

# Relation of Linguistic Indicators to Civic Engagement in Special Education

Chak Li  
Vanderbilt University  
chak.li@vanderbilt.edu

Scott Crossley  
Vanderbilt University  
scott.crossley@vanderbilt.edu

Meghan Burke  
Vanderbilt University  
meghan.burke@vanderbilt.edu

Zach Rossetti  
Boston University  
zsr@boston.edu

## ABSTRACT

This study explores the use of natural language processing to examine the relationship between linguistic features and reported levels of civic engagement among parents of individuals with intellectual and developmental disabilities. Specifically, linear regression analyses were conducted with multiple sentiment and semantic similarity variables extracted through linguistic analyses to predict parents' open-ended responses on four civic engagement scales: civic activity, electoral activity, political activity, and broad civic engagement. The resulting models were found to be significant for civic activity and electoral activity; specifically, the linguistic features explained 7% and 4% of the variance of the subscale scores, respectively. In particular, adverbs with positive sentiment were negatively correlated with civic activity while greater semantic similarity in terms of consistency within the text were positively correlated to civic activity. Further, parents' responses with greater semantic similarity when compared with words associated to civic engagement was a significant predictor for electoral activity.

## Keywords

Natural language processing, disability, civic engagement, parents

## 1. INTRODUCTION

With the impending reauthorization of the Individuals with Disabilities Education Act (IDEA, the federal special education law), it is important for parent input to inform legislative changes. In particular, it is essential to account for the perspectives and expertise of parents of individuals with intellectual and developmental disabilities (IDD) in disability legislation given that they often care for and advocate for their offspring across the lifespan [1]. However, parent legislative advocacy and civic engagement have been minimal in the past few decades despite parents of individuals with IDD spearheading several foundational disability policies [1]. Further, given the disparities in services and outcomes for children with IDD [5], it is necessary to identify correlates of civic activity.

Civic engagement encompasses the active participation of individuals in public life. Civic engagement activities may include voting, volunteering, protesting, and advocating. Civic engagement is often characterized in three areas: civic, electoral, and political [4, 10]. Individuals with IDD and their families, an often-marginalized group, have been found to experience improved outcomes on their advocacy, empowerment, and service access through civic

engagement [13]. Nevertheless, numerous barriers (e.g., insufficient: knowledge, resources, time, support, as well as discrimination, stigma, and hostility) pose challenges to civic engagement for parents of individuals with IDD [8, 12]. Understanding the linguistic features of responses made by parents of individuals with IDD presents an opportunity to gain insights into the language that parents produce about civic engagement. By understanding the language of parents about civic engagement, research can then use models to identify the likelihood of parents participating in civic engagement.

The present study aims to extend qualitative research about the civic engagement perspectives of parents of individuals with IDD by quantitatively investigating the connection between linguistic features of responses and self-reported measures on civic engagement. Specifically, we examined the open-ended responses regarding the qualities needed and the reasons for conducting civic engagement on a several linguistic features related to sentiment and semantic similarity derived from natural language processing (NLP) tools. The main goal of this study is to examine the extent to which the linguistic features from the parents' responses are predictive of their self-reported levels of civic engagement.

## 1.1 Language and Civic Engagement

Part-of-speech (POS) tagging involves the assignment of grammatical categories to each token (e.g., nouns, verbs, punctuations). This process allows for the identification and categorization of tokens that could be relevant for further analyses such as sentiment analysis. Sentiment analysis entails the identification and extraction of opinions, emotions, and attitudes expressed within a given text, facilitating an understanding of the emotional expressions of the writer. By extracting tokens which are more likely to convey a text's sentiment (e.g., adjectives, adverbs, and verbs), appropriate analysis can then be conducted to determine their sentiment associations (e.g., positive, negative) and be used for evaluation among other factors. Sentiment analysis via lexicon-based methods have been shown to be a useful tool for measuring and understanding public opinion relating to political discourse. For example, González-Bailón and Paltoglou [6] found that sentiments expressed from online posts were consistent with survey polls on the same topic.

Sentiment analysis may also be appropriate to consider civic engagement among parents of individuals with IDD. Parents of individuals with IDD often report relying on their lived experiences to promote systemic change in the disability community. When providing testimony to Congress, for example, about disability policy parents of individuals with IDD invoke emotional responses (which holds sentiment values) to legislators and other political actors [11]. Therefore, the examination of sentiment analysis on POS tags may assist with understanding the responses of parents of individuals with IDD towards civic engagement.

C. Li, S. Crossley, M. Burke, and Z. Rossetti. Relation of linguistic indicators to civic engagement in special education. In B. Paaßen and C. D. Epp, editors, *Proceedings of the 17th International Conference on Educational Data Mining*, pages 727–731, Atlanta, Georgia, USA, July 2024. International Educational Data Mining Society.

© 2024 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.  
<https://doi.org/10.5281/zenodo.12729930>

Semantic similarity, or how close the meanings of multiple words, phrases, or texts are, can be measured by indicators such as word embeddings or vector representations of the text and then attaining the average of the extracted scores via comparisons by word, sentence, paragraph, or document level. Analyzing the semantic similarity of words within text could provide information regarding its linguistic consistency. When considering parents' responses about civic engagement, linguistic consistency may be important for adequately conveying a core idea for systemic change especially in the context of a civic engagement [1]. Further, it may be important to determine the semantic similarity of responses to the overarching themes that encompasses the construct of civic engagement in the context of special education. By comparing the semantic similarity of parents' responses to civic engagement words (CEW; words that are thematically related to legislative advocacy and civic engagement) [6], we can explore the relations between perspectives and self-reported levels of civic engagement.

## 1.2 Current Study

Although previous research has provided various thematic understanding regarding perspectives on civic engagement from parents of individuals with IDD, none of the studies have investigated the correlation between parents' perspectives with their self-reported levels of civic engagement. For this study, we utilized an innovative approach by examining the linguistic features of open-ended responses of parents of individuals with IDD who demonstrated interest in participating in a training program for improving civic engagement and legislative advocacy. To obtain the relevant variables for analysis, we extracted several linguistic features from the parents' open-ended responses on the qualities needed and reasons for conducting civic engagement using two NLP methods. Specifically, the two methods involved parsing linguistic features including POS tags, sentiment, and semantic similarity. We sought to determine the relation between linguistic features of the responses and the desired outcome of actual civic engagement as measured by an established, self-report scale. The purpose of this study is to examine the association between linguistic features (sentiments and semantic similarity) from the open-ended responses by parents of individuals with IDD and their levels of self-reported civic engagement. We addressed the following research questions among parents of individuals with IDD: Are linguistic factors related to sentiment and semantic similarity significant predictors of civic activity, electoral activity, political activity, and broad civic engagement? Based on prior literature [1, 5], we hypothesized that parents' responses with greater indication of sentiment words and semantic similarity would predict higher civic engagement levels.

## 2. METHOD

### 2.1 Project Civic Leads

The corpus for this study comprised the Civic Leads Parent Survey dataset. Data were collected as part of a multi-state project to investigate the impact of a legislative advocacy program among 168 parents of individuals with disabilities. To be included in the study, participants must: have a child with a disability, being willing to attend a six-hour civic engagement program and be over the age of 18. The participants reflected six sites across the United States: Illinois (18.5%,  $n = 31$ ), Louisiana (10.7%,  $n = 18$ ), Maine (15.5%,  $n = 26$ ), New Mexico (19.0%,  $n = 32$ ), South Carolina (22.6%,  $n = 38$ ), and Washington D.C. (13.1%,  $n = 22$ ). Among the participants, 44.7% reflected racial minority backgrounds. On average, participants were 43.9 years of age ( $SD = 9.25$ ). For each site, we partnered with the Parent Training and Information Center (PTI), a federally-funded agency designed to educate and empower parents

of individuals with disabilities about their special education rights. Interested individuals were screened to ensure they met the inclusionary criteria. Eligible participants then completed a pre-survey asking questions about their demographic background, special education knowledge, civic engagement, and advocacy activities. For the civic engagement program, the six-hour training covered the historical context of special education, disability related policies, and ways to talk to legislators (for more information, see [2]). For this study, only the baseline data were analyzed. Approval from the University Institutional Review Board was obtained.

### 2.2 Open Ended Responses

The corpus for this study included text responses on two open-ended questions: (1) "What personal qualities do you need to conduct civic engagement?" and (2) "Why do you want to conduct civic engagement?" by participants prior to attending the training program. The researchers collapsed the responses from the two open-ended question for all of the following analyses. A total of 168 responses were collected, each approximately 40 words long.

### 2.3 Broad Civic Engagement Scale

The Broad Civic Engagement Scale [9] consists of 19 questions, with three subscales: civic (e.g., volunteering for a civic organization), electoral (e.g., being registered to vote), and political activities (e.g., taking part in a protest). A sample item was "Have you volunteered or done any voluntary community service for no pay?" Response options included: (a) *Yes, I have in the last 12 months*, (b) *Yes, once a month or more*, and (c) *Not within the last 12 months*. In prior studies, the reliability was high (ranging from .64 to 0.76; Sessa et al., 2013). In this study, all subscales and the full scale reported acceptable reliability; civic ( $\alpha = .71$ ), electoral ( $\alpha = .79$ ), political activities ( $\alpha = .83$ ), and the full scale ( $\alpha = .90$ ).

### 2.4 Linguistic Analysis

Two NLP methods were used to assess the linguistic features in the open-ended responses: sentiment analysis and semantic similarity. The exact code used can be found here: <https://t.ly/NXdNj>.

#### 2.4.1 Sentiment Analysis

The Bing Sentiment Lexicon [7] is a general-purpose English sentiment lexicon of binary categorization (i.e., positive or negative). The lexicon was created by mining and summarizing customer reviews of various products and contains a combination of 2006 positive and 4783 negative words. For this study, it was used for sentiment analysis with the aforementioned corpus via spaCy. Tokens within each text response that were tagged as adjectives, adverbs, or verbs were compared with the words from the positive and negative word lists, with incidences tabulated to their respective POS type sentiment count if true. The raw counts were normed by the number of words from the respective responses.

#### 2.4.2 Semantic Similarity

The spaCy English Core Web Large model was used to calculate semantic similarity among participant responses in two ways. First, content words (i.e., function words and punctuations were excluded) were extracted and counted for each response. For word similarity score (WSS), the *word2vec* algorithm via spaCy was utilized to determine semantic similarity scores across all extracted content word within each response. The average WWS was then calculated by dividing the sum of all semantic similarity scores across content words within a response with the total number of semantic similarity scores for that respective response. Incidences per averaged WSS were tabulated by response to the data frame.

**Table 1. Descriptive Statistics and Correlation of Variables**

Variable	1	2	3	4	5	6	7	8	9	10	11	$\bar{x}$	<i>SD</i>
1. Civic Activity	-											9.512	2.350
2. Electoral Activity	0.479	-										9.137	2.606
3. Political Activity	0.541	0.522	-									17.042	4.631
4. ADJ_POS	0.092	-0.078	0.026	-								.029	.050
5. ADV_POS	-0.204	-0.121	-0.135	-0.094	-							.004	.015
6. VERB_POS	-0.053	-0.026	-0.045	-0.136	-0.029	-						.016	.026
7. ADJ_NEG	-0.163	-0.074	-0.035	-0.037	0.141	-0.069	-					.005	.016
8. ADV_NEG	0.002	-0.079	-0.113	0.024	-0.029	-0.064	0.251	-				.001	.006
9. VERB_NEG	0.021	-0.085	-0.008	-0.005	-0.035	-0.062	0.013	-0.022	-			.002	.006
10. WSS	0.196	0.203	0.032	0.124	-0.208	-0.157	-0.078	-0.033	0.008	-		.303	.095
11. CEWS	0.177	0.098	0.036	0.111	0.049	-0.103	-0.103	0.000	-0.172	0.037	-	.083	.061

To calculate the semantic similarity score between CEW and the content words of each participant response in the corpus, a list of words (i.e., CEW) were derived based on findings from past research on civic engagement for parents of individuals with disabilities [5]. The CEW list was then imported and used for comparison with the content words extracted from each response via spaCy's semantic similarity process on the document level to obtain semantic similarity scores to CEW. The semantic similarity scores were then normed by dividing it with the respective number of content words in the response to calculate the final CEW semantic similarity scores (CEWS) and tabulated to the data frame.

## 2.5 Statistical Analysis

Preliminary analyses using descriptive statistics (e.g., means, standard deviations, standard errors) were conducted to characterize the data. Correlational analysis was then conducted among the independent variables: ADJ\_POS, ADV\_POS, VERB\_POS, ADJ\_NEG, ADV\_NEG, VERB\_NEG, WSS, and CEWS. Two methods were used to check for multicollinearity between the variables: correlations above .65 and Variance Inflation Factor (VIF) above 2.5 [10]. Since the correlations among the independent variables were all well below the recommended cutoff guidelines (i.e.,  $r$ 's = -.21~.25), multicollinearity was not a concern (see Table 1). To understand the contribution of different types of linguistic features (i.e., sentiment and semantic similarity) to civic engagement, regression analyses were conducted on each dependent variable (i.e., civic activity, electoral activity, political activity, and broad civic engagement). All statistical analyses were conducted using *R*.

## 3. RESULTS

### 3.1 Linguistic Factors on Civic Activity

The linear model regressing the eight linguistic factors onto civic activity revealed significant effects for adverbs with positive sentiment and CEW similarity scores. The overall model was significant,  $F(8, 159) = 2.584, p < .01, R^2 = .071$ , and the residuals were normally distributed. The linguistic variables explained approximately 7% of the variance of the civic activity subscale scores and indicated that parents' responses with fewer adverbs associated with positive sentiment and greater semantic similarity to the CEW list reported participating in more civic activities. See Table 2 for information on coefficients, standard error,  $t$  values, and  $p$  values for each of the included variable used for the analysis.

**Table 2. Regression for Linguistic Factors on Civic Activity**

Variable	Coefficient	Std. Error	$t$
ADJ_POS	1.55	3.57	0.43
ADV_POS	-25.26	12.23	-2.07*
VERB_POS	-1.53	6.86	-0.22
ADJ_NEG	-18.18	11.88	-1.53
ADV_NEG	12.32	30.02	0.41
VERB_NEG	16.44	27.87	0.59
WSS	3.47	1.92	1.81
CEWS	6.55	2.98	2.20*

Note: \*  $p < .05$

### 3.2 Linguistic Factors on Electoral Activity

The linear model regressing the eight linguistic factors onto electoral activity revealed significant effects for word similarity scores. The overall model was significant,  $F(8, 159) = 1.798, p < .01, R^2 = .037$ , and the residuals were normally distributed. The linguistic variables explained approximately 4% of the variance of the electoral activity subscale scores and indicated that parents' responses with greater semantic similarity across content words reported participating in more electoral type activities. See Table 3 for information on coefficients, standard error,  $t$  values, and  $p$  values for each of the included variable used for the analysis.

**Table 3. Regression for Linguistic Factors on Electoral Activity**

Variable	Coefficient	Std. Error	$t$
ADJ_POS	-6.38	4.03	-1.58
ADV_POS	-17.49	13.80	-1.27
VERB_POS	-1.73	7.75	-0.22
ADJ_NEG	-3.91	13.41	-0.29
ADV_NEG	-29.82	33.89	-0.88
VERB_NEG	-31.27	31.46	-0.99
WSS	5.13	2.17	2.37**
CEWS	3.93	3.36	1.17

Note: \*\*  $p < .01$

### 3.3 Linguistic Factors on Political Activity

The linear model regressing the eight linguistic factors onto political activity did not reveal any significant effects. The overall model was not significant,  $F(8, 159) = 0.769$ ,  $p = .631$ ,  $R^2 = 0$ . See Table 4 for information on coefficients, standard error,  $t$  values, and  $p$  values for each of the included variable used for the analysis.

**Table 4. Regression for Linguistic Factors on Political Activity**

Variable	Coefficient	Std. Error	$t$
ADJ_POS	0.53	7.35	0.07
ADV_POS	-45.60	25.13	-1.82
VERB_POS	-9.48	14.11	-0.67
ADJ_NEG	5.02	24.42	0.21
ADV_NEG	-96.41	61.71	-1.56
VERB_NEG	-9.38	57.29	-0.16
WSS	-0.55	3.94	-0.14
CEWS	2.80	6.13	0.46

### 3.4 Linguistic Factors on Civic Engagement

Although the linear model regressing the eight linguistic factors onto broad civic engagement revealed significant effects for adverbs with positive sentiment, the overall model was not significant,  $F(8, 159) = 1.418$ ,  $p = .193$ ,  $R^2 = .020$ . See Table 5 for information on coefficients, standard error,  $t$  values, and  $p$  values for each of the included variable used for the analysis.

**Table 5. Regression for Linguistic Factors on Civic Engagement**

Variable	Coefficient	Std. Error	$t$
ADJ_POS	-4.30	12.50	-0.34
ADV_POS	-88.35	42.74	-2.07*
VERB_POS	-12.74	24.00	-0.53
ADJ_NEG	-17.07	41.54	-0.41
ADV_NEG	-113.91	104.97	-1.09
VERB_NEG	-24.21	97.45	-0.25
WSS	8.05	6.71	1.20
CEWS	13.28	10.42	1.28

Note: \*  $p < .05$

## 4. DISCUSSION AND CONCLUSION

In this study, we took a novel approach in examining the relation between linguistic factors and civic engagement by using NLP methods to extract linguistic features from parents' responses about civic engagement and using inferential statistical analyses to then determine their impact on the levels of civic engagement as measured by the Broad Civic Engagement Scale [9]. Our findings indicated that linguistic features related to sentiment and semantic similarity accounted for 4~7% of the variance in two of the four civic engagement scores and contribute to the extant literature.

First, the findings demonstrate that the sentiment associations within open-ended responses provided by parents of children with IDD may not have a clear relationship direction with civic engagement scores as hypothesized. In particular, parents' responses with *less* adverbs that were found in the positive sentiment word list was

a significant predictor for higher levels of civic activities. Despite this contrasting finding to our hypothesis where greater sentiment associations in either direction (i.e., positive or negative) reflects similarly more emotional investment from the individual and subsequent indicator for civic engagement, the negative directionality of this relationship still has important implications to the field. Specifically, this may suggest that parents who expressed fewer positive modifiers to verbs (e.g., actions) in their responses are more realistic, critical, and/or dissatisfied with their offspring's current situation, and therefore are more motivated to engage in civic activities that can improve their conditions or rights [3]. However, it is important to note that this was the only predictor among the sentiment variables found significant across the four regression analyses conducted, therefore the practical significance of this finding is questionable especially given that it did not appear to be significant for any of the other regressions.

Second, the results from the regression analyses indicate that the two types of semantic similarity that were calculated within parents' responses were significant predictors for civic activity and electoral activity respectively, confirming our hypotheses. In terms of civic activity, these findings reveal that parents who have higher scores tend to produce responses that have more similarity with CEWs, which corroborates with previous studies [6] noting the importance of these thematic ideas to civic engagement. For electoral activity, parents' responses that had greater word similarity (which may imply they are more coherent, consistent, and clear in expressing their ideas) predicting higher scores for this subscale may indicate that these parents have a more unified vision or perspective on civic issues or actions. This is consistent with the underlying framework and activity (i.e., developing a clear political testimonial) which serves the core focus of the civic engagement training such as the one associated with this study [1]. The combination of these findings has direct implications for civic engagement training. Specifically, it is important for researchers and practitioners who are developing programs to increase the civic engagement of parents' of individuals with IDD to consider the integration of activities which sought to enhance the consistency in expressing ideas relating to civic engagement as well as repertoire of content specific lexicon (i.e., civic engagement words).

While this study provides an important launching point for using NLP approaches to understand the relationship between qualitative and quantitative data within the context of civic engagement by parents of individuals with IDD, some limitations should be noted. First, the absence of significant linguistic predictors towards political activity and broad civic engagement may indicate that the questions for the open-ended responses may not have adequately captured or represented the full spectrum of civic engagement attitudes and behaviors of parents of individuals with IDD as intended. Second, the low percentage of variance that were explained from the selected variables for the present regression analyses would suggest that there are other factors (both linguistically and non-linguistically related) that contributes to the participants' broad civic engagement scores. For future research, analyses which incorporates other linguistic features that were unexplored in the current study and/or more nuanced lexicon lists that were trained to be used specifically for the topic of civic engagement and legislative advocacy could be employed to better determine the relationship between perspectives on civic engagement and the reported civic engagement scores of parents of individuals with IDD.

## 5. ACKNOWLEDGMENTS

This research was supported by the Spencer Foundation (PI: Rossetti). The authors are grateful to the families who participated.

## 6. REFERENCES

- [1] Turnbull, H. R., Shogren, K. A., & Turnbull, A. P. (2011). Evolution of the parent movement: Past, present, and future. In J.M. Kauffman & D.P. Hallahan (Eds.), *Handbook of special education* (pp. 639–653). New York, NY: Routledge.
- [2] Burke, M. M., Rossetti, Z., Rios, K., Schraml-Block, K., Lee, J. D., Aleman-Tovar, J., & Rivera, J. (2020). Legislative Advocacy Among Parents of Children With Disabilities. *The Journal of Special Education*, 54(3), 169-179. <https://doi.org/10.1177/0022466920902764>
- [3] Harry, B., & Ocasio-Stoutenburg, L. (2021). Parent advocacy for lives that matter. *Research and practice for persons with severe disabilities*, 46(3), 184-198. <https://doi.org/10.1177/15407969211036442>
- [4] Shattuck, P. T., Narendorf, S. C., Cooper, B., Sterzing, P. R., Wagner, M., & Taylor, J. L. (2012). Postsecondary education and employment among youth with an autism spectrum disorder. *Pediatrics*, 129(6), 1042-1049. <https://doi.org/10.1542/peds.2011-2864>
- [5] Goldman, S. E., Burke, M. M., & Mello, M. P. (2019). The perceptions and goals of special education advocacy trainees. *Journal of Developmental and Physical Disabilities*, 31, 377-397. <https://doi.org/10.1007/s10882-018-9649-2>
- [6] González-Bailón, S., & Paltoglou, G. (2015). Signals of public opinion in online communication: A comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 95-107. <https://doi.org/10.1177/0002716215569192>
- [7] Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). <https://doi.org/10.1145/1014052.1014073>
- [8] Leiter, V., & Wyngaarden Krauss, M. (2004). Claims, barriers, and satisfaction: Parents' requests for additional special education services. *Journal of Disability Policy Studies*, 15(3), 135-146. <https://doi.org/10.1177/10442073040150030201>
- [9] Lopez, M. H., Levine, P., Both, D., Kiesa, A., Kirby, E., Marcelo, K., & Williams, D. (2006). *The 2006 civic and political health of the nation*. College Park, MD: The Center for Information and Research on Civic Learning and Engagement.
- [10] Tabachnik, B. G., & Fidell, S. L. (2007). Discriminant analysis. *Using multivariate statistics*, 201(3), 377-438.
- [11] Pearson, J. N., Stewart-Ginsburg, J. H., Malone, K., Manns, L., Mason Martin, D., & Sturdivant, D. (2023). Best FACES forward: Outcomes of an advocacy intervention for black parents raising autistic youth. *Exceptionality*, 31(2), 135-148. <https://doi.org/10.1080/09362835.2022.2100392>
- [12] Trainor, A. A. (2010). Diverse approaches to parent advocacy during special education home—school interactions: Identification and use of cultural and social capital. *Remedial and Special education*, 31(1), 34-47. <https://doi.org/10.1177/0741932508324401>
- [13] Turnbull III, H. R. (2005). Testimony to the Senate Committee on Health, Education, Labor, and Pensions, April 6, 2005: Health Care Provided to Nonambulatory Persons. *Research and Practice for Persons with Severe Disabilities*, 30(1), 38-41. <https://doi.org/10.2511/rpsd.30.1.38>