様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ９ 月 １１ 日現在

機関番号：３２６８９
研究種目：若手研究
研究期間：2022～2023
課題番号：２２Ｋ１８１５４

研究課題名（和文）Negative emotions in literature: a computational approach to tone and mood

研究課題名（英文）Negative emotions in literature: a computational approach to tone and mood

研究代表者
ＯＨＭＡＮ　Ｅｍｉｌｙ（Ohman, Emily）

早稲田大学・国際学術院・講師テニュアトラック

研究者番号：６０９０６５４３

交付決定額（研究期間全体）：（直接経費）　　1,500,000円

研究成果の概要（和文）：このプロジェクトでは、文学的テキストにおける気分を計算によって検出、認識、分析するための新しいツールを開発することができた。その結果
小説の第一章の最初の500語のみを用いて、小説の雰囲気を高い精度で計算によって検出できることを示すことができた。さらに、1,000のフィンランド語の小説と約10,000の英語小説の感情アークコーパスを作成した。
さらに、ネガティブな感情に注目し、特に英語と日本語における文化的概念としての恥と罪悪感を検討し、恥は公的な経験として描かれることが多く、罪悪感は悲しみに似た私的な要素と、公然と罪を認める動機のような公的な側面の両方を含む感情であることを示した。

研究成果の学術的意義や社会的意義
This research has added to the understanding of the literary concept of mood and how to detect it computationally. Additionally, the project has helped with the understanding of shame and guilt across and within cultures. It has also provided several open access corpora and other tools.

研究成果の概要（英文）：With this project we were able to develop new tools for the computational detection, recognition, and analysis of mood in literary texts. We were able to
to show that the mood of a novel can be computationally detected with high accuracy using only the first 500 words of the first chapter. Additonally, we created emotion arc corpora for 1,000 Finnish novels and nearly 10,000 English novels.
Furthermore, we focused on negative emotions and examined specifically shame and guilt as cultural concepts in English and Japanese to show that shame is often portrayed as a public experience and guilt as an emotion that encompasses both private elements, akin to sadness, and public aspects, such as the motivation to openly acknowledge a transgression.

研究分野：Computational humanities

キーワード：Emotion detection　sentiment analysis　digital humanities　literary analysis　NLP　machine learning

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

１．研究開始当初の背景

Existing sentiment analysis and computational literary analysis approaches felt inadequate as support tools for literary scholars whose research focus often differs from that of computer scientists using literature as a source of data. Specifically, the current focus on tone and mood in literary analysis has not become a research focus in computational literary studies and beyond attempts at sentiment analysis of novels, no impactful computational approaches to identifying mood or tone in literary texts had been attempted. Literary scholars tend to be suspiscious of quantitative approaches applied to their field where data is rich in subtle nuance and word play. Therefore, not only did it seem like computational literary studies were not focusing on the same topics as traditional literary scholars, but that the more traditional scholars were not really aware of developments in computational literary studies that could provide them with useful tools that would not ignore the layers of nuance in language and subtext in narrative texts.

２．研究の目的

This project focused on the construction of tone and mood in literary fiction with the aim of pinpointing how negative emotions in particular are expressed and how they affect the mood of the text using computational emotion analysis fine-tuned for literary studies. A part of the goal with this study was also to both create tools for literary scholars, and to disseminate information about such tools and methods to actual literary scholars and not just computer scientists.

３．研究の方法

First, emotion lexicons for Finnish were developed. These lexicons were then used to weight word embeddings, that is, vector representations of semantic context on models that were fine-tuned using a corpus of nearly 1,000 Finnish novels. This method was later used for a similar project with English novels and to study the different semantic contexts of shame and guilt in English and Japanese. The literary corpora were annotated with emotion arcs using this method to study the concept of affect as a proxy for mood.
Most papers related to this project were produced as part of international collaborations.

４．研究成果

We found that mood can be detected from the opening paragraphs of a book. We compared the results for the first 300 tokens, 500 tokens, and the first paragraph of each chapter combined with expert literary analyses and determined that a token chunk between 300-500 tokens was a reliable amount of text to measure emotion arcs produced by our method. In addition, this project yielded a corpus of emotion arcs for 975 Finnish novels and a python package to chapterize Finnish novels. We also found that this method held up well against large language model-aided sentiment analysis as well as human annotations (see Fig. 1). In this case we used fear as a proxy for negative and joy as a proxy for positive to compare to binary sentiment models.

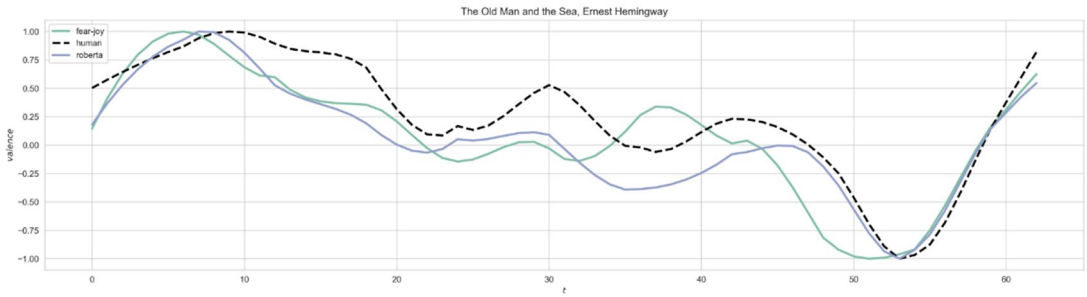

Figure 1 Comparison of different approaches for narrative dynamics.

We then compared how such emotions arks, specifically the entropy, as well as the skewness

and steepness of the arcs correlate with two proxies of literary reception: library holdings and GoodReads ratings indicating that specific types of emotion arcs and combinations are the key to capturing the reader and is closely linked to mood. We also published a corpus for over 9,000 English novels annotated with emotion arcs.
Finally, we looked at shame and guilt as cultural concepts contrasting Japanese and English. We looked at these emotion terms directly and identified semantically similar expressions with the help of semantic vector space models. We found some overlap but also distinct usages of these semantic concepts in the two languages and were able to provide some insights into how these emotion words can denote internal emotional states.

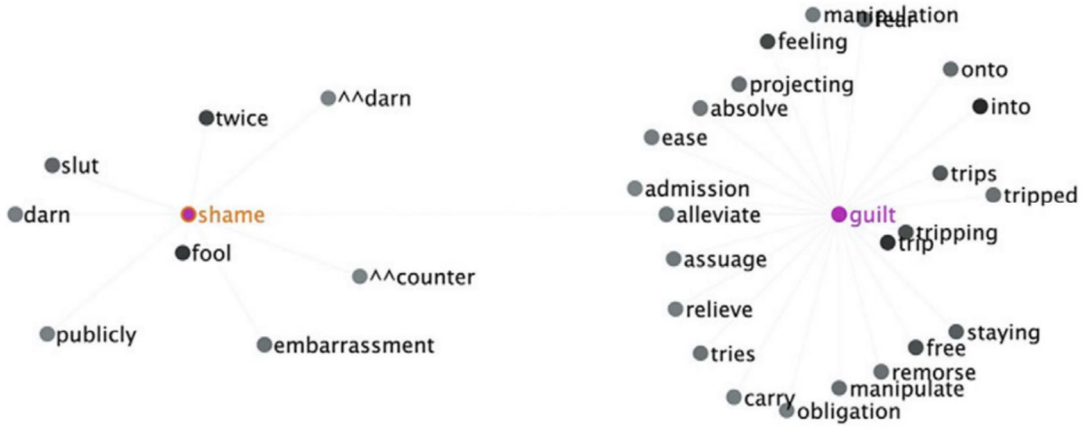

Figure 2 The semantic spaces of shame and guilt in English.

A key goal from the start was to disseminate this information to many different audiences. As NLP work is usually published in NLP-conference proceedings, they have wide reach within the community, but are less impactful to a wider audience. Thus, in addition to NLP conference proceedings we also chose to publish in interdisciplinary (Journal of Data Mining and Digital Humanities) and traditional journals (Language and Cognition, Scandinavian Studies Journal).
Because of this work, I was invited as a keynote speaker to the WASSA workshop at NAACL in 2023 and was thus able to reach an even larger audience with the new methodologies as well as the findings presented in this report.

５．主な発表論文等

〔雑誌論文〕　計8件（うち査読付論文　8件／うち国際共著　8件／うちオープンアクセス　8件）

１．著者名 Emily Ohman, Riikka Rossi	４．巻 97:3
２．論文標題 Combining Qualitative and Computational Approaches for Literary Analysis of Finnish Novels	５．発行年 2025年
３．雑誌名 Scandinavian Studies Journal	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Diegoli Eugenia、Ohman Emily	４．巻 -
２．論文標題 Contrasting the semantic space of ‘shame’ and ‘guilt’ in English and Japanese	５．発行年 2024年
３．雑誌名 Language and Cognition	６．最初と最後の頁 1～23
掲載論文のDOI（デジタルオブジェクト識別子） 10.1017/langcog.2024.6	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Emily Ohman, Yuri Bizzoni, Pascale Feldkamp Moreira, Kristoffer Nielbo	４．巻 -
２．論文標題 EmotionArcs: Emotion Arcs for 9,000 Literary Texts	５．発行年 2024年
３．雑誌名 Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)	６．最初と最後の頁 51-66
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Pascale Feldkamp Moreira, Yuri Bizzoni, Emily Ohman, Kristoffer Nielbo	４．巻 -
２．論文標題 Not just Plot(ting): A Comparison of Two Approaches for Understanding Narrative Text Dynamics	５．発行年 2023年
３．雑誌名 Proceedings of CHR 2023: Computational Humanities Research Conference	６．最初と最後の頁 191-205
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Ohman Emily、Rossi Riikka	4．巻 NLP4DH
2．論文標題 Affect as a proxy for literary mood	5．発行年 2023年
3．雑誌名 Journal of Data Mining & Digital Humanities	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.46298/jdmdh.11164	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Laaksonen Salla-Maaria、Paakkonen Juho、Ohman Emily	4．巻 -
2．論文標題 From hate speech recognition to happiness indexing: critical issues in datafication of emotion in text mining	5．発行年 2023年
3．雑誌名 Handbook of Critical Studies of Artificial Intelligence	6．最初と最後の頁 631～641
掲載論文のDOI（デジタルオブジェクト識別子） 10.4337/9781803928562.00064	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Emily Ohman	4．巻 3232
2．論文標題 SELF & FEIL: Emotion Lexicons for Finnish	5．発行年 2022年
3．雑誌名 Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB 2022)	6．最初と最後の頁 424-432
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1．著者名 Emily Ohman and Riikka Rossi	4．巻 2
2．論文標題 Computational Exploration of the Origin of Mood in Literary Texts	5．発行年 2022年
3．雑誌名 Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities	6．最初と最後の頁 8-14
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計8件（うち招待講演　5件／うち国際学会　5件）

1．発表者名
Emily Ohman

2．発表標題
Affective Datafication of Narratives: measuring affect, emotion, and mood in literary texts

3．学会等名
13th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis（招待講演）（国際学会）

4．発表年
2023年

1．発表者名
Emily Ohman

2．発表標題
Datafication of Affect: methodological, theoretical and epistemological concerns

3．学会等名
HelDig Research Seminar (University of Helsinki)（招待講演）

4．発表年
2024年

1．発表者名
Emily Ohman

2．発表標題
Mining Social Media Data for Insights into Language Use

3．学会等名
Language of Social Media course at University of Helsinki（招待講演）

4．発表年
2024年

1．発表者名
Emily Ohman

2．発表標題
Affective Queer Narratives on Japanese Online Fora

3．学会等名
Japanese Association of Digital Humanities

4．発表年
2023年

１．発表者名
Emily Ohman

２．発表標題
Computational Exploration of the Origin of Mood in Literary Texts

３．学会等名
Natural Language Processing for Digital Humanities（国際学会）

４．発表年
2022年

１．発表者名
Emily Ohman

２．発表標題
Sentiment analysis: Critical approaches

３．学会等名
Rajapinta computational social science unconference（招待講演）（国際学会）

４．発表年
2022年

１．発表者名
Emily Ohman

２．発表標題
SELF & FEIL Emotion Lexicons for Finnish

３．学会等名
Digital Humanities in the Nordic and Baltic Countries Conference 2022（国際学会）

４．発表年
2022年

１．発表者名
Emily Ohman

２．発表標題
Computational exploration of emotions in literature

３．学会等名
BINUS University Digital Humanities Symposium（招待講演）（国際学会）

４．発表年
2023年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

The emotion lexicons
https://github.com/Helsinki-NLP/SELF-FEIL
A Finnish version of the chapterize package
https://github.com/esohman/chapterize-fi
The literature corpus online
https://github.com/esohman/FinLit-corpus
The EmotionArcs online:
https://github.com/yuri-bizzoni/EmoArc

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究協力者	Rossi Riikka (Rossi Riikka)	University of Helsinki	
研究協力者	Bizzoni Yuri (Bizzoni Yuri)	Aarhus University	
研究協力者	Diegoli Eugenia (Diegoli Eugenia)	University of Bologna	
研究協力者	Feldkamp-Moreira Pascale (Feldkamp-Moreira Pascale)	Aarhus University	
研究協力者	Laaksonen Salla-Maaria (Laaksonen Salla-Maaria)	University of Helsinki	

6．研究組織（つづき）

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究協力者	Paakkonen Juho (Paakkonen Juho)	University of Helsinki	
研究協力者	Nielbo Kristoffer (Nielbo Kristoffer)	Aarhus University	

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計1件

国際研究集会	開催年
Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages	2023年～2023年

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
フィンランド	University of Helsinki			
デンマーク	Aarhus University			
イタリア	University of Bologna			