PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES

A syntax–lexicon trade-off in language production

Neguine Rezaii[a,1], Kyle Mahowald[b], Rachel Ryskin[c], Bradford Dickerson[a], and Edward Gibson[d,1]

Edited by Susan Goldin-Meadow, University of Chicago, Chicago, IL; received November 4, 2021; accepted March 28, 2022

Spoken language production involves selecting and assembling words and syntactic structures to convey one's message. Here we probe this process by analyzing natural language productions of individuals with primary progressive aphasia (PPA) and healthy individuals. Based on prior neuropsychological observations, we hypothesize that patients who have difficulty producing complex syntax might choose semantically richer words to make their meaning clear, whereas patients with lexicosemantic deficits may choose more complex syntax. To evaluate this hypothesis, we first introduce a frequency-based method for characterizing the syntactic complexity of naturally produced utterances. We then show that lexical and syntactic complexity, as measured by their frequencies, are negatively correlated in a large (n = 79) PPA population. We then show that this syntax–lexicon trade-off is also present in the utterances of healthy speakers (n = 99) taking part in a picture description task, suggesting that it may be a general property of the process by which humans turn thoughts into speech.

syntax–lexicon trade-off | syntactic complexity | syntax frequency | word frequency | primary progressive aphasia

Significance

This work evaluates the long-debated relationship between word selection and syntax selection during language production in a sample of 231 individuals (79 patients with primary progressive aphasia [PPA] and 152 healthy speakers). We first provide a characterization of syntactic complexity based on the frequency of the syntactic rules, which allows for the direct comparison of syntactic and lexical complexities. We then provide converging evidence for a syntax–lexicon complexity trade-off in utterance planning in picture description which might reflect a basic property of language production.

During language production, speakers turn complex thoughts into a linear sequence of words (1–4). In the process of selecting how to package a message into an utterance, speakers choose both the words and the syntactic frames that determine the order and morphology of the lexical items (5, 6). The nature of the relationship between word selection and structural selection has remained a long-standing topic of discussion in the psycholinguistic literature (3, 5, 7–10). The primary concern of this research is to delineate how the selection of words and syntactic structures affect each other in the process of language production.

A classic finding from the neuropsychological literature is that patients with post-stroke aphasia often fall within two broad categories: those with difficulties retrieving and producing lexical items rich in semantic content and those with agrammatism who have difficulties producing various morphosyntactic elements (11–16). A similar pattern has been observed among patients with primary progressive aphasia (PPA), a clinical syndrome arising from neurodegenerative disease that primarily affects language (17, 18). Two of the variants, the logopenic variant (lvPPA) and the semantic variant (svPPA), are characterized by deficits in lexicosemantic processing, whereas the nonfluent variant (nfvPPA) is characterized by production of simplified syntactic structures and/or effortful speech (19). At first glance, this double dissociation may seem to be evidence for a modular system where lexical and structural representations are stored separately and speech planning unfolds independently.

However, more fine-grained analyses hint at further interactivity and a trade-off in the planning of lexical and structural elements. For instance, patients categorized as agrammatic use semantically richer or heavier verbs such as fly and fewer light verbs such as go, whereas patients with semantic deficits rely more on light verbs (20–25). Thus, even the lexical items produced by patients with agrammatism seem to be semantically richer. Similarly, it has been shown that svPPA patients produce more high-frequency lexical items than nfvPPA patients and controls (26). They also produce a numerically higher rate of syntactic embeddings in connected speech samples than controls (refs. 27, 28 but also see ref. 29), which has been interpreted as a potential compensation for lexicosemantic deficits (30).

This clinical observation of a trade-off in complexity between words and syntactic structures during utterance assembly has the potential to inform psycholinguistic models of language production (26). However, the trade-off has not yet been rigorously characterized, nor has it been observed in language production among healthy speakers. Both limitations likely stem from the difficulty in operationalizing complexity in a manner that is general enough to capture both lexicosemantic and syntactic complexity. While lexicosemantic complexity has often been approximated with word frequency (5, 31, 32), no directly comparable metric has yet been used for syntactic structures.

Author affiliations: [a]Frontotemporal Disorders Unit, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; [b]Department of Linguistics, The University of Texas at Austin, Austin, TX 78712; [c]Department of Cognitive & Information Sciences, University of California, Merced, CA 95343; and [d]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Author contributions: N.R., R.R., and E.G. designed research; N.R., K.M., R.R., B.D., and E.G. performed research; N.R., K.M., R.R., and E.G. analyzed data; and N.R., K.M., R.R., B.D., and E.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: nrezaii@mgh.harvard.edu or egibson@mit.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2120203119/-/DCSupplemental.

Published June 16, 2022.

In the present work, we probe a potential syntax–lexicon trade-off in language production as follows. First, we provide a frequency-based method for characterizing the syntactic complexity of utterances in naturalistic speech. We then test whether lexical and syntactic complexity, as measured by their frequencies, trade off in a large ($n = 79$) sample of individuals with PPA engaged in a picture description task, which is commonly used by clinicians to diagnose patients with PPA. We hypothesize that patients with svPPA and lvPPA who have lexicosemantic deficits choose more complex syntax to make their meaning clear. For instance, if these patients are unable to access the low-frequency noun phrase "a sailboat," they might convey their meaning by producing the more complex structure "a boat that is moved by the wind." Conversely, patients with nfvPPA who have difficulty producing complex syntax as measured here by syntax frequency might choose semantically richer words. Therefore, we expect a negative correlation between the frequency of syntactic structure and the frequency of words. In addition, as PPA patients have persistent deficits in producing either complex syntactic or lexical items in their utterances, we expect the average syntax frequency and average word frequency to be negatively correlated at the subject level.

Finally, we test the generalizability of the syntax–lexicon trade-off in a large ($n = 99$) sample of healthy speakers taking part in the same picture description task. Here we expect a negative correlation between syntax frequency and word frequency at the utterance level. Unlike PPA patients, this population has the ability to shift between the use of complex syntactic or complex lexical items, perhaps depending on what is more accessible in the moment or what might facilitate comprehension.

A Frequency-Based Characterization of Syntactic and Lexical Complexity

Previous research has used a variety of approaches to assess syntactic complexity in language processing. In clinical populations, subjective assessments of syntactic complexity are based on the use of rating scales (33) or the best judgment of clinicians (34). Other approaches measure various features of an utterance such as the constituent or utterance length (35–37), the ratio of function words to content words (27, 28, 37, 38), the ratio of nouns to verbs (27, 37–39), the number of embedded clauses (27, 29, 38, 39), the proportion of inflected verbs (27, 38, 40), the use of certain noncanonical structures (41, 42), the left-branching depth of an utterance (27, 38, 43), the stage of development of structures during language acquisition (44, 45), and more recently, the variability of the structure of clauses (46, 47). Many of these measures are correlated and have been shown to differ between patient and control populations (48) and have been used to systematically study differences in patients with cognitive impairment (49–52).

In the psycholinguistics literature, much work has focused on the real-time consequences of processing a sentence containing syntactic ambiguity, which is often resolved after a few words. A good predictor of the difficulty of resolving a temporary ambiguity is syntactic surprisal, which reflects how unpredictable a word (or part of speech [POS]) is given the context (53, 54). A second line of work in psycholinguistics has investigated unambiguous sentence structures in both language comprehension and production. The difficulty of parsing or generating an unambiguous sentence left to right is proportional to the degree of center-embedding in the structure, the number of incomplete phrase structure rules (55), the number of incomplete sentences (56), and the maximal or average dependency distances within a structure (57–60). Whereas Gibson proposed a distance measure in terms of the number of discourse referents between a head and dependent, a simplified version from the literature counts the words between a head and dependent (61).

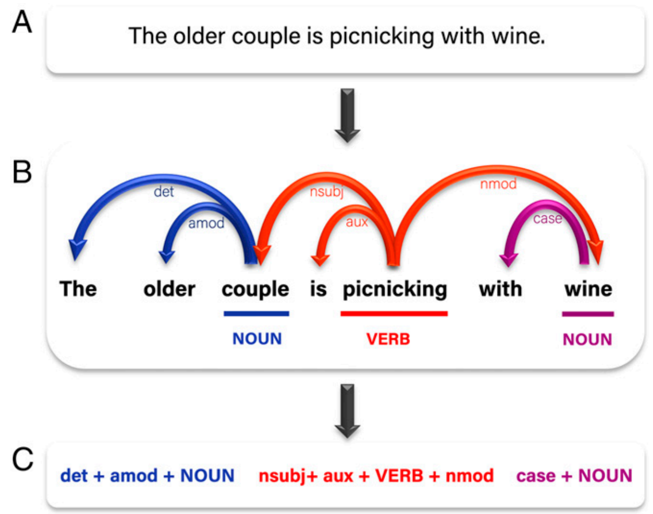

Fig. 1. Extracting syntactic rules of a dependency parse. (*A*) Sentences produced by participants were annotated and disfluencies removed. (*B*) Sentences were parsed using the Stanford Lexicalized Parser Package. Dependency heads, underlined, were identified by the parser. (*C*) The combinations of heads and dependents, preserving the order in which they appeared in the original sentence, were extracted as syntactic rules. This sentence resulted in three syntactic rules. The analyses we present here are by utterance (sentence). Thus, for this utterance, we would take the average syntax frequency of the three syntactic rules and the average word frequency for the content words in this sentence.

To explore a potential syntax–lexicon trade-off in language production, a syntactic complexity metric should be comparable to word frequency. We therefore propose the frequency of syntactic rules as a characterization of syntactic complexity. To measure the frequency of syntactic rules in spoken English, we use the Switchboard corpus (62). We first parsed the sentences in the corpus using the automated Stanford Lexicalized Parser (63) in order to extract headed syntactic rules (Fig. 1). 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. Thus, the syntactic rules extracted from the parse of the sentence "The older couple is picnicking with wine" in Fig. 1 are as follows.*

1) for the head NOUN "couple": det + amod + NOUN
2) for the head VERB "picnicking": nsubj + aux + VERB + nmod:with
3) for the head NOUN "wine": case + NOUN

These *n*-ary rules correspond to all the dependents of a word within a sentence. So the verb "picnicking" has three dependents: the subject (nsubj) headed by "couple," the auxiliary verb (aux) "is," and the modifier "wine" (nmod:with). The *n*-ary rules represent word collocations in language, corresponding to argument structures and common modifier structures (64, 65). In particular, the *n*-ary dependency formalism would have one rule for all dependents of a verb like "put," which almost always cooccurs with a subject noun, an object noun, and a goal noun (e.g., as in "Mary put the bucket in the closet"). Furthermore, as syntax frequency has not been previously evaluated in the

*The definition of common abbreviations of the dependency relations are provided in *SI Appendix*, Table S1.

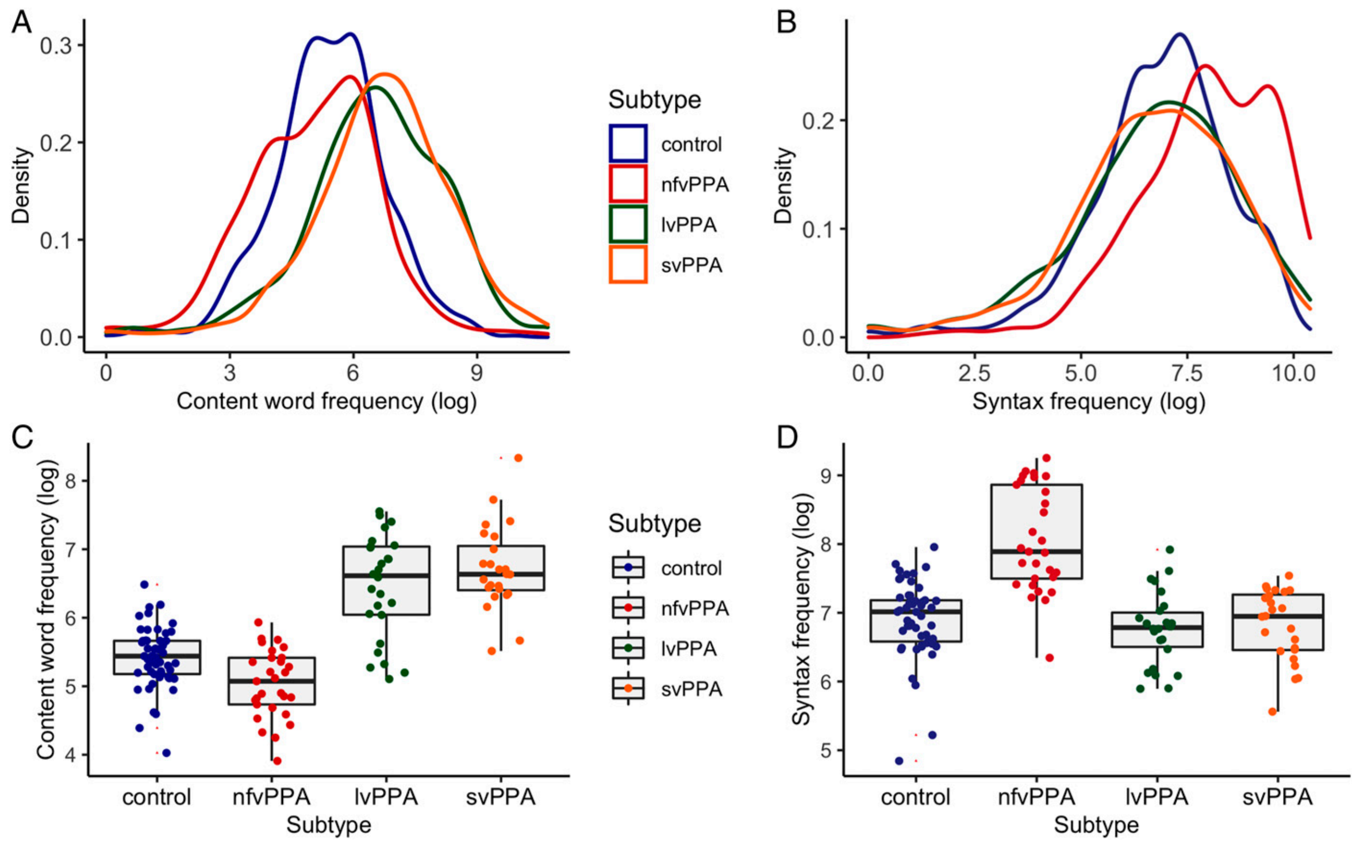

Fig. 2. (*A*) The density graphs of content word frequency and (*B*) syntax frequency of utterances. (*C*) The box plots of content word frequency per individual in each group. (*D*) The box plots of syntax frequency per individual in each group. The box plots show the median, 25th, and 75th percentiles of the data. Error bars represent 95% CI.

aphasia literature, we test its clinical relevance by examining its accuracy at classifying patients with nfvPPA.

For the same example sentence in Fig. 1, lexical complexity was measured as the average word frequency across the four content words (older, couple, picnicking, and wine). Mean frequency of content words (adjectives, adverbs, nouns, and verbs), as opposed to all words, was chosen as the primary measure because the frequency of nouns and verbs are the most commonly used in assessments of PPA patients (28, 66, 67). We present equivalent analyses using the mean frequency of all words in footnotes.

Results

Study 1: Word Frequency, Syntax Frequency, and the Syntax–Lexicon Trade-Off in the Utterances of Patients with PPA and Healthy Controls.

Word frequency. Here, we compare the average content word frequency at the utterance level across the four groups. We fit a mixed effects model with random effects for subjects to predict log word frequency with patient subtype (treatment-coded with healthy controls as reference level) and sentence length as predictors. We include random intercepts for subjects but no random slopes since models with random slopes failed to converge. The effect of patient subtype was significant by a likelihood ratio test, comparing the full model to a null model with random effects held constant but without the fixed effect for patient subtype [chisq (3) = 118.5, $P < 0.001$].

Replicating previous findings in the literature on language production in PPA, patients with lvPPA and svPPA produced utterances with higher-frequency content words when compared with healthy controls [$\beta = 1.134$, SE = 0.130, $t = 8.744$, $P < 0.001$ and $\beta = 1.350$, SE = 0.136, $t = 9.953$, $P < 0.001$, respectively). However, we found that patients with nfvPPA used lower-frequency words ($\beta = -0.282$, SE = 0.135, $t = -2.085$, $P = 0.039$) when compared with healthy controls.[†] Fig. 2*A* shows the density graphs of content word frequency of utterances across the four groups.

Syntax frequency.

Comparing syntax frequency of PPA variants with healthy controls. Fig. 3 illustrates the proportions of use of the 20 most common syntactic rules from the Switchboard corpus by patients with PPA and healthy controls. The last set of bars shows the proportion of use of syntactic rules for all other lower-frequency syntactic rules illustrating that nfvPPA patients produce more of the common syntactic rule types and fewer of the low-frequency syntactic rule types when compared to healthy controls and other PPA variants.

A mixed effects model predicting log syntax frequency was fit with patient subtype (treatment-coded with healthy controls as reference level) and sentence length as predictors and random intercepts for subjects (with no random slopes since they led to convergence failures). The effect of patient subtype was significant by a likelihood ratio test, comparing the full model to a null model with random effects held constant but without the fixed effect for patient subtype [chisq (3) = 60.364, $P < 0.001$]. Individuals with nfvPPA were more likely to use higher-frequency syntax rules when compared to healthy controls [$\beta = 0.704$, SE = 0.134, $t = 5.237$, $P < 0.001$], while patients with lvPPA and

[†]Analysis of all-word frequency, as opposed to just content word frequency, showed a similar pattern when we conducted the same analysis. Patients with lvPPA and svPPA produced utterances with higher-frequency content words when compared with healthy controls [$\beta = 0.404$, SE = 0.100, t = 4.022, p < 0.001 and $\beta = 0.557$, SE = 0.105, t = 5.320, p < 0.001, respectively], while patients with nfvPPA used lower-frequency words [$\beta = -0.336$, SE = 0.103, t = −3.255, p = 0.001] when compared with healthy controls.

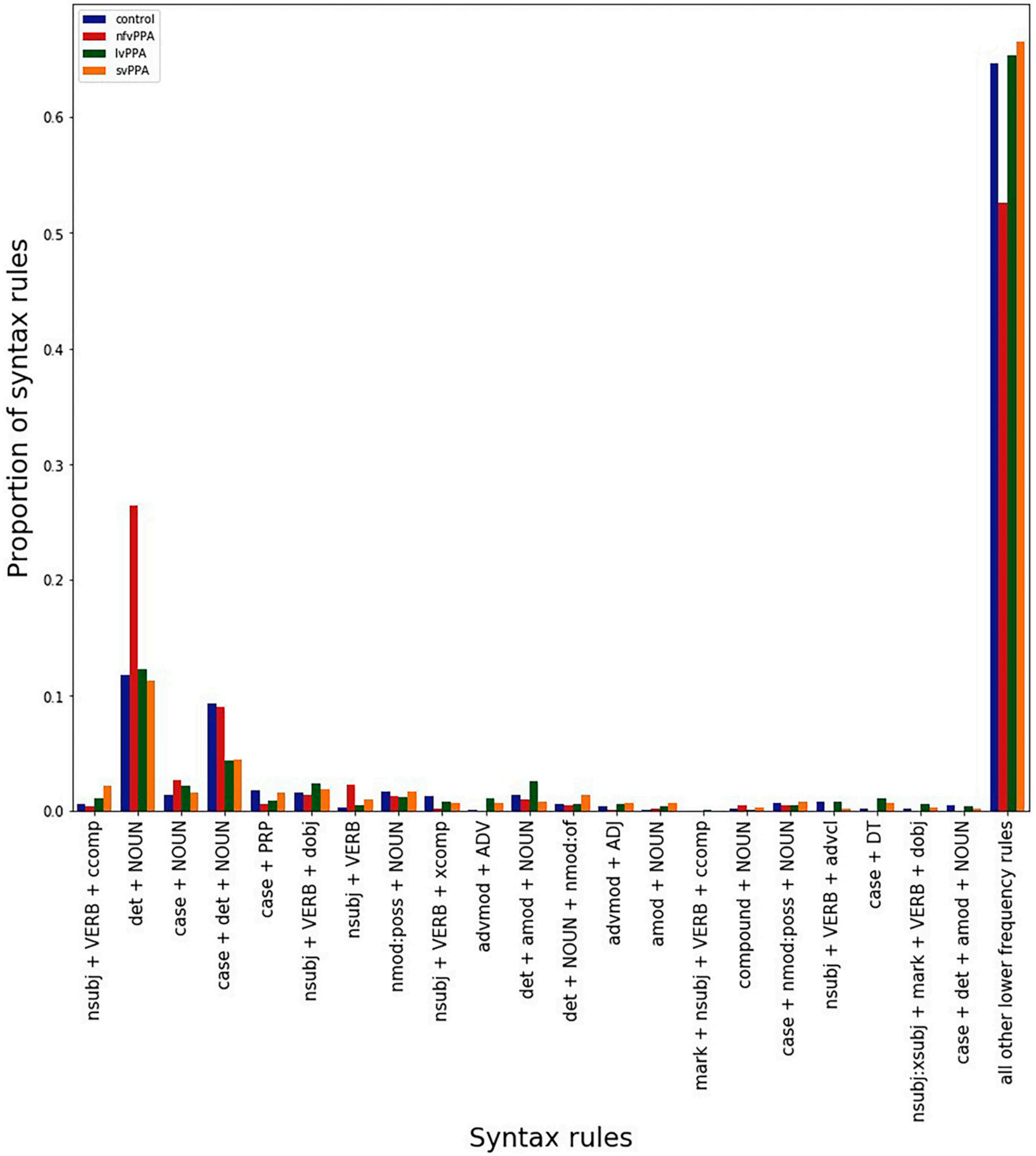

Fig. 3. The proportions of use of the 20 most common syntactic rules of the Switchboard by PPA patients and healthy controls. The last set of bars shows all other lower-frequency rules.

svPPA produced sentences with lower syntax frequency when compared to healthy controls [$\beta = -0.401$, SE = 0.124, $t = -3.238$, $P = 0.002$ and $\beta = -0.415$, SE = 0.129, $t = -3.213$, $P = 0.002$, respectively]. Fig. 2*B* shows the density plots of syntax frequencies of utterances in each group.

Comparing syntax frequency of PPA variants with healthy controls in language subsamples with equal sentence length distributions. As patients with nfvPPA tend to produce sparse speech when compared with other groups, it is possible that the production of syntactic rules with higher frequency results from the production of shorter sentences. In the previous section, we used sentence length as a predictor in the regressions. However, it is still possible that there is a nonlinear relationship between sentence length and syntax frequency which may drive the observed relationship and confound the results of the regression analyses. Consequently, we ran a more conservative version of the analysis where we randomly subsampled sentences from the language outputs of each of the four groups so that all groups would

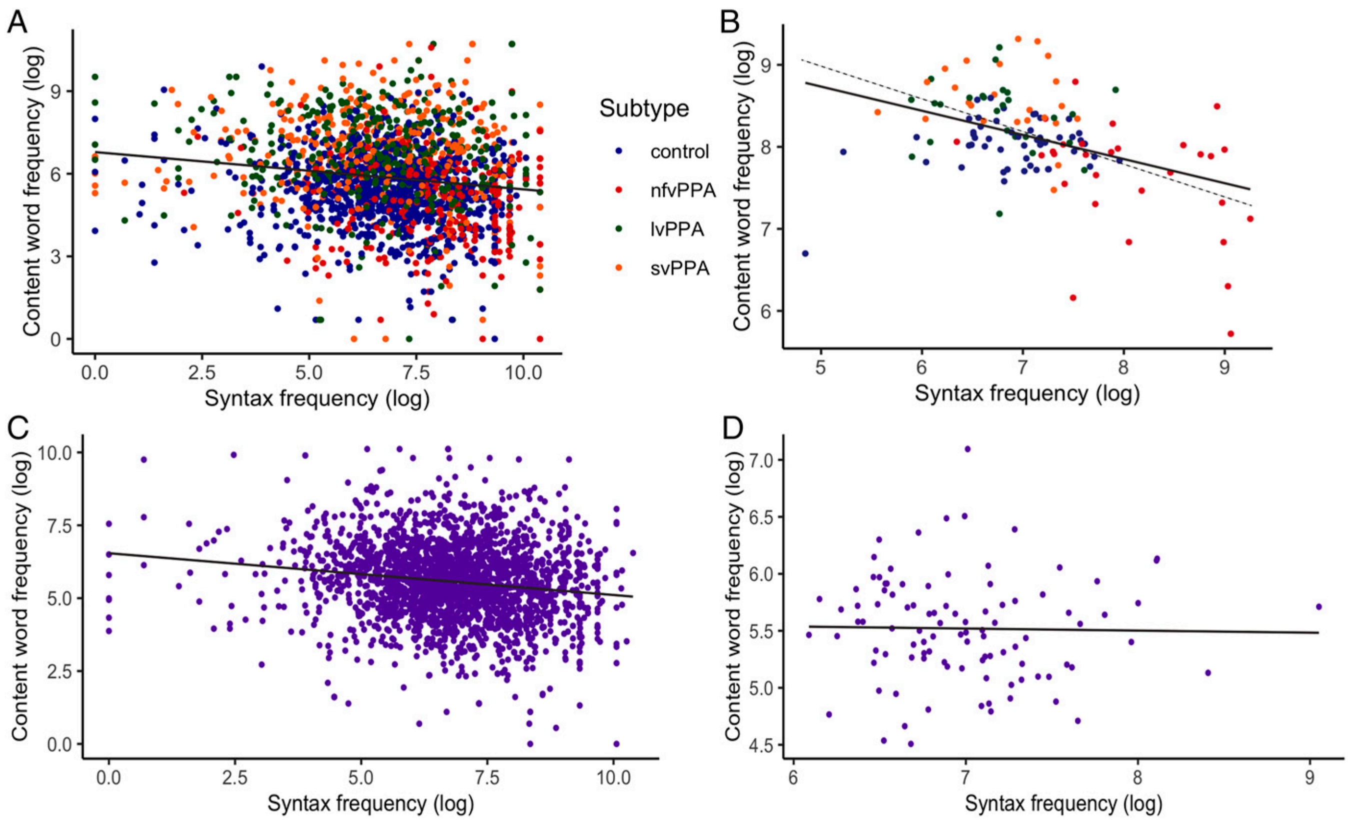

Fig. 4. Syntax–lexicon scatterplots. (*A*) Utterance level in PPA patients (n = 79) and healthy controls (n = 53). (*B*) Subject level in PPA patients (dashed line) and healthy controls (solid line represents all data points). (*C*) Utterance level in healthy individuals (n = 99). (*D*) Subject level in healthy individuals.

have the same distributions of sentence length. In this subsample, the linear regression analysis of syntax frequency with subtype as a dummy coded predictor (with the control group as the reference level) and the same random effect structure as described in *Comparing syntax frequency of PPA variants with healthy controls* continued to show that patients with nfvPPA produce syntactic rules with higher frequency (β = 0.654, SE = 0.183, t = 3.58, $P < 0.001$) and patients with lvPPA and svPPA produce sentences with lower syntax frequency when compared to healthy controls (β = −0.409, SE = 0.184, t = −2.23, $P < 0.05$ and β = −0.427, SE = 0.189, t = −2.286, $P < 0.05$, respectively).

Syntax-lexicon trade-off in PPA patients and healthy controls. At the utterance level, fitting a mixed effects model (68) that predicts content word frequency, with syntax frequency and sentence length as predictors and random intercepts for subject (with a random slope for syntax frequency but no correlations and no random slopes for length, in order to aid convergence), we found a main effect of syntax frequency (β = −0.09, SE = 0.020, t = −4.63, $P < 0.001$) but not sentence length (β = 0.013, SE = 0.008, t = 1.1.65, $P = 0.09$) (Fig. 4*A*). We do not include patient subtype as a predictor in this analysis, since we are investigating the utterance-level correlation across the patient population and do not presuppose a diagnosis.

At the subject level (n = 132), syntax and word frequency were inversely related to each other (r = −0.38, $P < 0.001$) (Fig. 4*B*). This inverse correlation was stronger when only PPA patients were analyzed (r = −0.65, $P < 0.001$) (the dashed line in Fig. 4*B*).[‡] Again, this analysis does not presuppose PPA subtypes.

Classifying nfvPPA versus other groups using syntax frequency. As syntax frequency has not been previously used in the aphasia literature, here we examine the measure's utility in differentiating patients with nfvPPA from healthy controls, lvPPA, and svPPA by using binary logistic regression to compare the classification accuracy of syntactic frequency and other measures of syntactic complexity commonly used in aphasia: sentence length, the maximum length of an incomplete dependency in an utterance (43), bigram entropy over POS tags (48), average dependency distance of a sentence (57), and the ratio of nouns to verbs. Table 1 shows the accuracy of the leave-one-out cross-validation of the classification using the average score of each measure of syntactic complexity per individual. We see that syntax frequency on its own achieves high classification accuracy (89%; baseline = 78%). All models that include both sentence length and lexical frequency and a syntactic complexity metric have high classification accuracy (92% or greater), independent of the choice of syntactic complexity metric. *SI Appendix*, Table S2, shows the accuracy of classification of all variants using multinomial logistic regressions, in which we use a four-way classification to categorize a particular individual as falling into the control group or one of the three patient groups. We observed broadly similar patterns to that of the logistic regression. All models that include both sentence length and lexical frequency and a syntactic complexity metric have high multinomial classification accuracy (66% or greater), independent of the choice of syntactic complexity metric.

[‡]We repeated similar analyses for the average frequency of all words in an utterance. At the utterance level, fitting a maximal mixed effects model with random effects for subject that predicts all-word frequency, with syntax frequency and sentence length as predictors, we found a main effect of syntax frequency (β = −0.035, SE = 0.012, t = −2.83, p < 0.01) but not sentence length (β = −0.001, SE = 0.005, t = −0.31, p = 0.76). At the subject level (n = 132), syntax and all-word frequency were inversely related to each other (r = −0.41, p < 0.001). This inverse correlation was stronger when only PPA patients were analyzed (r = −0.58, p < 0.001).

Table 1. Binary logistic regression with leave-one-out cross-validation to classify nfvPPA versus other groups (lvPPA, svPPA, and control)

Model	Classification accuracy, R^2 of the model	Classification accuracy, R^2 of the model adding sentence length as a predictor	Classification accuracy, R^2 of the model adding sentence length and word frequency as predictors
nfvPPA ~ sentence length	80%, 0.41	—	93%, 0.77
nfvPPA ~ maximum incomplete dependencies	85%, 0.43	83%, 0.44	92%, 0.78
nfvPPA ~ POS entropy	86%, 0.62	91%, 0.76	93%, 0.83
nfvPPA ~ dependency distance	87%, 0.46	85%, 0.48	93%, 0.78
nfvPPA ~ noun/(noun + verb)	79%, 0.16	87%, 0.57	94%, 0.80
nfvPPA~ syntax frequency	89%, 0.61	90%, 0.67	92%, 0.80

Nagelkerke's R^2 is reported for each model. Chance on the classification task is 78%.

We also obtained test scores on the Northwestern anagram test (NAT), a measure of syntactic complexity used in multiple studies of PPA, from 43 patients. In a binary logistic regression model that predicted the nfvPPA status from both NAT and syntax frequency as predictors, we found an effect for syntax frequency [$\beta = 5.072$ (1.64), $P = 0.002$] but not NAT [$\beta = -0.003$ (0.023), $P = 0.911$]. The inability of the NAT to classify PPA subtypes replicates results recently reported in a larger independent sample of patients (69).

Study 2: Syntax–Lexicon Trade-Off in Healthy Individuals.

The syntax–lexicon trade-off in healthy individuals describing the picnic picture. The 99 healthy individuals who completed the picture description task produced a total of 2,331 utterances. Fitting a maximal mixed effects model with random intercepts for subject and random slopes by subject for syntax frequency (but not for sentence length in order to make the models converge) that predicts content word frequency from syntax frequency and sentence length as predictors, we found a main effect of syntax frequency ($\beta = -0.146$, SE = 0.024, t = −5.98, $P < 0.001$) but not sentence length ($\beta = -0.004$, SE = 0.005, t = −0.65, $P = 0.51$) (Fig. 4*C*), indicating that utterances with higher syntax frequency contained lower-frequency content words and vice versa. We also examined the syntax–lexicon trade-off at the subject level. During the planning of each utterance, healthy speakers may select either low-frequency syntactic rules or low-frequency content words. Therefore, the correlation between syntax frequency and word frequency at the subject level was small (r = −0.02) and not significantly different from 0 ($P = 0.843$).[§]

Combining all the utterance level data from studies 1 and 2 (as shown in Fig. 4 *A* and *C*), we ran a regression predicting word frequency from syntax frequency, whether the subject was a patient, and the interaction of patient status and syntax frequency. We included a random intercept for subject but no random slopes to aid convergence. We found that whether the subject was a patient as opposed to healthy individual did not significantly moderate the effect of syntax frequency on word frequency, as evidenced by a nonsignificant interaction term (for the interaction term, $\beta = -0.018$, $P = 0.516$).

Combining all the subject-level data from studies 1 and 2 (as shown in Fig. 4 *B* and *D*), we ran a standard linear regression predicting content word frequency, based on patient status (i.e., whether the participant is a patient or a healthy control, binarized as a dummy coded variable), with syntax frequency, normalized sentence length, and the interaction between syntax frequency and patient status. We vary whether the patient group or control group is the baseline and use normalized sentence length so that the main effect of syntax frequency can be interpreted as the effect of syntax frequency for a sentence of average length, in the baseline group. With the patient group treated as the baseline, the main effect of syntax frequency was a strong predictor of word frequency ($\beta = -0.787$, $P < 0.001$), suggesting that within patients, word frequency is robustly related to syntax frequency. Treating healthy individuals as a baseline, we did not find a significant main effect for syntax frequency to predict word frequency ($\beta = -0.010$, $P = 0.918$).

The syntax–lexicon trade-off at the utterance level in a subset of healthy individuals describing their work. To examine the generalizability of our findings, we evaluated the syntax–lexicon trade-off in the language samples from a subset of MTurk healthy controls ($n = 26$) who described their work. Fitting a mixed effects model, with a random intercept for subject and a random slope for syntax frequency (additional random slopes prevented convergence) that predicts the average content word frequency of an utterance using the syntax frequency and length of the utterance as predictors, we found a numerical but nonsignificant trend for the predicted effect of syntax frequency ($\beta = -0.102$, SE = 0.067, t = −1.53, $P = 0.14$) and no clear effect of sentence length ($\beta = 0.0004$, SE = 0.113, t = 0.033, $P = 0.97$).[¶]

The syntax–lexicon trade-off at the utterance level in the Switchboard corpus. Here, we tested the syntax–lexicon trade-off in the utterances within the Switchboard corpus. We automatically removed disfluencies, such as "um" and "uh." Using a linear regression model to predict an utterance's content word frequency from its syntax frequency and sentence length, we found a main effect of syntax frequency ($\beta = -0.039$, SE = 0.004, t = −10.13, $P < 0.001$) and sentence length ($\beta = -0.025$, SE = 0.001, t = −25.88, $P < 0.001$). Whereas for the majority of our analyses, the results are qualitatively similar when we use content word frequency or all-word

[§]We further repeated the analyses using all-word frequency instead of content word frequency. Fitting a maximal mixed effects model with random effects for subject that predicts all-word frequency from syntax frequency and sentence length as predictors, we found a main effect of syntax frequency ($\beta = -0.057$, SE = 0.014, t = −4.186, p < 0.001) and sentence length ($\beta = -0.009$, SE = 0.003, t = −2.78, p < 0.01). At the subject level, the correlation between syntax frequency and all-word frequency was small (r = 0.06) and not significantly different from 0 (p = 0.540).

[¶]We repeated this analysis using all-word frequency. Fitting a maximal mixed effects model with random effects for subject that predicts the average content word frequency of an utterance using from the syntax frequency and sentence length of the utterance as predictors, we did not find a main effect of syntax frequency ($\beta = 0.004$, SE = 0.04, t = 0.115, p = 0.912) nor sentence length ($\beta = -0.008$, SE = 0.008, t = −0.992, p = 0.322).

frequency, in this case they differ: using a linear regression model to predict an utterance's all-word frequency from its syntax frequency and sentence length, we found a main effect of syntax frequency ($\beta = 0.009$, $SE = 0.002$, $t = 3.809$, $P < 0.001$) in the opposite direction of the predicted trade-off and a negative effect of sentence length ($\beta = -0.019$, $SE = 0.001$, $t = -30.494$, $P < 0.001$). Because all-word frequency is heavily dependent on the ratio of function words to content words, it measures something fundamentally different from content word frequency, and we leave it to future work to understand why this difference emerges in the Switchboard analysis but not in the analyses of the picture description task. In *Discussion*, we speculate about the potential differences between tasks that might give rise to the observed analysis differences.

Discussion

In this work, we introduced a frequency-based measure of syntactic complexity that can be used to characterize any language sample. Using this metric, we provide converging evidence for a trade-off between syntactic and lexical complexity in language productions for patients with PPA. Individuals with nfvPPA generate higher-frequency syntax and lower-frequency words. Conversely, patients with lvPPA and svPPA produce utterances with higher-frequency words and lower-frequency syntax. Impairment in the use of complex and thus lower-frequency syntax in patients with nfvPPA is consistent with previous reports of lower syntactic variability in the language production of these patients. The use of higher-frequency syntactic structures indicates that nfvPPA patients may have difficulty accessing a wide range of syntax rules. The negative correlation between syntax and lexical frequency held at both the utterance level and the subject level, consistent with the notion that the patients have a persistent deficit in either syntactic or lexicosemantic processing which in turn influences each utterance they produce. Crucially, we applied the same analysis to speech samples from a large number of healthy individuals. Unlike patients with PPA, healthy speakers do not have a persistent deficit that consistently constrains their production of either low-frequency syntax or low-frequency words for all utterances. Rather, healthy individuals vary this balance from utterance to utterance, with utterances that contain low-frequency words having higher-frequency syntax and vice versa. Taken together, these findings suggest that healthy individuals are able to dynamically produce language that balances syntactic and lexical complexity, while people with PPA are impaired in one of these dimensions yet still capable along the other dimension.

From a clinical perspective, syntax frequency—which performs as well in classifying patients as a number of well-attested measures of syntactic complexity like POS entropy and noun/verb ratio—may be a promising tool for discriminating patients with nfvPPA from other PPA subtypes. Unlike most other measures of syntactic complexity in the PPA literature, such as noncanonicality of sentence structures, which are applicable to a limited range of syntactic forms, syntax frequency provides a complexity metric which can be computed for any utterance. Additionally, the measurement of syntactic frequency can be fully automated resulting in analyses that are less labor-intensive to gather and less vulnerable to potential human biases.

The syntax–lexicon trade-off fits well with multiple current accounts in the language production literature. In what follows, we elaborate on a few possible explanations of this trade-off in terms of 1) communicative efficiency pressures and 2) access to lexical or structural elements. The data presented in this work do not allow us to distinguish among these explanations, which are not necessarily mutually exclusive, but we detail them here in the hopes that this will spur further empirical work. One intuitive explanation of why we observe this trade-off is that it reflects joint pressures to produce sentences that are unambiguous yet not overly verbose or redundant (70). As the speaker formulates their thoughts into utterances, they wish to be clearly understood, so for example, saying "thing" to describe a sailboat would not suffice. If they select a lower-frequency word such as "sailboat," the structure can be simple as the necessary information has already been conveyed. If they select a higher-frequency word, a more complex structure may be needed to convey the same information, as in "a boat that is moved by the wind." Note that this is not meant to imply that lexical items are chosen first. Combining both low-frequency words and structures may often lead to verbosity or redundancy.

Along these lines, the syntax–lexicon trade-off can be explained in the context of an information-theoretic account of language production which proposes a single pressure on the production system, that of distributing information evenly across the utterance (70, 71). The uniform information density (UID) hypothesis states that when multiple formulations are possible, the speaker will choose the one for which information is closest to uniformly distributed. This pressure avoids peaks in information density which could overwhelm the capacity of the channel and lead to information loss or troughs which could lead to miscommunication. Under this hypothesis, listeners are more likely to experience comprehension difficulty during a moment of high unpredictability [surprisal (53, 54, 72)], and speakers plan their utterances in such a way as to avoid causing such difficulty. For example, speakers use shorter word forms when they are more predictable (73, 74). At the syntactic level, speakers are more likely to include an optional complementizer, "that," when the upcoming clause is less predictable. The presence of the complementizer serves as a cue to the nature of the upcoming clause, thereby reducing its surprisal (ref. 70 but see ref. 75). From a UID perspective, lexical and syntactic elements are arranged in such a way as to efficiently convey the intended information. When an upcoming word will be high in surprisal such as a low-frequency word, the speaker may distribute the information more uniformly by using a simple structure such as a high-frequency one. Similarly, when the structure is complex, simpler words may be chosen to avoid peaks of information that would exceed the channel's capacity.

Alternatively, according to a more mechanistic account of language production, some lexical and structural elements are easier to access—more available—and are therefore more likely to be produced (76–79). Ease of access is influenced by a number of factors, including perceptual salience (80), the topic of conversation, idiosyncratic interest, conceptual simplicity, lifelong exposure, and recent experience (81–85). Lexical selection in this framework is typically modeled as spreading activation through a connectionist network from a meaning layer down to a lexical and finally a phonetic layer (5, 86, 87). The process of producing a word involves both the activation of the single most active representation and the inhibition of coactive competitors which is thought to engage domain-general cognitive control (88, 89). Similarly, neural networks have been used to model effects of syntactic persistence in production (90). Syntactic selection in these models involves the spread of activation from the meaning layer down to the syntactic state of the network and ultimately to the output layer (23, 91–93). Thus, via a similar mechanism of selection and competitor inhibition, it is plausible that cognitive control—possibly language-specific (94)—plays a role in syntactic selection as well. We can speculate that on this view, the planning of an utterance requires shared cognitive control resources to be allocated both for the selection of a lexical item and for that of a structure. Less cognitive control is needed to select an item that is higher in availability. Thus, if a lexical item is low in availability, such as a low-frequency word, its selection will require a large amount of cognitive control resources, and little will be available for the selection of a structure. In this scenario, only the most available structures—typically high-frequency structures—will be chosen. On the other hand, if a structure is highly available, its selection will require fewer cognitive control resources, and consequently, more will be available for the selection of low-frequency words. Further empirical work is needed to determine whether lexical items and structures do in fact compete for shared resources in the process of utterance formulation.

Interestingly, although the trade-off between syntax and lexical frequency was robust in the picture description task, the evidence in the work description task and in the Switchboard corpus was mixed (the trade-off holds for analyses including just content words but not all words). This may be explained, at least in part, by the smaller sample size of the work description task and the additional noise present in the Switchboard dataset (many false starts, such as "I've" or "I'm," remain as full sentences; many discourse fillers, such as "you know," also remain), relative to the picture description utterances which were filtered of discourse markers, false starts, and other disfluencies (see protocol in *SI Appendix*, supplementary material 3). More intriguingly, the cognitive demands and linguistic content present in the different types of speech samples may play a role. Much of the speech sampled in the Switchboard corpus consists of casual conversation between people. Similarly, the work description task involves generating narratives from the speaker's own personal context, which is easily accessible. Both of these speech samples may consist of language that is more familiar, internally elicited, and likely less demanding. It may be that the cognitive and linguistic demands imposed by on-the-fly description of an unfamiliar image are sufficiently demanding to reveal trade-offs in production processes. Thus, in order to investigate this lexicosyntactic trade-off in language production, we may need to focus on tasks like picture description. Note that many scenarios in which people use language in everyday life are in fact more similar to the unfamiliar picture description task than to the casual conversation about one's own life, for instance, a professor lecturing to a classroom, a person giving directions to someone else from a map, etc.

Overall, these studies suggest that language production involves a trade-off in complexity between lexical and syntactic elements of utterances. For healthy speakers, this appears to reflect an implicit utterance-by-utterance decision process, at least for tasks like picture description. We speculate that this trade-off results from the language system's need to manage the costs of producing informative language in light of information processing constraints both on the side of speakers and listeners. Further experimental work is needed to elucidate the underlying mechanisms driving this trade-off. In addition, use of a frequency-based specification of syntactic complexity not only allows for the evaluation of the syntax–lexicon trade-off but also provides a metric that is of clinical value for differentiating patients with agrammatism in PPA from other variants. Future work is needed to further explore the utility of syntax frequency in other languages as well as other research domains such as poststroke agrammatism or the development of syntax during normal language acquisition.

Methods

Participants.

Study 1. The first study includes 79 PPA patients and 53 healthy controls with details as follows and as shown in Table 2.

PPA patients. Seventy-nine patients with PPA were recruited from an ongoing longitudinal study being conducted in the PPA Program in the Frontotemporal Disorders Unit of Massachusetts General Hospital. Baseline clinical and language assessments were used to characterize patients and to subtype them into nfvPPA (n = 29), svPPA (n = 24), and lvPPA (n = 26). The participants of this study underwent a comprehensive clinical evaluation as previously described (95). The evaluation included a structure interview by a neuropsychiatrist and a neurological examination as well as speech and language assessment by a speech-language pathologist. The protocol for the participants of this study included the National Alzheimer's Coordinating Center Uniform Data Set measures (using version 2.0 for 65 of the assessments and version 3.0 for the remaining 5), including clinical dementia rating (CDR) scale supplementary language box ratings (96).

Healthy controls. Fifty-three healthy controls who were age matched with the PPA patients were included in the first part of this study. Twenty of the healthy controls were enrolled through the Speech and Feeding Disorders Laboratory at the Massachusetts General Hospital (MGH) Institute of Health Professions. These participants passed a cognitive screen, were native English speakers, and had no history of neurologic injury or developmental speech/language disorders. The remaining 33 healthy controls were recruited through Amazon's Mechanical Turk (MTurk). MTurk participants filled out the short and validated version of the everyday cognition test with 12 items, an informant-rated questionnaire designed to detect cognitive and functional decline (97). Only language samples from participants who were native English speakers, with no self-reported history of brain injury or speech/language disorder, either developmental or acquired, were included in the analyses. The healthy controls recruited from the clinic and MTurk did not differ in terms of age, gender, handedness, and years of education.

Study 2. The second study examined the syntax–lexicon trade-off in 99 healthy participants.

Table 2. Demographic and clinical characteristics of study 1 participants

	nfvPPA (n = 29)	lvPPA (n = 26)	svPPA (n = 24)	Healthy controls (n = 53)
Age at testing, years (SD)	68.91 (9.36)	69.18 (6.49)	65.61 (8.23)	65.11 (6.86)
Gender, % female	51.7	38.5	58.3	61.54
Education, years (SD)	16.07 (2.88)	16.38 (2.56)	16.36 (1.81)	15.29 (1.49)
Handedness, % left	3.5	11.5	12.5	15.4
CDR, global language (SD)	0.78 (0.49)	0.70 (0.35)	0.85 (0.46)	—
PASS, sum of boxes (SD)	4.81 (2.62)	5.04 (2.52)	4.68 (2.03)	—

There were no significant differences between healthy controls recruited from the clinic and MTurk in terms of age [clinic = 67.33 (8.86), MTurk = 63.59 (5.03)], % female (clinic = 55, MTurk = 65.6), years of education [clinic = 15.70 (1.17), MTurk = 14.94 (1.68)], and % left-handed (clinic = 25, MTurk = 9.38).

Healthy participants. A cohort of 99 unique users were recruited from Amazon's MTurk following a similar protocol described above. This group had an average age of 37.58 (SD = 9.19) and an average year of education of 14.90 (SD = 1.41). In this cohort, 38.2% were female, and 15.3% were left-handed.

The clinical data had been collected previously at MGH. All clinic participants gave written informed consent in accordance with guidelines established by the Massachusetts General Brigham Healthcare System Institutional Review Boards which govern human subject research at MGH. MTurk participants were consented according to the protocols of the Committee on the Use of Humans as Experimental Subjects (COUHES) at the Massachusetts Institute of Technology. This study was approved by the COUHES at the Massachusetts Institute of Technology.

Procedure.

Language samples. The participants were asked to look at a drawing of a family at a picnic from the Western Aphasia Battery–Revised (98) and describe it using as many full sentences as they could. Responses were audio-recorded using an Olympus VN-702PC Voice Recorder in a quiet room and later transcribed into text using the Microsoft Dictate application. The transcriptions were then manually checked for accuracy by a research collaborator who was blind to the grouping. Disfluencies of speech such as repetitions and use of fillers, such as "um," "you know," etc., were annotated according to the Codes for the Human Analysis of Transcripts (CHAT) Transcription Format (99) and removed from further analyses. More specific details on the filtering process are provided in *SI Appendix*, supplementary material 3.

Following the protocol in the B.D. laboratory for PPA diagnosis, we also asked all participants to describe their work. The patient data were not available for us to analyze (because of privacy issues), so we only had access to the data from the healthy MTurk individuals. Only a subset ($n = 26$) of the 99 participants chose to participate in this part of the study.

Constructing the syntactic rules. An illustration of the process of constructing syntactic rules is provided in Fig. 1. Sentences were parsed using the automated lexicalized dependency parser in the Stanford Lexicalized Parser Package (v3.9.2). The parser automatically determines the heads in each sentence as well as their dependencies. A syntactic rule was constructed by listing the dependencies of a head in the order of appearance in the original sentence. Thus, the syntactic rules extracted from the sentence "The older couple is picnicking with wine" are 1) for "couple," det + amod + NOUN; 2) for "picnicking," nsubj + aux + VERB + nmod:with; and 3) for "wine," case + NOUN. Descriptions of each of the dependency relations can be found in *SI Appendix*, Table S1, and also at https://universaldependencies.org/en/dep/.

Measuring the frequencies of syntactic rules and words. To measure syntactic rule and word frequencies, we used the Switchboard corpus (62), which consists of spontaneous telephone conversations averaging 6 min in length spoken by over 500 speakers of both sexes from a variety of dialects of American English. We use it to estimate word and syntax frequencies in spoken English, independently of our patient and control sample. The corpus contains 2,345,269 words. For word frequency, the full version of the Switchboard corpus was used with some filtering to remove disfluencies and function words. Content word frequencies reported in this study (which we use as our main measure of word frequency) are based on frequencies of main verbs (excluding auxiliary verbs and be, do, and have), nouns, adjectives, and adverbs. Word frequency in our main analyses excluded function words as some studies have reported group differences in the ratio of function to content words among PPA variants (28). In footnotes, we also report results for all-word frequency (not limited to just content words).

To measure syntax frequencies, we used a subset of the Switchboard that has been parsed and manually annotated, which allowed us to remove disfluencies that occur in the spoken language (100). This subset consists of 588,183 words. The resulting corpus was then parsed to extract structure frequencies as described in section *The syntax–lexicon trade-off at the utterance level in a subset of healthy individuals describing their work*. A total of 7,090 types of syntactic rules occurring more than once were extracted.

Our analyses consider utterances, not words, as a basic unit. The word frequency and syntax frequency of each utterance were calculated by taking the average log frequency (with +1 smoothing) of all content words and all syntactic rules within the utterance, respectively, based on the Switchboard corpus. Therefore, syntactic and lexical frequencies are obtained independently of the task itself, except for the analysis in *The syntax–lexicon trade-off at the utterance level in the Switchboard corpus*, in which the target of analysis is Switchboard itself.

Other measures of syntactic complexity.

Dependency distance. The dependency distance complexity of a text or speech transcript was measured as the average number of intermediate words between heads and dependents (including the dependent itself) in each utterance (57, 61). For example, in Fig. 1, the distance between the dependent "with" and its head "wine" is one word, whereas the distance between the dependent "couple" and its head "picnicking" is two words. The average dependency distance for this sentence is 1.5.

Maximum incomplete dependencies. The incomplete dependency score was calculated at each word position in a sentence, parsing from left to right. The score for a sentence is the maximum number of incomplete dependencies for any word in that sentence (43, 55).

POS entropy. Roark et al. (48) proposed POS bigram entropy: uncertainty about a POS given the previous POS. We compute this measure independently for each participant using the totality of their parsed output, measuring the average uncertainty of a POS tag given the immediately preceding POS tag (without reference to the out-of-sample Switchboard estimates for this analysis).

Noun/verb ratio. After processing, the number of nouns is divided by the total number of nouns plus verbs (28).

Sentence length. After tokenizing, the number of nonpunctuation tokens in the sentence was used as sentence length.

Progressive aphasia severity scale. The progressive aphasia severity scale (PASS) (34) uses clinicians' best judgment, integrating information from patient test performance in the office as well as a companion's description of routine daily functioning. PASS includes judgments about fluency, syntax, word retrieval and expression, repetition, auditory comprehension, single word comprehension, reading, writing, and functional communication. PASS syntax subscore is based on a clinician's judgment of the frequency of sentences with simple structures, paragrammatism, use of word forms (run, ran), functor words (the, an), and word order when forming phrases and sentences in primary modality (speech or writing).

NAT. The NAT (42) asks the participants to assemble individual word cards presented in scrambled order into meaningful sentences. The structure types tested in NAT are three canonical structures (active, subject-extracted wh- questions, and subject clefts) and three noncanonical structures (passives, object-extracted wh- questions, and object clefts).

Data Availability. Anonymized deidentified data, Python code, and R code have been deposited in Open Science Framework (https://osf.io/sr3ag/). Speech samples from the work description task could be considered personal information and are not included in the repository but are available upon request.

ACKNOWLEDGMENTS. We thank Matthew Honnibal and Mark Johnson for creating the parsed version of Switchboard that we used and Andrew Caines for sharing it. We thank Jordan Green and Claire Cordella for the provision of healthy control data from the Speech and Feeding Disorders Laboratory at the MGH Institute of Health Professions. We deeply appreciate the expert clinical assessments of these patients by Megan Quimby, MS, CCC-SLP; Daisy Hochberg, MS, CCC-SLP; and other members of the Frontotemporal Disorders Unit Speech-Language Pathology Program. We thank National Institute on Deafness and Other Communication Disorders Grant R01 DC014296, National Institute on Aging Grant R21 AG073744, Baer Foundation, and Tommy Rickles Endowed Chair in PPA Research for supporting this work and the patients and families who participated for their valuable contributions. We also appreciate the support from Harvard Catalyst | The Harvard Clinical and Translational Science Center (National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH Award UL1 TR002541). We thank Damian Blasi, Gary Dell, Ev Fedorenko, Vic Ferreira, Roger Levy, and audiences at the human sentence processing conference (2021) and Tedlab for comments on earlier versions of this work.

1. W. J. M. Levelt, *Speaking: From Intention to Articulation* (A Bradford Book, 1989).
2. H. Paul, *Prinzipien der sprachgeschichte* (M. Niemeyer, 1886).
3. W. Wundt, *Völkerpsychologie: Die Sprache* (Kröner-Engelmann, 1900).
4. J. K. Bock, Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychol. Rev.* **89**, 1–47 (1982).
5. W. J. Levelt, A. Roelofs, A. S. Meyer, A theory of lexical access in speech production. *Behav. Brain Sci.* **22**, 1–38, discussion 38–75 (1999).

6. K. Bock, V. S. Ferreira, "Syntactically speaking" in *The Oxford Handbook of Language Production*, M. Goldrick, V. S. Ferreira, M. Miozzo, Eds. (Oxford University Press, 2014), pp. 21–46.
7. J. K. Bock, "Coordinating words and syntax in speech plans" in *Progress in the Psychology of Language*, A. Ellis, Ed. (Psychology Press, 1987), pp. 337–390.
8. A. E. Konopka, K. Bock, Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognit. Psychol.* **58**, 68–101 (2009).
9. W. Levelt, *A History of Psycholinguistics: The Pre-Chomskyan Era* (Oxford University Press, 2021).
10. B. Macwhinney, Starting points. *Language* **53**, 152–168 (1977).
11. R. S. Berndt, F. Boller, J. Grafman, "More than just words: Sentence production in aphasia" in *Handbook of Neuropsychology: Language and Aphasia*, R. S. Berndt, Ed. (Elsevier, 2000), pp. 173–187.
12. H. Goodglass, E. Kaplan, *The Assessment of Aphasia and Related Disorders* (Lea & Febiger, 1983).
13. R. Jakobson, "Two aspects of language and two types of aphasic disturbances" in *Fundamentals of Language*, R. Jakobson, M. Halle, Eds. (Mouton, The Hague, 1956), pp. 55–83.
14. H. H. Kolk, C. Heeschen, Agrammatism, paragrammatism and the management of language. *Lang. Cogn. Process.* **7**, 89–129 (1992).
15. E. Rochon, E. M. Saffran, R. S. Berndt, M. F. Schwartz, Quantitative analysis of aphasic sentence production: Further development and new data. *Brain Lang.* **72**, 193–218 (2000).
16. E. M. Saffran, R. S. Berndt, M. F. Schwartz, The quantitative analysis of agrammatic production: Procedure and data. *Brain Lang.* **37**, 440–479 (1989).
17. M. M. Mesulam, Slowly progressive aphasia without generalized dementia. *Ann. Neurol.* **11**, 592–598 (1982).
18. M. M. Mesulam, Primary progressive aphasia. *Ann. Neurol.* **49**, 425–432 (2001).
19. M. L. Gorno-Tempini *et al.*, Classification of primary progressive aphasia and its variants. *Neurology* **76**, 1006–1014 (2011).
20. R. S. Berndt, C. C. Mitchum, A. N. Haendiges, J. Sandson, Verb retrieval in aphasia. 1. Characterizing single word impairments. *Brain Lang.* **56**, 68–106 (1997).
21. H. Bird, S. Franklin, Cinderella revisited: A comparison of fluent and non-fluent aphasic speech. *J. Neurolinguistics* **9**, 187–206 (1996).
22. S. D. Breedin, E. M. Saffran, M. F. Schwartz, Semantic factors in verb retrieval: An effect of complexity. *Brain Lang.* **63**, 1–31 (1998).
23. J. K. Gordon, G. S. Dell, Learning to divide the labor: An account of deficits in light and heavy verb production. *Cogn. Sci.* **27**, 1–40 (2003).
24. S. E. Kohn, M. P. Lorch, D. M. Pearson, Verb finding in aphasia. *Cortex* **25**, 57–69 (1989).
25. R. S. Berndt, A. N. Haendiges, C. C. Mitchum, J. Sandson, Verb retrieval in aphasia. 2. Relationship to sentence processing. *Brain Lang.* **56**, 107–137 (1997).
26. V. C. Zimmerer *et al.*, Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: An approach based on usage-frequency. *Cortex* **133**, 103–119 (2020).
27. K. Marcotte *et al.*, White matter disruption and connected speech in non-fluent and semantic variants of primary progressive aphasia. *Dement. Geriatr. Cogn. Disord. Extra* **7**, 52–73 (2017).
28. S. M. Wilson *et al.*, Connected speech production in three variants of primary progressive aphasia. *Brain* **133**, 2069–2088 (2010).
29. S. Ash *et al.*, Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology* **81**, 329–336 (2013).
30. C. K. Thompson, J. E. Mack, Grammatical impairments in PPA. *Aphasiology* **28**, 1018–1037 (2014).
31. A. K. Kittredge, G. S. Dell, J. Verkuilen, M. F. Schwartz, Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cogn. Neuropsychol.* **25**, 463–492 (2008).
32. R. C. Oldfield, A. Wingfield, Response latencies in naming objects. *Q. J. Exp. Psychol.* **17**, 273–281 (1965).
33. S. M. Wilson *et al.*, Syntactic processing depends on dorsal language tracts. *Neuron* **72**, 397–403 (2011).
34. D. Sapolsky, K. Domoto-Reilly, B. C. Dickerson, Use of the Progressive Aphasia Severity Scale (PASS) in monitoring speech and language status in PPA. *Aphasiology* **28**, 993–1003 (2014).
35. S. Ash *et al.*, Non-fluent speech in frontotemporal lobar degeneration. *J. Neurolinguist.* **22**, 370–383 (2009).
36. D. Bergeron *et al.*, Prevalence of amyloid-β pathology in distinct variants of primary progressive aphasia. *Ann. Neurol.* **84**, 729–740 (2018).
37. C. K. Thompson *et al.*, Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology* **26**, 20–43 (2012).
38. K. C. Fraser *et al.*, Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex* **55**, 43–60 (2014).
39. J. M. Harris, J. A. Saxon, M. Jones, J. S. Snowden, J. C. Thompson, Neuropsychological differentiation of progressive aphasic disorders. *J. Neuropsychol.* **13**, 214–239 (2019).
40. S. Ash *et al.*, Dissociation of quantifiers and object nouns in speech in focal neurodegenerative disease. *Neuropsychologia* **89**, 141–152 (2016).
41. C. K. Thompson *et al.*, Syntactic and morphosyntactic processing in stroke-induced and primary progressive aphasia. *Behav. Neurol.* **26**, 35–54 (2013).
42. S. Weintraub *et al.*, The Northwestern anagram test: Measuring sentence production in primary progressive aphasia. *Am. J. Alzheimers Dis. Other Demen.* **24**, 408–416 (2009).
43. V. H. Yngve, A model and an hypothesis for language structure. *Proc. Am. Philos. Soc.* **104**, 444–466 (1960).
44. L. Abbeduto, S. Rosenberg, "Linguistic communication and mental retardation" in Advances in Applied Psycholinguistics, S. Rosenberg, Ed. (Cambridge Monographs and Texts in Applied Psycholinguistics, Cambridge University Press, 1987), pp. 76–125.
45. H. Cheung, S. Kemper, Competing complexity metrics and adults' production of complex sentences. *Appl. Psycholinguist.* **13**, 53–76 (1992).
46. J. E. Mack, E. Barbieri, S. Weintraub, M.-M. Mesulam, C. K. Thompson, Quantifying grammatical impairments in primary progressive aphasia: Structured language tests and narrative language production. *Neuropsychologia* **151**, 107713 (2021).
47. N. Rezaii *et al.*, An information-theoretic characterization of language production in primary progressive aphasia. *J. Neuropsychiatry Clin. Neurosci.* **32**, E19 (2020).
48. B. Roark, M. Mitchell, K. Hollingshead, "Syntactic complexity measures for detecting mild cognitive impairment" in *Biological, Translational, and Clinical Language Processing*, K. B. Cohen, D. Demner-Fushman, Ca. Friedman, L. Hirschman, J. Pestian, Eds. (Association for Computational Linguistics, 2007), pp. 1–8.
49. F. S. Aronsson, M. Kuhlmann, V. Jelic, P. Östberg, Is cognitive impairment associated with reduced syntactic complexity in writing? Evidence from automated text analysis. *Aphasiology* **35**, 900–913 (2021).
50. S. V. Pakhomov *et al.*, A computerized technique to assess language use patterns in patients with frontotemporal dementia. *J. Neurolinguist.* **23**, 127–144 (2010).
51. J. E. Sung, S. Choi, B. Eom, J. K. Yoo, J. H. Jeong, Syntactic complexity as a linguistic marker to differentiate mild cognitive impairment from normal aging. *J. Speech Lang. Hear. Res.* **63**, 1416–1429 (2020).
52. S. Pakhomov, D. Chacon, M. Wicklund, J. Gundel, Computerized assessment of syntactic complexity in Alzheimer's disease: A case study of Iris Murdoch's writing. *Behav. Res. Methods* **43**, 136–144 (2011).
53. J. Hale, "A probabilistic Earley parser as a psycholinguistic model" in *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies, NAACL '01* (Association for Computational Linguistics, 2001), pp. 1–8.
54. R. Levy, Expectation-based syntactic comprehension. *Cognition* **106**, 1126–1177 (2008).
55. N. Chomsky, G. A. Miller, Introduction to the formal analysis of natural languages. *J. Symb. Log.* **33**, 299–300 (1968).
56. J. Kimball, Seven principles of surface structure parsing in natural language. *Cognition* **2**, 15–47 (1973).
57. E. Gibson, Linguistic complexity: Locality of syntactic dependencies. *Cognition* **68**, 1–76 (1998).
58. E. Gibson, "The dependency locality theory: A distance-based theory of linguistic complexity" in *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium*, A. Marantz, Y. Miyashita, W. O'Neil, Eds. (MIT Press, 2000), pp. 94–126.
59. R. L. Lewis, S. Vasishth, J. A. Van Dyke, Computational principles of working memory in sentence comprehension. *Trends Cogn. Sci.* **10**, 447–454 (2006).
60. R. L. Lewis, S. Vasishth, An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* **29**, 375–419 (2005).
61. J. A. Hawkins, *Efficiency and Complexity in Grammars* (Oxford University Press, 2004).
62. J. J. Godfrey, E. C. Holliman, J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development" in *Proceedings of the 1992 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'92* (IEEE Computer Society, 1992), vol. 1, pp. 517–520.
63. D. Klein, C. D. Manning, "Accurate unlexicalized parsing" in *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, 2003), pp. 423–430.
64. R. S. Jackendoff, *Semantic Structures* (MIT Press, 1990).
65. B. Levin, *English Verb Classes and Alternations* (The University of Chicago Press, 1993).
66. J. D. Rohrer, M. N. Rossor, J. D. Warren, Syndromes of nonfluent primary progressive aphasia: A clinical and neurolinguistic analysis. *Neurology* **75**, 603–610 (2010).
67. S. Cho *et al.*, Automated analysis of lexical features in frontotemporal degeneration. *Cortex* **137**, 215–231 (2021).
68. D. J. Barr, R. Levy, C. Scheepers, H. J. Tily, Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Mem. Lang.* **68**, 255–278 (2013).
69. A. M. Staffaroni *et al.*, Uniform data set language measures for bvFTD and PPA diagnosis and monitoring. *Alzheimers Dement. (Amst)* **13**, e12148 (2021).
70. T. F. Jaeger, Redundancy and reduction: Speakers manage syntactic information density. *Cognit. Psychol.* **61**, 23–62 (2010).
71. R. Levy, T. F. Jaeger, "*Speakers optimize information density through syntactic reduction" in* Advances in Neural Information Processing Systems 19, B. Schölkopf, J. Platt, T. Hoffman, Eds. (MIT Press, 2006), pp. 849–856.
72. N. J. Smith, R. Levy, The effect of word predictability on reading time is logarithmic. *Cognition* **128**, 302–319 (2013).
73. K. Mahowald, E. Fedorenko, S. T. Piantadosi, E. Gibson, Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition* **126**, 313–318 (2013).
74. S. T. Piantadosi, H. Tily, E. Gibson, Word lengths are optimized for efficient communication. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3526–3529 (2011).
75. M. Zhan, R. Levy, "Comparing theories of speaker choice using a model of classifier production in Mandarin Chinese" in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, M. Walker, H. Ji, A. Stent, Eds. (Association for Computational Linguistics, 2018), vol. 1, pp. 1997–2005.
76. V. S. Ferreira, "Ambiguity, accessibility, and a division of labor for communicative success" in *Psychology of Learning and Motivation*, B. H. Ross, Ed. (2008), pp. 209–246.
77. V. S. Ferreira, G. S. Dell, Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognit. Psychol.* **40**, 296–340 (2000).
78. K. Bock, D. E. Irwin, Syntactic effects of information availability in sentence production. *J. Verbal Learning Verbal Behav.* **19**, 467–484 (1980).
79. M. C. Macdonald, How language production shapes language form and comprehension. *Front. Psychol.* **4**, 226 (2013).
80. J. K. Bock, D. E. Irwin, D. J. Davidson, "Putting first things first" in *The Integration of Language, Vision, and Action: Eye Movements and the Visual World*, J. M. Henderson, F. Ferreira, Eds. (Psychology Press, 2008), pp. 249–278.
81. J. K. Bock, R. K. Warren, Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* **21**, 47–67 (1985).
82. K. Christianson, F. Ferreira, Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition* **98**, 105–135 (2005).
83. M. H. Kelly, J. K. Bock, F. C. Keil, Prototypicality in a linguistic context: Effects on sentence structure. *J. Mem. Lang.* **25**, 59–74 (1986).
84. K. H. Onishi, G. L. Murphy, K. Bock, Prototypicality in sentence production. *Cognit. Psychol.* **56**, 103–141 (2008).
85. M. Prat-Sala, H. P. Branigan, Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *J. Mem. Lang.* **42**, 168–182 (2000).
86. B. Rapp, M. Goldrick, Discreteness and interactivity in spoken word production. *Psychol. Rev.* **107**, 460–499 (2000).
87. G. S. Dell, A spreading-activation theory of retrieval in sentence production. *Psychol. Rev.* **93**, 283–321 (1986).

88. Z. Shao, A. S. Meyer, A. Roelofs, Selective and nonselective inhibition of competitors in picture naming. *Mem. Cognit.* **41**, 1200–1211 (2013).
89. N. Nozari, The dual origin of semantic errors in access deficit: Activation vs. inhibition deficit. *Cogn. Neuropsychol.* **36**, 31–53 (2019).
90. J. K. Bock, Syntactic persistence in language production. *Cognit. Psychol.* **18**, 355–387 (1986).
91. D. G. MacKay, The problems of flexibility, fluency, and speed–accuracy trade-off in skilled behavior. *Psychol. Rev.* **89**, 483–506 (1982).
92. F. Chang, G. S. Dell, K. Bock, Becoming syntactic. *Psychol. Rev.* **113**, 234–272 (2006).
93. G. S. Dell, F. Chang, Z. M. Griffin, Connectionist models of language production: Lexical access and grammatical encoding. *Cogn. Sci.* **23**, 517–542 (1999).
94. E. Fedorenko, M. K. Behr, N. Kanwisher, Functional specificity for high-level linguistic processing in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16428–16433 (2011).
95. D. Sapolsky *et al.*, Monitoring progression of primary progressive aphasia: Current approaches and future directions. *Neurodegener. Dis. Manag.* **1**, 43–55 (2011).
96. D. Knopman, S. Weintraub, V. Pankratz, Language and behavior domains enhance the value of the clinical dementia rating scale. *Alzheimers Dement.* **7**, 293–299 (2011).
97. S. Tomaszewski Farias *et al.*, The measurement of everyday cognition: Development and validation of a short form of the Everyday Cognition scales. *Alzheimers Dement.* **7**, 593–601 (2011).
98. A. Kertesz, A. Kertesz, J. C. Raven, *PsychCorp (Firm), WAB-R: Western Aphasia Battery-Revised* (PsychCorp, 2007).
99. B. Macwhinney, Tools for analyzing talk, part 1: The CHAT transcription format (2019). https://doi.org/10.21415/3MHN-0Z89. Accessed 19 July 2021.
100. M. Honnibal, M. Johnson, Joint incremental disfluency detection and dependency parsing. *Trans. Assoc. Comput. Linguist.* **2**, 131–142 (2014).