



# Zipf's law: Divergence from general content in online communities of peers affected by attention-deficit/hyperactivity disorder (ADHD) and anorexia nervosa

Livio TARCHI<sup>1\*</sup>, Giovanni CASTELLINI<sup>2</sup>, Valdo RICCA<sup>3</sup>

<sup>1</sup>MD, Psychiatry Unit. Department of Health Sciences, University of Florence, Florence, Italy. E-mail: livio.tarchi@unifi.it. ORCID: 0000-0002-9931-5621

<sup>1</sup> MD, Psychiatry Unit. Department of Health Sciences, University of Florence, Florence, Italy. E-mail: giovanni.castellini@unifi.it ORCID: 0000-0003-1265-491X

<sup>1</sup> MD, Psychiatry Unit. Department of Health Sciences, University of Florence, Florence, Italy. E-mail: valdo.ircca@unifi.it ORCID: 0000-0002-9291-2124

\*Correspondence

Cite this paper as: Tarchi L, Castellini G, Ricca V. Zipf's law: Divergence from general content in online communities of peers affected by attention-deficit/hyperactivity disorder (ADHD) and anorexia nervosa. Adv Med Psychol Public Health. 2024;1(2):82-91.  
Doi: 10.5281/zenodo.10637481

Received: 10 November 2023

Revised: 15 January 2024

Accepted: 30 January 2024

## Abstract

**Introduction:** Psycholinguistics studies the relationship between linguistic behavior and psychological factors. For what concerns ADHD (Attention-Deficit/Hyperactivity Disorder) and anorexia nervosa, psycholinguistics can offer insights into the language-related aspects of these conditions. The current study aimed to describe the relationship between rank frequency and frequency of use in written corpora retrieved from internet content.

**Methods:** Linguistic content was retrieved from online communities (Reddit). A total of 2,500 unique contributions were collected for each community (ADHD, anorexia nervosa, general content), composed of an initial script and each subsequent elicited comment. Zipf's law was adopted to estimate the constant and exponential terms describing the relationship between word frequency rank and total usage.

**Results:** A total of 17,853 and 10,049 written prompts were retrieved for ADHD and anorexia nervosa, respectively. This content was compared to 13,386 comments from a general community. The ADHD community exhibited a higher number of average comments per submission, while both ADHD and anorexia nervosa showed higher words per post than the general community. Zipf's law exponential and constant terms were also higher between ADHD and anorexia nervosa compared to general content, suggesting a higher frequency of rare terms in these communities. Notably, while general content emphasized etiquette and community moderation, ADHD and anorexia nervosa discussions predominantly revolved around psychopathological terms.

**Discussion:** This study underscores the unique linguistic characteristics within these communities, shedding light on the interplay between linguistic behavior and psychological factors in the context of ADHD and anorexia nervosa. Psycholinguistic tools could aid in the development of a better understanding of the lived experience of individuals suffering from a psychiatric disorder.

**Take-home message:** This study shows unique linguistic patterns in ADHD and anorexia nervosa online discussions, emphasizing psycholinguistics in understanding language use in online media.

**Keywords:** communication; neurodivergence; internet use; quantitative linguistics; psycholinguistics; social media.

---

## INTRODUCTION

In language, as in other realms of life, emergent properties can originate from the need to allocate scarce resources [1], [2]. For instance, in communication, in light of cognitive constraints, a balance must be reached between effectiveness, speed, specificity, and parsimony [3]. Empirical laws attempt to describe these emergent properties in language. Among them, Zipf's law formulates that the frequency of a word (in a written corpus of natural languages) is inversely proportional to its occurrence rank [4]. In other words, the most common word occurs twice as often as the second most common word, three times as often as the third, and so on. Early evidence showed that Zipf's law may not be maintained in speech or written forms of language in neurodevelopmental disorders, such as Autism Spectrum Disorder [5,6]. As another example of neurodevelopmental disorders, ADHD shows a similar onset and overlapping characteristics with Autism Spectrum Disorder [7,8], but no previous work has explored Zipf's law in this clinical population. However, several empirical accounts described a high rate of language deficits in children with ADHD [7,9], warranting a higher focus on this topic.

By contrast, Anorexia Nervosa (AN) has not been described as being associated with language deficits and does not show evidence in support of neurodevelopmental divergences [7]. Nonetheless, overlapping symptomatic characteristics have been observed between AN and Autism Spectrum Disorder [10]. Moreover, chronic calorie restriction is known to reduce cognitive functioning, but a complex mechanism might underlie differential longitudinal trajectories based on sex and onset of restriction [11]. In fact, animal studies offered evidence that late-onset restriction may worsen cognitive performances, while conversely, exposure to calorie restriction during development might be protective and beneficial [11]. However, the effect of calorie restriction seems selective for specific functions [12] and may not encompass all cognitive domains [13].

General and specific psychopathology might interfere in the before-mentioned allocation of scarce resources for what concerns language. Depression and social anxiety influence verbal expression by emotional status, interpersonal stances, and cognitive processes [14-16]. In ADHD, specific psychopathology might interfere with the law describing the relationship between rank and frequency of use for words in neurotypical linguistic corpora. Core symptoms such as inattention and hyperactivity might tax writing or reading skills [17,18]. Additionally, complex phenomena such as hyperfocus or proficiency in uncommon interests might stress vocabulary divergence [19-21]. To the authors' knowledge, no study has observed language deficits in clinical samples of patients affected by Anorexia Nervosa. Yet, the scientific literature has focused on the social aspect of eating disorders [22], and a specific lexicon, morally driven and morally informed, might stress vocabulary divergence and use in patients or the general public interacting with them [23].

### ***Aims and hypothesis***

For these reasons, a necessary discrimination between the contribution of general and specific psychopathology is needed before natural language processing might inform clinical practice. The current study aimed to describe the relationship between rank frequency and frequency of use in written corpora retrieved from internet content. Written corpora of online communities of peers affected by ADHD and Anorexia Nervosa were compared with general contents, unspecific of psychopathology.

The working hypothesis was that online communities devoted to specific psychopathologies would see an overinflation of most frequent words (those associated with conditions, needs, and symptoms) and less frequent use of singular vocabulary.

To control for these factors, the most distinct words for each community were computed and described.

## METHODS

### *Sample description and pre-processing*

The community of r/ADHD is an inclusive, disability-oriented support group of peers affected by ADHD. The community welcomes sharing stories, difficulties, and non-pharmaceutical strategies to cope with symptoms. The community has almost 1,5 million subscribers, with the highest growth between 2018 and 2021. The subreddit r/Anorexia Nervosa describes itself as a safe place for those affected by the disorder or in recovery. It counts approximately 36,720 subscribers and had the highest growth between 2019 and 2021. By contrast, the board of r/ShowerThoughts is devoted to sharing general wisdom gained during ordinary events. It mainly grew between 2014 and 2019 and counts over 25,126,800 to the present day.

These three reddit communities were searched, and a random sample of 2,500 posts was selected for each subreddit. All comments posted for each 2,500 submissions were retrieved, and a final written corpus of all posts and comments was defined for each community. The sample size and procedure reflect previous work on analyzing mental health correlates on social media platforms like Reddit [24]. To avoid over-sampling lexicon associated with COVID-19, pandemics, infectious diseases, contagions, or symptoms [25], only submissions and comments posted between 1 January 2019 and 31 December 2019 were included. All analyses were conducted in Python 3.8.13, with the support of the following libraries: “pandas” [26]; “numpy” [27]; “scipy” [28]; “matplotlib” [29]; “clean-text” [30]; “seaborn” [31]; “nltk” [32]; “genism” [33]; “sklearn” [34].

Each community's written corpus was pre-processed to remove the upper case, punctuation, control characters, URLs, line breaks, e-mails, phone numbers, numbers, digits, and currency symbols [30]. The overall written corpus was divided into words, and each word stemmed to retrieve its lexical meaning [32]. Stop words (e.g., “the”, “not”, “and” etc.) were removed [32,33]. Both stemming and word stops were performed according to standard practices in natural language processing [32].

### *Empirical distribution of frequency and rank*

Zipf's law is defined by the following:

$$f(N) = \frac{1}{C * N^Z}$$

That is, the frequency of the N<sup>th</sup> word is given by the power law with exponent Z. The standard Zipf's law describes Z = 1, while C is a constant that allows the distribution to be unitary. In order to fit Zipf's distribution to each corpus, the Maximum Likelihood Estimation (MLE) of Z was derived according to procedures defined by Moreno-Sánchez and colleagues [35]. The goodness of fit was evaluated by computing each curve's Mean Absolute Error (MAE; [36]) and Mean Squared Log Error (MSLE; [34]). The MAE of a model is given by the sum of differences between observed and predicted values divided by the sample size [36]. As exponential distributions, standard measurements such as Mean Squared Error (MSE) might be biased in the logarithmic scale. For these reasons, a variation of MSE, namely MSLE, which computes the relative difference between observed and predicted values and the percentual difference between them, was preferred [34].

### *Most distinctively frequent words*

In order to characterize the most distinctively frequent words, the Text Frequency – Inverse Document Frequency (TF-IDF) was adopted [32]. TF-IDF counts the number of times a term occurs in a document and divides it by the total number of documents where the same term appears. This procedure aims to reduce the weight of very frequent but not distinct terms in a document. A selection of the top 10 words by TF-IDF score was chosen to represent the most distinct words for each community.

## RESULTS

### Sample descriptives

A total of 2,500 submissions were retrieved, as planned, for each of the three online communities described. Comments posted for each submission were associated with the original body of text, and 41,288 comments were analyzed. In particular, the community of peers affected by ADHD had 17,853 comments and an average of 7.141 comments per submission. The community of peers affected by Anorexia Nervosa had a total of 10,049 comments and an average of 4.020 comments per submission. Finally, the general community, unspecific of psychopathology, had a total of 13,386 comments and an average of 5.354 comments per submission. Please see Table 1 for a tabular representation of the sample.

**Table 1.** Sample descriptive.

Subreddit	Total number of comments	Average comments per submission	Words per Posts
r/ADHD	17,853	7.141	27.649
r/AnorexiaNervosa	10,049	4.020	24.515
r/ShowerThoughts	13,386	5.354	11.108

*Note:* All subreddits were sampled for 2,500 submissions, with content posted from 1 January 2019 to 31 December 2019. Words per post given by both submissions and comments.

### Empirical distributions of words

The frequency of each unique word was calculated for each online community. ADHD was the community with the highest total words (493,616) but the lowest prevalence of unique words in the sample (4.87%). Anorexia Nervosa, on the other hand, had a total of 246,348 words and an intermediate prevalence of unique words (6.29%) between ADHD and general content. General content, in fact, had the lowest number of total words (148,696) but the highest prevalence of unique words (10.21%). Moreover, a gradient in the goodness of fit for Zipf's law was observed. The community of peers with ADHD showed the worst goodness of fit (MAE 22.056; MSLE 3.633), Anorexia Nervosa an intermediate goodness of fit (MAE 18.872; MSLE 3.477), and the general community the best (MAE 10.658; MSLE 2.624).

**Table 2.** Frequency of unique words, empirical distributions.

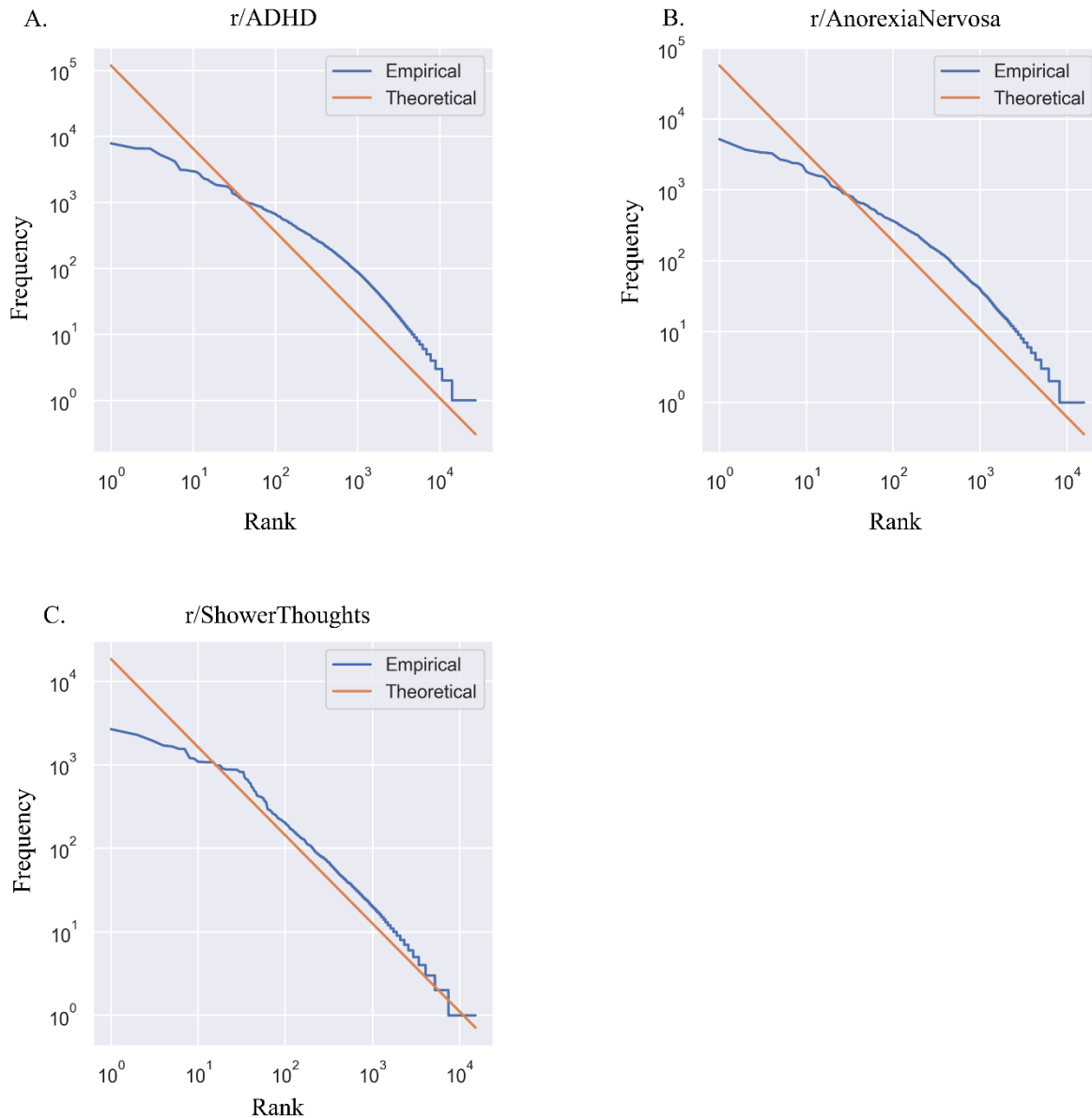
Subreddit	Total words	Unique words	Z exponent	C constant	MAE	MSLE
r/ADHD	493,616	27,037 (4.87%)	1.260	4.169	22.056	3.633
r/AnorexiaNervosa	246,348	15,514 (6.29%)	1.241	4.336	18.872	3.477
r/ShowerThoughts	148,696	15,187 (10.21%)	1.055	8.027	10.658	2.624

*Note:* Prevalence of unique words is given by dividing unique words by total words.

MAE: Mean Absolute Error; MSLE: Mean Squared Log Error.

The Z exponent and C constant of Zipf's law were fit according to MLE for each written corpus. The most used words were less frequent than expected across all communities. However, a higher frequency of rare words (the right-hand side of

the distribution) was observed only in communities of peers affected by ADHD or AN. The divergence from theoretical distribution (best fitted according to MLE as described above) interested words ranked 1000<sup>th</sup> or lower. The results can be found in Figure 1.



**Figure 1.** Relationship between frequency and rank of each word, theoretical and empirical distributions.

### ***Most distinct words***

The 10 most distinct words per community are represented in Table 3. Both the communities of peers affected by ADHD and AN were more distinctively represented by the words “time” and “people”. The ADHD community was represented mainly by its specific term (“ADHD”, TF-IDF 0.289), as the community of AN patients (“eating,” TF-IDF 0.210, “eat”, TF-IDF 0.195, “weight”, TF-IDF 0.212). By contrast, the general community of r/ShowerThought was distinctively represented by words pertaining to the realm of etiquette and moderation in the community (“removed,” “please”, “frequently asked questions, FAQ,” “rules,” “moderators”).

**Table 3.** Distinct word per community.

r/ADHD		r/AnorexiaNervosa		r/ShowerThoughts	
Word	TF-IDF	Word	TF-IDF	Word	TF-IDF
ADHD	0.289	Feel	0.236	Please	0.313
Get	0.259	Know	0.226	Removed	0.237
Time	0.206	Weight	0.212	FAQ	0.233
One	0.154	Eating	0.210	Rules	0.227
Really	0.152	Get	0.198	Reddit	0.224
Know	0.151	Eat	0.195	Wiki	0.208
Things	0.148	Really	0.165	Moderators	0.202
People	0.148	Want	0.158	Google	0.174
Feel	0.142	Time	0.146	Games	0.174
Work	0.132	People	0.144	Automatically	0.161

*Note:* TF-IDF: Text Frequency – Inverse Document Frequency.

## DISCUSSION

The current study aimed to explore the linguistic patterns in online communities of peers affected by ADHD and Anorexia Nervosa (AN) using Zipf's law as a framework. Contrary to the working hypothesis, the findings indicated no overinflation of the most frequent words in both communities. Instead, a higher frequency of unique, less-used words was observed. This divergence from expected linguistic patterns under Zipf's law may reflect the unique communicative needs and experiences of individuals in these groups.

In the ADHD community, the prevalence of unique words could be linked to the diverse and idiosyncratic nature of the disorder [37]. ADHD's core symptoms, like inattention and hyperactivity, might lead to a broader, more varied lexical choice as individuals seek to articulate their distinct experiences and coping strategies. This is further supported by the presence of words like "ADHD," "time," and "people" among the most distinct words, highlighting the community's focus on personal experiences and social interactions.

Similarly, the AN community displayed a unique lexical pattern, potentially reflecting the complex emotional and physical experiences associated with eating disorders [38-40]. Words like "eating," "weight," and "eat" were prominent, indicating a focused discussion on specific aspects of the disorder. The high prevalence of unique words in this community might indicate the nuanced and personal nature of discussing and coping with AN and sharing diverse personal stories and recovery journeys [41,42].

The current descriptive analysis using Zipf's law revealed that the ADHD community had a lower prevalence of unique words compared to the AN community. Still, both deviated from the general content regarding word frequency distribution.

This suggests that while both communities tend towards lexical diversity, the nature and context of their discussions differ significantly.

The TF-IDF analysis further illuminated these differences, revealing the most distinct words in each community. This analysis provides a deeper understanding of each group's thematic focuses and linguistic peculiarities [43,44].

### **Future directions**

The findings of this study open new avenues for future research. A deeper exploration into the psycholinguistic patterns of other neurodivergent communities could provide valuable insights into the unique ways these groups use language. Additionally, longitudinal studies could track changes in language use over time, particularly as members of these communities evolve in their understanding and experiences of their conditions.

### **Study limitations**

This study, however, is not without limitations. Firstly, the clinical samples are divergent and lack an objective definition of psychopathology or confirmation of diagnosis. This raises questions about the generalizability of our findings to all individuals with ADHD or AN. Furthermore, relying on online written text may not fully capture the complexities of language use in these communities, as online communication can significantly differ from other forms of expression.

### **CONCLUSIONS**

Contrary to the working hypothesis, neither of the two online communities of peers showed an overinflation of the most frequent words. By contrast, a higher frequency of unique, less-used words was observed in both. A divergence from the content retrieved by general communities was present in the online discourse of patients with ADHD and AN.

In conclusion, our study reveals intriguing linguistic patterns in online communities of peers affected by ADHD and AN, challenging the traditional understanding of Zipf's law in the context of neurodivergence. These findings underscore the importance of considering unique linguistic expressions in psychopathological research and highlight language use's rich, diverse nature in neurodivergent communities.

**Author Contributions:** Conceptualization: LT; methodology: LT, GC; software: LT; validation: VR; formal analysis, LT; investigation: LT, GC, VR; resources: LT; data curation: GC; writing—original draft preparation: LT, GC; writing—review and editing, VR, GC; visualization: LT; supervision: GC, VR; project administration: LT. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** None.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Disclaimer/Publisher's Note:** The Publisher remains neutral regarding jurisdictional claims in published maps and institutional affiliations. Additionally, the Publisher is not responsible for the accuracy, completeness, or validity of the content of scientific articles published herein. This statement exempts the Publisher from any responsibility regarding the content of scientific articles, which is solely the responsibility of the authors and peer reviewers.

### **References**

1. Ghazi S, Schneider M. Market Value of Rarity: A Theory of Fair Value and Evidence from Rare Baseball Cards. 29 June 2021. Available from: SSRN: <https://ssrn.com/abstract=4031091>. doi: 10.2139/ssrn.4031091.
2. Lazzardi S, Valle F, Mazzolini A, Scialdone A, Caselle M, Osella M. Emergent Statistical Laws in Single-Cell Transcriptomic Data. bioRxiv. p. 2021.06.16.448706. 4 February 2022. doi: 10.1101/2021.06.16.448706.
3. Marchman VA, Fernald A, Hurtado N. How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish–English bilinguals. *J Child Lang.* 2010;37(4):817–840. doi: 10.1017/S0305000909990055.

4. Newman M. Power laws, Pareto distributions and Zipf's law. *Contemp Phys*. 2005;46(5):323–351. doi: 10.1080/00107510500052444.
5. Beykikhoshk A, Arandjelović O, Phung D, Venkatesh S, Caelli T. Data-mining twitter and the autism spectrum disorder: A Pilot study. In 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), Aug. 2014, pp. 349–356. doi: 10.1109/ASONAM.2014.6921609.
6. Garcia-Camargo L, Kilfoyle J, De Palma P, VanDam M, Stover J. A First Look at Zipf's Law and the Speech of Children with Autism Spectrum Disorder. *Proceedings of the Annual Meeting of the Cognitive Science Society*. vol. 44, no. 44, 2022, Accessed: Jul. 09, 2022. [Online]. Available from: <https://escholarship.org/uc/item/4t02k3p6>.
7. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (5th ed.). 2013. Available from: <https://doi.org/10.1176/appi.books.9780890425596>.
8. Mayes SD, Calhoun SL, Mayes RD, Molitoris S. Autism and ADHD: Overlapping and discriminating symptoms. *Res Autism Spectr Disord*. 2012;6(1):277–285. doi: 10.1016/j.rasd.2011.05.009.
9. Bruce B, Thernlund G, Nettelbladt U. ADHD and language impairment. *Eur Child Adolesc Psychiatry*. 2006;15(1):52–60. doi: 10.1007/s00787-006-0508-9.
10. Westwood H, Eisler I, Mandy W, Leppanen J, Treasure J, Tchanturia K. Using the Autism-Spectrum Quotient to Measure Autistic Traits in Anorexia Nervosa: A Systematic Review and Meta-Analysis. *J Autism Dev Disord*. 2016;46(3): 964–977. doi: 10.1007/s10803-015-2641-0.
11. Prvulovic MR, Milanovic DJ, Vujovic PZ, Jovic MS, Kanazir SD, Todorovic ST, et al. Late-Onset Calorie Restriction Worsens Cognitive Performances and Increases Frailty Level in Female Wistar Rats. *J Gerontol A Biol Sci Med Sci*. 2022 May 5;77(5):947-955. doi: 10.1093/gerona/ab353.
12. Dal-Pan A, Pifferi F, Marchal J, Picq JL, Aujard F, R. Consortium. Cognitive Performances Are Selectively Enhanced during Chronic Caloric Restriction or Resveratrol Supplementation in a Primate. *PLOS ONE*. 2011;6(1):e16581. doi: 10.1371/journal.pone.0016581.
13. Martin CK, Anton SD, Han H, York-Crowe E, Redman LM, Ravussin E, et al. Examination of cognitive function during six months of calorie restriction: results of a randomized controlled trial. *Rejuvenation Res*. 2007 Jun;10(2):179-190. doi: 10.1089/rej.2006.0502.
14. Rowa K, Paulitzki JR, Ierullo MD, Chiang B, Antony MM, McCabe RE, et al. A false sense of security: safety behaviors erode objective speech performance in individuals with social anxiety disorder. *Behav Ther*. 2015 May;46(3):304-314. doi: 10.1016/j.beth.2014.11.004.
15. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preotiuc-Pietro D, et al. Facebook language predicts depression in medical records. *Proc Natl Acad Sci U S A*. 2018 Oct 30;115(44):11203-11208. doi: 10.1073/pnas.1802331115.
16. Bernard JD, Baddeley JL, Rodriguez BF, Burke PA. Depression, Language, and Affect: An Examination of the Influence of Baseline Depression and Affect Induction on Language. *J Lang Soc Psychol*. 2016;35(3):317-326. doi: 10.1177/0261927X15589186.
17. Rodriguez A, Järvelin MR, Obel C, Taanila A, Miettunen J, Moilanen I, et al. Do inattention and hyperactivity symptoms equal scholastic impairment? Evidence from three European cohorts. *BMC Public Health*. 2007 Nov 13;7:327. doi: 10.1186/1471-2458-7-327.
18. Swanson E, Barnes M, Fall AM, Roberts G. Predictors of Reading Comprehension Among Struggling Readers Who Exhibit Differing Levels of Inattention and Hyperactivity. *Read Writ Q*. 2018;34(2):132–146. doi: 10.1080/10573569.2017.1359712.
19. Hupfeld KE, Abagis TR, Shah P. Living "in the zone": hyperfocus in adult ADHD. *Atten Defic Hyperact Disord*. 2019 Jun;11(2):191-208. doi: 10.1007/s12402-018-0272-y. Epub 2018 Sep 28. Erratum in: *Atten Defic Hyperact Disord*.
20. Mottron L, Dawson M, Soulières I. Enhanced perception in savant syndrome: patterns, structure and creativity. *Philos Trans R Soc Lond B Biol Sci*. 2009 May 27;364(1522):1385-1391. doi: 10.1098/rstb.2008.0333.
21. Sedgwick JA, Merwood A, Asherson P. The positive aspects of attention deficit hyperactivity disorder: a qualitative investigation of successful adults with ADHD. *Atten Defic Hyperact Disord*. 2019 Sep;11(3):241-253. doi: 10.1007/s12402-018-0277-6.
22. Becker AE. Television, disordered eating, and young women in Fiji: negotiating body image and identity during rapid social change. *Cult Med Psychiatry*. 2004 Dec;28(4):533-559. doi: 10.1007/s11013-004-1067-5.



23. Griffin J, Berry EM. A modern day holy anorexia? Religious language in advertising and anorexia nervosa in the West. *Eur J Clin Nutr.* 2003 Jan;57(1):43-51. doi: 10.1038/sj.ejcn.1601511.
24. Schrading N, Ovesdotter Alm C, Ptucha R, Homan C. An Analysis of Domestic Abuse Discourse on Reddit. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, Sep. 2015, pp. 2577–2583. doi: 10.18653/v1/D15-1309.
25. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. *JMIR Public Health Surveill.* 2020 May 5;6(2):e19374. doi: 10.2196/19374.
26. Reback J, McKinney W, Van de Bossche J, Augspurger T, Cloud P, et al. Pandas-dev/pandas: Pandas 1.0.3. Zenodo, Mar. 18, 2020. doi: 10.5281/zenodo.3715232.
27. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature.* 2020 Sep;585(7825):357-362. doi: 10.1038/s41586-020-2649-2.
28. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. *Nat Methods.* 2020 Mar;17(3):261-272. doi: 10.1038/s41592-019-0686-2. Epub 2020 Feb 3. Erratum in: *Nat Methods.* 2020 Feb 24.
29. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng.* 2007;9(3):90–95. doi: 10.1109/MCSE.2007.55.
30. Rinker TW (2018), *textclean: Text Cleaning Tools*, version 0.9.3, Feb. 02, 2023 <https://github.com/trinker/textclean>.
31. Waskom M. Seaborn: statistical data visualization. *J Open Source Softw.* 2021;6(60):3021. doi: 10.21105/joss.03021.
32. Bird S, Klein E, Loper E. *Natural Language Processing with Python*. First edition. Beijing: Cambridge Mass; 2009.
33. Rehurek R, Gensim: Python framework for fast Vector Space Modelling. 2022. Accessed: 10 July 2022. Available from: <http://radimrehurek.com/gensim>.
34. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2011;12 (85):2825–2830.
35. Moreno-Sánchez I, Font-Clos F, Corral Á. Large-Scale Analysis of Zipf's Law in English Texts. *PLoS One.* 2016 Jan 22;11(1):e0147073. doi: 10.1371/journal.pone.0147073.
36. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res.* 2005;30(1):79–82. doi: 10.3354/cr030079.
37. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders : Fifth Edition Text Revision DSM-5-TR™*. American Psychiatric Association Publishing; 2022.
38. Tarchi L, Rossi E, Faldi M, Cassioli E, Ricca V, Castellini G. Linking Embodiment Disorder and Bulimia Nervosa. In *Eating Disorders*, Patel V and Preedy V, Eds. Cham: Springer International Publishing; 2021, pp. 1–19. doi: 10.1007/978-3-030-67929-3\_46-2.
39. Cassioli E, Rossi E, Martelli M, Arganini F, Giuranno G, Siviglia S, et al. Longitudinal Coupling between Eating Disorder Psychopathology and Depression in Patients with Anorexia Nervosa and Bulimia Nervosa Treated with Enhanced Cognitive Behavior Therapy: A One-Year Follow-Up Study. *Brain Sci.* 2023 Mar 24;13(4):535. doi: 10.3390/brainsci13040535.
40. Eshkevari E, Rieger E, Longo MR, Haggard P, Treasure J. Increased plasticity of the bodily self in eating disorders. *Psychol Med.* 2012 Apr;42(4):819-828. doi: 10.1017/S0033291711002091.
41. Bardone-Cone AM, Hunt RA, Watson HJ. An Overview of Conceptualizations of Eating Disorder Recovery, Recent Findings, and Future Directions. *Curr Psychiatry Rep.* 2018 Aug 9;20(9):79. doi: 10.1007/s11920-018-0932-9.
42. Eshkevari E, Rieger E, Longo MR, Haggard P, Treasure J. Persistent body image disturbance following recovery from eating disorders. *Int J Eat Disord.* 2014 May;47(4):400-409. doi: 10.1002/eat.22219.
43. Stanghellini G. The PHD Method for Psychotherapy: Integrating Phenomenology, Hermeneutics, and Psychodynamics. *Psychopathology.* 2019;52(2):75-84. doi: 10.1159/000500272.
44. Stanghellini G, Daga GA, Ricca V. From the patients' perspective: what it is like to suffer from eating disorders. *Eat Weight Disord.* 2021

Apr;26(3):751-755. doi: 10.1007/s40519-020-00913-8.



**Copyright:** © 2024 by the authors. Submitted for possible open-access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).