Distribution Agreement

Qinxin Yuan April 6, 2023

Hostile and Aggressive Semantic Themes in Children's Speech:

Discovering Linguistic Indicators for Aggression

by

Qinxin Yuan

Irwin Waldman
Adviser

Psychology

Irwin Waldman

Adviser

Phillip Wolff

Committee Member

Jinho Choi

Committee Member

2023

Hostile and Aggressive Semantic Themes in Children's Speech:

Discovering Linguistic Indicators for Aggression

By

Qinxin Yuan

Irwin Waldman

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Psychology

2023

Abstract

Hostile and Aggressive Semantic Themes in Children's Speech:

Discovering Linguistic Indicators for Aggression

By Qinxin Yuan

Specific and targeted indicators and predictors of psychopathology help identify individuals who are at risk for developing mental health problems. The present study investigated whether aggressive children have special linguistic indicators in their speech, reflecting their deficits in social perception. Data were collected on children's answers to a specific set of open-ended questions after watching video clips of one child ruining another child's play materials with varying intent. We transcribed children's recorded speech into text, and used a natural language processing tool to extract instances of noun, verb and adjective use. Then, we manually curated and used ChatGPT to generate aggression-related words and prosocial words. We predicted and found that children who are high in aggression present more semantic themes related to hostility, aggression, and threat - but not prosocial themes - than children who are low in aggression. The results for aggression-related themes held for manually curated but not ChatGPT generated themes. The results should encourage semantic analyses for understanding the causes and mechanisms underlying aggression and for the diagnosis, prevention and treatment of externalizing psychopathology. The results also raise questions about limitations of applying ChatGPT in psychology research.

Hostile and Aggressive Semantic Themes in Children's Speech:

Discovering Linguistic Indicators for Aggression

By

Qinxin Yuan

Irwin Waldman

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Psychology

2023

Acknowledgements

I would like to express my sincere gratitude to my thesis advisor, Dr. Irwin Waldman, for his guidance, support, and patience throughout my undergraduate studies. His mentorship and encouragement have been invaluable to me.

My thanks to the members of my thesis committee, Dr. Phillip Wolff and Dr. Jinho Choi, for their insightful feedback and constructive criticism.

I would like to extend my appreciation to the graduate students from Waldman Lab and Concept Mining Lab, Alireza Ehteshami and Linying Li, who provided me with technical assistance and insightful discussions.

Additionally, I want to express my appreciation to my family and friends for their encouragement and support throughout my academic journey.

Finally, thanks to myself for taking this commitment.

Table of Contents

Introduction

Mental illness impacts thousands of people around the world and their family, workplace and the society. An important dimension of psychopathology is Externalizing, which includes early-onset disorders such as Attention Deficit Hyperactivity Disorder, Oppositional Defiant Disorder, Conduct Disorder, and Antisocial Personality Disorder, and traits such as aggression and antisocial behavior, that continue to affect patients in adolescence and predict psychopathology in adulthood. Extracting indicators and predictors for specific disorders is the first step to aid in clinical diagnosis and could promote early prevention and treatment. Previous studies have found solid linguistic biomarkers for psychosis, internalizing psychopathology (e.g., depression), and neurodevelopmental disorders (e.g., Autism Spectrum Disorder) through health records, social media posts, and clinical interviews (e.g., Rezaii et al., 2019). However, few studies focus on externalizing behaviors in middle childhood. Thus, this study investigated whether children who exhibit aggression present more semantic themes related to hostility, aggression and threat than non-aggressive children, which might reflect their social perceptual deficits.

One major form of externalizing behaviors is aggression. Aggression refers to physical, verbal, relational and passive behaviors that cause harm. Two important types of aggression include Reactive aggression, which is an impulsive response to a perceived threat or provocation, often associated with high emotional arousal, anxiety, and anger, and Proactive aggression, which is more instrumental and organized, and might be motivated by the anticipation of reward (Raine et al, 2006). Dodge & Coie found that reactive and proactive aggression are highly reliable and show construct validity in their associations with other variables (Dodge & Coie, 1987).

Several attributional studies have shown that a child's tendency to interpret a provocation from a peer significantly predicts the likelihood of the child engaging in aggressive retaliation against the peer (Berkowitz, 1977), and aggression is very likely when the provocation is intentional rather than accidental (Rule & Duker, 1973). Specifically, attributional biases and deficits were found to be positively correlated with levels of reactive aggression, but not proactive aggression (Dodge & Coie, 1987). Although reactive-aggressive boys were viewed as less aggressive and bothersome and not intrusive compared to proactive-aggressive boys, they displayed more deficits in accurately interpreting their peers' benign intentions. Studies also showed that aggressive boys were much more likely than their average status peers to respond with aggressive behaviors to others' behavior perceived as nonhostile in intent (Waldman, 1996). These aggressive children show more hostile social perception biases while interpreting and reacting to intentions. Therefore, using different social perception cues with hostile/non-hostile intents, we could stimulate various social perceptions in children. Analyzing children's verbal communication and their variations in speech, we might catch indicative semantic features.

Besides the contextual intention, language in general is a media to externalize emotional or cognitive deficiencies. In a meta-analysis comprising 47 articles (63 153 participants), it was revealed that there is a difference in ratings of problem behaviors between children with language disorders and typically developing children, and the difference in problem behavior ratings intensifies as the child ages (Curtis et al., 2018). Many researchers also provided support for an early link between language and aggression due to executive function deficits. In a longitudinal approach, Silva et al. (1987) examined the long-term impact of language difficulties. The results revealed that those children who experienced delays in language development and subsequently exhibited slow reading skills, not only demonstrated these deficits at each evaluation at age 3, 7, 9, and 11, but also had notably higher scores on behavior problems at home and in the classroom when compared to the control group. Studies also showed that in preschool (39-75 months-old) children, there was a significant and negative relationship between children's receptive and expressive language skills and physical and relational aggression levels (Ersan, 2020).

In addition to focusing on detecting the linguistic characteristics of aggression in young children, this study also incorporates Natural Language Processing tools to extract nouns, verbs, and adjectives. Natural Language Processing (NLP), a subfield of linguistics and artificial intelligence, processes and analyzes natural human languages with the use of computer languages. Prior to this new system, language-processing systems were designed by symbolic methods, and researchers had to hand-code a set of rules prevalent in patients' semantics and sentence patterns to inspect new cases. Currently, scientists develop NLP methods based on Machine Learning (ML) algorithms to determine important items in speech and analyze sentence structures. This study used Stanza, a Python natural language analysis package containing a collection of efficient tools in language analysis (Qi et al., 2020). Lemmatization is the process of converting a word to its base or dictionary form, and the lemmatization module in Stanza normalizes the lemma form for each input word. The Part-of-Speech (POS) &

morphological features tagging module labels the word by nouns, verbs and adjectives, etc. These functions allow us to output instances of different word use in children's speech.

Another innovative approach for this study is that it also uses ChatGPT to categorize semantic themes. The unique approach of comparing manual curation and ChatGPT serves the purpose of demonstrating practical application of accessible tools in psychological research. ChatGPT is a model trained by OpenAI to follow an instruction in a prompt and provide a detailed response is gaining popularity in research field. The interactive conversation format makes it possible for ChatGPT to answer versatile questions related to natural languages, specifically, to group potential semantic themes. Although its specific use in children's aggression detection is unclear, there has been research on ChatGPT's detection of implicitly hateful tweets (Huang et al., 2023) showing an 80% agreement with human. The result demonstrates the potential of ChatGPT as a data annotation tool using a simple prompt design. As a result, we incorporated ChatGPT generated semantic themes to further examine the potential.

Therefore, our primary hypothesis is that aggressive/threatening themes in speech are correlated with children's aggression. Specifically, we predicted that (1) children displaying more aggressive behavior would employ more aggressive/threatening words in their verbal responses to questions about videotape scenarios of peer interactions. We also predicted that (2) children high in reactive aggression would produce more aggressive themes than children high in proactive aggression. At the same time, we predicted that (3) children's levels of aggression would be unrelated to their generation of prosocial and accidental themes. Finally, we predicted that (4) the associations of children's aggression levels with aggressive/threatening and prosocial themes generated by ChatGPT will be similar to that of manually curated group.

Methods

Participants

Participants were a subsample (*N* = 85 twin pairs) selected from the Georgia Twin Registry (*N* = 885 twin pairs; Ficks & Waldman, 2014). Researchers identified participants by obtaining birth records for all twins born in Georgia between 1980 and 1991. This is a representative sample of twins born in Georgia between 1980 and 1991 (M_{age} = 8.5, *SD* = 2.9, range 4 to 19, 51% female). According to parental report, the sample was 82% Caucasian, 11% African American, 1% Hispanic, and 6% of mixed ethnicity. Information on participants' psychopathology and personality were collected.

Families of twins were asked to join the Georgia Twin Registry (Ficks & Waldman, 2014; Singh & Waldman, 2010) and those who agreed to participate were mailed a set of questionnaires. Parent report shows that mothers completed 53% of the questionnaires alone, and the remaining questionnaires were completed either by fathers (1%) or both mothers and fathers jointly (46%) (Ficks & Waldman, 2014; Singh & Waldman, 2010). Parents were provided written informed consent to the study protocol. This study is approved by the institutional review board at Emory University (IRB number 436-95).

Materials and Measures

Parent-Ratings of Aggression

The questionnaire includes 12 statements that describe various types of aggressive behavior, such as proactive aggression, reactive aggression and non-specific aggression (Appendix 1). These statements were based on extensive observations of aggression in boys' peer groups (Dodge & Coie, 1987). Each statement was rated on a 0- to 4-point Likert scale, ranging from *never* to *almost always*, indicating how descriptive each statement was of the parents' child. For instance, children with the highest level of proactive aggression would receive 4 points on the statement "this child uses physical force in order to dominate other kids". Children with highest level of reactive aggression receive 4 points on the statement "when this child has been teased or threatened, he or she gets angry easily and strikes back".

Intention-Cue Detection Measure

This videotape measure examines children's detection and discrimination among various intentions in dyadic social interactions. Each video (lasting 30 seconds) shows one child entering a room and ruining another child's toy with either hostile, accidental, prosocial, ambiguous or merely present intentions. One child ruins the play material with varying intent. In the current study, the latter four intensions, accidental, prosocial, ambiguous or merely present, are intended to be interpreted as non-hostile.

The measure contains 2 tasks, a 10-item identification task and a 14-item discrimination task. In the identification task, the participants were asked to identify the actions and intentions of the child wrecking the other child’s toy after each of the 10 trials. Specifically, the questions asked of participants were “What did the child with the number on his/her t-shirt do?”, “Why did he/she do it?”. On 5 of the randomly selected trials, participants were asked about their own hypothetical behavioral responses to the scenario (“What would you do next if you were the child without the number on his/her t-shirt?”, “What else would you do?). In the discrimination task, participants were told to discriminate one of the three situations that displayed a different intention from other two. We only used responses to the Identification task in this study.

Procedure

Data were collected from a subset of twins from the Georgia Twin Study who came into the Waldman lab at Emory University for assessment on a variety of lab measures, including the Intention Cue Detection measure. Each participant completed 10 trials of the identification task in order and 14 trials of the discrimination task. Both interviewers’ questions and participants’ answers were recorded.

The 12-item Parent-Rating of Aggression Instrument was administered to parents of all 170 children (85 twin pairs) who took the video task. Total Aggression scale scores were calculated as the sums of the 12 aggression-related items, whereas Proactive and Reactive aggression scale scores were the sum of the 3 terms on each of those scales.

Speech Analysis

Speech to Text Transcription

Audio recordings of intension-cue interviews for all *N=170* participants' identification tasks were transcribed employing the transcription pipeline from Microsoft Azure Cognitive Services. The children's recorded speech was reported as a word-to-word script. The script was independently and manually reviewed by a research assistant to ensure the correctness. After manual revision, *N=102* participants' speech was clear and recognizable and thus usable in further analyses. The speech of participants was then separated from speech of interviewers.

Lemmatization NLP pipeline

Using the pretrained Stanza package in Python, each participant's use of nouns, verbs and adjectives were extracted. The word used, instance of specific words, and the feature of each word (i.e., whether it is a noun, verb or adjective) were reported on a spreadsheet. Fillers, prepositions, pronouns, or conjunctions were excluded.

Theme Classification

Based on the result of lemmatization, vocabularies used by participants were manually curated into aggressive/threatening themes, or prosocial and accidental themes. Words under aggressive/threatening theme include verbs describing hostile behaviors (e.g., hurt, smash, hit, kick) and derogatory adjectives (e.g., mean, stupid, selfish). The control group, or words under prosocial theme, contains commendatory adjectives (e.g., honest, polite, allowable) and prosocial actions (e.g., compromise, accept, tidy). The prototypical accidental phrases, such as "it was an accident," are combined with prosocial themes in this study because the child interprets intensions of action properly. Thus, words related to accidents were mixed with prosocial themes under the category prosocial and accidental themes. Words not included in either of the groups were treated as irrelevant and were thus ignored.

To compare manually curated groups with a “big data” approach, the word list was entered into ChatGPT for detecting aggressive/threatening or prosocial and accidental themes. ChatGPT was not able to find accidental themes. Thus, 4 groups of words ---- manually curated aggressive theme, aggressive theme ChatGPT, manually curated prosocial-accidental theme, and prosocial theme ChatGPT ---- were used in statistical analysis. Specific Aggressive/Threatening and Prosocial/Accidental words generated manually and by ChatGPT are listed in Table 2.

Statistical Analysis

Variable Computation

Using SPSS, we computed and analyzed four predictor variables and three outcome variables. Four predictor variables, Aggressive/Threatening Themes, Aggressive/Threatening Themes from ChatGPT, Prosocial/Accidental Themes and Prosocial Themes ChatGPT were computed. The Aggressive/Threatening Themes variable is the number of manually selected words in each child's speech considered as aggressive, threatening, and hostile. Aggressive/Threatening Themes from ChatGPT is the number of words in a child's speech identified as aggressive, threatening, and hostile by ChatGPT. The Prosocial/Accidental Themes variable refers to the number of prosocial or accidental words in children's speech. Since ChatGPT was not able to output any words related to accidental themes, the fourth variable, Prosocial Themes ChatGPT, only contains the number of prosocial words selected by ChatGPT in children's answers.

The three outcome variables in this study were aggtotal, reactiveaggsum and proactiveaggsum. Aggtotal, measuring the overall level of aggression, is the sum of 12 separate items. Reactiveaggsum and proactiveaggsum represent the level of children's reactive and proactive aggression respectively. They both were computed as the sum of 3 of the items contained within the broader variable of aggtotal.

The distribution of seven variables were plotted on bar-graphs, shown in Figure 1.

Correlation Analysis

This research aims to investigate the correlation between semantic themes produced by children and their aggression level. In SPSS, we generated a correlation matrix showing the correlation between all seven variables Aggressive/Threatening Themes, Aggressive/Threatening Themes from ChatGPT, Prosocial/Accidental Themes, Prosocial Themes ChatGPT, aggtotal, reactiveaggsum and proactiveaggsum. Spearman's correlation coefficient ρ, one-tailed significance p, and the number of pairs used to calculate the Spearman Correlation coefficient N was reported.

Results

In this study, our major hypothesis is that aggressive/threatening themes in speech are correlated with children's aggression. Specifically, we hypothesized that children displaying more aggressive behavior would employ more aggressive/threatening words in their verbal responses to questions about videotape scenarios of peer interactions. We also predicted that children high in reactive aggression would produce more aggressive themes than children high in proactive aggression. We also hypothesized that children's levels of aggression would be unrelated to their generation of prosocial and accidental themes. Finally, in a set of exploratory analyses, we conducted the same analyses examining the associations of children's aggression levels with aggressive/threatening and prosocial themes generated by ChatGPT (as opposed to being manually curated).

Descriptive Statistics

There are four predictor variables, Aggressive/Threatening Themes, Aggressive/Threatening Themes from ChatGPT, Prosocial/Accidental Themes and Prosocial Themes ChatGPT. The distributions of the 4 independent variables differed. For the manually selected aggressive words, more than half of children did not produce any aggression related words (Fig.1) and the distribution was right-skewed with a downward asymptote, indicating that aggression is not a common theme. Conversely, ChatGPT generated a more normally distributed set of responses for aggressive themes, with greater variability (Fig.2). Most children produced 3 to 8 aggression-related words, and the distribution is relatively more symmetrical.

On the contrary, instances of prototypical prosocial and accidental words have opposite distributions. For manually generated prosocial and accidental words, the distribution is disperse (Fig.3), with most children producing around 1 to 10 prosocial words, and few children producing none or many prosocial words. Although the graph is right skewed, indicating the presence of highly prosocial speech in few children, the graph is centered at a reasonable instance of prosocial and accidental themes (around 4 to 7 such words in conversation). However, ChatGPT generated a more asymptotic distribution with less variability. According to ChatGPT, most children did not produce any prosocial words (Fig.4). At the 50th percentile, children generated none of the prototypical prosocial word, and the frequency of producing more prosocial words steeply decreases.

The three outcome variables in this study were aggtotal, reactiveaggsum and proactiveaggsum. Aggtotal, measuring the overall level of aggression, consisted of 12 separate items. Reactiveaggsum and proactiveaggsum represent the level of children’s reactive and proactive aggression respectively. They both contain 3 of the items contained within the broader variable of aggtotal.

The distribution of aggtotal is right-skewed and bounded at 0, indicating deviation from a normal distribution (Fig.5). Most children have a low level of total aggression, and the number of children who are relatively more aggressive asymptotically decreases. Proactive and reactive aggression follow the same pattern, while proactive aggression has an even steeper asymptotic, right-skewed trend (Fig.6, Fig.7). These distributions show that aggressive behaviors in a population are rare, and the number of children taking the initiative to hurt others is even rarer than those who are non-aggressive.

Correlations between Semantic Themes and Aggression

Overall, the results are consistent with our first hypothesis: aggressive themes in speech are correlated with children's aggression level (see column 1 in Table 1). More manually classified aggressive themes in speech predict higher aggression levels, for total aggression (aggtotal; $\rho = 0.30$, $p = 0.002$) as well as both proactive and reactive aggression (proactiveaggsum: $\rho = 0.26$, $p = 0.008$) and reactiveaggsum: $\rho = 0.27$, $p = 0.006$), suggesting that aggressive language is positively correlated with both types of aggression. Unlike predicted, children who tend to reactively initiate aggressive behaviors did not use more aggressive and threatening words in their speech than children high on proactive aggression. In fact, the two variables, proactive and reactive are moderately correlated, with a Spearman's correlation of 0.422 ($p < 0.01$).

Reactive aggression level is highly correlated to total aggression level ($\rho = 0.88$, $p < 0.001$), and proactive aggression is moderately correlated with total aggression ($\rho = 0.52$, $p < 0.001$). The non-normal distribution of variables precluded the use of Pearson's correlation. Spearman's correlation, which uses ranks instead of correlation scores, is a more suitable method in this scenario.

In contrast to the results for manually curated Aggressive/Threatening themes, ChatGPT-generated aggressive themes showed only weak and statistically non-significant correlations with children's aggression. The lack of correlation was consistent for total aggression ($\rho = 0.12$, $p = 0.124$), proactive aggression ($\rho = 0.03$, $p = 0.401$), and reactive aggression ($\rho = 0.10$, $p = 0.165$). Despite having a moderate correlation with manually curated Aggressive/Threatening themes ($\rho = 0.423$, $p < 0.001$), aggressive themes selected by ChatGPT were not effective in distinguishing children who were high versus low in aggression.

For the control condition of prototypical prosocial/accidental words, the results were as predicted. Prosocial and accidental themes in children's speech were not correlated with children's total ($\rho = 0.141$, $p = 0.095$), proactive ($\rho = 0.01$, $p = 0.464$) or reactive aggressive behaviors ($\rho = 0.062$, $p = 0.284$) for manually generated themes. Consistently, prosocial and accidental themes in children's speech were not correlated with children's total ($\rho = 0.09$, $p = 0.202$), proactive ($\rho = -0.045$, $p = 0.338$) or reactive aggressive behaviors ($\rho = 0.022$, $p = 0.418$) for ChatGPT generated themes. The lack of correlation was consistently present for both manually curated and ChatGPT generated themes. Furthermore, manually selected accidental and prosocial words have a relatively high correlation with ChatGPT selected prosocial words, a Spearman's correlation of 0.636 ($p < 0.01$).

Another unexpected result is that children's use of aggressive, hostile and threatening words was positively correlated with children's use of prosocial and accidental words. For aggressive/threatening and prosocial/accidental themes generated manually and by ChatGPT, the correlations were 0.364 ($p < 0.01$) and 0.392 ($p < 0.01$) respectively. These correlations were thus consistent across the two methods of theme selection.

Discussion

This study found that children who display aggression present more semantic themes related to hostility, aggression and threat than non-aggressive children. This is consistent with previous studies which reported modest reciprocal associations between physical aggression and language performance in early childhood (Girard et al., 2014). It is also matches the presence of associations between conduct problems and language and executive deficits (Olson and Hoza, 1993). The high correlation between proactive, reactive and total level of aggression is also consistent with Dodge and Coie's study (1987). The result implies that certain verbal responses and externalizing behaviors are positively correlated in children. As children's lexical choices and their tendencies to initiate parallel actions are aligned, we might be able to predict aggression based on instances of hostile themes in children's speech. Thus, our approach provides initial results that may suggest the potential value of identifying aggressive behavioral tendencies through the analysis of linguistic indicators.

The study yielded additional insights regarding the performance of ChatGPT-generated aggressive themes in predicting children's aggression levels. Results showed that these ChatGPT-generated themes exhibited no statistically significant correlation with children's levels of aggression, suggesting that manually curated selection of aggressive themes may outperform ChatGPT in assessing psychopathology. In fact, ChatGPT generated a smaller set of aggressive and prosocial words compared to the manually curated set. This discrepancy may be attributed to the selective nature of ChatGPT's word choice, which tends to favor universally aggressive words, whereas our manual selection takes more contextual factors (e.g., knowledge of the specific task-relevant stimuli features) into account. Compared to psychology professionals' extensive time of research and specific expertise in children aggression research, ChatGPT's ability to detect semantic themes was outshone. Moreover, instances of aggressive words from ChatGPT followed a normal distribution, likely reflecting the algorithm's overgeneralized arithmetic. However, in reality, high levels of aggressiveness are relatively infrequent, approximately 3% to 7% of children and adolescents manifest aggressive signs (Zahrt & Melzer-Lange, 2011), which aligns more closely with the manually selected words. In that sense, ChatGPT should be applied to augment and support the work of psychology professionals, instead of replacing the expertise.

An unexpected result is that children’s use of aggressive, hostile and threatening words was positively correlated with children’s use of prosocial and accidental words. One possible explanation for these results could be variation of verbal abilities among children, such that the number of words produced by children is positively correlated with aggression, indicating that children who generate more words are more likely to use aggressive language. Children with higher verbal abilities or those who are simply more talkative may produce more speech overall, and consequently use a wider range of vocabulary, including both prosocial and aggressive language.

Several limitations must be acknowledged in this study. First, the speech samples were recorded before 2010. With the proliferation of social media and modern technologies, children's speech habits may have since evolved in new and unforeseen ways. Given this potential, there is a risk that the results may not be fully representative of current language use among children. Moreover, the selection of aggressive or prosocial words was based on current standards, and some words may have had different cultural or contextual meanings in previous decades.

Meanwhile, the sample size is relatively small. There are only 102 effective recordings included, and most of these children did not speak for more than five minutes. Such a small sample size might not be representative of a large population. Additionally, the topic of speech was restricted to the stimuli video. Although the interview questions were open-ended, children usually generated similar responses, thus decreasing the diversity of texts.

Additionally, to fully leverage the capabilities of ChatGPT, its utilization must be optimized. With the ability to process large scripts and contextualize language, ChatGPT might provide valuable insights by extracting specific types of words or phrases from long texts. Rather than generating themes from simple vocabulary lists, a more effective approach could be to prompt ChatGPT to identify the number of aggressive words present in the original script. This will allow for a more advanced analysis for using artificial intelligence in psychology research.

Furthermore, better Natural Language Processing (NLP) tools could be used in parsing words. While this study only focuses on identifying instances of nouns, verbs, and adjectives in the text, other approaches like the Bag-of-Words method might be more comprehensive and effective to all semantics. Bag-of-Words considers every word type by breaking down a piece of text into its constituent words, ignoring grammar and word order, and counting the frequency of each word in the text, resulting in a vector representation. In addition to the Bag-of-Words method, open-source libraries such as Scikit-learn could also be applied for sophisticated NLP tasks. Comparing to Stanza, it is more widely used in recent research.

In future research, machine learning techniques such as topic analysis could be used as a complementary approach to investigate the same topic from a different perspective. This will allow the machine to automatically categorize texts by topics and generate clustering of words without clarifying predefined themes. Researchers could then make sense of the meaning of each grouping. If more aggressive children have more clusters of threatening words, our hypothesis will better be supported. Indeed, we intend to pursue this unsupervised machine learning approach on this sample in further analyses.

To enhance the generalizability of the findings, future research should also aim to increase the sample size and include a wider range of children from different age groups, geographic locations, and ethnic and cultural backgrounds. Additionally, collecting data from children speaking on a more diverse set of topics could provide a more comprehensive understanding of the relations between semantic themes and psychological conditions.

Appendix

Appendix 1. Parent-Rating of Aggression Instrument Items

Proactive aggression

1. Uses physical force to dominate

2. Gets others to gang up on a peer

3. Threatens and bullies others

Reactive Aggression

4. When teased, strikes back

5. Blames others in fights

6. Overreacts angrily to accidents

Unclassified

7. Teases and name-calls

8. Starts fights with peers

9. Gets into verbal arguments

10. When frustrated, quick to fight

11. Breaks rules in games

12. Responds negatively when fails

Fig.1 Distribution of Manually Curated Aggressive Themes in Children's Speech

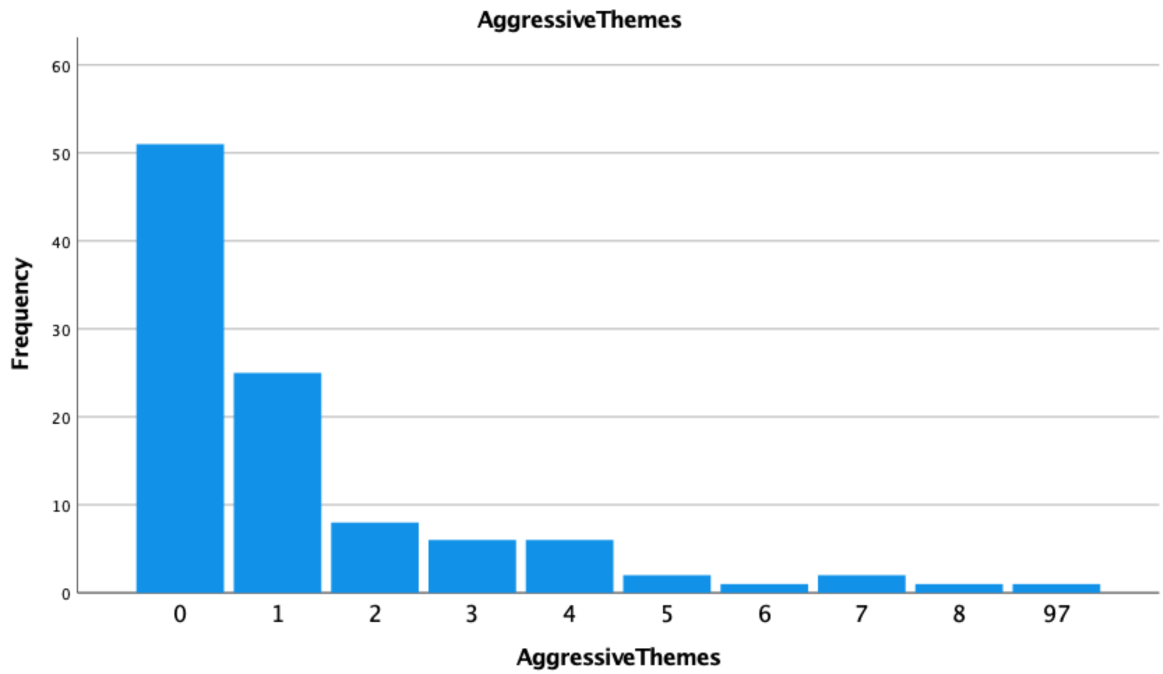


Fig.2 Distribution of ChatGPT generated Aggressive Themes in Children's Speech

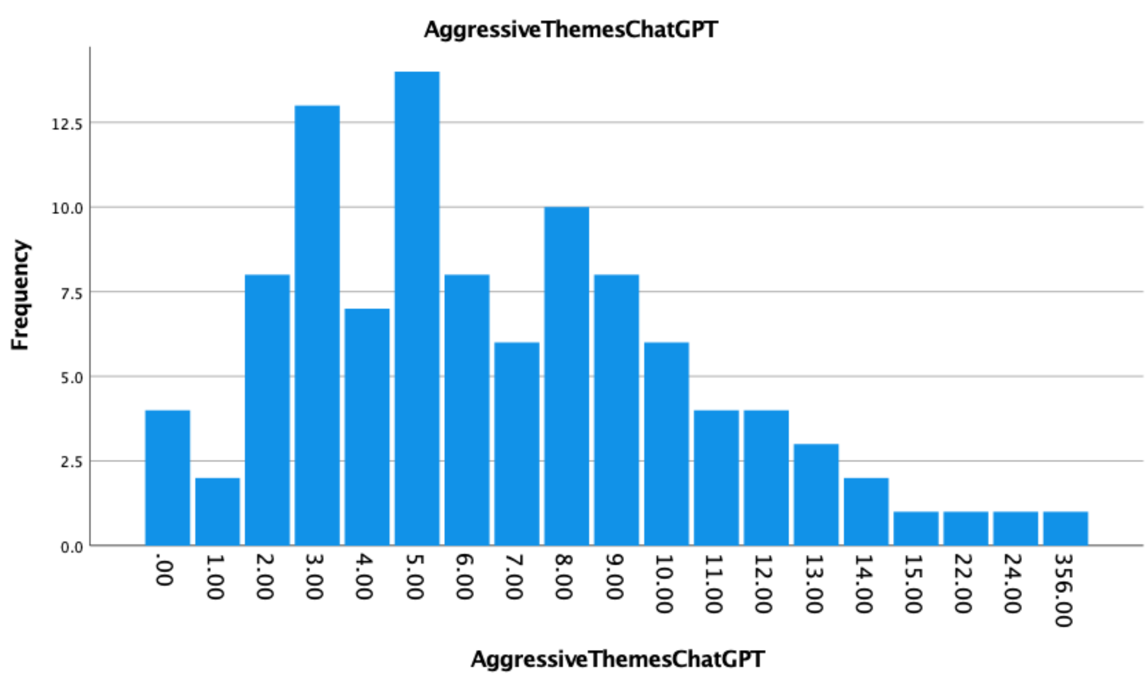


Fig.3 Distribution of Manually Curated Prosocial and Accidental Themes in Children's Speech

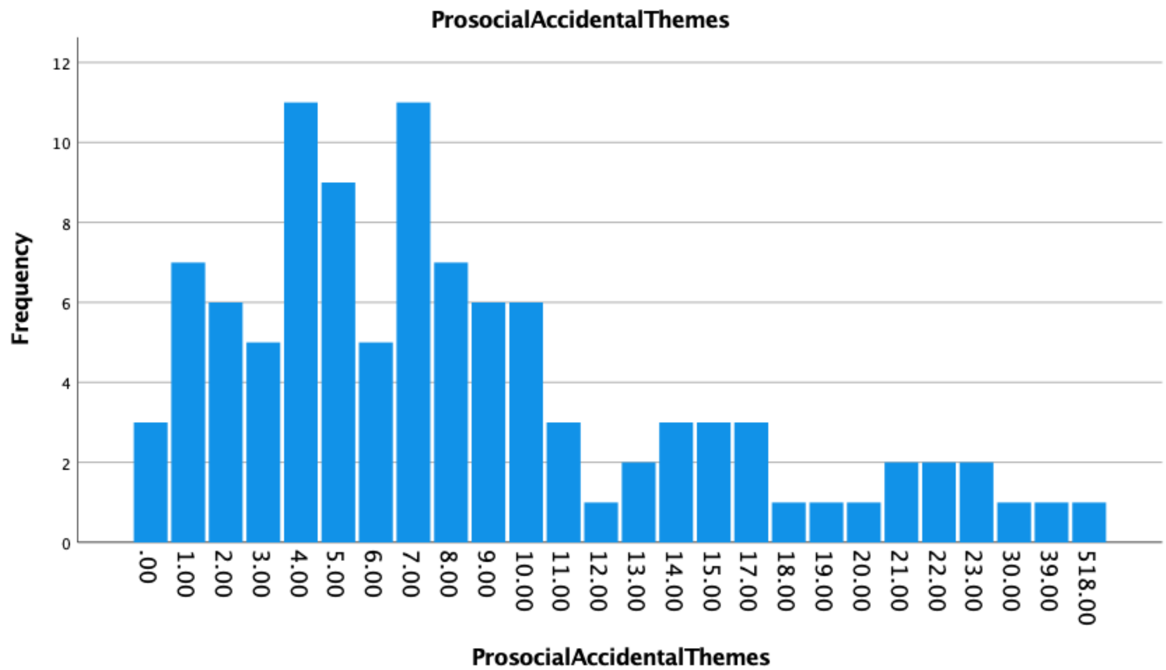


Fig.4 Distribution of ChatGPT Generated Prosocial Themes in Children's Speech

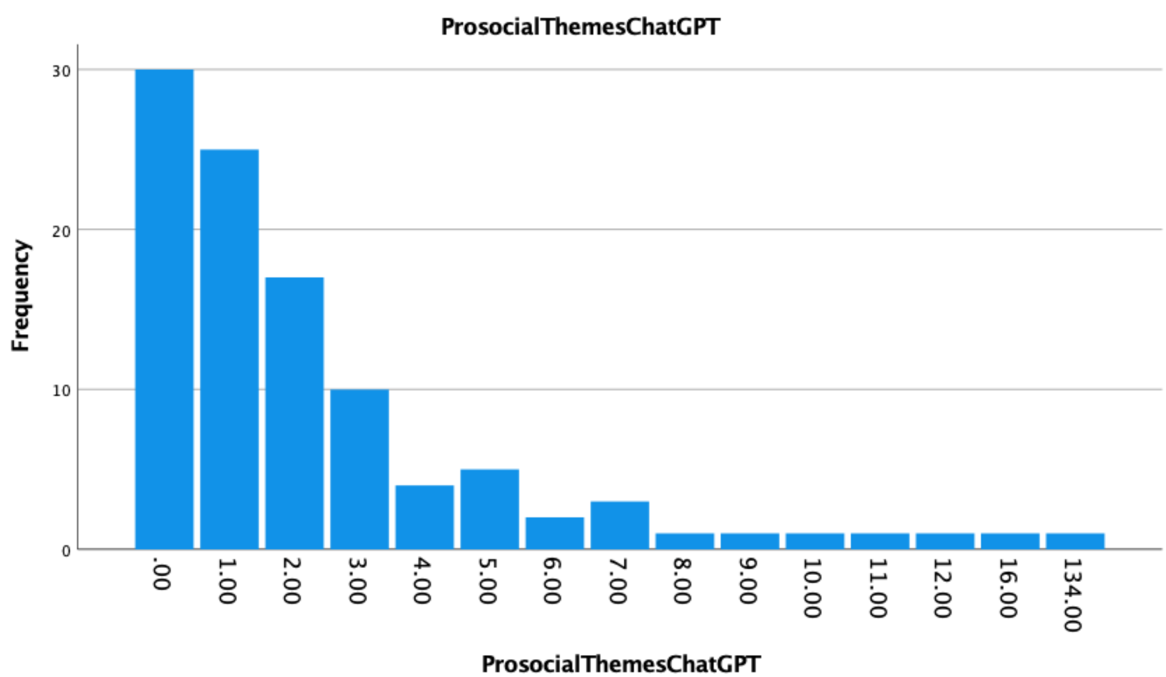


Fig.5 Distribution of Total Level of Aggression

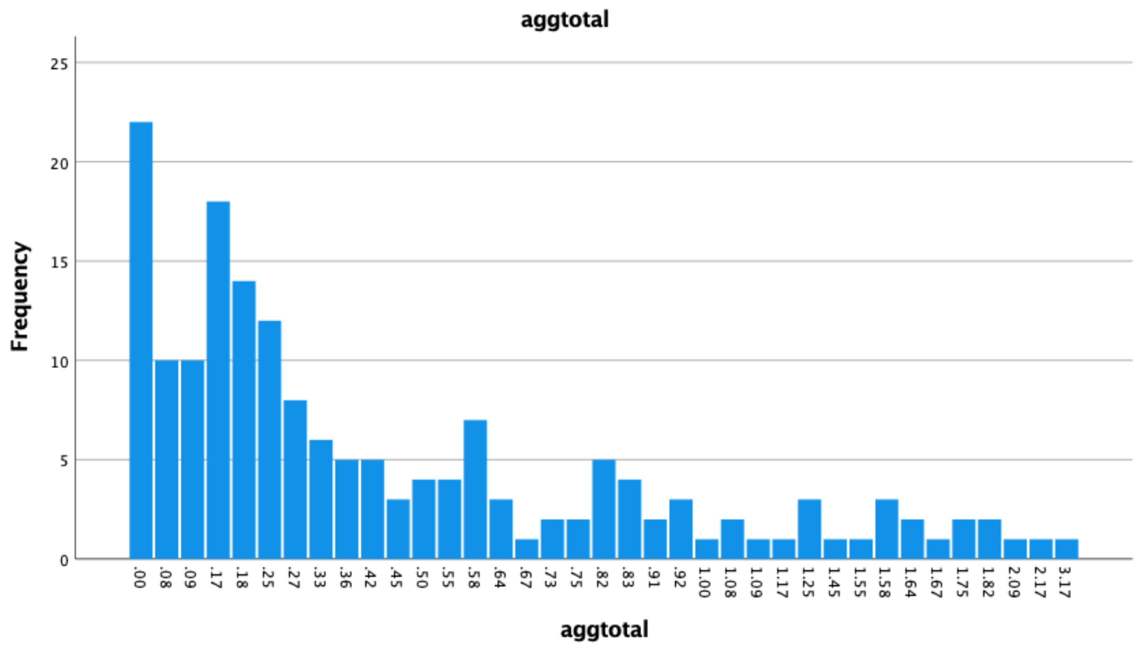


Fig.6 Distribution of Reactive Aggression

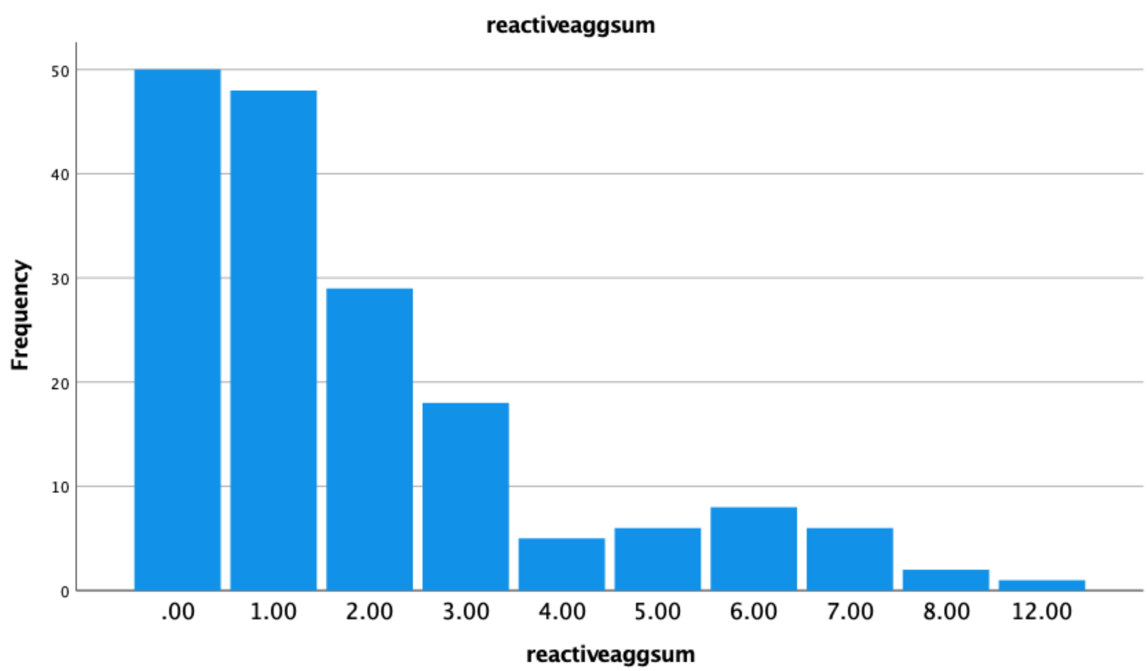
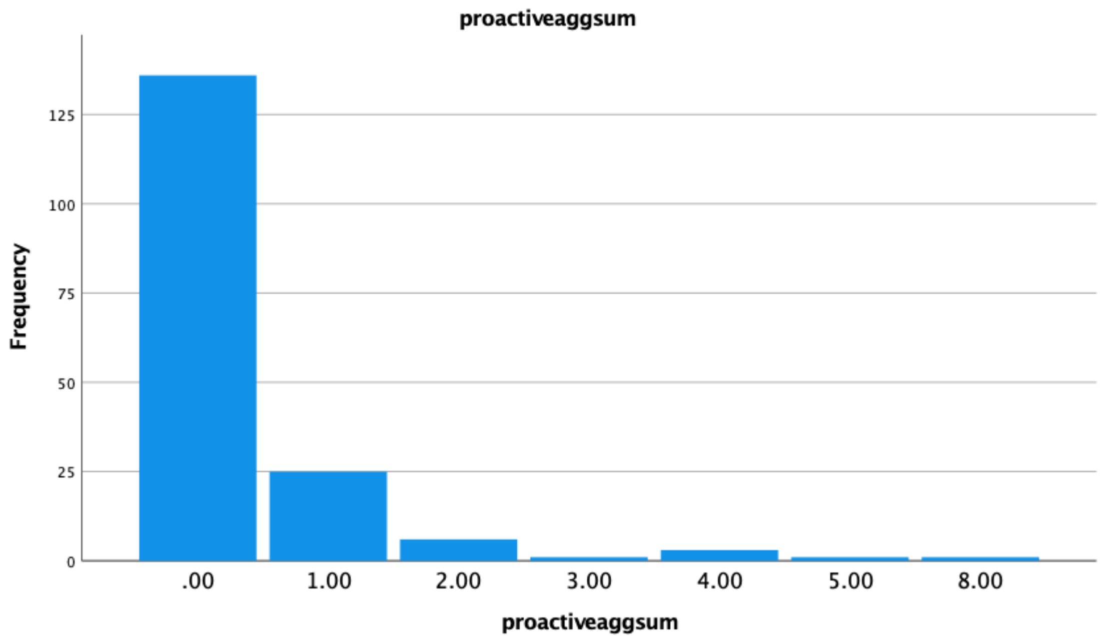


Fig.7 Distribution of Proactive Aggression



Correlations

			AggressiveThemes	AggressiveThemesChatGPT	ProsocialAccidentalThemes	ProsocialThemesChatGPT	aggtotal	aggtotal1	reactiveaggsum	proactiveaggsum
Spearman's rho	AggressiveThemes	Correlation Coefficient	1.000	.423**	.364**	.369**	.301**	.264**	.266**	.256**
		Sig. (1-tailed)	.	<.001	<.001	<.001	.002	.006	.006	.008
		N	103	103	103	103	88	88	88	88
	AggressiveThemesChatGPT	Correlation Coefficient	.423**	1.000	.559**	.392**	.124	.086	.105	.027
		Sig. (1-tailed)	<.001	.	<.001	<.001	.124	.213	.165	.401
		N	103	103	103	103	88	88	88	88
	ProsocialAccidentalThemes	Correlation Coefficient	.364**	.559**	1.000	.636**	.141	.058	.062	.010
		Sig. (1-tailed)	<.001	<.001	.	<.001	.095	.297	.284	.464
		N	103	103	103	103	88	88	88	88
	ProsocialThemesChatGPT	Correlation Coefficient	.369**	.392**	.636**	1.000	.090	-.005	.022	-.045
		Sig. (1-tailed)	<.001	<.001	<.001	.	.202	.481	.418	.338
		N	103	103	103	103	88	88	88	88
	aggtotal	Correlation Coefficient	.301**	.124	.141	.090	1.000	.891**	.875**	.519**
		Sig. (1-tailed)	.002	.124	.095	.202	.	<.001	<.001	<.001
		N	88	88	88	88	173	173	173	173
	aggtotal1	Correlation Coefficient	.264**	.086	.058	-.005	.891**	1.000	.975**	.580**
		Sig. (1-tailed)	.006	.213	.297	.481	<.001	.	<.001	<.001
		N	88	88	88	88	173	173	173	173
	reactiveaggsum	Correlation Coefficient	.266**	.105	.062	.022	.875**	.975**	1.000	.422**
		Sig. (1-tailed)	.006	.165	.284	.418	<.001	<.001	.	<.001
		N	88	88	88	88	173	173	173	173
	proactiveaggsum	Correlation Coefficient	.256**	.027	.010	-.045	.519**	.580**	.422**	1.000
		Sig. (1-tailed)	.008	.401	.464	.338	<.001	<.001	<.001	.
		N	88	88	88	88	173	173	173	173

**. Correlation is significant at the 0.01 level (1-tailed).

Table 1. Correlations between Semantic Themes and Aggression Levels

Table 2. Aggressive, Prosocial and Accidental Themes List

Aggressive Manual	Aggressive ChatGPT	Prosocial/Accidental Manual	Prosocial ChatGPT
mean_VERB_1	mean_VERB_1	accident_NOUN_1	help_VERB_1
hit_VERB_1	mess_VERB_1	help_VERB_1	apologize_VERB_1
wreck_VERB_1	wreck_VERB_1	like_VERB_1	kind_NOUN_1
pick_VERB_1	hit_VERB_1	sorry_ADJ_1	forgive_VERB_1
blame_VERB_1	ruin_VERB_1	nice_ADJ_1	give_VERB_1
ugly_ADJ_1	bad_ADJ_1	rebuild_VERB_1	friend_NOUN_1
knock_VERB_1	jealous_ADJ_1	careful_ADJ_1	care_VERB_1
rude_ADJ_1	hurt_VERB_1	fix_VERB_1	feel_VERB_1
jealous_ADJ_1	selfish_ADJ_1	redraw_VERB_1	happy_ADJ_1
mom_NOUN_1	bully_NOUN_1	apologize_VERB_1	understand_VERB_1
smashed_VERB_1	accuse_VERB_1	share_VERB_1	appreciate_VERB_1
fault_NOUN_1	crazy_ADJ_1	friend_NOUN_1	invite_VERB_1
ruin_VERB_1	upset_ADJ_1	care_VERB_1	participate_VERB_1
selfish_ADJ_1	hate_VERB_1	mistake_NOUN_1	perfect_ADJ_1
trouble_NOUN_1	push_VERB_1	helpful_ADJ_1	
upset_ADJ_1	block_VERB_1	redo_VERB_1	
hurt_VERB_1	shut_VERB_1	kind_NOUN_1	
kick_VERB_1	annoy_VERB_1	forgive_VERB_1	
bad_ADJ_1	discontinue_VERB_1	give_VERB_1	
bully_NOUN_1	stingy_ADJ_1	flatter_ADJ_1	
offend_VERB_1	ignorant_ADJ_1	alright_ADJ_1	
mean_ADJ_1	freak_VERB_1	like_ADJ_1	
wrong_ADJ_1	scream_VERB_1	apology_NOUN_1	

stupid_ADJ_1	remove_VERB_1	wait_VERB_1	
dad_NOUN_1	hurt_NOUN_1	help_NOUN_1	
fail_VERB_1	hit_NOUN_1	pickup_NOUN_1	
stop_NOUN_1	fight_NOUN_1	fun_NOUN_1	
break_VERB_1	revenge_NOUN_1	fair_ADJ_1	
force_VERB_1		happy_ADJ_1	
accuse_VERB_1		permission_NOUN_1	
inconsiderate_ADJ_1		redrew_VERB_1	
hate_VERB_1		tidy_VERB_1	
shoot_VERB_1		responsible_ADJ_1	
disagree_VERB_1		okay_ADJ_1	
damage_NOUN_1		compromise_VERB_1	
crumple_VERB_1		ask_VERB_one	
fight_VERB_1		fun_ADJ_1	
rip_VERB_1		explanation_NOUN_1	
freak_VERB_1		appreciate_VERB_1	
blame_NOUN_1		accidental_ADJ_1	
block_VERB_1		honest_ADJ_1	
bump_VERB_1		reasonable_ADJ_1	
crumble_VERB_1		hope_VERB_1	
stop_VERB_1		clean_ADJ_1	
angry_ADJ_1		offer_VERB_1	
throw_VERB_1		teach_VERB_1	
smash_VERB_1		pleasant_ADJ_1	
crazy_ADJ_1		accept_VERB_1	
destroy_VERB_1		suitable_ADJ_1	

mother_NOUN_1		beautiful_ADJ_1	
fight_NOUN_1		agreement_VERB_1	
parent_NOUN_1		careful_ADJ_one	
frustrate_VERB_1		give_VERB_one	
clumsy_ADJ_1		great_ADJ_1	
screw_VERB_1		remake_VERB_1	
push_VERB_1		improvement_NOUN_1	
shut_VERB_1		perfect_ADJ_1	
shot_NOUN_1		invite_VERB_1	
cut_VERB_1		change_NOUN_1	
revenge_NOUN_1		allowable_ADJ_1	
daddy_NOUN_1		cheer_VERB_1	
jerk_NOUN_1		redrawn_VERB_1	
tear_VERB_1		repaint_NOUN_1	
bang_VERB_1		consult_VERB_1	
crash_VERB_1		polite_ADJ_1	
crush_VERB_1		consider_VERB_1	
crack_VERB_1		funny_ADJ_1	
idiot_NOUN_1		thanks_NOUN_1	
annoy_VERB_1		prettier_ADJ_1	
stingy_ADJ_1		lovely_ADJ_1	
rubbish_VERB_1		love_VERB_1	
tarty_ADJ_1		excuse_VERB_1	
mash_VERB_1		thank_VERB_1	
scream_VERB_1			
wreck_NOUN_1			

silly_ADJ_1			
smack_VERB_1			
hit_VERB_one			
hateful_ADJ_1			
miserable_ADJ_1			
hit_NOUN_1			
heck_NOUN_1			
careless_ADJ_1			
kill_VERB_1			
complain_VERB_1			
harm_VERB_1			
confrontation_NOUN_1			
shut_NOUN_1			
bully_ADJ_1			
scapegoat_NOUN_1			
fire_VERB_1			

References

Berkowitz, L. (1977). Situational and personal conditions governing re- actions to aggressive cues. In D. Magnussen & N. S. Endler (Eds.), Personality at the crossroads: Current issues in interactional psychol- ogy(pp. 165-171). Hillsdale, NJ: Erlbaum.

Curtis, P. R., Frey, J. R., Watson, C. D., Hampton, L. H., & Roberts, M. Y. (2018). Language disorders and problem behaviors: A meta-analysis. *Pediatrics*, *142*(2). https://doi.org/10.1542/peds.2017-3551

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology*, *53*(6), 1146–1158. https://doi.org/10.1037/0022-3514.53.6.1146

Ersan, C. (2020). Early language development and child aggression. *World Journal of Education*, *10*(1), 1. https://doi.org/10.5430/wje.v10n1p1

Ficks, C. A., & Waldman, I. D. (2014). Candidate genes for aggression and antisocial behavior: A Meta-analysis of association studies of the 5HTTLPR and Maoa-uVNTR. *Behavior Genetics*, *44*(5), 427–444. https://doi.org/10.1007/s10519-014-9661-y

Girard, L.-C., Pingault, J.-B., Falissard, B., Boivin, M., Dionne, G., & Tremblay, R. E. (2014). Physical aggression and language ability from 17 to 72 months: Cross-lagged effects in a population sample. *PLoS ONE*, *9*(11). https://doi.org/10.1371/journal.pone.0112185

Huang, F., Kwak, H., & An, J. (1970, January 1). *[PDF] is CHATGPT better than human annotators? potential and limitations of chatgpt in explaining implicit hate speech: Semantic scholar*. ArXiv. Retrieved March 22, 2023, from https://www.semanticscholar.org/paper/Is-ChatGPT-better-than-Human-Annotators-Potential-Huang-Kwak/7cfb88ee7babd5cd717c525f8b642776c47c59cd

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020, April 23). *Stanza: A python natural language processing toolkit for many human languages*. arXiv.org. Retrieved March 22, 2023, from https://arxiv.org/abs/2003.07082

Raine A, Dodge K, Loeber R, Gatzke-Kopp L, Lynam D, Reynolds C, Stouthamer-Loeber M, Liu J. The Reactive-Proactive Aggression Questionnaire: differential correlates of reactive and proactive aggression in adolescent boys. Aggress Behav. 2006;32(2):159–71.

Rezaii, N., Walker, E., & Wolff, P. (2019). A machine learning approach to predicting psychosis using semantic density and latent content analysis. *Npj Schizophrenia*, *5*(1). https://doi.org/10.1038/s41537-019-0077-9

Rule, B. G., & Duker, P. (1973). The effects of intentions and conse- quences on children's evaluations of aggression. Journal of Personal- ityandSocialPsycholog~,27, 184-189.

Silva, P. A., Williams, S., & McGee, R. (1987). A longitudinal study of children with developmental language delay at age three: later intelligence, reading, and behavior problems. *Developmental Medicine and Child Neurology*, 29, 630–640.

Singh, A. L., & Waldman, I. D. (2010). The etiology of associations between negative emotionality and childhood externalizing disorders. *Journal of Abnormal Psychology*, *119*(2), 376–388. https://doi.org/10.1037/a0019342

Waldman, I. D. (1996). Aggressive boys' hostile perceptual and response biases: The role of attention and impulsivity. *Child Development*, *67*(3), 1015. https://doi.org/10.2307/1131877

Zahrt, D. M., & Melzer-Lange, M. D. (2011). Aggressive behavior in children and adolescents. *Pediatrics in Review*, *32*(8), 325–332. https://doi.org/10.1542/pir.32-8-325