Zero-Shot Cross-lingual Aphasia Detection using Automatic Speech Recognition

Gerasimos Chatzoudis[1], *Manos Plitsis*[1,3], *Spyridoula Stamouli*[1], *Athanasia–Lida Dimou*[1], *Athanasios Katsamanis*[1,2], *Vassilis Katsouros*[1]

[1]Institute for Language and Speech Processing, Athena Research Center, Athens, Greece
[2] Behavioral Signal Technologies, Los Angeles, CA, USA
[3] National and Kapodistrian University of Athens, Athens, Greece

{gerasimos.chatzoudis, manos.plitsis, pstam, ndimou, nkatsam, vsk}@athenarc.gr

Abstract

Aphasia is a common speech and language disorder, typically caused by a brain injury or a stroke, that affects millions of people worldwide. Detecting and assessing Aphasia in patients is a difficult, time-consuming process, and numerous attempts to automate it have been made, the most successful using machine learning models trained on aphasic speech data. Like in many medical applications, aphasic speech data is scarce and the problem is exacerbated in so-called "low resource" languages, which are, for this task, most languages excluding English. We attempt to leverage available data in English and achieve zero-shot aphasia detection in low-resource languages such as Greek and French, by using language-agnostic linguistic features. Current cross-lingual aphasia detection approaches rely on manually extracted transcripts. We propose an end-to-end pipeline using pre-trained Automatic Speech Recognition (ASR) models that share cross-lingual speech representations and are fine-tuned for our desired low-resource languages. To further boost our ASR model's performance, we also combine it with a language model. We show that our ASR-based end-to-end pipeline offers comparable results to previous setups using human-annotated transcripts.

Index Terms: Pathological speech assessment, aphasia, Greek, AphasiaBank, Disordered speech recognition, machine learning, zero-shot classification

1. Introduction

Aphasia is a common speech and language disorder resulting from a brain injury that is usually caused by a stroke. According to the National Aphasia Association, at least 2,000,000 people in the USA and 250,000 in the UK are currently being affected by aphasia and this number is increasing at a rate of 180,000 people who acquire aphasia each year [1]. People with Aphasia (PWA) face communication difficulties making social interaction inefficient due to their impaired speech production and/or language understanding. To mitigate these communication inefficiencies, PWA need to attend long-term, intensive targeted therapies with expertly trained Speech-Language Pathologists (SLPs) [2]. Such conventional in-clinic therapies are time-consuming, expensive, and, in some circumstances, unattainable for a lot of people due to the lack of local and long-term options.

Towards this end, several attempts have been made to detect aphasia from spontaneous speech [3], and develop end-to-end automatic assessment systems [4, 5]. These studies indicate that it is possible to automatically detect and evaluate aphasia using acoustic and linguistic features from either manual or ASR transcriptions. However, these works only focus on languages with sufficient data, such as English. Data scarcity of aphasic speech in low-resource languages makes it difficult to democratize these Machine Learning approaches, and enable aphasia assessment for everyone. One solution to the problem of data scarcity is given by cross-lingual methods that leverage language-agnostic features from resource-rich languages, and transfer them to low-resource settings. Previous attempts have been made to achieve few-shot and zero-shot aphasia detection from English to both closely related (e.g., French) and more distant languages (e.g., Mandarin) [6, 7]. However, these studies rely heavily on manual, human-annotated transcriptions by professional SLPs, a procedure that is costly and tedious, even for a small amount of data.

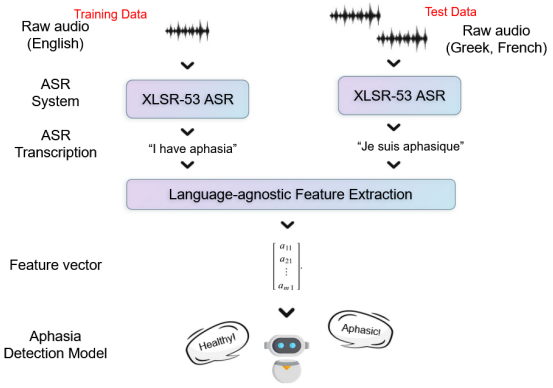

Figure 1: *Schematic diagram of the proposed end-to-end pipeline using language-agnostic linguistic features. Our classifier is trained using English data (left) and evaluated in our low-resource languages, i.e. French and Greek (right).*

For that purpose, we investigate extending current cross-lingual aphasia detection pipelines by incorporating Automatic Speech Recognition models for two closely related to English low-resource languages, i.e., French and Greek. Since training an ASR domain-specific model for these languages requires a considerable amount of collected speech data that is currently unavailable, we use large pre-trained ASR models that have demonstrated significant capabilities across several domains [8]. Specifically, we leverage pre-trained XLSR-53 architectures [9] based on the Wav2Vec2.0 audio representation [10], that have been proven to work for other low-resource languages on AphasiaBank, i.e., Spanish [11]. To augment ASR performance, we incorporate language models trained on out-

of-domain data for each language in our architecture.

Our main contribution is an end-to-end cross-lingual aphasia detection system for Greek and French, using language-agnostic linguistic features directly extracted from pre-trained ASR systems (as shown in Figure 1)[1]. This research activity is part of the ongoing project "PLan-V: A Speech and Language Therapy Platform with Virtual Agent", which aims to develop a technologically-assisted speech and language intervention platform for people with chronic neurogenic communication disorders, such as aphasia, integrating a system for the automated assessment of aphasia severity[2].

2. Related Work

The first attempts at aphasic speech recognition involved short, prompted speech segments [12]. One of the first papers to combine aphasic speech recognition with aphasia severity assessment was [4]. In this work, the researchers use human-annotated and ASR-based transcriptions from the English AphasiaBank, to extract certain linguistic features and develop a model that ultimately predicts a speaker's AQ score. A similar ASR-based approach, using features extracted with deep neural networks, was adopted by [5] for the Cantonese AphasiaBank. Both teams focused more on the robustness of the features used for the predicting model in their respective language, rather than on speech recognition performance. Both teams have also attempted to use out-of-domain speaker data to improve ASR performance in various ways [13, 14]. Recently, FacebookAI researchers have improved ASR performance in the English AphasiaBank by using neural models trained with vast amounts of data, and then adapted to various domains [8]. Semi-supervised learning was also used recently in an attempt to improve ASR performance in both English and Spanish [11].

Several attempts have also been made to classify different aphasia types from manual transcripts. Researchers in [15] and [16] use similar features extracted from the AphasiaBank manual transcriptions. Interestingly, in [16] the two features that were found most significant were the number of words and number of closed terms in the Cinderella story, which are respective measures of productivity and grammaticality, qualities that we also find important when assessing aphasic status, while in [15] we see similar patterns in the most important features.

Our work attempts to solve the problem of training an aphasia detection model from speech, for a low-resource language, for which there is a very limited amount of data available (or even no data at all). We are currently aware of two works that have attempted to solve this problem, both, however, using linguistic features extracted only from human-annotated textual transcriptions of available speech data [6, 7]. Both use transfer learning techniques to apply zero-shot and few-shot aphasia classification. Balagopalan et al. in [6] claim that it may be possible to transfer features from languages that are not distant and this is also validated in [7], where a wider set of linguistic features is also used. Last, in [17], the authors used cross-lingual linguistic features to assess morphosyntactic production in agrammatic aphasia.

3. Datasets

Our dataset consists of speech samples from the AphasiaBank database for English and French, as well as a Greek corpus collected by various researchers. The amount of data available for each language can be seen in Table 1.

AphasiaBank is a shared database of multimedia interactions with the purpose of studying aphasia communication [18, 19]. It consists of a collection of transcribed audio and/or video recordings of interviews between clinicians and aphasic or healthy subjects. Depending on the discourse language and the contributing research group, various speech elicitation protocols were followed, with the majority of protocols being variants of the AphasiaBank data collection protocol. In particular, we used speech samples from the English and French AphasiaBank subgroup, and specifically, only data that were collected according to the AphasiaBank protocol. AphasiaBank's transcriptions are produced using the CHAT transcription format [20] that includes, besides textual transcriptions, a variety of symbols denoting non-speech sounds such as laughs and coughs, as well as a system for transcribing paraphasias, or mispronounced words, in the International Phonetic Alphabet.

Our Greek data come from two main sources: a) the GREECAD corpus [21], a dataset of spoken narratives collected from 22 Greek-speaking PWA and 10 unimpaired controls using a protocol of 4 narrative tasks and b) the PLan-V project corpus [22], a dataset of spoken narratives collected from 10 unimpaired controls using an augmented narrative discourse elicitation protocol, which includes 4 tasks from the AphasiaBank protocol and 3 tasks from the GREECAD protocol. In both datasets, spoken language samples have been manually transcribed in an orthographic format and segmented into time-aligned utterances, following the CHAT guidelines for utterance segmentation. We chose not to incorporate the Greek datasets of the AphasiaBank repository, mainly because of their conversational form, in contrast to the GREECAD and PLan-V data, which represent different types of monologic narrative discourse, with minimal interventions from the investigator.

Table 1: *Dataset Statistics*

Language	Number of Speakers		
	Total	**Control**	**Aphasia**
English	705	245	460
French	27	14	13
Greek	42	20	22

4. Methods

4.1. Preprocessing

All available transcripts in each language contain various annotations, such as target phrases and pauses. These annotations, although useful, are unavailable in our automated assessment system, so we remove them from all transcripts using PyLangAcq [23]. We only retain the cleaned transcripts, with their corresponding duration information. We refer to these as oracle transcripts. For each oracle transcript, we downsample its corresponding audio to 16 kHz. Following data cleaning, we pass both oracle and ASR transcripts through a part-of-

[1]The source code for reproducing all our experiments can be found at https://gitlab.com/ilsp-spmd-all/public/crosslingual-aphasia-detection

[2]This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (Project title: PLan-V: A Speech and Language Therapy Platform with Virtual Agent, Project code: T2EDK-02159).

speech tagging, tokenization and lemmatization module. Tokens that correspond to out-of-vocabulary words are not omitted from the process. For English and French, we employ spaCy's transformer-based models[3], whereas for Greek, we use the Neural NLP toolkit for Greek [24].

4.2. Automatic Speech Recognition Models

Our suggested ASR model uses the XLSR-53 architecture [9] which is based on the Wav2Vec2.0 audio representation [10] and was trained with 56,000 hours of speech data in 53 languages. Because it is based on a language-agnostic representation of speech, this architecture is particularly well-suited for multi-language tasks, where it can be fine-tuned for each language using very few data instances. Specifically, we used fine-tuned XLSR-53 models for each language, that were pre-trained on the Common Voice dataset [25]. Hugging Face provides open-source versions of these models in English [26], French [27], and Greek [28].

To further optimize the performance of the fine-tuned XLSR-53 models, we combined them with a language model for each language and employed a beam search decoding strategy. Each experiment used a beam width of 10. To train the language model, we first collected textual data from Wikipedia [29] and then trained Kneser-Ney smoothed n-gram language models using the KENLM toolkit [30]. Specifically, we trained a 3-gram model for all languages. The language models are integrated into the XLSR-53 architecture using the pyctcdecode library[4], a fast and feature-rich CTC beam search decoder for speech recognition in Python. We evaluate our models on both Word-Error-Rate (WER) and Character-Error-Rate (CER) metrics [31].

4.3. Language-agnostic Feature Extraction

Previous works in automatic assessment of aphasia indicate that using linguistic features is considered standard practice (see Section 2). Given the condition's influence on the brain's language centers it is necessary for any reliable system to be able to access semantic and grammatical information. Over 500 linguistic indicators have been applied to assess spoken language ability and the efficacy of PWA interventions [22].

Following recent works in the analysis of discourse production [32, 33, 22], we focus on linguistic features that evaluate 6 language abilities, i.e., linguistic productivity, content richness, fluency, syntactic complexity, lexical diversity, and gross output. A complete list of the extracted features is presented in Table 2. These make up a vector of size 24 per speaker.

We chose to extract all features at the speaker level to minimize variability due to utterance length and number of stories between speakers. Our end-to-end approach can help minimizing the impact of this type of variability that is very common across datasets due to the different transcription guidelines that may have been followed in each case.

4.4. Cross-lingual aphasia detection

We initially examine the feasibility of aphasia detection in English prior to developing our transfer learning pipeline across languages. We compare multiple classification models in this monolingual scenario, including SVM [34], XGBoost [35], Decision Tree and Random Forest classifiers using Scikit-learn [36]. We evaluate our models' accuracy using a Leave-one-subject-out (LOSO) cross-validation approach.

We perform zero-shot cross-lingual binary classification of aphasic vs control subjects, from English to French and Greek respectively. We first test the performance of our models using oracle transcriptions in all languages. Because we introduce a generic end-to-end pipeline, we normalize the target languages' input features using Scikit-learn's Standard scaler [36] that was fitted in the source language, i.e., English.

Then, we investigate the integration of ASR transcripts under two settings. In the first setting, we train our classifier leveraging the English oracle transcripts, and we evaluate it using the ASR transcripts from the low-resource, target languages. In the second setting, we replace the English oracle training data with ASR transcripts in order to develop a fully end-to-end cross-lingual aphasia detection system. With the last experiment we want to explore whether such a system can be developed without any manually transcribed data. This would allow us to benefit from a much greater amount of data for training and potentially improve overall detection performance.

Table 2: *Language-agnostic linguistic features across 6 language ability categories.*

Language ability	**Feature**
Linguistic Productivity	Mean Length of Utterance
Content Richness	Verb/Word Ratio
	Noun/Word Ratio
	Adjective/Word Ratio
	Adverb/Word Ratio
	Preposition/Word Ratio
	Propositional density
Fluency	Words per Minute
Syntactic Complexity	Verbs per Utterance
	Noun Verb Ratio
	Open-closed class words
	Conjunction/Word Ratio
	Mean Clauses per utterance
	Mean dependent clauses
	Mean independent clauses
	Dependent to all clauses ratio
	Mean Tree height
	Max Tree depth
	Number of independent clauses
	Number of dependent clauses
Lexical Diversity	Lemma/Token Ratio
	Words in Vocabulary per Words
	Unique words in vocabulary per Words
Gross output	Number of words

5. Results and Discussion

5.1. Automatic Speech Recognition Performance

We present the performance of the ASR models for each language in Table 3, as measured by Word Error Rate (WER). As expected, ASR models perform worse in aphasic speech than in healthy speech. We observe that the English XLSR-53 model outperforms the other two, while the Greek model outperforms the French one in all categories. Adding a language model to the ASR system significantly drops WER in English and French, but exhibits inconsistent results in Greek.

[3]https://spacy.io/

[4]https://github.com/kensho-technologies/pyctcdecode

Table 3: *Pre-trained XLSR-53 ASR models performance in English, French and Greek (with and without a Language Model)*

Language	WER (%)		
	Total	Control	Aphasia
english-no-lm	52.57	41.45	62.74
english-with-lm	**47.14**	**37.46**	**55.98**
french-no-lm	71.97	66.67	85.71
french-with-lm	**66.88**	**62.78**	**77.52**
greek-no-lm	**65.66**	**61.98**	72.24
greek-with-lm	67.34	65.11	**71.33**

5.2. Monolingual Aphasia Detection - English Only

We compare the performance of various classification models trained and evaluated solely on English control and aphasic speakers, and we present the results in Table 4. Three different scenarios are evaluated, namely using oracle transcripts, ASR results, and ASR with LM results for both training and testing. As shown, the SVM classifier outperforms the XGBoost, Decision Tree and Random Forest classifiers in every setting. We conclude that, given our feature set, we can achieve satisfactory results for aphasia detection in English, even when using ASR transcripts.

Table 4: *Comparison of four classifiers in terms of accuracy (%) in the English-only experiment, using oracle and ASR transcriptions (with and without a Language Model).*

Model	Accuracy(%)		
	Oracle	ASR	ASR with LM
SVM	**97.45**	**93.90**	**93.76**
Decision Tree	92.06	87.38	89.22
XGBoost	96.31	93.62	93.48
Random Forest	96.03	92.34	92.20

Table 5: *Comparison of four classifiers in terms of accuracy (%) in the cross-lingual experiment using oracle transcriptions for all languages.*

Experiment	Accuracy(%)			
	SVM	Decision Tree	XGBoost	Random Forest
English → French	48.15	81.48	**100**	**100**
English → Greek	52.38	59.52	**97.62**	85.71

5.3. Cross-Lingual Aphasia Detection using Oracle Transcripts during Training

In Table 5, we show the results for the cross-lingual experiment using oracle transcripts for all languages. Since both the French and Greek datasets are balanced we only report accuracy. XGBoost classifier outperforms other classifiers in all settings and thus we chose to exclusively use it in the next experiments. We achieve almost perfect results in both Greek and French using oracle transcripts. As shown in Table 6, when using ASR transcripts for the test set, our classifier's accuracy drops by almost 19 % and 14 % in French and Greek respectively. The addition of the language model, while increasing accuracy in French, actually decreases performance in Greek.

Table 6: *Zero-shot cross-lingual aphasia detection results from English to French and Greek. The classifier is trained on oracle transcripts and tested on ASR output.*

Experiment	Transcriptions	Accuracy (%)
English → French	Oracle → Oracle	100
	Oracle → ASR	77.78
	Oracle → ASR with LM	81.48
English → Greek	Oracle → Oracle	97.62
	Oracle → ASR	83.33
	Oracle → ASR with LM	66.67

5.4. Cross-Lingual Aphasia Detection using ASR Transcripts during Training

In Table 7, we demonstrate that training our classifier on English ASR-derived transcriptions produces results comparable to those obtained using oracle transcriptions (Section 5.3), supporting our hypothesis that an end-to-end pipeline is achievable in the cross-lingual setting. Specifically, we achieve up to 88.88 % and 76.19 % accuracy in French and Greek respectively, depending on whether we use a language model for these languages. The addition of the language model for French and Greek worsened accuracy results in almost every case. Based on our results so far, we are unable to reach a resolution regarding the integration of a language model. This could be because the language models were trained on a generic data corpus (Wikipedia) rather than a dataset-specific domain corpus.

Table 7: *Zero-shot cross-lingual aphasia detection results from English to French and Greek. For each language pair, we evaluate the performance of our classifier in terms of accuracy using ASR-derived transcriptions for training too.*

Experiment	Transcriptions	Accuracy (%)
English → French	ASR→ Oracle	81.48
	ASR → ASR	77.78
	ASR → ASR with LM	62.96
	ASR with LM→ Oracle	88.88
	ASR with LM→ ASR	**88.88**
	ASR with LM→ ASR with LM	77.77
English → Greek	ASR→ Oracle	83.33
	ASR → ASR	73.81
	ASR → ASR with LM	66.67
	ASR with LM→ Oracle	92.86
	ASR with LM→ ASR	**76.19**
	ASR with LM→ ASR with LM	76.19

6. Conclusion and Future Work

In this work, we established an end-to-end pipeline for cross-lingual aphasia detection by using language-agnostic linguistic features from automatic speech recognition transcripts. We achieved up to 81 % and 83 % detection accuracy in French and Greek respectively, when training our model using human-annotated speech transcripts, and up to 88 % and 76 % for the end-to-end case when only using ASR output. Our results indicate that it is possible to apply transfer learning between two closely related languages even without oracle transcripts in both source and target languages. In the future, we aim to extend our research by developing a similar procedure for aphasia severity assessment and aphasia type classification, as well as study the effects of adapting the ASR and language models to domain-specific data.

7. References

[1] N. A. Association, "Aphasia." [Online]. Available: https://www.aphasia.org/

[2] M. C. Brady, H. Kelly, J. Godwin, P. Enderby, and P. Campbell, "Speech and language therapy for aphasia following stroke," *Cochrane database of systematic reviews*, no. 6, 2016.

[3] Y. Qin, T. Lee, and A. P. H. Kong, "Automatic Speech Assessment for Aphasic Patients Based on Syllable-Level Embedding and Supra-Segmental Duration Features," in *ICASSP*, Apr. 2018, pp. 5994–5998.

[4] D. Le, K. Licata, and E. Mower Provost, "Automatic quantitative analysis of spontaneous aphasic speech," *Speech Communication*, vol. 100, pp. 1–12, Jun. 2018.

[5] Y. Qin, T. Lee, and A. P. H. Kong, "Automatic Assessment of Speech Impairment in Cantonese-Speaking People with Aphasia," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 331–345, Feb. 2020.

[6] A. Balagopalan, J. Novikova, M. B. A. Mcdermott, B. Nestor, T. Naumann, and M. Ghassemi, "Cross-Language Aphasia Detection using Optimal Transport Domain Adaptation," in *Proceedings of the Machine Learning for Health NeurIPS Workshop*, vol. 116. PMLR, 2020, pp. 202–219.

[7] A. Shivkumar, J. Weston, R. Lenain, and E. Fristed, "Blabla: Linguistic feature extraction for clinical analysis in multiple languages," *Proc. Interspeech 2020*, pp. 2542–2546, 2020.

[8] A. Xiao, W. Zheng, G. Keren, D. Le, F. Zhang, C. Fuegen, O. Kalinli, Y. Saraf, and A. Mohamed, "Scaling ASR Improves Zero and Few Shot Learning," *arXiv:2111.05948 [cs, eess]*, Nov. 2021.

[9] A. Conneau, A. Baevski, R. Collobert, A. Mohamed, and M. Auli, "Unsupervised cross-lingual representation learning for speech recognition," *CoRR*, vol. abs/2006.13979, 2020.

[10] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 12 449–12 460.

[11] I. G. Torre, M. Romero, and A. Álvarez, "Improving Aphasic Speech Recognition by Using Novel Semi-Supervised Learning Methods on AphasiaBank for English and Spanish," *Applied Sciences*, vol. 11, no. 19, p. 8872, Jan. 2021.

[12] A. Abad, A. Pompili, A. Costa, I. Trancoso, J. Fonseca, G. Leal, L. Farrajota, and I. P. Martins, "Automatic word naming recognition for an on-line aphasia treatment system," *Computer Speech & Language*, vol. 27, no. 6, pp. 1235–1248, Sep. 2013.

[13] D. Le and E. M. Provost, "Improving automatic recognition of aphasic speech with aphasiabank." in *Interspeech*, 2016, pp. 2681–2685.

[14] Y. Qin, T. Lee, S. Feng, and A. P. H. Kong, "Automatic Speech Assessment for People with Aphasia Using TDNN-BLSTM with Multi-Task Learning," in *Interspeech 2018*. ISCA, Sep. 2018, pp. 3418–3422.

[15] K. C. Fraser, J. A. Meltzer, N. L. Graham, C. Leonard, G. Hirst, S. E. Black, and E. Rochon, "Automated classification of primary progressive aphasia subtypes from narrative speech transcripts," *Cortex*, vol. 55, pp. 43–60, Jun. 2014.

[16] D. Fromm, J. Greenhouse, M. Pudil, Y. Shi, and B. MacWhinney, "Enhancing the classification of aphasia: a statistical analysis using connected speech," *Aphasiology*, vol. 0, no. 0, pp. 1–28, Sep. 2021.

[17] C. Themistocleous, B. Ficek, K. Webster, D.-B. den Ouden, A. E. Hillis, and K. Tsapkini, "Automatic Subtyping of Individuals with Primary Progressive Aphasia," *Journal of Alzheimer's disease : JAD*, vol. 79, no. 3, pp. 1185–1194, 2021, publisher: IOS Press.

[18] B. Macwhinney, D. Fromm, M. Forbes, and A. Holland, "AphasiaBank: Methods for Studying Discourse," *Aphasiology*, vol. 25, pp. 1286–1307, Nov. 2011.

[19] M. M. Forbes, D. Fromm, and B. MacWhinney, "AphasiaBank: A resource for clinicians," *Seminars in speech and language*, vol. 33, no. 3, pp. 217–222, Aug. 2012.

[20] B. MacWhinney, "The childes project: Tools for analyzing talk: Volume i: Transcription format and programs, volume ii: The database," 2000.

[21] S. Varlokosta, S. Stamouli, A. Karasimos, G. Markopoulos, M. Kakavoulia, M. Nerantzini, A. Pantoula, V. Fyndanis, A. Economou, and A. Protopapas, "A greek corpus of aphasic discourse: collection, transcription, and annotation specifications," in *Proceedings of RaPID-2016 Workshop, May 2016*, no. 128, 2016.

[22] S. Stamouli, M. Nerantzini, I. Papakyritsis, A. Katsamanis, G. Chatzoudis, A.-L. Dimou, M. Plitsis, V. Katsouros, S. Varlokosta, and A. Terzi, "A web-based application for eliciting narrative discourse from greek-speaking people with aphasia." *(submitted for publication), Frontiers in Communication, Research Topic: Remote Online Language Assessment: Eliciting Discourse from Children and Adults*, 2022.

[23] J. L. Lee, R. Burkholder, G. B. Flinn, and E. R. Coppess, "Working with chat transcripts in python," Department of Computer Science, University of Chicago, Tech. Rep. TR-2016-02, 2016.

[24] P. Prokopidis and S. Piperidis, "A neural nlp toolkit for greek," in *11th Hellenic Conference on Artificial Intelligence*, ser. SETN 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 125–128.

[25] R. Ardila, M. Branson, K. Davis, M. Henretty, M. Kohler, J. Meyer, R. Morais, L. Saunders, F. M. Tyers, and G. Weber, "Common voice: A massively-multilingual speech corpus," in *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 2020, pp. 4211–4215.

[26] J. Grosman, "Xlsr wav2vec2 english by jonatas grosman," https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-english, 2021.

[27] ——, "Xlsr wav2vec2 french by jonatas grosman," https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-french, 2021.

[28] ——, "Xlsr wav2vec2 greek by jonatas grosman," https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-greek, 2021.

[29] W. Foundation. Wikimedia downloads. [Online]. Available: https://dumps.wikimedia.org

[30] K. Heafield, "Kenlm: Faster and smaller language model queries," in *Proceedings of the Sixth Workshop on Statistical Machine Translation*, ser. WMT '11. USA: Association for Computational Linguistics, 2011, p. 187–197.

[31] A. Morris, V. Maier, and P. Green, "From wer and ril to mer and wil: improved evaluation measures for connected speech recognition." 01 2004.

[32] B. C. Stark, "A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome," *American Journal of Speech-Language Pathology*, vol. 28, no. 3, pp. 1067–1083, 2019.

[33] D. Fromm, J. Greenhouse, M. Pudil, Y. Shi, and B. MacWhinney, "Enhancing the classification of aphasia: a statistical analysis using connected speech," *Aphasiology*, pp. 1–28, 2021.

[34] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[35] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 785–794.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.