
This is the **accepted version** of the journal article:

Salmons Llussa, Io; Muntané Sánchez, Helena; Gavarró Algueró, Anna. «An analysis of descriptions by Catalan-speaking individuals with aphasia».

Aphasiology, Vol. 40, Num. 2 (2024), p. 389-411 DOI
10.1080/02687038.2024.2428424

This version is available at <https://ddd.uab.cat/record/329199>

under the terms of the  license.

An analysis of descriptions by Catalan-speaking individuals with aphasia

Io Salmons^a, Helena Muntané-Sánchez^b and Anna Gavarró^a

^aDepartament de Filologia Catalana, Universitat Autònoma de Barcelona, Barcelona, Spain;

^bHospital de Sant Rafael, Barcelona, Spain

ARTICLE HISTORY

Compiled October 15, 2024

ABSTRACT

Background

Description tasks are used to evaluate and investigate the production abilities of people with language impairments such as aphasia.

Aims

The goal of the present study is to investigate the production abilities of Catalan-speaking individuals with aphasia (IWA), as well as a smaller sample of individuals with cognitive impairment (IWCI), in comparison with those of healthy controls. A second goal is to evaluate whether the scoring system of the Catalan version of the CAT (CAT-CAT; Salmons, Rofes, and Gavarró 2021) is helpful to discriminate different patterns of language impairment.

Methods and Procedure

In the study, 109 control subjects, 20 IWA and 4 IWCI were asked to orally describe a picture from the CAT-CAT. The scoring method consisted in a closed-rating system that evaluates productivity, discourse efficiency, fluency, grammatical complexity and grammaticality.

Results

The results show that the overall scores of the control subjects were significantly higher than those of the experimental groups; the difference between the two experimental groups was not significant. Yet, the group of IWA showed greater intersubject variability than the group of IWCI. Also, the IWA were consistently worse in fluency, grammaticality and grammatical complexity than the IWCI, which indicated a different pattern of performance between the two groups.

Conclusions

Our findings therefore show that the oral picture description of the CAT-CAT is sensitive to the language impairment of subjects with aphasia and cognitive impairment. Moreover, the rating system put forward allows us to uncover different patterns of language impairment, since it includes variables to evaluate separately content and discourse efficiency on the one hand, and grammatical features on the other hand.

KEYWORDS

description; Catalan; Comprehensive Aphasia Test; aphasia; cognitive impairment

1. Introduction

Picture description and narrative tasks are used to investigate and evaluate the production abilities of people with language impairments, such as aphasia or cognitive

impairment (among others, Andreetta and Marini 2015; Armstrong 2000; Cummings 2019; Forbes-McKay and Venneri 2005; Gordon 2008; Groves-Wright, Neils-Strunjas, Burnett, and O’Neill 2004; Hameister and Nickels 2018; Hazamy and Obermeyer 2020; Schönberger et al. 2014; Wright, Silverman, and Newhoff 2003). In particular, the picture description task of the *Comprehensive Aphasia Test* (CAT; Swinburn, Porter, and Howard 2005) has been shown to be suitable to assess subjects with brain damage both for research and clinical purposes (Kuvač Kraljević, Matić Škorić, and Lice 2023; Vandenborre, Visch-Brink, van Dun, Verhoeven, and Mariën 2018; Zakieline, Turk, Hashem, and Shafik 2020).

The aim of the present study is to investigate the production abilities of Catalan-speaking individuals with aphasia (IWA) and with cognitive impairment (IWCI) in comparison with those of controls without brain damage. A second goal is to test whether the scoring system of the Catalan version of the CAT (CAT-CAT; Salmons et al. 2021), which differs from the original in being a closed-rating system, can be helpful to discriminate language deficits from preserved language. Finally, we also explore whether the scoring system that we propose can be useful to qualitatively analyse different patterns of language impairment corresponding to different aphasia subtypes or language disorders.

This paper proceeds as follows: in the remainder of this section, we present the Catalan version of the CAT (1.1), and the picture description task of CAT-CAT which corresponds to its spoken version, as well as the scoring system that we propose (1.2). In section 2, we present the methods of the study we report; in section 3, we present the results; in section 4, we discuss the results and, in section 5, we present avenues for future research and draw our conclusions.

1.1. The Catalan version of the CAT

The CAT-CAT is the Catalan version of the *Comprehensive Aphasia Test* (CAT), originally developed for English (Swinburn et al. 2005; Swinburn, Porter, and Howard 2023). The CAT is a test that includes a large number of tasks that evaluate cognitive and linguistic skills and that can be used to diagnose aphasia and to evaluate the efficiency of therapy and intervention (e.g. Hoover, DeDe, and Maas 2021). It is an important tool in two respects: first, it not only evaluates language, but also cognitive verbal and nonverbal abilities, and therefore is broader in scope than other tools available (Howard, Swinburn, and Porter 2010a, 2010b); and, second, it has been developed or is currently under development for a large number of languages, which implies that it will have value as a research tool for comparative studies (Fyndanis et al. 2017; Martínez-Ferreiro, Arslan, et al. 2024). In fact, the CAT has already been published for Arabic (Abou El-Ella et al. 2013), Croatian (Kuvač Kraljević, Matić, and Lice 2020), Dutch (Visch-Brink, Vandenborre, de Smet, and Mariën 2014), Hungarian (Zakariás and Lukács 2022) and Turkish (Maviş, Tunçer, Selvi-Balo, Tokaç, and Özdemir 2021; Özdemir, Maviş, and Tunçer 2022). This is largely the result of the collaborative effort of the group Collaboration of Aphasia Trialists (IS1208, COST Action, PI: M. Brady, 2013-2017; and The Tavistock Trust for Aphasia, PI: M. Brady, 2017-2023).

To this, we can add that there is no diagnostic tool for aphasia in Catalan, with the only exception of the *Barcelona Test* (Peña-Casanova and Sánchez Benavides 2019), which was designed to diagnose different type of cognitive deficits and includes some linguistic tests that have been adapted specifically taking into account the grammat-

ical features of Catalan (Lluent Vallet, Peña-Casanova, and Böhm 2002). In current practice, speech therapists and clinicians resort to tests in Spanish or to informal literal translations from Spanish (Lluent Vallet et al. 2002). This procedure can be detrimental for the patient since the diagnosis and treatment are less accurate, given that the patient is not evaluated in his or her mother tongue. Furthermore, as one reviewer pointed out, in the case of bilingual speakers, assessing them in only one of the languages they speak could lead to overlooking certain difficulties that might surface in the unassessed language. Also, it is worth noting that not all Catalan-speakers are bilingual speakers of Catalan and Spanish. On the other hand, informal translations are problematic because they undermine the validity and reliability of the test, but also because they neglect the structural and cultural particularities of the language in which the participant is assessed (Paradis and Libben 1987). For this reason, Ivanova and Hallowell (2013, p. 8) argue that literal translations are never appropriate given the structural differences across languages, including those that show similarities and have a common origin, such as Catalan and Spanish. This is exemplified by the finding that Catalan-speaking children with developmental language disorder show a different pattern of impairment in Catalan than in Spanish, as only the former have difficulties in producing third person accusative clitics (Gavarró 2017). This is why one of the goals of the Working Group 2 of the Collaboration of Aphasia Trialists is to adapt the CAT to different languages considering the linguistic and cultural aspects of each language (Ivanova, Kuptsova, and Dronkers 2017; Martínez-Ferreiro, Arslan, et al. 2024).

Like other versions of the CAT, CAT-CAT consists of twenty-seven subtasks distributed in two parts: a short cognitive screening test and a linguistic test. The cognitive subpart includes subtests on line bisection, semantic memory, word fluency, memory recognition, object use gesture and arithmetic. This part is meant to screen for nonverbal cognitive deficits such as visual and memory deficits, or apraxias and dyscalculias. The language battery, on the other hand, evaluates language comprehension, production, writing and reading. The comprehension subtests assess the comprehension of spoken words, written words, spoken sentences, written sentences, and spoken paragraphs. On the production part, the test includes repetition of simple and complex words, nonword repetition, repetition of digit strings, sentence repetition, object and action naming, and spoken picture description subtests. Reading and writing are evaluated with subtests in simple and complex word reading, functional word reading, nonword reading, copying, written picture naming, writing to dictation and written picture description. A more recent version of CAT (Swinburn et al. 2023) retains the same tasks, and includes the *Aphasia Impact Questionnaire* (AIQ-21; Swinburn et al. 2018) that allows the clinician to learn the subjective perception of the patient on the impact of aphasia in his or her life. This part has also been adapted to Catalan and is available on line at <https://www.aiq-21.net/download>.

1.2. The description task

Subtest 19 assesses the oral production skills of participants. The participant is asked to describe the picture in Figure 1, which depicts a series of interrelated events (a cat reaching out for a fish in a tank and making the books on a bookshelf fall on top of a sleeping lady’s head, etc.) in order to allow for the production of a narrative discourse (Howard et al. 2010a). As the original authors of the test argued, this task allows for the collection of a narrative (Swinburn et al. 2005) without it actually being a measure



Figure 1. Image from the description task of the CAT-CAT (Salmons et al. 2021).

of functional communication, and without assessing nonverbal communication. Still, it does have the merit of providing information on a wide variety of aspects in language production. It has the added value to be easily administered by professionals with no linguistic background.

Both the picture and the scoring system of this subtest in the Catalan version slightly differ from the original one and from other versions of CAT (see Martínez-Ferreiro, Arslan, et al. (2024) for a more detailed discussion). As with all descriptions, the scoring method is essential in order to achieve the goals of the task. The original scoring system was designed to discriminate aphasia from other conditions with the following criteria: “The scored parameters are accuracy/appropriateness of information presented, speed/fluency of presentation, and syntactic accuracy and variety. Thus, the numbers of correct and appropriate information-carrying words/word-like units (e.g. potted plant, tape recorder) are scored as well as the number of inappropriate information-carrying words/word-like units. Syntactic ability is assessed by rating the variety of syntactic structures and grammatical well-formedness of the phrases used. Speed of response is also noted and contributes to the overall score” (Swinburn et al. 2005, p. 30).

This scoring system has received a number of criticisms by authors who have implemented the test in different languages and who have proposed alternative scoring methods (e.g. Martínez-Ferreiro, Arslan, et al. 2024; Vandenborre et al. 2018). One of its main problems is that the final score depends on the number of appropriate content words produced, favouring fluent speakers. Despite the fact that inappropriate words are to be subtracted from the number of appropriate words and that, as the authors argue, “fluent speakers often have a higher proportion of inappropriate words [...]” (Swinburn et al. 2005, p. 61), there are different types of fluent aphasia that show different patterns of impairment and compensatory strategies in oral production (Christiansen 1995). For example, a person with an anomic aphasia who is fluent may try to compensate his or her naming difficulties using circumlocutions or more generic words that cannot be counted as inappropriate words despite being of little informative weight.

This has been observed across languages in different studies. Christiansen (1995) observed that English-speaking people with fluent aphasia tended to produce irrelevant

and unrelated utterances in description and narrative tasks, such as comments on the task itself. Andreetta, Cantagallo, and Marini (2012), on the other hand, examined the descriptions of 10 Italian-speaking participants with anomic aphasia in comparison with 10 healthy participants, and concluded that descriptions of the group with aphasia contained “a number of unnecessary words, tangential speech or repetitions” (p. 1790). These results have been replicated more recently by Hazamy and Obermeyer (2020), who analysed the descriptions of 13 fluent and 18 nonfluent IWA and concluded that the former produced a greater proportion of non-specific and incorrect utterances (22.85% and 19.42% of their errors, respectively) than the latter (11.62% and 11.15%, respectively), whose errors consisted more frequently of omissions. Hence, although the original scoring system was meant to be unbiased with respect to fluency (Schönberger et al. 2014), it nevertheless favours the fluent type of aphasia and, in general, more talkative people, as it heavily relies on an absolute number of words produced, without a maximum score. In addition, (in)appropriateness of words is a rather vague concept, as is the concept of content words (e.g. some prepositions are considered content words and need to be counted, while other are considered function words and have to be excluded), despite the suggestions in Swinburn et al. (2005) in an attempt to be explicit.

On the other hand, even though the original task is not intended to evaluate the narrative abilities of the patients and, as a consequence, does not take into account the communicative efficiency of the discourse, we believe that a different type of scoring that assesses the informativeness of the productions can be useful for clinical purposes. In fact, some of the problems that the IWA experience in production can impact heavily on their overall communicative efficiency. This has been observed in the literature consistently and independently of the type of aphasia (see Linnik, Bastiaanse, and Höhle 2016 for a review). Despite the fact that there are different methods for assessing informativeness and communicative efficiency, one of the most widely used parameters is the number of relevant correct events (or content units) described or produced. Dalton and Richardson (2019), for instance, examined the descriptions of 38 IWA and concluded that their score on overall informativeness was lower than the controls’ score. The authors also argued that the description of main events could even distinguish IWA from people with brain injuries but without aphasia. Gordon (2008), on the other hand, analysed the descriptions of 8 fluent and 8 nonfluent English-speaking individuals with aphasia and did not observe any significant differences between the two groups in the number of correct content information units produced, despite differences in other aspects, such as lexical diversity. Andreetta and Marini (2015) examined the descriptions of Italian-speaking individuals with fluent aphasia and concluded that the informativeness of their productions was compromised. The differences between fluent and nonfluent speakers with aphasia seem to lie in the compensatory strategies, as discussed above, and the type of errors. For example, several authors highlight the fact that individuals with fluent aphasia tend to produce more paraphasias and neologisms, while the discourse of nonfluent speakers with aphasia is rather characterised by agrammatism and a greater number of apraxic errors (see Pallickal and Hema 2020 for a review). Informativeness is, therefore, a valid diagnostic criterion for all types of aphasia.

1.2.1. The scoring system

In contrast to the original scoring system, the one that we propose is a closed-type system with a maximum score of 28 (see Table 1). This maximum score results from

taking into account several variables that evaluate productivity, discourse efficiency, fluency and grammar (grammatical complexity and grammaticality). Productivity is assessed by counting the relevant number of nouns and verbs produced, that is, the number of relevant characters or objects (e.g. cat, woman, child...) and actions (e.g. sleep, point, reach out, fall...) identified, whereas the discourse efficiency is evaluated by counting the number of relevant events identified (e.g. a woman sleeping, a cat reaching out for a fish...). The total score of nouns, verbs and events identified constitutes the score on the content production abilities.

On the other hand, form is evaluated through the three criteria included in the original CAT, using a Likert-type scale: the fluency of the discourse, where pauses and delays in speech are taken into account; its grammaticality or the well-formedness of utterances, and the grammatical complexity or the variety of syntactic structures produced. The total score constitutes the score on the form production abilities. This system maintains the basic structure in the original version, which evaluates content and form separately in order to be able to “reveal the nature of the impairment” (Howard et al. 2010b, p. 3). A closed-scoring system like the one that we propose is more suitable to compare the performance of different subjects in keeping the differences between fluent and nonfluent speakers under a certain threshold. The original Catalan scoring sheet can be found at <https://ddd.uab.cat/record/273019>.

The evaluation of informativeness or discourse efficiency as a separate criterion is useful in the clinical setting in order to evaluate the communicative abilities of the participants, as enhancing communication is one of the main goals of rehabilitation; moreover, it allows us to detect content production alterations even in fluent aphasias. Thus, the addition of this criterion improves the task as it becomes more sensitive to production deficits. In fact, other versions of the CAT have also introduced similar changes: for example, the Norwegian version (Jensen, Norvik, and Gram Simonsen 2024) takes into account whether the participants mention four relevant events (man sleeping, girl pointing, cat trying to catch fish and books falling). The scoring is based on a Likert scale where missing information scores 0, incomplete or imprecise information scores 1, and complete and precise information scores 2. Other adaptations have adopted the same system (Martínez-Ferreiro, Arslan, et al. 2024).

However, unlike other adaptations of the CAT (e.g. Martínez-Ferreiro, Quique, Rodríguez, and Orellana 2024), we chose to differentiate productivity (number of relevant nouns and verbs produced) from communication efficiency (number of relevant content units mentioned) in order to get a more complete profile of the person evaluated in accordance with the objectives of the test and the task (Howard et al. 2010b). Productivity values give us additional information on whether the discourse is fluent or nonfluent, which is a more exact measure than only considering speech characteristics like the tempo or the duration of the pauses. Furthermore, it has the advantage of being more enlightening as it can indicate whether the problems are due to a lexical (productivity) or a grammatical (grammaticality and grammatical complexity) impairment. Given that different language disorders present lower content units in their descriptions, it is interesting to determine the underlying cause of the impairment or, in other cases, to establish the factors that increase the informativeness of the discourse. Actually, people with Alzheimer’s dementia have been shown to perform similarly to controls in the syntactic domain in oral description tasks, while the communication efficiency of their discourse is as impaired as those of the IWA (see Pallickal and Hema 2020 for a review). Finally, our scoring system also enables us to see whether the decreased productivity that some participants may show affects both nouns and verbs, or only one of these two categories.

Table 1. Scoring system of the description task of the CAT-CAT.

	Max	Description	Items
<i>Content</i>			
Nouns	8	Number of relevant of characters or objects identified	Woman, sofa, child, cat, fish, books, shelves, other elements in the background.
Verbs	5	Number relevant actions identified	Sleeping, reading, pointing, reaching out, falling.
Events	6	Number of relevant events identified	A woman sleeping, a child pointing, a cat reaching out for a fish, books falling, description of the background.
Total	19		
<i>Form</i>			
Fluency	3	Pauses, delays...	0 (consistent and significant delay), 1 (frequent and significant delay), 2 (occasional delay), 3 (without delay)
Grammaticality	3	Well-formedness of utterances	0 (absent), 1 (weak), 2 (medium), 3 (good)
Complexity	3	Variability of syntactic constructions	0 (mutism, stereotypy), 1 (one syntactic structure, list), 2 (2 syntactic structures), 3 (3 or more syntactic structures)
Total	9		
Overall score	28		

In the scoring system that we propose, the results are also easier and faster to score than in the original. This is also the main reason why other measures to evaluate discourse abilities –lexical variety, coherence, units of new information or others (Dietz and Boyle 2018)– have been discarded to avoid time-consuming assessments. As already argued before, the number of relevant events as an indicator of communication efficiency has been shown to be a good diagnostic criterion for aphasia and suffices for this purpose. For a more complete profile of the productions skills of the participants, the results can be individually and qualitatively analysed.

In scoring, only the production of the significant words or events listed in the scoring sheet were counted. So, for example, participant A5 scored only 1 on the content parameter (see Table 7) because he only mentioned one relevant noun (*Llibres. Conte.* ‘Books. Story-tale.’). Omissions or errors like semantic paraphasias or neologisms were excluded from the score. For example, participant A13 said *llet* ‘milk’ instead of *gat* ‘cat’ and *pobres* ‘the poor’ instead of *llibres* ‘books’, and these productions do not qualify as relevant nouns (*I mira el llet va a caure, fa caure tots els pobres.* ‘And look, the milk falls, makes all of the poor fall.’). Regarding the form parameter, we used the same system as in the original CAT: a scale from 0 to 3 (see Table 1).

If a participant responded in Spanish, something that happened on occasion, the response was evaluated in the same way as the Catalan responses. The advantage of the description task is that it could actually be administered in any language. The production of words or expressions in Spanish were not penalised given that it is common for some bilingual speakers of Catalan and Spanish to switch from one language to another. In fact, this behaviour was also observed in some control subjects, which indicates that it cannot be considered the result of an executive control deficit (see Lerman, Pazuelo, Kizner, Borodkin, and Goral 2019 for a review of the literature on this topic).

The transcriptions of the descriptions are available at <https://ddd.uab.cat/record/272356> and the individual scores are in Table 7.

2. Methods

The oral production skills of the participants were assessed with the Catalan version of the *Comprehensive Aphasia Test* (Salmons et al. 2021). The task was conducted with a sample of Catalan-speaking individuals with aphasia and healthy subjects, as well as a small group of individuals with cognitive impairment. Here we only report the participants’ characteristics, the procedure and the statistical analysis’ details, since the materials and the scoring system are described in section 1.2.

2.1. Participants

One hundred and nine non-brain-damaged subjects of ages between 18 and 90 years ($M = 50$, $SD = 19.15$), twenty people with aphasia of ages between 44 and 92 ($M = 63.3$, $SD = 15.24$) and four people with cognitive impairment of ages between 69 and 86 years ($M = 78$, $SD = 6.97$) participated in the present study. All the participants were speakers of Catalan, and they all reported that Catalan was their first language or that they were balanced bilingual speakers of Catalan and Spanish. Our participants had different educational backgrounds (Table 2). According to self-report, the control subjects had not suffered brain damage, cognitive disorders or psychiatric conditions prior to their participation in the study.

Table 2. Demographic characteristics of the participants

	Controls <i>n</i> =109	IWA <i>n</i> =20	IWCI <i>n</i> =4
Age (years)			
18-29	22	0	0
30-39	13	0	0
40-49	16	5	0
50-59	22	4	0
60-69	19	4	1
70-79	9	3	2
+80	8	4	1
Education			
Basic	28	3	1
Intermediate	43	9	2
Higher	38	8	1
Sex			
Female	65	9	1
Male	44	11	3

For the participants with aphasia or cognitive impairment, details regarding the diagnosis, etiology and time post-onset were collected from their medical health record (Table 3). Time post-onset is the time, in years, that has passed between brain damage and the evaluation. Except for three IWA (A13, A15 and A16), who were evaluated during the acute phase of aphasia, the other IWA presented chronic aphasia (≥ 6 months post-onset). According to Swinburn et al. (2005), the CAT can be administered from three weeks after brain damage.

The heterogeneity of the experimental groups is representative of the heterogeneity usually found in people with language impairments. The IWA and IWCI were diagnosed by speech pathologists and neurologists of several hospitals in the Barcelona area via clinical criteria and formal or informal tests mostly carried out in Spanish, since, as mentioned, there is a lack of standardised tools to diagnose language disorders in Catalan. In the case of IWCI, they were diagnosed with the Mini-Mental State Exam (MMSE; Folstein, Folstein, and McHugh 1975). The etiology of the cognitive impairment in IWCI also varies across patients: for example, C1’s cognitive impairment is due to a progressive supranuclear palsy, C2 and C3 suffer from Parkinson’s and Alzheimer’s disease, respectively, and the cause of the cognition impairment of C4 is unknown.

A Mann-Whitney test revealed that there are significant differences between the control and IWA groups ($U=696$, $p=0.01$) and the control and IWCI groups ($U=37.5$, $p=.005$) with respect to age. There are no significant differences among the three groups regarding sex ($\chi^2_2=3.1$, $p=.212$) or education ($\chi^2_2=.77$, $p=.679$) according to the Kruskal Wallis test.

2.2. Procedure

The study was approved by the Ethics Committee of the Universitat Autònoma de Barcelona (CEEAH 5656) and it was conducted following the Declaration of Helsinki ethical standards. All the participants provided their written informed consent prior to participation. The participants were first interviewed in order to collect the relevant personal data. They then completed the CAT-CAT (Salmons et al. 2021) in one or two individual sessions. For subtest 19 the experimenter asked the participants to orally describe what was happening in the picture shown in Figure 1. There was no

Table 3. Individual characteristics of participants with aphasia and cognitive impairment.

Subject	Age	Sex	Education	Handedness	TPO	Type	Etiology
<i>IWA</i>							
A1	78	F	3	R	12	anomic	ICVA
A2	48	M	2	R	6	unknown	CVA
A3	56	F	3	R	9	anomic	CVA
A4	60	M	3	R	8	unknown	CVA
A5	45	M	2	R	9	mixed TA	CVA
A6	54	F	2	R	10	Wernicke	ICVA
A7	47	F	3	R	5	global	ICVA
A8	49	F	3	R	11	unknown	HCVA
A9	51	F	3	R	24	unknown	tumour
A10	78	M	2	R	2	mixed TA	HCVA
A11	44	M	3	R	3	mixed TA	HCVA
A12	92	M	2	R	0.84	conduction	CVA
A13	86	F	2	R	0.33	mixed TA	HCVA
A14	71	M	2	R	0.66	sensory TA	ICVA
A15	85	M	1	R	0.33	anomic	ICVA
A16	62	F	2	R	0.4	anomic	HCVA
A17	53	M	2	R	3.5	mixed TA	HCVA
A18	80	F	1	R	0.7	anomic	ICVA
A19	66	M	1	R	3.5	mixed TA	HCVA
A20	61	M	3	R	11	global	ICVA
<i>IWCI</i>							
C1	69	M	2	R	1		PSP
C2	78	M	2	R	1		PD
C3	79	M	3	R	2		AD
C4	86	F	1	R	2		unknown

Sex: F = female; M = male; Education: 1 = basic; 2 = intermediate; 3 = higher; Handedness: R = right; TPO: time post-onset in years; Aphasia: TA = transcortical aphasia; Etiology: AD = Alzheimer's disease; CVA = cerebrovascular accident; H = hemorrhagic; I = ischemic; PD = Parkinson's disease; PSP = progressive supranuclear palsy.

time limit. The responses were recorded and transcribed for further analysis. The transcriptions and some of recordings have been published as part of the Corpus CAT-CAT at <https://filcat.uab.cat/einaling/> for educational purposes.

2.3. Statistical analysis

Descriptive statistics are presented as mean, standard deviation and range. Distribution for normality was assessed with the Shapiro-Wilk test. The Kruskal-Wallis test was used to compare the distributions of the three groups given the non-normal distribution of the data. In addition, the Mann-Whitney-Wilcoxon test was performed for the pairwise comparison of the groups with the Bonferroni correction method. Both Pearson and Spearman correlation coefficients were used to measure lineal and monotonic associations between different factors. The statistical level accepted as significant was 5% ($p < 0.05$). The statistical analysis was performed using the R Statistical Software (v4.1.2; R Core Team 2021). As mentioned before, the datasets analysed for this study can be found at the UAB Repository of Digital Documents (<https://ddd.uab.cat/record/272356>).

3. Results

The results by group appear in tables 4 (for controls and IWA) and 5 (for IWCI). Significant correlations between the total score and age (Pearson test = -0.21, $p = 0.027$) and education level ($r = 0.25$, $p = 0.01$) were found in the control group. On the other hand, no significant correlations were found between the total score and gender in the control group. With respect to the two experimental groups, no significant correlations were found between education level, gender or age and their overall scores.

The results show that the overall scores of the control subjects were higher than those of the experimental groups. According to the Kruskal-Wallis test, there were significant differences between the three groups in all the variables except for one, the verb score (nouns: $H=16.17$, $p < .001$; events: $H=11.11$, $p < .001$; total content: $H=17.25$, $p < .001$; fluency: $H=89.75$, $p < .001$, grammaticality: $H=76.7$, $p < .001$, complexity: $H=47.7$, $p < .001$, total form: $H=100.75$, $p < .001$; total score: $H=28.29$, $p < .001$).

The difference between the overall performance of control participants and IWA was significant (Wilcoxon rank test = 1805.5, $p < .001$). In particular, there were differences in the content total score ($W = 1607$, $p < .001$) and in the form total score ($W = 2012.5$, $p < .001$). With respect to specific variables, there were differences between the two groups both in the noun score ($W = 1613$, $p < .001$) and in the event content score ($W = 1510$, $p < .001$). Also, there were statistical differences in form in fluency ($W = 1907.5$, $p < .001$), grammaticality ($W = 1845$, $p < .001$) and complexity ($W=1526$, $p < .001$).

Even though the results of the group of IWCI have to be taken with caution given the small sample size, we also analysed them in comparison with the other two groups. In fact, the IWCI's overall performance differed significantly from the controls' performance ($W = 402.5$; $p < .001$). More specifically, statistically significant differences were found both for the total content score ($W = 388.5$, $p < .001$) and the total form score ($W = 379.5$; $p < .001$). If we look closely at specific items, there were significant differences between these two groups in the noun score ($W = 353$, $p = 0.028$), in flu-

Table 4. Scores by group in the description task of the CAT-CAT

	Group	Mean	SD	Range
<i>Content</i>				
Nouns	Control	6.86	0.94	5-8/8
	IWA	5.30	2.15	0-8/8
Verbs	Control	4.20	0.83	2-5/5
	IWA	3.75	1.25	0-5/5
Events	Control	4.71	1.02	2-6/6
	IWA	3.65	1.66	0-6/6
Total	Control	15.78	2.33	9-19/19
	IWA	12.70	4.23	1-19/19
<i>Form</i>				
Fluency	Control	3.00	0.00	3-3/3
	IWA	1.85	0.93	0-3/3
Grammaticality	Control	2.99	0.09	2-3/3
	IWA	2.00	0.92	0-3/3
Grammatical complexity	Control	3.00	0.00	3-3/3
	IWA	2.50	0.69	1-3/3
Total	Control	8.99	0.09	8-9/9
	IWA	6.35	2.32	1-9/9
Overall score	Control	24.75	2.35	18-28/28
	IWA	19.05	6.29	2-28/28

Table 5. Scores by IWCI in the description task of the CAT-CAT

	Mean	SD	Range
<i>Content</i>			
Nouns	5.00	1.83	3-7/8
Verbs	3.75	0.50	3-4/5
Events	3.50	1.29	2-5/6
Total	12.25	1.71	10-14/19
<i>Form</i>			
Fluency	2.50	0.58	2-3/3
Grammaticality	2.75	0.50	2-3/3
Grammatical complexity	3.00	0.00	3-3/3
Total	8.25	0.50	8-9/9
Overall score	20.50	1.91	18-22/28

Table 6. Wilcoxon coefficients with adjusted p values (with Bonferroni correction)

	Controls-IWA	Controls-IWCI	IWA-IWCI
<i>Content</i>			
Nouns	1613 $p = .0012$	353 $p = .0862$	46 $p = 1.0000$
Verbs	1316.5 $p = .34$	303.5 $p = .45$	44 $p = 1.0000$
Events	1510 $p = .013$	337.5 $p = .155$	44.5 $p = 1.0000$
Total	1607 $p = .0021$	388.5 $p = .0229$	46 $p = 1.0000$
<i>Form</i>			
Fluency	1907 $p < .001$	327 $p < .001$	24 $p = .6$
Grammaticality	1845 $p < .001$	270.5 $p = .0012$	20 $p = .3048$
Grammatical complexity	1526 $p < .001$	218 $p < .001$	24 $p = .3$
Total	2012.5 $p < .001$	379.5 $p < .001$	18 $p = .26$
Overall score	1804.5 $p < .001$	402.5 $p = .012$	37 $p = 1.0000$

ency ($W = 327$; $p < .001$) and in grammaticality ($W = 270.5$, $p < .001$). Unlike with the group of IWA, the performance of IWCI did not differ significantly from that of the control group in the verb and event score, nor in grammatical complexity score. After applying the Bonferroni correction, all the above mentioned p values remained significant except for the difference between the control and IWCI groups in the noun score (see the adjusted p values in Table 6).

If we compare the performance of the two experimental groups, no statistical significant differences were found under any of the criteria evaluated. However, if we look at the descriptive results in tables 4 and 5, the individuals with cognitive impairment presented greater difficulties and intersubject variability in the content subscores than in the form subscores: their scores on nouns, verbs and events was low, whereas their scores on grammaticality and complexity were higher than for the IWA. To further illustrate the variability within each group, we report the individual results of the participants of the two experimental groups in Table 7. As for the control group, intersubject variability was much smaller and only observed in content parameters (Table 4).

We also calculated the correlation between the scores in the different items and the time post-onset (in years) in the group of IWA to learn whether this factor could have had any impact on the results of the IWA. The Pearson coefficient of correlation, however, showed that there were no significant associations between any of the different items and the years that had passed after brain damage. However, a Spearman correlation coefficient of .47 ($p = .0358$) was obtained between the noun score and the TPO, which indicates a moderate association between these two factors.

Pearson coefficients were also calculated for the purpose of exploring possible associations between the different variables in the group of IWA (Table 8). The analysis revealed strong correlations between the total score of content and form and their re-

Table 7. Individual scores by IWA and IWCI in the description task of the CAT-CAT.

	N	V	E	<i>Content</i>	F	G	C	<i>Form</i>	<i>Total</i>
<i>IWA</i>									
A1	6	2	3	11	3	2	3	8	19
A2	7	4	4	15	2	2	2	6	21
A3	5	3	1	9	1	1	2	4	13
A4	7	4	5	16	2	2	2	6	22
A5	1	0	0	1	0	0	1	1	2
A6	8	5	6	19	3	3	3	9	28
A7	7	3	1	11	1	2	3	6	17
A8	6	5	5	16	2	3	3	8	24
A9	8	5	4	17	3	3	3	9	26
A10	5	5	5	15	2	3	3	8	23
A11	8	4	5	17	3	2	3	8	25
A12	4	4	3	11	1	2	3	6	17
A13	5	3	2	10	2	1	2	5	15
A14	6	5	6	17	2	2	3	7	24
A15	3	5	5	13	2	2	3	7	20
A16	5	4	4	13	3	3	3	9	22
A17	5	3	4	12	1	2	2	5	17
A18	4	4	4	12	2	3	3	8	20
A19	0	3	3	6	0	0	1	1	7
A20	6	4	3	13	2	2	2	6	19
<i>IWCI</i>									
C1	6	4	4	14	3	2	3	8	22
C2	3	4	3	10	2	3	3	8	18
C3	7	3	2	12	2	3	3	8	20
C4	4	4	5	13	3	3	3	9	22

N = nouns; V = verbs; E = events; F = fluency; G = grammaticality; C = grammatical complexity.

spective parameters, as expected, and also between these two measures ($r(18) = .83$, $p < .001$). The event score, on the other hand, is strongly associated with the verb score ($r = .82$, $p < .001$), whereas no significant correlation is observed between the event and noun scores. It is pertinent to consider the fact that the event score is significantly correlated with all other measures, though to a lesser degree than with the verb score.

If we examine the form parameters, they significantly correlate with other scores but to different degrees. Fluency, for instance, is strongly associated with the noun score ($r = .73$, $p < .001$) and the other form parameters. Grammaticality, on the other hand, shows a stronger relationship with the verb score ($r = .73$, $p < .001$) than with the noun score ($r = .64$, $p = .0024$), and with grammatical complexity and fluency ($r = .84$, $p < .001$ and $r = .74$, $p < .001$, respectively). Grammatical complexity is more *strongly* associated with the verb score ($r = .64$, $p = .0023$) than the noun score ($r = .57$, $p = .009$), though these results should be considered carefully due to the higher p -values. Also, grammatical complexity is more strongly correlated with grammaticality ($r = .84$, $p < .001$) than fluency ($r = .697$, $p < .001$).

Table 8. Pearson correlation coefficients with *p* values (IWA)

	N	V	E	CT	F	G	C	FT
N		.4783 p = .0329	.4127 p = .0705	.8128 p < .001	.7302 p < .001	.6389 p = .0024	.5679 p = .0090	.7139 p < .001
V	.4783 p = .0329		.8156 p < .001	.8598 p < .001	.5521 p = .0116	.7334 p < .001	.6417 p = .0023	.7016 p < .001
E	.4127 p = .0705	.8156 p < .001		.8444 p < .001	.6087 p = .0044	.6553 p = .0017	.5288 p = .0165	.66 p = .0015
CT	.8128 p < .001	.8598 p < .001	.8444 p < .001		.7743 p < .001	.7997 p < .001	.6867 p < .001	.8304 p < .001
F	.7302 p < .001	.5521 p = .0116	.6087 p = .0044	.7743 p < .001		.7374 p < .001	.6965 p < .001	.8993 p < .001
G	.6389 p = .0024	.7334 p < .001	.6553 p = .0017	.7997 p < .001	.7374 p < .001		.8333 p < .001	.9381 p < .001
C	.5679 p = .0090	.6417 p = .0023	.5288 p = .0165	.6867 p < .001	.6965 p < .001	.8333 p < .001		.9052 p < .001
FT	.7139 p < .001	.7016 p < .001	.66 p = .0015	.8304 p < .001	.8993 p < .001	.9381 p < .001	.9052 p < .001	

N = nouns; V = verbs; E = events; CT = content total; F = fluency; G = grammaticality; C = grammatical complexity; FT = form total.

4. Discussion

The goals of the present study were, on the one hand, to investigate the production abilities of individuals with aphasia and with cognitive impairment in comparison with those of control participants and, on the other hand, to evaluate the effectiveness of the scoring system of the Catalan version of the *Comprehensive Aphasia Test* when it comes to analysing descriptions. In what follows we discuss the results by group, and then move on to the discussion of the scoring proposed.

4.1. Control participants

The control group performed better than the experimental groups in the content and form parameters. Our results also revealed certain variability among healthy participants, with factors such as education level and age significantly correlating with their overall scores. Yet, the intersubject variability in the control group was observed only in the content parameters, but not in the form parameters (Table 4). This finding is consistent with previous literature (Capiluoto, Wright, and Wagovich 2005; Mackenzie 2000; Mackenzie, Brady, Norrie, and Poedjianto 2007; Smith, Heuerman, Wilson, and Proctor 2003), according to which the variable of age correlates mostly with the semantic and discourse aspects of oral descriptions. For example, Capiluoto et al. (2005) investigated the discourse ability of seventeen younger and seventeen older adults by analysing their responses to four story-telling tasks, and found that, on average, the younger group produced more accurate and relevant event descriptions than the older group. However, the debate is still ongoing, given that other authors have argued that older participants perform better in this type of tasks (see Obler et al. 1994).

In contrast to the controls, no significant correlations were found between age and the overall scores of the two experimental groups. This could be due to the fact that the samples of IWA ($M = 63.3$, $SD = 15.24$) and IWCI ($M = 4$, $SD = 6.97$) were more homogeneous in age than the control group ($M = 50$, $SD = 19.15$). However, the intersubject variability observed in the experimental groups was greater than in the control group in all variables, which mainly reflects their difficulties and the het-

erogeneity of the deficits. Hence, the significant differences found between the groups' performances cannot be solely attributed to between-group differences in age as the distribution of the variability shows. Also, the minimum score of the group of IWA is much lower than the controls'. If age was the cause of the poor performance of IWA, we wouldn't see these differences in the score range of the two groups, given that the control group also includes people of the same age.

4.2. *Individuals with aphasia*

The IWA's performance was lower than the controls' performance in all parameters (Table 4), which corroborates that the tools that we propose are sensitive to the linguistic impairments observed in aphasia despite intersubject variability. Interestingly, the performance of controls and IWA differed significantly in all the variables except for verb score. Previous studies have shown that the production of verbs is often compromised in aphasia which, in turn, has been found to be associated with a decrease of the informativeness of the discourse (see Kuvač Kraljević et al. 2023 for a review). In this study, the difference was not significant, although some participants produced a low number of relevant verbs (e.g. A1), whereas others mentioned all the relevant actions (e.g. A15). Most importantly, the verb score –but not the noun score– of IWA was actually found to be significantly correlated with the event score in line with previous studies (e.g. Hameister and Nickels 2018).

The variability within the group of IWA was very high: the scores of the IWA range from 2 to 28 (the maximum score). This reflects the presence of linguistic difficulties, as well as the heterogeneity of the sample with regards to etiology, aphasia subtype and severity. Other factors such the time post-onset have been shown not to be significantly associated with the results in the group of IWA. In the present study, the results by aphasia subtype or severity were not analysed given that participants were diagnosed using different methods, from informal diagnoses to tests in Spanish. The information on aphasia subtype and severity, hence, is not fully reliable. In addition, the CAT is not intended to discriminate among different types of aphasia (Swinburn et al. 2005).

With regard to intersubject variability in speech production and description tasks, Gordon (2008) analysed the speech of fluent and nonfluent IWA and concluded that the two groups did not differ in the proportion of correct content units. Instead, the factor that did correlate with the informativeness of their discourses was severity. If we take into account the information on the medical records and look at the descriptive results individually, some patterns of response seem to be explained by severity and aphasia type. To illustrate this, the participant with the lowest score, A5, was diagnosed with a nonfluent severe aphasia, while A6, who suffers from a very mild fluent aphasia, performed at ceiling. It is important to note that, despite the fact that the evaluation of spontaneous production can be decisive in the diagnostic process, at-ceiling performance in this subtest does not necessarily rule out an aphasia. Participant A6, for instance, scored low in the object naming subtest, in which she produced several semantic paraphasias (e.g. *plat* 'dish' and *got* 'glass' instead of *forquilla* 'fork'), and in the sentence comprehension task to a lesser extent. Yet, in subtest 19, she provided a detailed and comprehensive description in which she only paused for a moment before producing a couple of names without compromising the overall fluency of the discourse. In contrast, participant A5 performed poorly in this task, and also in several subtests from the CAT-CAT, such as the object and action naming tests that assess the production of names and verbs, respectively. His errors in these tasks were mostly

omissions, which is consistent with his two word description without any verbs.

Therefore, for the IWA, a more detailed analysis of the results can give additional relevant information regarding severity and type of aphasia, and the results in the other subtests (or in other language batteries) should also be taken into account. This is out of the scope of the present study, but we would like to argue that the proposed scoring system allows for that kind of analysis; Kuvač Kraljević et al. (2023) have already studied the descriptions of a group of IWA with differing degrees of severity with the Croatian version of the CAT.

Let us note that two IWA replied partly in Spanish even though the test is in Catalan. This might be taken to be in consonance with the finding that bilingual speakers with aphasia often present deficits in cognitive control functions that translate into involuntary switching or cross-language intrusions (Mooijman, Schoonen, Roelofs, and Ruiter 2021); the two participants also produced Spanish words in other productions tests, mainly in the verbal fluency subtest and the naming tasks. Other than that, participants occasionally produced a word or a sentence in Spanish; for example, one IWA finished her description with a sentence in Spanish (A3: *Que se caen los libros*. ‘The books are falling.’). However, it is difficult to ascertain that these were genuine intrusions, specially on those cases where the switching was occasional, given the fact that it is common for many bilingual speakers of Catalan and Spanish to occasionally mix the two languages. The fact that people without brain damage switch from one language to the other indicates that it cannot be automatically considered an intrusion triggered by a cognitive control deficit. In the particular case of PWA, code-switching may also be enhanced by the fact that most patients are rehabilitated in Spanish in spite of Catalan being their first language.

4.3. Individuals with cognitive impairment

The group of four IWCI also performed significantly worse than the control participants in the content and form total scores. Specifically, differences were observed between the two groups in the noun, fluency and grammaticality scores, and their results were slightly worse on all variables except for grammatical complexity, where they performed at ceiling (Table 5). All of these variables have been argued to be impaired in subjects with different types of cognitive impairment. Campbell, Yáñez Mesía, Docio-Fernandez, and García-Mateo (2021), among other authors, consider fluency as a landmark in Alzheimer’s diagnosis. Fluency is also affected in mild cognitive impairment: Kim, Kim, and Kim (2019) analysed the descriptions of healthy controls and subjects with cognitive impairment (with and without amnesia), and concluded that the latter presented a greater proportion of pauses than the former. People with cognitive impairment often present frequent pauses, articulation and vocabulary difficulties, and repetitiveness, which affect the fluency of the discourse. Hence, the task is sensitive to the difficulties that the IWCI present in production.

4.4. Scoring system

One of the advantages of the scoring system of CAT-CAT is that it includes a wider range of parameters than other tests and this allows us to have a more complete picture of the strengths and weaknesses of individuals. It being a closed-rating system facilitates the comparison among different types of aphasia and even between different types of language disorders.

4.4.1. Patterns of response

In fact, our findings show that each group tested is associated with a specific pattern of response. While control participants performed at ceiling, IWA and IWCI scored significantly lower than them in several form and content parameters. However, the magnitude of the differences between the control subjects and the two experimental groups is different (Table 6), since IWA tended to perform worse in comparison with control participants than IWCI. This was already evident in tables 4 and 5, with the descriptive results by group. Despite the fact that the difference between the two experimental groups was not statistically significant and that the sample size of the IWCI is very small, the descriptive results suggest two different patterns of response. In spite of the heterogeneity of the IWCI group, the four participants reached very high scores on the three grammatical subtests (scores from 8 to 9 out of a total of 9). The participants with cognitive impairment presented difficulties mostly with content: their responses were poor in productivity and discourse efficiency. On the other hand, IWA showed great difficulties both in the content and form parameters.

Furthermore, the performance of IWA entailed greater intersubject variability than that of IWCI: for IWA, the SD overall score was 6.29, while for IWCI the SD overall score was 1.91. The variability in the group of IWCI is mainly found in the content parameters (SD content score = 1.71, SD form score = 0.50), in which participants obtained lower scores as already pointed out. The variability among the participants of the experimental groups seems to reflect their linguistic deficits. If it was just the result of the intrinsic general heterogeneity of their disorders or the variability of the sample, we would observe the variability throughout all items in both groups. Variability is also observed in the control group; it affects the content parameters and is associated with demographic factors like age and educational level.

These results are consistent with previous literature, according to which IWA's deficits in spontaneous production concern especially their naming abilities and, very often, several grammatical aspects, regardless of the fact that the impairments are heterogeneous. IWCI, on the other hand, have difficulties with communicative efficiency, lexical diversity and other features of the discourse, such as coherence, whereas their grammatical abilities are relatively spared (Campbell et al. 2021; Cummings 2019).

4.4.2. Content and form

If we look at the individual results of IWA, other patterns can also be observed. For instance, some participants presented greater difficulties in the content part, even though content and form were found to be strongly correlated: this was the case for participants A1, A16 and A18, who were diagnosed with an anomic aphasia. The opposite pattern is more difficult to observe, as a deficit in form impacts heavily on communication efficiency. It may also be the case, as with subjects A3 and A19, that a person with great deficits in fluency and grammar still manages to name a few relevant events. Yet, their content score is still very low, which further corroborates that discourse efficiency relies on form abilities (Kuvač Kraljević et al. 2023). This is in line with the observation in the present study that the total content score of IWA strongly correlates with the form parameters, especially with grammaticality.

4.4.3. Productivity and informativeness

In this study, productivity and communicative efficiency are evaluated separately, which enables us to investigate whether the production of nouns and verbs equally

influence the informativeness of the aphasic discourse or not. According to our findings, the event score of IWA is strongly associated with the verb score, but not with the noun score. The analysis of these two aspects also reflects different patterns of impairment at an individual level: for example, two participants, A7 and A20, produced a high number of relevant nouns (scores of 7 and 6 out of 8) but their speeches were informatively very poor as they only mentioned 1 and 3 out of the 6 relevant events in the picture. In contrast, participant A15, diagnosed with an anomia aphasia, described 5 out of the 6 relevant events despite only mentioning 3 of the 8 relevant characters and objects pictured. Therefore, his description was less productive yet more informative than the descriptions of the patients mentioned before.

Similarly, not all patients had the same problems with the productivity of nouns and verbs. Participant A1, for example, produced 6 out of the 8 relevant nouns depicted, but only mentioned 2 of the 5 relevant actions; on the other hand, participant A15 showed greater difficulties in producing nouns (3/8) than verbs (5/5). These patterns have already been described in the literature. Several studies have been conducted in order to examine the production of verbs and have shown that IWA often show difficulties with verb production (see Faroqi-Shah and Venneri 2015 for a review).

4.4.4. *Fluency*

As we argued before, the introduction of other content variables and a closed-type system with a maximum score has the advantage of not favouring fluent speakers over nonfluent ones. For instance, the analysis of the results of the IWA showed a strong correlation between fluency and noun scores, which is consistent with the idea that fluent speakers tend to produce a larger number of lexical items, including nouns. However, their noun score was not associated with the informativeness of their discourse, as shown by the fact that their event score did not correlate with the noun score. A scoring system that relies heavily on the word count is less sensitive to the other problems that some IWA might present. As already mentioned in the introduction, we believe that the inclusion of the number of relevant content units produced as a criterion to evaluate is relevant for clinical purposes in consonance with other versions of the CAT (Jensen et al. 2024; Martínez-Ferreiro, Arslan, et al. 2024; Visch-Brink et al. 2014).

Fluency is also associated with other form parameters, especially with grammaticality, which, in turn, is strongly correlated with the verb count. This is consistent with the observation that individuals with nonfluent aphasia and agrammatism tend to have problems with verbal projections (among others, Bastiaanse 2008; Friedmann and Grodzinsky 1997; Gavarró and Martínez-Ferreiro 2011; Thompson 2003).

In conclusion, the scoring system that we propose captures heterogeneity in a way that can be used to diagnose aphasia, to investigate and compare the production abilities of individuals with different types of aphasia and severity, and to get an exhaustive picture of the spared and impaired abilities of each individual. It can also be used with other populations, such as individuals with cognitive impairment of different etiologies.

5. Prospects and conclusions

Here we have focused on the quantitative results of the study, but the qualitative analysis of the responses can also be useful in the clinical setting. As an example,

participant A7 produced semantic paraphasias and circumlocutions (e.g. *Els peixos de un... la piscina, bueno, de un... una piscina que es para los peces*. ‘The fish in a... pool, well, in a... a pool for the fish.’), while participant A13 produced neologisms (*La senyora que no porta res als parapeus i està veiruda*. ‘The woman who doesn’t wear anything on her *parafeet* and she’s *veiruda*.’), and participant A4, verbal perseverances (*I la música, i el mando, el mando, el mando, el mando, el mando, bueno, i la taula, els mandos i els fotos de, els fotos de, els fotos de el àlbum, l’àlbum*. ‘And the music, and the remote, the remote, the remote, the remote, well, and the table, the remotes and the pictures of, the pictures of, the pictures of the album, the album.’). On the other hand, participant A5, with a nonfluent severe aphasia, only uttered two nouns as mentioned before, and A19 used gestures and onomatopoeias to compensate for his anomia.

Like with the participants with aphasia, the qualitative analysis can provide useful information of IWCI. Regarding error types, even though this falls outside the scope of the present study, different error patterns were found. Despite their problems with naming, the participants did not produce neologisms or paraphasias, instead, they produced other error types and their discourse contained utterances aimed to compensate for their naming difficulties. For example, participants C3 and C4, used generic words such as *algo* ‘something’, *això* ‘this’, *allò* ‘that’, *aquí* ‘here’ or *aquest* ‘this’. The result is a rather unspecific description, like this description by C4: *Però en enfilat-se mira tot el que ha fet. I la senyora no se n’ha enterat. Està fent un d’allò fort*. ‘But after getting on, look at all it has done. And the woman hasn’t even realised it. It’s doing a very loud thing.’). This can be extended to other aspects that are not currently included in the scoring system; for example, individuals with cognitive impairment (see Cummings 2019 for discussion) mostly present pragmatic impairment in narrative and description tasks that affects, besides the discourse efficiency (poor description of the main events, lexical and semantic deficits), the discourse organization and cohesion.

We argue that the parameters that the Catalan version of the CAT includes in subtest 19 suffice to fulfill its goals, that is, on the one hand, to contribute to the diagnosis of aphasia and, on the other hand, to provide a detailed profile of the assessed participant. Moreover, our findings show that it can be useful to detect speech deficits in other populations such as people with cognitive impairment. The variables included in the scoring system preserve the structure of the original CAT (form/content), but are more exhaustive, as they involve items to evaluate productivity and discourse efficiency. This modification, together with the introduction of a maximum score, avoids favouring more talkative and fluent speakers. In fact, other versions of the CAT have already incorporated similar changes (Martínez-Ferreiro, Arslan, et al. 2024) for similar reasons. Yet, we also propose scoring the number of relevant events produced as a measure of informativeness. Communicative efficiency is useful to detect discourse deficits in more productive speakers. Also, it introduces a functional measure to evaluate communication. This is also relevant in rehabilitation. In fact, microlinguistic abilities like, for example, the production of verbs, have been shown not to be enough to rehabilitate macrolinguistic discourse aspects, such as communicative efficiency (Pallickal and Hema 2020). Hence, the introduction of a measure to evaluate the informativeness of the discourse can be useful in the clinical setting not only for diagnostic purposes.

To sum up, the two experimental groups in our study performed differently from the control in both content and form aspects. Therefore, our findings show that the oral picture description of the CAT-CAT is sensitive to the language impairment of subjects with aphasia and with cognitive impairment. The rating system put forward allows us to quantitatively and qualitatively analyse different patterns of impairment,

since it includes variables to evaluate separately content and discourse efficiency on the one hand, and grammatical features on the other hand. In the present study, the IWA were consistently worse in fluency, grammaticality and grammatical complexity than the IWCI, which indicated a different pattern of impairment between the two groups. Future research with a larger sample of IWCI is needed to confirm what at this point is only suggestive. Even though the CAT (Swinburn et al. 2005, 2023) does not aim to discriminate among different types of aphasia, we argue that these preliminary results show that more research in this direction could be undertaken in the future, together with additional measures to evaluate other aspects such as discourse organization. Our findings show that clinicians can use quantitative and qualitative results to draw a general picture of the type of impairment that the patient presents and, in consequence, design a personalised treatment.

Acknowledgements

The authors wish to thank the participants to the experiments reported, as well as the editor and the anonymous reviewers for their comments and suggestions. Any remaining errors are our own.

Disclosure statement

The authors declare no conflict of interest.

Funding

The authors acknowledge the funding of project PID2022-138413NB-100 by the *Ministerio de Ciencia, Innovación y Universidades del Gobierno de España*.

References

- Abou El-Ella, M., Alloush, T., El-Shobary, A., Hafez, N. E.-D., El-Halim, A. A., & El-Rouby, I. (2013). Modification and standardisation of Arabic version of the Comprehensive Aphasia Test. *Aphasiology*, 27(5), 599-614.
- Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia*, 50(8), 1787-1793.
- Andreetta, S., & Marini, A. (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. *Aphasiology*, 29(6), 705-723.
- Armstrong, E. (2000). Aphasic discourse analysis: the story so far. *Aphasiology*, 14(9), 875-892.
- Bastiaanse, R. (2008). Production of verbs in base position by Dutch agrammatic speakers: inflection versus finiteness. *Journal of Neurolinguistics*, 21(2), 104-119.
- Campbell, E. L., Yáñez Mesía, R., Docio-Fernandez, L., & García-Mateo, C. (2021). Paralinguistic and linguistic fluency features for Alzheimer disease detection. *Computer Speech and Language*, 68(12), 101198.
- Capiluoto, G., Wright, H. H., & Wagovich, S. A. (2005). CIU and main event analyses of the structured discourse of older and younger adults. *Journal of Communication Disorders*, 37, 431-444.

- Christiansen, J. (1995). Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language*, 51(2), 291-317.
- Cummings, L. (2019). Describing the cookie theft picture: sources of breakdown in Alzheimer’s dementia. *Pragmatics and Society*, 10(2), 151-174.
- Dalton, S. G., & Richardson, J. D. (2019). A large-scale comparison of main concept production between persons with aphasia and persons without brain injury. *American Journal of Speech-Language Pathology*, 28(1), 293-320.
- Dietz, A., & Boyle, M. (2018). Discourse measurement in aphasia research: have we reached the tipping point? *Aphasiology*, 32(4), 459-464.
- Farooqi-Shah, Y., & Venneri, A. (2015). Production of verb tense in agrammatic aphasia: a meta-analysis and further data. *Behavioural Neurology*, 983870.
- Folstein, M., Folstein, S., & McHugh, P. (1975). *Mini-Mental State Examination (MMSE)*. APA PsycTests.
- Forbes-McKay, K., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer’s disease with a picture description task. *Neurological Sciences*, 26, 243-254.
- Friedmann, N., & Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: pruning the syntactic tree. *Brain and Language*, 56(3), 397-425.
- Fyndanis, V., Lind, M., Varlokosta, S., Kambanaros, M., Soroli, E., Ceder, K., ... Howard, D. (2017). Cross-linguistic adaptations of The Comprehensive Aphasia Test: Challenges and solutions. *Clinical Linguistics and Phonetics*, 31(7-9), 697-710.
- Gavarró, A. (2017). A sentence repetition task for Catalan-speaking typically developing children and children with Specific Language Impairment. *Frontiers in Psychology*, 8, 1-13.
- Gavarró, A., & Martínez-Ferreiro, S. (2011). Cartography and agrammatic syntactic production in Ibero-Romance. In J. Berns, H. Jacobs, & T. Scheer (Eds.), *Romance Languages and Linguistic Theory 2009* (p. 115-132). John Benjamins Publishing.
- Gordon, J. K. (2008). Measuring the lexical semantics of picture description in aphasia. *Aphasiology*, 22(7-8), 839-852.
- Groves-Wright, K., Neils-Strunjas, J., Burnett, R., & O’Neill, M. J. (2004). A comparison of verbal and written language in Alzheimer’s disease. *Journal of Communication Disorders*, 37(2), 109-130.
- Hameister, I., & Nickels, L. (2018). The cat in the tree - using picture descriptions to inform our understanding of conceptualisation in aphasia. *Language, Cognition and Neuroscience*, 33(10), 1296-1314.
- Hazamy, A. A., & Obermeyer, J. (2020). Evaluating informative content and global coherence in fluent and non-fluent aphasia. *International Journal of Communication Disorders*, 55(1), 110-120.
- Hoover, E., DeDe, G., & Maas, E. (2021). A randomized controlled trial of the effects of group conversation treatment on monologic discourse in aphasia. *Journal of Speech, Language, and Hearing Research*, 64, 4861-4875.
- Howard, D., Swinburn, K., & Porter, G. (2010a). The CAT is now out: a response to the commentaries. *Aphasiology*, 21(1), 94-98.
- Howard, D., Swinburn, K., & Porter, G. (2010b). Putting the CAT out: what the Comprehensive Aphasia Test has to offer. *Aphasiology*, 24(1), 56-74.
- Ivanova, M. V., & Hallowell, B. (2013). A tutorial on aphasia test development in any language: key substantive and psychometric considerations. *Aphasiology*, 27(8), 891-920.
- Ivanova, M. V., Kuptsova, S. V., & Dronkers, N. F. (2017). A comparison of two working memory tasks in aphasia. *Aphasiology*, 31(3), 265-281.
- Jensen, B. U., Norvik, M. I., & Gram Simonsen, H. (2024). Statistics and psychometrics for the CAT-N: documenting the Comprehensive Aphasia Test for Norwegian. *Aphasiology*, 28(3), 412-439.
- Kim, B. S., Kim, Y. B., & Kim, H. (2019). Discourse measures to differentiate between mild cognitive impairment and healthy aging. *Frontiers in Aging Neuroscience*, 11(221).

- Kuvač Kraljević, J., Matić, A., & Lice, K. (2020). Putting the CAT-HR out: key properties and specificities. *Aphasiology*, 34(7), 820-839.
- Kuvač Kraljević, J., Matić Škorić, A., & Lice, K. (2023). Main concepts in the spoken discourse of persons with aphasia: analysis of a propositional and linguistic level. *Languages*, 8(2), 120.
- Lerman, A., Pazuelo, L., Kizner, L., Borodkin, K., & Goral, M. (2019). Language mixing patterns in a bilingual individual with non-fluent aphasia. *Aphasiology*, 33(9), 1137-1153.
- Linnik, A., Bastiaanse, R., & Höhle, B. (2016). Discourse production in aphasia: a current review of theoretical and methodological challenges. *Aphasiology*, 30(7), 765-800.
- Lluent Vallet, R., Peña-Casanova, J., & Böhm, P. (2002). Versión catalana de los subtests verbales del Test Barcelona. *Neurología*, 17(3), 124-131.
- Mackenzie, C. (2000). Adult spoken discourse: the influences of age and education. *International Journal of Language and Communication Disorders*, 35(2), 269-285.
- Mackenzie, C., Brady, M., Norrie, J., & Poedjianto, N. (2007). Picture description in neurologically normal adults: concepts and topic coherence. *Aphasiology*, 213/4(340-354).
- Martínez-Ferreiro, S., Arslan, S., Fyndanis, V., Howard, D., Kuvač Kraljević, J., Matić Škorić, A., ... Soroli, E. (2024). Guidelines and recommendations for cross-linguistic aphasia assessment: a review of 10 years of comprehensive aphasia test adaptations. *Aphasiology*, 1-25.
- Martínez-Ferreiro, S., Quique, Y., Rodríguez, V., & Orellana, C. M. (2024). Linguistic and cultural properties of the Spanish adaptation of the CAT (SP-CAT): pilot results from neurotypical subjects. *Aphasiology*, 1-24.
- Maviş, İ., Tunçer, M., Selvi-Balo, S., Tokaç, S. D., & Özdemir, Ş. (2021). The adaptation process of the Comprehensive Aphasia Test into CAT-Turkish: psycholinguistic and clinical considerations. *Aphasiology*.
- Mooijman, S., Schoonen, R., Roelofs, A., & Ruiter, M. B. (2021). Executive control in bilingual aphasia: a systematic review. *Bilingualism: Language and Cognition*, 1-16.
- Obler, L., Au, R., Kugler, J., Melvold, J., Tocco, M., & Albert, M. (1994). Intersubject variability in adult normal discourse. In L. Obler, S. De Santi, & J. Ehrlichs (Eds.), *Discourse analysis and applications: Studies in adult clinical populations* (p. 15-27). Lawrence Erlbaum Associates.
- Özdemir, Ş., Maviş, İ., & Tunçer, A. (2022). The validity and reliability of the language battery in Comprehensive Aphasia Test-Turkish (CAT-TR). *Journal of Psycholinguistic Research*, 51(4), 789-802.
- Pallickal, M., & Hema, N. (2020). Discourse in Wernicke's aphasia. *Aphasiology*, 34(9), 1138-1163.
- Paradis, M., & Libben, G. (1987). *The assessment of bilingual aphasia*. Psychology Press.
- Peña-Casanova, J., & Sánchez Benavides, G. (2019). *Test Barcelona-2*. Teià: Test Barcelona Services.
- R Core Team. (2021). R: A language and environment for statistical computing [Computer software manual].
- Salmons, I., Rofes, A., & Gavarró, A. (2021). *Prova integral d'afàsia. Llibre d'ítems*. Bellaterra: Servei de Publicacions de la UAB. Retrieved from <https://ddd.uab.cat/record/250143>
- Schönberger, E., Heim, S., Meffert, E., Pieperhoff, P., da Costa Avelar, P., Huber, W., ... Grande, M. (2014). The neural correlates of agrammatism: evidence from aphasic and healthy speakers performing an overt description tasks. *Frontiers in Psychology*, 5(246).
- Smith, R., Heuerman, M., Wilson, B. M., & Proctor, A. (2003). Analysis of normal discourse patterns. *Brain and Cognition*, 53, 368-371.
- Swinburn, K., Beeke, S., Cruice, M., Smith, L., Pearce Willis, E., Ledingham, K., ... McVicker, S. (2018). A concise patient reported outcome measure for people with aphasia: the Aphasia Impact Questionnaire 21. *Aphasiology*, 22(9), 1035-1060.
- Swinburn, K., Porter, G., & Howard, D. (2005). *Comprehensive Aphasia Test*. Hove: Psychology Press.
- Swinburn, K., Porter, G., & Howard, D. (2023). *Comprehensive Aphasia Test (2nd edition)*.

- Routledge.
- Thompson, C. (2003). Unaccusative verb production in agrammatic aphasia: the argument structure complexity hypothesis. *Journal of Neurolinguistics*, 16(2-3), 151-167.
- Vandenborre, D., Visch-Brink, E. G., van Dun, K., Verhoeven, J., & Mariën, P. (2018). Oral and written picture description in individuals with aphasia. *International Journal of Language and Communication Disorders*, 53(2), 294-307.
- Visch-Brink, E., Vandenborre, D., de Smet, D., & Mariën, P. (2014). *Comprehensive Aphasia Test - Nederlandse bewerking - Handleiding*. Pearson.
- Wright, H. H., Silverman, S. W., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, 17(5), 443-452.
- Zakariás, L., & Lukács, Á. (2022). The Comprehensive Aphasia Test-Hungarian: adaptation and psychometric properties. *Aphasiology*, 36(9), 1127-1145.
- Zakieldine, H. M., Turk, M. A., Hashem, R. F., & Shafik, M. A. (2020). The myth and truth of a comprehensive stroke scale. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, 56(8).