

Unsupervised Morphological Segmentation for Detecting Parkinson’s Disease

Elif Eyigoz,¹ Pablo Polosecki,¹ Adolfo M. García,²
Katharina Rogg,³ Juan Orozco-Arroyave,⁴ Sabine Skodda,⁵
Eugenia Hesse,² Agustín Ibáñez,² Guillermo Cecchi¹

¹IBM Research, ²Favaloro University, ³University of Würzburg, ⁴University of Antioquia, ⁵University of Bochum
{ekeyigoz, pipolose, gcecchi}@us.ibm.com, {adolfomartingarcia, euge.hesse}@gmail.com,
rafael.orozco@udea.edu.co, katharina.rogg@googlemail.com, sabine.skodda@kk-bochum.de, aibanez@ineco.org.ar

Abstract

The growth of life expectancy entails a rise in prevalence of aging-related neurodegenerative disorders, such as Parkinson’s disease. In the ongoing quest to find sensitive behavioral markers of this condition, computerized tools prove particularly promising. Here, we propose a novel method utilizing unsupervised morphological segmentation for accessing morphological properties of a speaker’s language. According to our experiments on German, our method can classify patients vs. healthy controls with 81 percent accuracy, and estimate the neurological state of PD patients with Pearson correlation of 0.46 with respect to the unified Parkinson’s disease rating scale. Our work is the first study to show that unsupervised morphological segmentation can be used for automatic detection of a neurological disorder.

1 Introduction

Automatic detection of aging related neurological diseases has gained a lot of interest as the population ages. These methods have a wide range of practical applications, such as aiding the diagnosis and tracking the progression of neurodegenerative diseases, which are essential for effective treatment planning. Parkinson’s disease (PD) is a neurodegenerative disease characterized by slowness of movement, rigidity, tremor, gait, and posture problems. Most Parkinson’s disease patients develop speech impairments affecting phonation, articulation, prosody and intelligibility (Ho et al. 1999). In the linguistic domain, PD involves deficits in semantic, syntactic, and prosodic processing abilities during language production and comprehension (Birba et al. 2017). For example, patients have a particularly salient impairment in processing action verbs (i.e., verbs denoting bodily movements) (Fernandino et al. 2013, Bocanegra et al. 2017), have disproportionate deficits in verb generation compared to noun generation (Péran et al. 2009, Silveri et al. 2012), and experience difficulties in processing syntactically complex sentences (Grossman et al. 1992, Natsopoulos et al. 1993). As PD patients experience a wide range of syntactic difficulties, and syntactic functions are frequently performed in the morphological domain in morphologically rich languages, PD presumably affects the morphology of a patient’s language.

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Since motor impairment directly affects speech production capabilities, most studies on automatic classification of PD through language analysis use acoustic features only (Orozco-Arroyave et al. 2015, Bayestehtashk et al. 2015). There are several studies which have analyzed spontaneous texts in PD using *non-automated* approaches (Illes 1989, Benke et al. 2000, McNamara and Durso 2003, Vanhoutte et al. 2012). The only study assessing spontaneous discourse in PD via automated tools is García et al. (2016). They explore differences in part-of-speech tag frequencies, word repetitions, performed semantic analysis, and obtained 75 percent classification accuracy using POS tags.

We propose a novel method utilizing unsupervised morphological segmentation for accessing morphological properties of a speaker’s language, validated with a classification and regression scheme. According to our experiments on interviews with PD patients and healthy controls in German, our method can classify patients vs. healthy controls with 81 percent accuracy, and estimate the neurological state of PD patients with Pearson correlation of 0.46 with respect to the unified Parkinson’s disease rating scale (UPDRS). UPDRS is the most commonly used scale for assessing the severity of the disease in the clinical study of PD (Ramaker et al. 2002). Our work is the first study to show that unsupervised morphological segmentation can be used for automatic-detection of a neurological disorder. The main advantage of using unsupervised morphological segmentation over other automated language analysis methods is that it can be applied to languages for which automated tools are not available to obtain linguistic analyses, as the only resource required to train an unsupervised morphological segmentation algorithm is large amounts of text. Therefore, the method we present in this paper can be applied to any language with a writing system.

Our work differs from most psycholinguistics and neuroscience studies on linguistic aspects of neurological disorders in multiple ways: First and foremost, our method is not intended for theoretical understanding of human language production and processing capabilities, but for practical purposes. Second, we performed automated analysis of the subjects’ language, as opposed to using human annotated data. Third, we present results that are generalizable to unseen data. Finally, the NLP technique we use, unsupervised morphological segmentation, by design, does not necessar-

ily generate analyses that are in line with the linguistic analyses of morphemes. Instead, it is an empirical data-driven method that estimates probability distributions of substring patterns, and uses them to find the optimal segmentation of the data it is trained on. Unsupervised morphological segmentation is generally used as a preprocessing step for other NLP tasks, where its usefulness is evaluated in terms of an increase in the performance of the overall system. Similarly, we used it merely as a tool for feature extraction, thus the linguistic irrelevance of the method is not a topic of concern for our purposes, unlike it is in theoretical fields.

Despite the aforementioned differences between our work and theoretical studies focused on the same phenomenon, the features that proved to be useful in our experiments are in line with the results of studies on speech disfluencies of PD patients. Therefore, we believe our results can provide insights for psycholinguistics and neuroscience, at least as a methodological framework to quantify speech patterns associated with PD.

The paper is organized as follows: We first present the unsupervised morphological segmentation tool we used in Section 2. We present our feature extraction method in Section 3, the data in Section 4, our feature-selection method in Section 5, and the performance of our method in Section 6. Finally, we provide a discussion on the features that proved to be useful in our experiments in Section 7.

2 Morphological Segmentation

Morphology is a discipline of linguistics which studies the combinatorial rules for combining morphemes, such as stems, prefixes and suffixes, into words. Morphemes are defined in linguistic theory as the smallest meaning-bearing units as well as the smallest elements of syntax (Matthews 1991).

We used Morfessor (Creutz and Lagus 2005), an unsupervised morpheme segmentation and morphology induction tool. Morphological segmentation breaks down words into substrings as ‘*morpheme-like*’ units. For example, the word ‘postprocessing’ can be segmented as ‘post-PREFIX’, ‘process-STEM’, and ‘ing-SUFFIX’. The quote on ‘*morpheme-like*’ is due to the fact that morphological segmentation is performed on orthographic representations. The orthography of German is fairly transparent. Therefore, it is possible to use orthographic strings as *approximations* to phonemes for practical NLP applications. The most common form of representation of language in NLP is orthographic. As a result, having a workable approximation to actual phonemes and morphemes through orthography is of major importance for practical NLP applications.

Morfessor assumes three *categories* of morphemes: prefix, suffix, and stem. Since Morfessor uses a probabilistic model to discover morphemes from unannotated text corpora, the category of a morpheme is inferred solely based on how frequently it occurs at the beginning/end of a word in the corpus. In this regard, we suggest the reader of this paper to consider ‘prefixes’ as substrings that frequently occur at the beginning words, and ‘suffixes’ as substrings that frequently occur at the end of words, as opposed to the linguistic conception of these terms.

Morfessor’s algorithm makes use of the following properties of morphemes to learn morpheme segmentations: the *frequencies* of morphemes, the *lengths* of morphemes in terms of number of characters, the *categories* of morphemes, and the predictability of a morpheme given the previous morpheme. Our method uses the same properties of morphemes for feature extraction. Consequently, we use the distributions of morphological patterns in the corpus Morfessor is trained on as a reference point from where we can assess the morphological properties of a speaker’s words.

3 Feature Extraction

We base our feature extraction method on the very popular notion of *word-frequency* in psycholinguistics, the study of which dates back as far as 1950s (Howes 1957, Rosenzweig 1956, Pollack, Rubenstein, and Decker 1959). Since then, word-frequencies have been used as a variable sensitive to and predictive of performance in fluency, competence, and difficulty of retrieval. As cognitive resources are diminished with aging, and more so in aging related neurodegenerative diseases, word-frequency affect has been observed to manifest itself in various aspects of language production (Tainturier, Tremblay, and Lecours 1989, Hodgson and Ellis 1998, Balota et al. 2002, Rayner et al. 2006).

As a generalization of words frequencies, our feature extraction method uses frequencies of morphological patterns as a proxy for Parkinsonian speech patterns. For this purpose, we estimated probabilities of morphological patterns from a corpus as follows: First, we trained Morfessor to learn morpheme segmentations from a corpus. Then, we used the learned morpheme segmentations to segment the words in the corpus and the interviews. Next, we estimated probability distributions of morphological patterns from the segmented corpus. Finally, we used the probability distributions estimated from the corpus to extract features from the interviews. Since we segmented the words in the corpus by the segmentation model Morfessor learned from it, the distributions we used for feature extraction are outcomes of an optimization over patterns of substrings in the corpus.

3.1 Morphological Features

Morpheme Categories We estimate the probability of a morpheme m given its category $p(m|c(m))$, where $c(m)$ stands for the *category* function: The category function $c(m)$ returns one of the categories in $\mathcal{C} = \{prefix, stem, suffix\}$. Similar to word-frequency, $p(m|c(m))$ means frequency of a morpheme in a corpus. However, unlike word-frequency, it is normalized for each morpheme category separately. For that matter, most of the morphological patterns explained in this section are conditioned on the morpheme categories, in that, we do not combine the frequencies of morphological patterns across categories in \mathcal{C} . We also estimate the probability of a category given the morpheme $p(c(m)|m)$.

We estimate the probability of a morpheme given the previous morpheme $p(m, c(m) | m_{-1}, c(m_{-1}))$, where m_{-1} stands for the morpheme that comes before morpheme m . This forward probability is again conditioned on the categories of m and m_{-1} , i.e., it is estimated for each category pair separately.

Morpheme given category c

$$\{p(m|c(m)) : m \in \mathcal{M} \wedge c(m) = c\} \quad (1)$$

Category c given morpheme

$$\{p(c(m)|m) : m \in \mathcal{M} \wedge c(m) = c\} \quad (2)$$

Morpheme given category c & length

$$\{p(m|c(m), l(m)) : m \in \mathcal{M} \wedge c(m) = c\} \quad (3)$$

Poisson on length per category c

$$\{p_{\text{pois}}(\lambda_c, l(m)) : m \in \mathcal{M} \wedge c(m) = c\} \quad (4)$$

Forward probability for category pair c_0 and c_1

$$\{p(m, c(m) | m_{-1}, c(m_{-1})) : m \in \mathcal{M} \wedge c(m) = c_1 \wedge c(m_{-1}) = c_0\} \quad (5)$$

Table 1: Morphological patterns involving probability distributions. \mathcal{M} stands for morphemes in a sample. $c(m)$ stands for the category function. $l(m)$ stands for the length function.

Morpheme Lengths We estimate the probability of observing a morpheme of a certain length $p(l(m))$, where $l(m)$ stands for the length function. We model $p(l(m)) \sim \text{Pois}(\lambda)$ as a Poisson distribution, where λ is estimated as the mean morpheme length in a corpus. As we estimate λ for each morpheme category separately, we use a subscript as λ_c when it is necessary to distinguish between different categories.

We set an upper bound on the length function to make it robust to noise. The distribution of lengths of morphemes may have a long tail towards long morphemes, due to noise in the corpus. As a result, one unlikely long morpheme in a subject’s interview, again due to noise, undesirably creates a feature that causes overfitting. To address this issue, we redefine the length function as follows:

$$l(m) = \min(\text{length}(m), U)$$

where U is the upper bound, \min means minimum, and $\text{length}(m)$ returns the number of characters in a morpheme. We calculate the upper bound U as follows:

$$U = \min_x F_{\text{pois}}(\lambda, x) > 0.9$$

where F_{pois} is the cumulative distribution function for the Poisson distribution, and λ is the mean stem length in the corpus.

Finally, we estimate the probability of a morpheme m , given its category and its length: $p(m|c(m), l(m))$. We observed that factoring out morpheme frequencies not only in terms of their categories, but also in terms of their lengths provides better learning performance.

Table 1 summarizes the morphological patterns presented so far, where we use set notation to express our collection of probabilities of morphological patterns in a given sample into sets. The sets are in turn used to calculate the following statistics: maximum, minimum, and standard deviation. In Table 1, m stands for a morpheme, and \mathcal{M} stands for all morphemes in an interview. All sets in Table 1 are multisets, i.e., they can have repetition of identical elements.

Patterns 1, 3, and 4 in Table 1 involve only one category c . Therefore, we compute a set using all morphemes m in \mathcal{M} for each morpheme category. Pattern 5, on the other hand,

involves two categories c_0 and c_1 , because it is defined over pairs of morphemes that are adjacent to each other. Therefore, they are computed for all category pairs that were observed adjacent to each other in the data. We consider only morphemes that are adjacent to each other within a word, not across words.

Finally, we use the rate of pairs of morpheme categories that occur adjacent to each other in an interview as a feature, as follows:

$$\frac{|\{(m, m_{-1}) : c(m) = c_1 \wedge c(m_{-1}) = c_0\}|}{|\mathcal{M}|} \quad (6)$$

such that $m \in \mathcal{M}$ and $m_{-1} \in \mathcal{M}$. $|\mathcal{M}|$ is the total number of morphemes in an interview, and (c_0, c_1) is the category pair for which the rate feature is computed. We use both the base and the logarithm of this value as a feature.

3.2 Word-based Features

We compare the performance of our model with a word-based baseline-model. We use the same feature-extraction method that we use for morphemes without utilizing morphological information as follows: Instead of morphological-pattern frequencies, we use word-frequencies, instead of morpheme-lengths, we use probabilities of word-lengths. Again, we compute sets of probability scores from each interview $\{p(w)|w \in \mathcal{W}\}$, where \mathcal{W} stands for the words in an interview, and $p(w)$ stands for word-frequency. Similarly, we use Poisson distribution to model lengths of words to compute sets of probability scores for each interview: $\{p_{\text{pois}}(\lambda, l(w))|w \in \mathcal{W}\}$.

We estimate the λ parameter for the Poisson distribution, and probabilities of words from the same corpus we used for the morphological patterns. As we obtain a score for each word in an interview, we again use the following statistics over the set of scores as features: maximum, minimum, standard deviation.

We use probability of word-length $p_{\text{pois}}(\lambda, l(w))$, and word-frequency $p(w)$ only in the baseline model, but not in the morpheme-based model in our experiments.

Estimating probability of word-length through morphemes If a word has a large number of characters, it may be because it has a large number for short morphemes. Alternatively, it may have a small number of long morphemes. A word-based model cannot make the distinction between these two cases. We model this discrepancy as follows: For each morpheme in a word, we compute their morpheme-length probabilities $p(l(m))$, as described in the previous section, and then multiply them to obtain a score for the entire word. We thus estimate the probability of the length of a word through the lengths of its morphemes. In our morpheme-based model, we use the estimated-word-length-probabilities.

As we obtain a score for each word in an interview, we again use the following statistics over these scores as features: maximum, minimum, and standard deviation.

4 Data

The interview data is a small spoken language corpus in German. 176 German native speakers, 88 patients with PD

(47 men, 41 women), and 88 healthy controls (44 men, 44 women) were recorded at the Knappschaftskrankenhaus of Bochum in Germany (Skodda, Visser, and Schlegel 2011). The subjects were demographically comparable in terms of gender, age, and education. The age of male patients ranged between 44 and 82 years (mean 66.7 ± 8.4), while the age of the female patients ranged from 42 to 84 years (mean 66.2 ± 9.7). The mean value of UPDRS was 22.7 ± 10.9 . Regarding the control group, the age of men ranged from 26 to 83 years (mean 63.8 ± 12.7), and the age of the women was between 54 and 79 years (mean 62.6 ± 15.2). The average time post PD diagnosis was 7.1 ± 5.8 years.

The participants were asked to briefly describe a typical day in their lives. The resulting monologues were hand transcribed. There was only one interview per subject. Patient interviews consist of a total of 4552 words, and 6172 morphemes, whereas healthy control interviews consist of 4552 words and 8267 morphemes. Since demarcation of sentences is not straightforward in spoken language, we do not provide any statistics over the number of sentences in the interviews.

For estimating probability distributions, we used the Open Subtitles corpus (Tiedemann 2009). Subtitles are presumably similar to the interviews, in that subtitles include a large amount of spoken language data. German Subtitles corpus has 187 million words and 210 million morphemes. 89 percent of the morphemes are stems, eight percent are prefixes, and the rest are suffixes. The average number of characters in a stem is 3.39, which is the λ parameter for the Poisson distribution that we used to calculate the upper bound on the length function. In the corpus, the average number of morphemes in a word is 1.13.

5 Feature-selection

Since we automatically generated a large number of features, eliminating features of low quality is essential to the performance of our method. We resorted to an experimental method for eliminating low quality features within a leave-one-out cross-validation setting (LOOCV). We split the data to folds of train-test sets. Within each fold, we performed feature selection as explained in this section, and predicted the label for the test sample using the selected features.

As initial filtering, we used univariate feature selection methods. We obtained a p-value for each feature by computing a t-test for classification, and computing Pearson r for regression. We eliminated features with p-value greater than 0.01. We then performed an ANOVA-F test, and modeled the decreasing scores as a logarithmic decay curve for classification, and modeled the decreasing p-values as a logarithmic decay curve for regression. We found a threshold on the curve using a non-parametric method as follows: for each feature index i on the x axis, we computed percent-change in the curve, and we computed whether the percent-change at position i is greater the percent-change in position $i - 1$. We used the first such position as a threshold to eliminate features with low scores.

We followed univariate selection methods with stability-selection (Meinshausen and Bühlmann 2010). We used the `scikit-learn` implementation of Randomized Logistic

	Acc	Prec	Rec	Med	Classifier
Morph	0.81	0.91	0.69	2	Linear SVM
Baseline	0.6	0.58	0.7	1	RBF SVM
Pearson r		Median		Regressor	
Morph	0.46		2		Lasso
Baseline	0.23		1		Lasso

Table 2: The ‘morph’ rows show the results obtained by using the morphological features. Baseline rows show results obtained by using only the word-based features. ‘Acc’ stands for accuracy, ‘prec’ for precision, ‘rec’ for recall, and ‘med’ for the median number of features across folds. The initialization code along with the grid search parameters of the estimators in the last column can be found in the Appendix.

Regression for classification. Stability selection returns a score for each feature, which we again modeled as a logarithmic decay curve, and eliminated features with low scores as explained above.

Next, we used recursive feature elimination (RFE) using `RFECV` in the `scikit-learn` package. Given an estimator, RFE selects features by recursively considering smaller and smaller sets of features. First, an estimator is trained on the initial set of features and weights are assigned to each of them. Then, features whose absolute weights are the smallest get pruned. This procedure is recursively repeated on the pruned set until the features are exhausted. `RFECV` performs RFE in a cross-validation loop to find an optimal set of features. As `RFECV` requires an estimator to obtain weights for the features, we used Linear SVM for classification, and Linear Regression for regression in `RFECV`.

As `RFECV` returned an optimal set of features, we rerun `RFECV` on the optimal set returned by the previous run, until it no longer returned a smaller set. In other words, we repeated `RFECV` until it converged.

6 Results

Table 2 shows the accuracy, the precision and the recall rates for the classification experiments. Pearson r correlation between the predicted values and the UPDRS scores are shown for the regression experiments. Table 2 also shows the median number of features across folds used by the estimators. Finally, the best performing estimator is given. All results are averages across folds in a LOOCV setting.

Our method performed significantly better than the word-based baseline-model for both regression and classification. For classification, the word-based baseline-model provided ten percent improvement over majority class predictor, which is 0.5, the chance level, as the number of patients and controls were balanced.

Table 3 shows the features that proved to be most useful for classification and regression, sorted in terms of the average absolute weight assigned to them across folds, which are given in the last column. The second column gives the number of the pattern in Table 1. We did not include the word-based baseline-model in Table 3, because all features involving word-frequencies were eliminated dur-

		Reference	Weights
Regression			
Probability of a prefix	standard deviation	(1)	-2.06
Probability of length of a prefix	maximum	(4)	-1.76
Classification			
Rate of prefix-prefix pairs	logarithm	(6)	0.51
Probability of length of a prefix	maximum	(4)	-0.33
Probability of a prefix	maximum	(1)	-0.25
Probability of a stem given a prefix in the previous position	standard deviation	(5)	0.20

Table 3: Features that were selected in at least 20 percent of the folds, and for which the average absolute weight across folds (as shown in the last column) was also among the highest.

ing feature-selection in both the classification and the regression experiments. The only feature the baseline-model used was the minimum word-length probability. Accordingly, the comparable feature in the morpheme-based model, the minimum estimated-word-length-probability, proved to be a useful feature for both classification and regression. Clearly, estimated-word-length-probabilities and probabilities of word-lengths are correlated. However, we observed that the features involving the estimated-word-length-probabilities are more significant than the features involving probabilities of word-lengths both in terms of univariate tests and learning performance.

7 Discussion

Features involving prefixes were disproportionately more useful than features involving stems, and no features involving suffixes proved to be useful in our experiments. Note that a ‘prefix’ in the context of unsupervised segmentation could either be a ‘prefix’ in the linguistic sense, or be a ‘stem’ that has never occurred as a free morpheme in the data Morfessor was trained on. In this regard, features involving prefixes in Table 3 can be interpreted more generally as features involving morphemes that occur at the beginning of words. This finding is in line with the results of studies on speech disfluencies in PD patients.

Stuttering-like speech disfluencies are not uncommon in Parkinson patients (Anderson et al. 1999), which are often attributed to speech motor initiation problems, as acoustic analysis of speech of PD patients indicate difficulty while starting of vocal articulation (Goberman, Blomgren, and Metzger 2010). In addition, the patients in our study were recorded when they were on L-DOPA medication, and it has been shown that PD patients on medication that increase dopamine levels, such as L-DOPA, may demonstrate an increase of speech disfluencies, and in particular stuttering (Wu et al. 1997).

Observation of features related to stems and prefixes in Table 3, but the lack of features related to suffixes in Table 3 can be explained by cognitive-load due to compositionality characteristics of morpheme categories. Prefixes are necessarily followed by other morphemes, whereas stems need not be followed by other morphemes. Suffixes, on the other hand, are very unlikely to be followed by other morphemes. As a result, morphological information attached to prefixes are larger than stems, which are in turn larger than suffixes. We observed that the greater the morphological re-

quirements leading to word completion upon articulation of a given morpheme category (prefix > stem > suffix), the greater the categories robustness to discriminate between patients and controls. This ranking of morpheme categories is in line with the observations in Table 3.

8 Conclusion

We used unsupervised morpheme segmentation for feature extraction to quantify Parkinsonian speech patterns. Our method used corpora to learn a segmentation model, which in turn was used to derive a measure to assess the morphological properties of a speaker’s language. We validated our method by classifying PD vs healthy controls, and predicting UPDRS scores. Our method can capture the alterations PD patients experience during speech production, as we obtained significant performance in classification accuracy, and prediction of UPDRS.

German is a fusional language, i.e., boundaries between morphemes are hard to determine. As a result, the discrepancy between the *morpheme-like* orthographic strings learned by unsupervised morphological segmentation and the actual morphemes is significant. Despite this discrepancy, our method was robust enough to perform well, even on fusional languages.

Our method has wide practical applicability in the medical domain, including speech assessment of PD, and other neurodegenerative diseases such as fronto-temporal dementia, Alzheimer’s and Huntington’s Disease. In addition to the practical value of our method, we believe that our results can also provide theoretical insights for psycholinguistics and neuroscience.

Appendix

Initialization parameters of the estimators from sklearn and lightning packages:

```
Linear SVM: LinearSVC(penalty="l2",
                      class_weight="balanced") GridCV: {"C":
[0.1, 0.5, 1, 2], loss:["hinge", "
squared_hinge"]}
RBF SVM: SVC("kernel": ["rbf"], probability=
True) GridCV:{"gamma": numpy.logspace
(-2, 0, 6).tolist() + numpy.logspace
(0,1,5)[1:].tolist(), "C": numpy.
logspace(-2, 2, 5).tolist()}
Lasso: CDRegressor(max_iter=200, tol=1e-3,
loss='squared', penalty='l1', debiasing=
```

```
True) GridCV:{"alpha": np.logspace(-2,
2, 5)}
```

References

- Anderson, J. M.; Hughes, J. D.; Rothi, L. J.; Crucian, G. P.; and Heilman, K. M. 1999. Developmental stuttering and Parkinson's disease: the effects of levodopa treatment. *Journal of Neurology, Neurosurgery & Psychiatry* 66(6):776–778.
- Balota, D. A.; Burgess, G. C.; Cortese, M. J.; and Adams, D. R. 2002. The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language* 46(1):199–226.
- Bayestehtashk, A.; Asgari, M.; Shafran, I.; and McNames, J. 2015. Fully automated assessment of the severity of Parkinson's disease from speech. *Computer Speech & Language* 29(1):172 – 185.
- Benke, T.; Hohenstein, C.; Poewe, W.; and Butterworth, B. 2000. Repetitive speech phenomena in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry* 69(3):319–324.
- Birba, A.; García-Cordero, I.; Kozono, G.; Legaz, A.; Ibáñez, A.; Sedeño, L.; and García, A. M. 2017. Losing ground: Frontostriatal atrophy disrupts language embodiment in Parkinson's and Huntington's disease. *Neuroscience & Biobehavioral Reviews*.
- Bocanegra, Y.; García, A. M.; Lopera, F.; Pineda, D.; Baena, A.; Ospina, P.; Alzate, D.; Buriticá, O.; Moreno, L.; Ibáñez, A.; and Cuertos, F. 2017. Unspeakable motion: Selective action-verb impairments in Parkinson's disease patients without mild cognitive impairment. *Brain and Language* 168:37 – 46.
- Creutz, M., and Lagus, K. 2005. Inducing the morphological lexicon of a natural language from unannotated text. In *Proceedings of the International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning (AKRR'05)*, 106–113. Helsinki University of Technology.
- Fernandino, L.; Conant, L. L.; Binder, J. R.; Blindauer, K.; Hiner, B.; Spangler, K.; and Desai, R. H. 2013. Parkinson's disease disrupts both automatic and controlled processing of action verbs. *Brain and Language* 127(1):65 – 74.
- García, A. M.; Carrillo, F.; Orozco-Arroyave, J. R.; Trujillo, N.; Bonilla, J. F. V.; Fittipaldi, S.; Adolphi, F.; Nöth, E.; Sigman, M.; Slezak, D. F.; Ibáñez, A.; and Cecchi, G. A. 2016. How language flows when movements don't: An automated analysis of spontaneous discourse in Parkinson's disease. *Brain and Language* 162:19 – 28.
- Goberman, A. M.; Blomgren, M.; and Metzger, E. 2010. Characteristics of speech disfluency in parkinson disease. *Journal of Neurolinguistics* 23(5):470–478.
- Grossman, M.; Carvell, S.; Stern, M. B.; Gollomp, S.; and Hurtig, H. I. 1992. Sentence comprehension in Parkinson's disease: The role of attention and memory. *Brain and Language* 42(4):347 – 384.
- Ho, A. K.; Ianseck, R.; Marigliani, C.; Bradshaw, J. L.; and Gates, S. 1999. Speech impairment in a large sample of patients with Parkinson's disease. *Behavioural Neurology* 11(3):131–137.
- Hodgson, C., and Ellis, A. W. 1998. Last in, first to go: Age of acquisition and naming in the elderly. *Brain and Language* 64(1):146–163.
- Howes, D. 1957. On the relation between the intelligibility and frequency of occurrence of English words. *The Journal of the Acoustical Society of America* 29(2):296–305.
- Illes, J. 1989. Neurolinguistic features of spontaneous language production dissociate three forms of neurodegenerative disease: Alzheimer's, Huntington's, and Parkinson's. *Brain and Language* 37(4):628 – 642.
- Matthews, P. H. 1991. *Morphology*. Cambridge England New York: Cambridge University Press.
- McNamara, P., and Durso, R. 2003. Pragmatic communication skills in patients with Parkinson's disease. *Brain and Language* 84(3):414 – 423.
- Meinshausen, N., and Bühlmann, P. 2010. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72(4):417–473.
- Natsopoulos, D.; Grouios, G.; Bostantzopoulou, S.; Mentenopoulos, G.; Katsarou, Z.; and Logothetis, J. 1993. Algorithmic and heuristic strategies in comprehension of complement clauses by patients with Parkinson's disease. *Neuropsychologia* 31(9):951 – 964.
- Orozco-Arroyave, J. R.; García, N.; Vargas-Bonilla, J. F.; and Nöth, E. 2015. Automatic detection of Parkinson's disease from compressed speech recordings. In *International Conference on Text, Speech, and Dialogue*, 88–95. Springer.
- Péran, P.; Cardebat, D.; Cherubini, A.; Piras, F.; Luccichenti, G.; Peppe, A.; Caltagirone, C.; Rascol, O.; Démonet, J.-F.; and Sabatini, U. 2009. Object naming and action-verb generation in Parkinson's disease: A fMRI study. *Cortex* 45(8):960 – 971.
- Pollack, I.; Rubenstein, H.; and Decker, L. 1959. Intelligibility of known and unknown message sets. *The Journal of the Acoustical Society of America* 31(3):273–279.
- Ramaker, C.; Marinus, J.; Stiggeibout, A. M.; and van Hilten, B. J. 2002. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. *Movement Disorders* 17(5):867–876.
- Rayner, K.; Reichle, E. D.; Stroud, M. J.; Williams, C. C.; and Polatsek, A. 2006. The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging* 21(3):448.
- Rosenzweig, M. R. 1956. Intelligibility as a function of frequency of usage. *The Journal of the Acoustical Society of America* 28(4):759–759.
- Silveri, M. C.; Ciccarelli, N.; Baldonero, E.; Piano, C.; Zinno, M.; Soleti, F.; Bentivoglio, A. R.; Albanese, A.; and Daniele, A. 2012. Effects of stimulation of the subthalamic nucleus on naming and reading nouns and verbs in Parkinson's disease. *Neuropsychologia* 50(8):1980 – 1989.
- Skodda, S.; Visser, W.; and Schlegel, U. 2011. Vowel articulation in Parkinson's disease. *Journal of Voice* 25(4):467–472.
- Tainturier, M.-J.; Tremblay, M.; and Lecours, A. 1989. Aging and the word frequency effect: A lexical decision investigation. *Neuropsychologia* 27(9):1197–1202.
- Tiedemann, J. 2009. News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces. In Nicolov, N.; Bontcheva, K.; Angelova, G.; and Mitkov, R., eds., *Relevant Advances in Natural Language Processing (vol V)*. Amsterdam/Philadelphia: John Benjamins. 237–248.
- Vanhoutte, S.; Letter, M. D.; Corthals, P.; Borsel, J. V.; and Santens, P. 2012. Quantitative analysis of language production in Parkinson's disease using a cued sentence generation task. *Clinical Linguistics & Phonetics* 26(10):863–881.
- Wu, J. C.; Maguire, G.; Riley, G.; Lee, A.; Keator, D.; Tang, C.; Fallon, J.; and Najafi, A. 1997. Increased dopamine activity associated with stuttering. *Neuroreport* 8(3):767–770.