比喩的表現と読み時間について

浅原 正幸　国立国語研究所　masayu-a@ninjal.ac.jp

概要

本研究は，日本語の読み時間と，広い意味での比喩的表現との関係を検討するものである。比喩言語処理に関する先行研究では，比喩的表現は追加的な認知的コストを課し，読み時間を延長させる可能性があると指摘されている [1, 2] 一方，文脈的・構造的要因によっては，比喩言語が字義通りの表現と同程度に効率的に処理されうることも示されている [3, 4]。本研究では客観的な結果を得るため，自己ペース読み課題に基づく BCCWJ-SPR2 コーパスの読み時間データと，比喩表現に関する包括的なアノテーション資源である BCCWJ-Metaphor を用いて評価を行った。分析の結果，比喩表現の開始位置および内部において特有の読解パターンが観察された。

1　はじめに

比喩的言語は，自然な談話における中心的な特徴であり，隠喩やイディオムなどの非字義通りの表現を通じて，話し手が複雑な意味を簡潔に伝えることを可能にする。心理言語学では，このような表現が字義通りの言語と比べて追加的な認知コストを課すのか，あるいは条件によっては理解を促進しうるのかについて，長年にわたり議論が続いてきた。初期の研究は，比喩的表現は余分な解釈ステップを要し，処理時間が長くなると主張している [1, 2]。これに対して，後続の研究では，慣習化された隠喩やイディオムは，文脈的または語彙的手がかりが十分に与えられる場合，字義通りの表現と同程度，あるいはそれ以上に効率的に処理されうることが示されている [5, 6, 3, 4]。神経認知的証拠も両方の見解を支持しており，比喩的意味に対しては神経活動の増大が報告される一方 [7, 8]，慣習的な比喩形式に対しては統合コストの減少が報告されている [9]。

このような議論は英語や他のヨーロッパ諸語で広く検討されてきたが，日本語における比喩処理を大規模に扱った研究は比較的少ない。その大きな要因として，心理言語学的証拠と比喩表現の詳細なアノテーションを同時に備えた資源の不足が挙げられる。しかし近年のコーパス開発の進展により，そのような分析が可能になってきた。現代日本語書き言葉均衡コーパス（Balanced Corpus of Contemporary Written Japanese; BCCWJ; [10]）は，複数の方向に拡張されている。BCCWJ-SPR2 は，理解度の測定を含む大規模自己ペース読み時間データを提供する [11] 一方，BCCWJ-Metaphor は，比喩表現に対する専門家およびクラウドソーシング型の包括的アノテーションを提供する [12, 13]。

本研究では，これら二つの資源をリンクさせ，日本語における比喩表現が読み時間にどのような影響を与えるかを検討する。具体的には，両コーパスに共通するレジスタである *PB*（書籍） を対象とし，比喩アノテーションを自己ペース読み時間と整合させるために，比喩表現の BIO ラベルを文節境界へ再割り当てした。この設計により，結合比喩・換喩・提喩・文脈比喩といった異なる種類の比喩表現が，オンラインの読解過程にどのような影響を与えるかを詳細に検討することが可能となる。

本研究の目的は，大規模な読み時間データと精緻な比喩アノテーションを統合することで，日本語における比喩言語処理を体系的かつコーパスベースに検証することである。これにより，比喩表現が一般に理解を妨げるのか，あるいはその種類や文中の位置によっては理解を促進しうるのかを明らかにすることを目指す。

2　リサーチクエスチョン

本稿では，以下の研究課題を扱う。

1. 日本語における比喩表現は，系統的に非比喩的表現よりも読み時間を増大させるのか。それとも場合によっては処理コストを低減させることがあるのか。
2. 比喩的表現の種類（結合比喩・換喩・提喩・文脈比喩）は，読み時間への影響という点でどのように異なるのか。
3. 比喩的表現が含まれる文節内部での位置（先頭 vs. 内部）は，処理コストにどのような違いをも

たらすのか。

本研究の貢献は三点にまとめられる。第一に，コーパスベースの心理言語学データを用いて，日本語における比喩処理を大規模に定量分析した初の研究である点である。第二に，比喩の効果は一様ではなく，表現の種類と文中の位置に体系的に依存することを示す。第三に，比喩アノテーションを自然な文脈における読み時間と統合することで，コーパス言語学と心理言語学実験を架橋する方法論的枠組みを提示する。これらの貢献を通じて，比喩言語の認知処理に新たな光を当てるとともに，非字義通りの意味に関する言語間比較の視点を拡張する。

3 関連研究

比喩表現の処理に関する研究は，大きく二つの対照的な立場を提示してきた。一方の立場では，比喩言語は字義通りの言語よりも多くの認知的努力を要するとされる。たとえば [1] は，隠喩やイディオムが字義通りの表現より処理に時間を要することを示し，[2] は，比喩理解の三段階モデルを支持する心理言語学実験の証拠を提示している。これらの結果は，読者がまず字義通りの解釈を試み，それが破綻したときにはじめて比喩的解釈へと移行するという語用論的な立場 [14, 15] と整合的である。神経認知研究でも，比喩処理中に前頭葉や側頭葉の活動が増大することが報告されており [7, 8]，比喩言語がしばしば追加的な処理コストを課し，読み時間の延長をもたらすことが示唆されている。

これとは対照的に，もう一方の研究潮流は，比喩言語が理解を促進しうる条件に焦点を当てる。Giora のサリエンス理論 [5, 6] は，特に慣習化された顕著な比喩表現は記憶内で語彙項目として保存されており，効率的に処理されると主張する。イディオム研究でも，イディオムが統語的に複雑な字義通りの句よりも一体の語彙項目としてアクセスされ，場合によってはより速く処理されることが示されている [16, 17, 3]。文脈による促進効果も報告されており，強い文脈的手がかりが与えられた場合，比喩表現は字義通りの表現よりも素早く統合される場合がある [18, 4]。ERP 研究も，慣習的な隠喩に対して N400 振幅の減少が観察されることから，意味統合の円滑さを示している [9]。理論的には，比喩表現は複雑な概念をコンパクトに圧縮する「認知的に経済的な」資源とみなされ，適切な文脈では処理負荷をむしろ軽減しうると論じられてきた [19, 20]。

総じて先行研究は，比喩表現が常に字義通りの言語よりも負荷が高いわけではなく，表現の種類，文脈的支援，認知的サリエンスといった要因に依存して，処理コストが増加する場合と減少する場合の両方が存在することを示している。

4 資料

日本語における比喩表現と読み時間の関係を検討するため，本研究では相補的な視点を提供する三つの資源——自己ペース読み時間コーパス，語義アノテーション付きコーパス，比喩表現データベース——を用いる。これらを組み合わせることで，読解行動を意味・比喩アノテーションと結びつけて分析することが可能になる。

4.1 現代日本語書き言葉均衡コーパス (BCCWJ)

現代日本語書き言葉均衡コーパス (BCCWJ) は，現代日本語の多様な書き言葉レジスタを代表するよう設計された 1 億語規模のコーパスである [10]。書籍，雑誌，新聞，白書，教科書など，多様な領域から収集されたテキストサンプルから構成されており，言語記述および実証的研究の双方に利用できるバランスのとれた資源である。BCCWJ は，その後の BCCWJ-SPR2，BCCWJ-WLSP，BCCWJ-Metaphor などの派生資源の基盤となっており，これらの資源はそれぞれ心理言語学的，意味的，比喩的なアノテーションを付与することでコーパスを拡張している。

4.2 BCCWJ-SPR2

BCCWJ-SPR2 は，理解度質問を伴う自己ペース読み実験によって収集された大規模な読み時間データセットである [11]。刺激文は，現代日本語書き言葉均衡コーパス（BCCWJ）に含まれる書籍・教科書・白書から抽出された。2,000 名以上の参加者が実験に参加し，語ごとの読み時間と理解度のデータが得られている。語レベルのデータに加えて，本コーパスでは日本語の基本的な句単位である 文節 レベルでも読み時間が集約・評価されている。文節レベルでの評価は，日本語文処理における基本単位を反映しており，語単位だけでは観察できない認知処理の側面を捉えるうえで特に有用である。したがって，本コーパスは比喩表現を含む多様な言語的特徴が読解行動に与える影響を検討するための強固な経験的基盤を提供する。

4.3 BCCWJ-WLSP

BCCWJ-WLSP は，国立国語研究所が開発した大規模日本語シソーラス『分類語彙表（Word List by Semantic Principles; WLSP』[21] に基づく意味ラベルを BCCWJ に付与した資源である。書籍，雑誌，新聞に含まれる 18 万語以上の内容語が，WLSP の意味カテゴリに従って手作業で語義判別・アノテーションされている。UniDic による形態素情報を WLSP カテゴリと対応づけることで，複数レジスタにまたがる精緻な意味情報が付与されている。BCCWJ-WLSP は，比喩表現の候補を同定するための意味的レイヤとして機能し，後述する BCCWJ-Metaphor における比喩アノテーションの準備段階として開発された。

4.4 BCCWJ-Metaphor

BCCWJ-Metaphor は，比喩言語に特化して開発された資源であり，加藤・菊地・浅原によって構築された [12, 13]。BCCWJ のうち書籍・雑誌・新聞のサンプルを対象とし，専門家アノテーションとクラウドソーシングによる評価を組み合わせている。比喩表現については，その位置，種類，可解釈性などがアノテーションされるとともに，クラウドワーカーによる評価は非専門読者が当該表現をどのように知覚・処理するかに関する示唆を与える。このように専門的分析と大規模な主観評価を組み合わせたアプローチにより，BCCWJ-Metaphor はコーパスベース研究と心理言語学実験を橋渡しするうえで重要な資源となっている。

さらに BCCWJ-Metaphor では，比喩表現を複数のタイプに細分類している。具体的には，選択制限の違反をともなう結合比喩・選択制限の違反をともなわない文脈比喩・換喩・提喩などである。各表現については，国語研短単位（Short Unit Word）区切りに基づき，BIO ラベリング方式で構成要素がアノテーションされる。これらの構成要素は *Metaphor Related Words (MRWs)* と呼ばれ、換喩・提喩については複数ラベルの付与が許される。

4.5 読み時間と比喩的表現アノテーションとの統合

本研究では，特に **BCCWJ-SPR2** と **BCCWJ-Metaphor** の重複部分に焦点を当てる。両コーパスはいずれも BCCWJ を基盤としているが，両者に共通するレジスタは書籍（Books）のみである。そのため，本研究の分析は書籍レジスタのサンプルに限定され，読み時間データと比喩表現アノテーションの比較可能性を担保している。この整合化により，自己ペース読み時間から得られた経験的証拠と，比喩言語に関する専門家・クラウドソーシング評価とを直接リンクさせ，比喩表現が日本語のオンライン読解過程に与える影響を検証することが可能になる。

さらに，BCCWJ-SPR2 の読み時間は 文節（*bunsetsu*） レベルで評価されているため，BCCWJ-Metaphor における比喩表現の BIO ラベルを文節境界に再割り当てした。具体的には，比喩表現の開始位置が文節の先頭と一致する場合にのみ，当該文節に B ラベルを付与し，比喩表現の内部に位置する文節には I ラベルを付与する形で再投影を行った。これにより，比喩関連語のアノテーション（MRWs）と，日本語文処理の心理言語学的単位である文節レベルの読み時間が同じ粒度で整合するようにした。

なお，BCCWJ を用いて読み時間と統語・意味カテゴリの相関を分析した関連研究として [22] などがあり，日本語におけるコーパスベース心理言語学アプローチの有効性を示している。

4.6 データライセンス

本研究で用いた BCCWJ 関連の資源はすべて，BCCWJ 購入者向けの配布条件に従って利用した。BCCWJ-SPR2 と BCCWJ-WLSP は BCCWJ 購入者に対して配布される。BCCWJ-Metaphor についても，元論文の刊行後，同様に購入者限定での配布が見込まれている。本研究では生データの再配布は行わず，再現に十分な集計統計量のみを提供する。

4.7 データプライバシーと安全性

本研究で用いたすべての資料は，BCCWJ に由来する公開された書き言葉テキストである。著者らの知る限り，これらの資料には人間被験者から収集された個人情報（PII）は含まれていない。また，本研究ではいかなる段階でもプライベートなユーザデータを処理していないそのため，匿名化やマスキングなどの作業を行う必要はなかった。タスク関連メタデータについても，個人を特定しうる情報を含めずに保存している。

5　データ処理と回帰分析

5.1　データ処理

本研究のデータセットは，**BCCWJ-SPR2** 書籍 (PB) と **BCCWJ-Metaphor** の重複部分から構成され，43 サンプル，111,983 語，40,974 文節を含む。比喩的表現アノテーションは，結合比喩（B = 864, I = 1,189），換喩（B = 297, I = 234），提喩（B = 58, I = 35），文脈比喩（B = 351, I = 298）を含む。これらのアノテーションは国語研短単位（Short Unit Word）レベルの BIO ラベルとして符号化されたものを，文節境界に再割り当てした。再割り当てに際しては，比喩表現の開始位置と文節の先頭が一致する場合にのみ B ラベルを付与し，継続部分に I ラベルを付与することで，文節レベルの読み時間と比喩アノテーションの整合性を担保した。

より具体的には，短単位語レベルの BIO タグを文節境界へ投影した。文節の先頭が比喩スパンの開始点と一致する場合は B，進行中のスパン内部に位置する場合は I，それ以外は O を付与した。比喩タイプが重なり合う場合もあるため，複数のラベルを同時に付与することを認めた。

読み時間データについては，二段階のクリーニング手続を適用した。第一に，文節レベルで，個々の読み時間が 100 ms 未満または 3,000 ms を超えるものを除外した。第二に，セッションレベル（1 名の参加者が 1 サンプルを読む単位）で，平均読み時間が 150 ms 未満または 2,000 ms を超えるセッションを丸ごと除外した。これらの手続きにより，局所的な外れ値と異常なセッションの双方が除去され，後続の回帰分析に適したデータセットが得られた。

最終的に，本データセットには異なり 404 名の参加者が含まれ，参加者–サンプルの組み合わせ（セッション）数は 6,875，クリーニング後の文節レベルのデータポイント数は合計 6,054,402 となった。

例として，表 1 には，サンプル `PB11_00006`（Sentence_ID = 37; subj_id = 72）から抽出した 4 文節分のスパンを示し，文節レベルの自己ペース読み時間とカテゴリ別 BIO ラベルを併記している。比喩スパンは文節 「一例を」から始まり，COMBI=**B** および METO=**B** が付与され，続く 「とるなら」 には COMBI=**I** と METO=**I** が付与される。一方，その後の文節は比喩領域の外側にあり，すべて **O** となっている。表には対応する読み時間（269, 328, 279, 364 ms）も示されており，比喩 BIO ラベルが再割り当て後に文節の先頭・内部にどのように整合されるか，また COMBI+METO のような複数ラベルがどのように表現されるかを例示している。

5.2　回帰分析

比喩表現が読み時間に与える影響を検討するため，文節（base phrase）レベルで線形回帰分析を行った。従属変数は対数変換した読み時間であり，反応時間分布を正規化するための標準的な手法である。独立変数としては，BCCWJ-Metaphor における比喩表現カテゴリ（Combination Metaphor, 換喩, 提喩, Contextual Metaphor）に加え，文節長（文字数・語数）や文中位置といった統制変数を導入した。

文節レベルの読み時間をモデル化するにあたり，比喩表現の効果を構造的要因を統制したうえで定量化することを目的とした。比喩関連の予測変数としては，文節先頭と内部の位置を区別した二値指標――$\textsc{Combination}_B$，$\textsc{Combination}_I$，$\textsc{Metonymy}_B$，$\textsc{Metonymy}_I$, $\textsc{Synecdoche}_B$, $\textsc{Synecdoche}_I$, $\textsc{Context}_B$, $\textsc{Context}_I$――を導入し，非比喩文節をベースラインとした。換喩・提喩ラベルは多重付与されているため，これらの指標は互いに排他的ではない。

また，読者間およびテキスト間のばらつきを考慮するため，被験者とテキストサンプルにランダム切片を持つ線形混合効果モデルを適用した（図 1 を参照）。

ここで，$u_{p[i]} \sim \mathcal{N}(0, \sigma_u^2)$ と $v_{s[i]} \sim \mathcal{N}(0, \sigma_v^2)$ はそれぞれ被験者 p およびサンプル s に対するランダム切片，$\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ は残差を表す。

予測変数のうち，最初のグループは読解材料の構造的要因を捉える。*Sentence_ID* はサンプル内における文の通し番号を表し，*Bunsetsu_ID* は文内における文節の通し番号を表す。これらの要因により，読者がテキストや文の後半に進むにつれて読み時間が短くなるという，一般的な適応・慣れの効果をモデル化できる。

Bunsetsu_length は各文節の文字列長を表し，より長い正書法単位に対して読み時間が増加する傾向を捉える。

DepPara_depnum は BCCWJ の係り受けアノテーション [23] に基づく指標であり，現在の文節に係る文節数を表す。日本語の統語構造では，係り元が一般にその係り先に先行するという特徴があり，多くの係り元を持つ文節は読者にとって予測しやすくな

表 1　例：PB11_00006（Sentence_ID = 37）における文節レベル読み時間（ms）とカテゴリ別比喩 BIO ラベル：subj_id = 72

表層（文節）	Sentence_ID	Bunsetsu_ID	RT (ms)	COMBI	METO	SYNEC	CONT
一例を	37	1	269	B	B	O	O
とるなら	37	2	328	I	I	O	O
ポケモンカードの	37	3	279	O	O	O	O
件です。	37	4	364	O	O	O	O

注. RT は自己ペース読み課題における文節レベルの読み時間を表す。BIO ラベルは BCCWJ-Metaphor におけるカテゴリ別ラベルであり，複数ラベル付与が許される。ここに示したデータは SPR2–Metaphor の整合結果から抽出したものである。

$$
\begin{aligned}
\log(\mathrm{RT}_i) \;=\; & \beta_1\,\mathrm{Sentence_ID} + \beta_2\,\mathrm{Bunsetsu_ID} \\
& + \beta_3\,\mathrm{Bunsetsu_length} + \beta_4\,\mathrm{DepPara_depnum} \\
& + \beta_5\,\textsc{Combination}_B + \beta_6\,\textsc{Combination}_I \\
& + \beta_7\,\textsc{Metonymy}_B + \beta_8\,\textsc{Metonymy}_I \\
& + \beta_9\,\textsc{Synecdoche}_B + \beta_{10}\,\textsc{Synecdoche}_I \\
& + \beta_{11}\,\textsc{Context}_B + \beta_{12}\,\textsc{Context}_I + \alpha \\
& + u_{p[i]} + v_{s[i]} + \varepsilon_i,
\end{aligned}
$$

図 1　線形混合効果回帰モデルの式

る場合がある。この予測変数は，このような統語的予測可能性が読み時間に与える影響を捉える。

分析手続としては二段階のアプローチをとった。まず，クリーニング済みデータセット（点ごとのトリミングとして 100 ms 未満・3000 ms 超の読み時間を除外，セッションごとのトリミングとして平均読み時間が 150 ms 未満または 2000 ms 超のセッションを除外）から，REML によりモデルを推定した。ついで標準化残差を算出し，|residual| > 3 SD の観測値を除外したうえで，再度モデルを推定した。このステップで 91,579 観測値（全体の 1.51%）が除外され，最終的に 5,962,823 件の文節レベル観測値が分析に用いられた。

このモデル仕様は，予測変数と従属変数の粒度を整合させるものである。比喩 BIO ラベルは文節境界に再割り当てされ，比喩表現の開始位置が文節先頭と一致する場合にのみ B ラベルを付与した。これにより，文節レベルの読み時間との整合性が担保される。得られたモデルは，各比喩カテゴリが非比喩ベースラインと比べてどの程度読み時間を増減させるかを示す回帰係数を提供し，その際，長さ，位置，ランダムなばらつきを統制している。

表 2 に示すように，統制変数は期待通りの挙動を示した。*Sentence_ID*（サンプル内の文の通し番号）と *Bunsetsu_ID*（文内の文節の通し番号）はいずれも有意な負の効果を持ち，読者がテキストや文の後半へ進むにつれて読み時間が短くなる一般的な適応効果を捉えている。*Bunsetsu_length*（文節の文字列長）は読み時間と正の関係を示し，より長い文字数ほど処理負荷が高まることを反映している。*DepPara_depnum*（BCCWJ の係り受けアノテーション [23] に基づく，現在の文節に係る文節数）は負の効果を示し，多くの依存文節を持つ文節ほど読み時間が短くなる傾向が見られた。これは，日本語において依存項が通常その支配語に先行するという性質により，当該文節が予測しやすくなるためと解釈できる。

比喩関連の予測変数に目を向けると，比喩表現を含む文節は，非比喩的な文節と比べて異なる処理パターンを示すことが明らかになった。より具体的には，結合比喩は表現の先頭で読み時間を延長させる一方，表現内部では読み時間を短縮させる傾向が見られ，比喩フレームを確立する過程で初期的な統合コストが生じるものの，いったんフレームが確立されると，その後の処理が促進されることを示唆している。換喩は全体として読み時間の延長と関連しており，意味統合に対する継続的な負荷を反映している可能性がある。これに対し，提喩は先頭位置でのみ読み時間の短縮が見られた。文脈比喩については，表現の開始位置において読み時間の延長が観察され，文脈から比喩解釈を導くための追加的な推論が必要であることと整合的である。

被験者およびテキストサンプルに対するランダム切片を導入することで，個人差およびテキスト固有の違いが適切に統制された。総じて，本回帰分析は，比喩表現が一様に処理負荷を増大させるのでは

表 2　線形混合効果回帰の結果

	従属変数:
	SPR_log_reading_time
Sentence_ID	−0.001***
	(0.00000)
Bunsetsu_ID	−0.003***
	(0.00002)
Bunsetsu_length	0.015***
	(0.0001)
DepPara_depnum	−0.012***
	(0.0001)
COMBINATIONB	0.004***
	(0.001)
COMBINATIONI	−0.007***
	(0.001)
METONYMYB	0.011***
	(0.001)
METONYMYI	0.006***
	(0.001)
SYNECDOCHEB	−0.007**
	(0.003)
SYNECDOCHEI	−0.001
	(0.004)
CONTEXTB	0.003***
	(0.001)
CONTEXTI	−0.002
	(0.001)
Constant	5.704***
	(0.014)
観測数	5,962,823
対数尤度	−460,486.100
赤池情報量規準	921,002.200
ベイズ情報量規準	921,206.200
注:	*p<0.1; **p<0.05; ***p<0.01

なく，その種類と文中位置に応じて処理コストが異なることを示している（表 2 を参照）。

5.3　ハードウェアと処理時間

すべての実験は，Intel Xeon W-2225 @ 4.10 GHz CPU と 64 GB RAM を搭載したワークステーション（GPU なし）上で実行した。この環境において，本研究のエンドツーエンドの処理パイプラインは 1 回の実行につき 1 分未満で完了した。このことから，本手法の計算コストは比較的低く，標準的なラップトップクラスの CPU 環境でも再現可能であるといえる。

6　考察

本研究の結果は，比喩言語が追加的な処理コストを課すのか，あるいは特定の条件下では理解を促進しうるのかという古くからの議論に対して，よりニュアンスに富んだ視点を提供する。従来の心理言語学的・神経認知的研究は，隠喩やイディオムが字義通りの表現よりも処理が困難であり，追加的な解釈ステップを要したり，前頭葉や側頭葉の活動が強まることを示してきた（例：[1, 2, 7, 8]）。これらの結果は，読者がまず字義通りの解釈を試み，それが不適切であることが判明した後に比喩解釈へと移行するという「字義通り優先」モデル [14, 15] を支持するものである。

一方で，他の研究は，比喩表現が効率的に処理されうる条件に注目してきた。[5, 6] によるサリエンス理論は，慣習化された顕著な隠喩は語彙項目として記憶に保存されており，迅速にアクセスされると主張する。同様に，イディオムに関する研究 [16, 17, 3] では，イディオムがしばしば同程度の構造的複雑さを持つ字義通りの句よりも速く処理されることが示されている。文脈による促進効果を検証した研究でも，強い文脈的期待が与えられた場合，比喩表現は字義通りの表現よりも円滑に統合されることが報告されている [18, 4]。ERP 研究は，慣習的な隠喩に対して N400 の振幅が減少することを示し，意味統合コストの軽減を裏づけている [9]。

こうした背景のもと，本研究の結果は，比喩表現の種類と位置に応じて異なるパターンが見られることを明らかにした。結合比喩は表現の先頭で読み時間を延長させる一方，表現内部では読み時間を短縮させる傾向があり，比喩フレームを確立するための初期コストと，フレームが活性化された後の処理促進という二段階の過程を示唆する。換喩は全般に読み時間の延長と結びついており，意味統合に対する持続的な負荷を反映している可能性がある。これに対して提喩は，先頭位置でのみ読み時間の短縮が観察され，文脈比喩は表現開始位置で読み時間を延長させることから，文脈に基づく比喩解釈の導出に追加的な推論が必要であることを示している。これらの結果は，いくつかの比喩形式（例えば結合比喩の内部）では処理が効率的でありうる一方，他の形式（例えば文脈比喩）では統合コストが増大するという点で，サリエンス理論と部分的に整合的である。

さらに本研究は，コーパス言語学と心理言語学をどのように橋渡しできるかを具体的に示した。大規模で自然な文脈における比喩アノテーションを実験的な読み時間データと整合させることで，従来の実験室ベースの研究では困難であったスケールと粒度で，比喩表現の処理コストを検討することが可能になった。比喩情報ラベル付きコーパスの構築も，自己ペース読み時間データの収集も時間と労力を

要するが，両者を統合することで，数百万件規模のデータポイントに基づき，比喩の種類が実時間の理解に与える影響を体系的にモデル化できる。このような方法論的シナジーは，比喩言語の微妙な認知的効果を，生態学的妥当性と定量的厳密性の双方を備えた形で検討する新たな道を開くものであり，将来的な言語間比較やマルチモーダル研究の基盤を提供する。

7　本研究の限界と今後の課題

本研究は，大規模自己ペース読み時間データと精緻な比喩アノテーションを活用しているものの，いくつかの制約が存在する。第一に，分析対象をBCCWJ-SPR2 と BCCWJ-Metaphor に共通するレジスタである書籍に限定している点である。この制約により，本研究の結果は新聞や雑誌といった他のレジスタにそのまま一般化できるとは限らない。これらのレジスタでは，比喩表現の用法や読解パターンが異なる可能性がある。第二に，本研究の方法は自己ペース読み課題に依拠しており，心理言語学では広く用いられているものの，自然な読書のダイナミクスを完全に再現するものではない。たとえば，視線計測を用いれば，より細やかで生態学的妥当性の高い瞬間的処理の指標が得られたはずである。また，脳活動イメージングによる検討は，より高い時間分解能を持つ補完的な証拠を提供しうる。第三に，本研究で扱った比喩カテゴリ（結合比喩, 換喩, 提喩, 文脈比喩）は中核的な現象をカバーしているが，その他の修辞表現形式は取り上げていない。第四に，我々の分析は文節レベルの二値指標に基づいており，比喩性評定や語彙サプライザルといった連続的な指標を導入すれば，より微妙な変動を捉えられる可能性がある。最後に，本研究は日本語に特化しており，日本語における比喩処理に光を当てる一方で，他言語・他書記体系における一般性を評価するためには，今後の言語間比較研究が必要である。

8　おわりに

本研究は，日本語における比喩言語処理をめぐる議論に対し，コーパスに対する比喩情報アノテーションと自己ペース読み時間データを統合することで新たな知見を提供した。結果として，比喩表現が字義通りの言語よりも一様に処理コストを増大させるわけではなく，表現の種類と位置に応じて異なる効果を示すことが明らかになった。具体的には，結合比喩は表現の先頭で読解を遅延させる一方，内部では処理を促進し，換喩は持続的なコスト増加をもたらし，提喩は先頭でのみコストを低減し，文脈比喩は表現の開始時に遅延を引き起こすことが示された。これらの知見は，比喩処理が均質ではなく多様であることを示すことで，先行研究における相反する結果を調停する。

8.1　理論的意義

理論的には，本研究は比喩処理のコストが一様ではなく，表現の種類と位置に体系的に依存することを明らかにした点で意義を持つ。これは，比喩言語が常によりコストの高い処理を要すると仮定する「字義通り優先」モデル [1, 2] に対して疑問を投げかけるとともに，サリエンスに基づくモデル [5, 6] を精緻化するものである。とりわけ，結合比喩の内部のように，ある種の比喩表現が処理をむしろ促進しうることを示した点は，英語など他言語で得られた知見を補完する日本語からの証拠として重要である。これらの結果は，比喩言語理解モデルにおいて，構造的要因と文脈的要因の双方を明示的に考慮する必要性を強調する。

8.2　方法論的意義

方法論的には，本研究はコーパスベースの比喩アノテーションを読み時間という心理言語学的指標と結びつけることの有用性を示した。具体的には，BCCWJ-Metaphor の BIO ラベルを BCCWJ-SPR2 の文節境界に再割り当てすることで，比喩カテゴリとオンライン処理データを直接リンクさせる枠組みを構築した。本アプローチは，自然な文脈におけるコーパス資源を活用して，比喩表現の処理を大規模かつ定量的に調査する方法を提示するものである。構築したデータを各用例を見ることで、より精緻な議論が可能となる。

8.3　実践的意義

実践的観点からも，本研究の知見は言語教育や自然言語処理などの応用分野に示唆を与える。言語教育において，比喩表現のどのタイプがより大きな処理コストを要するかを理解することは，比喩がしばしば大きな課題となる第二言語学習者向け教材の設計に役立つ。また，計算言語学の観点では，比喩アノテーションと読み時間のエビデンスを統合することで，比喩検出や比喩解釈のためのより認知的に妥

当なモデル構築が期待できる。これは，機械翻訳や読解支援ツール，教育テクノロジーなどの応用において，比喩表現の扱いを改善する上で重要である。心理言語学的エビデンスと応用的文脈とを結びつけることで，本研究は理論的・方法論的貢献にとどまらず，現実世界における比喩理解の支援に向けた基盤を提供する。

総じて，本研究は，日本語における比喩表現の認知処理を明らかにするとともに，比喩表現を対象とした理論的・方法論的・実践的検討のための土台を提供し，将来の言語間・モダリティ間研究への展望を開くものである。

利益相反の開示

本研究は JSPS 科研費 JP25K00459, JP23K21935, JP22K18483 および国立国語研究所共同研究プロジェクトの助成を受けたものです。企業などとの共同研究および企業などからの助成は受けていない。

参考文献

[1] Andrew Ortony, Diane L. Schallert, Richard E. Reynolds, and Samuel J. Antos. Interpreting metaphors and idioms: Some effects of context on comprehension. **Journal of Verbal Learning and Verbal Behavior**, Vol. 17, No. 4, pp. 465–477, 1978.

[2] Rebecca A. Janus and Thomas G. Bever. Processing of metaphoric language: An investigation of the three-stage model of metaphor comprehension. **Journal of Psycholinguistic Research**, Vol. 14, No. 5, pp. 473–487, 1985.

[3] Raymond W. Gibbs. Skating on thin ice: Literal meaning and understanding idioms in conversation. **Discourse Processes**, Vol. 9, No. 1, pp. 17–30, 1986.

[4] Sam Glucksberg. The psycholinguistics of metaphor. **Trends in Cognitive Sciences**, Vol. 7, No. 2, pp. 92–96, 2003.

[5] Rachel Giora. Understanding figurative and literal language: The graded salience hypothesis. **Cognitive Linguistics**, Vol. 8, No. 3, pp. 183–206, 1997.

[6] Rachel Giora. **On Our Mind: Salience, Context, and Figurative Language**. Oxford University Press, Oxford, 2003.

[7] Seana Coulson and Cyma Van Petten. Conceptual integration and metaphor: An event-related potential study. **Memory & Cognition**, Vol. 30, No. 6, pp. 958–968, 2002.

[8] Valentina Bambini, Claudia Gentili, Emiliano Ricciardi, Pier Marco Bertinetto, and Pietro Pietrini. Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. **Brain Research Bulletin**, Vol. 86, No. 3-4, pp. 203–216, 2011.

[9] Vicky T. Lai, Tim Curran, and Lise Menn. Comprehending conventional and novel metaphors: An ERP study. **Brain Research**, Vol. 1284, pp. 145–155, 2009.

[10] Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. **Balanced Corpus of Contemporary Written Japanese**. National Institute for Japanese Language and Linguistics, 2014.

[11] Masayuki Asahara. BCCWJ-SPR2: A Self-Paced Reading Corpus with Comprehension Questions. In **Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC 2022)**, pp. 5178–5186, Marseille, France, 2022. European Language Resources Association (ELRA).

[12] Sachi Kato, Rei Kikuchi, and Masayuki Asahara. Figurative Expression Information Database on 'Balanced Corpus of Contemporary Written Japanese'. Presented at RaAM15: Researching and Applying Metaphor Conference, 2022.

[13] Sachi Kato, Rei Kikuchi, and Masayuki Asahara. Expert Annotation and Crowdsourced Ratings of Figurative Expressions: A Study Using BCCWJ-Metaphor: A Cross-Domain Japanese Corpus of Books, Magazines, and Newspapers. Presented at RaAM17: Researching and Applying Metaphor Conference, 2025.

[14] John R. Searle. **Expression and Meaning: Studies in the Theory of Speech Acts**. Cambridge University Press, Cambridge, 1979.

[15] H. Paul Grice. Logic and Conversation. In Peter Cole and Jerry L. Morgan, editors, **Syntax and Semantics, Vol. 3: Speech Acts**, pp. 41–58. Academic Press, 1975.

[16] David A. Swinney and Anne Cutler. The access and processing of idiomatic expressions. **Journal of Verbal Learning and Verbal Behavior**, Vol. 18, No. 5, pp. 523–534, 1979.

[17] Raymond W. Gibbs. Spilling the beans on understanding and memory for idioms in conversation. **Memory & Cognition**, Vol. 8, No. 2, pp. 149–156, 1980.

[18] Steven Frisson and Martin J. Pickering. Obtaining a Figurative Interpretation of a Word: Support for Underspecification. **Metaphor and Symbol**, Vol. 16, No. 3-4, pp. 149–171, 2001.

[19] George Lakoff and Mark Johnson. **Metaphors We Live By**. University of Chicago Press, Chicago, 1980.

[20] Raymond W. Gibbs. **The Poetics of Mind: Figurative Thought, Language, and Understanding**. Cambridge University Press, Cambridge, 1994.

[21] Sachi Kato, Masayuki Asahara, and Makoto Yamazaki. Annotation of 'Word List by Semantic Principles' Labels for the Balanced Corpus of Contemporary Written Japanese. In **Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC 32)**, pp. 247–253, Hong Kong, 2018.

[22] Masayuki Asahara and Sachi Kato. Between Reading Time and Syntactic/Semantic Categories. In **Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017)**, pp. 404–412, Taipei, Taiwan, 2017. AFNLP.

[23] Masayuki Asahara and Yuji Matsumoto. BCCWJ-DepPara: A syntactic annotation treebank on the 'Bal-

anced Corpus of Contemporary Written Japanese'. In Koiti Hasida, Kam-Fai Wong, Nicoletta Calzorari, and Key-Sun Choi, editors, **Proceedings of the 12th Workshop on Asian Language Resources (ALR12)**, pp. 49–58, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.

付録のサンプル: 付録は独立に 1 ページだけ