙する方法が取られ、その規則性に対する徹底した検証があまりなされてこなかった。また、動詞と同様に感覚語彙として重要な形容詞については、いくつかのメタファー(e.g. Affection Is Warmth/Dislike Is Cold, Intelligence Is A Light Source.)の一部として論じられてきたのみにとどまり、これまであまり研究対象として取り上げられてこなかった。しかるに形容詞は、描写する際の正確性を増し、段階性をあらず特徴ゆえに、人間が本来どのように状況を解釈しているかを如実に反映している。その意味で、形容詞の意味拡張を探ることは、人間の認知的メカニズムを言語現象から探る極めて有効なアプローチとなる。

そこで本研究では、大規模コーパスに含まれるすべての形容詞を対象として、形容詞の意味をその修飾・叙述している名詞に注目することによって自動的に分類する方法を提案し、そのシステムを紹介する。このシステムにより、各形容詞が現代英語においてどのようなカテゴリーをもっとも頻繁に意味するかを調査する。また、このシステムによる検証の一例として、上に述べた英語の感覚形容詞の意味拡張のさまを示し、知覚動詞について提唱されている

ような、抽象概念との体系的メタファー的關係が感覚形容詞についても当てはまるのか否か、メタファー以外のどんな規則性が、感覚形容詞の意味拡張に見いだされるのかを探る。

2. データ抽出方法・システムの開発

形容詞の意味は、その形容詞が修飾している名詞、あるいは、その形容詞によって叙述されている名詞を特定することによって、把握できるものと考えられる。言葉の使用上、形容詞よりも名詞の方が使用されている意味が一意に決まるあいまい性が低いと考えられるからである。例えば、*sharp* + [noun]、あるいは、[noun] is *sharp* の場合、*sharp* - knife, *sharp* - pain, *sharp* - contrast, *sharp* - increase 等の用例が挙げられるが、*sharp* - knife の場合、knife は「切る道具」であるので、この場合の *sharp* は、「ある物理的な物体を描写している」と考えられる。同様に、pain は「身体的痛み」なのでこの場合の *sharp* は「ある生理的感覚を描写している」し、contrast は「対立した関係」なので *sharp* - contrast の *sharp* は「ある関係を描写している」し、increase は「変化、出来事」であるので *sharp* increase の *sharp* は「ある変化を描写している」というように判断できる。

このような前提に基づき、形容詞 名詞の限定用法・叙述用法両方の対データをBNCコーパスから抽出することとする。前回の研究では(進藤他 2004)、形容詞+名詞という隣接した共起対だけを取り出して調査したが、本研究では、Robust Accurate Statistical Parsing¹ (RASP, Briscoe and Carroll 2002)という各語彙間の文法関係を解析できるParserを用いることにより、限定用法・叙述用法の両方について、間に修飾語句等を含む用例についても、形容詞の意味対象となっている名詞を下記のように抽出した。

¹ RASPによるGrammatical Relationの抽出精度は、SUSANNE コーパスに対して、Precision 76.25%, Recall 76.77%, F-score 76.51% (Best parse)と報告されている (Carroll and Briscoe 2002)。

[限定用法(ncmod)²]の例

(|:2_PPIS1| |believe:3_VV0| |it:4_PPH1| |be+s:5_VBZ|
especially:6_RR		important:7_JJ		that:8_CST		
ACET:9_NN1		represent+s:10_VVZ		the:11_AT		
Church:12_NNL1		work+ing:13_VVG		in:14_II		the:15_AT
front:16_NN1		line:17_NN1		to:18_TO		provide:19_VV0
real:20_JJ		and:21_CC		practical:22_JJ		
support:23_NN1) 1 ; (-24.849)					
ncmod		real:20_JJ		support:23_NN1		
ncmod		practical:22_JJ		support:23_NN1		

[叙述用法・主格補語(ncsubj)-xcomp]³の例

(|Some:1_DD| |nation+s:2_NN2| |who:3_PNQS| |have:4_VH0|
|try+ed:5_VVN| |to:6_TO| |be:7_VB0| |serious:8_JJ|
|have:9_VH0| |end+ed:10_VVN| |up:11_RP| |be+ing:12_VBG|
|dull:13_JJ|) 1 ; (-18.760)
|ncsubj| |xcomp| |who:3_PNQS| (|be:7_VB0|)
|serious:8_JJ|
|ncsubj| |xcomp| |nation+s:2_NN2| (|be+ing:12_VBG|)
|dull:13_JJ|
(|The:1_AT| |market:2_NN1| |price:3_NN1| |end+ed:4_VVD|
|unchanged:5_JJ| |at:6_II| |169p:7_NN1|) 1 ; (-8.033)
|ncsubj| |xcomp| |price:3_NN1| (|end+ed:4_VVD|)
|unchanged:5_JJ|

[叙述用法・目的格補語(dobj)-xcomp]⁴の例

(|The:1_AT| |introduction:2_NN1| |of:3_IO| |higher:4_JJR|
fee+s:5_NN2		in:6_II		1956:7_MC		do+ed:8_VDD
not+:9_XX		solve:10_VV0		the:11_AT		District:12_NNL1
's+:13_\$		financial:14_JJ		problem+s:15_NN2		, :16_
only:17_RR		make:18_VV0		they+:19_PPH02		
manageable:20_JJ) 1 ; (-26.576)					
dobj		xcomp	(make:18_VV0)	they+:19_PPH02
manageable:20_JJ						
ncmod		higher:4_JJR		fee+s:5_NN2		
ncmod		financial:14_JJ		problem+s:15_NN2		

このようにして抽出した形容詞-名詞の対データの名詞の分類は、人手で行くと恣意的になるので、確立されたソーラスである WordNet 2.0 によって分類した。さらに、WordNet 2.0 は木構造をしており、その内部は細かく枝分かれしているため、各形容詞の意味対象となる名詞がどの下位分類に多く分布しているかを観察し、最も典型的に意味するカテゴリーを木構造の深度にかかわらず見つけ出すためのシステムを開発した。

このシステムは、二つの要素、「閾値」と「深度」をキーとして、その数値を変えてみることによって、どの下位分類のところに使用が多く集中しているかをさまざまに探索することができる。その最も意味分類の頻度の高い名詞カテゴリーは、濃い赤い色で示され、最も典型的な名詞は、緑色でマークされる。次にこのシステムを用いて行われた形容詞の意味分布調査の一例として、感覚形容詞の意味拡張の調査・分析を示す。

3. 感覚形容詞の意味拡張

3.1 調査方法

対象とする感覚形容詞は、Williams (1976)が共感覚表現の研究で用いた 63 個の形容詞に *plain* を加えた以下の 64 個の形容詞とした。

TEMPERATURE: cold, cool, hot, warm. (4 words); TOUCH: aspre, bitter, bland, cloying, coarse, crisp, dry, dull, grave, hard, harsh heavy, keen, mild, piquant, poignant, pungent, rough, sharp, smart, smooth, soft (22 words); TASTE: acrid, austere, brisk, dulcet, eager, mellow, sour, sweet, tart (9 words); DIMENSION: acute, big, deep, empty, even, fat, flat, full, high, hollow, level, little, low, plain, shallow, small, thick, thin (18 words); VISION: bright, brilliant, clear, dark, dim, faint, vivid (7 words); SOUND: loud, quiet, shrill, strident (4 words).

これらの形容詞が BNC のなかで用いられている用法について、RASP によって形容詞 名詞の対データを抽出し、WordNet 2.0 によって分類した。前回の調査(進藤他 2004)では、単なる共起対しか抽出できなかったため、限定用法の一部しか調査できなかったことになるが、本研究では、

に限定した: hold, keep, leave, call, confess, profess, pronounce, report, like, prefer, want, wish, believe, consider, deem, find, hold, imagine, judge, presume, rate, reckon, suppose, think, drive, get, make, prove, render, send, turn, certify, declare, proclaim. (*op.cit.*, 1196-7)。ただし、形容詞と名詞が前置詞で関係付けられている場合は抽出しない。また、形容詞の右が *that* の場合(it is JJ that ...の場合)も抽出しない。

² |ncmod|: RASPにより|ncmod|として抽出されるもののうち、形容詞と名詞の関係のみ、|ncmod|の関係として抽出した。ここで、形容詞は品詞がJJで始まる単語とし、名詞は品詞がNN、PN、VVG、PPのいずれかで始まる単語とする。また、品詞がNPで始まる単語は複合語の一部であることが多いため、品詞NPで始まる単語が形容詞と|ncmod|の関係にあり、かつ、名詞とも|ncmod|の関係にある場合には、その品詞NPで始まる単語を介して形容詞と名詞の間にも|ncmod|の関係があるものとし、その形容詞と名詞を|ncmod|の関係にあるものとして抽出した。

例：(|ncmod| |Cray:10_NP1| |Minneapolis-based:9_JJ|)

(|ncmod| |Research:11_NN1| |Cray:10_NP1|)
→ |ncmod| |Minneapolis-based:9_JJ| |Research:11_NN1|

形容詞と名詞が前置詞で関係付けられている場合は、前置詞がof の場合のみ|ncmod|の関係として抽出した。

³ |ncsubj| |xcomp|: 名詞と|ncsubj|の関係にある単語が、形容詞と|xcomp|の関係にある場合、その名詞と形容詞を|ncsubj| |xcomp|の関係にあるものとして抽出した。

例：(|ncsubj| |be+:12_VBR| |they:11_PPHS2| |_)
(|xcomp| |be+:12_VBR| |weak:15_JJ|)
→ |ncsubj| |xcomp| |they:11_PPHS2| (|be+:12_VBR|) |weak:15_JJ|
ただし、このとき名詞と形容詞を仲介している単語は、次の単語に限定した: be, find, become, make, seem, appear, feel, look, sound, smell, taste, remain, keep, stay, come, end-up, get, go, grow, prove, turn, turn out, wind-up, burn, lie, loom, play, plead, rest, stand, stand-up, blush, fall, fall-down, freeze, run, slam, spring, wax.(cf. Huddleston and Pullum 2002: 528, 530; Quirk *et al.* 1985: 1172)。ただし、形容詞と名詞が前置詞で関係付けられている場合は抽出しない。また、形容詞の右が *that* の場合(it is [形容詞] that ...の場合)も抽出しない。

⁴ |dobj| |xcomp|: 名詞と|dobj|の関係にある単語が、形容詞と|xcomp|の関係にある場合、その名詞と形容詞を|dobj| |xcomp|の関係にあるものとして抽出した。

例：(|dobj| |find+ing:22_VVG| |they+:23_PPH02| |_)
(|xcomp| |find+ing:22_VVG| |valuable:25_JJ|)
→ |dobj| |xcomp| (|find+ing:22_VVG|) |they+:23_PPH02| |valuable:25_JJ|
品詞 NP で始まる単語が、ある単語と|dobj|の関係にあり、かつ、名詞とも|ncmod|の関係にある場合には、その品詞 NP で始まる単語を|ncmod|の関係にある名詞と置き換えたのち、上記の|dobj| |xcomp|の関係を抽出した。
ただし、このとき名詞と形容詞を仲介している単語は、次の単語

限定用法・叙述用法をほぼ網羅的に抽出することができたので、形容詞の意味の分布をその用法にとられることなく、探ることができ、しかも限定用法と叙述用法の差異を見ることができた。

3.2 感覚領域毎の意味分布

各感覚領域毎に、どのような頻度分布をなすか、WordNet 2.0 の最も高いノード毎に限定・叙述用法別に頻度傾向を集計したのが表1である。これによると、感覚領域毎に特異的にある分類の意味の頻度が多いというようなことは起こっていないが、ピンクの色で示したように Psychological feature の割合が TASTE, VISION は多い、Event の割合が SOUND は多いなど、ある程度の差異が認められる。しかし、それぞれの感覚領域に含まれる形容詞の意味分布を個々に見ると、各形容詞がかなりばらばらの意味分布を示していることから、このような領域毎の差異は、出現数の多い形容詞の分布傾向に左右されているものと考えられる。例えば、TASTE(限定)の Psychological feature 14.1%は、きわめて出現頻度の多い *sweet* (限定: 1489)が、知覚的名詞 (*smell* (61), *taste* (37)) を多く修飾し、Psychological feature の分類が多い(16.6%)からであり、VISION の Psychological feature (限定: 14.3%、叙述: 16.7%) は、やはり出現頻度の多い *clear* (限定: 7225, 叙述: 3332) が、抽象名詞 *evidence* (180), *idea* (174), *message* (141), *indication* (136), *distinction* (130), *statement* (128)などを多く修飾し、Psychological feature を意味することが多い(限定: 23.8%、叙述: 19.8%)からと考えられる。また SOUND につ

いても、*quiet* (限定:2376), *loud* (限定:872)が特に Abstraction (*voice*, *music* 等), Event (*noise*, *explosion*, *knock* 等)の分類で多く使用されるから、この分布が多いのだと考えられる。

そこで各感覚形容詞それぞれの意味分布に注目することにし、本研究では、ほぼ同じ意味起源を持つ感覚形容詞が現代英語においてどのような意味拡張を示しているかを調査・分析する。

3.3 同じ起源を持つ2つの形容詞間の意味拡張の違い

[*hot vs. warm, cold vs. cool* の場合]

hot と *warm* は、本来の意味は温度のレベルの違いであるが、WordNet 2.0 による意味分布の階層構造の違いを見てみると、Entity の分類が *hot* は高く (*hot*: 59.6%, *warm*: 46.9%)、Abstraction の中の Relation の分類で *warm* は高くなっている (*hot*: 3.0%, *warm*: 6.3%)。 *cold, cool* も本来の意味の違いは、温度のレベルであるが、この場合も Entity の分類が *cold* は高く (*cool*: 43.8%, *cold*: 50.5%)、Abstraction の中の Relation の分類が *cool* は高くなっている (*cool*: 8.5%, *cold*: 3.0%)。これは、同じように感情を表す場合でも、*hot, cold* は *blood, sweat* など体の一部である物質の温度が高い/低いと表現することによって、その身体を持つ人の感情を表し、*warm, cool* は、*smile, reception, voice* などの人間関係に関わる領域を、我々を取り巻く空気のように見立てて直接表現することにより感情を表すように拡張しているためと考えられる。

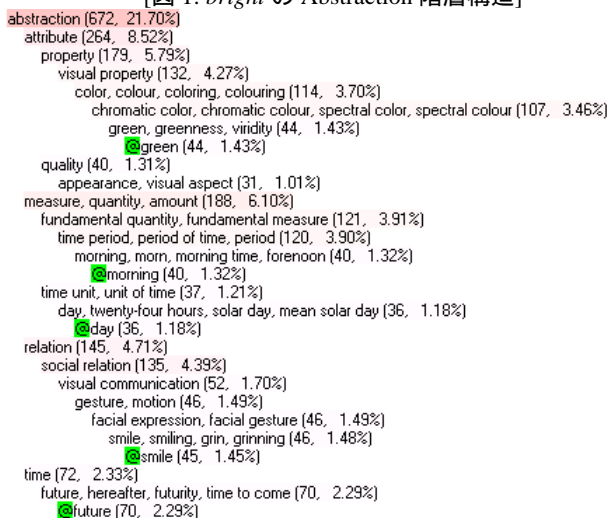
[表 1: BNC における、各感覚領域に起源を持つ形容詞の限定・叙述別意味分布]

	感覚領域	頻度	Entity	Psychological feature	Abstraction	State	Event	Human action	Group grouping	Possession	Phenomenon
限定用法	TEMPERATURE	13470	51.3%	4.7%	20.2%	5.5%	2.7%	5.8%	2.3%	0.3%	7.2%
	TOUCH	31953	39.8%	8.5%	20.6%	4.6%	5.9%	10.5%	4.2%	1.3%	4.4%
	TASTE	2748	44.4%	14.1%	20.8%	3.2%	3.5%	6.5%	4.0%	0.3%	3.1%
	DIMENSION	136744	37.4%	7.9%	24.0%	5.4%	3.9%	7.1%	8.1%	3.5%	2.5%
	VISION	19987	41.1%	14.3%	24.8%	3.3%	3.4%	6.3%	2.8%	0.4%	3.6%
	SOUND	3494	27.5%	7.9%	35.0%	5.9%	10.5%	7.7%	3.3%	0.5%	1.6%
	Total	208396	38.9%	8.5%	23.5%	5.1%	4.2%	7.5%	6.5%	2.6%	3.2%
叙述用法	TEMPERATURE	2091	56.5%	3.7%	20.2%	3.3%	1.5%	3.5%	3.9%	0.1%	7.4%
	TOUCH	4688	44.6%	7.8%	21.2%	4.4%	3.3%	7.0%	7.8%	0.9%	2.9%
	TASTE	542	38.5%	9.3%	15.9%	5.2%	6.0%	7.9%	13.4%	0.9%	3.0%
	DIMENSION	11893	36.1%	7.6%	26.6%	6.1%	3.2%	5.2%	5.5%	6.9%	2.7%
	VISION	4530	33.3%	16.7%	27.6%	5.0%	2.8%	7.3%	4.7%	0.4%	2.3%
	SOUND	792	43.0%	3.3%	29.3%	4.2%	5.6%	4.6%	7.5%	0.2%	2.2%
Total	24536	39.2%	8.9%	25.1%	5.3%	3.1%	5.8%	5.9%	3.6%	3.1%	
全体	TEMPERATURE	15561	52.0%	4.5%	20.2%	5.2%	2.6%	5.5%	2.5%	0.3%	7.2%
	TOUCH	36641	40.4%	8.4%	20.7%	4.6%	5.5%	10.1%	4.7%	1.3%	4.2%
	TASTE	3290	43.4%	13.3%	20.0%	3.6%	3.9%	6.8%	5.6%	0.3%	3.1%
	DIMENSION	148637	37.3%	7.9%	24.2%	5.5%	3.9%	7.0%	7.9%	3.7%	2.5%
	VISION	24517	39.7%	14.7%	25.3%	3.6%	3.3%	6.5%	3.1%	0.4%	3.4%
	SOUND	4286	30.3%	7.1%	34.0%	5.5%	9.6%	7.2%	4.1%	0.4%	1.7%
	Total	232932	39.0%	8.5%	23.6%	5.1%	4.1%	7.3%	6.4%	2.6%	3.2%

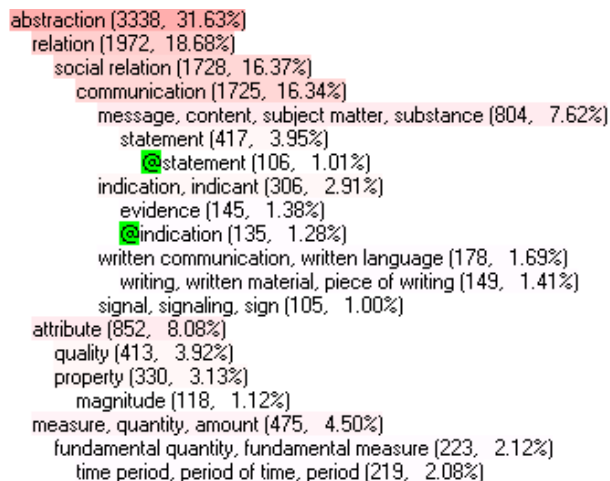
[bright vs. clear の場合]

bright, clearはもともとほとんど同じ意味起源を持つが⁵、その意味拡張は、Entityにおいてbrightは高く(52.2%)、clearは低い(22.7%)。Abstractionにおいてbrightは低く(21.7%)、clearは高く(31.6%)、その分布の構成も大きく異なる(図1, 2)。brightは人を表す名詞(boy, girl, child, pupilなど) clearは人への伝達内容を表す名詞(evidence, idea, message, indication, distinction, statementなど)を修飾・叙述することにより知性を表している。

[図1: bright の Abstraction 階層構造]



[図2: clear の Abstraction 階層構造]



[flat vs. plain の場合]

flat と plain もほぼ同じ意味起源を持つが⁶、特に

⁵ bright, clearの起源は次の通り: “Shining; emitting, reflecting, or pervaded by much light.” (OED. s.v. bright); “Expressing the vividness or intensity of light: Brightly shining, bright, brilliant.” (OED. s.v. clear)

⁶ flat, plainの起源は次の通り: “Horizontally level; without inclination.” (OED. s.v. flat); “Flat, level, even; free from elevations and depressions.” (OED. s.v. plain)

Psychological featureの分布において違いがあり、flatは低く(4.2%)、plainは高い(11.8%)。これはplainが、fact(55), truth(19), evidence(4)のように知識・情報に関わることを意味するように拡張しているためと考えられる。

4. 考察・今後の課題

以上の結果から、各感覚領域に含まれる形容詞が同じように抽象領域にメタファー写像されているわけではなく、各形容詞が人間の日常的身体経験に基づいて特徴的に意味拡張しており、そこには客観的・主観的視点等の認知能力が関わっていると言える。

本研究で用いた方法は、これまで見だし得なかった新たなカテゴリーの発見を可能にする。近年相次いでいる大規模コーパスに基づいた辞書の編纂では、その意味項目の設定は、研究者の内省により行われているが、本研究は、このような辞書の意味項目を検証することができる。

また今回は、既存の初期値として名詞を WordNet 2.0 でまず分類することによって形容詞のクラスタリングを行ったが、さらに、得られた形容詞のクラスタでもって名詞を分類することによって人手で作成している WordNet 2.0 の分類階層の検証が行える。例えば、今回抽出可能になった叙述用法でよく用いられる keen, eager の意味拡張を見ると、people, group, government, company などの集団の人間を含む分類である Group の他に、Entity 中の Person や Causal agent の分布が多くなっている。動作主としての階層を別に設けることも検討すべきと考える。

[謝辞]

本研究に用いたデータの処理に関しては、Timothy Baldwin 氏(University of Melbourne)に大変御助力をいただいた。ここに記して感謝致します。

[参考文献]

Briscoe, Ted, and John Carroll. 2002. “Robust Accurate Statistical Annotation of General Text.” *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)*, 1499-1504.

Carroll, John and Ted Briscoe. 2002. “High Precision Extraction of Grammatical Relations.” *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, 134-140.

Huddleston, Rodney D., and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

The Oxford English Dictionary. 1989. Second edition. Oxford: Oxford University Press. [OED]

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

進藤三佳、村田真樹、井佐原均 2004. 「感覚形容詞が意味拡張する抽象概念の調査・分析」『言語処理学会第10回年次大会発表論文集』、217-220.

Ullmann, Stephen 1951. *The Principles of Semantics*. Glasgow: Jackson, Son & Company.

Williams, Joseph M. 1976. “Synaesthetic Adjectives: A Possible Law of Semantic Change.” *Language* 52(2), 461-478.