

# An information-theoretic analysis of agrammatism in Broca's aphasia

Neguine Rezaii

## Abstract

Recently we proposed and tested a novel framework based on information theory about the features of agrammatism in patients with the nonfluent variant of primary progressive aphasia (nfvPPA). These features include short sentences, simplified structures, the omission of function words, reduced use of verbs, and increased use of heavy verbs. After distinguishing the syntactic and lexical features of agrammatism, we proposed that the lexical features are not defects but rather a compensatory response to structural deficits. We showed that although patients with nfvPPA have difficulty using complex structures, they use more informative words to facilitate the transfer of their intended message. In this work, we demonstrated the generalizability of our findings to patients with Broca's aphasia who share similar neuroanatomical involvements with patients with nfvPPA. We found that patients with Broca's aphasia use simpler syntactic structures as measured by higher syntax frequency. We also found that the use of simpler syntax predicts the use of lower-frequency words. Furthermore, our study showed that producing sentences of lower word frequency is likely achieved through the canonical features of agrammatism, i.e., higher proportions of content words over function words, nouns over verbs, and heavy verbs over light verbs.

## Author affiliation:

Frontotemporal Disorders Unit, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02129, USA

## Correspondence to:

Neguine Rezaii  
nrezaii@mgh.harvard.edu  
149 13th Street, Suite 10.011, Charlestown, MA 02129, USA

**Keywords:** Agrammatism, nonfluent aphasia, Broca's aphasia, word frequency, syntax frequency, syntax-lexicon trade-off, information theory

## Introduction

Lesions to the left hemisphere's inferior frontal gyrus are associated with a distinct style of language production known as agrammatism. These lesions may result from stroke or neurodegenerative disorders, such as primary progressive aphasia (PPA). Common symptoms of agrammatism include short, simplified sentences, the omission of function words, a decreased use of verbs relative to nouns, and increased use of heavy verbs than light verbs.<sup>1-9</sup> The predominant account of the disease suggests that these symptoms result from a deficit in syntax processing.<sup>6,10</sup> According to this account, the reduced use of function words and verbs in agrammatism can be attributed to their primary syntactic roles in a sentence, which often involve indicating grammatical relationships between content words.<sup>6,10-16</sup> Under this account, however, it remains unclear why patients with agrammatism use more heavy verbs, have intact online access to the verb lexicon,<sup>17,18</sup> and make minimal errors in using function words in sentence completion tasks.<sup>19</sup>

In a series of recent studies, first outlined by Rezaii et al. 2022,<sup>20</sup> we proposed an alternative view of agrammatism based on information theory. We demonstrated that the lexical features of agrammatism are not deficits but rather a compensatory response to short and simple sentences to increase sentence information. A key requirement of this proposal is to distinguish between the syntactic and lexical abnormalities of agrammatism. The syntactic abnormalities would be short sentences and simplified structures. To measure the complexity of syntactic structures, we introduced syntax frequency, which measures the frequency of a given syntactic structure based on a large corpus of spoken language.<sup>21,22</sup> Syntactic structures that are less common-- e.g., passive compared to active -- are more difficult to access, hence more complex. Using this metric, we showed that patients with the nonfluent variant of PPA (nfvPPA) use high-frequency syntax. On the other hand, the lexical abnormalities of agrammatism consist of increased use of content words over function words, nouns over verbs, and heavy verbs over light verbs. Our studies showed that patients with nfvPPA use lower-frequency words than healthy individuals. Crucially, we found a trade-off in the average frequency of words and the average frequency of syntax at the sentence level.<sup>23</sup> This trade-off was true for both all words and content words. The syntax-lexicon trade-off thus predicts that if a sentence contains simple structures, its words will be more complex. These analyses were based on the language samples of participants as they described a picture to keep the overall message of language constant.

To show whether the lexical features of agrammatism arise as a result of simplified structures, we asked healthy speakers to describe the same picture using short sentences of one to two words. We found that in this constrained task, similar lexical features of agrammatism emerged in the language of healthy speakers. Constrained sentences contained more content than function words, more nouns than verbs, and more heavy verbs than light verbs. Interestingly, short sentences of healthy individuals contained more verbs in *-ing* form, a finding consistently reported in the language of patients with nonfluent aphasia.<sup>24-27</sup> These lexical features of short sentences resulted in a lower average word frequency when compared with unconstrained language. Lastly, we showed that using words of lower frequency in sentence-making results in higher information content of sentences. We measured information using lexical entropy and surprisal.<sup>28,29</sup> The collective of our findings in nfvPPA suggests that the lexical profile of agrammatism is a compensatory response to short, simple sentences to increase sentence information.

In this study, we aim to test the generalizability of our findings to patients with Broca's aphasia using language samples from the AphasiaBank.<sup>30</sup> Previous research has shown considerable similarities between the language of patients with nfvPPA and Broca's aphasia, likely due to their similar neuroanatomical defect in the left inferior frontal gyrus.<sup>31,32</sup> Given the similar language characteristics of the two groups, we hypothesize that patients with Broca's aphasia 1) use simpler syntactic structures as measured by higher syntax frequency, 2) use lower frequency words as a compensatory response to simpler structures, and 3) show a trade-off between the frequency of words and syntactic structures which further highlights the lexical profile of agrammatism as a strategy to compensate for the loss of syntactic information.

## Methods

**Language samples.** We obtained language samples from the AphasiaBank.<sup>30</sup> We chose two topics for the picture description task, the cat rescue<sup>33</sup> (91 patients with Broca's aphasia and 256 healthy controls) and the broken window stories<sup>34</sup> (86 patients with Broca's aphasia and 271 healthy controls).

**Extracting language features.** To extract various language features, we used Quantitext, a language toolbox developed in the MGH FTD Unit, to increase precision while reducing human labor.<sup>35</sup> To determine the part of speech of words, the toolbox uses the automated Stanza Lexicalized Parser.<sup>36</sup> We measured the proportion of content words to all words by dividing the number of content words by the number of all words in a sentence. Nouns, verbs (except *be*, *have*, and *do*), adjectives, and adverbs were considered content words. All other words were classified as function words. The proportion of heavy verbs to all verbs was measured by dividing the number of heavy verbs by the total number of verbs in a sentence. The following verbs were classified as light verbs: 'be', 'go', 'take', 'come', 'make', 'get', 'give', and 'have' while excluding auxiliaries from this list.<sup>37</sup> To measure word frequency, we used the Switchboard corpus,<sup>38</sup> which consists of spontaneous telephone conversations averaging six minutes in length spoken by over 500 speakers of both sexes from a variety of dialects of American English. We use this corpus to estimate word frequency in spoken English, independently of the patient and control samples. The corpus contains 2,345,269 words. The word frequency of each sentence is calculated by taking the average log frequency of content words within that sentence. Therefore, in this manuscript, word frequency indicates the average log frequency of content words in a given sentence. To measure syntax frequency,<sup>21</sup> we first parsed the sentences in the corpus using Stanza to extract headed syntactic rules. A headed syntactic rule is determined by the head and all its dependents in a dependency parse, whether they occur on the left or right. We applied this method on Switchboard, which resulted in 954,616 rules, and measured the syntax frequency of each sentence of participants by calculating the average log frequency of the syntactic rules of that sentence based on the Switchboard counts.

**Statistical analysis.** For the statistical analyses of this study, we used the R software version 4.1.2. To compare language features at the sentence level across different groups, we used mixed-effects models with subject-specific random intercept via the lme4 package in R.<sup>39</sup> We used independent t-tests to compare the features of agrammatism between patients and healthy controls.

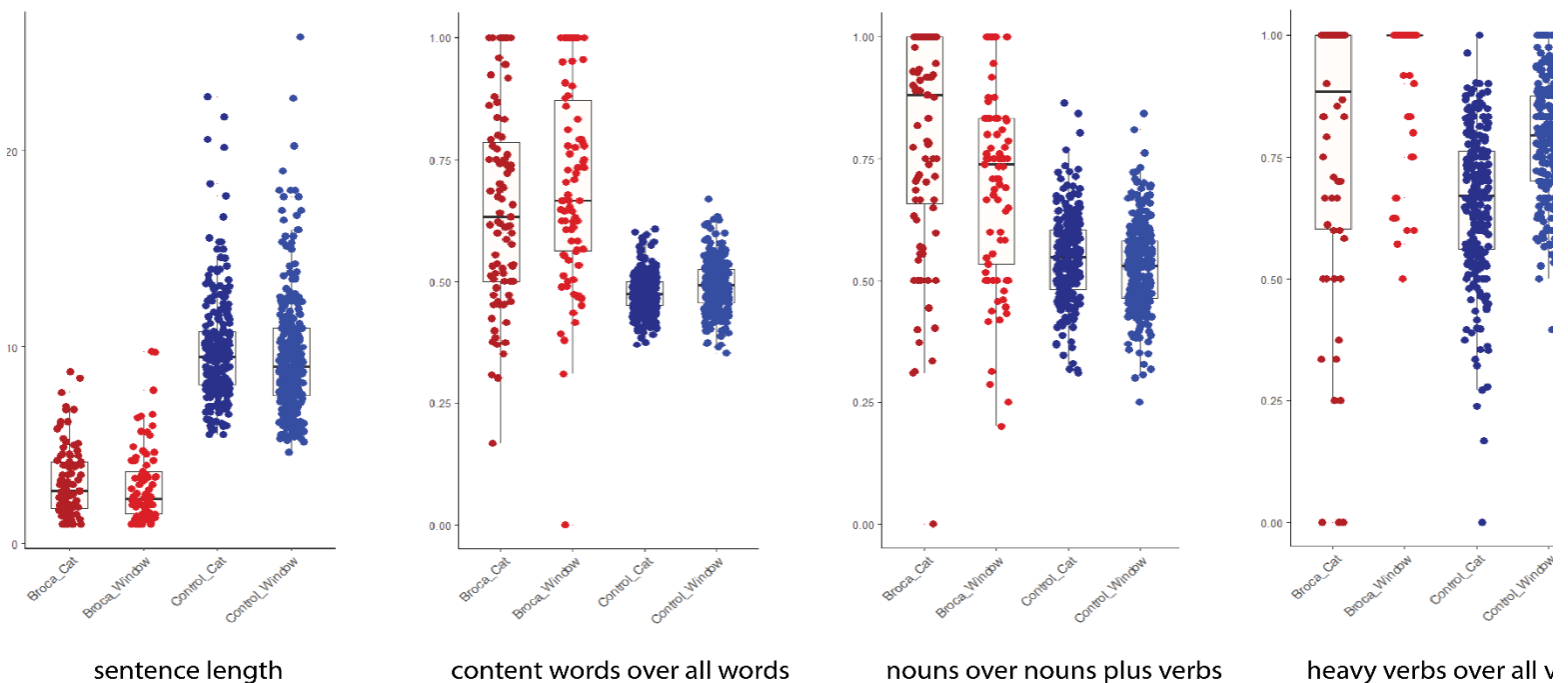
## Results

### 1. The conventional description of the language of patients with agrammatism

Here, we first analyze the language of patients with Broca's aphasia compared to healthy speakers to establish the existence of agrammatic features in their language. Following the convention of the field, these analyses were performed at the individual level (Figure 1).

Cat story. Compared to the language of healthy controls, the sentences of patients with Broca's aphasia were shorter ( $t(238.22) = -26.18, p < 0.001$ ), with a higher proportion of content words to all words ( $t(92.484) = 8.373, p < 0.001$ ), a higher proportion of nouns to the sum of nouns plus verbs ( $t(100.77) = 10.138, p < 0.001$ ) and a higher proportion of heavy verbs to all verbs ( $t(68.39) = 2.68, p = 0.009$ ).

Window story. We found similar results in analyzing language samples obtained from the window story. Compared to the language of healthy controls, the sentences of patients with Broca's aphasia were shorter ( $t(238.22) = -26.18, p < 0.001$ ), with a higher proportion of content words to all words ( $t(88.57) = 9.38, p < 0.001$ ), a higher proportion of nouns to the sum of nouns plus verbs ( $t(94.74) = 7.83, p < 0.001$ ) and a higher proportion of heavy verbs to all verbs ( $t(103.73) = 8.79, p < 0.001$ ).



**Figure 1.** The box plots of various features of agrammatism in patients with Broca's aphasia and healthy speakers across the cat and window stories

### 2. The information-theoretic analysis of language in Broca's aphasia

For the information-theoretic approach, we analyzed language at the sentence level rather than the individual level to increase the sensitivity of analyses.

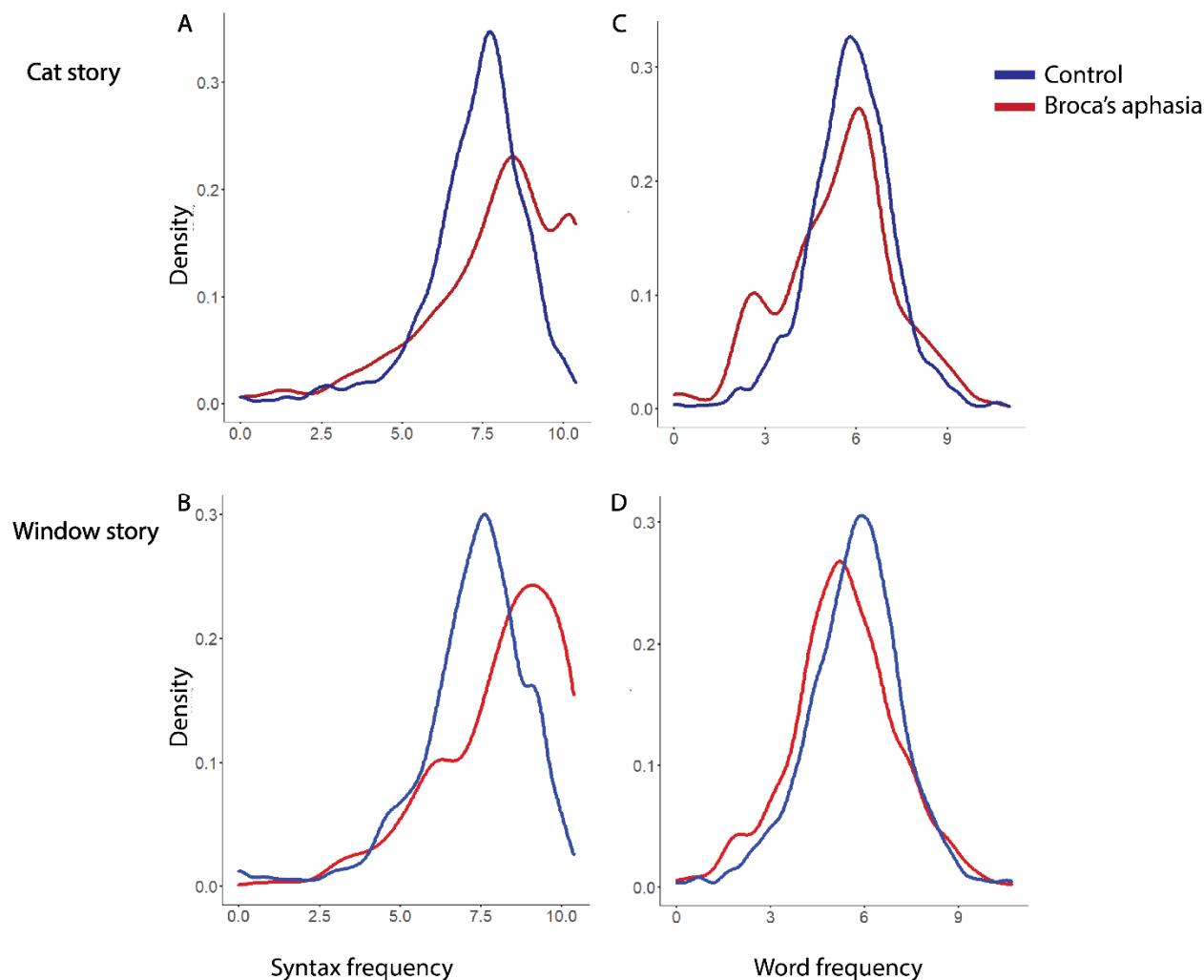
## 2.1. Syntax frequency

We used a linear mixed-effects model to examine the relationship between *syntax frequency* and two predictor variables: *group* and *sentence length*. The model included random intercepts for subjects to account for individual variability in the response variable and to address the correlation among the repeated measures of syntax frequency for each subject. The results of the linear mixed-effects model for the cat story showed significant effects of both *group* ( $\beta = 0.279$ ,  $t = 2.663$ ,  $p = 0.008$ ) and *sentence length* ( $\beta = -0.047$ ,  $t = -8.025$ ,  $p < 0.001$ ) on *syntax frequency* (Figure 2A). Specifically, the analysis revealed that patients with Broca's aphasia produce syntactic structures of higher frequency than healthy controls. We used the same model on the data obtained from the window story and found significant effects of both *group* ( $\beta = 0.466$ ,  $t = 3.904$ ,  $p < 0.001$ ) and *sentence length* ( $\beta = -0.058$ ,  $t = -8.025$ ,  $p < 0.001$ ) on *syntax frequency* (Figure 2B). The results suggest that regardless of the content of the picture across the two production tasks, patients with Broca's aphasia use simpler syntax than healthy controls, as evident from their higher average syntax frequency in a given sentence.

## 2.2. Word frequency

Similarly, we compared the average frequency of content words between the two groups. We used a linear mixed-effects model to examine the relationship between the average log frequency of content words and two predictor variables: *group* and *sentence length* with random intercepts for *subjects*. The results of the linear mixed-effects model for the cat story showed significant effects of both *group* ( $\beta = -0.254$ ,  $t = -2.897$ ,  $p = 0.004$ ) and *sentence length* ( $\beta = 0.003$ ,  $t = 5.140$ ,  $p < 0.001$ ) on *word frequency* (Figure 2C). We used the same model on the language samples obtained from the description of the window picture. Similarly, we found significant effects of both *group* ( $\beta = -0.182$ ,  $t = -1.984$ ,  $p = 0.048$ ) and *sentence length* ( $\beta = 0.003$ ,  $t = 4.906$ ,  $p < 0.001$ ) on *word frequency* (Figure 2D). The results suggest that regardless of the content of the picture across the two production tasks, patients with Broca's aphasia use lower-frequency words than healthy controls after controlling sentence length.

Here, we further show that the lower frequency of content words is associated with the canonical lexical features of agrammatism. We used a mixed effects model with random intercepts for *subjects* to predict word frequency from three predictor variables in the language of patients with Broca's aphasia. We found that the lower frequency of words in a sentence could be predicted from a higher proportion of content words to all words ( $\beta = -2.453$ ,  $t = -7.324$ ,  $p < 0.001$ ), a higher proportion of nouns over the sum of nouns plus verbs ( $\beta = -0.775$ ,  $t = -2.878$ ,  $p = 0.004$ ), and a higher proportion of heavy verbs to all verbs ( $\beta = -2.073$ ,  $t = -10.343$ ,  $p < 0.001$ ). These results suggest that the lexical profile of agrammatism reflects a strategy to lower word frequency, hence increasing sentence information.



**Figure 2.** The density plots of syntax frequency and word frequency obtained from the speech samples of healthy speakers and patients with Broca's aphasia describing the cat rescue and broken window pictures.

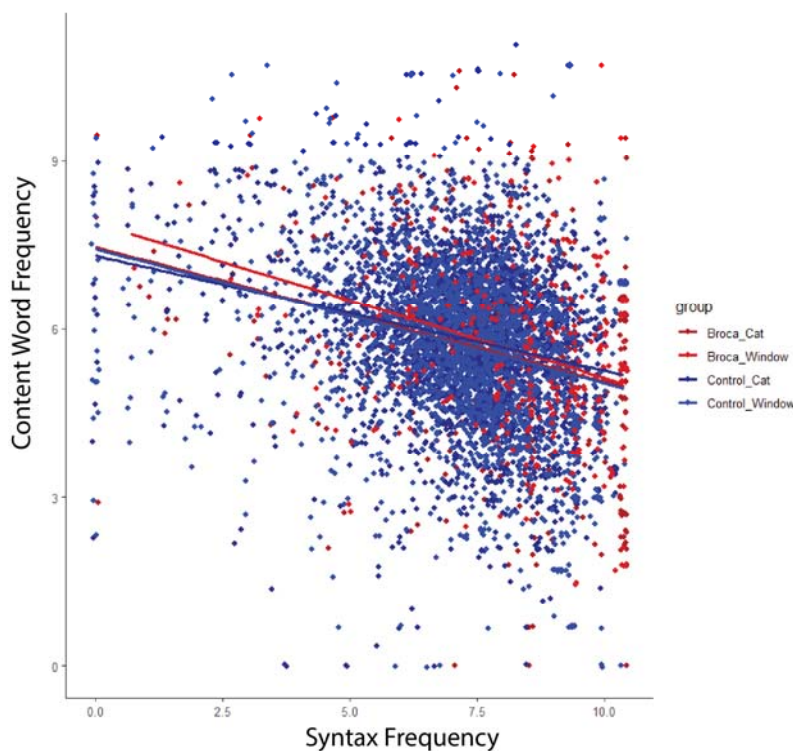
## 2.2. The syntax-lexicon trade-off

Here, we evaluated the trade-off between syntax frequency and word frequency at the sentence level. We used a linear mixed-effects model to examine the relationship between *syntax frequency* and two predictor variables, *word frequency* and *sentence length*, with random intercepts for *subjects* on the combined data from healthy individuals and patients with Broca's aphasia. The results of the linear mixed-effects model for the cat story showed significant effects of both *word frequency* ( $\beta = -0.024$ ,  $t = -13.115$ ,  $p < 0.001$ ) and *sentence length* ( $\beta = -0.005$ ,  $t = -8.768$ ,  $p < 0.001$ ) on *syntax frequency*. Similarly, for the window story, we found significant effects of both *word frequency* ( $\beta = -0.289$ ,  $t = -14.22$ ,  $p < 0.001$ ) and *sentence length* ( $\beta = -0.006$ ,  $t = -11.13$ ,  $p < 0.001$ ) on *syntax frequency*.

We further examined the potential interaction effect of task (cat story vs. window story) with *word frequency* on *syntax frequency*. After combining data from the two picture description tasks, we

found that the interaction term, *word frequency : task*, was not a significant predictor of *syntax frequency* ( $\beta = -0.004$ ,  $t = -1.629$ ,  $p = 0.103$ ), suggesting that the syntax-lexicon trade-off was independent of the content of the picture.

Further, we examined the potential interaction effect of *group* (Broca vs. healthy controls) with *word frequency* on *syntax frequency* and found that the interaction term, *word frequency : group*, was not a significant predictor of *syntax frequency* ( $\beta = 0.001$ ,  $t = -0.136$ ,  $p = 0.892$ ). This finding suggests that the trade-off is independent of the group because healthy controls and patients with Broca's aphasia show similar trade-offs (Figure 3).



**Figure 3.** The syntax-lexical trade-off in the language of patients with Broca's aphasia (in red) and healthy speakers (in blue) obtained from the description of cat and window stories.

## Discussion

In this work, we provided evidence synergistic with our information-theoretic proposal about agrammatism. Previously, we showed that the lexical profile of patients with nfvPPA is not a defect but rather a compensatory response to facilitate the transfer of their intended message. Specifically, patients with nfvPPA use more informative words to compensate for the loss of information that was supposed to be carried by syntactic structures.<sup>28,29</sup>

This study extended the generalizability of our previous findings on nfvPPA to patients with Broca's aphasia. This generalizability was expected, although not shown, given the overlapping neuroanatomical involvement between nfvPPA and Broca's aphasia. We first found that syntax frequency is an effective metric to capture structural deficits in patients with Broca's aphasia, as these patients produced high-frequency syntactic structures. We then found that patients with Broca's aphasia use lower-frequency content words compared to healthy individuals after controlling for sentence length. Producing sentences of lower-frequency words is likely achieved through the canonical lexical profile of agrammatism. The lexical profile that significantly correlated with a lower word frequency included higher proportions of content over function words, nouns over verbs, and heavy verbs over light verbs. Lastly, we found a syntax-lexicon trade-off in the language of patients with Broca's aphasia as well as healthy individuals. The collective findings suggest that individuals with nonfluent aphasia, including those with Broca's aphasia and nfvPPA, tend to select specific types of words to enhance the information content of their sentences. This compensation mechanism appears to be a response to their difficulty in generating long and complex sentences.

The results of this study were based on the description of two different pictures. Picture description tasks have the advantage of controlling the total amount of information needed to convey. We found similar results independent of the topic of the two pictures. Of note, the two pictures used in this study (cat rescue and broken window) were different from the picture we used in our studies on nfvPPA patients (picnic scene)<sup>40</sup> despite similar results. Future research is needed to investigate whether the information-theoretic features described in this work continue to exist in casual, everyday conversations. Furthermore, we conducted our analyses of language at the sentence level rather than the individual level. Previously, we showed that healthy speakers have the ability to shift between the use of complex syntax and complex words from sentence to sentence. For example, they might use simple words and complex syntax in one sentence while they adopt the opposite strategy in another. Therefore, analysis of language at the individual level that averages the syntax frequency and word frequency of all sentences would be insensitive to the normal variability used by healthy speakers.

From this work, it remains unclear what the origin of the syntactic impairment in patients with nonfluent aphasia is. For example, it is unclear whether it is the inability to make complex structures that result in shorter sentences or it is the inability to make long sentences that forces nonfluent patients to use simple syntax. Previous theories of compensation attributed the difficulty of making long, complex sentences to the cost of articulation.<sup>41</sup> However, we recently showed a high degree of similarity between the speech and writing of patients with nfvPPA<sup>28,42</sup> confirming previous studies showing similar patterns in patients with Broca's aphasia.<sup>43</sup> The similarity between writing and speaking excludes the cost of articulation to be the origin of the structural deficits in patients with nonfluent aphasia. Future work is required to delineate the functional locus of deficit as the bottleneck in producing syntactically complex structures.



## References

1. Damasio AR. Aphasia. *N Engl J Med*. 1992;326(8):531-539. doi:10.1056/NEJM199202203260806
2. Saffran EM, Berndt RS, Schwartz MF. The quantitative analysis of agrammatic production: Procedure and data. *Brain Lang*. 1989;37(3):440-479. doi:10.1016/0093-934X(89)90030-8
3. Goodglass H. Agrammatism in aphasiology. *Clin Neurosci N Y N*. 1997;4(2):51-56.
4. Goodglass H, Berko Gleason J. Agrammatism and Inflectional Morphology in English. *J Speech Hear Res*. 1960;3:257-267. doi:10.1044/jshr.0303.257
5. Zurif EB, Caramazza A, Myerson R. Grammatical judgments of agrammatic aphasics. *Neuropsychologia*. 1972;10(4):405-417. doi:10.1016/0028-3932(72)90003-6
6. Miceli G, Silveri MC, Villa G, Caramazza A. On the basis for the agrammatic's difficulty in producing main verbs. *Cortex J Devoted Study Nerv Syst Behav*. 1984;20(2):207-220. doi:10.1016/s0010-9452(84)80038-6
7. Goodglass H, Geschwind N. Language disorder (in aphasia). In: *Handbook of Perception, E. Carterette and M. Friedman*. Vol 7. Academic Press; 1976.
8. Kussmaul A. *Handbuch Der Speciellen Pathologie Und Therapie: Die Storungen Der Sprache Versuch Einer Pathologie Der Sprache*. Vol 12. FCW Vogel; 1877.
9. Deleuze JPF. *Histoire critique du magnétisme animal*. Chez Belin-Leprieur; 1819.
10. Bradley DC, Garrett MF, Zurif EB. Syntactic deficits in Broca's aphasia. In: *Biological Studies Of Mental Processes*. MIT Press; 1980.
11. Segalowitz SJ, Lane KC. Lexical access of function versus content words. *Brain Lang*. 2000;75(3):376-389. doi:10.1006/brln.2000.2361
12. Pulvermüller F. Brain mechanisms linking language and action. *Nat Rev Neurosci*. 2005;6(7):576-582. doi:10.1038/nrn1706
13. Zingeser LB, Berndt RS. Retrieval of nouns and verbs in agrammatism and anomia. *Brain Lang*. 1990;39(1):14-32. doi:10.1016/0093-934x(90)90002-x
14. Saffran EM, Schwartz MF, Marin OSM. The word order problem in agrammatism: II. Production. *Brain Lang*. 1980;10(2):263-280. doi:10.1016/0093-934X(80)90056-5
15. Lapointe SG. A theory of verb form use in the speech of agrammatic aphasics. *Brain Lang*. 1985;24(1):100-155. doi:10.1016/0093-934X(85)90100-2
16. Berndt R, Haendiges AN, Mitchum CC, Sandson J. Verb Retrieval in Aphasia. 2. Relationship to Sentence Processing. *Brain Lang*. 1997;56(1):107-137. doi:10.1006/brln.1997.1728
17. Shapiro LP, Levine BA. Verb processing during sentence comprehension in aphasia. *Brain Lang*. 1990;38(1):21-47. doi:10.1016/0093-934X(90)90100-U
18. Shapiro LP, Gordon B, Hack N, Killackey J. Verb-Argument Structure Processing in Complex Sentences in Broca's and Wernicke's aphasia. *Brain Lang*. 1993;45(3):423-447. doi:10.1006/brln.1993.1053
19. Kolk HHJ, Van Grunsven MJF, Keyser A. On Parallelism between Production and Comprehension in Agrammatism. In: *Agrammatism*. Elsevier; 1985:165-206. doi:10.1016/B978-0-12-402830-2.50011-0
20. Rezaei N, Ren B, Quimby M, Hochberg D, Dickerson B. Less is more in language production: Shorter sentences contain more informative words. Published online June 3, 2022:2022.06.02.22275938. doi:10.1101/2022.06.02.22275938
21. Rezaei Neguine, Mahowald Kyle, Ryskin Rachel, Dickerson Bradford, Gibson Edward. A syntax-lexicon trade-off in language production. *Proc Natl Acad Sci*. 2022;119(25):e2120203119. doi:10.1073/pnas.2120203119
22. Rezaei N, Mahowald K, Ryskin R, Dickerson B, Gibson E. Syntactic Rule Frequency as a Measure of Syntactic Complexity: Insights from Primary Progressive Aphasia, 34th Annual CUNY Conference on Human Sentence Processing. 34th Annual CUNY Conference on Human Sentence Processing. Accessed June 14, 2022. <https://www.cuny2021.io/2021/02/24/251/>
23. Rezaei N, Mahowald K, Ryskin R, Dickerson B, Gibson E. A syntax-lexicon trade-off in language production. *Proc Natl Acad Sci U S A*. 2022;119(25):e2120203119. doi:10.1073/pnas.2120203119

24. Bastiaanse R, Thompson CK. Verb and auxiliary movement in agrammatic Broca's aphasia. *Brain Lang*. 2003;84(2):286-305.
25. Centeno JG. Use of Verb Inflections in the Oral Expression of Agrammatic Spanish-Speaking Aphasics. *CUNY Acad Works*. Published online 1996:119.
26. Centeno J, Obler L. Agrammatic Verb Errors in Spanish Speakers and Their Normal Discourse Correlates. *Publ Res*. Published online January 1, 2001. [https://academicworks.cuny.edu/gc\\_pubs/55](https://academicworks.cuny.edu/gc_pubs/55)
27. Arslan S, Bamyacı E, Bastiaanse R. A characterization of verb use in Turkish agrammatic narrative speech. *Clin Linguist Phon*. 2016;30(6):449-469. doi:10.3109/02699206.2016.1144224
28. Rezaei N, Ren B, Quimby M, Hochberg D, Dickerson BC. Less is more in language production: An information-theoretic analysis of agrammatism in primary progressive aphasia. *Brain Commun Press*. Published online 2023.
29. Rezaei N, Michaelov J, Josephy-Hernandez S, et al. A computational approach for measuring sentence information via surprisal: theoretical implications in nonfluent primary progressive aphasia. *medRxiv*. Published online 2022:2022-11.
30. MacWhinney B, Fromm D, Forbes M, Holland A. AphasiaBank: Methods for studying discourse. *Aphasiology*. 2011;25(11):1286-1307. doi:10.1080/02687038.2011.589893
31. Grossman M, Irwin DJ. Primary Progressive Aphasia and Stroke Aphasia. *Contin Minneap Minn*. 2018;24(3, BEHAVIORAL NEUROLOGY AND PSYCHIATRY):745-767. doi:10.1212/CON.0000000000000618
32. Mesulam MM, Rogalski EJ, Wieneke C, et al. Primary progressive aphasia and the evolving neurology of the language network. *Nat Rev Neurol*. 2014;10(10):554-569. doi:10.1038/nrneurol.2014.159
33. Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *J Speech Hear Res*. 1993;36(2):338-350. doi:10.1044/jshr.3602.338
34. Menn L, Reilly KF, Hayashi M, Kamio A, Fujita I, Sasanuma S. The interaction of preserved pragmatics and impaired syntax in Japanese and English aphasic speech. *Brain Lang*. 1998;61(2):183-225. doi:10.1006/brln.1997.1838
35. Rezaei N, Wolff P, Price BH. Natural language processing in psychiatry: the promises and perils of a transformative approach. *Br J Psychiatry*. 2022;220(5):251-253. doi:10.1192/bjp.2021.188
36. Qi P, Zhang Y, Zhang Y, Bolton J, Manning CD. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Association for Computational Linguistics; 2020:101-108. doi:10.18653/v1/2020.acl-demos.14
37. Breedin SD, Saffran EM, Schwartz MF. Semantic factors in verb retrieval: an effect of complexity. *Brain Lang*. 1998;63(1):1-31. doi:10.1006/brln.1997.1923
38. Godfrey JJ, Holliman EC, McDaniel J. SWITCHBOARD: telephone speech corpus for research and development. In: *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing - Volume 1*. ICASSP'92. IEEE Computer Society; 1992:517-520.
39. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015;67:1-48. doi:10.18637/jss.v067.i01
40. Kertesz A, Kertesz A, Raven JC, PsychCorp (Firm). *WAB-R: Western Aphasia Battery-Revised*. PsychCorp; 2007.
41. Pick A. *Die Agrammatischen Sprachstörungen: Studien zur Psychologischen Grundlegung der Aphasielehre*. Springer-Verlag; 1913.
42. Josephy-Hernandez S, Rezaei N, Jones A, et al. Automated analysis of written language in the three variants of primary progressive aphasia. Published online July 25, 2022:2022.07.24.22277977. doi:10.1101/2022.07.24.22277977
43. Basso A, Taborelli A, Vignolo LA. Dissociated disorders of speaking and writing in aphasia. *J Neurol Neurosurg Psychiatry*. 1978;41(6):556-563. doi:10.1136/jnnp.41.6.556