An Analysis of Acknowledgments in NLP Conference Proceedings

Winston Wu
Computer Science and Engineering
University of Michigan
wuws@umich.edu

Abstract

While acknowledgments are often overlooked and sometimes entirely missing from publications, this short section of a paper can provide insights on the state of a field. We characterize and perform a textual analysis of acknowledgments in NLP conference proceedings across the last 17 years, revealing broader trends in funding and research directions in NLP as well as interesting phenomena including career incentives and the influence of defaults.

1 Introduction

A research project is seldom a solo endeavor. Different entities contribute ideas, expertise, labor, money, and many other factors that lead to a successful project. In a publication, the most salient contributors are the authors, whose names are front and center on page one. In this paper, we investigate the so-called "lesser" contributors, whose names exist in the acknowledgments section of a publication, typically right before the references. Specifically, we ask several research questions:

- How common are acknowledgments?
- Who are acknowledged?
- What are they acknowledged for?
- What else can we learn from acknowledgments?

Our analysis of acknowledgments in ACL and EMNLP conference proceedings presents a view of the state of the field of natural language processing, including:

- trends in the use of acknowledgments
- broader funding trends based on international government investment
- research trends based on industry gifts
- trends in grant life-cycle and productivity
- culture-specific career incentives of being a corresponding author
- the influence of defaults on authors' word choice

2 Related Work

Acknowledgments have been investigated in both the social sciences and computer science communities. Scrivener (2009) analyze acknowledgments in history students' dissertations. Tang et al. (2017) performed a cursory analysis of funding acknowledgments in Thomson Reuter's Web of Science database. Giles and Councill (2004) analyzed computer science articles from the CiteSeer database[1], identifying the most common acknowledged entities. Part of our work is similar in design but focuses specifically on the field of NLP rather than the broader field of computer science. Paul-Hus and Desrochers (2019) performed a qualitative analysis of acknowledgments, looking at word usage patterns. Our study goes into more depth, looking at specific entities that are acknowledged, and what they are acknowledged for.

Grant funding is typically acknowledged in the acknowledgments section, and there is some recent interest in identifying funding sources and grant numbers as an information extraction task (Dai et al., 2019; Bian et al., 2021). Our paper does not tackle the task of grant funding detection but rather analyzes general trends in grant funding, as well as other trends. Within the NLP community, a line of work has extracted insights from trends and citations in NLP publications (Mohammad, 2020a,b,c,d), but has not focused specifically on acknowledgments.

3 Data

We analyze proceedings of two conferences: the Annual Meeting of the Association for Computational Linguistics (ACL), and the Conference on Empirical Methods in Natural Language Processing (EMNLP). ACL and EMNLP are top-tier international NLP conferences with a broad scope and thus would be representative of the

[1] http://citeseer.ist.psu.edu

Proceedings of the Fifth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), pages 157 - 163
November 7, 2022

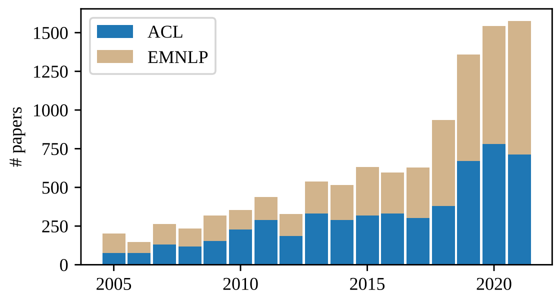

Figure 1: Number of papers in the ACL and EMNLP main conference proceedings from the past 16 years, highlighting the exponential growth of the field.

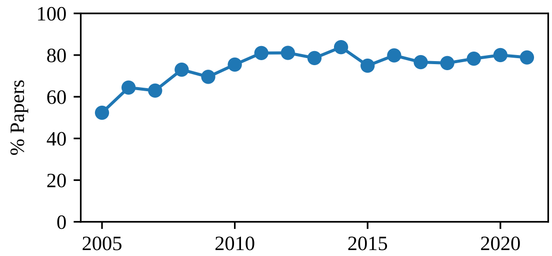

Figure 2: Percentage of papers containing an Acknowledgments section. The most recent years have stagnated around 79%.

broader NLP community. Specifically, we examine long and short papers published in the main conference proceedings from 2005–2021.[2] We download the proceeding PDFs from the ACL Anthology,[3] splitting the file into separate papers and extracting text using PyMuPDF.[4] We extract the acknowledgments section by searching for the word *Acknowledgments* and its spelling variants, followed by some manual cleaning efforts. We then perform dependency parsing and named entity recognition on all acknowledgments using spaCy's en_core_news_lg model.[5] Figure 1 presents the total number of ACL and EMNLP papers per year, from which we extract a total of 7,838 acknowledgments.

4 Characterizing Acknowledgments

This section, which forms the bulk of our paper, investigates several research questions that can be answered by analyzing papers' acknowledgments.

4.1 How common are acknowledgments?

In the nascent years of NLP, it was common to see papers published with a single author. For example, in the first iteration of EMNLP (1996), 7 of the 15 papers contained a single author, and 4 of the 15 papers contained acknowledgments (Melamed, 1996; Brants, 1996; Oflazer and Tur, 1996; Mooney, 1996). Nowadays, it is normal to see 4 or 5 author collaborations, and even more especially from large industry research groups. Thus is it interesting to see how often an acknowledgments section occurs at all.

Figure 2 presents the percentage of papers from each year containing an Acknowledgments section. Over time, the proportion of papers containing acknowledgments has slowly increased, though in recent years, the proportion has hovered around 79%. Acknowledgments are not mandatory, and it is difficult to investigate why authors do not include acknowledgments. Perhaps the publication was truly an isolated effort: the authors did not receive any funding, did not engage in any helpful conversations with others, and did not receive any useful feedback from the reviewers.

4.2 How long are acknowledgments?

Before diving into the contents of acknowledgments, we first investigate the surface-level question of how long are acknowledgments. The mean length of acknowledgments was 305.2 characters (roughly a fifth of a 2-column page), with a standard deviation of 172.6 characters. The shortest acknowledgment, in Singla et al. (2020), was 35 characters: "This work was supported by the NIH." The longest acknowledgment, in Nivre et al. (2007) was an impressive 2,408 characters; we will not reprint it here. A histogram of acknowledgment lengths is shown in Figure 3.

4.3 Who are acknowledged?

To identify acknowledged entities, we use spaCy to perform dependency parsing and named entity recognition on the acknowledgments. To account for variations in sentence structure and avoid overcounting, we (1) identify abbreviations for common government agencies, (2) ignore any names that are the subject of a "thanking" verb (*thank*, *acknowledge*, *appreciate*, *enjoy*), (3) ignore any names that are the subject of a passive "support-

[2] This excludes workshop papers, system demonstration papers, and student research papers. We exclude conference proceedings from 2004 and older because they have not been compiled into a single file in the ACL Anthology.

[3] `https://aclanthology.org/venues/acl` and `/emnlp`

[4] `https://github.com/pymupdf/PyMuPDF`

[5] `spacy.io`

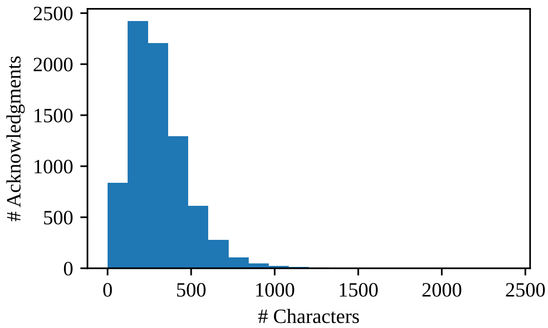

Figure 3: Histogram of the length of acknowledgments (in characters). The mean length of acknowledgments was 305 characters (std dev of 173 characters).

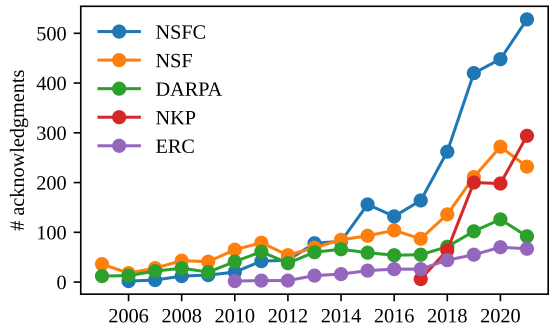

Figure 4: Top five government agency acknowledgments plotted over time. Within the last decade, there has been a drastic rise in Chinese government funding (NSFC and NKP).

Agency	Govt	Count
NSFC	China	2,408
National Science Foundation	USA	1,653
DARPA	USA	920
NKP	China	762
European Research Council	EU	348
EPSRC	UK	221
Air Force Research Laboratory	USA	161
IARPA	USA	158
Army Research Office	USA	154
Office of Naval Research	USA	147

Table 1: Most frequently acknowledged government funding agencies.[6]

ing" verb (*supported*, *funded*), (4) ignore any sentences containing *corresponding author* or *contact author* (see Section 4.3). In addition, we look for the words *reviewer* and *reviewers*, who are often acknowledged, because the conference review process includes a rebuttal phase where anonymous reviewers provide initial feedback to the authors.

Government Agencies. Government agencies fund the bulk of NLP research, largely through grants (Table 1). In the top 10 list of funders, government agencies in China, the US, and Europe are well-represented. The National Natural Science Foundation of China is the most frequently acknowledged funder, although when combined, US agencies have funded more publications. Notably, many papers are funded by military agencies, which may raise ethical concerns for some people. For example, in a recent survey of NLP researchers, 36% of respondents agree that it is plausible that AI could produce catastrophic outcomes in this century, on the level of all-out nuclear war (Michael et al., 2022).

Perhaps more interesting than aggregate counts is how the trends of funder acknowledgments have changed over the course of the past two decades (Figure 4). In the top five funders acknowledged, the last decade has seen a drastic rise in the number of Chinese government-funded publications, indicating heavy Chinese investment into NLP research. This also hints at a larger trend of global interest and participation in NLP research, which coincides with the recent (2020) creation of the Asian chapter of the ACL and the recent (2022) commitment of ACL to translate conference proceeding titles into numerous languages for greater worldwide multilingual access.

Aside: Tracking the Life-Cycle and Productivity of Grants. Acknowledgments also enable us to track a grant's life-cycle and productivity as measured by number of publications. Figure 5 shows the number of publications acknowledging several recent DARPA and ERC grants: DARPA CwC (2015), DARPA AIDA (2017-2021), DARPA MCS (2018-2023), and EU BroadSem (2016-2022). We see that it typically takes one year after the grant is announced before works funded under the grants are published. The number of publications across time also hints at the scope and success of the grants, with the number of papers decreasing as the grant comes to an end. While each funding source may keep track of such publication metrics resulting from their funds, we find that acknowledgments are another publicly available source of this information,

[6]NSFC = National Natural Science Foundation of China, DARPA = Defence Advanced Research Projects Agency, NKP = National Key Research and Development Program of China, EPSRC = Engineering and Physical Sciences Research Council, IARPA = Intelligence Advanced Research Projects Activity.

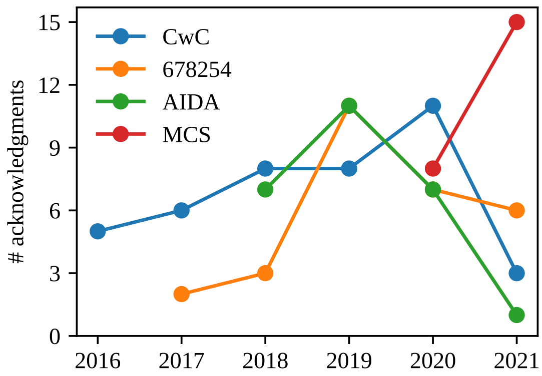

Figure 5: Grants that fund the most number of papers. The grants are DARPA CwC (Communicating with Computers), ERC Grant 678254 BroadSem (Induction of Broad-Coverage Semantic Parsers), DARPA AIDA (Active Interpretation of Disparate Alternatives), and DARPA MCS (Machine Common Sense).

which can be used to further study the impact of funding on publication rate and scientific productivity (e.g. Jacob and Lefgren, 2011).

Industry Funders Industry companies also fund a large portion of NLP research (Figure 6a). Most of these companies are acknowledged for providing including research awards, gifts, PhD fellowships. Notably, Nvidia[7] is acknowledged for grants and gifts of GPUs, which are vital resources for training neural networks. Perhaps not coincidentally, 2014, the first year Nvidia's gifts began to be acknowledged, was a year chock full of influential papers related to neural networks (e.g. Bahdanau et al., 2014; Kalchbrenner et al., 2014; Levy and Goldberg, 2014; Jia et al., 2014).

People. Figure 6b presents the most frequently acknowledged NLPers, who are all established researchers with thousands of citations. In addition, we find that the anonymous reviewers were thanked in over 51% of all acknowledgments. Peer review is important for upholding the quality of publications (Kelly et al., 2014), and it is heartening that many authors acknowledge and recognize reviewers' hard work.

Corresponding Authors. While performing this analysis, we identified a non-trivial number (185) of papers whose acknowledgments contained an indication of a paper's corresponding

[7]The NLP community does not have a consensus on the spelling of this company's name. In acknowledgments, it is alternately spelled Nvidia, NVidia, and NVIDIA.

Company	Count
Google	576
Nvidia	224
Microsoft	182
Amazon	161
Facebook	120
Bloomberg	77
Adobe	34
Salesforce	28
eBay	19
Apple	18

(a)

Person	Count
reviewers	4,065
Luke Zettlemoyer	46
Slav Petrov	36
Yoav Goldberg	29
Michael Collins	28
Tom Kwiatkowski	28
Ryan McDonald	27
Mark Yatskar	27
Kenton Lee	27
Chris Dyer	26

(b)

Figure 6: (a) Top 10 most frequently acknowledged industry companies. (b) Top 10 most helpful NLPers. The anonymous reviewers were thanked in over 51% of acknowledgments.

comment	2,861	provide	491
feedback	1,067	support	288
discussion	927	share	149
help	580	advice	119
suggestion	504	assistance	92

Table 2: The top 10 things (lemmatized) researchers are most thankful for.

author (e.g. *XX is the corresponding author of this paper*). While such sentences are common in journal articles (and are often on the first page of the paper), it is unusual to see this in NLP conference proceedings, and notably, these sentences only occur in papers published by Chinese institutions. There is a cultural explanation for the career incentive of being listed as a corresponding author: in China, promotions are heavily dependent on the number of published papers, but only papers where one is the first author or corresponding author counts toward this metric (Hvistendahl and Wang, 2014).

4.4 What are people acknowledged for?

The language in acknowledgments is highly regular, so to answer this question, we again utilize dependency parsing, identifying and lemmatizing the object of the preposition *for* in the text of the acknowledgments. The top 10 things researchers are most thankful for are listed in Table 2. The top two items, *comments* and *feedback*, are often provided by the reviewers (e.g. *We thank the reviewers for their helpful comments.*, while *discussion*, *help*, and *suggestions* are often *provided* by colleagues. *Sharing* of code, data, and results occur but is not nearly as prevalent, unfortunately.

4.5 How do you spell acknowledg... anyway?

This final question that we investigate has plagued countless authors: how is this word spelled?! We find four variants of the section title, shown in Figure 7. *Acknowledgements* is the traditional British spelling, while the American spelling omits the E. Our findings seems to indicate that most authors prefer the American spelling up until 2020, when suddenly the British spelling became more popular. However, this peculiarity has an explanation: it is likely due to a switch in the spelling of *Acknowledgments* in the paper templates[8,9] provided to the authors: the 2020 spelling (without the E) acquired an E in 2021.

Providing defaults. While the question of spelling may seem inconsequential, it raises a broader question of how the defaults provided to authors influence their choices. It is well-known that most people follow default choices (Thaler and Sunstein, 2009), and the trends in spelling usage of the word *Acknowledgments* reflect the defaults provided in the paper template. However, almost half of the acknowledgments section headers did *not* use the default spelling, indicating that these authors likely made a conscious choice: they probably deleted the section in the template and typed it back in when preparing the camera ready, rather than simply commenting out the section. Interestingly, a small minority of papers used the singular form *Acknowledgement*/*Acknowledgment*. To answer why, future work could investigate authors' writing process and workflow.

By providing default choices, institutions can influence individual's choices while not removing their freedom to choose. Recently, ACL conferences have been focusing heavily on ethics. The 2021 iteration of EMNLP required an additional section on ethical considerations in all submissions. Because this requirement was stipulated in the call for papers but was not included in the paper template, we found many variations of this section header in the proceedings, including *Ethical Considerations*, *Ethical Consideration*, *Broader Impact*, *Ethics and Broader Impact*, and *Ethics Statement*. However, the 2022 template includes *Limitations* and *Ethics Statement* sections, which we expect will be the predominant section titles in the 2022 proceedings. We also found that several papers include an addition section titled *Reproducibility* or *Code* with a link to the project's GitHub page, if the link was not already mentioned earlier in the paper. As a suggestion, if future *ACL conferences wish to emphasize other important issues such as reproducibility, they might consider adding an optional *Reproducibility* section to the paper template to nudge authors to consider this issue in their work.

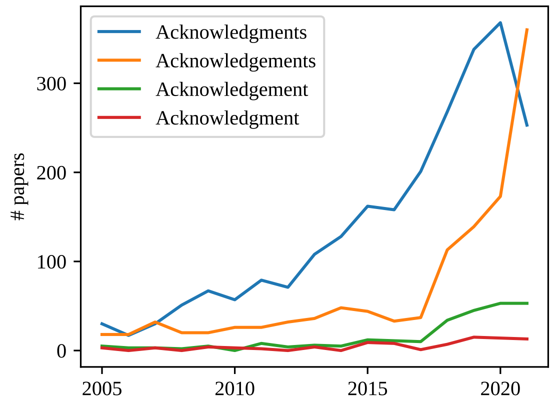

Figure 7: Spelling variation of the section header. The trend reversal between 2020 and 2021 is likely due to a switch in the spelling in the paper template provided to the authors.

5 Conclusion

While acknowledgments are seemingly insignificant and often entirely missing, in this paper we show that much can be gleaned from this short section in publications. Our analysis of acknowledgments in NLP conference proceedings reveal larger trends about the state of NLP research. Grant funding from government agencies and industry companies show increases in international participation and funding, especially from Chinese funding agencies. Grant acknowledgments also hint at the life-cycle and productivity of the grants. We identify the year 2014 as an important year of research using neural networks, corresponding with a dramatic increase in Chinese funding and industry GPU gifts. Textual analyses also reveal what researchers are most thankful for, and that some researchers indicate corresponding author, a career incentive specific to Chinese researchers. Finally, an analysis of spelling variation reveals the influence of defaults on the authors' choice of section headers. As the field of NLP continues to grow, followup analyses will help bring to light

[8]https://2020.emnlp.org/call-for-papers
[9]https://2021.emnlp.org/call-for-papers

more insights about the field and its behind-the-scenes contributors, without whom all these papers would not have been published.

Limitations

This paper investigates acknowledgments in proceedings of the ACL and EMNLP conferences, two of the largest, most prominent, international NLP conferences. This analysis unfortunately cannot account for the numerous projects that have been funded but rejected for publication. Our findings may also slightly differ for other types of publications (e.g. system demo papers, shared task papers), other venues with a geographical focus (e.g. AACL, EACL), or venues with a narrower research focus (e.g. workshops, or conferences such as LREC, CoNLL, WMT). These are all interesting avenues for investigation, and we leave these for future work.

Ethics Statement

All data used in this project is publicly and freely accessible. We do not see any ethical issues with this work.

Reproducibility

Code for acquiring the data and performing the analyses in this paper is available at `github.com/wswu/nlp-acks`.

Acknowledgments

It would be odd if a paper on acknowledgments did not include any acknowledgments. As usual, we would like to thank the anonymous reviewers for their helpful comments. We would also like to thank Frederick Zhang for helpful feedback on a draft of this paper, and Cicero Lu for early discussions that led to this work.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Junyi Bian, Li Huang, Xiaodi Huang, Hong Zhou, and Shanfeng Zhu. 2021. Grantrel: Grant information extraction via joint entity and relation extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2674–2685.

Thorsten Brants. 1996. Better language models with model merging. In *Conference on Empirical Methods in Natural Language Processing*.

Suyang Dai, Yuxia Ding, Zihan Zhang, Wenxuan Zuo, Xiaodi Huang, and Shanfeng Zhu. 2019. Grantextractor: Accurate grant support information extraction from biomedical fulltext based on bi-lstm-crf. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(1):205–215.

C Lee Giles and Isaac G Councill. 2004. Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences*, 101(51):17599–17604.

M Hvistendahl and MY Wang. 2014. China's publication bazaar (november, pg 1035, 2013). *Science*, 343(6167):137–137.

Brian A Jacob and Lars Lefgren. 2011. The impact of research grant funding on scientific productivity. *Journal of public economics*, 95(9-10):1168–1177.

Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678.

Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 655–665, Baltimore, Maryland. Association for Computational Linguistics.

Jacalyn Kelly, Tara Sadeghieh, and Khosrow Adeli. 2014. Peer review in scientific publications: benefits, critiques, & a survival guide. *Ejifcc*, 25(3):227.

Omer Levy and Yoav Goldberg. 2014. Neural word embedding as implicit matrix factorization. *Advances in neural information processing systems*, 27.

I. Dan Melamed. 1996. A geometric approach to mapping bitext correspondence. In *Conference on Empirical Methods in Natural Language Processing*.

Julian Michael, Ari Holtzman, Alicia Parrish, Aaron Mueller, Alex Wang, Angelica Chen, Divyam Madaan, Nikita Nangia, Richard Yuanzhe Pang, Jason Phang, et al. 2022. What do nlp researchers believe? results of the nlp community metasurvey. *arXiv preprint arXiv:2208.12852*.

Saif M. Mohammad. 2020a. Examining citations of natural language processing literature. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5199–5209, Online. Association for Computational Linguistics.

Saif M. Mohammad. 2020b. Gender gap in natural language processing research: Disparities in authorship and citations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7860–7870, Online. Association for Computational Linguistics.

Saif M. Mohammad. 2020c. NLP scholar: A dataset for examining the state of NLP research. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 868–877, Marseille, France. European Language Resources Association.

Saif M. Mohammad. 2020d. NLP scholar: An interactive visual explorer for natural language processing literature. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 232–255, Online. Association for Computational Linguistics.

Raymond J. Mooney. 1996. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. In *Conference on Empirical Methods in Natural Language Processing*.

Joakim Nivre, Johan Hall, Sandra Kübler, Ryan McDonald, Jens Nilsson, Sebastian Riedel, and Deniz Yuret. 2007. The CoNLL 2007 shared task on dependency parsing. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 915–932, Prague, Czech Republic. Association for Computational Linguistics.

Kemal Oflazer and Gokhan Tur. 1996. Combining hand-crafted rules and unsupervised learning in constraint-based morphological disambiguation. In *Conference on Empirical Methods in Natural Language Processing*.

Adèle Paul-Hus and Nadine Desrochers. 2019. Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015. *Plos one*, 14(12):e0226727.

Laurie Scrivener. 2009. An exploratory analysis of history students dissertation acknowledgments. *The Journal of Academic Librarianship*, 35(3):241–251.

Karan Singla, Zhuohao Chen, David Atkins, and Shrikanth Narayanan. 2020. Towards end-2-end learning for predicting behavior codes from spoken utterances in psychotherapy conversations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3797–3803, Online. Association for Computational Linguistics.

Li Tang, Guangyuan Hu, and Weishu Liu. 2017. Funding acknowledgment analysis: Queries and caveats. *Journal of the Association for Information Science and Technology*, 68(3):790–794.

Richard H Thaler and Cass R Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.