

災害報道文の特徴語抽出

砂岡和子 (早稲田大学政治経済学術院)

Raymond SHEN (早稲田大学人間科学研究科博士課程)

1 自然災害報道と共感醸成

大災害の復興支援は短期的な物資援助以外に、風評被害対策など長期的社会支援が重要である。2011年日本東北大地震後、大学授業でも被災者の苦難を思い復興支援への関心が強まった。筆者が担当する中国語授業では討論テーマの選定に配慮した。注目度が高い報道記事が被災者への共感を生むとは限らない。例えば、福島第一原発事故を受け中国での塩の買い占めを報じた中文報道¹を読んだ日本人学生は忌避感情を露にした。家族の絆は共感を呼ぶテーマだが、中国人学生推薦の四川大地震から自分の命と引き換えに子供を守った母親の記事²に、日本人学生の多くはマスコミが祭りあげた美談かと反発した。同じく四川地震で子供を亡くした親へ対する子作り特例に関する報道記事は³権力の人権介入が共感を減衰し、賛同者は25%に留まった〔砂岡和子2011〕参照〕。自然災害支援など地域を越えた共感醸成が求められる時、客観的な基準で報道文や討論テーマを抽出する方法はないだろうか。

2 報道記事特徴語の抽出手法

先行研究に、テキストマイニングの手法で地震災害発生後のメディア情報を自動的にアーカイブし、統計的処理に基づいて災害報道のキーワードや内容変化を時系列で提示し、正確な情報提供と風評被害対策に役立てようと試みがある。〔佐藤翔輔・林春男・牧紀男・井ノ口宗成(2007)、安田・海野(1977)、長尾光悦・大内東(2011)参照〕。これら社会システム工学が扱う語句は時の話題を反映し易いイベント名詞が主体となる。長尾他(2011)に従えば、頻出語(イベント名詞)はアルゴリズムが把握し易く抽出成功の確率が高いが、意味分類はカテゴリー枠が簡単過ぎて差異を検出できないという。そこで中文災害報道記事を収集し、特徴語抽出のための情報を付加したコーパスから動詞句(以後 VV と略記⁴)を抽出し、出現頻度や意味概念を分析した。別途、信頼性判定用に準備した一般中文テキスト中の動詞句と定量比較を行なった。VV を分析対象に選んだ理由は、実詞であること、人間活動の諸相を安定して反映する語彙と期待できるためである。以下、チャート〔図1〕に示すプロセスに沿って特徴語抽出の概要を説明する。

2.1 中文報道記事コーパス作成

最初に2種の中文報道記事の収集を行なった。1種は四川大地震の災害報道記事(「四川」または「S」と略記)で、①被災者の生の声や行動を伝える文章を、②政府報道に批判的な論調の新聞雑誌から、③地震発生からの経年バランスを考慮し、主観判断で集めた〔テキスト出典は砂岡和子

¹ 日本地震、中国抢盐，皇帝不急太监急？《南方报网》2011年3月17日

² 母亲用身体护住婴儿，婴儿被救出时仍熟睡，《网易》2008年5月17日

³ 艰难重生，《南方人物周刊》2009年5月11日

⁴ VV は形容詞句を含む6種の複合動詞の総称。詳細は中国語形態素解析機 CNP の品詞ラベル参照のこと

(2011)参照]. もう一種は一般報道記事で, 「四川コーパス」と同時期・同字数の「人民日報」テキスト⁵である(「人民」または「R」と略記). 最終的に四川地震発生直後, 半年後, 1年, 2年, 3年の経年別に, 各々約8万華字の中文報道記事データベースを作成した[表1].



[図1] 特徴語抽出プロセス

四川地震報道記事			人民日報記事		
発信年	篇数	総字数	発信年	篇数	総字数
地震発生直後	5	18200	2008年前半	20	17807
発生後6ヶ月	5	16000	2008年後半	16	15774
発生後1年	5	15400	2009年	9	15550
発生後2年	4	15000	2010年	13	14774
発生後3年	5	16400	2011年	18	16488
計	24	81000	計	76	80393
1篇1000-8000字			1篇147-3768字		

[表1] 2種中文報道コーパス内訳

2.2 形態素解析による動詞の抽出

上記2種中文テキストを中国語係り受け解析器 (A Chinese Dependency Parser, 略称CNP) で形態素解析を行い動詞語句を抽出した. 文字化けや固有名詞などCNPの誤解析を手作業で削除校正⁶, 最終的に「四川コーパス」と「人民コーパス」から抽出した動詞数量は各々, 8461語と8668語とほぼ同数になった. 中国語の常用動詞は補語や重ね型など多種の拡張形を派生する. 形態解析で自動抽出した動詞は, 例えば“上”と“上去”のような形式的異なりが別々の語彙として登録される. そこで先に同義語を整理し動詞数をそれぞれ約半分に統合した.

2.3 頻出動詞の意味分類

動詞の意味分析は機械的な自動処理が難しいためヒューリスティックな手法で進めた. 特徴語の傾向を大まかに把握するため, 「四川」と「人民」それぞれの時間帯で上位100語に入った動詞を対象とし, その出現頻度とランキング変動を調査した. その結果, 両コーパスの動詞上位100語の異なり動詞のうち顕著な差がある一方の動詞を優先して意味分類を進めた.

意味分類は日本語の『分類語彙表』[国立国語研究所編(2004)]の枠組を用い, 中国語に該当する同書の日本語分類番号を利用することで, 「四川」と「人民」から抽出した動詞をよりの確に分類でき, 大局的な概念帰属の理解が可能となった.

2.4 経年変化比較と特徴語の抽出

両コーパスの異なり動詞上位100語を4.3の手法で意味分類後, 主に目視と手作業で共起情報や頻度分析を行なった. 分析結果を以下に述べる.

3 災害報道記事の動詞の特徴

上位100異なり動詞につき意味分類を行なった結果, 次のような特徴が明らかになった.

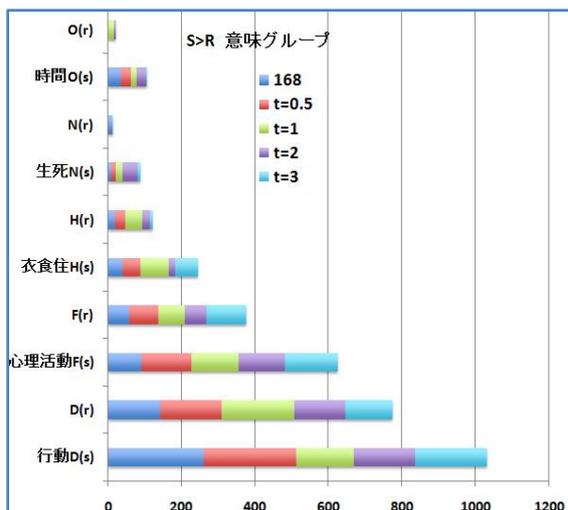
- 1) 「四川」には「行動」「心理活動」系動詞の出現頻度数が「人民」を大きく上回る ([図2] D, F).

⁵ 「人民コーパス」はランダム抽出した記事約135000語から更に「四川コーパス」の経年字数に合せて選出した

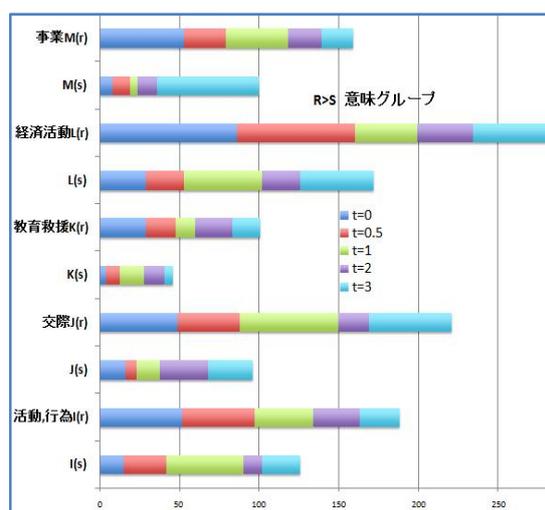
⁶ 誤解析による削除語彙は全体で約3%と極少数に留まる

ことに地震発生直後 0—2 年に顕著な差がある。例；(/後の数字は「四川」の経年数) “得知 (分かる) /0” “找到 (見つかる) /0” “发现 (見つける) /0.5” “哭 (泣く) /0.5” “叫 (叫ぶ) /0.5” “考虑 (思案する) /1” “回忆 (思い出す) /2” “找 (探す) /3”。「行動」「心理活動」動詞の増加は災害報道の特徴と言えよう。ランキングは下がるものの「衣食住」([図 2]H)，「生死」(同 N) “死, 死亡, 亡, 去世 (死ぬ) /0-3”，“怀孕, 生育, 生, (妊娠する, 生む) /1-3”，「時間」(同 O) 類の動詞が一貫して多出する。「人民」の上位には現れず，被災民にとり感情処理と復興が急務であることを反映する動詞と考えられる。

2) 「人民」の動詞には平時の社会生活である「活動, 行為」「交際」「経済活動」「事業」(それぞれ [図 3] I, J, L, M) に関する語彙が常に最上位を占める ([図 3])。「教育救済 (K)」が「四川」ではなく「人民」に多いのは“释放”“抓获”“安置”“维护”“安排”など刑事事件や治安維持関連の語彙が新聞記事に多いため、数は少ないが「四川」の (K) 枠には“复课”“复学”など教育復興関連の動詞が入っており、特徴語抽出に当たり [教育救済] など下位分類に留意が必要と分かる。



【図 2】(四川>人民) 経年特徴語比較



【図 3】(人民>四川) 経年特徴語比較

4 災害ストレス回復モデル

一般記事との比較で災害報道の意味的特徴を何点が掴むことができたが、「四川」全体 VV の構造特色はなにか？保坂隆 (2011, 66-94) によれば、一般人には危機に遭遇すると「衝撃・ショック」「防御的退行」「承認」「適応」の 4 段階を踏んで克服をはかり、被災者の心の復興もこれに酷似して「英雄期」「ハネムーン期」「幻滅期」「再建期」を経るという。「四川コーパス」各時期の高頻度動詞を配して見ると、中国被災者の心身の危機ストレスとその回復過程が保坂「危機モデル」の 4 段階に当てはまる ([図 4])。

日訳：“活 (生きる, 生きている)” “等候 (待つ)” “打飯 (食糧を手に入れる)” “挂职 (在職中の公務員が被災地へボランティアに出かける)” “抑郁 (うつ病になる)” “跳楼 (飛び降り自殺する)” “喜欢 (好き)” “恢复 (回復する)” “重建 (再建する)” “照顾 (世話する)” “赶到 (駆けつける)” “打工 (働く)” “遇难 (災害で犠牲になる)” “通往 (開通する)” “问 (尋ねる)” “读 (勉強する, 就学する)” “怕 (恐れる, 怖い)” “睡 (眠る)” “搬 (引っ越す)” “支援 (支援する)”

“搬（引っ越す）”“赚（稼ぐ）”

危機遭遇経過	災害ストレス復興経過	救済活動の重点	
第1段階	英雄期[衝撃・ショック期]	生命救助	情報・ 資源管理
第2段階	ハネムーン期[防衛的退行期]	社会フロー安定	
第3段階	幻滅期[承認期]	社会ストック再建	
第4段階	再建期[適応期]		
四川コーパスtime	各期高頻度動詞	通年高頻度動詞	
0-0.5年	“活”“等候”“打饭”“挂职”	“找”“逃”“跑”“送”	
1年	“等待”	“读”“哭”“怕”	
2年	“怀孕”“结婚”“搬”“抑郁”“跳楼”	“希望”“失去”“打工”	
3年	“死”“失踪”“遇难”	“笑”“喜欢”	

[図4]危機4段階モデルと四川記事の高頻度動詞⁷

「四川コーパス」には第2,第3段階でも依然として“埋”“死”“遇难”など被災直後の衝撃・ショック期と同じ動詞が多出し,心の再建に長い時間が必要と分かる.林春男(2001,52-67)は復興支援を「命を守る活動(72時間内)」「社会のフローを安定させる活動(100日)」「社会ストックの再建(10年)」「情報資源管理(ロジスティック)(0-10年)」に分け,これらが同時並行的に展開する必要性を説く.報道記事から被災者たちの心的ストレスのサインを読み取り,早期回復に向けた支援に活かすことができよう.

5 まとめと課題

異文化間コミュニケーションの忌避感情を緩和し,共感醸成を促進する災害報道記事の定量的選択基準を求めるべく,中国四川大地震の記事中の動詞について頻度分析と意味分類を行った.検証用の一般記事コーパス中の動詞との比較から,災害時の動詞の出現特徴を抽出した.分析の結果,中文コーパスの動詞は一般災害時の危機モデルや心的ストレス変化と近似する.抽出した頻出動詞は日本人学生が選好する共感キーワードとも概念的な一致率が高く,文化や言語の壁を越え被災者との共感醸成可能な特徴語と言える.災害報道特徴語を定量化し一定の規則を抽出するにはコーパスの範囲と量を増やし検証を進める必要がある.今後継続して分析する.

謝辞: 本稿の執筆にあたって中国北京語言大学の楊爾虹教授,同劉鵬遠博士,同鄒紅建氏から人民日報コーパスを提供頂いた.また高度言語情報融合フォーラム(ALAGIN)より中国語解析プログラムCNPの提供を受けた.ここに感謝する.本研究は平成22-24年度文部科学省科学研究費補助金〔基盤(C)課題番号:22520445:研究代表者砂岡和子〕および羅鳳珠代表「歴代語言知識庫建置計畫」九十八年度蔣經國國際學術交流基金會の助成で進行中の成果の一部である.

参考文献: 国立国語研究所編(2004),分類語彙表(増補改訂版)大日本図書

保坂隆編(2001),災害ストレス:直接被災と報道被害 角川書店

林春男(2001),率先市民主義:防災ボランティア論講義ノート 晃洋書房

長尾光悦,大内東(2011).観光地に対する風評被害の変遷と対応分析,人工知能学会誌(特集:観光と知能情報),26(3),264-271

佐藤翔輔・林春男・牧紀男・井ノ口宗成(2007).TFIDFを用いた災害・危機に関する言語資料体からのキーワード自動抽出手法の外的妥当性の検証 地域安全学会論文集,9,65-74

砂岡和子(2011).被災から復興へ中国語希望表現の特徴抽出,日本中国語学会第61回大会予稿集,280-284

⁷ 保坂隆(2011),林春男(2001)を参照し筆者が加工して作成.