NeuroImage: Clinical 42 (2024) 103602

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Discourse- and lesion-based aphasia quotient estimation using machine learning

Nicholas Riccardi [a,*], Satvik Nelakuditi [b], Dirk B. den Ouden [a], Chris Rorden [c], Julius Fridriksson [a], Rutvik H. Desai [c]

[a] Department of Communication Sciences and Disorders, University of South Carolina, United States
[b] Spring Valley High School, Columbia, SC, United States
[c] Department of Psychology, University of South Carolina, United States

ARTICLE INFO

Keywords:
Aphasia
Stroke
Discourse Production
Machine Learning
Lesion Symptom Mapping

ABSTRACT

Discourse is a fundamentally important aspect of communication, and discourse production provides a wealth of information about linguistic ability. Aphasia commonly affects, in multiple ways, the ability to produce discourse. Comprehensive aphasia assessments such as the Western Aphasia Battery-Revised (WAB-R) are time- and resource-intensive. We examined whether discourse measures can be used to estimate WAB-R Aphasia Quotient (AQ), and whether this can serve as an ecologically valid, less resource-intensive measure. We used features extracted from discourse tasks using three AphasiaBank prompts involving expositional (picture description), story narrative, and procedural discourse. These features were used to train a machine learning model to predict the WAB-R AQ. We also compared and supplemented the model with lesion location information from structural neuroimaging. We found that discourse-based models could estimate AQ well, and that they outperformed models based on lesion features. Addition of lesion features to the discourse features did not improve the performance of the discourse model substantially. Inspection of the most informative discourse features revealed that different prompt types taxed different aspects of language. These findings suggest that discourse can be used to estimate aphasia severity, and provide insight into the linguistic content elicited by different types of discourse prompts.

1. Introduction

Brain injury via stroke or neurodegenerative disease can often result in aphasia, defined as impaired language and communication. Aphasia can lead to significant declines in quality of life and well-being (Bullier et al., 2020; Spaccavento et al., 2014), as the ability to communicate effectively is vital for interpersonal relationships, employment, and navigating the world. A major part of this decline can be related to impairments in spoken discourse (Galski et al., 1998). Spoken discourse provides a wealth of information about linguistic ability that is related to aphasia severity. Hence, evaluation of discourse in persons with aphasia has gained increasing recognition for clinical assessment and treatment (Bryant et al., 2016; Stark & Fukuyama, 2021). The majority of current aphasia assessments, such as the Western Aphasia Battery-Revised (WAB-R; (Kertesz, 2007)) are rigorous but relatively demanding standardized tests that can be burdensome for survivors of stroke, their families, and clinicians. In the United States, it is often difficult for people to even be approved or financially supported for comprehensive baseline language evaluations post-stroke (Walker et al., 2022). Hence, a goal is to attempt to develop supplementary assessments that are brief but comparable. If reliable, such assessments could be used for triage purposes, measuring change in language abilities over time, or for individuals who have limited access to healthcare resources (e.g., rural or impoverished). In this context, discourse analysis is a promising line of research, given the rich set of microstructural (lexical, syntactic) and macrostructural (cohesion, coherence) elements in discourse.

Compared to a 45 minute to 2 hour standardized test, eliciting discourse is more tractable for a non-specialist, thanks to resources such as AphasiaBank (Fromm et al., 2020; Macwhinney et al., 2011). Tasks include an expositional description of a sequence of pictures (Broken Window), narrative discourse without visual aids ('tell me the story of Cinderella'), and procedural ('tell me how to make a peanut butter and jelly sandwich'). The tasks are brief (<5 min) and data collection could be done remotely via mobile phone applications or wearable monitors.

* Corresponding author at: Department of Communication Sciences and Disorders, United States.
E-mail address: riccardn@email.sc.edu (N. Riccardi).

https://doi.org/10.1016/j.nicl.2024.103602
Received 28 June 2023; Received in revised form 1 April 2024; Accepted 1 April 2024
Available online 6 April 2024

The prompts allow for more continuous and naturalistic output than other language assessments such as confrontation naming, sentence-picture matching, or production of isolated sentences. The variety of prompts (e.g., expositional vs. procedural) also allows for the inspection of relationships between linguistic and more domain-general cognitive processes such as procedural or episodic memory (Stark, 2019). These unique demands mean that discourse samples can measure language loss or recovery in a more naturalistic way than long-form standardized tests (Bryant et al., 2016).

Previous studies have noted that the time-intensive nature of discourse transcription and coding presents a significant barrier to using discourse analysis clinically and in research settings (Bryant et al., 2016; Cruice et al., 2020; Stark & Fukuyama, 2021). However, recent advances in computerized transcription and natural language processing are likely to aid automated transcription and coding in the coming years (Dalton et al., 2022; Liu et al., 2023), although these automated methods are not yet widely used in clinical or research settings. Some work has been done to develop brief aphasia screenings such as the mobile aphasia screening test (Choi et al., 2015), Quick Aphasia Battery (Wilson et al., 2018), Bedside WAB-R (Kertesz, 2007), or others designed for detection of paraphasia (Le et al., 2017) or primary progressive aphasia (PPA; (Fraser et al., 2014)). However, these have largely been designed with the binary goal of detecting the presence or absence of aphasia, specifically in severely affected patients (Kertesz, 2022), or classifying subtypes of PPA, instead of assessing the entire range of aphasia severity. Additionally, some of these brief measures have short and strict item-level response times, resulting in a tendency to overestimate aphasia severity in mildly affected patients (Wilson et al., 2018).

There is a growing body of research that uses discourse as an outcome measure of therapy (Bryant et al., 2016). Some work has been done to investigate higher-level conceptual properties (macrostructure) of spoken discourse, such as main concept production and informativeness metrics. These studies have found that people with aphasia tend to produce less informative speech and that different aphasia subtypes have differing levels of semantic and conceptual content in their speech (Dalton & Richardson, 2019; Gordon, 2008; Kong, 2009; Kong et al., 2016). Other work has focused more on discourse microstructure, the focus of the current manuscript. Microstructure here is defined as grammatical and 'lower-level' linguistic features, such as number of nouns, utterance length, etc., as opposed to the correctness/completeness of higher-level semantic or conceptual content of the speech.

For example, Stark (2019) quantitatively established that the different discourse prompt types (e.g., expositional, narrative, and procedural) tend to tax different aspects of the language system in both controls and people with aphasia. Narrative discourse was found to elicit the most content-rich speech. Procedural discourse, on the other hand, elicited the lowest syntactic complexity. These findings suggest that using multiple prompt types may be important for discourse-based language assessment.

In related work, Fromm et al. (2016) investigated the relationship between proposition density and aphasia severity. Proposition density is the number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words, and is an indication of communicative efficiency or adequacy. While they found that people with aphasia have significantly different proposition density than controls, there was no clear relationship between proposition density and aphasia severity. This was due to people with fluent aphasia having low AQs, but inflated proposition density, while people with nonfluent aphasia would have equally low AQs and low proposition density. However, Bryant et al. (2013) found a significant positive relationship between AQ and proposition density, but the frequency of aphasia type in that sample was not reported, making it possible that most of those participants had nonfluent aphasia. These findings highlight the difficulty of using quantitative discourse analysis. There are complex, multidimensional, and collinear relationships between microstructural variables, aphasia type, and overall language ability, which are difficult (or impossible) to capture with traditional univariate statistics. For this reason, dimension reduction, multivariate, or machine learning methods may be well-suited for discourse analysis.

For example, Stark and Fukuyama (2021) used two dimension reduction techniques to investigate the relationship between discourse prompt type and microstructural variables in speech produced by those with and without aphasia. This work demonstrated that produced microstructural speech features are largely dissociable depending on the type of discourse prompt used for speech elicitation. Casilio and colleagues (2019) used factor analysis to reduce the dimensionality of 27 auditory-perceptual ratings of connected speech samples from AphasiaBank. They found four factors (paraphasia, logopenia, agrammatism, and motor speech) that emerged which explained 79 % of the variance in the connected speech samples. However, predicting AQ using discourse features was not the main aim of those studies.

Finally, from a neuroanatomical perspective, a large body of work has been conducted to map lesion features to specific discourse qualities, aphasia subtypes, or language measures such as sentence processing, naming, word comprehension, and semantic knowledge (Fridriksson et al., 2018; Fridriksson et al., 2016; Hillis et al., 2017; Kristinsson et al., 2020; Magnusdottir et al., 2013; Riccardi et al., 2022; Riccardi et al., 2020; Riccardi et al., 2019; Riccardi et al., 2023; Schwen Blackett et al., 2022). Regarding discourse specifically, Mirman and colleagues (2019) used lesion-symptom mapping to investigate the neural correlates of discourse-elicited articulatory and grammatical deficits. They found that these discourse deficits were related to damage to motor and frontal cortices, and that those deficits were correlated with aphasia severity (AQ) and WAB fluency scores. These works highlight that various aspects of language performance can be mapped to brain damage in specific areas, and, in the context of the current study, begs the question of whether inclusion of lesion features can aid in the estimation of aphasia severity above and beyond discourse features alone.

In sum, the research on microstructural speech features elicited from discourse production has established that: 1) different prompts elicit different speech features, 2) some of these features are related to damage in specific areas of the brain, 3) these features cluster together in ways that differ between aphasia subtypes and can distinguish people with aphasia from controls, and 4) some of these features are clearly correlated with aphasia severity (AQ) while others, like proposition density, have a more complex relationship with aphasia severity. Given the complex relationships between speech features with each other and with aphasia severity, a relevant question is whether machine learning techniques can be used to quantify and predict aphasia severity (AQ) using speech microstructure. If WAB-R AQ can be quantified from discourse prompts that are short (<5 min), more naturalistic, and can be administered by non-clinicians, then it is a good proof-of-concept that these discourse tasks can be used as brief, yet quantitative, language evaluations in situations when full administration of longer tests is undesirable or unfeasible. This would be especially useful for rural or at-risk populations. For example, researchers or clinicians may want a quantitative estimate of a person's AQ as part of a 'check-in' or longitudinal assessment, but the person does not have a means of transportation, or a trained clinician is unavailable to administer the WAB-R. Discourse samples can be quickly and easily recorded remotely, even by a non-clinician, and then used to estimate AQ, while simultaneously collecting the rich data that accompanies recordings of connected speech.

Here, we used expositional, narrative, and procedural discourse tasks in a group of 71 stroke survivors with available structural neuroimaging scans. Our first aim was to use a machine learning approach with microstructural speech features to predict aphasia severity (WAB-R AQ). A second aim was to examine how the inclusion of lesion features impacts model performance. Lesion features may provide additional information about AQ beyond speech features, boosting estimation accuracy. While this may be less clinically useful if structural MRI collection and analysis is not feasible, it is a valuable question to

consider in the context of understanding the neurobiology of language post-stroke. If lesion features significantly boost model estimation, it would mean that structural MRI is capturing information about aphasia severity above and beyond what is being captured by discourse microstructure. If lesion features do not aid model performance, then it indicates that discourse features alone are able to capture variance in WAB-R scores. A final aim was to investigate feature weights in the predictive models to determine which discourse or lesion features are most informative for the estimation of AQ. This would provide novel information about what features, in a machine learning context, are capturing the most information related to aphasia severity and how these features may differ according to the discourse prompt being used. We used Support Vector Regression (SVR) to predict AQ and assess feature importance.

2. Materials and methods

2.1. Participants

Speech recordings were obtained from 71 unilateral left-hemisphere chronic (>12 months post-stroke, mean = 60 months, range = 12 – 237) stroke survivors by the Center for Study of Aphasia Recovery (C-STAR), as part of a multi-day data collection battery (see Spell et al. (2020)). Demographics are shown in Table 1. Participants were a mean of 61.7 years old (range = 29 – 80, SD = 10.7). Participants were screened for conditions other than stroke via self and/or caregiver report of no history or diagnosis of dementia and or neurological disorders. The battery included structural and functional neuroimaging, administration of the WAB-R (Kertesz, 2007) by licensed speech-language pathologists, discourse collection, and other cognitive and language testing. Although beyond the scope of the current manuscript, other measures collected include Philadelphia Naming Test (Roach et al., 1996), Northwestern Assessment of Verbs and Sentences (Thompson, 2012), health history questionnaires, Pyramids and Palm Trees (Howard & Patterson, 1992), and various in-house tasks designed to measure language abilities (Riccardi et al., 2022; Riccardi et al., 2020; Riccardi et al., 2019).

Mean WAB-R AQ score was 65.9 (range = 14.5 – 100, SD = 23.5). Aphasia subtype was based on WAB-R AQ classification. Among these participants, 10 did not suffer from aphasia, while the rest had different types of aphasia: Broca's (28), Anomic (14), Conduction (11), Global (4), Wernicke's (3), and Transcortical Motor (1). A classification of 'no aphasia' was defined by the recommended cutoff of an AQ greater than 93.8 (Kertesz, 2007). The 10 participants without aphasia suffered left hemisphere stroke but did not have self-reported or clinician-identified language problems. We included them in the analysis as it provides the statistical models with information about lesion/discourse features that are associated with high AQ. All participants signed informed consent, and the research was approved by the University of South Carolina Institutional Review Board.

2.2. Behavioral data

At intake, each participant was prompted by a clinician to narrate the Cinderella story, describe how to make a peanut butter and jelly (PBJ) sandwich, and explain the sequence of events shown in a picture, referred to as Broken Window, according to AphasiaBank prompt directions (Macwhinney et al., 2011). Their discourse was video recorded. For each of the discourse tasks, graduate students trained by certified speech-language pathologists generate transcripts, separate utterances into communication units and code the transcripts for specific linguistic variables (e.g., word repetitions, semantic and phonemic paraphasias, fillers, etc.) using the CHAT transcription format for automatic analyses by the CLAN program. All transcripts were then rated by second trained study staff member, and a final consensus was made on any disagreements before running any CLAN analyses on the transcripts. Inter and intra rater reliability is completed every spring on 10 % of all transcripts

Table 1
Participant information, sorted by WAB-R Aphasia Type.

ID	WAB-R AQ	WAB-R Aphasia Type	Days Post-Stroke	Age	Gender
1013	91.8	Anomic	730	44	Male
1014	85.8	Anomic	5432	60	Male
1026	88.8	Anomic	1552	48	Female
1028	91.3	Anomic	371	71	Male
1029	77.2	Anomic	476	69	Male
1033	80.3	Anomic	442	76	Male
1046	82.3	Anomic	578	58	Male
1049	82.7	Anomic	421	69	Female
1059	91.3	Anomic	3035	71	Female
1065	85	Anomic	381	75	Male
1069	92.6	Anomic	2332	64	Male
1099	79.9	Anomic	1323	66	Female
1103	86.5	Anomic	417	71	Male
1104	82.2	Anomic	485	60	Male
1002	72.2	Broca's	2722	58	Male
1004	52.1	Broca's	444	60	Male
1005	41	Broca's	4363	56	Male
1006	45.9	Broca's	7115	71	Male
1008	37.6	Broca's	425	70	Female
1012	40.6	Broca's	2706	50	Male
1015	52.9	Broca's	2439	55	Male
1016	36.3	Broca's	829	66	Male
1030	73.9	Broca's	502	43	Male
1031	57.8	Broca's	884	73	Male
1035	54.8	Broca's	4334	43	Female
1036	53	Broca's	400	50	Male
1039	56.3	Broca's	1952	54	Male
1040	65.5	Broca's	446	37	Female
1044	53.7	Broca's	3349	66	Male
1050	27.7	Broca's	691	68	Female
1062	64	Broca's	6335	64	Female
1063	25.4	Broca's	1744	76	Female
1064	57.9	Broca's	470	62	Male
1072	67	Broca's	2022	62	Female
1080	71.4	Broca's	787	78	Male
1081	30.5	Broca's	3004	49	Male
1082	71.7	Broca's	1445	59	Male
1096	77.4	Broca's	427	63	Male
1101	56.2	Broca's	472	80	Female
1102	44.5	Broca's	1659	44	Male
1106	32.4	Broca's	6579	76	Male
1112	63	Broca's	369	58	Female
1001	63.4	Conduction	2530	74	Female
1023	39.8	Conduction	1170	69	Female
1024	79.2	Conduction	802	29	Female
1034	68.9	Conduction	927	69	Male
1041	60.5	Conduction	4571	76	Male
1045	86	Conduction	1432	51	Male
1056	51.1	Conduction	453	52	Female
1060	63.1	Conduction	540	75	Female
1077	57.7	Conduction	363	59	Female
1079	88	Conduction	721	61	Female
1114	51.9	Conduction	1459	57	Female
1017	25.2	Global	2544	53	Male
1051	31.3	Global	1094	49	Male
1058	14.5	Global	630	65	Female
1092	25.3	Global	518	67	Female
1073	96.4	No Aphasia	4822	54	Female
1074	96.8	No Aphasia	4696	42	Female
1076	97.8	No Aphasia	2667	67	Female
1078	99.1	No Aphasia	4092	67	Male
1085	98.4	No Aphasia	3619	66	Male
1088	100	No Aphasia	3130	71	Female
1093	95.4	No Aphasia	517	65	Female
1097	99.2	No Aphasia	1635	62	Male
1098	98	No Aphasia	1984	65	Male
1119	98	No Aphasia	561	78	Female
1087	78.2	Transcortical Motor	448	60	Male
1003	33.9	Wernicke's	2362	63	Male
1047	30.3	Wernicke's	750	66	Male
1089	67.8	Wernicke's	366	57	Male

(see Spell et al. (2020)). Using Computerized Language ANalysis (CLAN) software (MacWhinney, 2000), we first ran the MOR command to analyze the morphosyntax and grammatical relations of all utterances in the transcripts. We then ran the EVAL command, which extracts a number of those indices for evaluating language in people with aphasia (Table 2). Examples of five of these completed CHAT transcripts are provided in the Supplementary Materials. All resulting indices were used as features in the machine learning algorithm, as our statistical approach (Section 2.5) was designed to automatically remove uninformative features during training of the model. See (Table 3).

2.3. MRI data acquisition and preprocessing

MRI data were obtained with a Siemens 3 T Trio System with a 12-channel head coil and a Siemens 3 T Prisma System with a 20-channel coil. Participants underwent two anatomical MRI sequences: (i) T1-weighted imaging sequence with a magnetization-prepared rapid-gradient echo (MPRAGE) turbo field echo (TFE) sequence with voxel size = 1 mm3, field of view (FOV) = 256 × 256 mm, 192 sagittal slices, 9° flip angle, repetition time (TR) = 2,250 ms, inversion time (TI) = 925 ms, echo time (TE) = 4.15 ms, generalized autocalibrating partial parallel acquisition (GRAPPA) = 2, and 80 reference lines; and (ii) T2-weighted MRI with a 3D sampling perfection with application optimized contrasts by using different flip angle evolutions protocol with the following parameters: voxel size = 1 mm3, FOV = 256 × 256 mm, 160 sagittal slices, variable flip angle, TR = 3,200 ms, TE = 212 ms, and no slice acceleration. The same slice center and angulation were used as in the T1 sequence.

Lesions were defined in native space by a neurologist in MRIcron (Rorden et al., 2012) on individual T2-weighted images. Preprocessing started with coregistration of the T2-weighted images to match the T-weighted images, allowing the lesions to be aligned to native T1 space. Images were warped to standard space using enantiomorphic (Nachev et al., 2008) segmentation-normalization (Ashburner & Friston, 2005) custom Matlab script (https://github.com/rordenlab/spmScripts/blob/master/nii_enat_norm.m) to warp images to an age-appropriate template image found in the Clinical Toolbox for SPM (https://www.nitrc.org/scm/?group_id=881). The normalization parameters were used to reslice the lesion into standard space using linear interpolation, with subsequent lesion maps stored at 1 × 1 × 1-mm resolution and binarized using a 50 % threshold. (Because interpolation can lead to fractional probabilities, this step confirms that each voxel is categorically either lesioned or unlesioned without biasing overall lesion volume.) Normalized images were visually inspected to verify quality.

Table 2
The list of discourse features extracted for each of the prompts. The last 16 features starting from Nouns to WordErrors are included as both absolute numbers and relative percentages, amounting to a total of 45 discourse features per prompt.

Name	Description
Duration	Total duration of discourse (sec)
Total Utts	Total utterances
MLU Utts	Total #utterances for calculating MLU below
MLU Words	Mean number of words per utterance
MLU Morphs	Mean number of morphemes per utterance
FreqTypes	Number of word types used
FreqTokens	Number of unique words used
FreqTTR	Ratio of types to tokens
Words/Min	Words per minute
Verbs/Utt	Number of verbs per utterance
Density	Proposition idea density
Retracing	Number of self-corrections during speech
Repetition	Number of word repetitions
Nouns	Words that were nouns
Prep	Words that were prepositions
Adj	Words that were adjectives
Adv	Words that were adverbs
Conj	Words that were conjunctions
Det/Art	Words that were determiners or articles
Pro	Words that were pronouns
Aux	Words that were auxiliaries
Verbs	Words that were verbs
3S	Verbs that were 3rd person singular
1S/3S	Verbs with same form for first/third person
Past	Verbs that were past tense
PastPart	Verbs that were past participles
PresPart	Verbs that were present participles
Plurals	Nouns that were plural
WordErrors	Sound, verbal, or mixed paraphasias

Table 3
Results summary for predicted AQ compared to observed AQ for each model. The Lesion Only column has only a single model, where all lesion features (and no discourse features) were included.

Prompt	Discourse	Discourse + Lesion	Lesion Only
Broken Window	r = 0.79 RMSE = 14.53 MAE = 10.44	r = 0.76 RMSE = 15.47 MAE = 11.29	~
Cinderella	r = 0.75 RMSE = 15.50 MAE = 12.14	r = 0.83 RMSE = 12.94 MAE = 9.53	~
PBJ	r = 0.70 RMSE = 16.75 MAE = 13.47	r = 0.72 RMSE = 16.56 MAE = 12.95	~
All Combined	r = 0.83 RMSE = 13.09 MAE = 9.77	r = 0.82 RMSE = 13.29 MAE = 10.03	r = 0.64 RMSE = 18.53 MAE = 14.69

2.4. Lesion feature extraction

The lesion incidence map is shown in Fig. 1. The average lesion volume was 112,736 mm^3 (SD = 84,360 mm^3). The resulting images were parcellated according to the Johns Hopkins University atlas (Faria et al., 2012; Mori et al., 2005; Wakana et al., 2004). For each participant, the percent of voxels damaged within each of these regions was calculated, and areas that were undamaged in all participants were removed from further analysis, resulting in 64 lesion features considered in this study (Supplementary Materials).

2.5. Machine learning approach

Our goal was to predict the AQ of a participant based on their discourse and/or lesion features with the help of machine learning. Two key design choices in developing a machine learning system are the learning algorithm and the feature set. We chose linear Support Vector Machines (SVM) which is a popular machine learning method that is known to perform well on relatively small datasets, (Mahmoud et al. 2021) and is resistant to overfitting. An appropriate subset of given features was selected through recursive feature elimination and cross-validation. Specifically, we used leave-one-out (LOO) to split the participants into a set of 70 for training and 1 for testing. Using the 70 samples in the training set, all the features were ranked using recursive feature elimination. We then selected a combination of top features through cross-validation as follows. By employing LOO again, the training set of 70 samples was further split into 69 for training and 1 for validation. By training the SVM on the 69 samples, we predicted the AQ for the one in the validation set. This is done 70 times with each participant in the validation set once. The predicted AQ values were compared against the true AQ values to compute an R^2 score. This process was repeated for each combination of top-k features, with k limited to 10. When the features are highly correlated, as in the current study, a feature set close to the square root of the sample size is often ideal for SVM (Hua et al., 2005), and the inclusion of too many features in a relatively small sample can lead to overfitting. The feature combination with the highest R^2 score was then used to train the model with

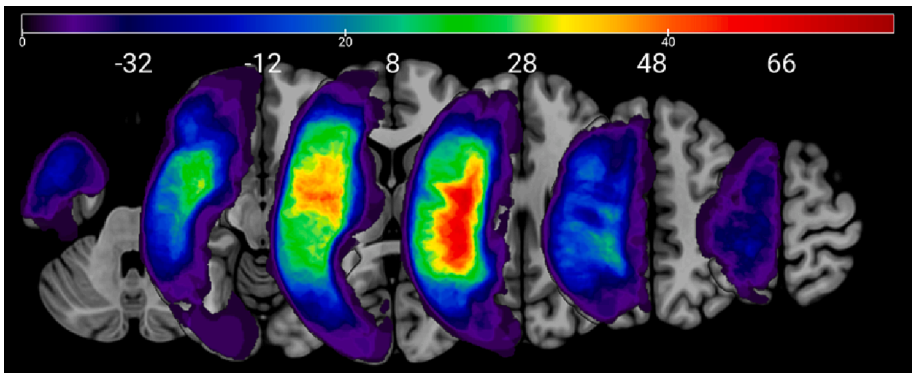

Fig. 1. Lesion incidence map (max = 54).

70 samples to predict the AQ for the one in the test set. We capped predicted AQ values at 100, and set the minimum to 20. Observed AQs below 20 are rare in clinical studies (Walker et al., 2022), and differences in numerical AQ below this cutoff are unlikely to be clinically relevant. This process was repeated with each participant in the test set once. Note that with this procedure, we avoid 'peeking', and no information about the left-out participant is used for feature selection. We then computed Pearson's correlation (r), root mean squared error (RMSE), and mean absolute error (MAE) between predicted AQ and observed AQ to evaluate estimation accuracy. The SVM model uses a hyper-parameter C for regularization. We varied C from 0.01 to 100 and chose the C that yielded the highest correlation coefficient.

We conducted analyses with (1) Discourse features only from each of the three prompts individually, and the combined set of features from the three prompts, (2) Lesion features combined with the discourse features in (1), and (3) lesion features only.

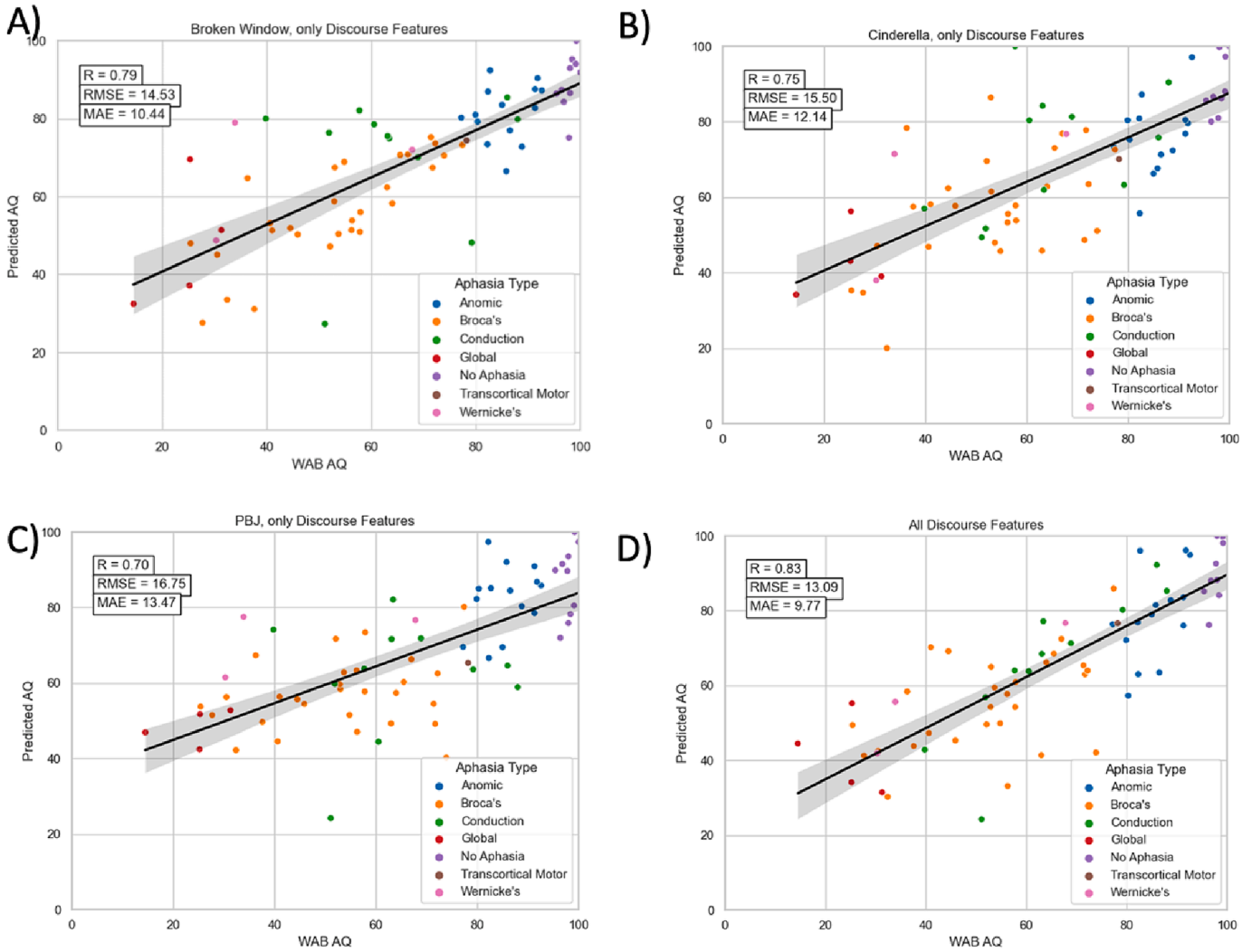

Fig. 2. AQ prediction using only discourse features; A) Broken Window, B) Cinderella, C) PBJ, and D) all features combined.

3. Results

Table 2 and Figs. 2-4 summarize the results. For each prompt, and for all sets of features (discourse only, discourse + lesion, lesion only), the correlation between predicted and actual AQ was significant (all p's < 0.001, Pearson's r range 0.64 – 0.83).

We used a two-tailed Hotelling's *t*-test for dependent correlations (Weiss, 2011) to examine whether any of the models were significantly better at predicting AQ. This test compares the Pearson's r between predicted and observed AQ for a given pair of models (e.g., Cinderella vs. Broken Window), while considering that the values come from the same group of participants. All features combined was significantly better than PBJ ($t(68) = 2.89$, $p = .005$). No other pairwise tests were significant.

When lesion features were added to discourse features, the performance for BrokenWindow, PBJ, and all features combined was not altered significantly. However, lesion features significantly boosted the performance of Cinderella, as determined by a Hotelling's *t*-test for dependent correlations ($t(68) = -2.05$, $p = .04$).

The lesion-feature only model was significantly outperformed by all models except for PBJ (with or without lesion) and Cinderella without lesion features (all p's < 0.05). We also calculated the 10 most informative features for each model (Table 4). Inspecting these features allows us to examine how informative linguistic features change depending on the prompt type.

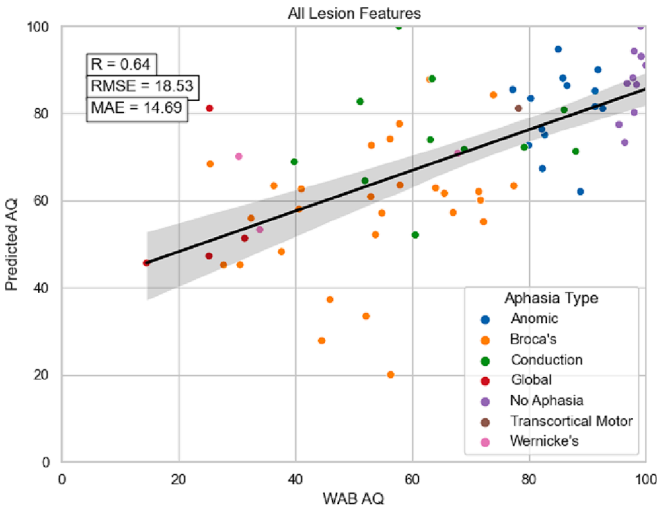

Fig. 4. AQ prediction using only lesion features.

4. Discussion

Here, we used features extracted from discourse analysis to quantify aphasia severity. Using these features, we were able to build models that

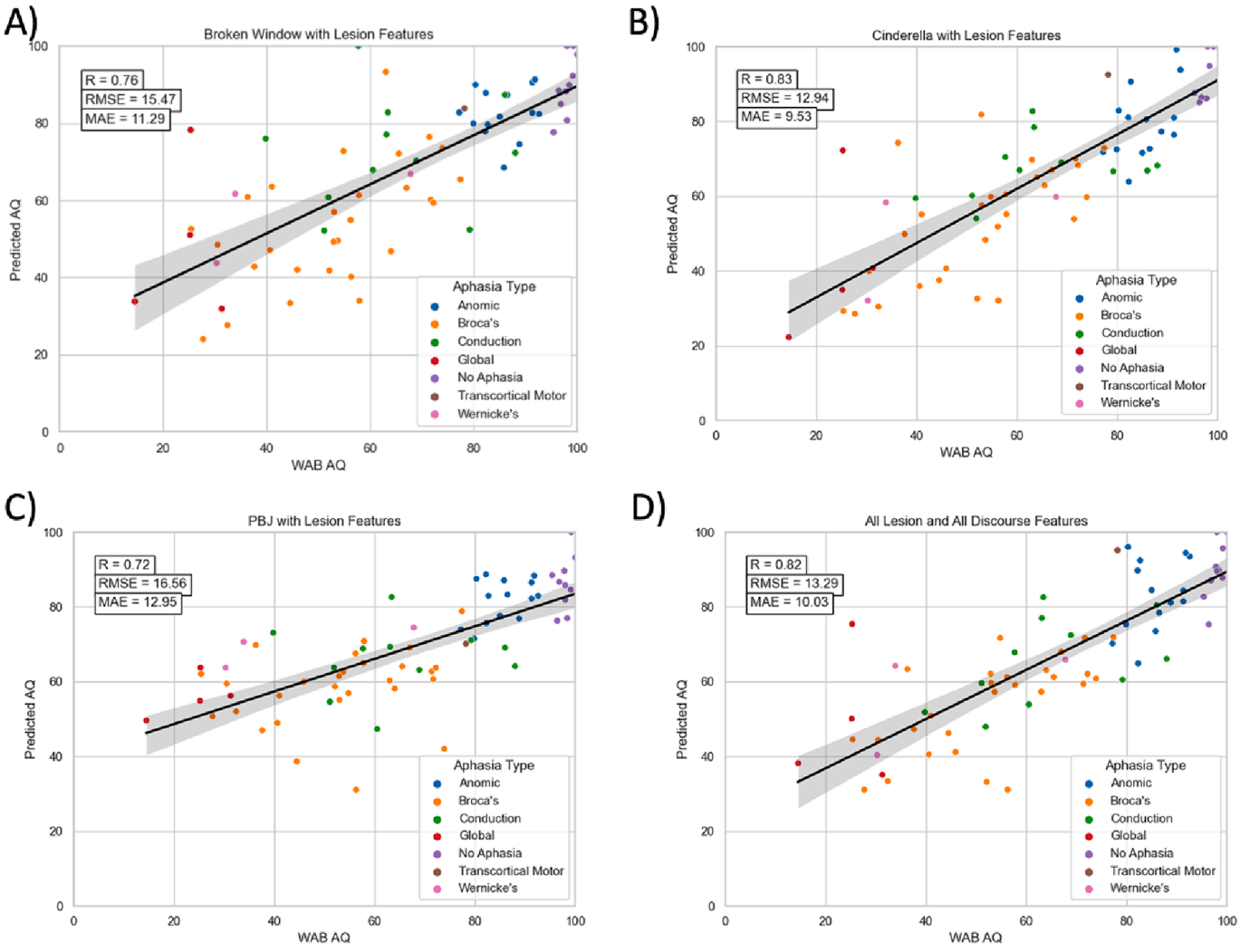

Fig. 3. AQ prediction using discourse plus lesion features.

Table 4
Top 10 features for each model. In parenthesis, the first number is the percent of times that feature was chosen as a top 10 (across 71 LOO cross validation folds), and the second number is the median rank that feature had. e.g., FreqTypes (100, 1) means that the number of different types of words used was a top 10 feature 100% of the time and had a median importance rank of #1.

Broken Window	Cinderella	PBJ	All Prompts Combined
% Det/Art (100, 1)	FreqTypes (100, 1)	% Prep (100, 1)	% Prep-PBJ (100, 2)
Total Utts (100, 3)	% PastPart (100, 2)	MLU Morphs (100, 2)	# WordErrors-BW (82, 3)
% Nouns (100, 4)	% Past (100, 3)	# Prep (100, 5)	# Nouns-BW (70, 1)
Density (100, 6)	% Nouns (100, 4)	% 3S (100, 8)	% Nouns-Cind (70, 6)
% Word Errors (99, 7)	% PresPart (97, 6)	# Nouns (99, 3)	% Conj-BW (66, 9)
Freq Types (97, 4)	Words/Min (92, 5)	Words/Min (99, 6)	MLU Utts-BW (59, 7)
MLU Utts (97, 7)	# Repetition (85, 9)	% Nouns (94, 6)	% Past-BW (51, 10)
Words/Min (87, 9)	% Det/Art (80, 8)	MLU Words (77, 6)	% Det/Art-PBJ (48, 10)
% 1S/3S (61, 10)	# PresPart (70, 7)	Verbs/Utt (49, 10)	MLU Utts-Cind (38, 12)
% Adj (58, 8)	MLU Morphs (56, 10)	% Past (49, 10)	% Det/Art-BW (34, 14)

Lesion Only	+ Broken Window	+ Cinderella	+ PBJ	+ All Prompts Combined
Superior Longitudinal Fasciculus (100, 1)	Superior Longitudinal Fasciculus (100, 1)	% Nouns (100, 3)	Superior Longitudinal Fasciculus (100, 1)	% Nouns-Cind (100, 4)
External Capsule (100, 2)	Density (100, 2)	% PastPart (100, 4)	MLU Morphs (100, 2)	FreqTypes-Cind (99, 2)
Middle Occipital Gyrus (100, 3)	% WordErrors (99, 3)	FreqTypes (97, 1)	% Prep (100, 3)	% Prep-PBJ (97, 3)
Splenium of Corpus Callosum (99, 4)	External Capsule (99, 4)	Middle Fronto-orbital Gyrus (97, 5)	# Nouns (100, 4)	Superior Longitudinal Fasciculus (94, 1)
Fusiform Gyrus (99, 5)	% Det/Art (93, 5)	Superior Longitudinal Fasciculus (90, 2)	Middle Occipital Gyrus (100, 8)	Posterior Superior Temporal Gyrus (72, 8)
Body of Corpus Callosum (79, 10)	% Nouns (83, 6)	% Past (89, 8)	Retrolenticular Part of Internal Capsule (99, 7)	# WordErrors-BW (61, 6)
Retrolenticular Part of Internal Capsule (77, 6)	# Repetition (75, 8)	% Det/Art (87, 6)	Words/Min (97, 6)	# Nouns-BW (61, 8)
Superior Corona Radiata (77, 9)	% Adj (39, 10)	Posterior Superior Temporal Gyrus (85, 7)	MLU Words (97, 8)	Middle Fronto-orbital Gyrus (54, 9)
Lenticular Fasciculus (75, 8)	Posterior Superior Temporal Gyrus (32, 10)	Caudate (59, 9)	% WordErrors (96, 6)	% Past-Cind (51, 10)
Middle Fronto-orbital Gyrus (68, 7)	Lenticular Fasciculus (30, 13)	# WordErrors (30, 11)	# Prep (39, 10)	% Det/Art-BW (25, 18)

provided person-specific aphasia severity estimates, with predicted AQ scores being significantly correlated with observed AQ. We also investigated which discourse or lesion features are most predictive of AQ. From these top features, we can draw conclusions about: 1) which linguistic (or lesion) features are most important for estimating aphasia severity, and 2) differences in how the language system is taxed by different prompt types.

4.1. Prompt types and discourse features

Using only discourse features, Broken Window had the highest prediction accuracy, while PBJ was numerically the lowest. Although this difference was not significant, it aligns well with the findings of Stark (2019), who suggested that PBJ has lower syntactic demands than narratives or expositional picture description tasks, even when inspecting discourse output from healthy adults. The top features for the PBJ model demonstrate that it captures somewhat different linguistic properties than Broken Window and Cinderella, especially related to the use of prepositions, which turned out to be the only feature selected 100 % of the time when combining all discourse features into a single model. This result is consistent with Stark and Fukuyama (2021), who found that prepositions were one of the main features that separated PBJ from other prompt types (expositional, narrative) when examining discourse output using between-class analysis, a dimension reduction technique.

Some patterns emerged relating to the task demands of expositional picture sequence description in the Broken Window prompt. First, the total number of different word types used, percent of nouns, and total utterances were important features, demonstrating that it encourages participants to display their general mastery of language by eliciting the use of different types of words and more utterances (Dalton & Richardson, 2015). This is perhaps also reflected in the importance of the determiners and articles feature, which may reflect preservation of grammar during speech (Matchin et al., 2020; Zhang & Hinzen, 2022).

Proposition density – a measure of content richness - being chosen in 100 % of Broken Window models was somewhat surprising. Past research has shown that while expositional tasks elicit the most diverse language, as measured by TTR, they elicit the lowest amount of content richness (measured by propositional density) of the 3 task types (expositional, narrative, procedural; Stark (2019)). However, its inclusion in the model suggests that, while picture description may not elicit particularly high proposition density, proposition density in an individual person's expositional discourse sample provides information about their aphasia severity. This finding can also be linked to Fromm et al. (2016), who found that proposition density differed in people with aphasia compared to controls, but that there was no straightforward relationship between proposition density and aphasia severity. This was due to fluent and non-fluent participants having equivalent AQ but dissimilar proposition density. The multivariate, machine learning approach used in the current manuscript was able to leverage information captured by proposition density, in combination with other features, to predict AQ. A caveat here being that proposition density in the current manuscript was informative for an expositional task, while the findings of Fromm et al. (2016) were from a narrative task (see further discussion in Limitations). Finally, word errors were also important for Broken Window models, reflecting the naming processes elicited by picture description tasks (e.g., having to name various objects or characters in the picture).

Similar to Broken Window, an important feature elicited by Cinderella for predicting AQ was the number of different word types used. However, results suggest that the usage of the past tense is what separates Cinderella from the other prompt types, with past tense and past participle use being among the most important features for predicting AQ in Cinderella-based models. Several studies have found that production and comprehension of past tense can be especially difficult for people with aphasia (Faroqi-Shah & Friedman, 2015; Jonkers & de Bruin, 2009; Ullman et al., 2005). Cinderella, a narrative recall task,

forces participants to use the past tense in their retelling of the Cinderella story, while Broken Window and PBJ can be completed using the present tense.

The findings demonstrate that, while each prompt can be used to predict AQ, the top ten features differ – providing insights about the unique linguistic demands of each prompt type. Indeed, combining all features from all prompts into a single model yielded numerically the highest AQ prediction accuracy, suggesting that the prompts make unique contributions to aphasia severity estimation. However, this comes with the drawback that there is less consistency among the top ten features chosen (evidenced by lower percentages and more variable median ranks for the top ten features), due to the expanded feature selection space. This could be ameliorated by simply using each discourse model individually, and then averaging the predicted AQs together for each participant. This maintains top ten feature consistency from each prompt type, while also allowing each prompt to contribute to prediction. We tested this averaging method in a supplementary analysis (Supplementary Materials, Fig. 1), and it yielded prediction accuracies virtually identical to the results generated from combining all features from all prompts into a single model (Fig. 2D and 3D).

4.2. Lesion features and aphasia severity

We also investigated the relative importance of the lesion features in AQ assessment. The superior longitudinal fasciculus (SLF) was the most frequently selected top ranked feature when only lesion features was used. Moreover, SLF is among the most frequently selected top two features even in the discourse plus lesion models. The SLF is a white matter tract that connects portions of the occipital, posterior temporal, and parietal lobes to the frontal cortex (Bernal & Altman, 2010; Kamali et al., 2014). Our finding that the SLF is an important feature for predicting aphasia severity aligns with previous research demonstrating that degradation of the SLF in a variety of etiologies has been linked to impaired language or executive abilities that contribute to language (Madhavan et al., 2014; Nagae et al., 2012; Rizio & Diaz, 2016; Shinoura et al., 2013).

The other features chosen 100 % of the time as a top 10 feature in the lesion-only model, the external capsule (EC) and middle occipital gyrus (MOG), are somewhat surprising as they are not considered classic 'language areas' in most neurobiological models (Desai & Riccardi, 2021; Hickok & Poeppel, 2004). However, EC integrity has been implicated in executive dysfunction (Nolze-Charron et al., 2020), and is considered by some to be a part of the ventral language stream (Axer et al., 2013), although this is debated. EC tracts are also adjacent to portions of SLF (Schmahmann et al., 2009), raising the possibility that these two pathways are commonly damaged together in stroke affecting the middle cerebral artery. It is also possible that the EC contributes to language via subcortical connections that support language either directly or through domain-general processes (Kuljic-Obradovic, 2003; Sharif et al., 2022). The MOG, on the other hand, may be related to visual identification of items and objects near the 'beginning' of the ventral language stream (Fridriksson et al., 2016; Hickok & Poeppel, 2004; Hickok & Poeppel, 2016). People with MOG damage may perform poorly on visual aspects of the WAB-R such as object naming or picture description, making it an informative feature when predicting aphasia severity.

Aphasia severity, as measured by the WAB-R, is measured using three principal criteria: production (spontaneous speech and naming), comprehension, and repetition. Production accounts for 60 % of one's AQ score, with repetition and comprehension comprising an additional 20 % each. The discourse production task nominally includes only production (with perhaps some overlap with repetition), and hence might be expected to perform relatively poorly in estimating AQ. The success of discourse tasks in predicting AQ suggests that discourse production is a rich and multi-faceted task, while also aligning well with the WAB-R's relatively heavier weighting of production skills compared to comprehension.

While it is somewhat surprising that adding lesion features to the discourse models did not boost the accuracy of aphasia severity estimation, it demonstrates the effectiveness of discourse tasks in estimating AQ. The discourse prompts require some of the same language skills that are measured by WAB-R (e.g., picture description and object naming), which was sufficient for discourse tasks to have a high predictive value, as suggested above. Lesion features, on the other hand, are comparatively more 'indirect' representatives of language ability. When considering the future use of discourse features to estimate aphasia severity, it is a net positive that lesion features do not contribute significantly above and beyond discourse features. If discourse features alone could not estimate aphasia severity and MRI scans were required, then it would negate the advantage of the discourse method as less demanding in terms of resources.

4.3. Limitations and future directions

One limitation, which holds true for many discourse-related investigations, pertains to the discussion of findings in comparison to other studies that used different discourse prompts or extracted different speech features. Although it is typically agreed that there are three 'families' of discourse tasks (expositional, narrative, procedural), there are many different flavors within each family which may lead to discrepancies between studies. For example, although we used Broken Window as our expositional prompt, the use of a different expositional prompt, such as Cat Rescue, may produce different results. Investigations such as Stark and Fukuyama (2021) have demonstrated that different prompts within a given family tend to elicit similar speech features, but our specific findings using Broken Window, Cinderella, and PBJ may not hold true for other expositional, narrative, or procedural prompts, respectively. Future studies could seek to replicate our findings with other prompts.

Another difficulty in comparing results between studies are the compositions of participant samples. For example, our study included 'controls' (survivors of stroke with no language impairment), while other studies may only include those with aphasia. There are also differences between studies in the proportion of the various aphasia subtypes, as well as mean AQ scores, which are factors that may have complex relationships with discourse features. Future work could seek to replicate and expand these findings in diverse cohorts of people with aphasia.

Here, our focus was on using mostly microstructural discourse features to estimate aphasia severity. Adding macro-level features such as main concept analysis or demographic features could boost the model prediction (Johnson et al., 2022). Regarding anatomical features, it is possible that other measures of brain health, such as resting state connectivity (Kristinsson et al., 2021) or brain age (Busby et al., 2023; Kristinsson et al., 2022) could also be useful estimators of aphasia severity. Understanding how these factors relate to aphasia severity could improve our understanding of the neuroanatomical and behavioral correlates of aphasia. It could also lay the groundwork for new behavioral or neurostimulation-based interventions, or for building models designed to predict long-term language recovery trajectories post-stroke, as opposed to AQ at a single time point.

Clinical use of discourse-based aphasia severity estimation relies on improved automation of transcription and coding of impaired speech in the coming years, as current automated transcription methods perform relatively poorly in people with aphasia (Mahmoud et al., 2023). Another goal for future work would be to develop a'pre-trained' model, which could be trained on a large amount of discourse from people with aphasia, and then disseminated to researchers and clinicians for aphasia quotient estimation without the need for them to develop or train their own models. These steps towards automation and ease-of-use would be necessary before discourse-based aphasia quotient estimation could be used in clinical settings.

Future work could also investigate how, instead of prompts, other naturalistic discourse paradigms could be used to assess language abilities and their neural correlates (Birba et al., 2022; Riccardi & Desai, 2022). Finally, in the current study, even though the prediction was accurate overall with Pearson's correlation between measured and predicted AQ near 0.8, and the models were accurate for the majority of the participants, they were relatively inaccurate for a handful of cases. Understanding the characteristics of individuals that lead to lower model prediction performance may help improve models even further.

5. Conclusion

The present study showed that microstructural features elicited from three AphasiaBank discourse prompts can be used to estimate aphasia severity. Even a single prompt, containing only a few minutes (or sometimes less than a minute) of speech output, was sufficient to estimate AQ reasonably well for most individuals. Each prompt elicited different informative features, demonstrating potential differences between prompts. Lesion features can also be used to estimate aphasia severity, although with lower accuracy than the discourse-based models. An important role for superior longitudinal fasciculus integrity in aphasia severity is suggested. Discourse-based aphasia severity estimation is promising as a supplemental language measurement that is ecologically valid and less resource-intensive. The current study provides important first steps towards mapping how discourse features can quantify aphasia severity.

CRediT authorship contribution statement

Nicholas Riccardi: Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Satvik Nelakuditi:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Dirk B. den Ouden:** Writing – review & editing, Resources, Project administration, Funding acquisition, Data curation. **Chris Rorden:** Software, Resources, Project administration, Funding acquisition, Data curation. **Julius Fridriksson:** Resources, Project administration, Funding acquisition. **Rutvik H. Desai:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by NIH/NIDCD R01DC017162 and P50DC014664.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2024.103602.

References

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26 (3), 839–851. https://doi.org/10.1016/j.neuroimage.2005.02.018.

Axer, H., Klingner, C.M., Prescher, A., 2013. Fiber anatomy of dorsal and ventral language streams. Brain Lang. 127 (2), 192–204. https://doi.org/10.1016/j.bandl.2012.04.015.

Bernal, B., Altman, N., 2010. The connectivity of the superior longitudinal fasciculus: a tractography DTI study. Magn. Reson. Imaging 28 (2), 217–225. https://doi.org/10.1016/j.mri.2009.07.008.

Birba, A., Fittipaldi, S., Cediel Escobar, J.C., Gonzalez Campo, C., Legaz, A., Galiani, A., Diaz Rivera, M.N., Martorell Caro, M., Alifano, F., Pina-Escudero, S.D., Cardona, J.F., Neely, A., Forno, G., Carpinella, M., Slachevsky, A., Serrano, C., Sedeno, L., Ibanez, A., Garcia, A.M., 2022. Multimodal neurocognitive markers of naturalistic discourse typify diverse neurodegenerative diseases. Cereb. Cortex 32 (16), 3377–3391. https://doi.org/10.1093/cercor/bhab421.

Bryant, L., Spencer, E., Ferguson, A., Craig, H., Colyvas, K., Worrall, L., 2013. Propositional idea density in aphasic discourse. Aphasiology 27 (8), 992–1009.

Bryant, L., Ferguson, A., Spencer, E., 2016. Linguistic analysis of discourse in aphasia: a review of the literature. Clin. Linguist. Phon. 30 (7), 489–518. https://doi.org/10.3109/02699206.2016.1145740.

Bullier, B., Cassoudesalle, H., Villain, M., Cogne, M., Mollo, C., De Gabory, I., Dehail, P., Joseph, P.A., Sibon, I., Glize, B., 2020. New factors that affect quality of life in patients with aphasia. Ann. Phys. Rehabil. Med. 63 (1), 33–37. https://doi.org/10.1016/j.rehab.2019.06.015.

Busby, N., Wilmskoetter, J., Gleichgerrcht, E., Rorden, C., Roth, R., Newman-Norlund, R., Hillis, A. E., Keller, S. S., de Bezenac, C., Kristinsson, S., Fridriksson, J., & Bonilha, L. (2023). Advanced Brain Age and Chronic Poststroke Aphasia Severity. Neurology, 100(11), e1166-e1176. https://doi.org/10.1212/WNL.0000000000201693.

Choi, Y., Park, H.K., Ahn, K., Son, Y., Paik, N., 2015. A telescreening tool to detect apahsia in patients with stroke. Telemedicine and e-Health 21 (9), 729–734.

Cruice, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., Dipper, L., 2020. UK speech and language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. Int. J. Lang. Commun. Disord. 55 (3), 417–442.

Dalton, S.G., & Richardson, J.D. (2015). Core-lexicon and main-concept production during picture-sequence description in adults without brain damage and adults with aphasia. Am J Speech Lang Pathol, 24(4), S923-938. https://doi.org/10.1044/2015_AJSLP-14-0161.

Dalton, S.G.H., Richardson, J.D., 2019. A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. Am. J. Speech Lang. Pathol. 28 (1S), 293–320. https://doi.org/10.1044/2018_AJSLP-17-0166.

Dalton, S.G., Stark, B.C., Fromm, D., Apple, K., MacWhinney, B., Rensch, A., Rowedder, M., 2022. Validation of an automated procedure for calculating core lexicon from transcripts. J. Speech Lang. Hear. Res. 65 (8), 2996–3003. https://doi.org/10.1044/2022_JSLHR-21-00473.

Desai, R.H., & Riccardi, N. (2021). Cognitive neuroscience of language. In The Routledge handbook of cognitive linguistics (pp. 615-642).

Faria, A.V., Joel, S.E., Zhang, Y., Oishi, K., van Zjil, P.C., Miller, M.I., Pekar, J.J., Mori, S., 2012. Atlas-based analysis of resting-state functional connectivity: evaluation for reproducibility and multi-modal anatomy-function correlation studies. Neuroimage 61 (3), 613–621. https://doi.org/10.1016/j.neuroimage.2012.03.078.

Faroqi-Shah, Y., & Friedman, L. (2015). Production of verb tense in agrammatic aphasia: A meta-analysis and further data. Behavioural neurology, 2015.

Fraser, K.C., Meltzer, J.A., Graham, N.L., Leonard, C., Hirst, G., Black, S.E., Rochon, E., 2014. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 55, 43–60. https://doi.org/10.1016/j.cortex.2012.12.006.

Fridriksson, J., Yourganov, G., Bonilha, L., Basilakos, A., Den Ouden, D.B., Rorden, C., 2016. Revealing the dual streams of speech processing. PNAS 113 (52), 15108–15113. https://doi.org/10.1073/pnas.1614038114.

Fridriksson, J., den Ouden, D.B., Hillis, A.E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., Bonilha, L., 2018. Anatomy of aphasia revisited. Brain 141 (3), 848–862. https://doi.org/10.1093/brain/awx363.

Fromm, D., Greenhouse, J., Hou, K., Russell, G.A., Cai, X., Forbes, M., Holland, A., MacWhinney, B., 2016. Automated proposition density analysis for discourse in aphasia. J. Speech Lang. Hear. Res. 59 (5), 1123–1132. https://doi.org/10.1044/2016_JSLHR-L-15-0401.

Fromm, D., Forbes, M., Holland, A., MacWhinney, B., 2020. Using aphasiabank for discourse assessment. Semin. Speech Lang. 41 (1), 10–19. https://doi.org/10.1055/s-0039-3399499.

Galski, T., Tompkins, C., Johnston, M., 1998. Competence in discourse as a measure of social integration and quality of life in persons with traumatic brain injury. Brain Inj. 12 (9), 769–782.

Gordon, J.K., 2008. Measuring the lexical semantics of picture description in aphasia. Aphasiology 22 (7–8), 839–852.

Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92 (1–2), 67–99. https://doi.org/10.1016/j.cognition.2003.10.011.

Hickok, G., Poeppel, D., 2016. Neural basis of speech perception. Neurobiology of Language 299–310.

Hillis, A.E., Rorden, C., Fridriksson, J., 2017. Brain regions essential for word comprehension: Drawing inferences from patients. Ann. Neurol. 81 (6), 759–768. https://doi.org/10.1002/ana.24941.

Howard, D., & Patterson, K.E. (1992). The pyramids and palm trees test.

Hua, J., Xiong, Z., Lowey, J., Suh, E., Dougherty, E.R., 2005. Optimal number of features as a function of sample size for various classification rules. Bioinformatics 21 (8), 1509–1515.

Johnson, L., Nemati, S., Bonilha, L., Rorden, C., Busby, N., Basilakos, A., Newman-Norlund, R., Hillis, A.E., Hickok, G., Fridriksson, J., 2022. Predictors beyond the lesion: health and demographic factors associated with aphasia severity. Cortex 154, 375–389. https://doi.org/10.1016/j.cortex.2022.06.013.

Jonkers, R., de Bruin, A., 2009. Tense processing in Broca's and Wernicke's aphasia. Aphasiology 23 (10), 1252–1265.
Kamali, A., Flanders, A.E., Brody, J., Hunter, J.V., Hasan, K.M., 2014. Tracing superior longitudinal fasciculus connectivity in the human brain using high resolution diffusion tensor tractography. Brain Struct. Funct. 219 (1), 269–281. https://doi.org/10.1007/s00429-012-0498-y.
Kertesz, A., 2007. Western Aphasia Battery-Revised. Pearson.
Kertesz, A., 2022. The western aphasia battery: a systematic review of research and clinical applications. Aphasiology 36 (1), 21–50.
Kong, A.-P.-H., 2009. The use of main concept analysis to measure discourse production in Cantonese-speaking persons with aphasia: a preliminary report. J. Commun. Disord. 42 (6), 442–464.
Kong, A.-P.-H., Whiteside, J., Bargmann, P., 2016. The main concept analysis: Validation and sensitivity in differentiating discourse produced by unimpaired English speakers from individuals with aphasia and dementia of Alzheimer type. Logoped. Phoniatr. Vocol. 41 (3), 129–141.
Kristinsson, S., Thors, H., Yourganov, G., Magnusdottir, S., Hjaltason, H., Stark, B.C., Basilakos, A., den Ouden, D.B., Bonilha, L., Rorden, C., Hickok, G., Hillis, A., Fridriksson, J., 2020. Brain damage associated with impaired sentence processing in acute aphasia. J. Cogn. Neurosci. 32 (2), 256–271. https://doi.org/10.1162/jocn_a_01478.
Kristinsson, S., Zhang, W., Rorden, C., Newman-Norlund, R., Basilakos, A., Bonilha, L., Yourganov, G., Xiao, F., Hillis, A., Fridriksson, J., 2021. Machine learning-based multimodal prediction of language outcomes in chronic aphasia. Hum. Brain Mapp. 42 (6), 1682–1698. https://doi.org/10.1002/hbm.25321.
Kristinsson, S., Busby, N., Rorden, C., Newman-Norlund, R., den Ouden, D.B., Magnusdottir, S., Hjaltason, H., Thors, H., Hillis, A.E., Kjartansson, O., Bonilha, L., Fridriksson, J., 2022. Brain age predicts long-term recovery in post-stroke aphasia. Brain Commun 4 (5), fcac252. https://doi.org/10.1093/braincomms/fcac252.
Kuljic-Obradovic, D.C., 2003. Subcortical aphasia: three different language disorder syndromes? Eur. J. Neurol. 10 (4), 445–448. https://doi.org/10.1046/j.1468-1331.2003.00604.x.
Le, D., Licata, K., Provost, E.M., 2017. 2017. A Preliminary Study Interspeech, Automatic Paraphasia Detection from Aphasic Speech.
Liu, H., MacWhinney, B., Fromm, D., Lanzi, A., 2023. Automation of language sample analysis. J. Speech Lang. Hear. Res. 1–13.
Macwhinney, B., Fromm, D., Forbes, M., Holland, A., 2011. AphasiaBank: methods for studying discourse. Aphasiology 25 (11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893.
MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. transcription format and programs (Vol. 1). Psychology Press.
Madhavan, K.M., McQueeny, T., Howe, S.R., Shear, P., Szaflarski, J., 2014. Superior longitudinal fasciculus and language functioning in healthy aging. Brain Res. 1562, 11–22. https://doi.org/10.1016/j.brainres.2014.03.012.
Magnusdottir, S., Fillmore, P., den Ouden, D.B., Hjaltason, H., Rorden, C., Kjartansson, O., Bonilha, L., Fridriksson, J., 2013. Damage to left anterior temporal cortex predicts impairment of complex syntactic processing: a lesion-symptom mapping study. Hum. Brain Mapp. 34 (10), 2715–2723. https://doi.org/10.1002/hbm.22096.
Mahmoud, S.S., Kumar, A., Li, Y., Tang, Y., Fang, Q., 2021. Performance evaluation of machine learning frameworks for aphasia assessment. Sensors 21 (8), 2582.
Mahmoud, S.S., Pallaud, R.F., Kumar, A., Faisal, S., Wang, Y., Fang, Q., 2023. A comparative investigation of automatic speech recognition platforms for aphasia assessment batteries. Sensors (Basel) 23 (2). https://doi.org/10.3390/s23020857.
Matchin, W., Basilakos, A., Stark, B.C., den Ouden, D.B., Fridriksson, J., Hickok, G., 2020. Agrammatism and paragrammatism: a cortical double dissociation revealed by lesion-symptom mapping. Neurobiol Lang (camb) 1 (2), 208–225. https://doi.org/10.1162/nol_a_00010.
Mirman, D., Kraft, A.E., Harvey, D.Y., Brecher, A.R., Schwartz, M.F., 2019. Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. Cognitive, Affective, & Behavioral Neuroscience 19, 1286–1298.
Mori, S., Wakana, S., Van Zijl, P.C., Nagae-Poetscher, L., 2005. MRI atlas of human white matter. Elsevier.
Nachev, P., Coulthard, E., Jager, H.R., Kennard, C., Husain, M., 2008. Enantiomorphic normalization of focally lesioned brains. Neuroimage 39 (3), 1215–1226. https://doi.org/10.1016/j.neuroimage.2007.10.002.
Nagae, L.M., Zarnow, D.M., Blaskey, L., Dell, J., Khan, S.Y., Qasmieh, S., Levy, S.E., Roberts, T.P., 2012. Elevated mean diffusivity in the left hemisphere superior longitudinal fasciculus in autism spectrum disorders increases with more profound language impairment. AJNR Am. J. Neuroradiol. 33 (9), 1720–1725. https://doi.org/10.3174/ajnr.A3037.
Nolze-Charron, G., Dufort-Rouleau, R., Houde, J.C., Dumont, M., Castellano, C.A., Cunnane, S., Lorrain, D., Fulop, T., Descoteaux, M., Bocti, C., 2020. Tractography of the external capsule and cognition: a diffusion MRI study of cholinergic fibers. Exp. Gerontol. 130, 110792 https://doi.org/10.1016/j.exger.2019.110792.
Riccardi, N., & Desai, R.H. (2022). Discourse and the brain. In The Routledge Handbook of Semiosis and the Brain (pp. 174-189). https://doi.org/10.4324/9781003051817-14.
Riccardi, N., Yourganov, G., Rorden, C., Fridriksson, J., Desai, R.H., 2019. Dissociating action and abstract verb comprehension post-stroke. Cortex 120, 131–146. https://doi.org/10.1016/j.cortex.2019.05.013.
Riccardi, N., Yourganov, G., Rorden, C., Fridriksson, J., Desai, R., 2020. Degradation of praxis brain networks and impaired comprehension of manipulable nouns in stroke. J. Cogn. Neurosci. 32 (3), 467–483. https://doi.org/10.1162/jocn_a_01495.
Riccardi, N., Rorden, C., Fridriksson, J., Desai, R.H., 2022. Canonical sentence processing and the inferior frontal cortex: is there a connection? Neurobiol Lang (Camb) 3 (2), 318–344. https://doi.org/10.1162/nol_a_00067.
Riccardi, N., Zhao, X., den Ouden, D.-B., Fridriksson, J., Desai, R.H., Wang, Y., 2023. Network-based statistics distinguish anomic and Broca's aphasia. Brain Struct. Funct. 1–17.
Rizio, A.A., Diaz, M.T., 2016. Language, aging, and cognition: frontal aslant tract and superior longitudinal fasciculus contribute toward working memory performance in older adults. Neuroreport 27 (9), 689–693. https://doi.org/10.1097/WNR.0000000000000597.
Roach, A., Schwartz, M., Martin, N., Grewal, R.S., Brecher, A., 1996. The philadelphia naming test: scoring and rationale. Clinical Aphasiology 24, 121–133.
Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., Karnath, H.O., 2012. Age-specific CT and MRI templates for spatial normalization. Neuroimage 61 (4), 957–965. https://doi.org/10.1016/j.neuroimage.2012.03.020.
Schmahmann, J. D., Schmahmann, J., & Pandya, D. (2009). Fiber pathways of the brain. OUP USA.
Schwen Blackett, D., Varkey, J., Wilmskoetter, J., Roth, R., Andrews, K., Busby, N., Gleichgerrcht, E., Desai, R.H., Riccardi, N., Basilakos, A., Johnson, L.P., Kristinsson, S., Johnson, L., Rorden, C., Spell, L.A., Fridriksson, J., Bonilha, L., 2022. Neural network bases of thematic semantic processing in language production. Cortex 156, 126–143. https://doi.org/10.1016/j.cortex.2022.08.007.
Sharif, M.S., Goldberg, E.B., Walker, A., Hillis, A.E., Meier, E.L., 2022. The contribution of white matter pathology, hypoperfusion, lesion load, and stroke recurrence to language deficits following acute subcortical left hemisphere stroke. PLoS One 17 (10), e0275664.
Shinoura, N., Midorikawa, A., Onodera, T., Tsukada, M., Yamada, R., Tabei, Y., Itoi, C., Saito, S., Yagi, K., 2013. Damage to the left ventral, arcuate fasciculus and superior longitudinal fasciculus-related pathways induces deficits in object naming, phonological language function and writing, respectively. Int. J. Neurosci. 123 (7), 494–502. https://doi.org/10.3109/00207454.2013.765420.
Spaccavento, S., Craca, A., Del Prete, M., Falcone, R., Colucci, A., Di Palma, A., Loverre, A., 2014. Quality of life measurement and outcome in aphasia. Neuropsychiatr. Dis. Treat. 10, 27–37. https://doi.org/10.2147/NDT.S52357.
Spell, L.A., Richardson, J.D., Basilakos, A., Stark, B.C., Teklehaimanot, A., Hillis, A.E., Fridriksson, J., 2020. Developing, implementing, and improving assessment and treatment fidelity in clinical aphasia research. Am. J. Speech Lang. Pathol. 29 (1), 286–298. https://doi.org/10.1044/2019_AJSLP-19-00126.
Stark, B.C., 2019. A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. Am. J. Speech Lang. Pathol. 28 (3), 1067–1083.
Stark, B.C., Fukuyama, J., 2021. Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. Language, Cognition and Neuroscience 36 (5), 562–585.
Thompson, C.K. (2012). Northwestern assessment of verbs and sentences (NAVS).
Ullman, M.T., Pancheva, R., Love, T., Yee, E., Swinney, D., & Hickok, G. (2005). Neural correlates of lexicon and grammar: evidence from the production, reading, and judgment of inflection in aphasia. Brain Lang, 93(2), 185-238; discussion 239-142. https://doi.org/10.1016/j.bandl.2004.10.001.
Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C., Mori, S., 2004. Fiber tract-based atlas of human white matter anatomy. Radiology 230 (1), 77–87. https://doi.org/10.1148/radiol.2301021640.
Walker, G.M., Fridriksson, J., Hillis, A.E., Den Ouden, D.B., Bonilha, L., Hickok, G., 2022. The severity-calibrated aphasia naming test. Am. J. Speech Lang. Pathol. 31 (6), 2722–2740.
Weiss, B.A., 2011. Hotelling's t Test and Steiger's Z test calculator. In. https://blogs.gwu.edu/weissba/teaching/calculators/hotellings-t-and-steigers-z-tests/.
Wilson, S.M., Eriksson, D.K., Schneck, S.M., Lucanie, J.M., 2018. A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. PLoS One 13 (2), e0192773.
Zhang, H., Hinzen, W., 2022. Grammar in 'agrammatical'aphasia: what's intact? PLoS One 17 (12), e0278676.