

英単語の意味内容とCEFRレベル：フレーム意味論に基づいた学習用語彙リストの分析

内田, 諭
九州大学大学院言語文化研究院

<https://hdl.handle.net/2324/1933263>

出版情報：言語科学. 52, pp.75-85, 2017-03-31. Faculty of Languages and Cultures, Kyushu University
バージョン：
権利関係：



英単語の意味内容と CEFR レベル

—フレーム意味論に基づいた学習用語彙リストの分析—

内田 諭

1. はじめに

英単語を学習する際（あるいは教授する際）、道標の一つとなるのが語彙リストである。今日、新 JACET 8000¹、NGSL²、English Vocabulary Profile³、CEFR-J Wordlist⁴、ALC SVL⁵など多くの語彙リストが存在し、教育・研究の手がかりとして活用されている。語彙リストの利点は、単語が難易度順に示され、レベルごとにグルーピングされているため、学習の進捗に合わせて学ぶべき語（あるいは提示すべき語）を明確にすることができるということである。これらの語彙リストの多くは、コーパスにおける単語の頻度に依拠している。例えば、新 JACET 8000 は BNC⁶、COCA⁷ の 10 ジャンルに共通する高頻度語を基にベースリストを作成し、英語検定教科書、英語能力試験、学術入門書などから作成したコーパスを補正資料として用いている（大学英語教育学会基本語改訂特別委員会[編]2016）。多くの語彙リストではこのように単語の頻度がベースとなり、教育的な配慮を加味して配列・分類されている。語彙リストは教育現場で広く使われているものであるが、レベル別の語彙の特徴に関する検証は少なく、特に「どのような意味内容の語がどのレベルに適しているか」という点については明らかにされてこなかった。この点が明らかになると、例えば「色」を表す単語と「変化」を表す単語はどちらのほうが学びやすいかということが明確になり、特に教材選択や教材作成において有益な情報を提供することになるだろう。

本稿は、CEFR レベルと単語の意味内容の関連をフレーム意味論の観点から調査し、CEFR の各レベルに特徴的なフレームを明らかにすることを試みる。具体的には CEFR-J Wordlist の単語に対して、FrameNet⁸に基づいて意味フレームを付与し、それぞれのレベル (A1, A2, B1, B2) に特徴的なフレームを抽出する。また、フレーム間関係に基づいてフレームのタイプを一般化し、対応分析によってそれぞれのレベルに特徴的な意味タイプを明らかにする。以降、2 節で CEFR-J について概観し、3 節で FrameNet についての概要を述べる。4 節で本研究の手法と実験結果を提示し、5 節はまとめである。

¹ 大学英語教育学会基本語改訂特別委員会[編] (2016)

² New General Service List (<http://www.newgeneralservicelist.org/>)

³ <http://www.englishprofile.org/wordlists>

⁴ 『CEFR-J Wordlist Version 1.3』 東京外国語大学投野由紀夫研究室 (<http://www.cefr-j.org/download.html>)

⁵ <http://www.alc.co.jp/vocgram/article/svl/>

⁶ The British National Corpus (<http://www.natcorp.ox.ac.uk/>)

⁷ Corpus of Contemporary American English (<http://corpus.byu.edu/coca/>)

⁸ <https://framenet.icsi.berkeley.edu/fndrupal/>

2. CEFR-J の概要

CEFR (Common European Framework of Reference for Languages) はヨーロッパの複言語主義を背景に開発された言語到達度指標で、A1, A2, B1, B2, C1, C2 の6段階からなる。Aレベルは基礎段階の言語使用者、Bレベルは自立した言語使用者、Cレベルは熟達した言語使用者と定義され、言語活動を念頭に置いた Can-Do Statement で記述されているという点が特徴的である。例えば、B1レベルの一部を引用すると、「仕事、学校、職場で普段出会うような身近な話題について、標準的な話し方であれば主要点を理解できる」(日本語は吉島・大橋[編訳]2004より引用)などと規定されている。それぞれのレベルはさらにリーディング、リスニング、スピーキング・インタラクション、スピーキング・プロダクション、ライティングの5技能に下位分類され、より詳細な Can-Do 記述が提供されている。

CEFR-J は CEFR のフレームワークを基に、日本の教育環境に特化して開発されたものである(投野[編]2013)。日本における小学校から大学までの英語カリキュラムのバックボーンとなる目的で開発され、ウェブで公開されている⁹。Negishi, Takada and Tono (2012)は日本人の8割がAレベルに位置していることを指摘しているが、このことは日本人向けには CEFR の初級・中級レベルにより手厚い記述が必要であることを示している。CEFR-J では Pre-A1 レベルを設定し、A1 を A1.1, A1.2, A1.3 の3区分、また A2 を A2.1, A2.2、B1 を B1.1, B1.2、B2 を B2.1, B2.2 の下位レベルにそれぞれ分類している。細分化した独自の Can-Do 記述を開発し、教員によるアンケートや専門家による精査などを経てそれぞれのレベルに対して対応付けられている。

CEFR-J Wordlist は A1, A2, B1, B2 レベルの語彙を日本における英語教育の文脈に照らし合わせて選定したもので、ウェブサイトからダウンロードできる。Ver. 1.3 には 7814 語が含まれており、A1 には 1165 項目、A2 には 1416 項目、B1 には 2451 項目、B2 には 2782 項目含まれている。単語は品詞ベースで収録されており、例えば前置詞の above は A1、形容詞は B1 に位置付けられている。また、単語には CEFR 開発のベースとなった Threshold (van Ek and Trim 1990) で示されている概念カテゴリーの情報が付与されているが、部分的でありカバレッジは全体の約 2 割程度にとどまる。

3. FrameNet の概要

FrameNet はフレーム意味論 (Fillmore 1982, 1985 など) を基盤として開発されている「フレームの辞書」である。オンラインで公開されており、現在も構築が進行中である。2017年2月現在約 13000 の単語の語義に対して「どのフレームが喚起されるか」という情報が含まれている。例えば、buy という単語であれば Commerce_buy、sell であれば Commerce_sell というフレームをそれぞれ喚起すると規定されている。フレームとは単語の理解の礎となる背景知識を指し、buy であれば参加者(buyer)がお金(money)と引き換えに売り手(seller)から商品(goods)を得る、という一連の商取引に関わる知識があってはじめて単語の意味理解が達成されると想定されている。

フレームを喚起する最小の単位は Lexical Unit と呼ばれ、おおよそ単語の語義に対応する(厳密な定義については Ruppenhofer et al. 2010 を参照)。従って、一つの単語は複数のフレームを

⁹ <http://www.cefr-j.org/>

喚起する場合もある（例えば cook(v)は Apply_heat フレーム、Cooking_creation フレームなどを喚起する）。また、逆に特定のフレームは複数の Lexical Unit によって喚起される。例えば、Apply_heat フレームは boil(v), cook(v), fry(v)などの単語によって喚起される。この性質により、意味内容（フレーム）による単語のグルーピングが可能となる。

FrameNet におけるフレームの特徴の一つは、フレーム間関係によってフレーム同士が関連付けられていることである。フレーム間関係で最も厳密なのは Inheritance の関係である（その他の関係については Ruppenhofer et al. 2010 を参照）。この関係にあるフレームは上位フレームと下位フレームで構造が類似しており、上位フレームが下位フレームの一般化された意味タイプを示すと理解することができる。例えば Commerce_buy は上位に Inheritance の関係で Getting フレームを持つ（図 1）。この関係より「買う」という行為の上位概念は「得る」と言い換えることができる。フレーム間関係を上位にたどることさらに意味の抽象化が可能となり、より大局的な単語のグルーピングが可能となる。

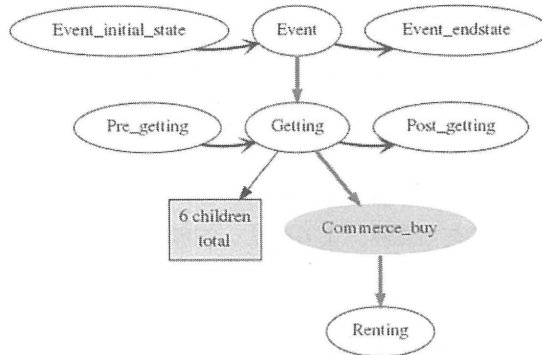


図 1 Commerce_buy フレームのフレーム間関係 (FrameGrapher より)

4.1 研究手法

本研究では、CEFR-J Wordlist に掲載された単語に意味フレームを付与し、レベルごとにその頻度を集計し、特徴フレームを抽出するという手法をとる。まず FrameNet のデータから見出し語、品詞、フレームを抽出し、CEFR-J Wordlist に対して見出しと品詞の情報から機械的にフレームを付与した。その後、対応付けの間違いを修正するため目視でリストを確認した。多義語の場合、フレームの付与が難しい場合があったが、CEFR-J Wordlist に掲載された概念カテゴリーなどを手がかりに可能な限り特定し、判断がつかない場合はフレーム付与を行わなかった。例えば、bank (A1)の場合、「銀行」、「土手」などの意味があり、前者は Business フレーム、後者は Relational_natural_features が喚起されると規定されているが、CEFR-J Wordlist の概念カテゴリーとして Things in the town, shops and shopping が提示されているため、前者の Business のフレームを採用した。また、FrameNet は現在進行中のプロジェクトであり、単語によってはフレームを付与できない場合がある。例えば、race に関しては「人種」という意味で Type フレームを喚起すると規定されているが、「(競技などの) レース」の意味では FrameNet には記載がな

い。そのため、CEFR-J Wordlist で Hobbies and pastimes のカテゴリーが振られている race (B1) は、該当フレームがないという判断になる。

これらのフレームを付与した CEFR-J Wordlist に対して、各レベルにおける特徴的なフレームを抽出するため、レベルごとの頻度を集計した。また、「出現頻度 5 以上でいずれかのレベルで占有率が 50%を超えるもの」を特徴フレームと定義し、レベルごとに抽出した（全リストは Appendix を参照）。さらにフレームの上位カテゴリーを検証するため、Inheritance の関係を用いてフレームを可能な限り一般化し、その集計結果に対して多変量解析の手法の一つである対応分析を行い、CEFR レベルと意味タイプの関係性を検証した。

4.2 結果

CEFR-J の 7814 語に対して、FrameNet をベースに意味フレームを付与できたのは 3843 件 (49.2%) であった。フレームの種類は 761 で、3843 件の単語に対して圧縮率は 19.8%である。平均して 5 語ごとに意味フレームをベースにグルーピングできたことになる。フレームが付与できた単語について Inheritance の関係を用いて上位フレームに一般化できたものは 2429 件であった（フレーム付与済みの単語数に対して約 63%）。なお、この 2429 件の単語は 39 種類のフレームに集約される（圧縮率約 1.6%）。

特徴フレーム（出現頻度 5 以上でいずれかのレベルで占有率が 50%を超えるもの）は 111 件抽出され、その内訳は A1:16 件、A2:16 件、B1:36 件、B2:43 件となった。CEFR レベルが上がるにつれて特徴フレームの数が増えており、より幅広い範囲の意味内容が出現するようになっていることが読み取れる。

4.3 CEFR レベルごとのフレームの特徴

A1 レベルの特徴フレームで最も頻度が高いものは Food フレームである。このフレームには banana, bread, cake, chicken, coffee などの語が含まれている。次に頻度が多いものは Calendric_unit で、January, February, morning, night などを含む。さらに brother, child, dad, family などが喚起する Kinship フレームが続く。これらの傾向から、A1 レベルは食事、時間、家族など日常生活において身近に感じられるものに関するフレームが特徴的であるといえることができる。

A2 レベルで最も頻度が高い特徴フレームは Direction ある。east, left, forward, north, up などがこのフレームを喚起する。また、Organization フレームも特徴フレームの一つで、agency, committee, government などが関連する単語である。さらに Posture フレーム (lie, position, stand など)、Suitability フレーム (appropriate, proper, suitable など) なども特徴的なフレームであると位置付けられる。これらのフレームは、A1 レベルと比較すると、「方向」、「組織」など英語使用者の外部にあるものと関連していることがわかる。また、「判断」を表す単語も多くなり、英語を通してより高度な知覚活動を行うことになることがわかる。

B1 で最頻出の特徴フレームは Frequency フレームである。このフレームを喚起する語には constantly, frequently, normally, occasionally などの副詞が多く含まれ、副詞表現を使ったより詳細な事態の描画がこのレベルの特徴だとわかる。また、Change_position_on_a_scale フレーム

が次いで頻度の多いフレームであるが、このフレームを喚起する単語 (decrease, fall, increase, rise など) は複雑な変化の事態を表す。例えば、In the first year-on-year increase since last July, West European car sales ROSE for April by 8.4% to 1.3 m, compared with April 1990. (FrameNet より) という文では、変化の割合 (by 8.4%) や変化の終点 (to 1.3m) などを含み、統語的にも複雑なものとなることが多い。また、Sounds フレームも特徴的で、bang, bark, whistle など、特定の状況と密接な関係を持つ語が含まれている。

B2 レベルでは Cause_harm フレームがリストのトップにくる。chop, crack, slap, smash など文脈が限定される多くの単語と関連する。また、Expertise フレーム (competence, expert, mastery など)、document フレーム (certificate, contract, testimony など)、Categorization (categorization, classification, interpretation など) など、専門性の高い文書などで使われると考えられる意味内容が特徴的に見られる。

以上の分析から、A1 レベルから B2 レベルに上がるにつれて、語彙リストの意味内容は高度化していることはわかる。日常生活に関連するフレームから始まり、発話者の身の回りのことや判断に関わるフレーム、複雑な事態を表すフレーム、専門分野と関連のあるフレームへと意味内容が具体的なものから抽象的・文脈依存的なものへと変遷している。このことは学習レベルに応じた適切な意味フレームが存在することを示しており、教材の選定や作成の際に意味内容も考慮すべきであることを示唆する。

4.4 上位フレームでの分析

次に、フレームの意味タイプをさらに一般化するため、フレームの上位レベルでの分析を行う。3 節で述べた通り、フレームはフレーム間関係によってリンクされており、それを上位にたどることでさらに大きなグルーピングが可能となる。ここでは最も厳密なフレーム間関係である Inheritance に絞って一般化に利用する。一つ例を挙げると、bay (A2) は Natural_features フレーム、campus (A2) は Locale_by_use フレーム、district (B1) は Political_locales フレームを喚起するが、これは Inheritance の関係で Locale フレームと結びついているため (図 2)、Locale タイプとしてまとめることが可能である。

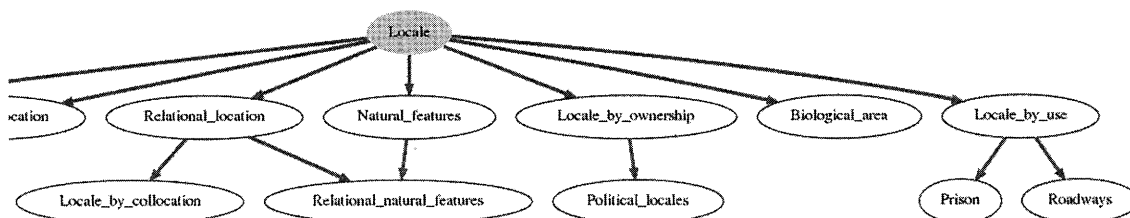


図 2 Locale フレームの Inheritance の関係 (FrameGrapher より)

本研究では、CEFR レベルと上位フレームの関係をみるため、レベルごとの上位フレームの集計表に対して多変量解析の手法の一つである対応分析を行った¹⁰。上位フレームの集計に際しては、フレーム付与済みの単語についてフレームの **Inheritance** の関係を上位にたどり、可能な限り一般化した。その上で上位フレームが存在しなかったフレームとともに集計し、総頻度が 10 以上のものを分析の対象とした。その結果、3125 件（フレーム付与済みの語の約 81%）が対象となり、これらは **Attributes, Cardinal_numbers, Entity, Event** など 48 の意味タイプに集約された。

対応分析には統計ソフト R を使用し、MASS ライブラリの `corresp` 関数によって計算した ($nf=4$)。描画には `wordcloud` ライブラリの `textplot` 関数を用い、視認性を高めるため一部のラベルの位置を調整した。結果を図 3 に示す。なお、第 1 次元 (X 軸) の寄与率は 75.77%、第 2 次元 (Y 軸) の寄与率は 16.04% で、2 次元における累積寄与率は約 91.8% と非常に高い数値を得た。

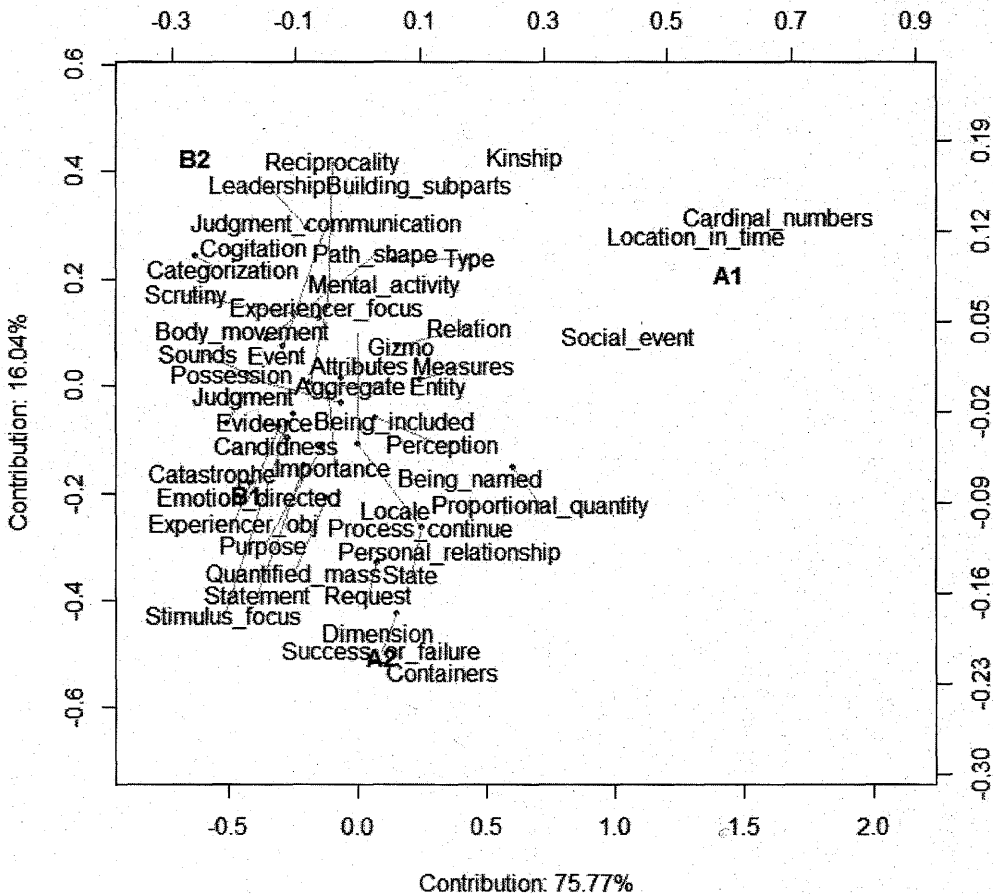


図 3 対応分析の結果

¹⁰ 対応分析は言語研究でも広く用いられ、その有効性が示されている（後藤 2007、田畑 2009 など）。主成分分析との違いについては水本（2009）に詳しい。

図3から読み取れることは、X軸（横軸）の大小がCEFRレベルの上下とほぼ一致しているということである。この値が大きければCEFRレベルは低く（A1に近づく）、逆に低ければレベルが高くなる（B2に近づく）。Y軸（縦軸）はAレベル、Bレベル内の細分化に寄与しているといえるだろう。

それぞれのレベルの付近にプロットされたフレームを観察すると、A1では *Cardinal_numbers*, *Location_in_time*, *Social_event*, *Kinship*（第1次元＋、第2次元＋）などが見られ、前節における観察結果と符合する。A2は *Dimension*, *Success_or_failure*, *Container*, *Request*（第1次元＋、第2次元－）などの意味タイプが特徴的で、やはり発話者の観察や判断、社会的な関わりなどを表す意味内容が特徴的であることが読み取れる。また、B1は *Statement*, *Catastrophe*, *Quantified_mass*（両次元－）、B2は *Leadership*, *Judgment_communication*, *Categorization* フレーム（第1次元－、第2次元＋）が密接に関連していることがわかる。これらはAレベルよりも詳細な状況を表したり、抽象的な内容を表すフレームであるといえる。

5. 結語

英単語が喚起するフレームを特定し、CEFRレベルごとに集計することで各レベルの意味的な特徴が明らかになった。具体的な意味内容を表す単語が初級段階に登場し、それが使用者の外部・思考を表す単語へと移行し、抽象的・専門的な領域へと発展していく。およそ直感に沿う結果であるが、フレームによって意味内容が理論的に「具体化」されたことに大きな意味がある。これにより、各CEFRレベルに特化した教材の開発やタスクの選定などの際に適切な意味内容やトピックを選択する一つの基準となるだろう。

本研究はFrameNetに依拠したため、分析の範囲はFrameNetに掲載されている語にとどまる。今後、フレームを付与できなかった語を含めて包括的な検証を行う予定である。また、CEFR-J WordlistはA1~B2までの4レベルに限定されているが、English Vocabulary Profileなど他のリストも併用してC1, C2レベルの特徴も明らかにしたいと考えている。

謝辞

本研究はJSPS科研費（15K16798）「CEFRレベルと意味内容の対応付け：フレーム意味論の観点から」の助成を受けたものである。

References

- 大学英語教育学会基本語改訂特別委員会[編](2016)『大学英語教育学会基本語リスト：新JACET8000』東京：桐原書店。
- Fillmore, C. J. (1982) Frame semantics. In Yang, I. (ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981*. pp.111-137. Seoul: Hanshin.
- Fillmore, C. J. (1985) Frames and the semantics of understanding. *Quaderni di Semantica*, 6 (2). pp.222-254.

- 後藤一章 (2007) 「統語機能別頻度分布に基づく名詞的特徴的コロケーションの発見」『多変量解析を用いたテキスト分析研究』統計数理研究所共同研究レポート 201. pp.1-23.
- 水本篤 (2009) 「コーパス言語学研究における多変量解析手法の比較—主成分分析 vs. コレスポネンデンス分析—」『コーパス言語学研究における量的データ処理のための統計手法の外観』統計数理研究所共同研究レポート 232. pp.53-64.
- Negishi, M., T. Takada and Y. Tono (2012) A progress report on the development of the CEFR-J. *Studies in Language Testing*, 36. pp.137-157.
- Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. R. Johnson and J. Scheffczyk (2010) *FrameNet II: Extended Theory and Practice*. <<https://framenet2.icsi.berkeley.edu/docs/r1.5/book.pdf>> [Accessed: Feb. 7, 2017].
- van Ek, J. A. and J. L. M. Trim (1991) *Threshold 1990*. Cambridge: Cambridge University Press.
- 田畑智司 (2009) 「*Gentleman in Dickens*—多変量アプローチで見る文体意匠としてのコロケーション—」『多変量アプローチによるテキストの計量研究』統計数理研究所共同研究レポート 231. pp.1-22.
- 投野由紀夫[編](2013) 『CAN-DO リスト作成・活用：英語到達度指標 CEFR-J ガイドブック』東京：大修館書店.
- 吉島茂・大橋理枝[編・訳](2004) 『外国語の学習、教授、評価のためのヨーロッパ共通参照枠』東京：朝日出版社.

Appendix: CEFR レベルごとの特徴フレーム（総計数順）

特徴レベル	Frame	A1	A2	B1	B2	総計
A1	Food	26	10	8	7	51
A1	Calendric_unit	30	5	3	2	40
A1	Kinship	16	3	4	9	32
A1	Cardinal_numbers	19	3			22
A1	People	8		5	1	14
A1	Social_event	7	1	3		11
A1	Color	8	1			9
A1	Measure_duration	6	2		1	9
A1	Performers_and_roles	5	1	2	1	9
A1	Age	4	3			7
A1	Noise_makers	4	1		2	7
A1	Animals	6				6
A1	Aesthetics	4			1	5
A1	Artifact	3	1		1	5
A1	Intentionally_act	3	1	1		5
A1	Temporal_subregion	3	1		1	5
A2	Direction	2	7			9
A2	Organization	1	5	2		8
A2	Posture	1	4	1	2	8
A2	Suitability	1	5	1	1	8
A2	Attempt	1	4	2		7
A2	Attack	1	3	1	1	6
A2	Biological_area		4	2		6
A2	Contacting	2	4			6
A2	Custom	1	3	1	1	6
A2	Finish_competition		3	2	1	6
A2	Hair_configuration	1	3	1	1	6
A2	Level_of_force_exertion	1	3	1	1	6
A2	People_by_age	2	3		1	6
A2	Predicament	1	3	1	1	6
A2	Subjective_influence		3	2	1	6
A2	Intentionally_create		3		2	5
B1	Frequency	5	4	13	1	23
B1	Change_position_on_a_scale	1	2	12	6	21
B1	Sounds		2	8	6	16
B1	Perception_active	4		8	2	14
B1	Catastrophe		1	7	3	11
B1	Substance	1	3	6	1	11
B1	Chemical-sense_description	3		5	2	10
B1	Attaching		1	5	3	9
B1	Capability		1	5	2	8

特徴レベル	Frame	A1	A2	B1	B2	総計
B1	Departing		2	4	2	8
B1	Fields	1		4	3	8
B1	Cause_to_amalgamate	1	1	4	1	7
B1	Contingency		2	4	1	7
B1	Deny_or_grant_permission		2	4	1	7
B1	Obviousness		2	5		7
B1	Being_obligated		1	4	1	6
B1	Cause_change_of_position_on_a_scale		1	3	2	6
B1	Communicate_categorization	1	2	3		6
B1	Competition		2	3	1	6
B1	Destroying		1	3	2	6
B1	Domain			3	3	6
B1	First_rank		1	4	1	6
B1	Ingestion	2		3	1	6
B1	Judgment_direct_address	2		3	1	6
B1	Motion_noise		2	3	1	6
B1	Partiality		1	3	2	6
B1	Being_at_risk		1	3	1	5
B1	Being_employed	2		3		5
B1	Breathing	1		3	1	5
B1	Correctness	1		3	1	5
B1	Directional_locative_relation	2		3		5
B1	Grooming	1		3	1	5
B1	Impact		1	3	1	5
B1	Participation			3	2	5
B1	Prevarication			4	1	5
B1	Secrecy_status			5		5
B2	Cause_harm	3	2	3	13	21
B2	Placing	2	1	6	9	18
B2	Expertise	1	4	2	8	15
B2	Documents		2	4	7	13
B2	Categorization		1	5	6	12
B2	Cause_expansion			4	7	11
B2	Cotheme		1	4	6	11
B2	Sufficiency	1	3	1	6	11
B2	Expectation		3	2	5	10
B2	Filling	1	2	2	5	10
B2	Removing	1		2	7	10
B2	Cause_change	1		3	5	9
B2	Cure			4	5	9
B2	Architectural_part	2	1	1	4	8
B2	Duration_description	1		3	4	8
B2	Manufacturing		3	1	4	8

特徴レベル	Frame	A1	A2	B1	B2	総計
B2	Morality_evaluation	1		1	6	8
B2	Rite	1		3	4	8
B2	Supply		1	3	4	8
B2	Communication_noise		1	2	4	7
B2	Compliance		1	2	4	7
B2	Assessing	1			5	6
B2	Communication_manner	1		1	4	6
B2	Distributed_position	1	1	1	3	6
B2	Earnings_and_losses	1	1	1	3	6
B2	Employing	1	1		4	6
B2	Emptying			2	4	6
B2	Legality		1	2	3	6
B2	Make_noise	1		1	4	6
B2	Protecting	1		2	3	6
B2	Addiction				5	5
B2	Change_of_leadership		1	1	3	5
B2	Compatibility	1	1		3	5
B2	Hindering			2	3	5
B2	Manipulate_into_doing				5	5
B2	Part_piece	1	1		3	5
B2	Perception_body	1		1	3	5
B2	Perception_experience	1		1	3	5
B2	Progression			2	3	5
B2	Prohibiting_or_licensing			2	3	5
B2	Strictness	1		1	3	5
B2	Theft		1	1	3	5
B2	Willingness			2	3	5