Original Research

Lexical and syntactic deficits analyzed via automated natural language processing: the new monitoring tool in multiple sclerosis

Martin Šubert, Michal Novotný, Tereza Tykalová, Barbora Srpová, Lucie Friedová, Tomáš Uher, Dana Horáková and Jan Rusz

Abstract

Background: Impairment of higher language functions associated with natural spontaneous speech in multiple sclerosis (MS) remains underexplored.

Objectives: We presented a fully automated method for discriminating MS patients from healthy controls based on lexical and syntactic linguistic features.

Methods: We enrolled 120 MS individuals with Expanded Disability Status Scale ranging from 1 to 6.5 and 120 age-, sex-, and education-matched healthy controls. Linguistic analysis was performed with fully automated methods based on automatic speech recognition and natural language processing techniques using eight lexical and syntactic features acquired from the spontaneous discourse. Fully automated annotations were compared with human annotations.

Results: Compared with healthy controls, lexical impairment in MS consisted of an increase in content words ($p = 0.037$), a decrease in function words ($p = 0.007$), and overuse of verbs at the expense of noun ($p = 0.047$), while syntactic impairment manifested as shorter utterance length ($p = 0.002$), and low number of coordinate clause ($p < 0.001$). A fully automated language analysis approach enabled discrimination between MS and controls with an area under the curve of 0.70. A significant relationship was detected between shorter utterance length and lower symbol digit modalities test score ($r = 0.25$, $p = 0.008$). Strong associations between a majority of automatically and manually computed features were observed ($r > 0.88$, $p < 0.001$).

Conclusion: Automated discourse analysis has the potential to provide an easy-to-implement and low-cost language-based biomarker of cognitive decline in MS for future clinical trials.

Keywords: automated linguistic analysis, language, multiple sclerosis, nature language processing, spontaneous discourse

Correspondence to:
Jan Rusz
Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, 160 00 Prague, Czech Republic.

Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic

Department of Neurology and ARTORG Center for Biomedical Engineering Research, Inselspital (Bern University Hospital), University of Bern, Bern, Switzerland
rusz.mz@gmail.com

Martin Šubert
Michal Novotný
Tereza Tykalová
Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic

Barbora Srpová
Lucie Friedová
Tomáš Uher
Dana Horáková
Department of Neurology and Centre of Clinical Neuroscience, First Faculty of Medicine, Charles University and General University Hospital, Prague, Czech Republic

Introduction

Multiple sclerosis (MS) is an inflammatory and neurodegenerative disease characterized by demyelination of axons, axonal and neuronal cells loss in the spinal cord and brain, including cortical gray matter.[1–3] Among sensorimotor and other physical dysfunctions, cognitive decline represents a major cause of disability and affects a great proportion of patients with MS.[4–6] Cognitive dysfunction has been reported in all MS phenotypes and may be present from the very early stage of the disease.[6,7] The most affected cognitive domains include information processing speed, complex attention, memory, and executive function.[8] The impact of cognitive decline on quality of life can be enormous, affecting family, social, and/or professional status.[9,10] However, cognitive relapses or worsening can occur independent of neurological worsening or subjective cognitive deficits,[11–13] and therefore early, regular cognitive monitoring can help clinicians recognize MS disease activity and tailor treatment recommendations.

Language decline is one of the primary domains reflecting the potential progression of cognitive dysfunction. Indeed, a recent study has reported that 75% of MS individuals self-reported some form of language impairment.[14] However, the majority of research concerning language disorders in MS is based on verbal fluency or naming deficits.[15] These simple tests, though revealing certain linguistic problems,[15–17] do not give a complete picture of potential language dysfunction in MS. It is thus conceivable that the assessment of the natural speech production during spontaneous discourse, comprising both the syntactic and lexical components of language, may draw a novel and more comprehensive frame to uncover potential MS linguistic deficits than naming or fluency tests.

While the syntax domain covers the rules of the word, clause, and utterance order, including the relationship among sentence elements and the language's grammar, the lexical domain reflects the vocabulary and meaning of particular words. However, how the lexical and syntactic language domains in MS are affected still remains poorly understood. Only two previous studies based on a low sample of MS patients have reported a decline in language grammar, shorter sentence length, and limited syntactic complexity and vocabulary.[18,19]

Nowadays, progress in digital technologies provides new possibilities in language processing, including automatic speech recognition (ASR) and natural language processing (NLP) methods. ASR involves transforming spoken language into text format, which requires the application of statistical models and algorithms to transcribe audio recordings into written text. In contrast, NLP uses machine learning algorithms to analyze the structure of sentences, words, and phrases to gain an understanding of the context and meaning behind them. Combining these techniques makes it possible to comprehensively analyze potential language abnormalities in patients and identify early signs of cognitive deterioration. Indeed, language assessments have already been extensively studied in other neurodegenerative disorders, such as Alzheimer's disease or mild cognitive impairment.[20–22] Such objective language-based biomarkers may significantly benefit from automated data analysis acquired even outside the laboratory environment and provide screening tool to detect and monitor disease progression,

helping professionals make necessary interventions[23] and improve patient outcomes. However, patients with MS may present not only with cognitive alterations but also with a motor speech disorder,[24] which both might pose difficulty to ASR and NLP techniques to transcribe and annotate speech correctly.

Importantly, the new concept of progression independent of relapse activity (PIRA) that emerged in recent years is today considered a critical part of disease progression in MS patients.[25,26] Unfortunately, the PIRA is still often undetected not only due to the low sensitivity of measures monitoring tools, but also challenges associated with assessment burden (staff, time, financial support). In this context, identifying new sensitive and automated tools for monitoring cognitive/language performance is extremely important. As the first step in this effort, the present pilot and proof of concept study aims to assess whether the syntactic and lexical language abnormalities derived from natural language samples can be detected in MS based on objective, fully automated assessment using ASR and NLP techniques. The additional aim was to explore the sensitivity of discourse analysis in MS through a fully automated approach using state-of-the-art ASR and NLP techniques in comparison with manual text transcription and annotation.

Materials and methods

Study design and participants
Consecutive Czech patients with a clinically confirmed diagnosis of definite MS according to the revised McDonald Criteria 2010[27] were consecutively recruited at the Charles University and General University Hospital, Prague, Czech Republic. The inclusion criteria for MS patients were (1) relapse-free state for at least 30 days prior to testing, (2) completion of at least elementary education lasting 8 years, and (3) no neurological disorder (e.g. stroke, epilepsy, Huntington's disease) or communication disorders that would significantly interfere with recording speech protocol (e.g. stuttering, aphasia, apraxia of speech) unrelated to MS. Each patient was ranked by a board-certified neurologist according to the Expanded Disability Status Scale (EDSS).[28] The neuropsychological assessment included the symbol digit modalities test (SDMT)[29] and the Rao adaptation of the Paced Auditory Serial Addition Test-3

(PASAT-3).[30] The SDMT assesses rapid information processing, working memory, and visual scanning, while the PASAT-3 represents multifactorial analysis that assesses auditory information processing speed and flexibility, cognitive processing speed, sustained and also divided attention and working memory functions, and calculation ability.[31,32] The Beck Depression Inventory-Second Edition (BDI-II) was used for the assessment of depressive symptoms[33] and Fatigue Severity Scale (FSS) for the evaluation of the impact of fatigue.[34]

In addition, age-, sex-, and education-matched healthy control group with no history of neurological or communication disorders was included solely to facilitate the evaluation of the extent of severity of language disorder in MS.

Speech examination

Speech data were recorded in one session in a quiet room with low ambient noise using a head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed approximately 5 cm from the participant's mouth. Recordings were sampled at 48 kHz with 16-bit resolution.

Each participant underwent a speech assessment guided by a trained speech professional (M.N., T.T., J.R.). Participants were instructed to perform spontaneous discourse on a neutral, freely chosen topic. If the examinator observed the signs of anxiety or excitement, the task was repeated based on another self-chosen topic. The content of the discourse task was monitored and categorized into seven topics, including description of (1) current day (19%), (2) hobby (15%), (3) holiday (15%), (4) past event (14%), (5) childhood (12%), (6) work (11%), and (7) others (14%). The mean length of recordings was 123 s [standard deviation (SD) 16], and the mean number of words was 216 (SD 64). The task's length was comparable with previous studies on lexical and syntactic features in patients with dementia.[22,35]

Speech transcription and annotation

The speech recordings were transcribed into text files using Google Cloud Speech-to-Text API,[36] chosen based on the comparison of the leading ASR software[37] based on the referred accuracy, documentation quality, and implementation difficulty.

The paralinguistic phenomena, including empty and filled pauses (e.g. 'mmm', 'ehm'), repetitive speech phenomena (i.e. repetition of phonemes or syllables), and non-verbal phenomena (i.e. laughing, coughing), were excluded from the transcripts; these paralinguistic phenomena were presented scarcely and did not substantially influence recording lengths. Text files were further processed using MorphoDiTa tool[38] with the Czech language model available from LIDNAT/CLARIN center.[39] Each word in the text was labeled with the corresponding word type. If a non-existing word was detected, it was eliminated from further analyses.

In addition to the automated analysis, we also conducted manual transcription and annotation. A speech specialist (M.Š.) transcribed and annotated each audio recording manually.

Linguistic analysis

We proceed with the analysis using eight linguistic features that have been studied in neurodegenerative diseases in previous literature[22,35,40,41] The criteria for feature selection were as follows (1) covering complex aspects of the lexical domain, syntax, and distinctive patterns of cognitive impairment (i.e. features' computational principle should be distinctive and thus low correlation among the parameters could be expected), and (2) the possibility for easy and robust implementation using available NLP tools to fully automate the analysis process.

We selected four following features to cover the lexical part of the language: content words,[40] function words,[40] moving-average type–token ratio,[42] and reference rate to reality;[35] for moving-average type–token ratio the window size was set to 57 words in accordance with our available data set and recommendations for determining the subject's vocabulary. In addition, we investigated four syntactic features: n-grams,[43] coordinate clauses,[41] subordinate clauses,[41] and mean length of utterance.[44] See Table 1 for a detailed feature description.

Accepting only a limited number of features, we lower the probability of a type II error and reduce potential overfitting in the regression analysis. The linguistic features were found only weakly correlated (Pearson: $|r| < 0.49$), except the content words and function words (Pearson: $r = -0.88$;

Table 1. Description of lexical and syntactic features.

Feature	Description
Lexical	
Content words (CW)	Bring information to the sentence. Computed as a sum of content words (nouns, verbs, adjectives, and adverbs) divided by the total number of words (N) CW = content words$/N$
Function words (FW)	Necessary for grammatically correct sentences. Computed as a sum of function words (pronouns, prepositions, conjunctions, auxiliary verbs, and determiners) divided by the total number of words (N) FW = function words$/N$
Moving-average type–token ratio (MATTR)	Measures lexical diversity. Function (F) loops through the text with a chosen window size (W) and step size of 1, counting the number of unique words (V) in each window and dividing by the total number of words in the window (W). The resulting values are averaged to obtain the final score MATTR = MEAN($F\{V/W\}$)
Reference rate to reality (RRtR)	Investigate the proportion of nouns to verbs RRtR = nouns/verbs
Syntactic	
n-grams (NG)	Investigate the repetitiveness of words/phrases. The sum of n-grams repetitions, that is, bigrams (BI), trigrams (TRI), and fourgrams (FOUR), normalized to the number of words in the text (N) NG = SUM(BI, TRI, FOUR)$/N$
Coordinate clauses (CC)	Represent the sentence complexity. The number of coordinate clauses (clause connected with coordinating conjunctions – e.g. and, for, but) normalized to the total number of clauses CC = coordinate clauses/total number of clauses
Subordinate clauses (SC)	Represent the sentence complexity. The number of subordinate clauses (clauses connected with subordinating conjunctions – e.g. although, after, because) normalized to the total number of clauses (C) SC = subordinate clauses/total number of clauses
Mean length of utterance (MLU)	A measure of linguistic productivity by the total number of words (N) to a total number of utterances (U) MLU = N/U

despite a high correlation, we preferred to preserve the content words and function words due to the completeness of the lexical analysis. The analysis was conducted in MATLAB (MathWorks, Natick, MA).

Statistical analysis

A one-way analysis of covariance (ANCOVA) was used to discriminate MS from the control group based on linguistic features. Effect sizes were determined using eta squared (η^2), with $\eta^2 > 0.009$ indicating small, $\eta^2 > 0.059$ indicating medium, and $\eta^2 > 0.139$ indicating a large effect.

The analysis was adjusted to the content of discourse (covariate). Partial correlation was used to assess the relationship between linguistic features and clinical data within the MS group, with age, sex, education, and content of discourse as covariates. Pearson correlation was applied to test for significant relationships between pairs of linguistic features obtained from the automated and manual datasets. The magnitude of agreement between features obtained from both data sets was measured as the root mean squared error normalized by the mean observed value (NRMSE).[45] All analyses used a two-tailed $p < 0.05$ threshold for statistically significant

Table 2. Clinical characteristics of MS patients.

Multiple sclerosis	(n = 120, 89 females)
Age (years)	43.8 (10.9, 18–74)
Education (years)	14.7 (2.9, 10–22)
Disease duration (years)	14.5 (7.6, 2–37)
EDSS score	3.8 (1.3, 1–6.5)
PASAT-3 score	45.0 (15.9, 0–60)
SDMT score	52.9 (13.0, 14–88)
BDI-II score	8.4 (8.2, 0–37)
FSS score	35.3 (15.0, 5–63)

Data are shown as mean (SD, range).
BDI-II, back depression inventory-second edition; EDSS, Expanded Disability Status Scale; FSS, Fatigue Severity Scale; PASAT-3, Paced Auditory Serial Addition Test-3; SD, standard deviation; SDMT, symbol digit modalities test.

differences. Moreover, we performed a binary logistic regression followed by leave-one-subject-out cross-validation to evaluate the ability of linguistic feature combinations to differentiate between groups (i.e. accuracy, sensitivity, and specificity). Several different classification scenarios were evaluated, including classifiers based on the combination yielding the best accuracy. The subset of language features providing the best accuracy was searched automatically using a grid-search approach. Overall diagnostic accuracy was reported as the area under the curve (AUC), determined from the receiver operating characteristic curve.

Results

Participants
The MS group consisted of 120 participants (89 females) with a mean age of 43.8 (SD 10.9, range 18–74) years and a mean education of 14.7 (SD 2.9, range 10–22; see Table 2). A total of 94 patients were diagnosed with relapsing-remitting MS, 15 with secondary progressive MS, 8 with primary-progressive MS, and 3 with clinically isolated syndrome. MS patients had EDSS scores between 1.0 and 6.5. In addition, the healthy control group consisted of 120 age- and sex-matched participants (88 females) with a mean age of 45.8 (SD 17.6, range 18–73) years and mean education of 15.0 (SD 2.9, range 8–23). No between-group

significance between MS and control groups was found for age ($p = 0.45$), education ($p = 0.39$), or sex ($p = 1$).

Linguistic features and sensitivity analysis
Considering lexical features, the MS group was significantly discriminated from healthy controls by an increase in content words ($p = 0.037$, $\eta^2 = 0.018$) and a decrease in function words ($p = 0.007$, $\eta^2 = 0.030$; see Figure 1). Further word type analysis showed that the increase of content words in MS was predominantly caused by the significantly higher number of verbs in MS compared with controls ($p = 0.003$, $\eta^2 = 0.037$; see Table S1). A lower reference rate to reality in the MS group compared with healthy controls ($p = 0.047$, $\eta^2 = 0.016$) confirms the overuse of verbs at the expense of nouns. No significant difference was found between MS and healthy controls using moving-average type–token ratio ($p = 0.22$).

For syntactic features, MS showed shorter mean utterance length ($p = 0.002$, $\eta^2 = 0.039$) as well as low number of coordinate clauses ($p < 0.001$, $\eta^2 = 0.086$) compared with controls. No significant difference between MS and healthy controls was found for subordinate clauses ($p = 0.43$) and n-grams ($p = 0.61$).

Considering sensitivity analysis (see Figure 2), combining six features, including content words, function words, reference rate to reality, coordinate clauses, subordinate clauses, and mean length of utterance, led to the best achieved AUC of 0.70 (accuracy of 63.3%, specificity of 63.6%, and sensitivity of 63.1%).

Correlations between language and clinical markers
Examination of the relationship between linguistic features and clinical scales including EDSS, SDMT, and PASAT-3 revealed a correlation between mean length of utterance and SDMT (Partial correlation: $r = 0.25$, $p = 0.008$; see Figure 3). No significant relationships were observed between linguistic features and BDI-II and FSS.

Comparing results from automated and manual data set
The ASR system achieved a word error rate of 22.4% compared with manually transcribed texts.
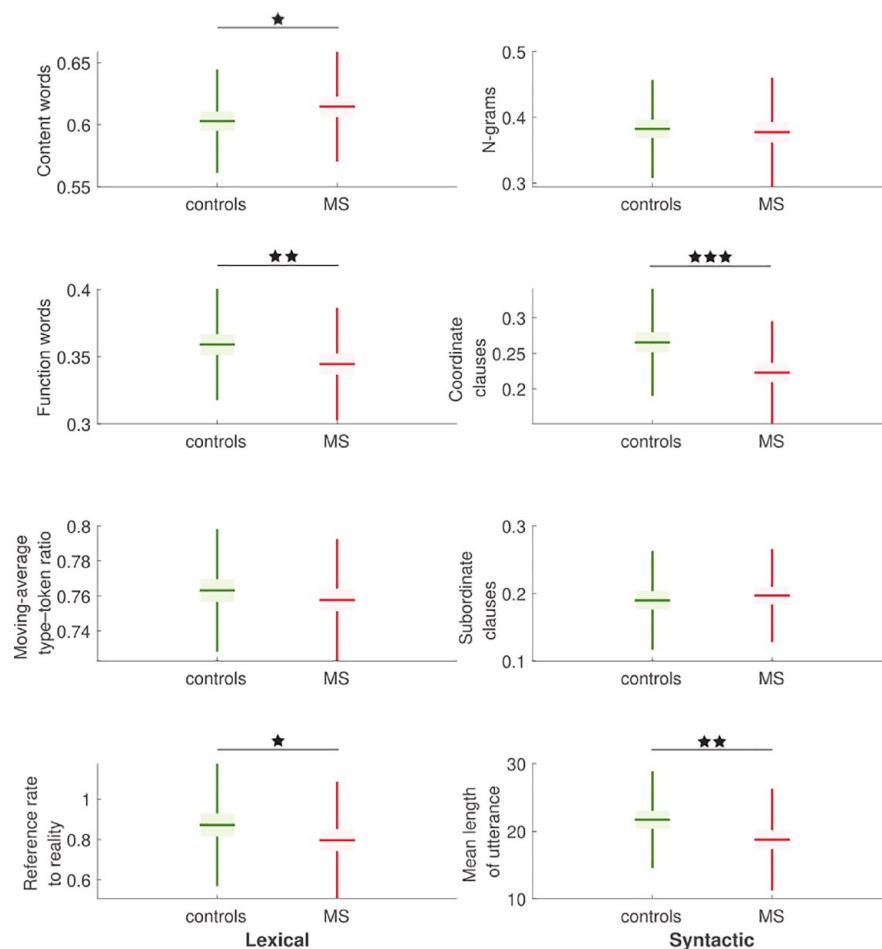
Figure 1. Results of linguistic analysis for lexical and syntactic features.
The left panel shows data for lexical features, and the right panel data for syntactic features. Horizontal lines represent the means, boxes represent the 95% confidence interval, and whiskers represent the standard deviation. Statistically significant differences between groups after analysis of covariance: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. All results are adjusted for the content of discourses.
MS, multiple sclerosis.

Our analysis revealed very strong correlations (Pearson: $r > 0.88$, $p < 0.001$) between features extracted from the automated and manual data sets with NRMSE lower than 0.18, except for the mean length of utterance, where we found a moderate correlation (Pearson: $r = 0.58$, $p < 0.001$) with NRMSE of 0.44 (see Figure 4). The ANCOVA analysis allowed us to significantly differentiate MS patients from controls using the same linguistic features as in the case of automated analysis, except for content words, where we only achieved a trend toward significance ($p = 0.086$; see Figure S1).

Discussion

This study proved the hypothesis of higher language function deficits in both lexical and syntactic domains in MS patients determined based on discourse analysis of spontaneous speech. Considering lexical features, the MS group showed a decrease in function words, demonstrating lower stress on grammar and proper sentence structure. The absence of function words in MS is compensated with an increased occurrence of content words. Indeed, detailed word type analysis revealed that MS patients tend to use verbs more frequently. Such overuse of verbs in MS was also demonstrated by decreased reference rate to reality, assessing the frequency of verbs related to nouns. For instance, the sentences '*Yesterday, I went to the store and picked up some groceries like salad and carrots. Then I came back home and made dinner.*' could be formulated by MS patient as '*I went to store yesterday. I bought*
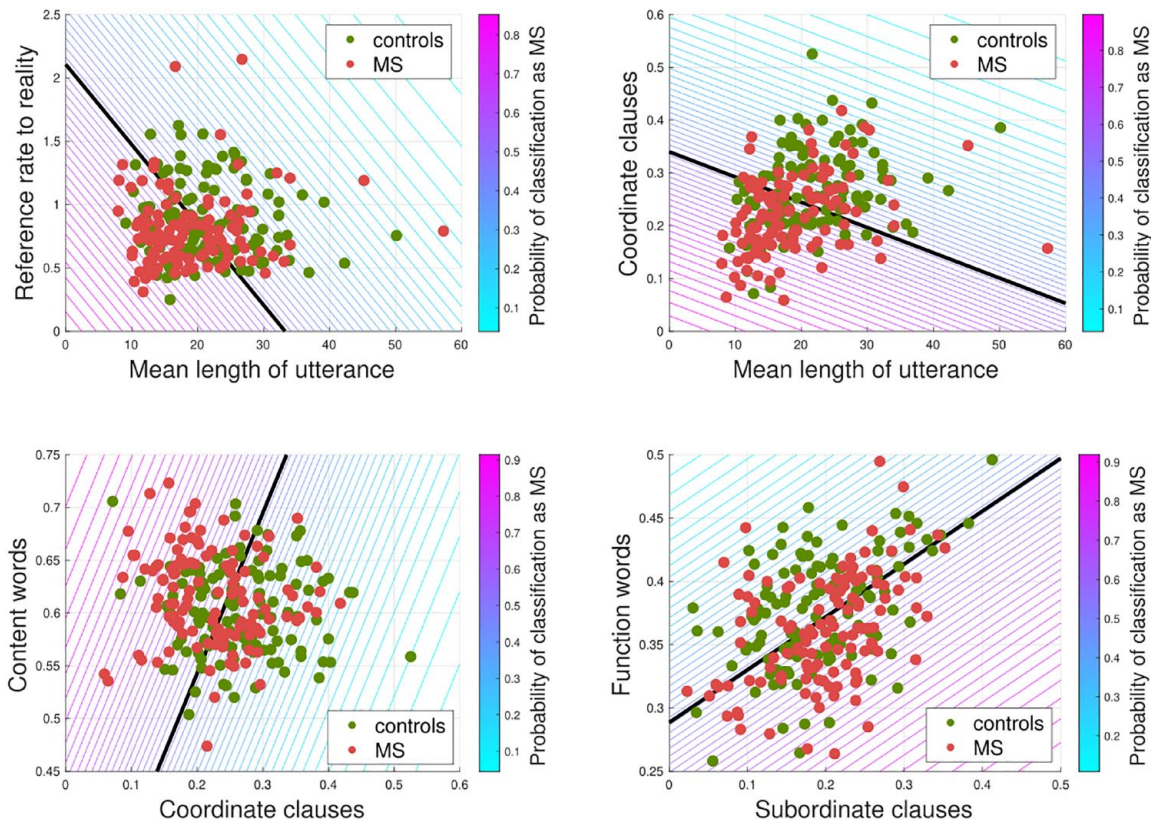
Figure 2. Selected pairs of linguistic features contributing to the best classification accuracy with classification boundaries separating MS from controls.
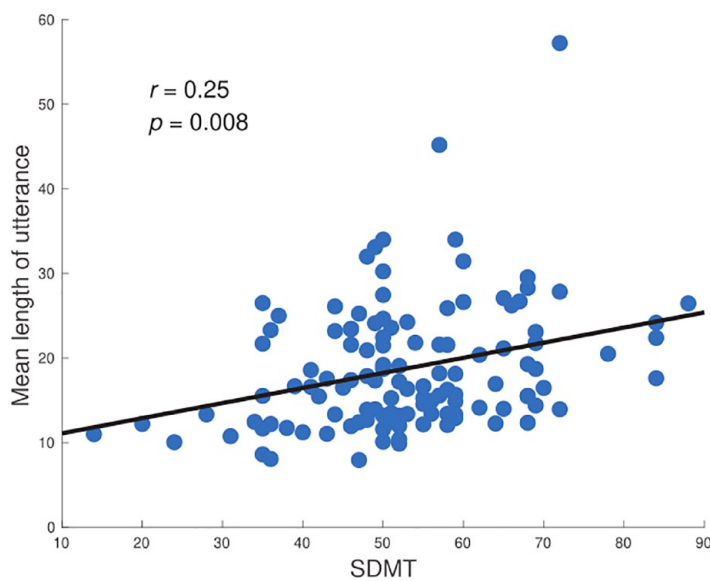MS, multiple sclerosis.

Figure 3. Significant correlation between SDMT and mean length of utterance.
The blue circles demonstrate the real, uncorrected linguistic and neuropsychological values, while the correlation coefficient *r* and its corresponding *p*-value are corrected to age, sex, education, and content of discourse.
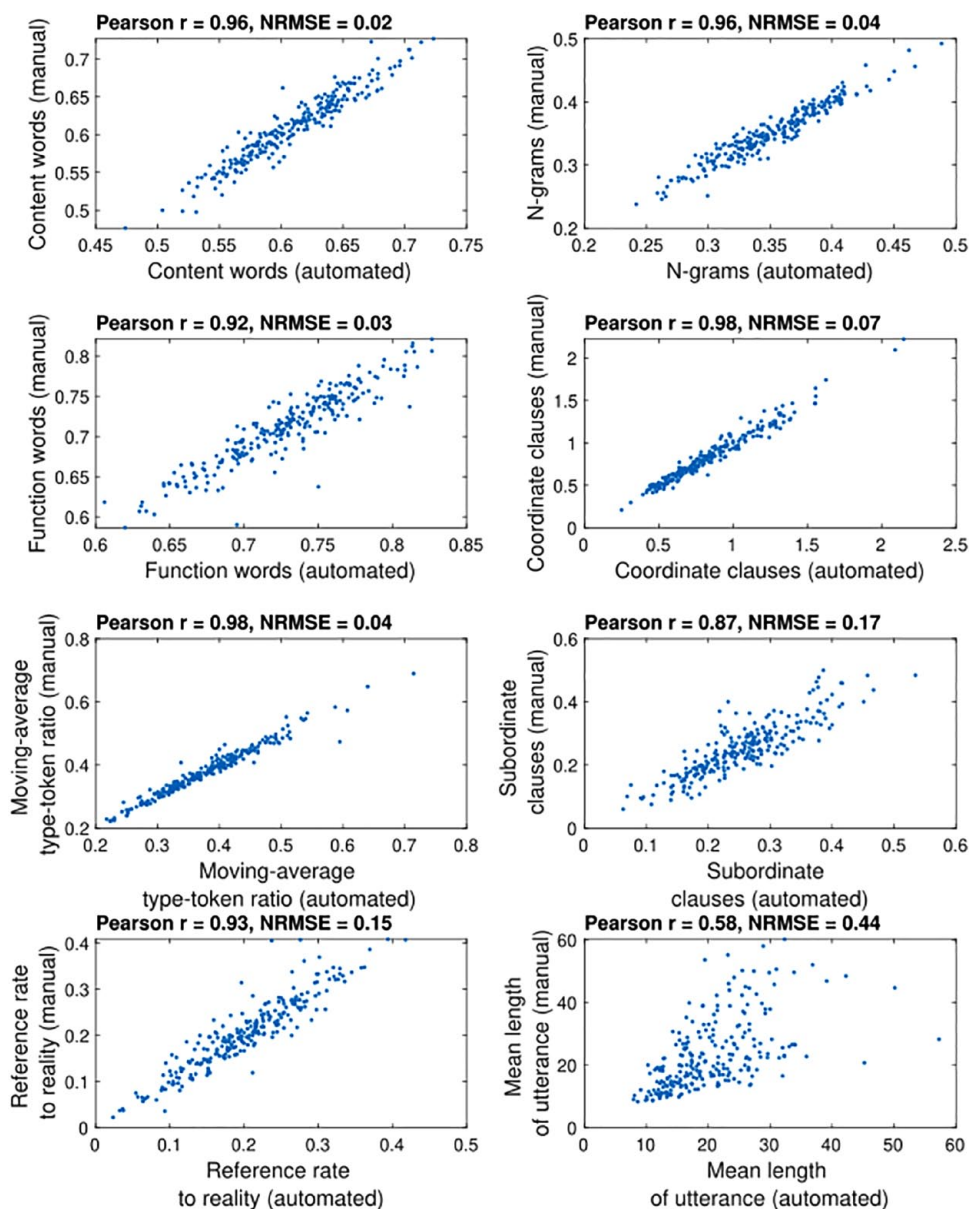SDMT, symbol digit modalities test.

Figure 4. Relationship between features extracted from automated and manual data set.
NRMSE, normalized root mean square error; Pearson *r*, Pearson correlation.

salad, carrot, came back home, and made dinner.'. In this example, there is lower use of function words such as '*and*' or '*then*' and instead higher use of necessary content words, to express patient thought. No significant difference was found between MS and controls using moving-average type–token ratio, indicating that the vocabulary of MS patients is not restricted. Even though previous studies report some deficiencies in the lexical domain, their occurrence in MS seems to be rather rare.[18,19,46] Thus, we may hypothesize that our findings in the lexical area are predominantly

associated with grammatical deficits. This would also be in agreement with the previous study, indicating difficulty with grammar in MS.[18]

Regarding the syntactic domain, the MS group manifests significantly shorter sentence lengths without the intention to evolve more complex sentences with various clauses, which is suggested by a decline in coordinate clauses. The difficulties with sentence structure in MS are also supported by reduced function words. For instance, the sentence '*I went to the store and bought some red apples*

but forgot to add oranges, so I wasn't happy.' could be reformulated by the MS patient as '*I went to the store and bought apples. I didn't buy oranges.*'. In this example, we can see that the thought is similar, but the sentences are more fragmented without the intention to expand them into a broader sentence with minor additional information. In line with our findings, syntax deficiencies, including shorter utterance length and difficulty with sentence comprehension, have already been reported in MS.[18,46,47] No statistical difference between MS and controls was found using *n*-grams, indicating no difficulty with word finding and the necessity of repetition of the same phrase. This might be expected since reduced vocabulary in MS was not notable.

The alteration of language function in MS seems to reflect different neuropathological processes across brain regions rather than the cognitive decline captured by standardized neuropsychological assessment. Indeed, we could only find a weak but significant correlation between the mean length of utterance and SDMT, which captures a hallmark cognitive deficit in MS. The relevance of this finding can be supported by a previous study reporting that the frontal lobe region, known to be involved in cognitive impairment in MS,[48] plays a critical role in both syntactic processing and speed of processing.[49] Also, based on differences in speech performance between children and adults, the previous study has shown links between higher-order processing demands and syntactic complexity with utterance length.[50] However, our cognitive assessment only examined visual and auditory information processing speed. Hence, further studies are needed to investigate relationships between language performance and other cognitive domains, such as memory or executive functions, frequently affecting MS patients.[51,52]

This study aimed to systematically select linguistic features covering most of the language domains, which could be possible to analyze using known techniques, including ASR and NLP, and with minimum human control. The advantage of our solution based on only eight language features with interpretable neurocognitive behavior was that it did not require a large sample size to train the specific model and is ready to use for future clinical trials. Although we could find a high statistical significance between MS and controls across several language features, our classification experiment reached a relatively low discrimination accuracy with an AUC of 0.70. If diagnostic utility should be the primary endpoint, combining more features that quantify different properties of language impairment would likely lead to enhancement of reported accuracy. However, such accuracy seems to be still reasonable considering our relatively short recordings with an average number of 216 words. In the future, we expect significant enhancement in this approach by mass screening via smartphones outside a laboratory environment, allowing us to collect considerably longer audio recordings.[53]

Despite a relatively high word error rate of 22.4% in the automated transcription, our automated approach yielded highly comparable results with manual analysis in terms of strong correlations and significant group differences between MS and controls among the language features in both data sets. This is crucial from the clinical point of view, as it is more important to achieve a correct estimate of the patient's language performance than to obtain the precise transcription of individual words. The only exception was the mean length of utterance, characterized by the lowest correlation between manual and automated labels. This discrepancy is understandable, given the challenging task of predicting punctuation in spontaneous speech.[54] Although Google Cloud Speech-to-Text API presents a state-of-the-art ASR system, it is trained mainly using data sets with healthy speech, which might present a notable disadvantage for our tasks. Motor speech disorder, that is, spastic-ataxic dysarthria typically encountered in MS,[24] may cause more complications to the acoustic model of the ASR system that attempts to recognize a word based on acoustic information. In addition, we have found that MS manifests different word selection and sentence structures from the healthy controls. This might lead to difficulties for the language model that finds the probability of one word following another. Since the research in tuning ASR system using data sets with impaired speech already suggests significant improvements,[55] the accuracy of automated annotation could likely be significantly improved in the future. Furthermore, Czech poses a more challenging and complex task for ASR compared with more common languages such as English or German,[56] suggesting that follow-up research in such languages will likely already result in better automated speech transcription of patients with MS.

This is a proof-of-concept study showing that automated language analysis could potentially detect pathological cognitive performance in MS. Current cognitive measures dedicated to MS care have low sensitivity to detect cognitive change, are influenced by the learning effect, and other factors such as depression or fatigue.[52,57] Furthermore, traditional monitoring tools, such as brief international cognitive assessment for MS, are relatively time-consuming, personnel-intensive, and sometimes not well tolerated by patients, especially when administered frequently. Therefore, there is an urgent need for sensitive and easily applied biomarkers able to monitor disease activity. The proposed automated analysis of lexical and syntactic deficits could represent a new, potentially more sensitive, easy-to-implement, and low-cost language-based biomarker of cognitive decline in MS for future clinical trials and routine clinical practice. However, future longitudinal studies are needed to describe the evolution and dynamics of language performance over time, to investigate whether the decline in language performance reflects PIRA,[25,58] and to explore whether the decline in language-based biomarkers increases the sensitivity for the detection of disease activity beyond established assessment tools such as EDSS or quantitative measures of walking and hand function. In addition, technical improvement of language-based biomarker software and passive, unconscious (e.g. without required active participation of patients) screening of language performance through mobile application could further improve the sensitivity of this technique to detect cognitive change and enable very frequent or continuous monitoring. In summary, language-based biomarkers might have very high potential to revolutionize the way on how to detect disease progression, including PIRA, and thus enable clinicians to modify treatment as soon as possible to prevent further disability progression, leading to the improvement of everyday functioning, quality of life, and clinical care.

A potential limitation of this study is that our results are based solely on the Czech language. It is desirable to reproduce the findings across different languages to ensure the robustness of used linguistic features characterizing the cognitive language decline in MS. Also, we did not investigate BDI-II and FSS in healthy controls; therefore, we cannot entirely exclude the effect of depression and fatigue on cognitive linguistic function.[59] However, we did not reveal any significant correlation between linguistic features and BDI-II or FSS, indicating rather low influence of depression and fatigue on language performance in our MS cohort. Admittedly, since the extent of cognitive dysfunction may differ across various clinical subtypes of MS[60] and the majority of our sample had a relapsing-remitting course, future studies are encouraged to assess the sensitivity of syntactic and lexical language features across these differing MS subtypes.

In conclusion, we present an objective, fully automated method for discriminating MS patients from healthy controls using linguistic analysis. MS participants manifest a decline in both lexical and syntactic language domains. Language impairments in MS are prevalent in shorter, undeveloped sentences and grammatical incorrectness. If confirmed in future studies and improved sensitivity, this approach might provide a fully automated, easy-to-implement, and low-cost language-based biomarker of cognitive decline and disease progression in MS for future clinical trials and clinical practice.

Declarations

Ethics approval and consent to participate

Each participant provided written, informed consent for participation, and the study was approved by the Ethics Committee of the General University Hospital in Prague, Czech Republic (approval number 6/15 Grant GACR VFN).

Consent for publication

Not applicable.

Author contributions

Martin Šubert: Formal analysis; Investigation; Methodology; Visualization; Writing – original draft.

Michal Novotný: Investigation; Writing – review & editing.

Tereza Tykalová: Investigation; Writing – review & editing.

Barbora Srpová: Investigation; Writing – review & editing.

Lucie Friedová: Investigation; Writing – review & editing.

ORCID iD
Jan Rusz 🔘 https://orcid.org/0000-0002-1036-3054

Supplemental material
Supplemental material for this article is available online.

References
1. Eshaghi A, Young AL, Wijeratne PA, *et al*. Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat Commun* 2021; 12: 2078.

2. Bergsland N, Horakova D, Dwyer MG, *et al*. Gray matter atrophy patterns in multiple sclerosis: a 10-year source-based morphometry study. *Neuroimage Clin* 2018; 17: 444–451.

3. Lassmann H. Pathogenic mechanisms associated with different clinical courses of multiple sclerosis. *Front Immunol* 2019; 9: 3116, https://www.frontiersin.org/articles/10.3389/fimmu.2018.03116

4. Benedict RHB and Zivadinov R. Risk factors for and management of cognitive dysfunction in multiple sclerosis. *Nat Rev Neurol* 2011; 7: 332–342.

5. Cruccu G, Deuschl G and Federico A. Scientific publications of European neurologists: a survey commissioned by the European Academy of Neurology. *Eur J Neurol* 2018; 25: 1128–1133.

6. DeLuca J, Chiaravalloti ND and Sandroff BM. Treatment and management of cognitive dysfunction in patients with multiple sclerosis. *Nat Rev Neurol* 2020; 16: 319–332.

7. McNicholas N, O'Connell K, Yap SM, *et al*. Cognitive dysfunction in early multiple sclerosis: a review. *QJM* 2018; 111: 359–364.

8. Oreja-Guevara C, Ayuso Blanco T, Brieva Ruiz L, *et al*. Cognitive dysfunctions and assessments in multiple sclerosis. *Front Neurol* 2019; 10: 581, https://www.frontiersin.org/articles/10.3389/fneur.2019.00581

9. Rao SM, Leo GJ, Ellington L, *et al*. Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning. *Neurology* 1991; 41: 692–696.

10. Shevil E and Finlayson M. Perceptions of persons with multiple sclerosis on cognitive changes and their impact on daily life. *Disabil Rehab* 2006; 28: 779–788.

11. Pardini M, Uccelli A, Grafman J, *et al*. Isolated cognitive relapses in multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2014; 85: 1035–1037.

12. Benedict RH, Morrow S, Rodgers J, *et al*. Characterizing cognitive function during relapse in multiple sclerosis. *Mult Scler* 2014; 20: 1745–1752.

13. Motyl J, Friedova L, Vaneckova M, *et al*. Isolated cognitive decline in neurologically stable patients with multiple sclerosis. *Diagnostics* 2021; 11: 464.

14. El-Wahsh S, Ballard K, Kumfor F, *et al*. Prevalence of self-reported language impairment in multiple sclerosis and the association with health-related quality of life: an international

survey study. *Mult Scler Relat Disord* 2020; 39: 101896.

15. Renauld S, Mohamed-Saïd L and Macoir J. Language disorders in multiple sclerosis: a systematic review. *Mult Scler Relat Disord* 2016; 10: 103–111.

16. Ntoskou K, Messinis L, Nasios G, *et al*. Cognitive and language deficits in multiple sclerosis: comparison of relapsing remitting and secondary progressive subtypes. *Open Neurol J* 2018; 12: 19–30.

17. Lebkuecher AL, Chiaravalloti ND and Strober LB. The role of language ability in verbal fluency of individuals with multiple sclerosis. *Mult Scler Relat Disord* 2021; 50: 102846.

18. Sonkaya AR and Bayazit ZZ. Language aspects of patients with multiple sclerosis. *EJMI* 2018; 2: 133–138.

19. Feenaughty L. Linguistic performance during monologues and correlates of neuropsychological function for adults with multiple sclerosis. *Aphasiology*. Epub ahead of print 14 July 2022. DOI: 10.1080/02687038.2022.2099527.

20. Asgari M, Kaye J and Dodge H. Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimers Dement* 2017; 3: 219–228.

21. Eyigoz E, Mathur S, Santamaria M, *et al*. Linguistic markers predict onset of Alzheimer's disease. *eClinicalMedicine* 2020; 28: 100583.

22. Šubert M, Šimek M, Novotný M, *et al*. Linguistic abnormalities in isolated rapid eye movement sleep behavior disorder. *Mov Disord* 2022; 37: 1872–1882.

23. Guilloton L, Camdessanche JP, Latombe D, *et al*. A clinical screening tool for objective and subjective cognitive disorders in multiple sclerosis. *Ann Phys Rehabil Med* 2020; 63: 116–122.

24. Rusz J, Benova B, Ruzickova H, *et al*. Characteristics of motor speech phenotypes in multiple sclerosis. *Mult Scler Relat Disord* 2018; 19: 62–69.

25. Kappos L, Wolinsky JS, Giovannoni G, *et al*. Contribution of relapse-independent progression vs relapse-associated worsening to overall confirmed disability accumulation in typical relapsing multiple sclerosis in a pooled analysis of 2 randomized clinical trials. *JAMA Neurol* 2020; 77: 1132–1140.

26. Giovannoni G, Popescu V, Wuerfel J, *et al*. Smouldering multiple sclerosis: the 'real MS'. *Ther Adv Neurol Disord* 2022; 15: 17562864211066751.

27. Polman CH, Reingold SC, Banwell B, *et al*. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann Neurol* 2011; 69: 292–302.

28. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; 33: 1444–1452.

29. Smith A. Symbol digit modalities test. Los Angeles, CA: Western Psychological Services, 1973.

30. Rao SM. *A manual for the brief repeatable battery of neuropsychological tests in multiple sclerosis*. New York: National Multiple Sclerosis Society.

31. Tombaugh TN. A comprehensive review of the paced auditory serial addition test (PASAT). *Arch Clin Neuropsychol* 2006; 21: 53–76.

32. Benedict RH, Amato MP, Boringa J, *et al*. Brief international cognitive assessment for MS (BICAMS): international standards for validation. *BMC Neurol* 2012; 12: 1–7.

33. Ciharova M, Cígler H, Dostálová V, *et al*. Beck depression inventory, second edition, Czech version: demographic correlates, factor structure and comparison with foreign data. *Int J Psychiatry Clin Pract* 2020; 24: 371–379.

34. Krupp LB, LaRocca NG, Muir-Nash J, *et al*. The fatigue severity scale. Application to Patients with Multiple Sclerosis and Systemic Lupus Erythematosus. *Arch Neurol* 1989; 46: 1121–1123.

35. Beltrami D, Gagliardi G, Rossini Favretti R, *et al*. Speech analysis by natural language processing techniques: a possible tool for very early detection of cognitive decline? *Front Aging Neurosci* 2018; 10: 369, https://www.frontiersin.org/articles/10.3389/fnagi.2018.00369

36. Google. Speech-to-text: automatic speech recognition. https://cloud.google.com/speech-to-text (accessed 30 August 2022).

37. Filippidou F and Moussiades L. A benchmarking of IBM, Google and Wit automatic speech recognition systems. In: Maglogiannis I, Iliadis L and Pimenidis E (eds) *Artificial intelligence applications and innovations*. Vol. 583. Cham: Springer, 2020, pp. 73–82.

38. Straka M and Straková J. MorphoDiTa: morphological dictionary and tagger. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, 2014. http://hdl.handle.net/11858/00-097C-0000-0023-43CD-0

39. Straka M and Straková J. Czech models (MorfFlex CZ 160310+ PDT 3.0) for MorphoDiTa 160310. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, 2016. http://hdl.handle.net/11234/1-1674

40. Corver N and Riemsdijk H van. *Semi-lexical categories: the function of content words and the content of function words*. Berlin; Boston, MA: De Gruyter, Inc., 2013.

41. De Lira JO, Ortiz KZ, Campanha AC, *et al*. Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *Int Psychogeriatr* 2011; 23: 404–412.

42. Covington MA and McFall JD. Cutting the gordian knot: the moving-average type–token ratio (MATTR). *J Quantit Linguist* 2010; 17: 94–100.

43. Orimaye S, Wong J, Golden K, *et al*. Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinform* 2017; 2017: 34.

44. Ezeizabarrena M-J and Garcia Fernandez I. Length of utterance, in morphemes or in words?: MLU3-w, a reliable measure of language development in early Basque. *Front Psychol* 2018; 8: 2265.

45. Vogel AP, Rosen KM, Morgan AT, *et al*. Comparability of modern recording devices for speech analysis: smartphone, landline, laptop, and hard disc recorder. *Folia Phoniatr Logop* 2014; 66: 244–250.

46. Arrondo G, Sepulcre J, Beatriz D, *et al*. Narrative speech is impaired in multiple sclerosis. *Eur Neurol J* 2010; 2: 11–40.

47. Laakso KB, Brunnegård K, Hartelius L, *et al*. Assessing high-level language in individuals with multiple sclerosis: a pilot study. *Clin Linguist Phon* 2000; 14: 329–349.

48. Curti E, Graziuso S, Tsantes E, *et al*. Correlation between cortical lesions and cognitive impairment in multiple sclerosis. *Brain Behav* 2018; 8: e00955.

49. Caplan D, Waters G and Alpert N. Effects of age and speed of processing on rCBF correlates of syntactic processing in sentence comprehension. *Hum Brain Mapp* 2003; 19: 112–131.

50. Maner KJ, Smith A and Grayson L. Influences of utterance length and complexity on speech motor performance in children and adults. *J Speech Lang Hear Res* 2000; 43: 560–573.

51. Chiaravalloti ND and DeLuca J. Cognitive impairment in multiple sclerosis. *Lancet Neurol* 2008; 7: 1139–1151.

52. Uher T, Vaneckova M, Sormani MP, *et al*. Identification of multiple sclerosis patients at highest risk of cognitive impairment using an integrated brain magnetic resonance imaging assessment approach. *Eur J Neurol* 2017; 24: 292–301.

53. Sateli B, Cook G and Witte R. Smarter mobile apps through integrated natural language processing services. In: *Proceedings of 10th international conference, MobiWIS 2013*, Paphos, Cyprus, 26–29 August 2013, pp. 187–202. Berlin: Springer.

54. Nguyen B, Nguyen VBH, Nguyen H, *et al*. Fast and accurate capitalization and punctuation for automatic speech recognition using transformer and chunk merging. In: *2019 22nd conference of the oriental COCOSDA international committee for the co-ordination and standardisation of speech databases and assessment techniques (O-COCOSDA)*, Cebu, Philippines, 25 October 2019, pp. 1–5. New York: IEEE.

55. MacDonald B, Jiang P-P, Cattiau J, *et al*. Disordered speech data collection: lessons learned at 1 million utterances from project Euphonia. In: *INTERSPEECH 2021*, Brno, Czechia Republic, 30 August – 3 September 2021, pp. 4833–4837, https://www.isca-speech.org/archive/pdfs/interspeech_2021/macdonald21_interspeech.pdf

56. Nouza J, Zdansky J, Cerva P, *et al*. Challenges in speech processing of slavic languages (case studies in speech recognition of Czech and Slovak). In: Esposito A, Campbell N, Vogel C, *et al*. (eds) *Development of multimodal interfaces: active listening and synchrony: second COST 2102 international training school*, Dublin, 23–27 March 2009, revised selected papers, pp. 225–241. Berlin and Heidelberg: Springer.

57. Chitnis T, Gonzalez C, Healy BC, *et al*. Neurofilament light chain serum levels correlate with 10-year MRI outcomes in multiple sclerosis. *Ann Clin Transl Neurol* 2018; 5: 1478–1491.

58. Portaccio E, Bellinvia A, Fonderico M, *et al*. Progression is independent of relapse activity in early multiple sclerosis: a real-life cohort study. *Brain* 2022; 145: 2796–2805.

59. Golan D, Doniger GM, Wissemann K, *et al*. The impact of subjective cognitive fatigue and depression on cognitive function in patients with multiple sclerosis. *Mult Scler* 2018; 24: 196–204.

60. Denney DR, Sworowski LA and Lynch SG. Cognitive impairment in three subtypes of multiple sclerosis. *Arch Clin Neuropsychol* 2005; 20: 967–981.

Visit Sage journals online
journals.sagepub.com/home/tan

S Sage journals