

Table S1. Description of monologue speech task

Thematic category	Prompt (English equivalent)
Description of the previous day	<i>"Try to describe your day yesterday in as much detail as possible."</i>
Favorite places/food/activities	<i>"Describe the town/village where you live. What interesting places nearby would you recommend to someone visiting from another region?"</i>
	<i>"Describe the town/village where you spent your childhood. What interesting places nearby would you recommend to someone visiting from another region?"</i>
	<i>"What was your most interesting summer vacation? What was the itinerary, and what interesting places did you visit?"</i>
	<i>"What is your favorite food? How often and where do you eat it? Who prepares it for you? Which local restaurant would you recommend?"</i>
	<i>"Describe your favorite TV show/series. What type of movies do you like to watch?"</i>
	<i>"What is your favorite hobby? What do you like to do in your free time?"</i>
	<i>"Which school did you attend in the past? Which school subjects did you enjoy or dislike? Which classmates do you fondly remember?"</i>
Hypothetical reasoning, imagination, and autobiographical memory	<i>"What was the happiest moment in your life?"</i>
	<i>"What was the silliest thing you have ever done?"</i>
	<i>"What is your biggest bad habit?"</i>
	<i>"Which person in your life has influenced you the most, and why?"</i>
	<i>"If you could be an animal, which one would it be and why?"</i>
	<i>"If you won 20 million in the lottery, how would you use the winnings?"</i>
	<i>"If you could live in another historical era, which one would it be and why?"</i>

Each session included three questions. Each question was chosen from one thematic category.

Table S2. Description of retelling speech task

Language	Fairy tale (English equivalent)	Number of words in the reference fairy tale
Czech	<i>Hansel and Gretel</i>	269
	<i>Sleeping Beauty</i>	225
	<i>The Emperor's New Clothes</i>	235
	<i>Snow White and the Seven Dwarfs</i>	258
	<i>Cinderella</i>	282
	<i>The Three Billy Goats Gruff</i>	262
	<i>Little Red Riding Hood</i>	267
German	<i>Hansel and Gretel</i>	210
	<i>Sleeping Beauty</i>	223
	<i>The Emperor's New Clothes</i>	228
	<i>Rumpelstiltskin</i>	276
	<i>The Bremen Town Musicians</i>	221
	<i>The Frog King</i>	261
	<i>Rapunzel</i>	264

Each session included one fairy tale.

Table S3 Predictive performance of linguistic features for clinical markers with adjustment for Diagnostic Confidence Level variable as predictor in monologue tasks

Clinical scale	Predictors selected through feature selection (β , p-value)	R^2	ΔR^2	p-value
UHDRS-TMS	Syntactic complexity ($\beta=-0.61$, $p<0.001$) Czech/German language ($\beta=0.73$, $p<0.001$) DCL ($\beta=0.53$, $p<0.001$)	0.76	0.18	<0.001
UHDRS-TFC	Vocabulary range ($\beta=0.42$, $p<0.001$) Phrase patterns ($\beta=0.36$, $p=0.0071$) Sentence length ($\beta=0.41$, $p=0.0057$) DCL ($\beta=-0.50$, $p=0.0021$)	0.63	0.25	<0.001
SDMT	Sentence length ($\beta=1.08$, $p<0.001$) Sentence development ($\beta=-0.89$, $p<0.001$) Czech/German language ($\beta=-0.36$, $p=0.0083$) DCL ($\beta=-0.43$, $p=0.0021$)	0.62	0.22	<0.001
Lexical verbal fluency	Phrase patterns ($\beta=-0.20$, $p=0.19$) Sentence length ($\beta=1.50$, $p<0.001$) Sentence development ($\beta=-0.82$, $p=0.010$) Syntactic complexity ($\beta=-0.86$, $p<0.001$)	0.41	0.24	<0.001
Categorical verbal fluency	Sentence length ($\beta=0.90$, $p<0.001$) Sentence development ($\beta=-0.45$, $p=0.044$) Czech/German language ($\beta=-0.29$, $p=0.035$) DCL ($\beta=-0.37$, $p=0.0094$)	0.61	0.19	<0.001
Stroop Color Test	Vocabulary range ($\beta=0.35$, $p=0.020$) Reference rate to reality ($\beta=-0.30$, $p=0.045$) Sentence length ($\beta=0.39$, $p=0.011$) DCL ($\beta=-0.41$, $p=0.0090$)	0.54	0.17	<0.001
Stroop Word Test	Vocabulary range ($\beta=0.14$, $p=0.29$) Sentence length ($\beta=1.02$, $p<0.001$) Sentence development ($\beta=-0.70$, $p=0.0052$) DCL ($\beta=-0.25$, $p=0.12$)	0.50	0.30	<0.001
Stroop Interference Test	Vocabulary range ($\beta=0.15$, $p=0.32$) Sentence length ($\beta=0.62$, $p=0.030$) Sentence development ($\beta=-0.37$, $p=0.17$) DCL ($\beta=-0.41$, $p=0.026$)	0.37	0.08	0.002
Trail Making Test A	Phrase patterns ($\beta=0.35$, $p=0.0094$) Sentence length ($\beta=-1.29$, $p<0.001$) Sentence development ($\beta=1.13$, $p<0.001$)	0.56	0.43	<0.001
Trail Making Test B	Phrase patterns ($\beta=0.32$, $p=0.031$) Sentence length ($\beta=-0.93$, $p=0.013$) Sentence development ($\beta=0.96$, $p=0.0031$) Syntactic complexity ($\beta=-0.23$, $p=0.25$)	0.45	0.17	<0.001
CAG	Sentence length ($\beta=-0.24$, $p=0.16$) Age ($\beta=-0.49$, $p=0.0015$) DCL ($\beta=0.35$, $p=0.040$)	0.38	0.004	0.002

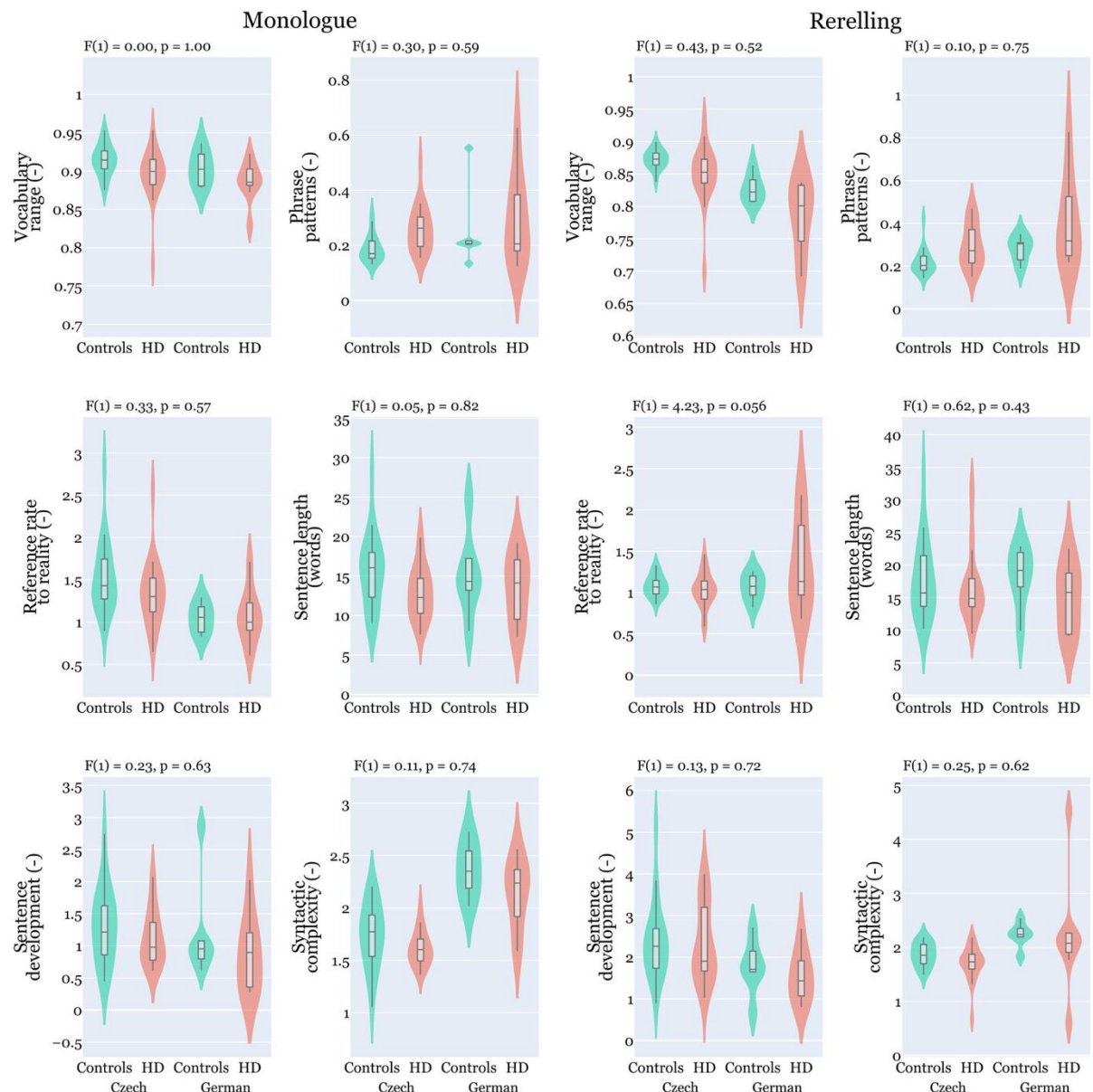
β = standardized beta value of individual predictors, ΔR^2 = increase in R^2 from the model including only demographic variables (age, sex, and Czech/German language) and DCL, and the model additionally including linguistic features, DCL = Diagnostic Confidence Level, UHDRS-TMS = Unified Huntington's Disease Rating Scale Total Motor Score, UHDRS-TFC = Unified Huntington's Disease Rating Scale Total Functional Capacity, SDMT = Symbol Digit Modalities Test, CAG = cytosine-adenine-guanine.

Table S4 Predictive performance of linguistic features for clinical markers with adjustment for Diagnostic Confidence Level variable as predictor in retelling tasks

Clinical scale	Predictors selected through feature selection (β , p-value)	R^2	ΔR^2	p-value
UHDRS-TMS	Vocabulary range ($\beta=-0.20$, $p=0.047$) Phrase patterns ($\beta=0.45$, $p<0.001$) DCL ($\beta=0.52$, $p<0.001$)	0.81	0.23	<0.001
UHDRS-TFC	Vocabulary range ($\beta=0.16$, $p=0.18$) Phrase patterns ($\beta=-0.52$, $p<0.001$) Reference rate to reality ($\beta=0.30$, $p=0.0040$) DCL ($\beta=-0.47$, $p<0.001$)	0.71	0.33	<0.001
SDMT	Phrase patterns ($\beta=-0.52$, $p<0.001$) Sentence development ($\beta=-0.27$, $p=0.031$) Syntactic complexity ($\beta=0.12$, $p=0.27$) DCL ($\beta=-0.61$, $p<0.001$)	0.65	0.25	<0.001
Lexical verbal fluency	Vocabulary range ($\beta=0.60$, $p<0.001$) Sentence development ($\beta=0.25$, $p=0.12$)	0.33	0.16	<0.001
Categorical verbal fluency	Vocabulary range ($\beta=0.36$, $p=0.0073$) Sentence development ($\beta=0.21$, $p=0.094$) Syntactic complexity ($\beta=0.23$, $p=0.053$) DCL ($\beta=-0.47$, $p=0.0011$)	0.62	0.20	<0.001
Stroop Color Test	Phrase patterns ($\beta=-0.57$, $p<0.001$) Reference rate to reality ($\beta=0.27$, $p=0.035$) DCL ($\beta=-0.47$, $p<0.001$)	0.60	0.23	<0.001
Stroop Word Test	Phrase patterns ($\beta=-0.77$, $p<0.001$) Reference rate to reality ($\beta=0.30$, $p=0.018$) Language ($\beta=0.33$, $p=0.0081$) DCL ($\beta=-0.29$, $p=0.019$)	0.59	0.40	<0.001
Stroop Interference Test	Phrase patterns ($\beta=-0.66$, $p<0.001$) Reference rate to reality ($\beta=0.30$, $p=0.051$) Age ($\beta=-0.22$, $p=0.12$) DCL ($\beta=-0.39$, $p=0.0067$)	0.53	0.24	0.002
Trail Making Test A	Phrase patterns ($\beta=0.49$, $p=0.013$) Reference rate to reality ($\beta=-0.29$, $p=0.091$) Language ($\beta=-0.47$, $p=0.0080$) DCL ($\beta=0.32$, $p=0.058$)	0.26	0.13	<0.001
Trail Making Test B	Language ($\beta=-0.50$, $p=0.0037$) DCL ($\beta=0.44$, $p=0.0096$)	0.27	0	<0.001
CAG	Vocabulary range ($\beta=-0.14$, $p=0.41$) Phrase patterns ($\beta=0.29$, $p=0.072$) Age ($\beta=-0.42$, $p=0.0086$) DCL ($\beta=0.31$, $p=0.049$)	0.44	0.06	0.002

β = standardized beta value of individual predictors, ΔR^2 = increase in R^2 from the model including only demographic variables (age, sex, and Czech/German language) and DCL, and the model additionally including linguistic features, DCL = Diagnostic Confidence Level, UHDRS-TMS = Unified Huntington's Disease Rating Scale Total Motor Score, UHDRS-TFC = Unified Huntington's Disease Rating Scale Total Functional Capacity, SDMT = Symbol Digit Modalities Test, CAG = cytosine-adenine-guanine.

Figure S1. Normative data of language characteristics across two languages for monologue and retelling tasks



Violin plots depict the kernel density of the data, overlaid with boxplots, which represent the interquartile range (IQR; 25th–75th percentile), with the horizontal line indicating the median. Whiskers extend to the most extreme values within 1.5×IQR. Results above each panel report the group per language interaction based on analysis of covariance, with age and sex as covariates. HD = Huntington's disease.