

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/343425228>

Word Error Analysis in Aphasia: Introducing the Greek Aphasia Error Corpus (GRAEC)

Article in *Frontiers in Psychology* · August 2020

DOI: 10.3389/fpsyg.2020.01577

CITATIONS

0

READS

8

6 authors, including:



Dimitrios Kasselimis

National and Kapodistrian University of Athens

75 PUBLICATIONS 331 CITATIONS

[SEE PROFILE](#)



Maria Varkanitsa

Massachusetts General Hospital

25 PUBLICATIONS 48 CITATIONS

[SEE PROFILE](#)



Georgia Angelopoulou

National and Kapodistrian University of Athens

17 PUBLICATIONS 16 CITATIONS

[SEE PROFILE](#)



Ioannis Evdokimidis

Eginition Hospital Athens

187 PUBLICATIONS 3,362 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Headaches and Socioeconomical distress [View project](#)



The Disintegration of Spatio-temporal and Semantic Memory Binding in Mild Cognitive Impairment [View project](#)



Word Error Analysis in Aphasia: Introducing the Greek Aphasia Error Corpus (GRAEC)

Dimitrios Kasselimis^{1*†}, Maria Varkanitsa^{2†}, Georgia Angelopoulou¹, Ioannis Evdokimidis¹, Dionysis Goutsos³ and Constantin Potagas¹

¹ Neuropsychology and Language Disorders Unit, 1st Department of Neurology, Eginition Hospital, National and Kapodistrian University of Athens, Athens, Greece, ² Sargent College of Health and Rehabilitation Sciences, Boston University, Boston, MA, United States, ³ Department of Linguistics, National and Kapodistrian University of Athens, Athens, Greece

Keywords: Greek, corpora, aphasia, errors, discourse, narration

OPEN ACCESS

Edited by:

Carlo Semenza,
University of Padova, Italy

Reviewed by:

Silvia Martínez Ferreiro,
Université de Toulouse, France
Mira Goral,
The City University of New York,
United States

*Correspondence:

Dimitrios Kasselimis
dkasselimis@gmail.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 06 February 2020

Accepted: 12 June 2020

Published: 04 August 2020

Citation:

Kasselimis D, Varkanitsa M,
Angelopoulou G, Evdokimidis I,
Goutsos D and Potagas C (2020)
Word Error Analysis in Aphasia:
Introducing the Greek Aphasia Error
Corpus (GRAEC).
Front. Psychol. 11:1577.
doi: 10.3389/fpsyg.2020.01577

INTRODUCTION

Since the pioneering work of Paul Broca and Carl Wernicke, it has become clear that the interaction of aphasia research and theoretical linguistics can be beneficial for both disciplines: (1) in order to understand the nature of aphasia as a language disorder, it is crucial to understand the nature of language; its internal rules and principles, (2) linguistic analysis of aphasic speech can also provide some evidence on the relation between brain and language, (3) neurolinguistic data can be used to distinguish between competing linguistic theories, and (4) linguistic analysis of aphasic speech often leads to the design of linguistic-specific treatment programs for aphasia (for more details, see Avrutin, 2001).

One of the most exciting recent developments in linguistics has been the widespread use of electronic corpora, both as a methodology and a theoretical viewpoint on language (see e.g., McEnery and Hardie, 2012, for an overview). In parallel, in aphasia research, large-scale data collection and group studies allow generalizations about the population from which the participants have been drawn, leading to useful findings (see Grodzinsky et al., 1999) that can complement single case studies, which allow for a detailed description of aphasic speech patterns and inferences about the language system in non-brain damaged individuals (see amongst others Badecker and Caramazza, 1985; Caramazza, 1986; Caramazza and Badecker, 1991). However, recruiting patients with aphasia on a large scale is difficult. Even when permission for collecting and using data by patients with aphasia has been obtained, considerable resources are required to move patients through the steps of consenting, screening and testing. A solution to this problem could be data sharing, as is increasingly realized in recent bibliography, which has evidenced a surge in corpora of language datasets from speakers with various disorders, including aphasia, in several languages such as Dutch (Westerhout and Monachesi, 2007), Cantonese (Kong and Law, 2019), Russian (Khudyakova et al., 2016), Croatian (Kuvač Kraljević et al., 2017), and, of course, English (Mirman et al., 2010; Williams et al., 2010; MacWhinney et al., 2011; Laures-Gore et al., 2016). Despite such attempts of developing corpora widely available to researchers, the need for additional open data banks from different languages still remains. For instance, for Greek a recent study has presented a detailed methodology for the transcription and annotation of aphasic speech samples (Varlokosta et al., 2016); although the authors describe an elaborate pipeline, no data has been available yet.

Apart from the importance of data sharing discussed above, there is a methodological issue related to aphasic discourse analysis that is worth mentioning, namely, the method of eliciting a speech sample, which will be then used to evaluate a patient's linguistic competence on the basis of several indices, such as type and frequency of errors, semantic content, speech rate, mean length of utterance, etc. Given the large number of genres used in studies assessing aphasic narration

ability (for an overview, see Müller et al., 2008), one must acknowledge the possible effects of the chosen elicitation task on the qualitative and quantitative characteristics of speech output (Armstrong, 2000), and, subsequently, the importance of evaluating verbal production across such genres (Armstrong et al., 2011).

Moreover, there has been a well-established tradition of comparing data from speakers with aphasia with general corpus data, used as controls for a variety of purposes (e.g., Schwartz et al., 1994; Gahl, 2002; Fraser et al., 2015). As reference corpora become widely available for many languages, including Greek (Goutsos, 2010), there is an increasing need for developing resources with specialized data from speakers with disorders.

To that end, we have developed the Greek Aphasia Error Corpus (GREAC), which is a large, searchable, web-based corpus of patients' performance on two different elicitation tasks, i.e., picture description and free narration, also including background language testing, and clinical/demographic information. The corpus is available at <http://aphasia.phil.uoa.gr/>, while a pilot sample of the data has been included in AphasiaBank (<http://talkbank.org/AphasiaBank/>).

COMPILING THE GREAC

To our knowledge, this is the first publicly available corpus with data from Greek patients with aphasia. We present the first data from 50 right-handed monolingual Greek patients, with left stroke-induced aphasia, assessed at the Neuropsychology and Language Disorders Unit of the 1st Neurology Department of the National and Kapodistrian University of Athens, at Eginition Hospital. The participants (16 women) were 30–86 years old, with 4–20 years of formal schooling.

Background language testing included the Boston Diagnostic Aphasia Examination–Short Form (BDAE-SF) adapted for Greek (Goodglass and Kaplan, 1983; Tsapkini et al., 2009), and the Boston Naming Test (Kaplan et al., 1983), standardized in Greek (Simos et al., 2011), CT and/or MRI scans were obtained for each patient, and two independent neuroradiologists identified lesion sites, which were then coded according to previously reported methodology (Kasselimis et al., 2017). These reports are part of the publicly available database. At this point, the structural MRIs of the patients are not included in GREAC. Demographic and speech sample information are shown in **Table 1**. Informed consent for participation in the study and publication of the data (ensuring anonymity) was obtained from all participants according to the Ethics Committee of Eginition Hospital. No individually identifying information—apart from time post onset, brain lesion loci, tests' performance, and basic demographic information, including sex, age, and years of formal schooling—about the patients is contained in the corpus, and individual patients are listed by random codes (see in **Supplementary Tables 1, 2**, for individual information regarding lesions and BDAE scores, respectively).

At present, GREAC includes 17,507 words (counting only those produced by patients) with 2,397 annotated errors. GREAC is an on-going project, aiming at a corpus of approximately

50,000 words produced by 120 patients in the following 5 years. The data included in GREAC are derived from a thorough neuropsychological assessment, during which patients were first asked to talk about their illness in the form of a semi-prompted monolog (stroke story interview) and then describe the Cookie Theft picture (Picture Description task) from the BDAE-SF (Goodglass and Kaplan, 1983). All assessments were performed by a psychologist/clinical neuropsychologist in a quiet room at the Neuropsychology and Language Disorders Unit of Eginition Hospital. The examiner first initiated a short discussion with the patient, then proceeded to medical history taking, and explained in short the process of the neuropsychological assessment. During this initial interaction, the examiner made all possible efforts to establish Rapport, and make the patient feel comfortable. After that, the speech samples were obtained. First, for the stroke story, the examiner asked the patient to describe the story of their illness: "Please tell me what happened to you when you had the stroke." Then, the patient was asked to describe the Cookie Theft picture: "Please look carefully and describe whatever you see happening in this image." The first task was chosen in order to elicit more natural speech data, while picture elicitation was employed to ensure more controlled discourse samples, since participants have to generate a possible story from the picture without any additional requirements on memory. It must be noted that these two genres correspond to the first two of four suggested in the AphasiaBank protocol¹. These are standard tasks, widely used in the literature (see Linnik et al., 2016 for an overview) and therefore have also been employed in GREAC in order to maximize the comparability and generalizability of findings.

Patients were given as much time as needed in both tasks with minimal prompting from the examiner when absolutely necessary. Furthermore, neurotypical adults performed the same tasks, with the only difference being that in the stroke story they were asked to narrate the stroke incident of another person (usually, a person with aphasia they accompany). We have already collected 50,000 words from 60 participants on these tasks, which at a later stage can be used as a reference corpus. GREAC will also include follow up data to allow for longitudinal studies investigating the nature of connected speech impairment in aphasia. The length of patients' connected speech samples ranges from 38 to 613 s. However, their actual speech is often less due to pausing and false starts. The Cookie Theft recordings range between 69 and 486 s.

Stroke Story and Picture Description tasks were audio-recorded. All collected material was orthographically transcribed and checked for accuracy by a second transcriber. Transcriptions included both patients and examiners' speech; however, the examiner did not interfere in patient's narration, except from the case that patients needed to be encouraged to continue their story. Standard spelling conventions were maintained to increase consistency. However, sometimes it was necessary to deviate from standard conventions, in order to transcribe as accurately as possible what was said, like in cases of unfinished words or neologisms. Fluency problems, voiced and unvoiced

¹Retrieved at <http://aphasia.talkbank.org/protocol/>

TABLE 1 | Demographic and sound files information for the patients with aphasia.

No	Code	Age	Education (years)	Sex	TPO (months)	Stroke story					Cookie theft picture								
						Number of words	Duration (s)	Number of errors					Number of words	Duration (s)	Number of errors				
								PH	MS	L	N	C			PH	MS	L	N	C
1	A1	56	12	Male	16	87	143	4	6	0	0	12	39	102	2	0	0	1	0
2	A2	71	6	Female	16	439	204	9	8	6	0	15	311	203	11	2	2	0	22
3	A4	78	14	Female	40	218	152	2	0	0	0	2	379	484	2	0	4	0	12
4	A5	58	10	Male	1	165	218	2	4	0	0	1	127	207	4	4	12	0	0
5	A6	49	9	Male	21	113	101	4	38	0	6	0	74	140	8	16	2	0	2
6	A9	56	17	Male	3	154	268	34	6	0	6	2	60	103	8	2	6	0	2
7	A11	50	12	Male	20	68	233	12	6	2	16	0	88	197	26	0	0	4	2
8	A12	63	9	Male	1	536	238	4	6	4	0	2	304	183	0	4	0	0	6
9	A14	77	12	Male	1	NA	NA	NA	NA	NA	NA	NA	7	20	0	0	0	7	0
10	A15	71	8	Male	1	384	314	2	6	8	0	22	379	480	8	4	12	0	16
11	A19	64	6	Female	1	357	195	6	10	0	0	2	241	180	6	6	0	2	2
12	A20	74	12	Female	4	462	250	28	16	8	0	2	202	181	6	2	4	0	2
13	A26	67	10	Male	43	5	60	5	0	0	0	0	4	60	0	0	0	0	0
14	A29	63	6	Male	29	83	169	4	4	2	0	14	87	203	2	0	0	0	4
15	A32	73	12	Male	1	175	145	10	6	0	24	6	318	386	10	6	8	8	12
16	A33	34	9	Female	1	319	198	2	4	10	4	4	282	218	32	8	8	38	0
17	A35	86	12	Female	1	174	240	10	2	4	0	2	188	260	8	9	1	0	2
18	A37	58	6	Female	2	629	257	4	0	0	0	12	270	262	2	4	6	0	4
19	A38	72	12	Female	45	281	187	30	36	4	0	12	361	255	20	18	2	30	0
20	A42	55	12	Male	13	361	431	12	4	2	4	8	76	109	2	2	2	0	0
21	A43	45	14	Male	51	35	125	22	2	1	4	0	61	190	4	0	4	8	0
22	A46	50	6	Female	6	23	65	2	6	0	0	0	25	97	4	2	6	2	0
23	A51	79	6	Male	2	1040	480	40	50	16	10	22	329	190	16	28	28	2	6
24	A52	59	6	Male	10	8	80	0	0	0	0	0	24	160	0	0	0	0	6
25	A53	84	12	Male	1	509	310	18	22	8	34	12	624	429	26	22	10	61	16
26	A55	78	12	Female	3	328	200	8	22	4	4	8	152	100	2	0	6	6	8
27	A59	79	6	Female	6	102	83	12	2	0	0	6	93	80	8	2	0	10	0
28	A61	51	6	Male	9	173	117	2	0	2	0	6	61	60	2	0	0	0	2
29	A63	60	6	Female	1	163	98	16	12	0	0	2	74	70	4	6	2	0	2
30	A64	56	10	Male	13	5	60	0	0	0	0	0	3	60	0	3	0	0	0
31	A65	57	9	Male	20	390	325	4	3	0	0	6	130	70	1	2	1	0	3
32	A66	41	15	Male	1	43	110	4	2	0	10	0	16	68	0	2	0	0	4
33	A68	56	11	Male	45	10	60	0	4	8	0	5	6	60	2	4	0	0	6
34	A69	52	16	Female	2	269	182	34	6	0	4	4	85	120	22	10	0	0	2
35	A71	37	15	Male	1	3	60	2	2	0	0	2	5	60	2	2	2	0	4
36	A74	61	16	Male	4	499	320	16	4	0	0	2	150	90	8	2	4	0	2
37	A77	59	20	Male	1	54	85	0	0	2	2	4	49	130	4	0	0	2	18
38	A100	67	15	Female	4	142	175	32	2	8	4	0	NA	NA	NA	NA	NA	NA	NA
39	A103	30	17	Male	2	392	460	0	6	2	0	4	56	53	0	0	4	0	2
40	D1	51	10	Male	1	201	90	0	2	0	2	0	88	52	0	0	0	0	0
41	D4	62	9	Female	1	170	210	0	0	0	0	2	77	140	0	4	0	0	0
42	D6	39	14	Male	17	27	110	10	4	0	4	0	86	175	0	10	2	0	2
43	D10	54	12	Male	19	288	645	2	18	10	0	2	NA	NA	NA	NA	NA	NA	NA
44	D11	60	13	Male	4	154	190	20	14	2	4	2	107	150	14	2	0	0	0
45	D21	55	4	Male	33	44	80	0	0	0	0	2	38	105	0	0	0	0	2
46	D23	61	8	Male	2	290	132	10	10	0	6	0	154	110	4	2	2	0	2
47	D24	72	12	Male	1	10	60	0	0	0	0	4	5	60	0	0	0	0	2
48	D25	79	6	Male	1	41	170	6	8	0	6	2	91	135	20	13	2	14	10
49	D26	73	12	Male	1	494	325	10	20	4	2	0	166	140	2	10	0	2	2
50	D28	64	12	Male	1	9	60	0	0	0	2	0	29	60	3	4	2	6	2

PH, phonological errors, MS, morpho-syntactic errors, L, lexical errors; N, neologisms, C, circumlocutions.

starters and fillers, pauses, repetitions, and other phenomena of spoken interaction such as noise from the outside, coughing etc. were carefully noted, following conventions for spoken data transcription (Georgakopoulou and Goutsos, 2004: vii; and for Greek: Georgakopoulou and Goutsos, 1999, p. 70–72). All interjections were also transcribed to give an indication of the effortful speech of patients with aphasia. Transcribed files were named by using the patient's code and the type of interaction (f for spontaneous data, p for picture description). Preliminary findings of the corpus have been previously presented at Actas del III Congreso Internacional de Lingüística de Corpus (Goutsos et al., 2011a).

ANNOTATION FOR SPEECH ERRORS

The texts included in the Corpus are kept in two different formats, plain and annotated for speech errors. The typology of errors follows the standard distinction between phonological, morphological and lexical/semantic errors found in the literature (e.g., Saling, 2007; Schwartz and Dell, 2016, cf. Schwartz et al., 1994). Following the relevant bibliography we have restricted annotation to lemma level errors, omitting e.g., pronoun referent or coherence errors (see Marini et al., 2011; Harris Wright and Capilouto, 2012) (Syntactic and other sentence level errors are included in morphosyntactic errors in order to avoid unnecessary repetition, since morphosyntactic marking is obligatory in Greek). First, participants' responses were recorded and then transcribed by transcribers trained in transcribing aphasic speech samples. During error annotation, transcribers indicated all words, phrases or sentences that they found to differ from the target word, phrase or sentence expected based on the task at hand. A second check by a different researcher was then performed in order to ascertain whether the decision was correct, excluding for instance dialectal forms or other instances of variation (e.g., learned forms used by older speakers). All discrepancies were discussed and resolved.

Error classification followed, on the basis of phonological, morphological and syntactic properties of the Greek language. Error types, along with representative examples, are summarized in **Table 2** (in several cases the distinction between two types of errors is impossible; in this case both types of error are annotated). Error frequencies for each patient are shown in **Table 1**. Further details of error annotation can be found in Goutsos et al. (2011a,b). A speech sample from the Cookie theft Picture description task, including annotations according to error types, is presented in **Supplementary Table 3**. Individual data on error subtypes in the present sample are provided in **Supplementary Table 4**.

CONTRIBUTION OF THE CORPUS

The development of GREAC puts a much-needed emphasis on spontaneously produced data and the analysis of speech errors in their discourse context. Apart from the examination of speech errors, GREAC can be immensely helpful in the study of Greek aphasia in several other ways. First, information can be retrieved

TABLE 2 | Error categories in the GRAEC.

Category	Type	Example
<i>Phonological errors</i> Errors affecting isolated phonemes or syllables	PH1: phoneme omission	"peni" instead of "perni," transl. "she takes"
	PH2: phoneme addition	"axarti" instead of "xarti," transl. "paper"
	PH3: phoneme substitution	"gromiko" instead of "vromiko," transl. "dirty"
	PH4: syllable omission/addition/substitution	"eninda" instead of "eneninda," transl. "ninety"
<i>Morpho-syntactic errors</i> Errors affecting grammatical morphemes	MS1: morpheme omission	"vrisi" instead of "i vrisi," transl. "the tap"
	MS2: morpheme addition	not available in Greek for structural reasons
	MS3: general morpheme substitution	"plenete" transl. "washes herself" instead of "pleni," transl. "washes"
	MS4: aspect substitution	"plini" (perfective aspect) instead of "pleni" (imperfective aspect), transl. "washes"
	MS5: tense substitution	"valo" (present) instead of "evala" (past), transl. "put"
	MS6: agreement substitution	"mia (feminine article) proi (neuter noun)" instead of "ena proi" (neuter article-noun), transl. "one morning"
	MS7: other morpho-syntactic errors	–
<i>Lexical errors</i> Substitution of a word by another pre-existing word	L1: substitution by a word that is similar in form	"plakaki," transl. "tile," instead of "neraki," transl. "water"
	L2: substitution by a word that is similar in meaning	"ruxo," transl. "cloth," instead of "petseta," transl. "towel"
	L3: substitution by a nonsimilar word	"numera," transl. "numbers," instead of "biskota," transl. "biscuits"
<i>Neologisms</i> Errors in which more than half of the target word was incorrect, resulting in a non-existing word	N1: Neologisms that retain the structure of a Greek word and can be classified in terms of part of speech	"jerevitis" instead of "neroxitis," transl. "sink"
	N2: Neologisms that are non-recognizable and unclassified words	"dilepona"
<i>Circumlocutions</i>	Phrases used instead of a target word	"afto pu exi to nero," transl. "the one that has the water," instead of "vrisi," transl. "the tap"

from the corpus on the frequency and types of phonological and lexical errors in Greek, including neologisms and other semantically related errors. Also, comparisons between GREAC and a reference corpus of Greek, such as the Corpus of Greek Texts (CGT, see Goutsos, 2010), or a similar corpus that contains

patients' data from another language, such as the Cambridge Cookie-Theft Corpus (Williams et al., 2010), could result in interesting findings. A further interesting aspect of aphasic speech that could be explored using GREAC could be the use of combination of words or lexical bundles in terms of Biber et al. (1999). In GREAC the most frequent word combinations include phrases such as "I cannot/could not say/understand it," "how to say it/what can I say," "it must be," "these things/this thing over here" (for further examples of errors, see **Table 2**). These findings are significant not only in revealing the discourse strategies followed by speakers with aphasia (e.g., avoidance, modality, periphrasis), but also for a further exploration of formulaic language in aphasia, which, as known, is processed in different ways than the rest of the vocabulary (e.g., Wray, 2002). More generally, extended data from aphasic discourse in languages like Greek are expected to contribute to the investigation of its linguistic properties in comparison with other languages; for example, the pilot version of GREAC has been compared to English and Hungarian data, suggesting that word frequency distribution is similar to non-aphasic discourse, whereas differences between languages can be related to languages' morphological properties and particular language impairments (Neophytou et al., 2017).

The detailed error annotation can also provide important evidence for the distribution of error types, especially the pervasive phonological vs. semantic distinction (Schuchard et al., 2017; McKinnon et al., 2018; Harvey et al., 2019), as well as of sub-categories of error types, that is the relative frequency of substitution, omission, addition etc. in order to test the findings of earlier linguistic studies of aphasic discourse (e.g., Blumstein, 1973; Lesser, 1995). More details can be obtained for e.g., the distribution of phonetic vs. phonemic errors (Ash et al., 2010), semantic errors vs. errors of omission (Bormann et al., 2008), the characteristics of errors of omission (vs. errors of commission, Chen et al., 2019), the target relatedness of neologistic errors (Pilkington et al., 2017) etc. Moreover, individual information on speech sample characteristics, such as total number of words and duration, could be used by researchers for participant selection according to specific exclusion criteria, or as covariates in statistical analyses. Finally, by relating data to metadata, including the level of severity of aphasia, GREAC can contribute to the development of a baseline for Greek for the automatic recognition of aphasic speech (cf. Le and Mower Prevost, 2016 for English).

Furthermore, the question of aphasia types can be studied on a much firmer basis. Different speech errors have been associated in the literature with different aphasia types (Goodglass, 1981). For example, errors in tense and agreement marking have been associated with non-fluent types of aphasia (e.g., Friedmann and Grodzinsky, 1997), whereas phonological errors and neologisms have been associated with fluent types of aphasia (e.g., Schwartz et al., 2004; Stenneken et al., 2008). However, group studies have shown that patients belonging to different diagnostic categories often made similar errors (e.g., Ardila and Rosselli, 1993). By keeping a separate file on metadata such as the demographic and clinical characteristics of patients, we would be able to link language problems with the clinical

assessment of aphasic deficits. Thus, it would be possible to revisit the criteria of distinguishing between phenotypes of aphasia on the basis of findings from linguistic errors, instead of following the traditional taxonomy; in this sense, openly shared databases like GREAC could aid in the effort to cut the traditional aphasia classification cord, and move forward toward more progressive schemas (see also Schwartz, 1984; Caplan, 1993; Basso, 2000; Charidimou et al., 2014; Tremblay and Dick, 2016; Kasselimis et al., 2017). Finally, follow-up data would allow for longitudinal studies on the nature of connected speech impairment in different types of aphasia.

Two issues remain to be addressed. The first one is the justification of the existence of GREAC as a standalone database. There are several reasons that led us to the decision to create GREAC. First, the number of participants is much greater compared to the Greek sample included in the AphasiaBank for instance. Second, the addition of metadata is important; as stated above, apart from demographics, GREAC includes individual scores on BDAE, as well as lesion information. The inclusion of such variables in statistical analyses could strengthen the findings of any aphasiological study that would utilize our database. Third, as data collection progresses, we will be able to add data from more patients, as well as data from follow-up assessments from patients already included in the corpus. Our Unit is mainly focusing on language disorders, and therefore several patients with aphasia are referred to us by other Units inside Eginition Hospital, but also by other collaborating clinics. Moreover, we regularly perform follow-up assessments for clinical and research purposes, i.e., monitor the course of aphasic deficits for individual patients or investigate the recovery pattern and possible predictors of recovery at the group level (see for example, our small scale study conducted a few years ago, which included data from the acute and the chronic phase: Chatziantoniou et al., 2015). Such follow-up data have already been collected, and will gradually be incorporated in GREAC.

The second issue is that of sample size. There have been several databanks published in other disciplines, usually in the framework of large epidemiological studies, which include tens or even hundreds of thousands of participants. However, the GREAC is not an epidemiological databank. Its purpose is to make speech data from Greek patients with aphasia available to any researcher who wants to study aphasic errors in Greek language. To the best of our knowledge, aphasiological studies (usually in the field of psycho- or neuro-linguistics) presenting rather interesting results on Greek aphasia have samples that do not exceed the number of 20 participants (e.g., Stavrakaki and Kouvava, 2003; Koukoulioti and Stavrakaki, 2014). We argue that similar studies in the future would have much more robust and generalizable results by using a greater sample derived from GREAC. Moreover, the fact that interested researchers would have the opportunity to select samples with specific characteristics on the basis of the metadata included in GREAC, could lead to more focused studies. Considering how difficult patient recruiting is, let alone sampling that results in a homogenous group of participants, we believe that the present databank will aid researchers to save time

and allocate their resources to aspects other than baseline testing, identifying patients suitable for their study, and speech data collecting.

To summarize, the GREAC is a unique data source for Greek that provides a rich resource for future research in many aspects of language deficits in aphasia. It allows for studying large amounts of naturally occurring data, by focusing on actual language use. The data included in GREAC come from conditions which are closer to conversation or natural discourse than experimental elicitation data, based on comprehension and production tests. Therefore, although they are not of the same ecological validity as data derived from natural verbal interaction, they can help us identify phenomena that could not have occurred if a more traditional experimental design was followed. It also allows for assessing “the relative probability of particular symptom patterns and their possible etiology” (Bates et al., 1987, p. 25) and statistically evaluating aspects of actual language usage (e.g., Wright et al., 2003). Thus, we can both generalize across patients’ linguistic symptoms, by treating their discourse as a coherent whole, and study individual variation by setting it against the general pattern.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author. GREAC is available at <http://aphasia.phil.uoa.gr/>.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Eginition Hospital Ethics Committee, National and Kapodistrian Athens, School of Medicine, Greece. The

patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DK contributed to the conceptualization and design of the study, performed clinical language testing, and wrote the manuscript. MV contributed to the conceptualization and design of the study, performed linguistic data processing, and wrote the manuscript. GA performed linguistic data processing, and revised the manuscript. IE contributed to the design of the study and revised the manuscript. DG conceived and designed the study, supervised linguistic data processing, and wrote the manuscript. CP contributed to the conceptualization and design of the study, recruited patients, supervised clinical language testing, and revised the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank the patients who participated in this study. We also acknowledge the financial contribution of the Dean of the School of Philosophy, through the Special Account for Research Grants of the University of Athens. DK was supported by IKY Foundation co-financed by ESF and Greek national funds through action MIS5033021 of the Operational Programme Human Resources Development Program, Education and Lifelong Learning of the NSRF 2014–2020.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01577/full#supplementary-material>

REFERENCES

- Ardila, A., and Rosselli, M. (1993). Language deviations in aphasia: a frequency analysis. *Brain Lang.* 44, 165–180. doi: 10.1006/brln.1993.1011
- Armstrong, E. (2000). Aphasic discourse analysis: the story so far. *Aphasiology* 14, 875–892. doi: 10.1080/02687030050127685
- Armstrong, E., Ciccone, N., Godecke, E., and Kok, B. (2011). Monologues and dialogues in aphasia: some initial comparisons. *Aphasiology* 25, 1347–1371. doi: 10.1080/02687038.2011.577204
- Ash, S., McMillan, C., Gunawardena, D., Avants, B., Morgan, B., Khan, A., et al. (2010). Speech errors in progressive non-fluent aphasia. *Brain Lang.* 113, 13–20. doi: 10.1016/j.bandl.2009.12.001
- Avrutin, S. (2001). Linguistics and agrammatism. *Glott Int.* 5, 1–11.
- Badecker, W., and Caramazza, A. (1985). On considerations of method and theory governing the use of clinical categories in neurolinguistics and cognitive neuropsychology: the case against agrammatism. *Cognition* 20, 97–125. doi: 10.1016/0010-0277(85)90049-6
- Basso, A. (2000). The aphasias: fall and renaissance of the neurological model? *Brain Lang.* 71, 15–17. doi: 10.1006/brln.1999.2199
- Bates, E., Friederici, A., and Wulfeck, B. (1987). Comprehension in aphasia: a cross-linguistic study. *Brain Lang.* 32, 19–67. doi: 10.1016/0093-934X(87)90116-7
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Blumstein, S. E. (1973). *A Phonological Investigation of Aphasic Speech*. The Hague: Mouton. doi: 10.1515/9783110887433
- Bormann, T., Kulke, F., Wallech, C.-W., and Blanken, G. (2008). Omissions and semantic errors in aphasic naming: is there a link? *Brain Lang.* 104, 24–32. doi: 10.1016/j.bandl.2007.02.004
- Caplan, D. (1993). Toward a psycholinguistic approach to acquired neurogenic language disorders. *Am. J. Speech Lang. Pathol.* 2, 59–83. doi: 10.1044/1058-0360.0201.59
- Caramazza, A. (1986). On drawing inferences about the structure of normal cognitive processes from patterns of impaired performance: the case for single patient studies. *Brain Cognit.* 5, 41–66. doi: 10.1016/0278-2626(86)90061-8
- Caramazza, A., and Badecker, W. (1991). Clinical syndromes are not God’s gift to cognitive neuropsychology: a reply to a rebuttal to an answer to a response to the case against syndrome-based research. *Brain Cognit.* 16, 211–227. doi: 10.1016/0278-2626(91)90007-U
- Charidimou, A., Kasselimis, D., Varkanitsa, M., Selai, C., Potagas, C., and Evdokimidis, I. (2014). Why is it difficult to predict language impairment and outcome in patients with aphasia after stroke? *J. Clin. Neurol.* 10, 75–83. doi: 10.3988/jcn.2014.10.2.75
- Chatziantoniou, L., Kasselimis, D., Kyrozis, A., Ghika, A., Kourtidou, P., Peppas, C., et al. (2015). Lesion size and initial severity as predictors of aphasia outcome. *Stem Spraak en Taalpathologie* 20, 34–35.

- Chen, Q., Middleton, E., and Mirman, D. (2019). Words fail: Lesion-symptom mapping of errors of omission in post-stroke aphasia. *J. Neuropsychol.* 13, 183–197. doi: 10.1111/jnp.12148
- Fraser, K. C., Ben-David, N., Hirst, G., Graham, N. L., and Rochon, E. (2015). “Sentence segmentation of aphasic speech.” in *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL* (Denver, CO: Association for Computational Linguistics), 862–871. doi: 10.3115/v1/N15-1087
- Friedmann, N., and Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: pruning the syntactic tree. *Brain Lang.* 56, 397–425. doi: 10.1006/brln.1997.1795
- Gahl, S. (2002). Lexical biases in aphasic sentence comprehension: an experimental and corpus linguistic study. *Aphasiology* 16, 1173–1198. doi: 10.1080/02687030244000428
- Georgakopoulou, A., and Goutsos, D. (1999). *Text and Communication* [In Greek]. Athens: Ellinika Grammata.
- Georgakopoulou, A., and Goutsos, D. (2004). *Discourse Analysis. An Introduction, 2nd Edn.* Edinburgh: Edinburgh University Press. doi: 10.3366/edinburgh/9780748620456.001.0001
- Goodglass, H. (1981). The syndromes of aphasia: similarities and differences in neurolinguistic features. *Top. Lang. Disord.* 1, 1–14. doi: 10.1097/00011363-198109000-00004
- Goodglass, H., and Kaplan, E. (1983). *The Assessment of Aphasia and Related Disorders, 2nd Edn.* Philadelphia, PA: Lea and Febiger.
- Goutsos, D. (2010). The corpus of Greek texts: a reference corpus for Modern Greek. *Corpora* 5, 29–44. doi: 10.3366/cor.2010.0002
- Goutsos, D., Potagas, C., Kasselimis, D., Varkanitsa, M., and Evdokimidis, I. (2011a). “The corpus of Greek aphasic speech: design and compilation,” in *Las tecnologías de la información y las comunicaciones: Presente y futuro en el análisis de corpora. Actas del III Congreso Internacional de Lingüística de Corpus. Valencia: Universitat Politècnica de València*, eds M. L. Carrió Pastor and M. A. Candel Mora (Valencia: Universitat Politècnica de València), 77–86.
- Goutsos, D., Potagas, C., Kasselimis, D., Varkanitsa, M., and Evdokimidis, I. (2011b). “Studying paraphasias in a corpus of Greek aphasic discourse [In Greek],” in *Language and Memory*, eds C. Potagas and I. Evdokimidis (Athens: Synapses), 23–47.
- Grodzinsky, Y., Piñango, M., Zurif, E., and Drai, D. (1999). The critical role of group studies in neuropsychology: comprehension regularities in Broca’s aphasia. *Brain Lang.* 67, 134–147. doi: 10.1006/brln.1999.2050
- Harris Wright, H., and Capilouto, G. J. (2012). Considering a multi-level approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology* 26, 656–672. doi: 10.1080/02687038.2012.676855
- Harvey, D. Y., Massa, J. A., Shah-Basaka, P., Wurzman, R., Faseyitana, O., Sacchetti, D. L., et al. (2019). Continuous theta burst stimulation over right pars triangularis facilitates naming abilities in chronic post-stroke aphasia by enhancing phonological access. *Brain Lang.* 192, 25–34. doi: 10.1016/j.bandl.2019.02.005
- Kaplan, E., Goodglass, H., and Weintraub, S. (1983). *Boston Naming Test*. Philadelphia, PA: Lea and Febiger.
- Kasselimis, D. S., Simos, P. G., Peppas, C., Evdokimidis, I., and Potagas, C. (2017). The unbridged gap between clinical diagnosis and contemporary research on aphasia: a short discussion on the validity and clinical utility of taxonomic categories. *Brain Lang.* 164, 63–67. doi: 10.1016/j.bandl.2016.10.005
- Khudyakova, M., Bergelson, M., Akinina, Y., Iskra, E., Toldova, S., and Dragoy, O. (2016). “Russian CLiPS: a corpus of narratives by brain-damaged individuals,” in *Proceedings of LREC 2016 Workshop. Resources and Processing of Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive/Psychiatric Impairments (RaPID-2016)* (Linköping: Linköping University Electronic Press).
- Kong, A. P. H., and Law, S. P. (2019). Cantonese AphasiaBank: an annotated database of spoken discourse and co-verbal gestures by healthy and language-impaired native Cantonese speakers. *Behav. Res. Methods* 51, 1131–1144. doi: 10.3758/s13428-018-1043-6
- Koukouloti, V., and Stavrakaki, S. (2014). Producing and inflecting verbs with different argument structure: evidence from Greek aphasic speakers. *Aphasiology* 28, 1320–1349. doi: 10.1080/02687038.2014.919561
- Kuvač Kraljević, J., Hržica, G., and Lice, K. (2017). CroDA: a Croatian discourse corpus of speakers with aphasia. *Hrvatska revija za rehabilitacijska istraživanja* 53, 61–71. doi: 10.31299/hrri.53.2.5
- Laures-Gore, J., Russell, S., Patel, R., and Frankel, M. (2016). The Atlanta motor speech disorders corpus: motivation, development, and utility. *Folia Phoniatrica Logopaedica* 68, 99–105. doi: 10.1159/000448891
- Le, D., and Mower Prevost, E. (2016). Improving automatic recognition of aphasic speech with AphasiaBank. *Interspeech* 2681–2685. doi: 10.21437/Interspeech.2016-213
- Lesser, R. (1995). *Linguistic Investigations of Aphasia*. London: Whurr.
- Linnik, A., Bastiaanse, R., and Höhle, B. (2016). Discourse production in aphasia: a current review of theoretical and methodological challenges. *Aphasiology* 30, 765–800. doi: 10.1080/02687038.2015.1113489
- MacWhinney, B., Fromm, D., Forbes, M., and Holland, A. (2011). AphasiaBank: methods for studying discourse. *Aphasiology* 25, 1286–1307. doi: 10.1080/02687038.2011.589893
- Marini, A., Andreetta, S., del Tin, S., and Carlomagno, S. (2011). A multi-level approach to the analysis of narrative language in aphasia. *Aphasiology* 25, 1372–1392. doi: 10.1080/02687038.2011.584690
- McEnery, T., and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511981395
- McKinnon, E. T., Fridriksson, J., Basilakos, A., Hickok, G., Hillis, A. E., Spampinato, M. V., et al. (2018). Types of naming errors in chronic post-stroke aphasia are dissociated by dual stream axonal loss. *Nat. Sci. Rep.* 8:14352. doi: 10.1038/s41598-018-32457-4
- Mirman, D., Strauss, T. J., Brecher, A., Walker, G. M., Sobel, P., Dell, G. S., et al. (2010). A large, searchable, web-based database of aphasic performance on picture naming and other tests of cognitive function. *Cognit. Neuropsychol.* 27, 495–504. doi: 10.1080/02643294.2011.574112
- Müller, N., Guendouzi, J. A., and Wilson, B. (2008). Discourse analysis and communication impairment. In: *The handbook of Clinical Linguistics*. eds M. J. Ball, M. R. Perkins, N. Müller, and S. Howard (Blackwell Oxford), 3–31. doi: 10.1002/9781444301007.ch1
- Neophytou, K., van Egmond, M., and Avrutin, S. (2017). Zipf’s law in aphasia across languages: a comparison of English, Hungarian and Greek. *J. Quant. Linguist.* 24, 178–196. doi: 10.1080/09296174.2016.1263786
- Pilkington, E., Keidel, J., Kendrick, L. T., Saddy, J. D., Sage, K., and Robson, H. (2017). Repetition: perseverative, neologistic, and lesion patterns in jargon aphasia. *Front. Hum. Neurosci.* 11:225. doi: 10.3389/fnhum.2017.00225
- Saling, M. M. (2007). “Disorders of language,” in *Neurology and Clinical Neuroscience*, ed A. H. V. Schapira (Amsterdam: Elsevier), 31–42. doi: 10.1016/B978-0-323-03354-1.50007-9
- Schuchard, J., Middleton, E. L., and Schwartz, M. F. (2017). The timing of spontaneous detection and repair of naming errors in aphasia. *Cortex* 93, 79–91. doi: 10.1016/j.cortex.2017.05.008
- Schwartz, M. F. (1984). What the classical aphasia categories can’t do for us, and why. *Brain Lang.* 21, 3–8. doi: 10.1016/0093-934X(84)90031-2
- Schwartz, M. F., and Dell, G. S. (2016). “Word production from the perspective of speech errors in aphasia,” in *Neurobiology of Language*, eds G. Hickok and S. L. Small (Amsterdam: Elsevier), 701–715. doi: 10.1016/B978-0-12-407794-2.00056-0
- Schwartz, M. F., Saffran, E. M., Blocch, D. E., and Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain Lang.* 47, 52–88. doi: 10.1006/brln.1994.1042
- Schwartz, M. F., Wilshire, C. E., Gagnon, D. A., and Polansky, M. (2004). Origins of nonword phonological errors in aphasic picture naming. *Cognit. Neuropsychol.* 21, 159–186. doi: 10.1080/02643290342000519
- Simos, P. G., Kasselimis, D., and Mouzaki, A. (2011). Age, gender, and education effects on vocabulary measures in Greek. *Aphasiology* 25, 475–491. doi: 10.1080/02687038.2010.512118
- Stavrakaki, S., and Kouvava, S. (2003). Functional categories in agrammatism: evidence from Greek. *Brain Lang.* 86, 129–141. doi: 10.1016/S0093-934X(02)00541-2
- Stenneken, P., Hofmann, M. J., and Jacobs, A. M. (2008). Sublexical units in aphasic jargon and in the standard language: comparative analyses of neologisms in connected speech. *Aphasiology* 22, 1142–1156. doi: 10.1080/02687030701820501

- Tremblay, P., and Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain Lang.* 162, 60–71. doi: 10.1016/j.bandl.2016.08.004
- Tsapkini, K., Vlahou, C. H., and Potagas, C. (2009). Adaptation and validation of standardized aphasia tests in different languages: lessons from the Boston Diagnostic Aphasia Examination. *Behav. Neurol.* 22, 111–119. doi: 10.1155/2010/423841
- Varlokosta, S., Stamouli, S., Karasimos, A., Markopoulos, G., Kakavoulia, M., Nerantzini, M., et al. (2016). “A Greek corpus of aphasic discourse: collection, transcription, and annotation specifications,” in *Proceedings of LREC 2016 Workshop. Resources and Processing of Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive/Psychiatric Impairments (RaPID-2016), Monday 23rd of May 2016* (No. 128) (Linköping: Linköping University Electronic Press).
- Westerhout, E., and Monachesi, P. (2007). *A Pilot Study for a Corpus of Dutch Aphasic Speech (CoDAS)*. Available online at: <http://citeseerx.ist.psu.edu/viewdoc/download?>
- Williams, C., Thwaites, A., Buttery, P., Geertzen, J., Randall, B., Shafto, M., et al. (2010). “The Cambridge Cookie-Theft Corpus: a corpus of directed and spontaneous speech of brain-damaged patients and healthy individuals,” in *Proceedings of the International Conference on Language Resources and Evaluation* (Valletta: European Language Resources Association (ELRA)).
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511519772
- Wright, H. H., Silverman, S., and Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology* 17, 443–452. doi: 10.1080/02687030344000166

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Kasselimis, Varkanitsa, Angelopoulou, Evdokimidis, Goutsos and Potagas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.