Feasibility of using an automated analysis of formulation effort in patients' spoken seizure descriptions in the differential diagnosis of epileptic and nonepileptic seizures

Nathan Pevy[1], Heidi Christensen[2], Traci Walker[3], Markus Reuber[4]

[1] Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK
[2] Department of Computer Science, University of Sheffield, Sheffield, UK
[3] Division of Human Communication Sciences, University of Sheffield, Sheffield, UK
[4] Academic Neurology Unit, University of Sheffield, Royal Hallamshire Hospital, Sheffield, UK

Correspondence

Nathan Pevy, Sheffield Institute of Translational Neuroscience (SiTraN), University of Sheffield, Sheffield, UK
Email: ndpevy1@sheffield.ac.uk

Keywords

Epilepsy, Transient Loss of Consciousness, Classification, Natural Language Processing, Speech Analysis, Diagnosis

Acknowledgements

This research project was funded by Epilepsy Research UK

Conflict of interest

None of the authors have any conflict of interest to disclose

Ethics in publishing

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Abstract

Objective: There are three common causes of Transient Loss of Consciousness (TLOC), syncope, epileptic and psychogenic nonepileptic seizures (PNES). Many individuals who have experienced TLOC initially receive an incorrect diagnosis and inappropriate treatment. Whereas syncope can be distinguished relatively easily with a small number of “yes”/”no” questions, the differentiation of the other two causes of TLOC is more challenging. Previous qualitative research based on the methodology of Conversation Analysis has demonstrated that the descriptions of epileptic seizures contain more formulation effort than accounts of PNES. This research investigates whether features likely to reflect the level of formulation effort can be automatically elicited from audio recordings and transcripts of speech and used to differentiate between epileptic and nonepileptic seizures.

Method: Verbatim transcripts of conversations between patients and neurologists were manually produced from video and audio recordings of interactions with 45 patients (21 epilepsy and 24 PNES). The subsection of each transcript containing the patient’s account of their first seizure was manually extracted for the analysis. Seven automatically detectable features were designed as markers of formulation effort. These features were used to train a Random Forest machine learning classifier.

Result: There were significantly more hesitations and repetitions in descriptions of epileptic than nonepileptic seizures. Using a nested leave-one-out cross validation approach, 71% of seizures were correctly classified by the Random Forest classifier.

Discussion: This pilot study provides proof of principle that linguistic features that have been automatically extracted from audio recordings and transcripts could be used to distinguish between epileptic seizures and PNES and thereby contribute to the differential diagnosis of TLOC. Future research should explore whether additional observations can be incorporated into a diagnostic stratification tool and compare the performance of these features when they are 4 Pevy et al. combined with additional information provided by patients and witnesses about seizure manifestations and medical history

1. Introduction

Transient loss of consciousness (TLOC) is defined as a loss of awareness characterised by amnesia, abnormal motor control, loss of responsiveness and a short duration (Brignole *et al.*, 2018) . Over 90% of TLOC presentations are attributable to one of three causes: epileptic seizures, psychogenic nonepileptic seizures (PNES) or syncope (Kotsopoulos *et al.*, 2003) . The differentiation between these three mechanisms can be challenging as TLOC manifestations have usually subsided before patients present to health services, investigations carried out after TLOC events are often unhelpful or misleading and because there are no TLOC manifestations that are pathognomonic for a particular cause. In most cases the diagnosis ultimately rests on an expert interpretation of the history from patients and witnesses (Plug and Reuber, 2009). Unfortunately, around 20% of patients initially receive an incorrect diagnosis (Xu *et al.*, 2016). Misdiagnosing patients can be dangerous, for instance, if cardiac syncope or epilepsy are missed, and patients may be exposed to the negative consequences associated with a given diagnosis unnecessarily, for example the unpleasant side effects of medication or a driving ban (Xu *et al.*, 2016). Furthermore, there are considerable costs associated with referring patients for the incorrect tests and prescribing unnecessary treatments (Juarez-Garcia *et al.*, 2006).

Ongoing research is investigating the feasibility of using a clinical decision tool to standardise the collection and interpretation of a patient's seizure history (Stiell and Bennett, 2007; Wardrope, Newberry and Reuber, 2018). One such tool, the iPEP, a computer-analysed questionnaire including a series of yes/no questions about TLOC manifestations, questions about patients' medical history and some additional questions to observers (Reuber *et al.*, 2016; Chen *et al.*, 2019), has shown considerable diagnostic promise in a modeling study (Wardrope *et al.*, 2020). Using a three-class multinomial Random Forest classifier, the iPEP correctly predicted the underlying diagnosis with an accuracy of 86%. The differentiation between epileptic and nonepileptic seizures was found to be more challenging than that of syncope from seizures, as the model correctly identified all patients with syncope and the 14% of patients who were misdiagnosed either had epilepsy or PNES. These findings suggest that the model could benefit from further refinement, especially of its ability to discriminate between epilepsy and PNES.

One way to improve an automated clinical decision or diagnostic stratification tool (for instance for the planning of disorder-appropriate investigations) involves the incorporation of automated language analysis. Previous qualitative research using conversation analysis (CA) has described differences in how patients with epilepsy (PWE) and patients with nonepileptic seizures (PWNES) talk to clinicians about their seizure (Schwabe, Howell and Reuber, 2007; Schwabe *et al.*, 2008; Plug, Sharrack and Reuber, 2009a, 2010; Robson *et al.*, 2012). The utility of these observations for diagnostic purposes has been demonstrated by multiple blinded, multi-rater research studies where linguists or psychologists correctly predicted the diagnosis of epilepsy and PNES by studying transcripts of interactions with an accuracy ranging between 80-90% (Reuber *et al.*, 2009; Cornaggia *et al.*, 2012; Papagno *et al.*, 2017; Yao *et al.*, 2017; Biberon *et al.*, 2020). Although it has been shown that diagnostically relevant interactional and linguistic can also be made in real time by clinicians while they speak to patients, considerable expertise is required on the part of the clinician to make the relevant observations (Jenkins *et al.*, 2015). However, previous research focusing on patients with memory problems has demonstrated the feasibility of automating the analysis of language for diagnostic purposes (Mirheidari *et al.*, 2017). The automation of diagnostic language analysis involves the definition of semantic and acoustic features approximating the conversation analytic observations, programming a computer to detect these features in transcripts produced using automatic speech recognition and identifying the most discriminant features using machine learning models.

One of the most important differentiating features between the speech of PWE and PWNES previously described by the qualitative studies mentioned above is the amount of formulation effort typically expended by patients when they describe their seizure experiences (Schwabe *et al.*, 2008). In this context, formulation effort refers to the number and extent of hesitations, reformulations, restarts, repairs, and changes in grammatical construction (Schwabe, Howell and Reuber, 2007). Whereas speech in which PWE describe their seizure experiences is characterised by a high level of formulation effort as they struggle to communicate how exactly they experience their seizures, formulation effort is largely absent from the seizure accounts of PWNES (Schwabe *et al.*, 2008). Hesitations are a prominent aspect of formulation effort and can be detected using automated language analysis (Mirheidari *et al.*, 2017). We hypothesised that it is possible to automate the detection of hesitations as a marker of formulation effort in records of clinic conversations with seizures and that our findings would replicate those previously achieved using qualitative analyses.

Another potentially automatable method for measuring formulation effort involves the identification and analysis of pauses within the interaction. Pauses could be an indicator of formulation effort because they may reflect the difficulties the patient is facing with the accurate description of their complex seizure experiences (Plug, Sharrack and Reuber, 2009a). The automatic detection of pauses in speech has previously been used as an indicator of dementia (Mirheidari *et al.*, 2017). We hypothesised that the inclusion of one or several measures based on pauses would improve the classification performance.

In summary, the present study investigates whether features that can be automatically extracted from audio recordings and transcripts of speech as measures of formulation effort can be used to differentiate between epileptic and nonepileptic seizures. We hypothesise that it will be possible to differentiate between seizure accounts provided by PWE and PWNES using automatically measurable markers of formulation effort. We will explore the classification performance of a combination of these features using the Random Forest algorithm. Furthermore, we will explore to what extent independent features contribute to the classification performance using independent comparisons between groups and exploring the performance of the algorithm using different combinations of features.

2. Method

2.1 Participants

This study used recordings of doctor-patient interactions collected at the Royal Hallamshire Hospital in Sheffield between 2005 and 2013. The recordings have been used in previous CA research. Some of the interactions took place while patients were staying on a video-electroencephalography (EEG) unit (Schwabe, Howell and Reuber, 2007), whereas others were “naturally occurring” consultations conducted in an outpatient setting (Robson *et al*., 2012; Jenkins *et al*., 2015). Participants who were currently under examination to determine the cause of their seizures at the Royal Hallamshire Hospital were eligible for the original CA research and volunteered to participate. The final diagnoses for patients were determined using clinical assessment and/or a video-EEG recording of a typical seizure. Participants had not received a final diagnosis at the time of participating in the CA research. Our analysis incorporated the subset of the interview recordings from the three previous studies that resulted in a final diagnosis of epilepsy or PNES. We manually extracted a subsection of each interview in which the neurologist asked the patient to describe their first seizure using an open question. Previous research has observed that the questions that neurologists ask in an outpatient setting can be more restrictive due to the time pressures associated with these interactions, and that this can reduce the presence of CA observations that are important for the differential diagnosis process (Ekberg and Reuber, 2015). We chose this question because it gives the patient many appropriate response options. Focussing on this question allowed us to create the largest possible corpus of interviews (N=45, PNES n=24, PWE n=21), while ensuring that patients have been provided the opportunity to describe one particular seizure experience freely. We defined the end of the target subsection as the point when the neurologist either changed or accepted a change in topic agenda (Fehlenberg, 1986) away from the first seizure by asking questions unrelated to this topic. Changes in topic agenda introduced by patients could be an example of resistance to the question being asked which is a feature identified by previous CA research as indicative of a PNES description (Schwabe, Howell and Reuber, 2007).

2.2 Preprocessing

Audio-recordings were extracted from video recordings of the encounters, transcribed manually and further processed into Extensible Markup Format (XML). XML is a machine and human readable text format that is used to structure information. The transcripts were manually demarcated into individual turns within the conversation and each turn labelled with a speaker identifier, the start time and end time. The start and end time of the target subsection were noted and a new audio file consisting only of the target subsection was created using the AudioSegment function from Pydub (Hu and Wang, 2007). The raw text was converted to lowercase, punctuation and numerical digits were removed, contractions were expanded, and all words were converted to the corresponding lemma through lemmatization using the natural language toolkit (NLTK) in Python (Loper and Bird, 2002).

2.3 Feature Extraction

Seven features were designed as markers of formulation effort. Three of the features involved searching for a given word or word pair within the transcript. These features were the total number of hesitations (e.g. “hmm” or “erm”), the total number of repetitions (e.g. “I I don’t know”) and the presence or absence of words that suggest uncertainty (e.g. “sort of” or “might”). Four features involved measuring pauses within the interaction. Pauses were detected using the WebRTC Voice Activity Detector (VAD) from Google which checked whether each 10ms window contained speech or not. Only pauses greater than 30 milliseconds were included. Pauses in the speech of patients (patient pauses) were identified using a manually created function that aligned each pause with the turn labels on the XML transcript. Between speaker pauses were defined as pauses that crossed the turn allocation boundary. The four pause features were the ‘frequency of patient pauses’, 'average length of patient pauses', 'total length of patient pauses', and 'average length of between speaker pauses'.

2.4. Statistical Analysis

Group differences for each feature were compared using an independent t-test, Mann Whitney U test, or chi squared test as appropriate. The alpha level was set at 0.05. A Bonferroni correction was performed to reduce the risk of a type 1 error and resulted in an adjusted alpha level of 0.007 (0.05/7).

2.5. Classification

The Random Forest (Breiman, 2001) machine learning algorithm was used to investigate whether the features designed as markers of formulation effort were capable of differentiating between descriptions of epileptic or nonepileptic seizures. Random Forest is an algorithm that involves training many uncorrelated decision trees and subsequently making predictions based on the majority vote of all decision trees within the forest. The correlation between each decision tree is reduced by using a random sample of training data points to create each decision tree and selecting from a random subsample of features at each node within the decision tree. Reducing the correlation between trees improves the performance of the Random Forest algorithm (Breiman, 2001). The Random Forest algorithm was trained by applying the nested "leave-one-out" cross validation method (Vabalas *et al.*, 2019) and using the Scikit-learn toolkit in Python (Cournapeau *et al.*, 2011).

3. Result

3.1 Participants and seizure descriptions

A chi squared test of independence was performed to examine the relationship between gender and diagnosis. The relationship between these variables was significant, X^2 (1, N = 45) = 13, p <0.01. The PNES group included a higher proportion of women than the epilepsy group (women = 82.6% vs 23.8%). A Mann Whitney U test showed that there was no significant difference between the epilepsy and PNES groups in terms of vocabulary size (PWE median=101 vs. PWNES median = 100, U=212.5, p=0.187) and word count (PWE median=257 vs. PWNES median=210, U=187, p=0.071). A Chi squared test showed that there was no significant difference in word length distribution, X^2 (14, N = 45) = 15.2, p = 0.365.

3.2 Feature Comparison

There were significantly more hesitations and repetitions in the speech of PWE than that of PWNES (**Table 1**). There was no significant difference in terms of average pause length, pause frequency, total pause time, average length of between speaker pause and the presence or absence of key words associated with uncertainty (**Table 1**).

Table 1*: The mean (parametric tests) or median (non-parametric tests) for each variable*

Feature	PWNES	PWE	T value	P value
Hesitations †	2 (7)	9 (10)	U = 117.5	0.001
Repetitions †	2 (2.25)	3 (7)	U = 133.0	0.003
Pause frequency	45.4 (22.1)	46.4 (26.9)	0.135	0.893
Pause average †	0.996 (0.188)	0.786 (0.385)	159.000	0.018
Pause total	45.1 (24.5)	43.1 (31.6)	-0.238	0.813
Between speaker pause average †	1.15 (0.544)	0.922 (0.543)	198.000	0.112
Uncertainty keyword §	13/21 (61.9%)	10/24 (41.7%)	X^2 = 1.115	0.291

Note: results indicate mean (SDs) unless otherwise indicated. Adjusted alpha set at p<0.007.
t value given unless otherwise specified

† Mann Whitney U, median, and Interquartile range are reported because the variable is not normally distributed

§ Chi squared test, count and percentage because the variable is categoric

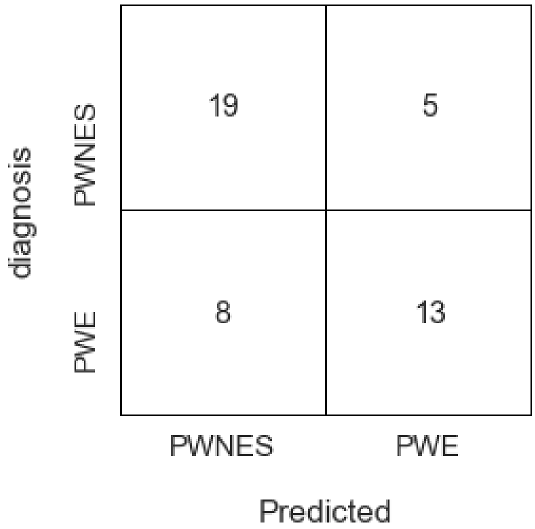
3.3 Random Forest performance

We compared the performance of the Random Forest algorithm using different combinations of the features (Table 2). The best performance was achieved when all formulation effort features were used (accuracy = 71%)(Figure 1), followed by hesitations and repetitions alone (accuracy = 68.9%), hesitations, repetitions, and the presence of uncertainty related words (accuracy = 64.5%) and all pause features (accuracy = 48.9%).

Table 2*: The accuracy, sensitivity, and specificity of the Random Forest algorithm trained using Leave-One-Out Cross Validation and different combinations of features*

Features	**Accuracy**	**Sensitivity**	**Specificity**	**F1 Score**
All features (7)	71%	61.9%	79.2%	67%
Hesitations & Repetitions (2)	68.9%	66.7%	70.8%	69%
Hesitations, Repetitions & Uncertainty (3)	64.5%	52.4%	75%	62%
Pause features (4)	48.9%	42.9%	54.2%	49%

Figure 1 *- A confusion matrix for differentiating between PWE and PWNES using the Random Forest model trained using all seven formulation effort features.*



diagnosis	PWNES	PWE
PWNES	19	5
PWE	8	13

4. Discussion

Our analysis supports the hypothesis that indicators of formulation effort that are automatically extracted from audio recordings and transcripts of seizure descriptions and inspired by observations made by CA can contribute to the differential diagnosis of epilepsy and PNES. PWE demonstrate significantly more formulation effort as indicated by hesitations and repetitions while describing their first seizure compared to PWNES. Furthermore, the seven features that were designed based on manual annotations of formulation effort were able to differentiate between PWE and PWNES with an accuracy of 71% using the Random Forest algorithm.

Previous qualitative research reported no difference in pause frequency and pause duration between PWE and PWNES (Walker *et al.*, 2020). Our findings that there was no significant difference in ‘patient pause frequency’, ‘total pause time’, ‘average pause length’, and ‘average length of between speaker pauses’ supports this finding. However, we observed an improvement in the performance of the Random Forest algorithm when the patient pause features were incorporated into the model. These findings illustrate the complex interaction between different linguistic and interactional features and demonstrate how a particular feature may make a diagnostic contribution in a particular context.

Our findings provide the basis of a more detailed exploration into possible contributions a fully automated analysis of language could make to the differentiation of epilepsy and PNES. We note that the discrimination we achieved by automated analysis of manually produced transcripts and audio clips was less accurate than the fully manual, qualitative approach based on the analysis of complete interactions and taking account of a wider range of potentially diagnostic features (Reuber *et al.*, 2009; Cornaggia *et al.*, 2012; Papagno *et al.*, 2017; Yao *et al.*, 2017; Biberon *et al.*, 2020). However, our study provides proof of principle that qualitatively described features can be translated into observations which can be made by a computer. The findings of this study provide encouragement for efforts to develop equivalent methods for other discriminating qualitative observations such as differences in the metaphoric conceptualisations of seizure experiences preferentially used by PWE and PWNES, or the extent to which subjective seizure experiences are volunteered, and how periods of unconsciousness are described (Schwabe, Howell and Reuber, 2007; Schwabe *et al.*, 2008; Plug, Sharrack and Reuber, 2009b).

In clinical practice, a TLOC stratification tool would be unlikely to be based on the predictive performance of language features alone as in this paper. In a clinical system, these features could be used alongside symptom checklists to train a classifier (which may also be more diagnostic if used with a Random Forest Classifier than regression based approaches)(Wardrope *et al.*, 2020). Future research should therefore explore the performance of a classifier trained using these features in tandem. Another reason for such a combined approach is that a clinical TLOC classification tool should not only be capable of predicting likely diagnoses of epilepsy and PNES but also of syncope. While little is known about the typical linguistic and interactional profile of patient descriptions of syncope, as stated above, this cause of TLOC can be differentiated very well from the two types of seizure based on symptom checklists.

The inclusion of language features into a fully automated clinical decision or stratification tool will require the use of an automatic speech recognition module. Although such systems will generate transcripts that are far less accurate than the manually produced transcripts used in this study, experience with a fully automatic “digital doctor” system, programmed to ask patients questions about memory problems and analyse their answers, suggests that remarkably high correct classification levels can be achieved with somewhat “faulty” transcripts (Mirheidari *et al.*, 2019; O’Malley *et al.*, 2021). While the switch from a conversation between clinician and patient to one between a talking head on a computer screen and the patient is likely to have significant consequences on how patients speak about their seizures, there are many similarities between the speech of patients between these two contexts (Walker *et al.*, 2020), so this aspect of automation may actually improve the diagnostic accuracy of a fully automatic classification system.

There are several limitations to this study. Firstly, the features used to approximate formulation effort may not capture all instances of formulation effort within the data. The features used in our analysis may suggest that patients are having difficulty describing their seizures by hesitating more, but another way that people can express formulation effort is by using meta-talk (Schiffrin, 1980) to describe their difficulties explaining their seizures (Schwabe, Howell and Reuber, 2007). Secondly, the sample size was small and the context in which the spoken seizure descriptions were recorded were heterogeneous. Although we used cross validation to demonstrate the machine learning algorithm’s ability to generalise to unseen data and to accommodate for the small dataset available for this analysis, it is difficult to evaluate the variance of a machine learning model using the “leave-one-out” cross validation method. Therefore, a larger sample size is required before we can be confident that this level of performance will be exhibited across other datasets. Finally, the analysis does not consider the type or severity of the seizures and future research should explore whether this influences the level of formulation effort that patients exhibit.

4. Conclusion

Our results provide evidence that PWE demonstrate increased formulation effort compared to PWNES when they describe their seizure experiences to a clinician and that features reflecting formulation effort, which can be extracted automatically from transcripts and audio recordings, can be used to differentiate between epilepsy and PNES. While in isolation, the accuracy of the method described in this study is lower than the analysis of full transcripts and recordings by trained experts. However, the described features can be incorporated into fully automated clinical decision tools also taking account of other data and may improve the diagnostic performance of these tools, especially in terms of the particularly challenging differentiation between epilepsy and PNES. Moreover, this analysis is faster to compute and cost effective to deliver at scale. Therefore, future research should explore the performance of these features alongside information about symptoms, patient history, and witness observations and whether these results can be maintained when data is collected using a fully automated methodology.

4. References

Biberon, J. *et al.* (2020) ‘Differentiating PNES from epileptic seizures using conversational analysis on French patients: A prospective blinded study’, *Epilepsy and Behavior*, 111, p. 107239. doi: 10.1016/j.yebeh.2020.107239.

Breiman, L. (2001) ‘Random forests’, *Machine Learning*, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324.

Brignole, M. *et al.* (2018) ‘2018 ESC Guidelines for the diagnosis and management of syncope’, *Kardiologia polska*, 76(8), pp. 1119–1198. doi: 10.5603/KP.2018.0161.

Chen, M. *et al.* (2019) ‘Value of witness observations in the differential diagnosis of transient loss of consciousness’, *Neurology*, 92(9), pp. E895–E904. doi: 10.1212/WNL.0000000000007017.

Cornaggia, C. M. *et al.* (2012) ‘Conversation analysis in the differential diagnosis of Italian patients with epileptic or psychogenic non-epileptic seizures: A blind prospective study’, *Epilepsy and Behavior*, 25(4), pp. 598–604. doi: 10.1016/j.yebeh.2012.09.003.

Cournapeau, D. *et al.* (2011) ‘Scikit-learn: Machine Learning in Python’, *Journal of Machine Learning Research*, 12(7), pp. 2825–2830.

Ekberg, K. and Reuber, M. (2015) ‘Can conversation analytic findings help with differential diagnosis in routine seizure clinic interactions?’, *Communication and Medicine*, 12(1), pp. 13–24. doi: 10.1558/cam.v12i1.26851.

Fehlenberg, D. (1986) *The Discourse of Medicine: Dialectics of Medical Interviews (Book).*, *Sociology of Health & Illness*. Greenwood Publishing Group. doi: 10.1111/1467-9566.ep11340200.

Hu, G. and Wang, D. (2007) ‘Auditory segmentation based on onset and offset analysis’, *IEEE Transactions on Audio, Speech and Language Processing*, 15(2), pp. 396–405. doi: 10.1109/TASL.2006.881700.

Jenkins, L. *et al.* (2015) ‘A brief conversation analytic communication intervention can change history-taking in the seizure clinic’, *Epilepsy and Behavior*, 52, pp. 62–67. doi: 10.1016/j.yebeh.2015.08.022.

Juarez-Garcia, A. *et al.* (2006) 'The costs of epilepsy misdiagnosis in England and Wales', *Seizure*, 15(8), pp. 598–605. doi: 10.1016/j.seizure.2006.08.005.

Kotsopoulos, I. A. W. *et al.* (2003) 'The diagnosis of epileptic and non-epileptic seizures', *Epilepsy Research*, 57(1), pp. 59–67. doi: 10.1016/j.eplepsyres.2003.10.014.

Loper, E. and Bird, S. (2002) 'NLTK: The Natural Language Toolkit', *arXiv preprint cs/0205028*. doi: 10.3115/1225403.1225421.

Mirheidari, B. *et al.* (2017) 'Toward the Automation of Diagnostic Conversation Analysis in Patients with Memory Complaints', *Journal of Alzheimer's Disease*, 58(2), pp. 373–387. doi: 10.3233/JAD-160507.

Mirheidari, B. *et al.* (2019) 'Dementia detection using automatic analysis of conversations', *Computer Speech and Language*, 53, pp. 65–79. doi: 10.1016/j.csl.2018.07.006.

O'Malley, R. P. D. *et al.* (2021) 'Fully automated cognitive screening tool based on assessment of speech and language', *Journal of Neurology, Neurosurgery and Psychiatry*, 92(1), pp. 12–15. doi: 10.1136/jnnp-2019-322517.

Papagno, C. *et al.* (2017) 'Differentiating PNES from epileptic seizures using conversational analysis', *Epilepsy and Behavior*, 76, pp. 46–50. doi: 10.1016/j.yebeh.2017.08.034.

Plug, L. and Reuber, M. (2009) 'Making the diagnosis in patients with blackouts: It's all in the history', *Practical Neurology*, 9(1), pp. 4–15. doi: 10.1136/jnnp.2008.161984.

Plug, L., Sharrack, B. and Reuber, M. (2009a) 'Conversation analysis can help to distinguish between epilepsy and non-epileptic seizure disorders: A case comparison', *Seizure*, 18(1), pp. 43–50. doi: 10.1016/j.seizure.2008.06.002.

Plug, L., Sharrack, B. and Reuber, M. (2009b) 'Seizure metaphors differ in patients' accounts of epileptic and psychogenic nonepileptic seizures', *Epilepsia*, 50(5), pp. 994–1000. doi: 10.1111/j.1528-1167.2008.01798.x.

Plug, L., Sharrack, B. and Reuber, M. (2010) 'Seizure, fit or attack the use of diagnostic labels by patients with epileptic or non-epileptic seizures', *Applied Linguistics*, 31(1), pp. 94–114. doi: 10.1093/applin/amp012.

Reuber, M. *et al.* (2009) 'Using interactional and linguistic analysis to distinguish between epileptic and psychogenic nonepileptic seizures: A prospective, blinded multirater study', *Epilepsy and Behavior*, 16(1), pp. 139–144. doi: 10.1016/j.yebeh.2009.07.018.

Reuber, M. *et al.* (2016) 'Value of patient-reported symptoms in the diagnosis of transient loss of consciousness', *Neurology*, 87(6), pp. 625–633. doi: 10.1212/WNL.0000000000002948.

Robson, C. *et al.* (2012) 'Catastrophising and normalising in patient's accounts of their seizure experiences', *Seizure*, 21(10), pp. 795–801. doi: 10.1016/j.seizure.2012.09.007.

Schiffrin, D. (1980) 'Meta- Talk: Organizational and Evaluative Brackets in Discourse', *Sociological Inquiry*, 50(3–4), pp. 199–236. doi: 10.1111/j.1475-682X.1980.tb00021.x.

Schwabe, M. *et al.* (2008) 'Listening to people with seizures: How can linguistic analysis help in the differential diagnosis of seizure disorders?', *Communication and Medicine*, 5(1), pp. 59–72. doi: 10.1558/cam.v5i1.59.

Schwabe, M., Howell, S. J. and Reuber, M. (2007) 'Differential diagnosis of seizure disorders: A conversation analytic approach', *Social Science and Medicine*, 65(4), pp. 712–724. doi: 10.1016/j.socscimed.2007.03.045.

Stiell, I. G. and Bennett, C. (2007) 'Implementation of Clinical Decision Rules in the Emergency Department', *Academic Emergency Medicine*, 14(11), pp. 955–959. doi: 10.1197/j.aem.2007.06.039.

Vabalas, A. *et al.* (2019) 'Machine learning algorithm validation with a limited sample size', *PLoS ONE*, 14(11), p. e0224365. doi: 10.1371/journal.pone.0224365.

Walker, G. *et al.* (2020) 'On the potential of phonetic analysis to distinguish between people with epilepsy and non-epileptic seizures', *International Journal of Applied Linguistics (United Kingdom)*, 30(1), pp. 92–109. doi: 10.1111/ijal.12268.

Wardrope, A. *et al.* (2020) 'Machine learning as a diagnostic decision aid for patients with transient loss of consciousness', *Neurology: Clinical Practice*, 10(2), pp. 96–105. doi: 10.1212/CPJ.0000000000000726.

Wardrope, A., Newberry, E. and Reuber, M. (2018) 'Diagnostic criteria to aid the differential diagnosis of patients presenting with transient loss of consciousness: A systematic review', *Seizure*, 61, pp. 139–148. doi: 10.1016/j.seizure.2018.08.012.

Xu, Y. *et al.* (2016) ‘Frequency of a false positive diagnosis of epilepsy: A systematic review of observational studies’, *Seizure*, 41, pp. 167–174. doi: 10.1016/j.seizure.2016.08.005.

Yao, Y. *et al.* (2017) ‘Conversation analysis in differential diagnosis between epileptic seizure and psychogenic nonepileptic seizure’, *Chinese Journal of Neurology*, 50(4), pp. 266–270. doi: 10.3760/cma.j.issn.1006-7876.2017.04.007.