

# 日本語テキスト解析のための感情辞書標準版の作成と公開： ビッグデータ時代の心理分析に向けて

研究代表者 五十嵐 祐 名古屋大学大学院 教育発達科学研究科 准教授  
共同研究者 笹原 和俊 名古屋大学大学院 情報学研究科 講師

## 1 問題

心理学の領域では、個人の行動傾向や価値観の分析を行うために、従来、実験や調査といったアプローチを用いることが主であった。一方、高度情報化社会の発展に伴って、人々の社会的活動をオンラインでの行動履歴を通じて把握する、ビッグデータの活用がますます重要となりつつある(総務省, 2017)。中でも、言語行動をデジタル化したテキストデータから個人の心理傾向を推測する、言語内容に関する分析的アプローチは、今日、マーケティング戦略の策定から「国民感情」の推定に至るまで、世界的に大きな注目を集めている。国内でも、道徳基盤理論に基づく Twitter のツイート内容の解析 (Matsuo, Sasahara, Taguchi, & Karasawa, 2018) など、近年、テキストデータを用いた心理学的な研究が活発に行われている。その一方で、共通の基準に則った国際的な比較を可能とする網羅的な感情辞書の日本語版は、未だに学術利用が可能な形で一般に公開されていない。そのため、日本におけるテキストデータ分析の心理学的応用は、国内の一部サービスが独自に開発した内容に限定されているのが現状である。

本研究の目的は、心理学の領域におけるデファクト・スタンダードの感情辞書である LIWC (Linguistic Inquiry and Word Count; <http://www.liwc.net/>) (Pennebaker, Boyd, Jordan, & Blackburn, 2015) の日本語版を作成することである。感情辞書とは、日常よく使用される単語に対して、専門家によって予め分類された言語学的なカテゴリー(助動詞や接続詞などの文法的機能を示す機能語と、名詞や動詞などの実質的な内容を持つ内容語の両方を含む)と、心的プロセスに関するカテゴリー(感情的プロセス、認知的プロセス、社会的プロセスなどを含む)を複合的に割り当てたものである(広義には、機械学習や自然言語処理などの領域で、単語の極性(感情価、センチメント)をアルゴリズムに基づいて分類したものも含まれる)。一般的な感情辞書では、ポジティブ、ネガティブの2カテゴリーに語句を分類するが、LIWCでは、例えば「jealous」という単語に「ネガティブ感情」、「怒り」という要素が、「ambition」という単語に「欲求」、「権力」、「仕事」、「達成」といった多様なカテゴリーが割り当てられる。また、カテゴリーは階層性をもち、例えば「ネガティブ感情」カテゴリーは「怒り」や「悲しみ」カテゴリーの上位に位置する。さらに、単語の語尾にワイルドカード(\*)を含めることもでき、例えば「occupa\*」という単語は、「occupancy」、「occupant」、「occupation」、「occupational」という単語をカバーする。テキストデータに感情辞書を適用することで、そのテキストの書き手の心理傾向を多面的に推測することができるため、LIWCは特に近年、議事録の発話者の心的態度の推定 (Frimer, Aquino, Gebauer, Zhu, & Oakes, 2015)、TwitterやFacebookといったSNS利用者の政治的態度の予測 (Tumasjan, Sprenger, Sandner, & Welpe, 2010)、さらには、SNSでのいわゆる「炎上」の把握 (Cheng, Caverlee, & Lee, 2010) といったさまざまな方面での活用が進んでいる。

LIWCは、1997年に初版の英語版が作成され、その後、2001年、2007年、2015年に、語彙やカテゴリーの追加を含むバージョンアップが行われてきた。最新版のLIWC2015は、約6400の英単語、73のカテゴリーを含む辞書データベースである。表1に、カテゴリーの一覧と各カテゴリーに含まれる単語の例を示す。LIWCは各国の研究者によってローカライズされ、現在までに、スペイン語、ドイツ語、簡体字中国語、繁体字中国語、ロシア語、アラビア語、フランス語、イタリア語、ポルトガル語、セルビア語、トルコ語に翻訳されている。また、旧版のLIWC2001、LIWC2007については、日本語版の作成の試みがこれまでになされている(那須川・上條・山本・北村, 2016; 柴田, 2018; 山本・那須川・上條・北村, 2016)。

本研究では、開発者のJames E. Pennebakerの許諾を得て、最新版のLIWC2015の辞書データの提供を受け、英語版に含まれる単語の翻訳と、カテゴリーとの対応関係について検討を行い、LIWC2015日本語版(J-LIWC2015)のプロトタイプ版の作成を試みる。また、内容的妥当性の検証のために、クラウドソーシングサービス上で心理学実験を行う。実験では、参加者に自身の感情体験を記述してもらい、J-LIWC2015で記述内容の解析を行うことで、想定される感情状態に対応した心的要素が記述内容から適切に抽出できるかどうかの検討を行う。

表 1 LIWC2015、J-LIWC2015 のカテゴリー

カテゴリー	略称	単語数		単語例 (J-LIWC2015)
		J-LIWC2015 (日本語)	LIWC2015 (英語)	
言語的側面				
機能語		1051	491	
代名詞	Pronoun	151	153	
人称代名詞	Ppron	61	93	
一人称単数	I	22	24	私, 僕, 自分
一人称複数	We	14	12	我ら, 我々, 私たち
二人称	You	18	30	あなた, 君, 汝
三人称単数	SheHe	5	17	彼, 彼女
三人称複数	They	4	11	それら, 彼ら, 彼女ら
不定代名詞	Ipron	89	59	これら, どの, どれ
前置詞	Prep	332	74	について, ほぼ, 未満
助動詞	Auxverb	71	141	べき, なる, だ
副詞	Adverb	521	140	あるいは, しかも, すでに
接続詞	Conj	134	43	だから, すると, ところが
否定詞	Negate	39	62	ない, ず, 不
その他の文法				
疑問詞	Interrog	55	48	いつ, なぜ, どこ
数詞	Number	47	36	一級, シングル, 多数
数量詞	Quant	146	77	いくつ, トータル, ほぼ
心理的プロセス				
感情的プロセス	Affect	1999	1393	
ポジティブ感情	Posemo	890	620	ありがたい, すばらし*, 歓喜
ネガティブ感情	Negemo	1098	744	
不安	Anx	210	116	ひどい, パニック, 緊張
怒り	Anger	380	230	むかつ*, 嫉妬, 憎らし*
悲しみ	Sad	227	136	かわいそう, 気落ち, 哀*
社会的プロセス	Social	960	756	
家族	Family	112	118	おかあさん, 兄, 血縁
友人	Friend	131	95	友, 同士, 連中
女性	Female	103	124	乙女, 婦人, ママ
男性	Male	111	116	王子, 主人, パパ
認知的プロセス	CogProc	1368	797	
洞察	Insight	423	259	検査, 記憶, 見つけ*
原因	Cause	279	135	なぜ, 解明, 反応
食い違い	Discrep	114	83	理想, 誤解, 要求
あいまい表現	Tentat	341	178	ああいう, らしい, 概ね
確かさ	Certain	254	113	はっきり, 事実, 明快
差異化	Differ	191	81	別, 違い, 対立

(表 1 の続き)

知覚的プロセス	Percept	652	436	
視覚	See	190	126	光, 風景, 目
聴覚	Hear	142	93	歌, 笑, ひそひそ
感覚	Feel	206	128	硬い, 厚い, 鋭利
生体的プロセス	Bio	900	748	
身体	Body	262	215	筋肉, 胸, ヒップ
健康	Health	371	294	肥満, 免疫, 元気
性	Sexual	149	131	妊娠, 性器, セックス
摂取	Ingest	224	184	アルコール, 砂糖, 摂取
動因	Drives	1866	1103	
親和	Affiliation	395	248	愛*, コミュニティ, 談話
達成	Achieve	439	213	勤勉, パーフェクト, 報酬
パワー	Power	982	518	エリート, 威力, 争い
報酬	Reward	185	120	功績, 恩恵, 給付
リスク	Risk	190	103	トラブル, 危険, 回避
相対性	Relativ	1734	974	
動作	Motion	435	325	ズーム, 運転, 回す
空間	Space	786	360	あちこち, あたり, そば
時間	Time	631	310	たちまち, ごろ, から
個人的関心				
仕事	Work	896	444	ディレクター, 学会, 業績
余暇	Leisure	460	296	カフェ, 映画, 趣味
家	Home	162	100	ワンルーム, 車庫, 住宅
金銭	Money	341	226	為替, 給料, 金庫
宗教	Relig	237	174	クリスマス, モスク, 儀礼
死	Death	138	74	喪*, 根絶, 訃報
インフォーマル	Informal	458	380	
罵倒	Swear	154	131	アホ, 役立たず, クズ
ネットスラング	Netspeak	228	209	www*, リツイート, 乙
間投詞	Assent	80	36	あー, よし, わあ
感動詞 (間投詞)	Nonflu	46	19	ふーん, そのー, うっ
フィラー	Filler	44	14	ほら, さあね, ええと

注: 「\*」はワイルドカード; 上位カテゴリーのみに含まれる語が存在するため、下位カテゴリーの単語数の合計は上位カテゴリーの単語数とは一致しない; LIWC2015に含まれていて、J-LIWC2015に含まれていないカテゴリーは、冠詞 (Article)、比較級 (Compare)、動詞 (Verb)、形容詞 (Adverb)、時制 (TimeOrient/Focuspresent/Focuspast/Focusfuture) である

## 2 辞書データの翻訳

まず、LIWC2015 に含まれるすべての英単語について、心理学を専攻し、英語圏での生活経験がある大学院生及び大学生 5 名 (日本人学生 4 名、留学生 1 名) と、研究実施者 2 名による翻訳作業を行った。翻訳の手順は、繁体字中国語版の LIWC の作成手順 (Huang et al., 2012) を参考に、下記の手順で行った。

### 2-1 機械翻訳

LIWC に含まれる全ての英単語について、Weblio (<https://ejje.weblio.jp/>) の機械翻訳を用いて日本語訳を抽

出した。この段階では、1つの英単語に対応するできるだけ多くの訳語を抽出した。ワイルドカードが含まれる語については、OneLook (<https://www.onelook.com/>) で該当する英単語を抽出し、明らかに適切でない英単語を手動で除去した後、Weblio で訳語を抽出した。訳語については、現代日本語書き言葉均衡コーパス (BCCJW (少納言、中納言) ; [http://pj.ninjal.ac.jp/corpus\\_center/bccwj/](http://pj.ninjal.ac.jp/corpus_center/bccwj/))、筑波ウェブコーパス (TWC; <http://nlt.tsukuba.lagoinst.info/>) に含まれるかどうかを確認し、いずれかのコーパスに含まれる語のみを残した。

## 2-2 人力翻訳

機械翻訳で得られた語について、まず、明らかに適切でない日本語訳を手動で除去した。その後、元の英単語に付与されている LIWC のカテゴリーを参考に、訳語の確認・修正を行った。作業にあたっては、次の点に配慮した。

### (1) 訳語の表記

訳語は、辞書の項目をそのまま用いるのではなく、実際に書き言葉で使われる形式となるように配慮した。抽象的な意味の英単語や、日本語に直接対応する語彙がないと思われる英単語については、実際にその語が使われる文脈での表現をできるだけ含めるようにした。一語で直接対応する訳語がないと思われる英単語は削除した。また、日本語では漢字2字以上で構成される熟語(複合語)が1つの単語となるため、1つの英単語の意味(カテゴリー)を直接的に表す漢字一文字の訳語が、複数の熟語の構成要素に含まれるケースも多い(例えば、「angry」→「怒」は、「怒号」、「怒鳴る」といった熟語の構成要素となり、いずれの語も「ネガティブ感情」や「怒り」といったカテゴリーに対応する)。さらに、語形変化が規則的(語幹+活用語尾)である語も多くみられる(例えば、「怒る」、「怒り」、「怒って」)。これらの語については、ワイルドカード(「\*」)を用いて訳語を同定した(例えば、「怒\*」)。ただし、ワイルドカードは個別の単語指定よりも優先されるため、国語辞典と漢和辞典をもとに、同じ漢字が熟語ごとに異なる意味をもたないと考えられる場合にのみ、ワイルドカードを使用した。また、語尾についてはワイルドカードを使用できないため、該当すると思われる語をできるだけ含めた(例えば、「激怒」)。

### (2) 文化的差異

英語圏と日本語圏では、文化的な背景が異なるため、「宗教」や「レジャー」のカテゴリーに含まれる英単語(例えば、「yiddish」→「イディッシュ(東欧のユダヤ人が用いる言語)」、「DVR」→「デジタルビデオレコーダー」)や、サービスや商品名などの固有名詞(例えば、「zoloft」→「ゾロフト(抗うつ剤)」)は、日本語への翻訳は可能なものの、実際にはほとんど使用されていないものもみられた。これらの語については、原版との対応を考慮してそのまま残した。また、一般的な語に置き換え可能な場合は、そちらを日本語訳として含めた(例えば、「prozac」→「抗うつ剤」)。

### (3) 分かち書きと複合語の登録

LIWC では、半角スペースを一区切りとして単語を検出するため、日本語のテキストデータをそのまま処理することができない。そのため、前処理として MeCab (Kudo, Yamamoto, & Matsumoto, 2004) などのソフトウェアによる形態素解析を行い、文章を形態素ごとに区切った分かち書きの形式に変換する必要がある (Matsuo et al., 2018)。本研究では、MeCab のシステム辞書である IPA 辞書および mecab-ipadic-NEologd (佐藤・橋本・奥村, 2016) を用いて形態素解析を行った。また、形態素解析の際には、J-LIWC2015 に含まれる語が複数の単語に分割されてしまうケースがある(例えば、「cancer」→「悪性腫瘍」は、IPA 辞書による形態素解析では「悪性」と「腫瘍」に分かれてしまう)。そこで、これらのケースにあてはまる語については、単語が分割されないよう、ユーザー辞書を用いて個別に語の登録を行った(単語のコストは便宜的にすべて1とした)。登録した複合語は1,372語であった。

### (4) 人力翻訳に用いた辞書

人力翻訳の作業には、Longman Dictionary of Contemporary English, 5<sup>th</sup> Edition (Pearson Education)、Collins Cobuild English Dictionary for Advanced Learners, 3<sup>rd</sup> Edition (HarperCollins)、新英和大辞典第6版(研究社)、新和英大辞典第5版(研究社)、新編英和活用大辞典(研究社)、リーダーズ英和辞典第2版(研究社)、新英和中辞典第7版(研究社)、新和英中辞典第5版(研究社)、ランダムハウス英語辞典第2版(小学館)、ジーニアス英和辞典第4版(大修館)・ジーニアス和英辞典第2版(大修館)、NEW 斎藤和英大辞典(日外アソシエーツ)、ビジネス技術実用英語大辞典 V5 (プロジェクトポトス)、広辞苑第6版(岩波書店)、日本語大シソーラス第2版(大修館)、ライフサイエンス辞書(LSDプロジェクト)、英辞郎(アルク)、漢字源改訂第5版(学研プラス)、Urban Dictionary (<https://www.urbandictionary.com/>) を用いた。また、各種書籍や Web サイト

の情報も適宜参照した。

### 2-3 訳語とカテゴリの選択

日本語と英語ではさまざまな点で語法が異なるため、訳語、カテゴリの選択や辞書の作成にあたっては、以下の点に配慮した：(1) 今回の作業では、LIWC2015に含まれない新たなカテゴリは加えなかった。これは、LIWC2015の構成をなるべく忠実に反映したベースラインの感情辞書を作成することを念頭に置いたためである。(2) 冠詞 (a, an, the)、比較級、時制、動詞、形容詞のカテゴリは、J-LIWC2015には含めなかった。冠詞のカテゴリについては、日本語に同様の語法が存在しないこと、比較級と時制のカテゴリについては、日本語では副詞(「より」)や助動詞(「た」)として情報が付与されること、動詞のカテゴリについては、日本語では名詞+「する」という複合語によって構成されるケースが多いこと(例えば、「信頼」→「信頼する」)、形容詞のカテゴリについては、語形変化によって派生語としての名詞になるケースが多いこと(例えば、「小さい」→「小ささ」)を考慮したためである。(3) 助動詞の短縮形(例えば、「shouldn't」)は翻訳対象から除外した。(4) 英単語として含まれている顔文字(例えば、「:」)については、直接対応する日本語の顔文字を同定することが困難なため、翻訳対象から除外した。(5) 同一の単語で、過去形、過去完了形、進行形、および三単現の「s」が付与されるといったバリエーションが存在し、それぞれ個別に登録されており、対応するカテゴリが重複するケース(例えば、associate, associated, associating, associates)に関しては、すべて一つの単語として扱い、訳語を対応させた。(6) インフォーマル、ネットスラングのカテゴリが付与されており、それらに直接対応する訳が見つからない語は、翻訳対象から除外した(例えば、「prob」は「おそらく」の意味だが、インフォーマル、ネットスラングのカテゴリに対応する適切な訳語がないため削除した)。(7) 同じ訳語が異なる英単語に対応するケース(例えば、「利益」→benefit(動因、報酬のカテゴリ)、profit(動因、報酬、仕事のカテゴリ))については、それぞれの英単語に紐付けられているカテゴリをすべて合成して訳語に対応させた(例えば、「利益」→動因、報酬、仕事のカテゴリを付与)。

上記の基準を用いて、J-LIWC2015のプロトタイプ版を作成した。プロトタイプ版に含まれるカテゴリや単語の情報を表1に示す。総単語数は8,564であった(LIWC2015の総単語数は6,549)。

## 3 J-LIWC2015の妥当性：感情状態の操作による検討

J-LIWC2015の妥当性を検証するため、クラウドソーシングサービスで募集した実験参加者にテーマを指定して文章を記述してもらい、記述時の感情状態がJ-LIWC2015による内容分析の結果と対応するかどうかを検討した。なお、先行研究では、研究実施時に十分な配慮を行うことで、オンライン上で参加者が自身の感情体験を言語的に報告するという手続きを用いた場合に、ポジティブな感情体験、ネガティブな感情体験のいずれについても、参加者のストレスレベルの上昇や心理的健康の低下が直接引き起こされるわけではないことが示されている(Baikie, Geerlings, & Wilhelm, 2012; Stockton, Joseph, & Hunt, 2014)。

### 3-1 方法

#### (1) 参加者

クラウドソーシングサービス(ランサーズ)で募集した実験参加者のうち、研究内容と手続きを理解した上で参加に同意し、回答完了後にも回答を分析に用いることに同意した者のうち、記述課題の内容がナンセンスであった2名を除き、全ての設問に回答した500名(女性305名、男性195名、平均年齢37.9歳( $SD=9.85$ ))のデータを分析に用いた<sup>1</sup>。参加者の学歴は、4年制大学卒・在学中252名、高卒100名、専門学校卒・在学中77名、大学院卒・在学中21名、2年制大学卒・在学中39名、中卒7名、未回答(答えたくない)4

<sup>1</sup> サンプルサイズの決定にあたっては、G\*Power 3.1を用いて、3条件(ポジティブ感情条件、ネガティブ感情条件、統制条件)の一要因分散分析に関する検定力分析を行った。本研究は、J-LIWC2015の妥当性に関する探索的な検討であること、参加者1名あたりの記述量が少ないこと、さらにクラウドソーシングサービスで募集した参加者のバックグラウンドの多様性を考慮して、検定力分析のパラメータを有意水準 $\alpha=0.05$ 、検定力 $(1-\beta)=0.8$ 、効果量 $f=0.15$ (効果量小 $f=0.10$ と効果量中 $f=0.25$ の間)にそれぞれ設定した。分析の結果、必要なサンプルサイズは432名(各条件144名)となり、この人数を目安に参加者の募集を行った。

名であった。参加者の職業は、主婦・主夫 127 名、会社員・団体職員 116 名、自営業 87 名、パート・アルバイト 61 名、無職 30 名、学生 19 名、契約社員・派遣社員 23 名、公務員 3 名、会社役員・団体職員 2 名、その他 23 名、未回答（答えたくない）9 名であった。参加報酬は 1 名あたり 410 円で、回答完了後、回答を分析に用いることの同意の有無にかかわらず支払われた。平均回答時間は 25.5 分 ( $SD = 24.0$ ) であった。

## (2) 手続き

及川・及川 (2012) の手続きを参考に、感情状態を操作した。課題はすべて、オンラインアンケート作成サービスである Qualtrics で作成・提示した。なお、以下では本研究の分析に用いた測定項目についてのみ報告する。

まず、参加者は、研究内容の説明（個人的に体験した出来事について、指定されたテーマでの記述を求めると）と、研究参加にあたっての注意事項（回答は匿名で行われ、研究実施者が個人を特定することは決していないこと；回答は強制ではなく、参加者の自由意志に基づいていること；記述のテーマは作業前に確認でき、回答したくない場合はその場で回答を終了できること）を読んで、参加への同意をクリックした後、デモグラフィック項目（年齢、性別、学歴、職業）について回答した。その後、参加者は、ポジティブ感情条件、ネガティブ感情条件、統制条件のいずれかにランダムに割り当てられた。望まないテーマでの回答を避けるため、参加者には、最近経験したポジティブな出来事（ポジティブ感情条件）、ネガティブな出来事（ネガティブ感情条件）、特に何も無い平凡な一日の出来事（統制条件）の各テーマのうち、割り当てられたテーマのみを予め提示し、次に行う作業（テーマに関する説明や、出来事を体験した際の感情状態の記述）についての説明を行った後、課題を継続してもよいかを確認した。この段階で課題の継続を希望しない参加者については、直ちにアンケートを終了した（この場合、参加者には報酬が支払われなかった）。課題の継続を希望した参加者は、次の画面で、指定されたテーマに関する出来事の詳細や、そのときの感情状態について、200 文字以上での記述を求めた。記述の完了後、参加者は、操作チェックとして出来事の快・不快感（1 項目 6 件法）、インパクト（2 項目 6 件法）について回答し、感情状態の測定のために顕在的感情（ポジティブ感情、ネガティブ感情）を測定する PANAS（16 項目 6 件法）に回答した。その後、パーソナリティ（外向性、誠実性、神経症傾向、開放性、調和性）に関する質問（並川他, 2012; 29 項目 7 件法）についての回答を求めた後、アンケートは終了した。

終了画面では、参加者に研究の目的について説明した後、コメントや意見等がある場合は自由記述で回答してもらい、回答を分析に用いることに同意するかどうかを尋ね、全ての質問を終えた。非同意を選択した参加者は 15 名であり、これらの回答は分析から除外した。募集時の画面とアンケート終了時の画面には、研究実施者の所属、氏名、連絡先メールアドレス（大学ドメイン）を明記し、アンケートに関する質問や疑問にはランサーズの個別連絡機能で対応した。なお、個別連絡（5 件）はいずれも回答後の報酬の受領手続きに関するものであり、コメントや意見の自由記述においても、研究内容に対する不満や回答後の苦痛を訴えた参加者はいなかった。

## (3) 記述内容の解析

記述内容の解析に先立って、McCab による形態素解析を行い、半角スペースで形態素を区切った形式に記述内容を変換して、参加者ごとに個別のテキストファイルで保存した（エンコード：UTF-8、BOM あり）。これらのファイルについて、LIWC2015 のソフトウェア（Version 1.2）で辞書に J-LIWC2015 を指定して解析を行い、記述内容に含まれる各カテゴリーの単語の出現割合を算出した。

### 3-2 結果

各テーマに割り当てられた参加者は、ポジティブ感情条件 164 名、ネガティブ感情条件 179 名、統制条件 157 名であった。

#### (1) 操作チェック変数および心理変数の分析

アンケートで測定した心理変数について、条件別の記述統計量と条件間の平均値の比較結果（一要因分散分析（参加者間計画）の結果、Tukey 法による多重比較結果）を表 2 に示す。

出来事の快・不快感については、ポジティブ感情条件の参加者 > 統制条件の参加者 > ネガティブ感情条件の参加者の順に、出来事がポジティブであると評定していた。インパクトについては、ポジティブ感情条件の参加者 > ネガティブ感情条件の参加者 > 統制条件の参加者の順に強いと評定していた。そこで、インパクトの得点範囲（2～12 点）を踏まえて、中点（7 点）を母平均とする 1 サンプルの  $t$  検定を条件別に行ったところ、ポジティブ感情条件およびネガティブ感情条件では、いずれもインパクト得点が中点よりも有意に高

かった (ポジティブ感情条件 :  $t(163)=27.0, p<.001$ ; ネガティブ感情条件 :  $t(178)=14.0, p<.001$ ) のに対し、統制条件ではインパクト得点と中点との間に有意な差はみられなかった ( $t(156)=0.16, p=.87$ )。これらのことから、記述した出来事の主観的な快・不快度の評価は感情状態の操作に対応しており、さらにポジティブ感情条件、ネガティブ感情条件のいずれの参加者も、記述した出来事について強めのインパクトを感じていたと考えられる。

表 2 心理変数に関する条件別の平均値 ( ) 内は標準偏差) と平均値の比較結果

	出来事の 快・不快度	インパクト ( $\alpha = .69$ )	ポジティブ 感情 ( $\alpha = .88$ )	ネガティブ 感情 ( $\alpha = .91$ )	外向性 ( $\alpha = .87$ )	誠実性 ( $\alpha = .93$ )	神経症傾向 ( $\alpha = .88$ )	開放性 ( $\alpha = .83$ )	調和性 ( $\alpha = .76$ )
統制条件 (C)	3.83 (1.24)	7.03 (2.47)	20.20 (7.30)	19.90 (8.92)	19.50 (6.42)	27.90 (7.79)	25.60 (5.73)	25.50 (6.43)	23.20 (5.83)
ポジティブ 感情条件 (P)	5.38 (1.14)	10.30 (1.58)	26.80 (7.63)	20.10 (8.34)	19.50 (5.90)	28.40 (6.68)	24.00 (6.36)	25.20 (5.99)	22.10 (5.05)
ネガティブ 感情条件 (N)	1.61 (0.94)	9.20 (2.10)	17.80 (7.11)	28.30 (9.37)	18.80 (6.35)	27.10 (7.21)	25.30 (5.54)	24.40 (6.16)	22.90 (5.69)
$F(2, 497)$	503.2***	104.9***	68.0***	49.5***	0.70	0.20	3.40*	0.22	0.16
多重比較 (Tukey)	P>C>N	P>N>C	P>C>N	N>C, N>P			C>P		

\*\*\* $p < .001$ ; \* $p < .05$ . ポジティブ感情条件164名、ネガティブ感情条件179名、統制条件157名

また、出来事について記述した直後の顕在的ポジティブ感情は、ポジティブ感情条件の参加者>統制条件の参加者>ネガティブ感情条件の参加者の順に高かった。顕在的ネガティブ感情は、ネガティブ感情条件の参加者が統制条件とポジティブ感情条件の参加者よりも高かった。したがって、テーマの設定による感情状態の操作は成功していたと考えられる。

神経症傾向については、統制条件の参加者>ポジティブ感情条件の参加者であった。外向性、誠実性、開放性、調和性については、条件間で有意な差はみられなかった。

## (2) 記述内容に含まれる語のカテゴリーの分析

記述内容における各カテゴリーの単語の出現割合について、条件別の記述統計量と条件間の平均値の比較結果 (一要因分散分析 (参加者間計画) の結果、Tukey 法による多重比較結果) を表 3 に示す。ここでは、本研究の実験操作に直接関連すると考えられる、感情プロセスのカテゴリーについてのみ下位カテゴリーを含めて分析の対象とし、その他のカテゴリーについては最上位のカテゴリーのみを分析の対象とした。

記述の総文字数および辞書に含まれている語の割合については、条件間で有意な差はみられなかった。記述の総語数については、ネガティブ感情条件よりもポジティブ感情条件の方が多かった。機能語の割合は、ポジティブ感情条件、統制条件に比べて、ネガティブ感情条件で多かった。

感情的プロセスのカテゴリーに属する語の割合は、ポジティブ感情条件>ネガティブ感情条件>統制条件の順に多かった。その一方で、下位カテゴリーごとの分析の結果から、ポジティブ感情のカテゴリーに属する語の割合は、ポジティブ感情条件が最も多く、ネガティブ感情とその下位カテゴリー (不安、怒り、悲しみ) のカテゴリーに属する語の割合は、ネガティブ感情条件が最も多かった。

社会的プロセスおよび動因のカテゴリーに属する語の割合は、統制条件に比べて、ポジティブ感情条件、ネガティブ感情条件で多かった。認知的プロセスのカテゴリーに関する語の割合は、ネガティブ感情条件>ポジティブ感情条件>統制条件の順に多かった。これに対して、生体的プロセスおよび相対性のカテゴリーに属する語の割合は、ポジティブ感情条件、ネガティブ感情条件よりも統制条件で多かった。インフォーマルのカテゴリーに属する語の割合は、統制条件よりもネガティブ感情条件で多かった。知覚的プロセスのカテゴリーに属する語の割合については、条件間で有意な差はみられなかった。

以上の結果は、感情状態の操作に伴って、想起した出来事の記述内容が変化し、操作に対応すると考えられるカテゴリーに属する語が、記述内で多く用いられたことを示すものである。

表3 記述内容における各カテゴリー語の出現割合の条件別平均値 ( )内は標準偏差) と平均値の比較結果

	統制条件 (C)	ポジティブ感情条件 (P)	ネガティブ感情条件 (N)	<i>F</i> (2, 497)	多重比較 (Tukey)
総文字数 <sup>a</sup>	297.0 (107.0)	281.1 (97.2)	309.8 (153.2)	2.34	
総語 (形態素) 数 <sup>a</sup>	158.9 (55.3)	150.9 (54.0)	171.3 (83.7)	4.10*	N>P
辞書に含まれている語	63.0% (5.84)	62.0% (5.28)	62.8% (4.84)	1.63	
機能語	38.1% (6.62)	39.1% (4.98)	41.4% (5.08)	15.7***	N>C, N>P
感情的プロセス	5.36% (2.71)	8.34% (2.90)	7.09% (2.52)	48.8***	P>N>C
ポジティブ感情	3.65% (2.19)	6.77% (2.68)	4.04% (1.91)	91.5***	P>C, P>N
ネガティブ感情	1.33% (1.19)	1.26% (1.15)	2.80% (1.69)	68.5***	N>C, N>P
不安	0.33% (0.53)	0.44% (0.66)	0.71% (0.90)	11.9***	N>C, N>P
怒り	0.35% (0.65)	0.24% (0.46)	0.71% (0.84)	22.7***	N>C, N>P
悲しみ	0.24% (0.44)	0.26% (0.47)	0.83% (0.89)	45.7***	N>C, N>P
社会的プロセス	3.66% (2.58)	5.73% (3.33)	5.68% (3.41)	22.8***	P>C, N>C
認知的プロセス	15.8% (5.44)	17.1% (3.97)	18.5% (4.22)	14.3***	N>P>C
知覚的プロセス	2.77% (1.43)	2.69% (1.78)	2.36% (1.65)	3.11*	
生体的プロセス	4.21% (2.60)	1.85% (2.07)	2.38% (2.23)	46.5***	C>P, C>N
動因	6.57% (2.70)	9.85% (4.05)	9.34% (2.90)	46.9***	P>C, N>C
相対性	22.6% (4.71)	19.5% (4.41)	20.0% (4.01)	23.1***	C>P, C>N
インフォーマル	3.09% (1.94)	3.33% (1.84)	3.74% (1.80)	5.24**	N>C

注：表中の値は総語数に対する割合 (0-100%); <sup>a</sup>実数; \*\*\**p* < .001, \*\**p* < .01, \**p* < .05. ポジティブ感情条件164名、ネガティブ感情条件179名、統制条件157名

#### 4 考察

本研究では、感情辞書の LIWC2015 を日本語に翻訳した J-LIWC2015 のプロトタイプ版を作成し、その妥当性の検証を行った。翻訳にあたっては、原版との対応を重視し、機械翻訳と人力翻訳を組み合わせることで訳語を抽出し、カテゴリーについても日本語の特徴を踏まえて取捨選択を行った。その後、クラウドソーシングサービスで募集した参加者に対して、感情状態を操作した後に個人的な出来事についての記述を行ってもらった後、記述内容を J-LIWC2015 で分析し、記述に含まれる語のカテゴリーの割合について条件間で比較検討を行った。その結果、記述内容に多く含まれる語のカテゴリーは、操作された感情状態に対応していた。すなわち、ポジティブな感情を喚起した場合は、記述内容に占めるポジティブな語の割合が多くなり、ネガティブな感情を喚起した場合は、記述内容に占めるネガティブな語の割合が多くなっていた。また、特定の感情を喚起せず、平凡な出来事を記述した場合に比べて、ポジティブあるいはネガティブな感情を喚起した場合には、さまざまな心理的プロセスを反映した語や、動因に関する語の使用割合が多くなっていた。

言語を用いた感情的な体験の再体制化 (意味づけ) には、体験に関する洞察や原因の探求などの認知的プロセスが関与し、特にネガティブな感情体験については、自己理解に基づく再体制化のプロセスが重要な役割を果たす (羽鳥・石村, 2016)。また、日本人の感情的な体験は他者との社会的関係に根ざしていることが多い (Kitayama, Mesquita, & Karasawa, 2006)。本研究でも、統制条件と比較して、ネガティブ感情条件とポジティブ感情条件では、認知的カテゴリーや社会的カテゴリーに属する語が記述内容に占める割合が多かった。これらの結果は、本研究の参加者が感情体験を言語的に報告する際、その内容に対する意味づけを積極的に行っていたことを示唆するものである。また、LIWC2007 を用いて筆記療法での記述内容と精神的健康との

関連を検討した研究 (Niederhoffer & Pennebaker, 2009) では、ポジティブ感情のカテゴリーに属する語の使用割合の増加が、健康状態の増進につながる事が報告されている。本研究では、ネガティブ感情条件と統制条件との間で、ポジティブ感情のカテゴリーに属する語の使用割合に有意な差はみられなかった。このことから、本研究におけるネガティブ感情条件の参加者は、感情体験に対するポジティブな意味づけを一定程度行っていたと解釈できる。

今後の課題としては、人力評定によって、語のカテゴリーを日本語の意味や文脈に即したものに改めて割り当てる必要がある。他言語版の LIWC ではカテゴリーの再割り当てを行った上で辞書が公開されており、J-LIWC2015 についても、公開に先立って、他言語版と同様の基準で辞書の内容を吟味することが急務である。また、プロトタイプ版では、感情体験の記述内容に含まれていた語に対する辞書のカバー率は決して高くはない (約 60%)。この点に関しては、登録単語数の拡充とともに、元データに含まれるノイズ (誤字・脱字、ひらがな一文字で J-LIWC2015 に含まれない語) の除去を前処理の段階で行うことで、カバー率を向上させることが必要である。今後は、日本語の大規模コーパスや Google N-gram コーパスなどを利用して、頻度情報に基づく頻出単語の新規追加や、既存の単語の取捨選択を行い、必要に応じて日本語に適したカテゴリーの追加 (丁寧語など) を行うことも求められる。また、J-LIWC2015 に登録する語は、形態素解析に適した形式に調整する必要があるが、日本語の場合、形態素解析に用いる辞書やアルゴリズムによって分析のベースとなる語形が変化するため、汎用性の高い辞書を作成することが難しい。機能語に関しては、既存の形態素解析ソフトウェアの解析結果をベースに解釈を行うことも有効と考えられる。

このように、いくつかの課題は残されているものの、J-LIWC2015 は、日本語テキストの多面的評価のベースとなる標準的な感情辞書として、高い有用性を持つことが示された。ひらがな、カタカナ、漢字の 3 種類の文字体系を持つ日本語では、漢語、和語、カタカナ語が混在し、オノマトペのバリエーションも豊富であるため、語彙数が他のヨーロッパ系言語に比べて多く、平均的な日本語母語話者 (大学生) では 30,000 語から 50,000 語とも言われている (佐藤・田島・橋本・松下・笹尾, 2017)。今後、感情辞書のカバー率を増やすためには、機械学習を用いた類語の自動登録などの新たな技術が不可欠であり、J-LIWC2015 は学習のベースとなる教師データとして重要な意味を持つ。J-LIWC2015 は、カテゴリーの再割り当てや登録語の再吟味を行った後、他言語版と同様、LIWC2015 のサイトで公開予定である。本辞書の公開によって、日本語テキストデータを対象とした心理的側面の解析がさらに発展することを願う。

## 【参考文献】

- Baikie, K. A., Geerlings, L., & Wilhelm, K. (2012). Expressive writing and positive writing for participants with mood disorders: An online randomized controlled trial. *Journal of Affective Disorders, 136*, 310-319. doi:10.1016/j.jad.2011.11.032
- Cheng, Z., Caverlee, J., & Lee, K. (2010). *You are where you tweet: a content-based approach to geo-locating twitter users*. Paper presented at the Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada.
- Frimer, J. A., Aquino, K., Gebauer, J. E., Zhu, L., & Oakes, H. (2015). A decline in prosocial language helps explain public disapproval of the US Congress. *Proceedings of the National Academy of Sciences, 112*, 6591-6594. doi:10.1073/pnas.1500355112
- 羽鳥 健司・石村 郁夫 (2016). 筆記と語りに関するポジティブ心理学的研究の概観 東京成徳大学大学院心理学研究科臨床心理学研究, *16*, 205-212.
- Huang, C.-L., Chung, C. K., Hui, N., Lin, Y.-C., Seih, Y.-T., Lam, B. C. P., . . . Pennebaker, J. W. (2012). The development of the Chinese linguistic inquiry and word count dictionary. *Chinese Journal of Psychology, 54*, 185-201.
- Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology, 91*, 890-903. doi:10.1037/0022-3514.91.5.890
- Kudo, T., Yamamoto, K., & Matsumoto, Y. (2004). Applying conditional random fields to Japanese morphological analysis. *IPSJ SIG Notes, 2004*, 89-96.
- Matsuo, A., Sasahara, K., Taguchi, Y., & Karasawa, M. (2018). *Development of the Japanese Moral Foundations Dictionary: Procedures and applications*. (1804.0087). arXiv.

- 並川 努・谷 伊織・脇田 貴文・熊谷 龍一・中根 愛・野口 裕之 (2012). Big Five 尺度短縮版の開発と信頼性と妥当性の検討 心理学研究, 83, 91-99.
- 那須川 哲哉・上條 浩一・山本 眞大・北村 英哉 (2016). 日本語における筆者の性格推定のための言語的特徴の調査 言語処理学会第 22 回年次大会発表論文集, pp. 1181-1184.
- Niederhoffer, K. G., & Pennebaker, J. W. (2009). Sharing one's story: On the benefits of writing or talking about emotional experience. In S. J. Lopez & C. R. Snyder (Eds.), *The Oxford Handbook of Positive Psychology (2nd ed.)* (pp. 621-632). New York: Oxford University Press.
- 及川 晴・及川 昌典 (2012). 感情抑制が顕在モードと潜在モードに及ぼす影響 社会心理学研究, 28, 24-31.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The Development and psychometric properties of LIWC2015*. The University of Texas at Austin.
- 佐藤 尚子・田島 ますみ・橋本 美香・松下 達彦・笹尾 洋介 (2017). 使用頻度に基づく日本語語彙サイズテストの開発 : 50,000 語レベルまでの測定の試み 千葉大学国際教養学研究, 1, 15-25.
- 佐藤 敏紀・橋本 泰一・奥村 学 (2016). 単語分かち書き用辞書生成システム NEologd の運用一文書分類を例にしてー 自然言語処理研究会研究報告, 229, 1-14.
- 柴田 大作 (2018). 自然言語処理による認知症スクリーニング 奈良先端科学技術大学院大学修士論文 (<http://library.naist.jp/dspace/handle/10061/12412>)
- 総務省 (2017). 平成 29 年版情報通信白書 (<http://www.soumu.go.jp/johotsusintokei/whitepaper/h29.html>)
- Stockton, H., Joseph, S., & Hunt, N. (2014). Expressive writing and posttraumatic growth: An Internet-based study. *Traumatology: An International Journal*, 20, 75-83. doi:10.1037/h0099377
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 10, 178-185.
- 山本 眞大・那須川 哲哉・上條 浩一・北村 英哉 (2016). LIWC2001 手作業翻訳の方針と半自動翻訳手法の提案 言語処理学会第 22 回年次大会発表論文集, 1185-1188.

### 〈 発 表 資 料 〉

題 名	掲載誌・学会名等	発表年月