HHS Public Access
Author manuscript
Cortex. Author manuscript; available in PMC 2021 November 01.

Published in final edited form as:
Cortex. 2020 November ; 132: 191–205. doi:10.1016/j.cortex.2020.08.020.

From discourse to pathology: Automatic identification of Parkinson's disease patients via morphological measures across three languages

Elif Eyigoz[1,#], **Melody Courson**[2,#], **Lucas Sedeño**[3], **Katharina Rogg**[4], **Juan Rafael Orozco-Arroyave**[5,6], **Elmar Nöth**[5], **Sabine Skodda**[7], **Natalia Trujillo**[8,9], **Mabel Rodríguez**[10], **Jan Rusz**[11], **Edinson Muñoz**[12], **Juan F. Cardona**[13], **Eduar Herrera**[14], **Eugenia Hesse**[3,15], **Agustín Ibáñez**[3,15,16,17,18], **Guillermo Cecchi**[1], **Adolfo M. García**[3,12,15,18,19,*]

[1]IBM Research, T. J. Watson Research Center, New York, USA

[2]Department of Psychology, Université de Montréal, CRIUGM Research Center, Montréal, Canada

[3]National Scientific and Technical Research Council (CONICET), Buenos Aires, Argentina

[4]Department of Social Psychology, University of Würzburg, Würzburg, Germany

[5]Pattern Recognition Lab, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany

[6]GITA Lab, Faculty of Engineering, University of Antioquia (UdeA), Medellín, Colombia

[7]Department of Neurology, Knappschaftskrankenhaus, Ruhr-University, Bochum, Germany

[8]Neuroscience Group, Faculty of Medicine, University of Antioquia (UdeA), Medellín, Colombia

[9]School of Public Health, University of Antioquia (UdeA), Medellín, Colombia

[10]Neurocognition Unit, National Institute of Mental Health, Prague, Czech Republic

[11]Department of Circuit Theory, Faculty of Electrical Engineering, Czech Technical University in Prague, Czech Republic

[*]**Corresponding author:** Adolfo M. García, Ph. D., Universidad de San Andrés & CONICET; Vito Dumas 284, B1644BID Victoria, Buenos Aires, Argentina; Phone: (54-11) 4725-7000. adolfomartingarcia@gmail.com.
[#]First authors

CRediT author statement

Elif Eyigoz: Methodology, Software, Formal analysis, Data curation, Writing - Original draft. **Melody Courson:** Writing - Original draft. **Lucas Sedeño:** Resources, Writing -Review & Editing. **Katharina Rogg:** Data curation, Resources. **Juan Rafael Orozco-Arroyave:** Data curation, Project administration, Writing - Review & Editing. **Elmar Nöth:** Data curation, Methodology. **Sabine Skodda:** Data curation, Methodology. **Natalia Trujillo:** Data curation, Methodology. **Mabel Rodríguez:** Data curation, Methodology. **Jan Rusz:** Data curation, Resources. **Edinson Muñoz:** Writing - Review & Editing, Resources. **Juan F. Cardona:** Writing - Review & Editing, Resources. **Eduar Herrera:** Writing - Review & Editing, Resources. **Eugenia Hesse:** Writing -Review & Editing, Resources. **Agustín Ibáñez:** Writing - Review & Editing, Resources. **Guillermo Cecchi:** Methodology, Formal analysis. **Adolfo M. García:** Conceptualization, Validation, Resources, Methodology, Writing - Original draft, Writing - Review & Editing, Supervision, Project administration.

Conflict of interest
None to declare

[12]Departamento de Lingüística y Literatura, Facultad de Humanidades, Universidad de Santiago de Chile, Santiago, Chile

[13]Instituto de Psicología, Universidad del Valle, Cali, Colombia

[14]Departamento de Estudios Psicológicos, Universidad Icesi, Cali, Colombia

[15]Universidad de San Andrés, Buenos Aires, Argentina

[16]Universidad Autónoma del Caribe, Barranquilla, Colombia

[17]Center for Social and Cognitive Neuroscience (CSCN), School of Psychology, Universidad Adolfo Ibáñez, Santiago de Chile, Chile

[18]Global Brain Health Institute, University of California, San Francisco, United States

[19]Faculty of Education, National University of Cuyo (UNCuyo), Mendoza, Argentina

Abstract

Embodied cognition research on Parkinson's disease (PD) points to disruptions of frontostriatal language functions as sensitive targets for clinical assessment. However, no existing approach has been tested for crosslinguistic validity, let alone by combining naturalistic tasks with machine-learning tools. To address these issues, we conducted the first classifier-based examination of morphological processing (a core frontostriatal function) in spontaneous monologues from PD patients across three typologically different languages. The study comprised 330 participants, encompassing speakers of Spanish (61 patients, 57 matched controls), German (88 patients, 88 matched controls), and Czech (20 patients, 16 matched controls). All subjects described the activities they perform during a regular day, and their monologues were automatically coded via morphological tagging, a computerized method that labels each word with a part-of-speech tag (e.g., noun, verb) and specific morphological tags (e.g., person, gender, number, tense). The ensuing data were subjected to machine-learning analyses to assess whether differential morphological patterns could classify between patients and controls and reflect the former's degree of motor impairment. Results showed robust classification rates, with over 80% of patients being discriminated from controls in each language separately. Moreover, the most discriminative morphological features were associated with the patients' motor compromise (as indicated by Pearson r correlations between predicted and collected motor impairment scores that ranged from moderate to moderate-to-strong across languages). Taken together, our results suggest that the morphological patterning, an embodied frontostriatal domain, may be distinctively affected in PD across languages and even under ecological testing conditions.

Keywords

Parkinson's disease; linguistic assessments; morphology; automated speech analysis; cross-linguistic validity

1. Introduction

Recent translational studies couched in the embodied cognition framework point to disruptions of frontostriatal language domains in Parkinson's disease (PD) as sensitive targets for clinical assessment (Birba et al., 2017; Cardona et al., 2014; García & Ibáñez, 2014; García & Ibáñez, 2018). Although biochemical, genetic, and neuroimaging tests have long proven quite successful at identifying and classifying patients with this disease, they are either limited due to their invasiveness, elevated costs, or dependence on highly specialized equipment that not all clinical centers possess. Ecological discourse-level assessments thus emerge as a promising complement, since they afford a simple, non-fatiguing, scalable, and cost-effective framework for patient discrimination (García et al., 2016a, 2018). So far, however, this approach has been employed in very few studies and it has not been tested for cross-linguistic validity, let alone via sophisticated classification tools. To address these issues, we report the first classifier-based examination of morphological processing, a core frontostriatal function (Carota, Bozic, & Marslen-Wilson, 2016; Nevat, Ullman, Eviatar, & Bitan, 2017; Newman, Supalla, Hauser, Newport, & Bavelier, 2010), in spontaneous monologues from PD patients and controls across three typologically different languages.

Neurolinguistic research on PD has been recently fueled by insights from the embodied cognition framework. Succinctly, this perspective posits that diverse higher-order processes are grounded in sensorimotor networks subserving functionally akin operations (Borghi & Cangelosi, 2014; Buccino, Colagè, Gobbi, & Bonaccorso, 2016; Pulvermüller, 2005; Pulvermüller & Fadiga, 2010). In particular, processing of action verbs (i.e., words denoting bodily motion) and morphosyntax (i.e., sequencing of hierarchically organized morphemes and words) involves differential recruitment of frontostriatal motor mechanisms underlying the preparation and execution of actions (García et al., 2019; Pulvermüller, 2013; Vigliocco, Vinson, Druks, Barber, & Cappa, 2011) as well as their organization into hierarchically organized sequences (Casado et al., 2018; Pulvermüller, 2014; Pulvermüller & Fadiga, 2010; Ullman, 2001). From an embodied perspective, this suggests that such linguistic domains involve reusing brain mechanisms specialized for processing similar types of information (Puvermüller, 2018).

As proposed in recent works (for a review, see Birba et al., 2017; Gallese & Cuccio, 2018), these linguistic domains should be distinctively impaired in PD patients, given their predominantly frontostriatal atrophy and diverse motor initiation and sequencing disorders (Dujardin et al., 2013; Helmich, Hallett, Deuschl, Toni, & Bloem, 2012; Liu et al., 2006; McKinlay, Grace, Dalrymple-Alford, & Roger, 2010; Rodriguez-Oroz et al., 2009; Samii, Nutt, & Ransom, 2004). Indeed, embodied research on this population has consistently revealed selective or differential deficits in accessing words denoting bodily movements or graspable objects (Bocanegra et al., 2017; Boulenger et al., 2008; Buccino et al., 2018; Cardona et al., 2014; Cotelli et al., 2007; Fernandino et al., 2013a, 2013b; García et al., 2018; Péran et al., 2009; Peran et al., 2013), and in processing diverse syntactic patterns (Bocanegra et al., 2015; García et al., 2018; Grossman, Carvell, & Peltzer, 1993; Grossman, Carvell, Stern, Gollomp, & Hurtig, 1992; Hochstadt, Nakano, Lieberman, & Friedman, 2006). In particular, high classification rates at the individual-patient level have been obtained by tracking these domains in naturalistic textual tasks (García et al., 2016a, 2018). This suggests that assessments of embodied language functions via discourse-level data could inform the cognitive characterization of PD in a simple, non-fatiguing, scalable, and cost-effective setting.

While the available studies on naturalistic texts have focused on action language and syntax, a highly relevant and under-explored target can be found in morphology –i.e., the internal structural organization of words (García, Sullivan, & Tsiang, 2017). Processing of inflectional and derivational morphology has been repeatedly associated with activity in various frontostriatal structures affected early in PD (Carota et al., 2016; Nevat et al., 2017; Newman et al., 2010). In fact, despite certain inconsistencies (García et al., 2020), previous word- and sentence-level research on PD has revealed morphological impairments in both comprehension (Grossman, 1999; Kemmerer, 1999; Terzi, Papapetropoulos, & Kouvelas, 2005) and production (Silveri et al., 2018; Terzi et al., 2005; Ullman, Corkin, Coppola, Hickok, Growdon, & Koroshetz, 1997; Zanini, Tavano, & Fabbro, 2010) tasks. Thus, the analysis of morphological patterns in spontaneous discourse might also index the impact of frontostriatal disruptions in PD.

Though certainly promising, research on these domains in PD is mainly limited by its lack of cross-linguistic validation. Word- and sentence-level studies have targeted only 11 separate languages, whereas text-level studies have been conducted in only four –and none of these reports has assessed more than a single language at a time (for a review, see Birba et al., 2017). Moreover, the only two studies that have performed automated classification analyses based on text-level performance have focused on Spanish only (García et al., 2016a, 2018). Given that the neurocognitive mechanisms of linguistic processing may vary widely depending on the typological properties of particular languages (Evans & Levinson, 2009; Kemmerer, 2014; Kemmerer & Eggleston, 2010), their assessment needs to be supported by findings from various languages (Calvo, Ibáñez, Muñoz, & García, 2017). To meet this imperative, the present study targeted morphological patterns in speech samples from Spanish (a Romance language), German (a Germanic language), and Czech (a Slavic language).

Briefly, then, our study aimed to assess whether morphological usage patterns are systematically affected in PD patients across languages. To this end, we employed automated text analysis (Bedi et al., 2014, 2015; Cohen, Alpert, Nienow, Dinzeo, & Docherty, 2008; Elvevåg, Foltz, Weinberg, & Goldberg, 2007), an approach that allows detecting sensitive features in spontaneous discourse to discriminate between healthy subjects and patients with different neuropsychiatric disorders (Bedi et al., 2014, 2015), including PD (García et al., 2016a). Specifically, we evaluated whether specific clusters of derivational and inflectional morphemes can classify PD patients and controls in each of our three target languages, and whether morphological usage patterns correlate with the patients' motor symptoms. Succinctly, then, our study seeks to open a clinically relevant, cross-cultural avenue for the neurolinguistics of movement disorders.

2. Methods

We report how we determined our sample size, all data exclusions (if any), all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. Participants

The study comprised 320 participants, recruited at three different international centers: Clínica Noel in Medellín, Colombia; the Knappschaftskrankenhaus of Bochum, Germany; and the General University Hospital of Prague, Czech Republic. The sample size for each language group proved larger than or similar to those of most previous studies in the field (for a review see Birba et al., 2017). None of these 320 participants was excluded from the reported analyses. Recruitment at each center encompassed non-demented PD patients and sociodemographically matched healthy controls (see Table 1). All participants were native speakers of each country's official language (Spanish, German, and Czech, respectively). The data used in the study belongs to systematic databases used in previous studies on Spanish (Orozco-Arroyave, Arias-Londoño, Vargas-Bonilla, González-Rátiva, & Nöth, 2014), German (Skodda, Gronheit, & Schlegel, 2011), and Czech (Rusz et al., 2013).

Spanish-speaking subjects included 61 patients (27 women) and 57 controls (28 women) [sex: $\chi^2(117) = 9.62$, $p = 1$; age: $t(117) = 0.54$, $p = .30$); education: $t(117) = 0.8855$, $p = .38$]. The German sample was composed of 88 patients (41 women) and 88 controls (44 women) [sex: $\chi^2(175) = 0$, $p = 1$; age: $t(175) = 2.06$, $p = .02$). Speakers of Czech included 20 patients (all male) and 16 controls (all male) [age: $t(35) = .19$, $p = .39$]. For further details about each sample, see Table 1.

Clinical diagnosis of PD was made by expert neurologists at each institution, in accordance with the United Kingdom PD Society Brain Bank criteria (Hughes, Daniel, Kilford, & Lees, 1992). Motor impairments in all patients were assessed with section III of the Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS-III) (Goetz et al., 2004) and the Hoehn &Yahr (H&Y) scale (Hoehn & Yahr, 1967). As reported in previous works based on the same groups of subjects (Orozco-Arroyave et al., 2014; Rusz et al., 2013; Skodda et al., 2011), and as revealed by individual clinical reports, patients exhibited canonical motor symptoms (mainly bradykinesia, freezing of gate, and resting tremor, but also rigidity and postural instability in several cases). As reflected in H&Y scores, and in line with MDS-UPDRSIII outcomes, most patients presented with bilateral or midline involvement (without balance impairment) and good recovery on pull test (Hoehn & Yahr, 1967). Some of them also reported non-motor dysfunctions, such as hyposmia, fatigue, weight loss, and sleep problems (these abnormalities were confirmed by caregivers, when present). No major visual or auditory compromise was reported. Importantly, no patient exhibited or had received treatment for primary speech or language disorders (including orofacial and abdominothoracal dyskinesias), and all of them were capable of providing fully coherent and cohesive verbal responses across different tests and in formal clinical interviews. The patients had no symptoms of Parkinson-plus and they lacked a history of other neurological or psychiatric disorders. None of them underwent deep brain stimulation or presented with signs of depression of cognitive dysfunction that could interfere with the measurements. Healthy controls reported no history of neurological or psychiatric disorders or substance abuse, and they did not have a background of motor symptomatology.

All participants gave written informed consent in accordance with the Declaration of Helsinki. The study was approved by the ethics committees of Universities of each country, namely, the University of Antioquia in Medellín (Colombia), Ruhr University Bochum (Germany), and the General University Hospital in Prague (Czech Republic). No part of the study procedures or analyses was pre-registered prior to the research being conducted.

2.2. Data collection

The data collection protocol began with the recording session. Roughly 30 minutes later, participants underwent the neurological evaluation, including administration of the MDSUPDRS-III and the H&Y tests. In the case of the Spanish- and German-speaking patients, the recording session began approximately 60 minutes after the morning dose of medication, to ensure the "on" state. Dopaminergic medication in these patients remained unchanged for at least four weeks before the examination (Orozco-Arroyave et al., 2014; Skodda et al., 2011).

Spanish-speaking participants were recorded in a sound-proof booth at the Clínica Noel in Medellín (Colombia), through a dynamic omni-directional microphone and a professional audio card. German-speaking and Czech-speaking participants were recorded in quiet rooms at the Knappschaftskrankenhaus of Bochum (Germany) and the General University Hospital of Prague (Czech Republic), with a Plantronics 550 head-set microphone and the microphone of a Panasonic NV-GS 180 videocamera, respectively. The sampling frequency of the recordings was 44.1 kHz with 16-bit resolution for the Spanish sample, 16 kHz with 16-bit resolution for the German sample, and 48 kHz with a 16-bit resolution for the Czech sample.

As in previous research on PD (García et al., 2016a), each participant was asked to talk about the activities they perform during a regular day in order to induce spontaneous speech. Average durations of the monologues for patients and controls were statistically similar in each country, with respective means of 45.24 (SD = 23.65) and 48.03 (SD = 28.85) seconds for Spanish [t(98) = 0.5269, p = .60], 31.16 (SD = 5.18) and 32.76 (SD = 6.21) seconds for German [t(174) = 1.7529, p = .08], and 129.50 (SD = 50.99) and 111.19 (SD = 61.76) seconds for Czech [t(34) = 0.9750, p = .33]. Spanish- and German-speaking patients were recorded during the "on" phase of antiparkinsonian medication. Recordings of the Czech patients were obtained during the "off" phase.

Audio recordings of each language were transcribed verbatim by native experts in linguistics from each country. Transcribed texts were punctuated following standard norms of each language –as endorsed by Real Academia Española (http://www.rae.es/) for Spanish, the Council for German Orthography (http://www.rechtschreibrat.com/) for German, and the Serbski Institut (https://www.serbski-institut.de) for Czech. Of note, inter-sentential elements (i.e., full stops) were identified and transcribed based strictly on grammatical criteria. The rare occurrences of unintelligible words were discarded from the transcripts.

The data used for this study has not been made publicly available because it contains identifiable information of the participants (derivable from acoustic and linguistic features of their recorded and transcribed speeches). Interested parties can contact the corresponding author, who will provide access to these records, exclusively for research purposes, upon signature of a formal data transfer agreement to be co-signed by the coauthors responsible for data collection.

2.3. Morphological tagging

Each participant's monologue was automatically coded via morphological tagging (MTag) (Jurafsky, 2018). This computerized method labels each word with a part-of-speech (POS) tag (e.g., noun, verb, adjective) and specific morphological tags (e.g., person, gender, number, case, tense, voice, mood, negation), based on well-established human-annotated corpora containing a few million words. Importantly, since most isolated word forms are ambiguous (i.e., they can manifest more than one POS), tagging relies on statistical algorithms, such as HMMs and MEMMs (Jurafsky, 2018), which factor in grammatical and/or semantic attributes of the words surrounding the target item to estimate the probability of a tag in its current context. The assigned tags are thus morphosyntactically disambiguated. For example, the sentence "*So I have some small businesses there, like cattle*" (from the Spanish "*Entonces yo tengo unos negocitos ahí, como de ganado*"*)* is tagged as shown in Table 2, which shows each lexical item, its corresponding POS, the morphological tags assigned, and the probability score with which the latter were assigned given the word's linguistic context. As shown in Table 2, MTag provides a list of labels capturing each word's overall morphological attributes, thus offering substantial information about a text's grammatical and semantic properties (Bertram, Pollatsek, & Hyönä, 2004; de Gispert & Mariño, 2008; Eyigöz, Gildea, & Oflazer, 2013; Habash & Rambow, 2007; Hajic et al., 2009; Shrivastava, Agrawal, Mohapatra, Singh, & Bhattacharya, 2005).

MTag of Spanish and German was conducted with Freeling (Carreras, Chao, Padró, & Padró, 2004; Padró & Stanilovsky, 2012), which yields a tagging accuracy of roughly 97% (Carmona et al., 1998). MTag of Czech was performed with Morphodita (Hajič, 2004; Straková, Straka, & Hajic, 2014) –based on the pre-trained linguistic model included in the package–, which warrants an accuracy of 95.03% on the whole tag set (Straková et al., 2014). Of note, Freeling uses morphological tags based on the proposal by EAGLES (Ide & Véronis, 1993), which encodes morphological features for most European languages. Therefore, the morphological tags used by Freeling are similar to those used by Morphodita. Links to the complete lists of tags used by each software can be found in the Supplementary Material (section 1).

2.4. Feature extraction method

Each participant's monologue was tagged separately, and so was each word within the monologues (i.e., in no case was a single morphological tag assigned to a sequence of words, such as nominal compounds). For Spanish and German, Freeling uses the following POS categories: noun, verb, adjective, adverb, pronoun, determiner, conjunction, adposition, and interjection. For a given POS category, a list of attributes is specified, which in turn can be assigned a value from a set of possible values. For example, a Spanish verb may have the attribute 'tense', and its value can be 'present', 'past' or 'future'. A verb may also have the attribute 'person', which can have the value 'first', 'second' or 'third'. A noun may have the attribute 'case', which can be either 'nominative', 'accusative', 'dative', or 'genitive'. Similarly, an adjective can be tagged with a degree, which can be superlative or comparative.

Each POS category was combined with one of its attributes and the value of that attribute (e.g., verb in past tense, noun in nominative case). The number of POS and attribute-value pairs in a monologue was computed as a percentage to adjust for interindividual differences in fluency (i.e., number of words per minute). Attribute extraction was similar for Czech, as the morphological tags used by Morphodita and the tags used by Freeling overlap significantly.

As shown in Table 2, Freeling provides a score between 0 and 1 for each tagged word, indicating the likelihood of the assigned tag (Carreras et al., 2004; Padró & Stanilovsky, 2012). Since such scores are not computed automatically by Morphodita, we computed them manually for Czech. To this end, we counted frequencies of POS-attribute-value pairs and POS tags in a large Czech corpus. To obtain a score between 0 and 1, we divided the frequencies of POS-attribute-value pairs by the frequency of their POS tags. Through this normalization procedure, frequencies were converted to probability scores, indicating the probability estimate of observing an attribute-value pair given the POS tag of the word –e.g., the probability that the word *gone* is in past tense (POS is verb, attribute is tense, value is past), given that the word *gone* is a verb. Similarly, we divided the frequencies of POS tags by the total number of words in the corpus, to obtain probability estimates of all POS tags (Manning & Schütze, 1999).

For all languages, the following measures were computed on the scores for each POS-attribute-value pair: minimum, maximum, mean, standard deviation, skewness and kurtosis. Standard deviation, skewness and kurtosis can be interpreted as three different measures of the stability of the score. Standard deviation represents the dispersion of values around the mean, skewness represents the asymmetry of the distribution of values, and kurtosis represents the presence/absence of outlying values. Therefore, they constitute complementary measures of morphological consistency across the datasets, showing how consistently they figured in the linguistic output of each sample and, hence, how robust they prove as potential discriminatory features. In particular, the statistical capabilities of the framework allow detecting covert patterns distributed throughout the texts. Thus, for example, the consistency of a morphological feature (e.g., present-tense suffixation) can be operationalized in terms of mean, minimal, and maximal scores, rate, or stability indices (standard deviation, kurtosis, and skewness).

This information was computed for four sets of features in each language, namely: (i) rates of *POS categories* (e.g., verb, noun); rates of *POS-attribute-value pairs* (e.g., verbs in past tense); (iii) *POS* scores (e.g., mean score of verbs); and (iv) *POS-attribute-value pair* scores (e.g., mean scores of verbs in past tense). Also, given that Morphodita provides information about each word's frequency of use in different styles (including the most frequent contemporary style, a less frequent but still standard style, or colloquial, archaic or bookish styles), this feature was also considered in the analysis of Czech texts.

2.5. Data analysis: prediction and inference

For each language, we tested three different classifier algorithms (stochastic gradient descent, support vector machines, and logistic regression) with a leave-one-out-cross-validation (LOOCV) scheme to evaluate which of them was the best to model the data from each language, and learn to discriminate whether each monologue belonged to a PD patient or a healthy control. Moreover, feature selection was performed during this step to optimize classification performance (see section 2.5.1 for methodological details). Next, for interpretation purposes, we determined the minimal set of best features for classifying between PD patients and controls in each language using recursive feature elimination (RFE) (Guyon, Weston, Barnhill, & Vapnik, 2002)–section 2.5.2. Finally, the minimal set of features was used to predict the participants' degree of motor impairment (as revealed by MDS-UPDRS-III scores). A set of regression analyses was then conducted between the predicted and collected MDS-UPDRS-III scores of PD patients of each language –section 2.5.3. The complete process, from tagging to prediction, is summarized in Figure 1.

2.5.1. Classification setup—In a LOOCV setting, as used in previous spontaneous speech analyses (Bedi et al., 2014, 2015; García et al., 2016a), each sample (i.e., each monologue) is tested against the rest of the dataset, which is called a fold (i.e., round) of cross-validation; the number of folds is equal to the number of samples. In the present case, each speech sample was classified as PD or control as the result of a prediction based on all the other speech samples of the dataset. The sample that was left-out for testing is called the test sample, and all other samples are called the training dataset of a cross-validation fold. For each algorithm applied in this step (stochastic gradient descent, support vector machines, and logistic regression), we used a grid search approach (an exhaustive method that builds a model for every combination of hyperparameters specified, evaluates each resulting model, and determines the best one) to explore the best hyperparameters for each model –see details of the initialization of hyperparameters in the Supplementary Material (section 2).

Feature selection (i.e., elimination of non-discriminative features) was performed to optimize classification performance. Feature selection was implemented in the LOOCV scheme as follows: discriminative features in each fold were identified by using only the training samples of the fold, and then a classifier was trained using only the selected features of the training dataset, which in turn was used to predict the label (PD, control) of the left-out test sample of that fold. Given the high-dimensionality and the relative small sample sizes of our data, we combined multiple feature selection methods because this approach yields robust results and provide more robust feature subset than a single feature selection technique (Saeys, Abeel, & Van de Peer, 2008; Saeys, Inza, & Larrañaga, 2007). As an initial filtering, we used univariate feature selection methods. First, we obtained a *p*-value for each feature by computing a *t*-test between the samples from PD patients and the samples from controls. Then, we eliminated features with a *p*-value higher than .01 corrected by FDR via Bejamini-Hochberg's procedure. Next, each feature that passed the FDR correction was analyzed via an ANOVA *f*-test. We eliminated features with low *f*-test scores using a non-parametric method for setting a threshold for elimination (see Supplementary material, section 3). Finally, we implemented a stability-selection procedure (Meinshausen & Bühlmann, 2010), which computes a score for each feature indicating its importance. We eliminated features with low importance scores using a non-parametric method for setting a threshold for elimination (Supplementary material, section 3).

In summary, we performed the following feature-selection methods in LOOCV folds subsequently: (1) *t*-test, (2) ANOVA *f*-test, (3) stability-selection, such that each step was applied on the features that were selected in the prior step. Feature-selection was performed using only the training data of a given fold, so that feature-selection did not observe the test sample. Therefore, feature-selection in each fold did not observe the entire dataset. In contrast, we present a feature-selection method that used the entire dataset in the following section, for interpretation purposes.

Finally, the values obtained through the LOOCV schemes were used to compute accuracy, recall, precision, area under the curve (AUC) scores, and the confusion matrix for the discrimination between PD patients and controls in each language. As stated before, such analyses were performed through different classifiers. In this sense, note that since the three languages differed in their number of samples, subjects, and features, each of them could reach its highest classification results based on different classifiers (Hastie, Tibshirani, & Friedman, 2009). In the Results section, only the best performing classifier is reported for each language (see Table 3).

2.5.2. Interpretation of features—The features obtained in the previous analysis (section 2.5.1) were further scrutinized to detect the most informative subset of features for each language. To this end, we first selected a subset of the features through RFE. If the subset was smaller than the set of features RFE was run on, then we reran RFE on the selected subset. We performed this procedure until RFE no longer returned a smaller subset of features –in other words, until RFE *converged*. The most informative subset of features thus selected was then used in LOOCV experiments.

The outcome of this method was the set of features yielding optimal classification performance between patients and controls for each language in the dataset. The classification performance of this method is presented for interpretation purposes and should not be taken to generalize to other datasets. In order to emphasize this distinction between the previous analysis and the method presented in this section, below we refer to the latter as a “non-generalizable” analysis.

2.5.3. Correlations between morphological features and motor compromise —We further assessed whether morphological usage patterns correlated with the patients’ degree of motor compromise as indexed by MDS-UPDRS-III scores. To this end, we computed a predicted MDS-UPDRS-III score for each participant by fitting a multiple regression model to the collected MDS-UPDRS-III scores with the minimal set of features previously selected in the classification analysis (section 2.5.2). To estimate the *p*-value for the correlation between inferred and actual MDS-UPDRS-III score, we performed 100,000 permutations of the inferred values and computed the correlation with the actual ones. We then used this distribution of random correlations to estimate the probability of finding by chance the same or higher values than the obtained inferred correlation. The association between the predicted and the collected MDS-UPDRS-III scores was calculated via Pearson’s r index. A high correlation between predicted and actual MDS-UPDRS-III scores would reveal a strong link between morphological usage patterns and motor compromise across individual patients.

3. Results

3.1. Classification of speech samples

A high classification accuracy was achieved in each language separately. The best performing classifiers discriminating PD patients' monologues from the ones of the controls yielded an accuracy rate of 71% and an AUC of 73 (with LR) for Spanish, an accuracy rate of 71% and an AUC of 76 (with SGD) for German, and an accuracy rate of 80% and an AUC of 83 (with SGD) for Czech. As expected, metrics were even higher for the 'non-generalizable' analysis (in which we selected the minimal set of features through the RFE method), with values of accuracy and AUC of: 82% and 89 for Spanish (with SVM), 81% and 84 for German (with SVM), and 94% and 97 for Czech (with LR) (see Table 3 for a summary of all the classification metrics for the cross-validation and the 'non-generalizable' analysis).

In addition, a set of four main features in the classification was extracted for each language. Each set of four features contains the features that were the most important in the classification of speech samples for each language. Results show that the four main features for speech sample classification vary across languages (Table 4).

Importantly, complementary analyses showed that similar classification rates are obtained even when participants are pooled apart depending on their sex, age, and education level (when available). For details, see Supplementary material (section 5).

3.2. Correlations between morphological features and motor compromise levels

The predicted MDS-UPDRS-III scores obtained fitting a multiple regression model with the minimal set of features (section 2.5.2) showed a significant correlation between the automated analysis of naturalistic speech and the degree of motor compromise as indexed by MDS-UPDRS-III scores (Figure 3). Correlations ranged from moderate to moderate-to-strong across languages (Spanish: Pearson's $r = 0.35$, $p < .01$; German: Pearson's $r = 0.26$, $p = .01$; Czech: Pearson's $r = 0.61$, $p < .001$) –for statistical details about each feature from multiple regressions, see Supplementary material (section 6).

Of note, correlations with MDS-UPDRS-III scores yielded similar results even when patients were analyzed in separate groups differing in sex, age, and education level (when available), as well as MDS-UPDRS-III scores, H&Y scores, and years since diagnosis. For details, see Supplementary material (section 5).

4. Discussion

This is the first cross-linguistic investigation of spontaneous speech on PD. Using automated analyses of Spanish, German, and Czech monologues we found that different clusters of morphological patterns consistently discriminated between patients and controls with high accuracy. Moreover, those differential patterns were significantly correlated with the patients' degree of motor compromise in each language. Taken together, these results underscore the embodied domain of morphological usage as an ecologically and trans-linguistically valid target for clinical research on PD.

Across the three languages tested, specific collections of morphological features allowed classifying patients from controls with over 70% accuracy, an outcome that actually surpassed 80% when only the optimal discriminatory features were considered. Given the physiopathology of early-stage PD (Rodriguez-Oroz et al., 2009; Samii et al., 2004), this pattern supports the overarching view that frontostriatal networks are critically involved in morphological processing (Carota et al., 2016; Nevat et al., 2017; Newman et al., 2010). More particularly, it aligns with previous evidence that PD patients differ from controls in several morphological skills, such as detection of obligatory affixes (Grossman, Carvell, Stern, Gollomp, & Hurtig, 1992), word derivation for specific lexical classes (Silveri et al., 2018), and past-tense inflection (Longworth, Keenan, Barker, Marslen-Wilson, & Tyler, 2005; Terzi et al., 2005; Ullman, Corkin, Coppola, Hickok, Growdon, Koroshetz, et al., 1997). Our results extend these findings by showing that morphological assessments in PD might also possess two crucial features: ecological and trans-linguistic validity.

Note, in this sense, that our approach captures differences in usage rather than deficits proper. This is noteworthy given that experiments testing for morphological impairments in PD have yielded a mixture of significant (Grossman et al., 1992; Longworth et al., 2005; Silveri et al., 2018; Terzi et al., 2005; Ullman, Corkin, Coppola, Hickok, Growdon, Koroshetz, et al., 1997) and non-significant (Grossman, 1999; Longworth et al., 2005; Macoir et al., 2013; Silveri et al., 2018) differences between patients and controls. Therefore, classification analyses capturing dissimilarities in morphological patterning across texts might outperform typical controlled tasks in their capacity to identify PD patients. In fact, previous analyses of grammatical features in spontaneous monologues yielded 75% accuracy in classifying individuals with and without this disease (García et al., 2016a). Here lies another potential advantage of favoring more naturalistic set-ups in the linguistic assessment of patients with movement disorders (Birba et al., 2017; García et al., 2018).

The distinct sensitivity of morphological usage to the impact of PD was further underscored by the regression analyses. In fact, in each language, the minimal set of discriminative morphological features was significantly correlated with the patients' degree of motor compromise, as tapped by the MDS-UPDRS-III. This finding mirrors previous results from controlled tasks, showing that inflectional morphology deficits in PD correlate with the patients' symptoms of hypokinesia (Ullman, Corkin, Coppola, Hickok, Growdon, Koroshetz, et al., 1997). Moreover, it also extends outcomes from previous spontaneous speech analyses in PD showing that syntagmatic properties of their texts allowed predicting MDS-UPDRS-III scores with 77% accuracy (García et al., 2016a). Our results afford a promising synthesis of these antecedents, showing that morphological patterns might reflect the motoric impact of PD even in ecological verbal settings.

Note that the obtained *r* values were smallest for German, followed by Spanish and then by Czech. This gradient was inverse to that of the sample sizes in each case (German > Spanish > Czech). This is likely because the ordinary least squares regression aims to minimize the sum of the squares of the errors –i.e., the differences between the observed MDS-UPDRS-III scores and those predicted by the linear function. In this setting, the in-sample estimates of the mean squared error (MSE) has been shown to decrease as the number of samples used for training increases (Friedman, Hastie, & Tibshirani, 2001). Yet, beyond sample size differences, this variance may also be partly explained by other factors. In particular, the range of MDS-UPDRS-III scores was much wider for the Spanish-speaking group (the 25th and 75th percentiles are 28 and 52) than for the German- speaking group (the 25th and 75th percentiles are 14.75 to 30). This might also influence the differential correlation outcomes in each of these languages, as variables with narrower ranges are harder to fit. While this remains speculative and calls for further research, the emergence of robust results in each language despite this variability speaks to the apparent sensitivity of the approach.

Given their Romance, Germanic, and Slavic roots, the languages involved in our study are structurally dissimilar. It is, therefore, unsurprising that the specific morphological features affording the above results rates varied considerably across languages, with different top features for Spanish (e.g., use of proper nouns and present tense), German (e.g., case-marking for determiners and gender in nominal groups), and Czech (e.g. person-marking and gender). Despite such idiosyncrasies, it is interesting to note that those top features, across languages, mainly point to specific word classes and inflectional (as opposed to derivational) morphemes. Though preliminary, this observation mirrors previous spontaneous speech analyses in PD that underscored specific word classes (e.g., pronouns, negative adverbs) as contributing to patient/control discrimination (García et al., 2016a). In addition, it aligns and with the fact that the morphological tasks yielding more consistent deficits in this population involved inflectional operations –in particular, past-tense formation (Longworth et al., 2005; Terzi et al., 2005; Ullman, Corkin, Coppola, Hickok, Growdon, Koroshetz, et al., 1997). More importantly, the robustness of our results despite major typological variability suggests meets the imperative of crosslinguistic validity, a cornerstone to support any claim of broad generalizability in the study of language mechanisms, in general (Evans & Levinson, 2009; Kemmerer, 2014; Kemmerer & Eggleston, 2010), and their dysfunction in neurodegenerative conditions, in particular (Calvo et al., 2017).

It is also worth noting that the above findings, across all three languages, proved consistent even when samples were partitioned in terms of sex, age, and education (when data thus allowed), and in terms of the patients' MDS-UPDRS-III scores, H&Y scores, and years since diagnosis. Thus, the relation between morphological usage and motor dysfunction would not seem to be biased by particular sociodemographic and clinical profiles. Although a number of caveats must still be noted in this regard (see "Limitations" below), this opens fruitful avenues to further explore the generalizability of our results in future research.

Additionally, we showed that our main results, as obtained with a LOOCV approach, were highly consistent in 25-fold settings. This further attests to the robustness of our findings. As it happens, although LOOCV has been successfully used with relatively small samples in previous spontaneous speech analyses (Bedi et al., 2014; 2015; García et al., 2016a), this method does not allow for repetition of the cross-validation experiments, potentially leading to over-fitting issues (Gareth, Hastie, Witten, & Tibshirani, 2013). This is not the case when more than only one sample is left out for testing. Upon performing multiple variations of the latter approach (over 25 cross-validation folds for each language with 100 combinations per metric), we found strikingly similar (and very high) classification results, suggesting that our findings were not an artifact of over-fitting in a LOOCV setting.

From a larger theoretical perspective, these findings have implications for fine-tuning our understanding of the role of motor circuits in language processing. In line with Birba et al. (2017), we propose that the implication of frontostriatal networks in morphology reflects the embodied nature of this domain: just like these networks are crucial for processing *hierarchically organized sequences of actions* so do they prove crucial to process *hierarchically organized sequences of morphemes*. This claim, which represents a straightforward extension of the so-called Disrupted Motor Grounding Hypothesis (Birba et al., 2017; García & Ibáñez, 2018), implies that the intimate relation between morphology and frontostriatal networks would be a manifestation of ‘grounding’, namely: the recycling of lower-level sensorimotor networks supporting functionally germane operations (Dehaene & Cohen, 2007) –for similar claims, see Ullman (2001). Incidentally, this postulation highlights the contributions of the embodied cognition framework as an organizing principle for understanding the neurocognitive particularities of patients with neurodegenerative motor disorders (Gallese & Cuccio, 2018; García & Ibáñez, 2018).

Our results also have clinical implications. Disruptions of embodied processes have been proposed as potential signatures of motor-network degeneration in early disease stages (Abrevaya et al., 2017; Bocanegra et al., 2015), irrespective of the patients’ overall cognitive status (Bocanegra et al., 2017; García et al., 2016a, 2016b, 2018), and even in preclinical stages (García et al., 2017a, 2017b; Kargieman et al., 2014). While these claims have been advanced by reference to other embodied domains (syntax and action-language processing), our results point to morphology as a sensitive domain for neurolinguistic research on PD. Note, in this sense, that despite presenting similar H&Y scores our patient samples featured great heterogeneity in terms of their years since diagnosis and their MDS-UPDRS-III scores. Moreover, whereas the Spanish and German samples were tested during the “on” phase of antiparkinsonian medication, Czech patients were evaluated in the “off” phase. Moreover, all three patient samples had similar H&Y scores despite differing in their years since diagnosis. Yet, high classification rates were nonetheless obtained in each language. Tentatively, this indicates that morphological patterns might robustly discriminate between patients and controls despite major variability in disease progression and dopamine bioavailability –an important milestone in this research field (Birba et al., 2017; García & Ibáñez, 2018).

Of course, the gold standard for diagnosis, prognosis, and monitoring of PD involves a combination of validated clinical tests complemented by biochemical, neuroimaging, and (when appropriate) genetic biomarkers (e.g. Rodriguez-Oroz et al., 2009; Samii et al., 2004). Far from a replacement of any of these elements, our approach might represent a useful complement to them –especially in institutions from low-income countries lacking neuroimaging, biochemical, or genetic expertise (Parra et al., 2018). In particular spontaneous speech tasks are undemanding, non-stressful, and non-fatiguing, which sets them apart from several standardized clinical tests (García et al., 2016a). Moreover, they involve virtually no costs, and they can be administered remotely and massively. Finally, our automated approach is also advantageous in that it can handle vast amounts of data in little time, circumventing the biases of human analysis. Moreover, its relevance for translational research has been demonstrated in previous studies yielding high classification rates for psychiatric populations, including ecstasy users (Bedi et al., 2014), schizophrenics, maniacs (Mota et al., 2012), and bipolar subjects (Mota, Furtado, Maia, Copelli, & Ribeiro, 2014). Together with the only previous application of this approach to PD (García et al., 2016a), our novel findings indicate that this framework can also afford breakthroughs in the context of neurological disorders.

5. Limitations and avenues for further research

Admittedly, however, our study presents a number of limitations, mainly due to differences in the standard clinical protocols adopted in each international center. First, the size of our samples was not homogeneous across the three languages tested. However, even the *n* of the smallest sample (i.e., the Czech cohort) was similar to or even larger than that of previous morphological studies on PD (Longworth et al., 2005; Macoir et al., 2013; Silveri et al., 2018; Terzi et al., 2005). While this speaks to the relevance of present results even for our smallest sample, future applications of our approach should aim to recruit groups of comparable size across languages.

Second, we lacked data on the participants' education level in two countries. Still, patients and controls were matched for education in the large Spanish sample, which yielded similar results to those obtained in for German and Czech. Moreover, results in the Spanish sample remained proved consistent even when participants were binned into lower and higher education subgroups. Also, note that the significant correlation between morphological patterns and motor impairments is a result that can hardly be explained principally by potential educational confounds –and, more generally, morphological processing in spontaneous discourse is a domain that would not seem to depend on formal education. Still, it would be useful to replicate this investigation while controlling for this factor in all language groups.

Third, Spanish- and German-speaking patients were recorded during the "on" phase of antiparkinsonian medication, while recordings of the Czech patients were obtained during the "off" phase. Although our study did not involve statistical comparisons among the three patient samples and robust results were obtained in all of them, levodopa or dopamine agonists are known to modulate other linguistic domains, such as lexical access (Boulenger et al., 2008) and word fluency (Herrera and Cuetos, 2012). Therefore, future cross-linguistic assessments of morphology in this disease should aim to homogenize dopaminergic levels across samples to circumvent this caveat of our research.

Fourth, the software used to analyze Czech data (Morphodita) was not the same as that used for the other two languages (as it happens, Freeling is not available for Czech). While it is true that Morphodita and Freeling yield very similar tagging accuracies in their respective languages (95% and 97%, respectively), and even though both programs use nearly identical tags, software-specific discrepancies may have introduced a source of inconsistency across results for each language. This reservation should be acknowledged in the present results and addressed through the use of identical analysis programs in future research.

Finally, potential extensions of our study should also include more detailed clinical characterizations of the patients, crucially controlling for their overall cognitive status. Although previous evidence suggests that embodied language disruptions in PD hold irrespective of the patients' domain-general (dys)functions (Birba et al., 2017; García & Ibáñez, 2018), direct testing of this factor would represent an important complement to the data reported herein. Moreover, valuable insights could be gained by factoring in physiopathological information ideally by combining linguistic assessments with anatomo-functional brain measures, as done in previous PD research (Grossman et al., 2003; Isaacs, McMahon, Angwin, Crosson, & Copland, 2019; Magdalinou et al., 2018; Pereira et al., 2009). Finally, given that other aspects of language have revealed deficits in very early and even preclinical stages of PD (Birba et al., 2017; García et al., 2017a), our approach should tested on *de novo* patients, with continual monitoring in the course of disease.

6. Conclusion

This study has offered unprecedented evidence that morphological patterning in spontaneous discourse can represent a robust and cross-linguistically valid target for research on PD. Such a finding represent a promising extension of recent proposals capitalizing on embodied cognition principles to advance innovations in clinical neuroscience. Further efforts in this direction may lead to fruitful synergies at the crossing of theoretical and applied research in the field.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is partially supported by grants from CONICET; CONICYT/FONDECYT Regular (1170010); FONCYT-PICT 2017-1818; FONCYT-PICT 2017-1820; FONDAP 15150012; COLCIENCIAS (1106-744-55314); CODI at the University of Antioquia (PRG2017-15530); the European Union's Horizon 2020 Research and Innovation Programme under Marie Sklodowska-Curie grant agreement No 766287; the Czech Ministry of Health (NV19-04-00120); OP VVV MEYS (Research Center for Informatics, CZ.02.1.01/0.0/0.0/16_019/0000765); Sistema General de Regalías (BPIN2018000100059), Universidad del Valle (CI 5316); Programa Interdisciplinario de Investigación Experimental en Comunicación y Cognición (PIIECC), Facultad de Humanidades, USACH; GBHI ALZ UK-20-639295; and the Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat), funded by the National Institutes of Aging of the National Institutes of Health under award number R01AG057234, an Alzheimer's Association grant (SG-20-725707-ReDLat), the Rainwater Foundation, and the Global Brain Health Institute. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health, Alzheimer's Association, Rainwater Charitable Foundation, or Global Brain Health Institute.

References

Abrevaya S, Sedeño L, Fitipaldi S, Pineda D, Lopera F, Buritica O, … García AM (2017). The road less traveled: Alternative pathways for action-verb processing in Parkinson's disease. Journal of Alzheimer's Disease, 55(4), 1429–1435. doi: 10.3233/jad-160737

Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, … Corcoran CM (2015). Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophrenia, 1, 15030. doi: 10.1038/npjschz.2015.30 [PubMed: 27336038]

Bedi G, Cecchi GA, Slezak DF, Carrillo F, Sigman M, & de Wit H (2014). A window into the intoxicated mind? Speech as an index of psychoactive drug effects. Neuropsychopharmacology, 39(10), 2340–2348. doi: 10.1038/npp.2014.80 [PubMed: 24694926]

Bertram R, Pollatsek A, & Hyönä J (2004). Morphological parsing and the use of segmentation cues in reading Finnish compounds. Journal of Memory and Language, 51(3), 325–345. doi: 10.1016/j.jml.2004.06.005

Birba A, García-Cordero I, Kozono G, Legaz A, Ibáñez A, Sedeño L, & García AM (2017). Losing ground: Frontostriatal atrophy disrupts language embodiment in Parkinson's and Huntington's disease. Neuroscience and Biobehavioral Reviews, 80, 673–687. doi: 10.1016/j.neubiorev.2017.07.011 [PubMed: 28780312]

Bocanegra Y, García AM, Lopera F, Pineda D, Baena A, Ospina P, … Cuetos F (2017). Unspeakable motion: Selective action-verb impairments in Parkinson's disease patients without mild cognitive impairment. Brain and Language, 168, 37–46. doi: 10.1016/j.bandl.2017.01.005 [PubMed: 28131052]

Bocanegra Y, García AM, Pineda D, Buritica O, Villegas A, Lopera F, … Ibáñez A (2015). Syntax, action verbs, action semantics, and object semantics in Parkinson's disease: Dissociability, progression, and executive influences. Cortex, 69, 237–254. doi: 10.1016/j.cortex.2015.05.022 [PubMed: 26103601]

Borghi AM, & Cangelosi A (2014). Action and language integration: From humans to cognitive robots. Topics in Cognitive Science, 6(3), 344–358. doi: 10.1111/tops.12103 [PubMed: 24943900]

Boulenger V, Mechtouff L, Thobois S, Broussolle E, Jeannerod M, & Nazir TA (2008). Word processing in Parkinson's disease is impaired for action verbs but not for concrete nouns. Neuropsychologia, 46(2), 743–756. doi: 10.1016/j.neuropsychologia.2007.10.007 [PubMed: 18037143]

Buccino G, Colagè I, Gobbi N, & Bonaccorso G (2016). Grounding meaning in experience: A broad perspective on embodied language. Neuroscience and Biobehavioral Reviews, 69, 69–78. doi: 10.1016/j.neubiorev.2016.07.033 [PubMed: 27477443]

Buccino G, Dalla Volta R, Arabia G, Morelli M, Chiriaco C, Lupo A, … Quattrone A (2018). Processing graspable object images and their nouns is impaired in Parkinson's disease patients. Cortex, 100, 32–39. doi: 10.1016/j.cortex.2017.03.009 [PubMed: 28413070]

Calvo N, Ibáñez A, Muñoz E, & García AM (2017). A core avenue for transcultural research on dementia: on the cross-linguistic generalization of language-related effects in Alzheimer's disease and Parkinson's disease. International Journal of Geriatric Psychiatry. doi: 10.1002/gps.4712

Cardona JF, Kargieman L, Sinay V, Gershanik O, Gelormini C, Amoruso L, … Ibáñez A (2014). How embodied is action language? Neurological evidence from motor diseases. Cognition, 131(2), 311–322. doi: 10.1016/j.cognition.2014.02.001 [PubMed: 24594627]

Carmona J, Cervell S, Marquez L, Martí MA, PadrO L, Placer R, … Turmo J (1998). An environment for morphosyntactic processing of unrestricted Spanish text. Proceedings of the First International Conference on Language Resources and Evaluation (LREC 98), Granada, Spain.

Carota F, Bozic M, & Marslen-Wilson W (2016). Decompositional Representation of Morphological Complexity: Multivariate fMRI Evidence from Italian. Journal of Cognitive Neuroscience, 28(12), 1878–1896. doi: 10.1162/jocn_a_01009 [PubMed: 27417206]

Carreras X, Chao I, Padró L, & Padró M (2004). FreeLing: An Open-Source Suite of Language Analyzers. Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004), Granada, Spain.

Casado P, Martin-Loeches M, Leon I, Hernandez-Gutierrez D, Espuny J, Muñoz F, … de Vega M (2018). When syntax meets action: Brain potential evidence of overlapping between language and motor sequencing. Cortex, 100, 40–51. doi: 10.1016/j.cortex.2017.11.002 [PubMed: 29212607]

Cohen AS, Alpert M, Nienow TM, Dinzeo TJ, & Docherty NM (2008). Computerized measurement of negative symptoms in schizophrenia. Journal of Psychiatric Research, 42(10), 827–836. [PubMed: 17920078]

Cotelli M, Borroni B, Manenti R, Zanetti M, Arévalo A, Cappa SF, & Padovani A (2007). Action and object naming in Parkinson's disease without dementia. European Journal of Neurology, 14(6), 632–637. doi: 10.1111/j.1468-1331.2007.01797.x [PubMed: 17539940] de Gispert A, & Mariño JB (2008). On the impact of morphology in English to Spanish statistical MT. Speech Communication, 50(11–12), 1034–1046. doi: 10.1016/j.specom.2008.05.003

Dehaene S, & Cohen L (2007). Cultural recycling of cortical maps. Neuron, 56(2), 384–398. [PubMed: 17964253]

Dujardin K, Tard C, Duhamel A, Delval A, Moreau C, Devos D, & Defebvre L (2013). The pattern of attentional deficits in Parkinson's disease. Parkinsonism and Related Disorders, 19(3), 300–305. doi: 10.1016/j.parkreldis.2012.11.001 [PubMed: 23196037]

Elvevåg B, Foltz PW, Weinberg DR, & Goldberg TE (2007). Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. Schizophrenia Research, 93(1–3), 304–316. [PubMed: 17433866]

Evans N, & Levinson SC (2009). The myth of language universals: language diversity and its importance for cognitive science. Behavioral and Brain Sciences, 32(5), 429–448; discussion 448–494. doi: 10.1017/S0140525X0999094X [PubMed: 19857320]

Eyigöz E, Gildea D, & Oflazer K (2013). Simultaneous word-morpheme alignment for statistical machine translation. Paper presented at the NAACL HTL, Atlanta, Georgia, USA.

Fernandino L, Conant LL, Binder JR, Blindauer K, Hiner B, Spangler K, & Desai RH (2013a). Parkinson's disease disrupts both automatic and controlled processing of action verbs. Brain and Language, 127(1), 65–74. doi: S0093–934X(12)00140-X [PubMed: 22910144]

Fernandino L, Conant LL, Binder JR, Blindauer K, Hiner B, Spangler K, & Desai RH (2013b). Where is the action? Action sentence processing in Parkinson's disease. Neuropsychologia, 51(8), 1510–1517. doi: 10.1016/j.neuropsychologia.2013.04.008 [PubMed: 23624313]

Friedman J, Hastie T, & Tibshirani R (2001). The Elements of Statistical Learning. New York: Springer.

Gallese V, & Cuccio V (2018). The neural exploitation hypothesis and its implications for an embodied approach to language and cognition: Insights from the study of action verbs processing and motor disorders in Parkinson's disease. Cortex, 100, 215–225. doi: 10.1016/j.cortex.2018.01.010 [PubMed: 29455947]

García AM, & Ibáñez A (2014). Words in motion: Motor-language coupling in Parkinson's disease. Translational Neuroscience, 5(2). doi: 10.2478/s13380-014-0218-6

García AM, Sullivan W, & Tsiang S (2017). An Introduction to Relational Network Theory: History, Principles, and Descriptive Applications. London: Equinox.

García AM, Bocanegra Y, Birba A, Orozco-Arroyave JR, Sedeño L & Ibáñez A (2020). Disruptions of frontostriatal language functions in Parkinson's disease In Martin C & Preedy VR (eds.), The Neuroscience of Parkinson's Disease, volume 2 (pp. 413–430). London: Elsevier Academic Press.

García AM, Carrillo F, Orozco-Arroyave JR, Trujillo N, Vargas Bonilla JF, Fittipaldi S, … Cecchi GA (2016a). How language flows when movements don't: An automated analysis of spontaneous discourse in Parkinson's disease. Brain and Language, 162, 19–28. doi: 10.1016/j.bandl.2016.07.008 [PubMed: 27501386]

García AM, Abrevaya S, Kozono G, Cordero IG, Córdoba M, Kauffman MA, … Ibáñez A (2016b). The cerebellum and embodied semantics: evidence from a case of genetic ataxia due to STUB1 mutations. Journal of Medical Genetics, 54, 114–124. doi: 10.1136/jmedgenet-2016-104148 [PubMed: 27811304]

García AM, Sedeño L, Trujillo N, Bocanegra Y, Gomez D, Pineda D, … Ibáñez A (2017a). Language deficits as a preclinical window into Parkinson's disease: Evidence from asymptomatic parkin and dardarin mutation carriers. Journal of the International Neuropsychological Society, 22, 150–158.

García AM, Bocanegra Y, Herrera E, Pino M, Muñoz E, Sedeño L, & Ibáñez A (2017b). Action-semantic and syntactic deficits in subjects at risk for Huntington's disease. Journal of Neuropsychology. doi: 10.1111/jnp.12120

García AM, Bocanegra Y, Herrera E, Moreno L, Carmona J, Baena A, … Ibáñez A (2018). Parkinson's disease compromises the appraisal of action meanings evoked by naturalistic texts. Cortex, 100, 111–126. doi: 10.1016/j.cortex.2017.07.003 [PubMed: 28764852]

García AM, & Ibáñez A (2018). When embodiment breaks down: Language deficits as novel avenues into movement disorders. Cortex, 100, 1–7. doi: 10.1016/j.cortex.2017.12.022 [PubMed: 29496261]

García AM, Moguilner S, Torquati K, García-Marco E, Herrera E, Muñoz E, … Ibáñez A (2019). How meaning unfolds in neural time: Embodied reactivations can precede multimodal semantic effects during language processing. NeuroImage, 197, 439–449. doi: 10.1016/j.neuroimage.2019.05.002 [PubMed: 31059796]

Gareth J, Hastie T, Witten D, & Tibshirani R (2013). An Introduction to Statistical Learning. New York: Springer.

Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, … Seidl L (2004). Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. Movement Disorders, 19(9), 1020–1028. doi: 10.1002/mds.20213 [PubMed: 15372591]

Grossman M (1999). Sentence Processing in Parkinson's Disease. Brain and Cognition, 40(2), 387–413. doi: 10.1006/brcg.1999.1087 [PubMed: 10413567]

Grossman M, Carvell S, & Peltzer L (1993). The sum and substance of it: the appreciation of mass and account quantifiers in Parkinson's disease. Brain and Language, 44, 351–384. [PubMed: 8319078]

Grossman M, Carvell S, Stern MB, Gollomp S, & Hurtig HI (1992). Sentence comprehension in Parkinson's disease: the role of attention and memory. Brain and Language, 42, 347–384. [PubMed: 1611464]

Grossman M, Cooke A, DeVita C, Lee C, Alsop D, Detre J, … Hurtig HI (2003). Grammatical and resource components of sentence processing in Parkinson's disease: an fMRI study. Neurology, 60(5), 775–781. [PubMed: 12629232]

Guyon I, Weston J, Barnhill S, & Vapnik V (2002). Gene selection for cancer classification using support vector machines. Machine learning, 46, 389–422.

Habash N, & Rambow O (2007). Arabic diacritization through full morphological tagging. Paper presented at the NAACL HLT, Dorchester, NY, USA.

Hajič J (2004). Disambiguation of rich inflection: computational morphology of Czech: Vydavatel: Karolinum.

Hajic J, Ciaramita M, Johansoon R, Kawahara D, Marti MA, Marquez L, … Zhang Y (2009). The CoNLL-2009 shared task: syntactic and semantic dependencies in multiple languages. Paper presented at the Thirteenth conference on computational natural language learning (CoNLL): shared task, Boulder, Colorado, USA.

Hastie T, Tibshirani R, & Friedman J (2009). The elements of statistical learning: data mining, inference, and prediction: Springer Science & Business Media

Helmich RC, Hallett M, Deuschl G, Toni I, & Bloem BR (2012). Cerebral causes and consequences of parkinsonian resting tremor: a tale of two circuits? Brain, 135(Pt 11), 3206–3226. doi: 10.1093/brain/aws023 [PubMed: 22382359]

Herrera E, & Cuetos F (2012). Action naming in Parkinson's disease patients on/off dopamine. Neuroscience Letters, 513(2), 219–222. doi: 10.1016/j.neulet.2012.02.045 [PubMed: 22387157]

Hochstadt J, Nakano H, Lieberman P, & Friedman J (2006). The roles of sequencing and verbal working memory in sentence comprehension deficits in Parkinson's disease. Brain and Language, 97(3), 243–257. doi: 10.1016/j.bandl.2005.10.011 [PubMed: 16332387]

Hoehn MM, & Yahr MD (1967). Parkinsonism: onset, progression and mortality. Neurology, 17(5), 427–442. [PubMed: 6067254]

Hughes AJ, Daniel SE, Kilford L, & Lees AJ (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. Journal of Neurology, Neurosurgery and Psychiatry, 55(3), 181–184.

Ide N, & Véronis J (1993). Background and context for the development of a Corpus Encoding Standard. EAGLES working paper. Available at <http://www.cs.vassar.edu/CES/CES3.ps.gz>.

Isaacs ML, McMahon KL, Angwin AJ, Crosson B, & Copland DA (2019). Functional correlates of strategy formation and verbal suppression in Parkinson's disease. NeuroImage: Clinical, 22, 101683. doi: 10.1016/j.nicl.2019.101683 [PubMed: 30711682]

Jurafsky D, Martin JH (2018). Speech and Language Processing. New York: Prentice Hall.

Kargieman L, Herrera E, Baez S, García AM, Dottori M, Gelormini C, … Ibáñez A (2014). Motor-Language Coupling in Huntington's Disease Families. Frontiers in Aging Neuroscience, 6, 122. doi: 10.3389/fnagi.2014.00122 [PubMed: 24971062]

Kemmerer D (1999). Impaired comprehension of raising-to-subject constructions in Parkinson's disease. Brain and Language, 66, 311–328. [PubMed: 10190993]

Kemmerer D (2014). Body ownership and beyond: connections between cognitive neuroscience and linguistic typology. Conscious Cogn, 26, 189–196. doi: 10.1016/j.concog.2014.03.009 [PubMed: 24759681]

Kemmerer D, & Eggleston A (2010). Nouns and verbs in the brain: Implications of linguistic typology for cognitive neuroscience. Lingua, 120(12), 2686–2690. doi: 10.1016/j.lingua.2010.03.013

Liu W, McIntire K, Kim SH, Zhang J, Dascalos S, Lyons KE, & Pahwa R (2006). Bilateral subthalamic stimulation improves gait initiation in patients with Parkinson's disease. Gait Posture, 23(4), 492–498. doi: 10.1016/j.gaitpost.2005.06.012 [PubMed: 16098748]

Longworth CE, Keenan SE, Barker RA, Marslen-Wilson WD, & Tyler LK (2005). The basal ganglia and rule-governed language use: evidence from vascular and degenerative conditions. Brain, 128(Pt 3), 584–596. doi: 10.1093/brain/awh387 [PubMed: 15659423]

Macoir J, Fossard M, Merette C, Langlois M, Chantal S, & Auclair-Ouellet N (2013). The role of basal ganglia in language production: evidence from Parkinson's disease. Journal aof Parkinson's Disease, 3(3), 393–397. doi: 10.3233/JPD-130182

Magdalinou NK, Golden HL, Nicholas JM, Witoonpanich P, Mummery CJ, Morris HR, … Warren JD (2018). Verbal adynamia in parkinsonian syndromes: behavioral correlates and neuroanatomical substrate. Neurocase, 24(4), 204–212. doi: 10.1080/13554794.2018.1527368 [PubMed: 30293517]

Manning CD, Schütze H (1999). Foundations of statistical natural language processing: Cambridge, Masschussetts: MIT Press.

McKinlay A, Grace RC, Dalrymple-Alford JC, & Roger D (2010). Characteristics of executive function impairment in Parkinson's disease patients without dementia. Journal of the International Neuropsychological Society, 16(2), 268–277. doi: 10.1017/s1355617709991299 [PubMed: 20003582]

Meinshausen N, & Bühlmann P (2010). Stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72, 417–473.

Mota NB, Furtado R, Maia PPC, Copelli M, & Ribeiro S (2014). Graph analysis of dream reports is especially informative about psychosis. Scientific Reports, 4, 3691. doi: 10.1038/srep03691 [PubMed: 24424108]

Mota NB, Vasconcelos NA, Lemos N, Pieretti AC, Kinouchi O, Cecchi GA, … Ribeiro S (2012). Speech graphs provide a quantitative measure of thought disorder in psychosis. PLoS One, 7(4), e34928. doi: 10.1371/journal.pone.0034928 [PubMed: 22506057]

Nevat M, Ullman MT, Eviatar Z, & Bitan T (2017). The neural bases of the learning and generalization of morphological inflection. Neuropsychologia, 98, 139–155. doi: 10.1016/j.neuropsychologia.2016.08.026 [PubMed: 27575853]

Newman AJ, Supalla T, Hauser P, Newport EL, & Bavelier D (2010). Dissociating neural subsystems for grammar by contrasting word order and inflection. Proceedings of the National Academy of Science USA, 107(16), 7539–7544. doi: 10.1073/pnas.1003174107

Orozco-Arroyave JR, Arias-Londoño DD, Vargas-Bonilla JF, González-Rátiva MC, & Nöth E (2014). New Spanish speech corpus database for the analysis of people suffering from Parkinson's disease. Paper presented at the Proceedings of the 9th Language Resources and Evaluation Conference (LREC), Málaga, Spain.

Padró L. i., & Stanilovsky E (2012). Freeling 3.0: Towards wider multilinguality. Paper presented at the LREC2012, Málaga, Spain.

Parra MA, Baez S, Allegri R, Nitrini R, Lopera F, Slachevsky A, … Ibáñez A (2018). Dementia in Latin America: Assessing the present and envisioning the future. Neurology, 90(5), 222–231. doi: 10.1212/WNL.0000000000004897 [PubMed: 29305437]

Péran P, Cardebat D, Cherubini A, Piras F, Luccichenti G, Peppe A, … Sabatini U (2009). Object naming and action-verb generation in Parkinson's disease: A fMRI study. Cortex, 45(8), 960–971. doi: 10.1016/j.cortex.2009.02.019 [PubMed: 19368905]

Peran P, Nemmi F, Meligne D, Cardebat D, Peppe A, Rascol O, … Sabatini U (2013). Effect of levodopa on both verbal and motor representations of action in Parkinson's disease: a fMRI study. Brain and Language, 125(3), 324–329. doi: 10.1016/j.bandl.2012.06.001 [PubMed: 22841350]

Pereira JB, Junque C, Marti MJ, Ramirez-Ruiz B, Bartres-Faz D, & Tolosa E (2009). Structural brain correlates of verbal fluency in Parkinson's disease. Neuroreport, 20(8), 741–744. doi: 10.1097/WNR.0b013e328329370b [PubMed: 19349926]

Pulvermüller F (2005). Brain mechanisms linking language and action. Nature Reviews Neuroscience, 6(7), 576–582. doi: 10.1038/nrn1706 [PubMed: 15959465]

Pulvermüller F (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. Trends in Cognitive Sciences, 17(9), 458–470. doi: 10.1016/j.tics.2013.06.004 [PubMed: 23932069]

Pulvermüller F (2014). The syntax of action. Trends in Cognitive Sciences, 18(5), 219–220. doi: 10.1016/j.tics.2014.01.001 [PubMed: 24582824]

Puvermüller F (2018). Neural reuse of action perception circuits for language, concepts and communication. Progress in Neurobiology, 160, 1–44. doi: 10.1016/j.pneurobio.2017.07.001 [PubMed: 28734837]

Pulvermüller F, & Fadiga L (2010). Active perception: sensorimotor circuits as a cortical basis for language. Nature Reviews Neuroscience, 11(5), 351–360. doi: 10.1038/nrn2811 [PubMed: 20383203]

Rodriguez-Oroz MC, Jahanshahi M, Krack P, Litvan I, Macias R, Bezard E, & Obeso JA (2009). Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. The Lancet Neurology, 8(12), 1128–1139. doi: 10.1016/s1474-4422(09)70293-5 [PubMed: 19909911]

Rusz J, Cmejla R, Tykalova T, Ruzickova H, Klempir J, Majerova V, … Ruzicka E (2013). Imprecise vowel articulation as a potential early marker of Parkinson's disease: effect of speaking task. Journal of the Acoustical Society of America, 134(3), 2171–2181. doi: 10.1121/1.4816541 [PubMed: 23967947]

Saeys Y, Abeel T, & Van de Peer Y (2008). Robust feature selection using ensemble feature selection techniques. Paper presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases.

Saeys Y, Inza I, & Larrañaga P (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, 23(19), 2507–2517. [PubMed: 17720704]

Samii A, Nutt JG, & Ransom BR (2004). Parkinson's disease. The Lancet, 363(9423), 1783–1793. doi: 10.1016/S0140-6736(04)16305-8

Shrivastava M, Agrawal N, Mohapatra B, Singh P, & Bhattacharya P (2005). Morphology based natural language processing tools for Indian languages. Paper presented at the 4th Annual Inter Research Institute Student Seminar in Computer Science, IIT, Kanpur, India.

Silveri MC, Traficante D, Lo Monaco MR, Iori L, Sarchioni F, & Burani C (2018). Word selection processing in Parkinson's disease: When nouns are more difficult than verbs. Cortex, 100, 8–20. doi: 10.1016/j.cortex.2017.05.023 [PubMed: 28669510]

Skodda S, Gronheit W, & Schlegel U (2011). Intonation and speech rate in Parkinson's disease: general and dynamic aspects and responsiveness to levodopa admission. Journal of Voice, 25(4), e199–205. doi: 10.1016/j.jvoice.2010.04.007 [PubMed: 21051196]

Straková J, Straka M, & Hajic J (2014). Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. Paper presented at the ACL (System Demonstrations).

Terzi A, Papapetropoulos S, & Kouvelas ED (2005). Past tense formation and comprehension of passive sentences in Parkinson's disease: evidence from Greek. Brain and Language, 94(3), 297–303. doi: 10.1016/j.bandl.2005.01.005 [PubMed: 16098379]

Ullman MT (2001). A neurocognitive perspective on language: The declarative/procedural model. Nature Reviews Neuroscience, 2(10), 717–726. doi: 10.1038/35094573 [PubMed: 11584309]

Ullman MT, Corkin S, Coppola M, Hickok G, Growdon JH, & Koroshetz WJ (1997). A neural dissociation within language: evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. Journal of Cognitive Neuroscience, 9(2), 266–276. [PubMed: 23962016]

Vigliocco G, Vinson DP, Druks J, Barber H, & Cappa SF (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. Neuroscience & Biobehavioral Reviews, 35(3), 407–426. doi: 10.1016/j.neubiorev.2010.04.007 [PubMed: 20451552]

Zanini S, Tavano A, & Fabbro F (2010). Spontaneous language production in bilingual Parkinson's disease patients: Evidence of greater phonological, morphological and syntactic impairments in native language. Brain and Language, 113(2), 84–89. doi: 10.1016/j.bandl.2010.01.005 [PubMed: 20188407]

Author Manuscript

Author Manuscript

Author Manuscript

Highlights

- We examined morphology in natural speech from Parkinson's disease (PD) patients.
- Our study comprised speakers of three languages: Spanish, German, and Czech.
- Morphological features classified > 80% of patients and controls in each language.
- The most discriminative features correlated with the patients' motor compromise.
- This embodied frontostriatal domain may afford crosslinguistic signatures of PD.

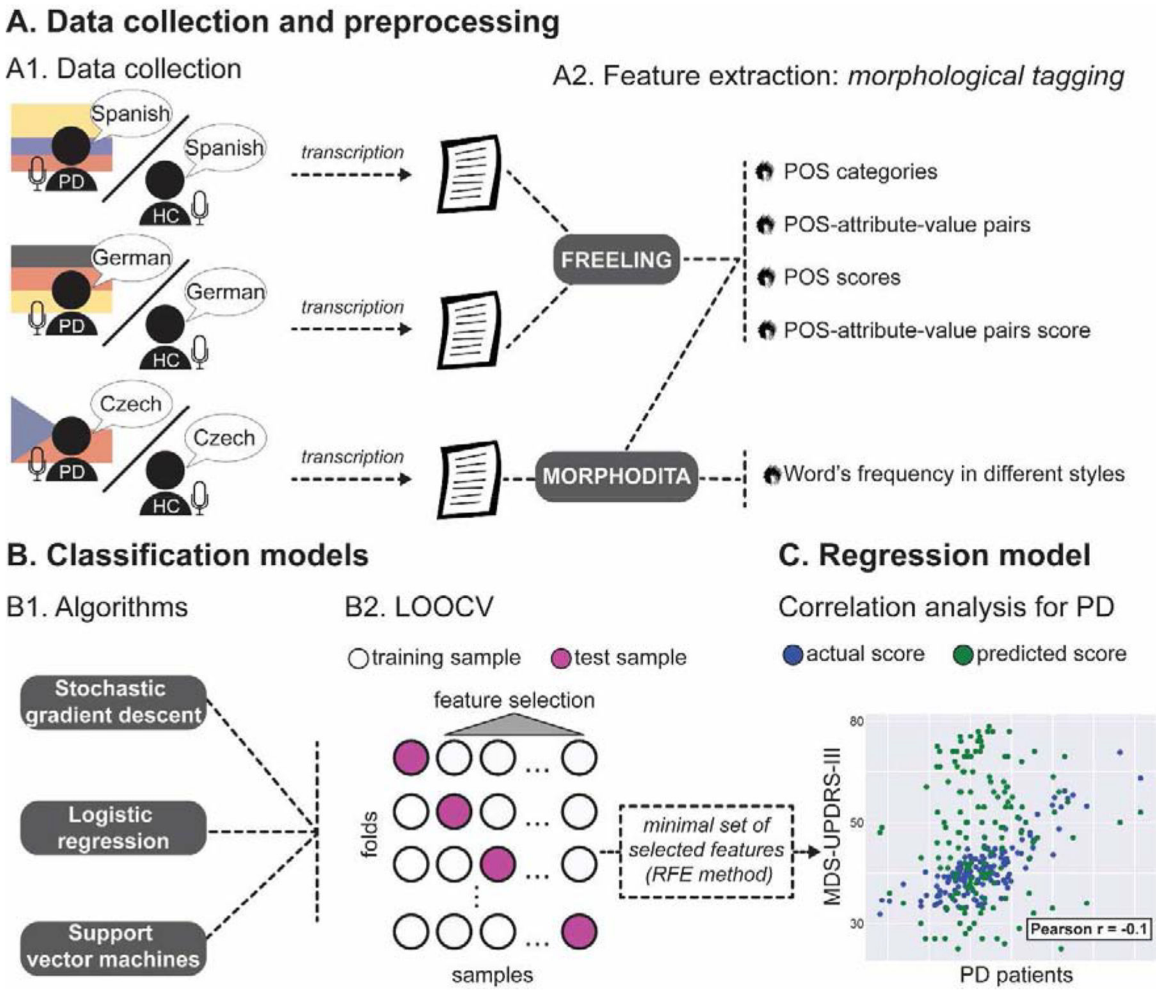

Figure 1.

Flowchart summarizing the pipeline for morphological tagging and cross-validation analysis of the participants' monologues. **A. Data collection and preprocessing.** In sound-proof rooms, PD patients and healthy controls (HCs) from each language group were asked to narrate the activities they perform during a regular day (**A1**). Morphological tagging (MTag) for each monologue was conducted with Feeling for Spanish and German, and with Morphodita for Czech (**A2**). **B. Classification models.** For each dataset in each language, we predicted each participant's monologue as belonging to a PD patient or a HC using three classification models: stochastic gradient descent (SGD), support vector machine (SVM), and logistic regression (LR) (**B1**). For cross-validation, we applied a leave-one-out-cross-validation (LOOCV) approach, combined with the elimination of non-discriminative features in each fold, considering only the training samples of each fold (based on the analysis of each feature through a *t*-test, an ANOVA *F*-test, and a stability-selection method). Then, the minimal set of best features for the classification in each language from the previous step was identified based on a recursive feature elimination (RFE) method (these correspond to the 'non-generalizable' analysis from section 2.5.2). (**B2**). **C. Regression model.** Finally, these minimal set of selected features were used to predict the participant's degree of motor impairment (as measured by the MDS-UPDRS-III scores), and these predicted scores were used to perform correlation analyses between the selected feature set and the actual scores of the PD patients (**C1**).

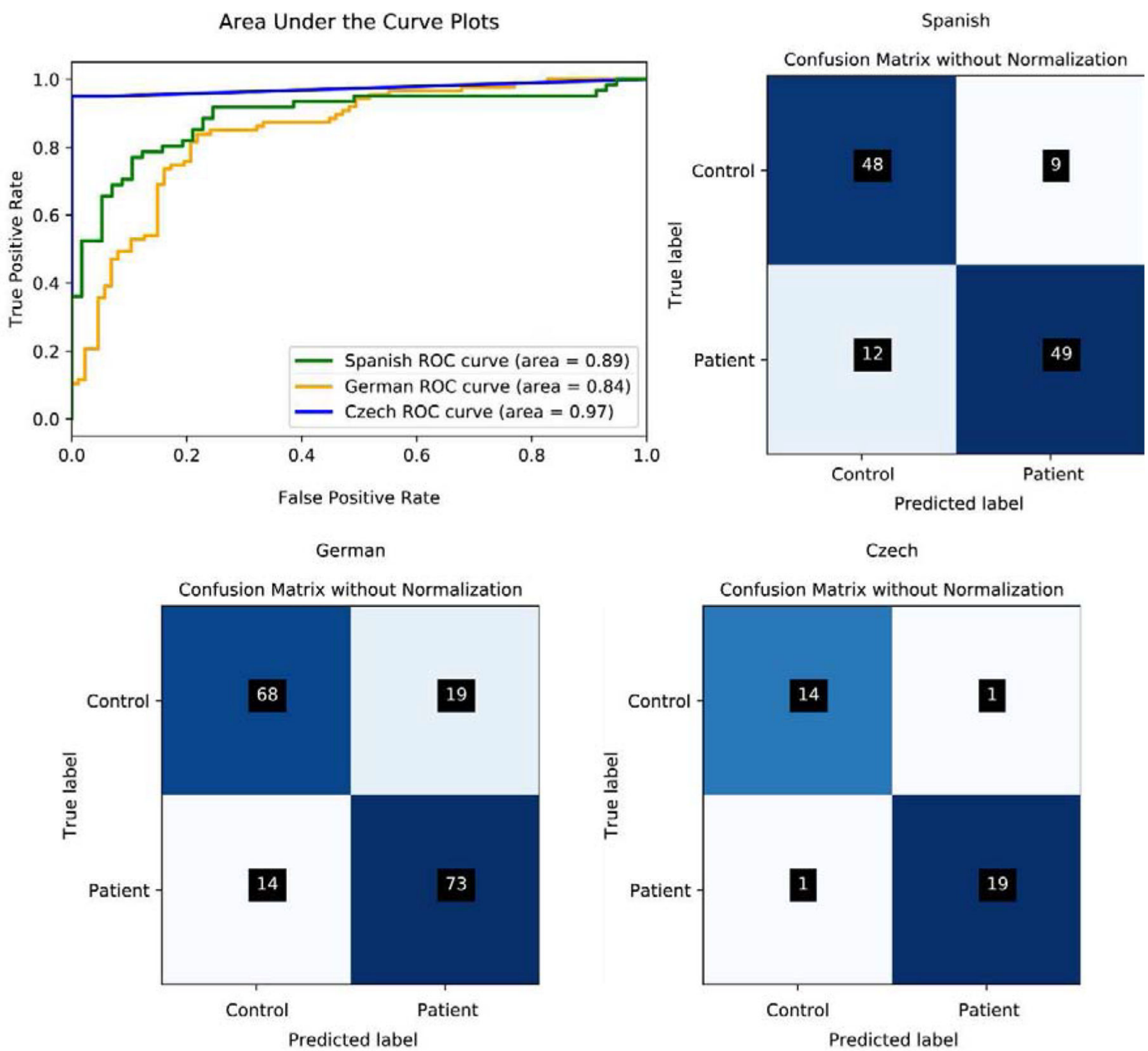

Figure 2.

Results from the non-generalizable analyses. The top left panel shows the AUC scores obtained after RFE procedure (section 2.5.1.2). The remaining three panels show the confusion matrices based on the same analysis settings for each language (top right: Spanish; bottom left: German; bottom right: Czech).

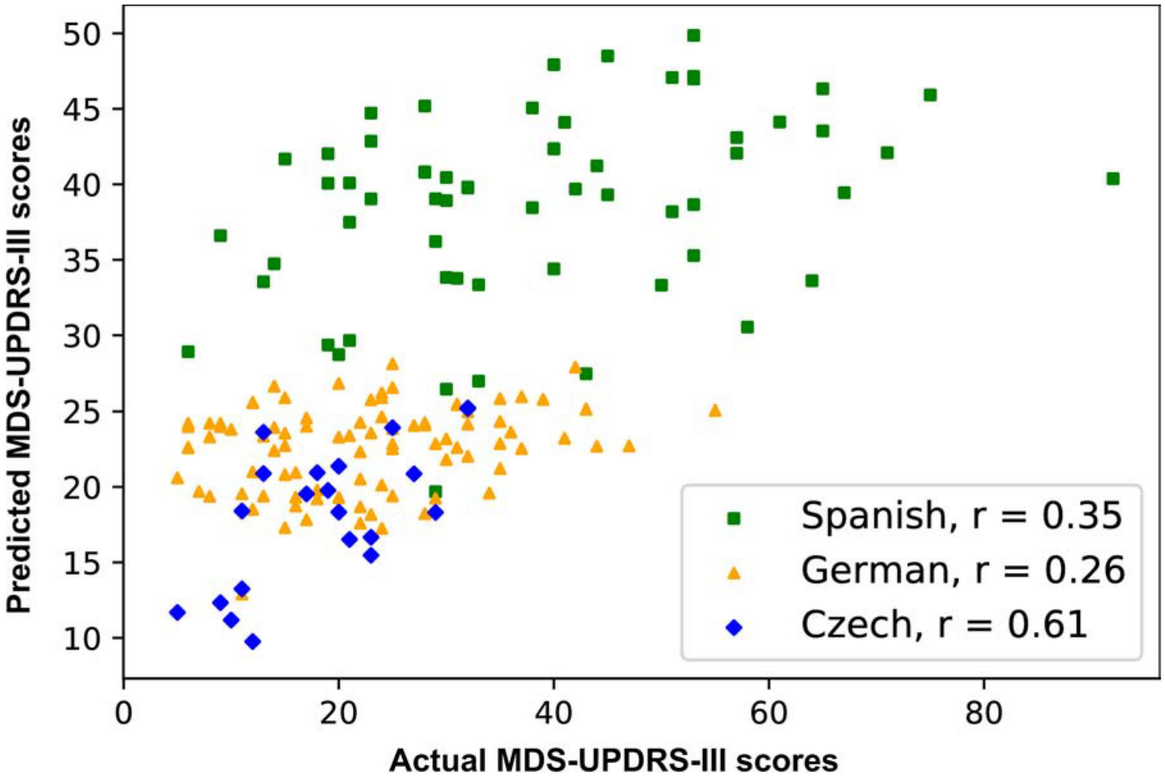

Figure 3.

Correlation between actual motor compromise and predicted motor compromise based on morphological usage patterns. A predicted MDS-UPDRS-III score was computed for each participant by fitting a multiple regression model to the collected MDS-UPDRS-III scores with the minimal set of features selected by the classification analysis. The actual and predicted MDS-UPDRS-III scores for each PD patient are shown for Spanish, German, and Czech. Pearson's *r* correlations between the predicted and the collected MDS-UPDRS-III score were 0.35 for Spanish, 0.26 for German, and 0.61 for Czech.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Demographic and clinical data.

Language	Group	*N*	Gender (F:M)	Years of age	Years of education	MDS-UPDRS-III	H&Y	Years since diagnosis
Spanish	PD patients	61	27:34	62.0 (10.0)	11.30 (4.36)	38.4 (19.4)	2.2 (0.9)	10.9 (9.1)
	Controls	57	28:29	61.7 (9.9)	10.56 (4.64)	-----	-----	-----
German	PD patients	88	41:47	66.5 (9.0)	-----	22.7 (10.9)	2.4 (0.6)	6.6 (5.9)
	Controls	88	44:44	63.2 (14.0)	-----	-----	-----	-----
Czech	PD patients	20	0:20	61.0 (12.0)	-----	17.9 (7.3)	2.2 (0.5)	2.4 (1.7)
	Controls	16	0:16	61.8 (13.2)	-----	-----	-----	-----

MDS-UPDRS-III: Movement Disorder Society-sponsored revision of the Unified Parkinson’s Disease Rating Scale. H&Y: Hoehn & Yahr scale.

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Example of tagging of a sentence in the Spanish corpus.

Word	Part of speech	Attribute-value pairs	Correct tagging probability score
Entonces (*So*)	Adverb	Type: General	0.998
Yo (*I*)	Pronoun	Type: Personal, Person: 1st; Gender: Common; Number: Singular; Case: Nominative	1
Tengo (*have*)	Verb	Type: Main; Mood: Indicative; Tense: Present; Person: 1st; Number: Singular	1
Unos (*some*)	Determiner	Type: Indefinite; Gender: Masculine; Number: Plural	0.96
Negocitos (*small businesses*)	Noun	Type: Common; Gender: Masculine; Number: Plural; Degree: Evaluative	1
Ahí (*there*)	Adverb	Type: General	1
como (*like*)	Conjunction	Type: Subordinating	0.967
De (…)	Adposition	Type: Preposition	1
ganado (*cattle*)	Noun	Type: Common; Gender: Masculine; Number: Singular	0.246
0	Punctuation	Period	1

Table 3.

Classification results for each language.

Feature selection within cross-validation folds					
	Accuracy	**Recall**	**Precision**	**AUC**	**Classifier**
Spanish	71%	70%	73%	73%	LR
German	71%	68%	73%	76%	SGD
Czech	80%	90%	78%	83%	SGD
Results from the non-generalizable analyses					
	Accuracy	**Recall**	**Precision**	**AUC**	**Classifier**
Spanish	82%	80%	84%	89%	SVM
German	81%	84%	79%	84%	SVM
Czech	94%	95%	95%	97%	LR

SGD: stochastic gradient descent; SVM: linear support vector machine; LR: logistic regression.

Table 4.

The most important four features for classification for each language.

Language	Morphological tag description	Feature specification	weight	*p*-value
Spanish	Subordinating conjunction	Rate of tag	0.47	< .001
	Proper noun	Rate of tag	−0.46	0.019
	Present tense	Mean of probability scores	−0.42	.040
	Proper noun	Skewness of probability scores	0.41	.070
German	Verb person not specified	Skewness of probability scores	0.36	< .001
	Determiner in accusative case	Skewness of probability scores	−0.29	.012
	Neuter gender in pronouns	Kurtosis of probability scores	0.25	.002
	Feminine gender in nouns	*SD* of probability scores	−0.22	.010
Czech	Person not specified	Skewness of probability scores	39	< .001
	Use of 2nd most frequent variant	Kurtosis of probability scores	27	.001
	Personal pronoun	Rate of tag	19	< .001
	Masculine gender	Skewness of probability scores	−19	.024

The weight column shows the average weight assigned to each feature by the classifiers, which indicates the importance of the feature in classification decision. The rightmost column shows the *p*-value of the *t*-test for each feature.