Effects of modality, administration and stimuli on picture descriptions in adults

By

Anke Lingscheid

Dissertation

Submitted to the College of Arts, Humanities and Social Sciences

in partial fulfilment of the

Requirements for the Degree of

Doctor of Philosophy

The University of Edinburgh

August 2022

Abstract

Background

Picture descriptions are a commonly used tool in the diagnostic process of aphasias, including vascular as well as neurodegenerative disorders affecting language production. However, this widespread clinical use contrasts with the relative scarcity of empirical data exploring the role of different variables which can influence the results, particularly in cases in which the original procedure has been modified. This thesis addresses three areas in which such modifications need a much stronger empirical basis.

Firstly, (1) the traditional default approach has been an oral picture description. However, a substantial group of patients, particularly those with neurodegenerative disorders affecting motor functions, cannot produce oral language. In such cases, an assessment of written language could offer a valuable alternative. Such a procedure would require, however, a systematic comparison between linguistic material collected in the spoken and written description.

Another little-explored field is (2) the difference between traditional in-person versus remote testing. While remote testing has been possible for several decades, in the recent pandemic it became often the only available option. Research has been conducted on remote testing, but only for video-call testing settings. Automated online testing has not yet been assessed.

Finally, in (3) the increasingly globalised world diagnostic material is used in different countries with diverse cultures and routines. Linguistic stimuli are translated (although not always adapted) but this is rarely the case for picture material. Therefore, the picture stimulus itself might be victim to misinterpretation if the patient is not culturally familiar with the depicted scene. However, the influence of the cultural familiarity of the picture material has not been addressed in research yet.

Aim

The present thesis addressed the current gaps in research focussing on these three aspects: (1) the influence of production modality (spoken vs. written [handwritten vs. typed], (2) testing modality (in-person vs. video-call vs. automated online) and (3) picture stimulus (North American [traditional] vs. Indian) on the picture description task using a mix of literature reviews and empirical data.

Method

A total of 100 healthy participants were tested (1) in two different production modalities, (2) in three different testing modalities and (3) with two different picture stimuli. The picture description was complemented by seven additional tests assessing naming, grammar, possible underlying dyslexia, concept formation, visual-spatial abilities and intelligence. The collected data were interpreted using an extended list of linguistic symptoms, statistical comparisons and linear models.

Results

(1) Spoken picture descriptions were significantly longer, containing more words and sentences, whereas written descriptions were more concise, syntactically complex and lexically diverse. Unlike aphasic patients, healthy participants very rarely produced phonetic, syntactic or semantic errors. Isolated language functions and other cognitive abilities did not have a major influence on the picture description task. (2) The testing modality influenced the linguistic performance of the participants. Differences were more pronounced in the written task and could be traced back to differences between handwriting and typing. (3) The non-familiar scene produced a higher lexical diversity, but also a less complex syntax.

Discussion

For clinical purposes but also research, (1) written picture descriptions should be considered for diagnosis as it is a more cost-effective method that can lead to conclusions of comparable relevance. Performance on tests of isolated linguistic domains and cognitive

abilities do not have a significant influence on picture description tasks in healthy individuals. Therefore, picture descriptions cannot be used as a substitute for the assessment of isolated linguistic functions and cognitive abilities, or vice versa. (2) Equally, picture descriptions cannot be easily replaced by other tests, as they contain other types of information of considerable relevance. Remote testing is a useful tool that can be used in the diagnostic process, however, there are caveats that need to be considered before testing and interpretation. (3) A culturally non-familiar picture stimulus can lead to misinterpretation of the presented scene which might lead to false diagnoses. Therefore, the cultural origin of the stimulus needs to be taken into account when interpreting the results.

Conclusion

Overall, all three examined modifications can be clinically feasible and useful. However, they cannot and should not be understood as identical to traditional methods, as they might have different limitations, but also offer different opportunities.

Keywords: Picture Description Task, Neurodegenerative Disorders, Spoken vs. Written Language, Handwriting vs. Typing, Influence of Cultural Differences

Lay Summary

Background

Every year stroke and degenerative conditions of the brain, such as Alzheimer's or Parkinson's, affect millions of people worldwide. These conditions can lead to problems with speaking and writing. In the clinic patients often have to complete picture description tasks as part of the diagnostic procedure. Usually, the patient tells the doctor what is happening in the picture. This stands however in contrast with current research that raises the importance of other factors that could influence the diagnosis. The present thesis discusses three areas that could influence the diagnosis of people with degenerative conditions: (1) Spoken versus written picture descriptions; (2) in-person, versus video call versus online testing and (3) the influence of our cultural background.

First, (1) the traditional approach is to obtain the picture descriptions in a spoken form. However, people with degenerative conditions of the brain might have problems with this task as they often have trouble speaking. However, their written language is often better than their spoken language. In this case, a suitable alternative would be to get a written picture description from the patient. However, there is currently little data about spoken and written picture descriptions of patients which makes a systematic comparison of spoken and written picture descriptions necessary.

The Covid-19 pandemic has also drawn attention to another little-explored field: (2) differences between traditional in-person testing and remote testing. Remote testing, for example over video-call apps, has been technically possible for several decades and has been the only option in the recent pandemic. Another possibility would be automated online testing in which patients would complete a test by themselves that would be analysed later by a professional. However, so far research has only been conducted for video-call testing settings. Therefore, an assessment of automated online testing is necessary.

Finally, (3) increased globalisation has led to diagnostic material which was initially developed for and in, for example, North America, to be used in different countries, such as India, with different cultures and different routines. While stimuli concerning language have been translated and also adapted to new countries, this was very rarely the case for exercises using pictures. Consequently, the picture might be misinterpreted. However, research has not yet addressed the influence of cultural background on picture description tasks.

Aim

The present thesis addressed the current gaps in research by focussing on three aspects: (1) the influence of production modality (spoken vs. written [handwritten vs. typed], (2) testing modality (in-person vs. video-call vs. automated online) and (3) picture stimulus (North American [traditional] vs. Indian) on the picture description task using a mix of methods by comparing and interpreting previous research results as well as performing novel research and statistical tests.

Method

A total of 100 participants were tested. Their performance in describing (3) two different pictures (the traditional Cookie Theft Picture and the Indian Street Scene), both (1) once spoken and once written was tested. 30 participants took part in in-person testing settings, (3) 30 participants completed the testing in a video-call setting and 40 participants participated in the online testing condition (2). All participants completed furthermore additional tests that tested their naming and grammar abilities, possible underlying reading and writing problems, the way they can recognise patterns, their hand and eye coordination as well as their intelligence.

Results

(1) Spoken picture descriptions were longer; they contained more words and sentences. Written picture descriptions were shorter but consisted of more content. The sentence structure was also more complex, using for example more

subclauses. Furthermore, the participants used more different words in their written picture descriptions compared to their spoken picture descriptions. It was also observed that the healthy participants made fewer pronunciation or spelling errors such as mixing up letters. Healthy participants were also less likely to have grammar errors and they are less likely to mix up words, compared with previous studies of aphasic patients. (2) It also made a difference whether participants were tested in person, via video-call or online. The main difference was that participants who typed their picture descriptions had fewer errors in their writing as their computers might have corrected errors automatically. Finally, (3) the participants had more difficulties describing the non-familiar Indian Street Scene. As they needed to use more different words their sentences became less complex with, for example, fewer subclauses.

Discussion

In clinical work but also in research, (1) the written picture description task should be considered as a diagnostic method as it is cheaper but can still lead to the same diagnosis. It is also important that the results of single tests, such as a naming test, for example, cannot give information about the entire language abilities of a person as language is more complex. Vice versa, picture descriptions cannot be used as a substitution for tests testing naming abilities only, for example. (2) When diagnosing a patient, remote testing can be useful. However, the diagnostician has to consider the circumstances in which the patient was diagnosed. If the patient was diagnosed using video-call testing, for example, the written responses may have been autocorrected. So, it might not be entirely clear whether a patient needs support with spelling, for example. Furthermore, (3) if a picture that shows a scene that is not familiar to the patient is used in the diagnostic process, the patient might misinterpret the scene and give unintentionally wrong information that might lead to a wrong diagnosis. Therefore, when diagnosing patients, the diagnostician has to take into account the potential match or mismatch between the picture and the cultural background of the patient if possible.

Conclusion

Overall, the thesis shed light on three areas that could influence the diagnosis of people with degenerative conditions. Written instead of spoken picture descriptions might be an alternative when it is hard for a patient to speak. Video-call testing might be a preferred method in times in which patients might have to isolate. However, the diagnostician has to be aware that technical equipment might mask errors in the patient's answer. Finally, pictures might have to be chosen according to the cultural background of the patient to ensure that the picture will be described without possible errors due to a misunderstanding of the depicted scene. All presented modifications in this thesis can be helpful in clinical settings and research. However, they should not be understood as equal to traditional methods. The presented modification should be seen more as opportunities that come with their own opportunities as well as limitations.

Acknowledgements

First and foremost, I am extremely grateful to my supervisors, Dr Thomas Bak and Dr Paul Hoffman for their invaluable advice, continuous support and patience during my PhD study. Their immense knowledge and plentiful experience have encouraged me all the time in my academic research and daily life.

I would also like to thank Dr Pauranik for providing me with picture material from India and my friends from the Speech and Language Therapy Summer School, Marco Pepito Simoni, Laetitia Bayard, Robyn Alexandra and Margot Schmidt for their detailed description of the use of picture material in the diagnostic process for adults in Italy, France, Spain, Belgium, Netherlands, Luxembourg and Great Britain. Without their help, this thesis would be lacking important cultural diversity.

I would like to thank all the members in the Psychology Department. Especially Javiera Alforo Chat and the members of the PhD Psychology Committee Lasse Schaefer and Samuel Henry as well as my very missed office desk partner Melissa Thye. It is their kind help and support that have made my study and life in Edinburgh a wonderful time.

Furthermore, I would like to thank the Black and Minority Ethnic Campaign of the University of Edinburgh and Leyla Gadema for advising me on inclusive language. I am also deeply grateful to my friends Shin Huang, Laura Alonso, Kostas Tsampourakis, Ruben Verdoodt, Helen Jones and Samuel Ratcliffe for taking part in countless hours of pilot testing. I would also like to express my sincere gratitude to all my volunteer participants, without whom this thesis would have not been possible. I spend some very fun hours doing all the in-person and online testing sessions with you. I would like to offer special thanks to the Funds for Women Graduates to supported me financially with a grant.

Finally, I would like to express my gratitude to my parents, Klaudia and Arno Lingscheid, my siblings Meike and René Lingscheid and last but not least my grandmas Katharina Doppelfeld and Helene Lingscheid. Without their tremendous understanding and encouragement in the past few years, it would have been impossible for me to complete my study.

Danke!

Thank you!

Merci!

Grazie!

Gracias!

Dank je wel!

谢谢!

Σας ευχαριστώ!

धन्यवाद!

Table of Contents

Page

List of Tables

List of Figures

1. Introduction

This chapter will give an overview of the use of picture description tasks in the diagnosis of aphasia and neurodegenerative diseases.

To understand why and how picture descriptions are used in diagnosis and research, first, the history of the diagnostis of aphasia will be presented and discussed. Furthermore, the different syndromes of aphasia and their prevalent pathology will be described as the diagnosis of neurodegenerative diseases such as dementia are built upon this foundation.

Following the foundations of picture description in aphasia, the use of picture description tasks in the diagnosis and research of neurodegenerative diseases will be presented and analysed. The use of picture description tasks in different neurodegenerative disorders such as dementia and motor disorders will be discussed.

Finally, difficulties of conducting picture description tasks in the traditional oral administration and opportunities for improvement will be described.

1.1. Brief history of picture description as a tool in the assessment of aphasia

Picture description tasks are a method to elicit language with a target stimulus. Although picture description tasks seem to many current practitioners as one of the best-established diagnostic procedures, they were introduced relatively late. First attempts at assessing aphasia were made by individual physicians, reporting methods for testing or treating single cases. Various relevant techniques were described, some of which are still in use, for example, the Three-Paper Test of Marie, in which the patients have to perform three unrelated tasks with three different-sized papers. They are instructed to crumple the largest one, give the middle-sized one to the diagnostician and put the smallest one in their pocket (Marie, 1906). The first popular standardised testing procedure was a description of language evaluation presented by Head (1926) in Aphasia and Kindred Disorders (Head, 1926), including naming and recognition tasks, a reading test as well as language comprehension and imitation tasks.

Building upon Head's (1926) work, Weisenburg and McBride (1934/1964) designed tests for the evaluation of aphasic language, representing the first attempt at a psychometric aphasia battery. It was the first of its kind to use standardised procedures and to compare results from healthy individuals and people with aphasia.

After WWII new test batteries were introduced, varying in comprehensiveness but sharing the same approach. Eisenson's (1954) Examining for Aphasia, Wepman & Jones's, (1961) The Language Modalities Test for Aphasia, Schuell's (1965) Minnesota Test for the Differential Diagnosis of Aphasia, Spreen and Benton's (1968) Neurosensory Center Comprehensive Examination for Aphasia, Sarno's (1969) Functional Communication Profile (FCP), Porch's (1967) Porch Index of Communicative Ability and Holland's (1980) Assessment of Communicative Activities Relevant to Daily Living all test naming, writing, reading, repetition and matching except for the FCP relying on observations.

It was not until the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1972) that a picture description task was introduced as a subtest for the assessment of the connected spoken language of individuals with aphasia (see Figure 1 for Cookie theft Picture). The Western Aphasia Battery (WAB; Kertesz 1979, 1982) was introduced later, with a shorter language test (see Figure 1 for a picture of the Picnic Scene) modelled on the BDAE.

From then on picture description became a much-used standard in the diagnosis of aphasia but also in the research of aphasia symptoms and syndromes as well as treatment studies.

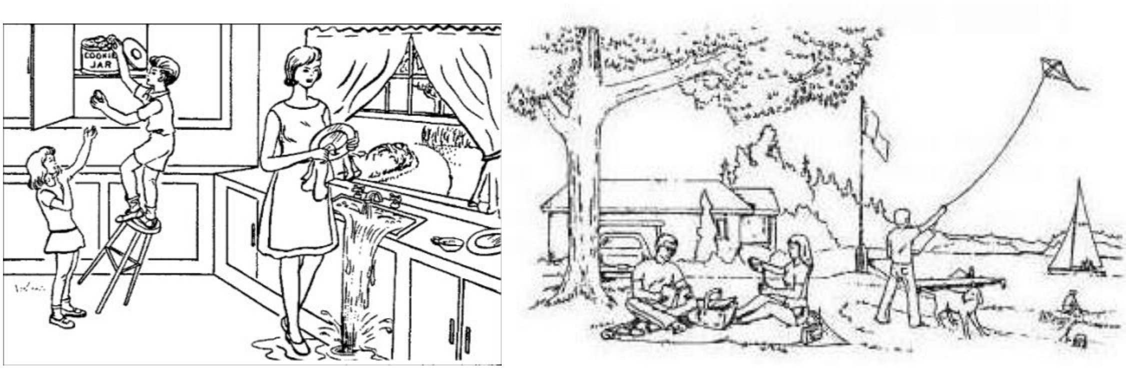


Figure 1: Boston Cookie Theft Picture (left) and Picnic Scene (right)

To systematically quantify the characteristics of speech that are elicited in a picture description task, several standardised metrics of speech properties have been developed. Yorkston and Beukelman (1980) took the first measures using the Cookie Theft Picture to present means and standard deviations for syllables per minute, content units per minute and content units in total. They were introduced as methods for quantifying connected speech samples systematically and efficiently specially to observe how patients recover from mild to moderate aphasia. An inverse relationship between the amount of information that was conveyed, and the severity of aphasia was later confirmed by Craig et al. (1993) replicating the study of Yorkston and Beukelman (1980). In general people with severe aphasia produce short and relatively poor discourse while people with mild aphasia produce a discourse that is longer and better in quality (Ulatowska et al., 2003).

Wright et al. (2003) added measures for Type-Token-Ratio (TTR), the number of words and a variable D (Malvern & Richards, 2002) to differentiate fluent from non-fluent aphasia using the Picnic Scene from the WAB. This was backed up by Gordon (2008) stating that TTR is only influenced by the grade of severity in people with fluent aphasia. Gordon (2008) also added that information units can reflect aphasia severity though cannot differentiate between fluent and non-fluent aphasia. In an earlier study, using the same picture material, Gordon (2006) collected data to identify quantitative variables most predictive for agrammatism by using Quantitative Production Analysis (QPA) but found that QPA does not reflect agrammatism. These examples show that although there are instructions on how to assess and evaluate picture descriptions in the BDAE and WAB, there still seems to be no universally agreed procedure for the assessment and evaluation of picture description tasks.

Yet, there seems to have formed a habit that the Cookie Theft Picture and the picture set by Brookshire and Nicholas (1994) are generally used in research. The picture set of Brookshire and Nicholas (1994) consists of the cat and flood rescue picture as well as the umbrella and window picture sequences and was first used in a study analysing test-retest stability of connected speech. They were later included in the talkbank/aphasiabank. Still, the use of the previously mentioned pictures is not a reasoned decision. None of the studies presents a theoretical or scientifically reasoned justification for the picture choice. It seems that pictures were used either because they were already in testing material or on the other hand already used in previous studies.

The **Cookie Theft Picture** is very popular as a baseline and follow-up test in treatment studies (for example Edwards & Tucker, 2006; Gordon, 2007) as well as in basic (fundamental) research such as differences of certain characteristics in aphasia (Yorkston & Beukelman, 1980; Slobin, 1991; Ardila & Rosselli, 1993; Mortensen, 2005; Patterson et al., Hodges, 2006) or influences of aphasia on bilingualism and vice versa (Abuom & Bastiaanse, 2012; Kambanaros, 2009). Brookshire and Nicholson (1994) analysed the test-retest stability of connected speech samples. They developed two single pictures ("***Cat Rescue***" and "***Birthday Cake***") and two picture sequences ("***The Argument***" and "***Directions***"; for pictures see Figures 2, 3 and 4). Nicholas and Brookshire (1994) concluded that speech samples should be taken from a variety of elicitation stimuli to be more representative of a person's everyday connected speech. Their picture material was later used in studies to compare the proportion of main events used in descriptions of people with and without aphasia (Capilouto et al., 2006), treatment baseline testing (Boyle, 2004; Boyle, 2015; Wambaugh & Ferguson, 2007; Antonucci, 2008; Hoover, Caplan, & Waters, 2014) and for another study measuring the variability of repeated sampling (Cameron et al., 2010) reporting greater variability than initially described by Brookshire and Nicholas (1994). Cameron et al. (2010) focussed more on individual variability whereas Brookshire and Nicholas (1994) reported mostly temporal stability for all measures but did not publish individual data. Therefore, the findings of Cameron et al. (2010) would be more important for clinical purposes as they identified the importance of individual baseline testing prior to treatment initiation to document changes in the patient’s performance.

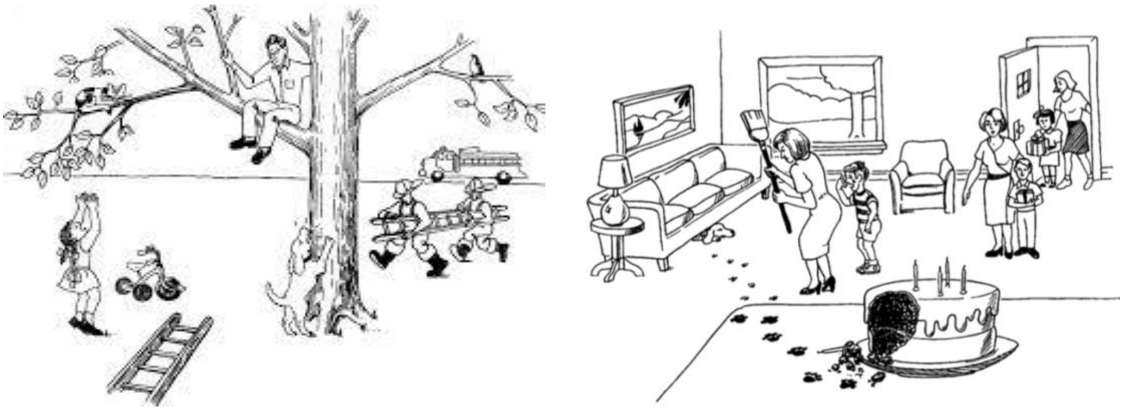


Figure 2: Cat Rescue Scene (left) and Birthday Cake (right)

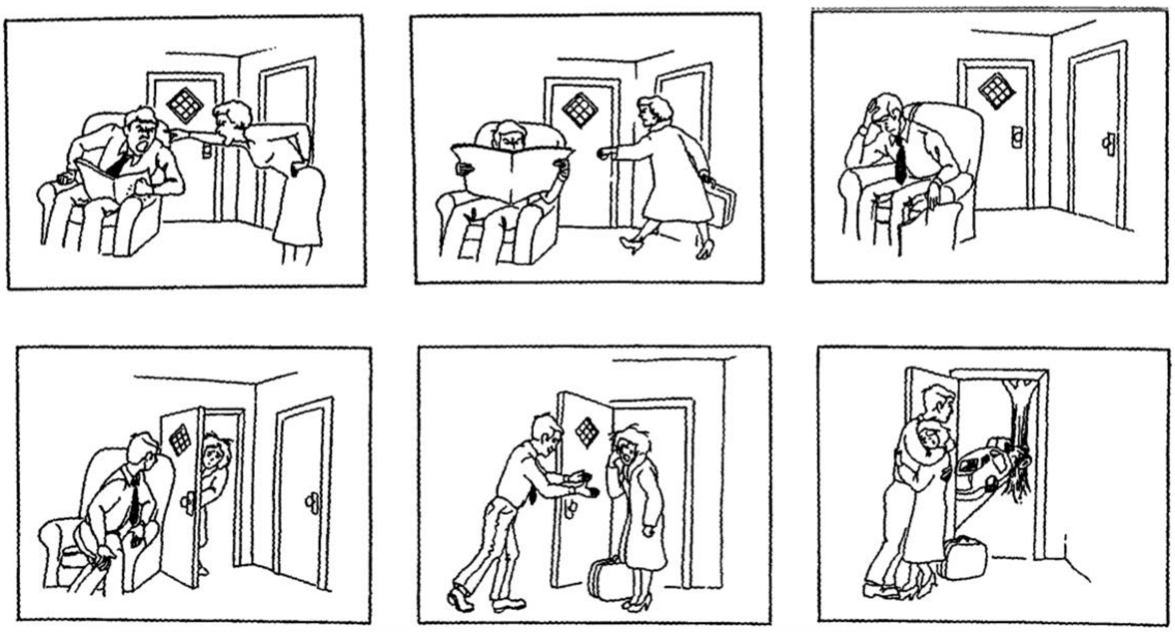


Figure 3: The Argument

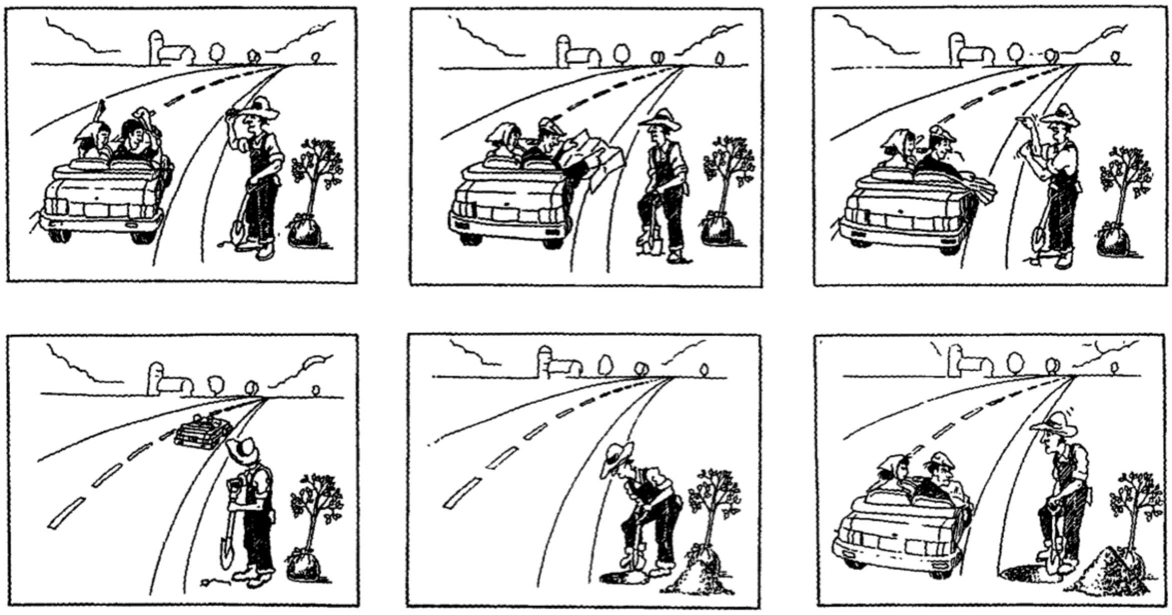


Figure 4: Directions

Since 2011 the "***Cat Rescue***" picture from Brookshire and Nicholas (1994) is included in the picture stimuli of the talkbank/aphasia bank protocol and used for standardised testing for aphasia research. The talkbank protocol also includes a photograph by Annie Wells of an emergency rescue of a girl from flood water (Rubin & Newton, 2001; see Figure 5) as well as the picture sequences "***Broken Window***" (see Figure 6) and "***Umbrella***" (see Figure 7) (MacWhinney et al., 2011). Data from the talkbank/aphasia bank was among others used for language comparisons (Fergadiotis et al., 2011; Fergadiotis et al., 2013), comparing different test forms for aphasia testing (Sung et al., 2016) and developing a checklist for propositions to make testing easier and more comparable (Hudspeth Dalton & Richardson, 2019).

Figure 5: Rescue from Flood by Annie Well

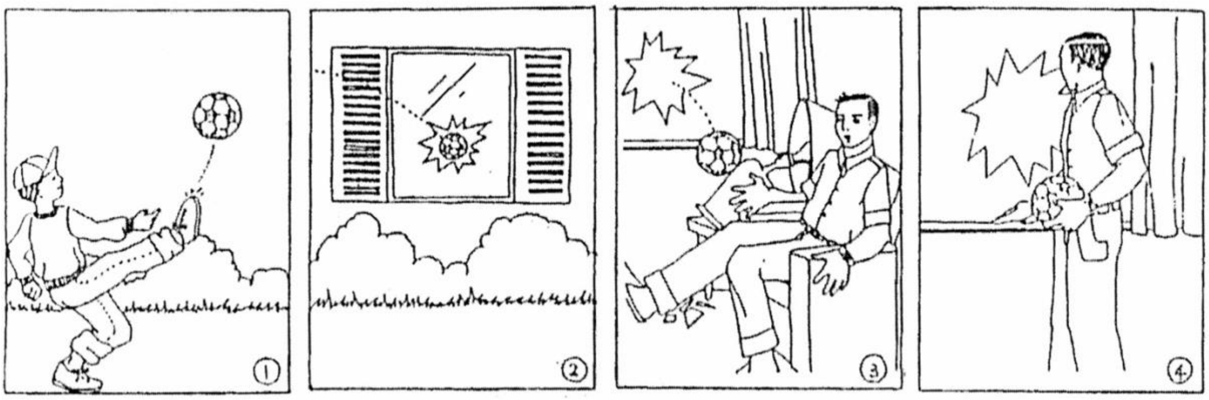


Figure 6: Broken window

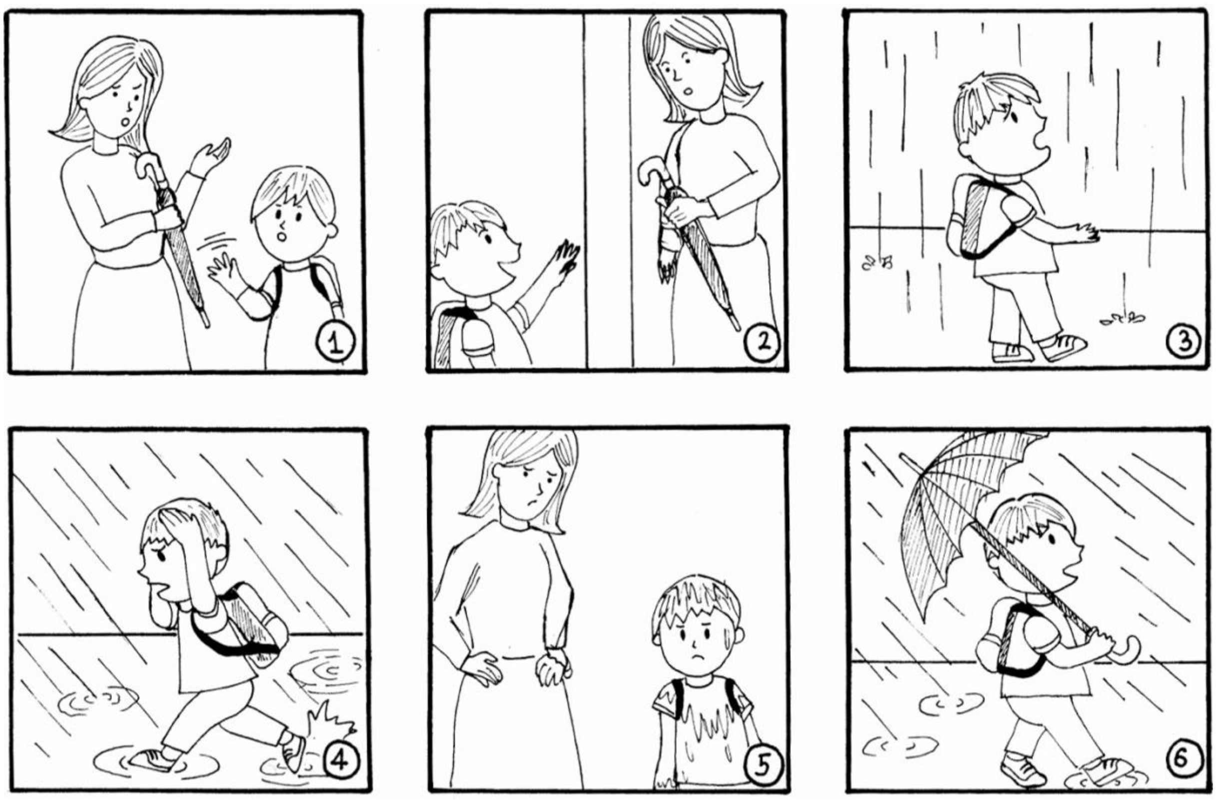


Figure 7: Umbrella

Cherepski and Drummond (1987) used pictograms and the Cookie Theft Picture to compare the different modes of elicitation in five patients with non-fluent aphasia. Although the pictograms, which were taken from the Sunday comic strip of a local newspaper, elicited a greater variety of verbal expressions, they found that the pictograms could be too cluttered for some patients and that single pictures might be easier to describe. Olness (2006) confirmed Cherepski and Drummond's (1987) assumption in their study, comparing the influence of ethnicity on the quantity and quality of discourse in African Americans and Caucasians with aphasia. All groups produced more descriptive discourse in the single picture whereas picture sequences elicit more narrative discourse. According to Olness et al. (2002), descriptive discourse only mentions characters and actions while narrative discourse also includes temporal progression. Therefore, single pictures are easier to describe, especially for people with language impairment. Nonetheless, neither single pictures nor picture sequences lead to a higher number of words as this varies individually from person to person already in premorbid processing but can also be influenced by aphasia severity (Ulatowska et al., 2003). McNeil et al. (2007) compared descriptions of single pictures and picture sequences as well as story retelling and procedural descriptions from 20 people with aphasia finding no differences in most linguistic measures. If natural discourse is the point of concern, single pictures are limited in reflecting discourse according to Olness (2006) as patients mostly produce listings of actors and action and cannot, therefore, reflect coherence. Though single picture description tasks can be sensitive enough to record subtle changes in mild aphasia as described by Hickin et al. (2015) in a single case treatment study. Concluding, single pictures are often used in diagnostics and are suitable for a first impression but should be complemented by other tests to gain a more detailed diagnosis and prognosis as well as for treatment planning.

All in all, picture description tasks are, since their introduction in 1972 by Goodglass and Kaplan, commonly used in the diagnostis of aphasia. Picture description tasks find use in fundamental research and treatment success control. Although there are standardized protocols for the application, there is disagreement over which variables should be included in the assessment and their interpretation. However, there exists a common tenor that picture sequences produce more output but do not necessarily reveal more symptoms, so that single pictures are sufficient to get a first impression of the language profile of a person with aphasia. Researchers seem also to agree on the picture material used in studies. The Cookie Theft Picture and the picture set of the talkbank are most widely used by different researchers, which makes the comparison of results easier.

1.2. Picture descriptions as part of the classical aphasia batteries

Picture description tasks are widely used in aphasia batteries to diagnose post-stroke aphasia. They all are mainly based on the BDAE and have been adapted in different countries.

Translated versions of the BDAE are used in France, Finland, Greece, India, Italy, Netherlands, Portugal and Spain. But also, the Frenchay Aphasia Screening Test (FAST; Enderby et al., 1986) was adapted and is in use in different countries amongst others Germany, India and Italy.

While the FAST only measures whether aphasia is present, the BDAE and tests based on the BDAE, such as the WAB, Comprehensive Aphasia Test (CAT; Howard et al., 2010), Esame al letto del malato (ELLM; Allibrio et al., 2009) and Esame del linguaggio 2 (Ciurli & Basso, 1996) fit tested patients in syndromes based on the Wernicke-Geschwind model (Geschwind, 1985), which was later known as the Boston classification (Benson, 1979).

One key diagnostic distinction of the Boston Classification is that between fluent and non-fluent aphasia, based on speech fluency which comprises prosody, melody, rate and pauses. Subtypes are then further distinguished by repetition and language comprehension as well as the type of paraphasias present. Paraphasias are divided into four categories: 1. Literal/phonemic (shooshbruss/toothbrush, tevilision/television), 2. Verbal/semantic (table for chair, cranberry for teapot), 3. Neologism (chantlast/fridge) and 4. Perseverative (comb, fork/toothbrush, comb/key). Syndromes are then further divided into Anomic, Broca’s, Wernicke’s, Conduction, Transcortical Motor, Transcortical Sensory, Mixed Transcortical and Global aphasia.

Broca’s Aphasia, which is characterized by effortful speech and sentence production, is usually seen as the prototypical non-fluent aphasia. People with Broca’s Aphasia are often only able to produce three to four words at a time and have limited vocabulary and word-finding problems, which is also usually reflected in their writing. The typical profile of a person with Broca’s Aphasia is presented in Figure 8. A transcript of the picture description of the Cookie Theft picture from a person with Broca’s aphasia looks for example like this:

“Water dripping.... Boy.... Girl... Okay.... Okay.... Mother...Mother” (Clinician prompt..What’s going on?) “No”.

	1	2	3	4	5	6	7
Articulation	Unable			Sometimes clumsy			Never impaired
Phrase length	1 word			4 words			7 words
Grammatical form	No syntactic word groupings			Simplified/incomplete			Normal range
Paraphasias	Word by word			Limited to short phrases			Normal
Melodic line	Present in every utterance			1-2 instances per minute			Absent
Word finding	Fluent but empty			Information proportionally to fluency			Primarily content words
Sentence repetition	0-20	30	40	50	60	70-80	90-100
Auditory comprehension	0-20	30	40	50	60	70-80	90-100

Figure 8: Broca's Aphasia Rating Scale Profile of Speech Characteristics Boston Diagnostic Aphasia Examination (purple line marking the classical case results and red lines marking the boundaries)

The most severe form of aphasia, *Global aphasia*, is also categorized as non-fluent aphasia. Patients with Global aphasia show severe language deficits in all modalities. They usually cannot speak many words or understand speech and cannot write or read.

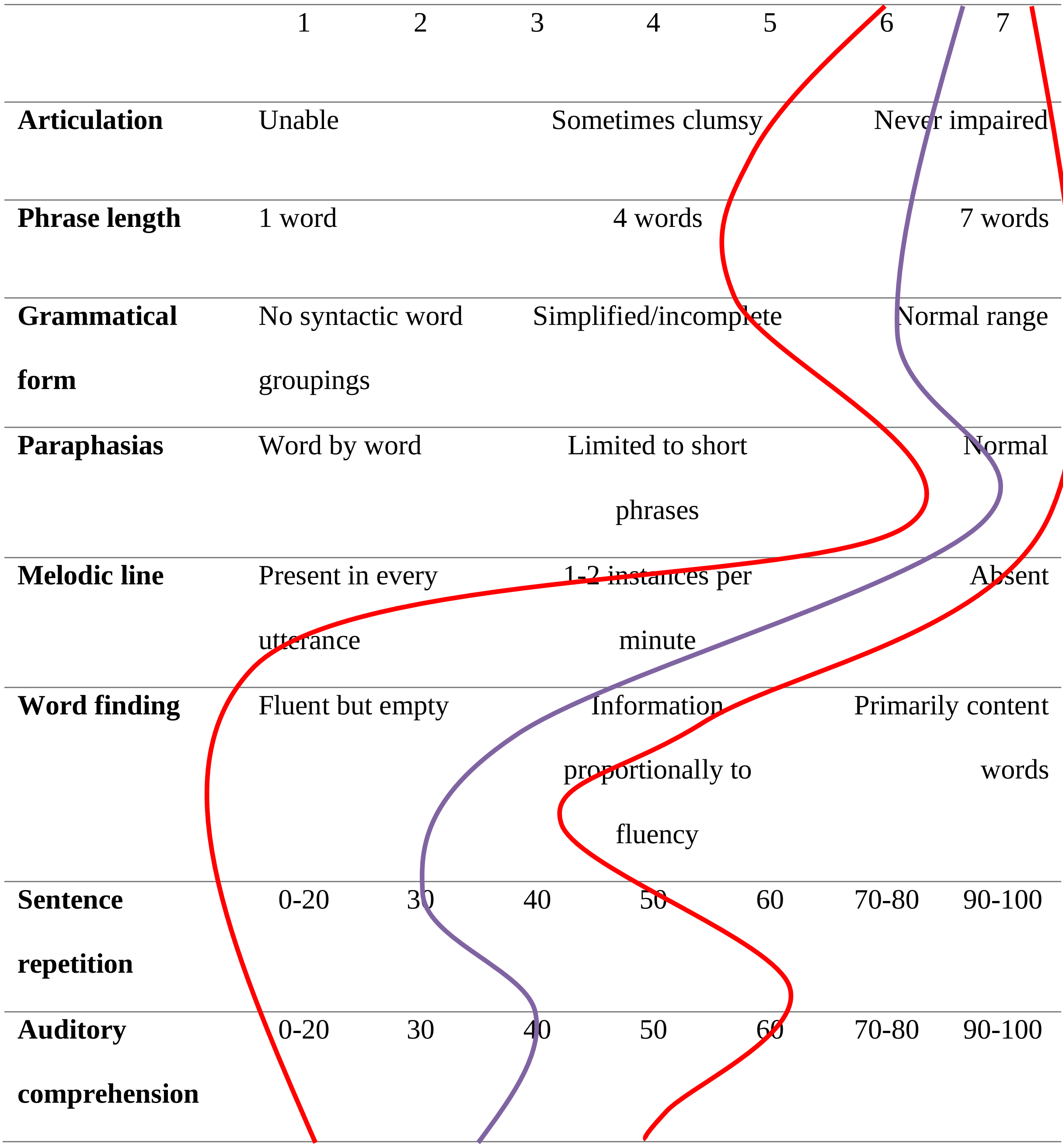
Another syndrome categorized as non-fluent aphasia is *Transcortical Motor aphasia* which is characterized by a reduced speech output, good comprehension skills and an impressive ability to repeat language.

A mixed non-fluent aphasia resembles *Broca's aphasia* in language output but differs in the comprehension modality. The language output of all non-fluent variants of aphasia is comparable to Broca aphasia and is usually characterized by short sentence fragments or list-type descriptions. Subtypes can only be distinguished by further tests such as repetition and comprehension of words and sentences.

The prime example of fluent aphasia is *Wernicke's Aphasia*. The language of people with Wernicke's Aphasia is more effortless. Speaking is not difficult for them, and words seem to pour out of their mouth, though the words and sentences are not coherent. Persons with Wernicke's aphasia are also lacking awareness of their speech production. The typical profile of a patient with Wernicke's Aphasia is presented in Figure 9. Here is an example of a picture description of the Cookie Theft picture from a person with Wernicke's aphasia:

> That's on fairble my own. Clinician: yes, your family. Stuck at that feek already.. On the fff..starting to goof uf already....Clinician: ok, do you have a large family? Do you have a big family? No, yes well there's 3, 4-4 all told but we only see masically once of a time at home and 2 of them occasionally at home..the other 3rd well he's always away at ff foam but masically on on the boys always failing...he's living at home and belonging to the future show. Clinician: They're not all at home all at once? Just one at home all the time, he goes to... goes to school?..it's a high school , it's like a hymn school like I do not know what you call it...ffffforgetting and Kathy's

she's a married man and has a daughter at home.. a new daughter.. The other girl is seeing her boyfriend but going give a poy part with her parents. And what does your wife do?

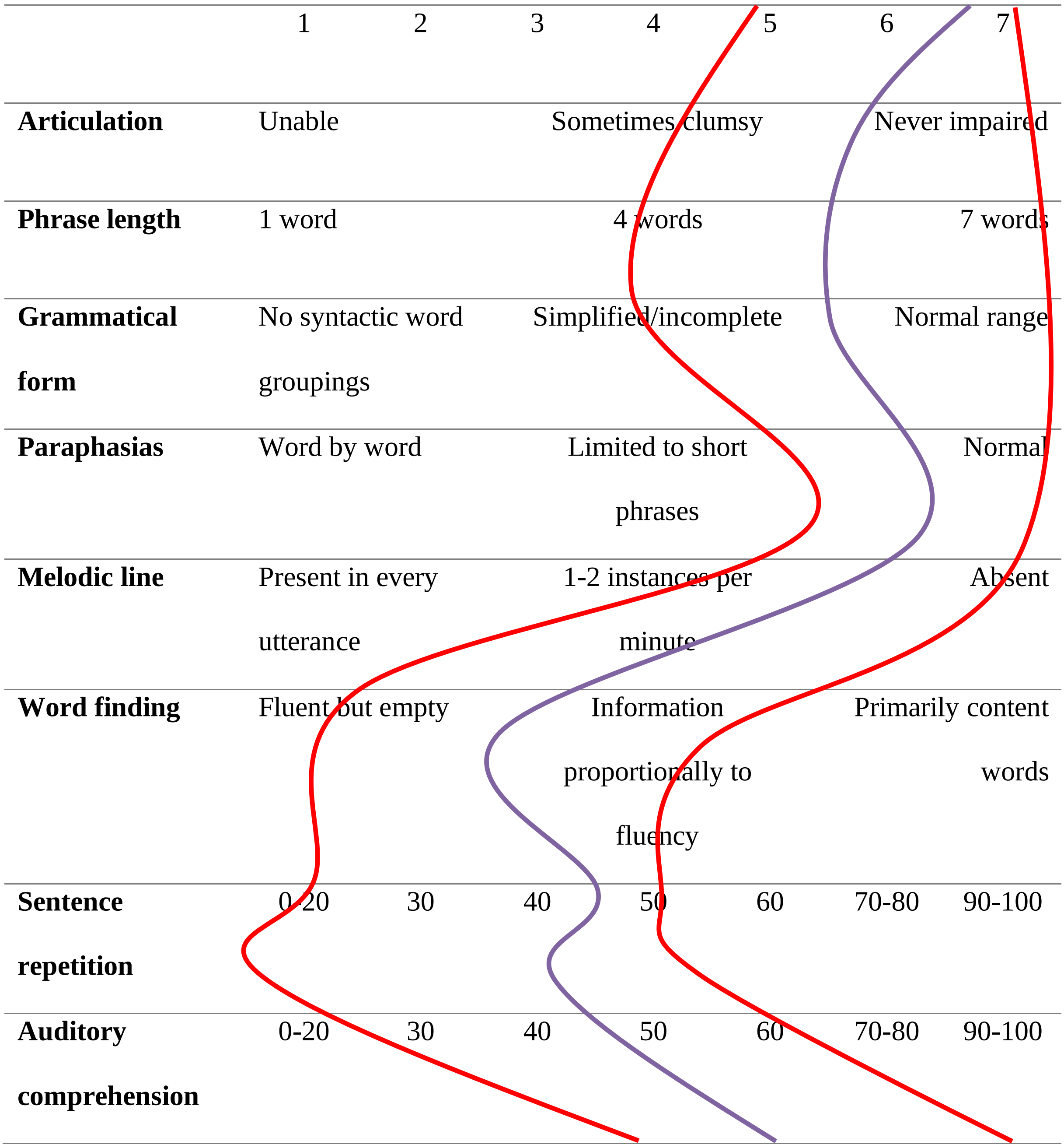


	1	2	3	4	5	6	7
Articulation	Unable			Sometimes clumsy			Never impaired
Phrase length	1 word			4 words			7 words
Grammatical form	No syntactic word groupings			Simplified/incomplete			Normal range
Paraphasias	Word by word			Limited to short phrases			Normal
Melodic line	Present in every utterance			1-2 instances per minute			Absent
Word finding	Fluent but empty			Information proportionally to fluency			Primarily content words
Sentence repetition	0-20	30	40	50	60	70-80	90-100
Auditory comprehension	0-20	30	40	50	60	70-80	90-100

Figure 9: Wernicke's Aphasia Rating Scale Profile of Speech Characteristics Boston Diagnostic Aphasia Examination (purple line marking the classical case results and red lines marking the boundaries)

Conduction aphasia is another subtype of fluent aphasia. The language of persons with Conduction Aphasia is characterized by word finding problems, conduites d'approches (lumpily, lutikly, … , likely) and literal/phonemic paraphasias, but they seem to have awareness and tend to self-corrections. The typical profile of a patient with Conduction Aphasia is presented in Figure 10. A description of the Cookie Theft Picture from a person with Conduction aphasia looks for example like this (translated from French):

> It's the kitchen .... uh it's the kitchen ..... uh..the sink uh which with the water ok ... the water, the water then she has it ... the 'an .... the a the reed..the reed..ok..she wipes the fla wipes the sla with ok ...... with .... then ... uh .. .yes ... ok .... ok ... ok ..... no ok .... no there is something else ... uh .. there is something else there .. there is the water which the calf, which the floor good ok it is the boy, the boy and the leaf, the girl, then the she is, she faal, the dru, the drum, the drum the drum is going da, go down in any case ..... it's scary! .... it's scary! Yes boy, well he wants too he wants to ga spoil spoil the cakes ... uh ..... then that's it.



	1	2	3	4	5	6	7
Articulation	Unable			Sometimes clumsy			Never impaired
Phrase length	1 word			4 words			7 words
Grammatical form	No syntactic word groupings			Simplified/incomplete			Normal range
Paraphasias	Word by word			Limited to short phrases			Normal
Melodic line	Present in every utterance			1-2 instances per minute			Absent
Word finding	Fluent but empty			Information proportionally to fluency			Primarily content words
Sentence repetition	0-20	30	40	50	60	70-80	90-100
Auditory comprehension	0-20	30	40	50	60	70-80	90-100

Figure 10: Conduction Aphasia Rating Scale Profile of Speech Characteristics Boston Diagnostic Aphasia Examination (purple line marking the classical case results and red lines marking the boundaries)

Patients with *Anomic Aphasia*, a fluent type of aphasia, have mostly word-finding problems and tend to use similar words or vague fillers. They are very aware of their difficulties. When asked what happened instead of saying: “I had an operation on my head.”, they would for example say: “I had one of them up there.” or “I had one of them where my hair is.” The typical profile of a patient with Anomic Aphasia is presented in Figure 11.

	1	2	3	4	5	6	7
Articulation	Unable			Sometimes clumsy			Never impaired
Phrase length	1 word			4 words			7 words
Grammatical form	No syntactic word groupings			Simplified/incomplete			Normal range
Paraphasias	Word by word			Limited to short phrases			Normal
Melodic line	Present in every utterance			1-2 instances per minute			Absent
Word finding	Fluent but empty			Information proportionally to fluency			Primarily content words
Sentence repetition	0-20	30	40	50	60	70-80	90-100
Auditory comprehension	0-20	30	40	50	60	70-80	90-100

Figure 11: Anomic Aphasia Rating Scale Profile of Speech Characteristics Boston Diagnostic Aphasia Examination (purple line marking the classical case results and red lines marking the boundaries)

Another fluent type is *transcortical sensory aphasia* which resembles a severe Wernicke's aphasia though people with transcortical sensory aphasia have preserved repetition skills whereas persons with Wernicke's aphasia have poor repetition abilities. All fluent aphasia syndromes comprise fluent, well-articulated speech with good prosody. Subtypes can be distinguished based on the language output but for some types such as transcortical sensory aphasia further tests are needed.

Mixed transcortical aphasia is sometimes considered a fluent but also non-fluent type. This is due to its origin: multiple lesions in anterior and posterior border zones. Characteristics of transcortical mixed aphasia are symptoms similar to global aphasia but with better repetition abilities.

Figure 12 presents an overview of all presented aphasia types sorted by fluent and non-fluent and further abilities.

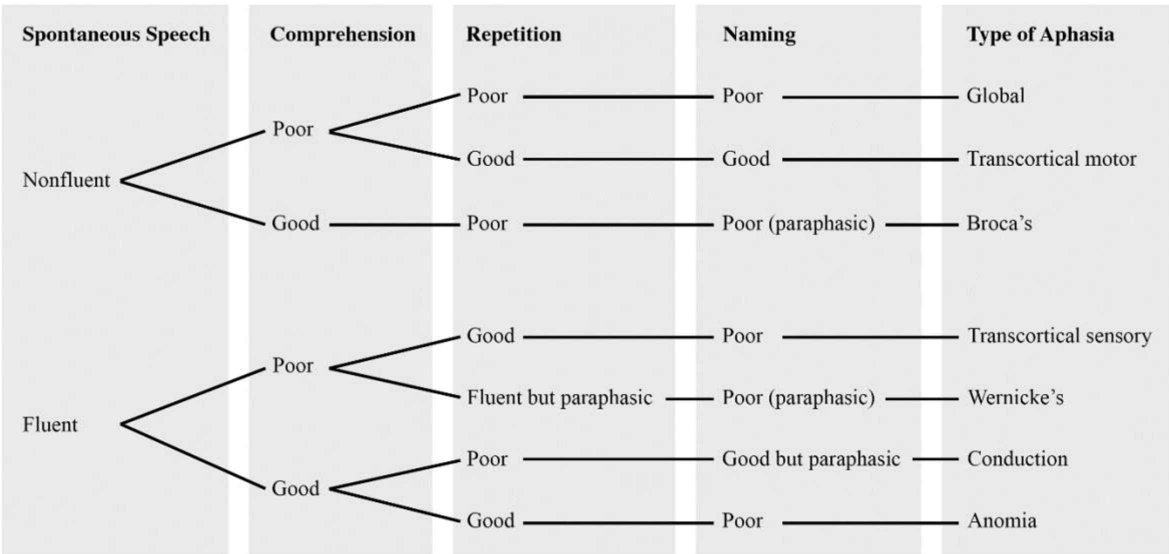


Figure 12: Overview of aphasia types and symptoms according to the Aachener Aphasie Test

Table 1
Overview of aphasia types and symptoms according to the Boston Classification

Aphasic Syndrome	Conversational speech	Auditory Comprehension	Auditory Speech Repetition	Confrontation Naming	Reading		Writing
					Aloud	Comprehension	
Anomic	Fluent, empty	Good to mild	Good	Severely abnormal	Good or abnormal	Good or abnormal	Good or abnormal
Broca’s	Nonfluent	Good	Abnormal	Abnormal	Abnormal	Good or abnormal	Abnormal
Wernicke’s	Fluent, paraphasic	Abnormal	Abnormal	Abnormal	Abnormal	Abnormal	Abnormal
Conduction	Fluent, paraphasic	Good	Abnormal	Usually good	Abnormal	Good	Abnormal
Transcortical motor	Nonfluent	Good	Good	Abnormal	Abnormal	Often good	Abnormal
Transcortical sensory	Fluent, paraphasic echolalic	Severely abnormal	Good	Abnormal	Abnormal	Abnormal	Abnormal
Mixed transcortical	Nonfluent with echolalia	Severely abnormal	Good	Severely abnormal	Abnormal	Abnormal	Abnormal

Overall, picture description tasks are a helpful instrument to diagnose post-stroke aphasia. However, some aphasia syndromes cannot be diagnosed solely by using picture description tasks. While types of non-fluent aphasia and fluent aphasia can be distinguished from one another, syndromes such as Broca's and transcortical motor or mixed aphasia can only be diagnosed with further tests like for example speech repetition, confrontation naming or writing.

1.3. Picture Description task in Neurodegenerative disorders

Picture description tasks are also a method to elicit connected language with a target stimulus in neurodegenerative disorders. They are often chosen as they are a constrained task relying on semantic retrieval and semantic knowledge as opposed to episodic memory which is particularly important in Alzheimer's dementia (AD) and normal ageing. Therefore, picture description tasks are ideally suited for eliciting connected speech from individuals with neurodegenerative diseases as patients only have to describe a scene directly in front of them. Pictures that are mostly used in picture description tasks in neurodegenerative disorders are The Cookie Theft Picture from the BDAE (Goodglass et al., 1983), The Tripping Woman Picture (Semenza & Cipolotti, 1989), The Traffic Chaos Scene (Forbes-McKay & Venneri, 2005), The Picnic Scene from the WAB (Kertesz, 1982) and the Picture of the CAT (Swinburn et al., 2004). Picture description tasks have certain advantages and limitations in neurodegenerative diseases.

As with post-stroke aphasia, the administration of picture description in neurodegenerative disorders is relatively easy. During testing, patients are asked to describe the simple or complex depicted scenes. In healthy participants, this takes usually two to five minutes, which is short compared to naming tests like the Graded Naming Test (McKenna & Warrington, 1983) or grammar tests like Northwestern Anagram Test (Thompson et al., 2011). After transcription, scoring is easier compared to spontaneous speech, due to predefined contents. Therefore, it is easier as well to compare results across languages and subjects.

Picture description tasks have been suggested to be useful in assessing lexico-semantic deficits, especially of deixis (pronouns to refer to a specific time, place, or person in a context such as tomorrow, there and they) and nouns (March et al., 2006) and are thus used to identify word finding deficits and semantic difficulties (Sajjadi et al., 2012). However, elicited syntactic structures are usually limited in variety and restricted to simple constructions such as declarative present tense statements (Garrard & Forsyth, 2010). As the speech output only delivers the information, key elements and semantic units represented in the picture, it might be difficult to assess discourse or narrative symptoms. Parallel tests of pictures used for elicitation found correlations on most measures, for example, grammatical form, error monitoring and information content (Forbes-McKay & Venneri, 2005). March et al. (2006) found that the Cookie Theft Picture was more sensitive to reduced noun use compared to describing a cartoon or map. Low correlations were shown for word-finding delays and pictorial themes, which was likely due to ceiling effects as the measures were taken from a young and healthy group (Forbes-McKay & Venneri, 2005).

Only one study was conducted for re-test reliability by Forbes-McKay and Venneri (2005) which produced low test-retest correlations distinguishing AD from controls. These results can be explained by the short retest delay of one week, which might have caused practice effects, and ceiling effects as the cohort was relatively young and healthy. Interrater reliability is another factor that makes the comparability and analysis of semi-spontaneous speech harder but is only reported in some studies. Coefficients of interrater reliability for content units are ranging between .78 and .99 (Gordon 2006; Boucher et al., 2021).

The process of transcription adds another factor of variability (Garrard et al., 2001; MacWhinney et al., 2011). Nevertheless, Forbes-McKay and Venneri (2005) performed multiple regressions adjusting for gender, age and education to provide normative scores with cut-off scores for healthy individuals up to 90 years for the measures in the BDAE discourse coding scheme (Goodglass et al., 1983) as well as measures for information content in general and specific to the picture, speech monitoring and response to word finding delays. Kavé and Goral (2016) compared language performance across Cookie Theft Picture description, picture naming and semantic verbal fluency tasks in people with AD and healthy controls. Measures of mean frequency of words, percentage of content words, nouns and pronouns as well as mean word length correlated with picture naming scores, however only the percentage of nouns correlated with semantic verbal fluency. Kavé and Goral (2016) trace this back to the fact that the picture description and picture naming tasks have visual support whereas verbal fluency tasks depend on executive functions. Picture description tasks have also been found to be sensitive to clinical between-group differences. Ahmed et al. (2013) analysed data from people with autopsy-confirmed AD at different stages of the disease and discovered that measures of syntactic complexity and semantic and lexical content portray the stages of progression of disease from MCI to AD. Picture description tasks can also be used to distinguish between AD and SD (Sajjadi et al., 2012), AD and vascular dementia (Nicholas et al., 1985) and AD and depression (Murray, 2010).

In the assessment of neurodegenerative conditions, clinicians have often distinguished between five domains of language production: discourse-pragmatic, lexico-semantic, phonetic-phonological, morpho-syntactic and syntactic. The domain of discourse and pragmatics describes all processes necessary for the continuation of a conversation such as coherence, cohesion, information content, relevance and adaptation to communication partner (Lai, 2014). Impairments at word and content levels are included in lexico-semantic features. Errors in this domain can be classified as lexical or semantic. Lexical errors comprise revisions, repetition, word-finding problems and neologisms (Croisile, et al., 1996; de Lira et al., 2011). Semantic errors usually consist of substitutions of a word with a word that is semantically related like coordinate or superordinate terms which are classified as paraphasia. At speech and sound levels variables are gathered under the phonetic and phonological domains. This includes in addition to phonematic elisions, additions, substitutions, metathesis, neologism, jargon and conduite d'approche/d'écart, different acoustic measures like the time needed to produce words, syllables or phonemes or the quantification of filled pauses, silent pauses and speech-pause-ratio (Szatloczki et al., 2015). The morpho-syntactic domain includes errors of word agreement and inflexion like gender, number, person, tense, aspect and mood. On a word level, this applies to an existent erroneous choice of a morphological form as well as the inappropriate use of non-existent words. On a grammatical level, morphological errors consist of incorrect use of tense in verbs as well as inappropriate use or absence of functors. The syntactic domain entails incomplete sentences and general structural violations as well as quantitative measures of for example utterances, words per clause, embeddings, passive constructions, dependents, simple clauses or a calculated syntactic complexity. It should be noted that these domains only describe the level a variable is situated at and are not an indication of the nature of a deficit. A pathological feature in a specific linguistic domain can be caused by different cognitive impairments. Revisions, where the speaker corrects a proceeding error, can be caused for example by an impaired lexico-semantic domain or discourse planning deficits.

The next section will discuss individual neurodegenerative diseases in more detail with their pathology and linguistic symptoms.

1.3.1. Alzheimer's Dementia

Alzheimer's is a chronic neurodegenerative disease, starting relatively slowly but gradually worsening over time (World Health Organization, 2020). One of the most common early symptoms is difficulties in remembering recent events. In later stages, problems with language as well as mood, motivation, behaviour, orientation and self-care can arise (Burns & Iliffe, 2009).

Most studies describing picture description tasks in Alzheimer's Dementia (AD) investigate the phonetic and phonologic, lexico-semantic, morpho-syntactic and discourse and pragmatic domains. On the phonetic and phonological level a low speech rate with frequent hesitations is reported but rarely phonetic errors (Hoffman, et al., 2010; Sajjadi et al., 2012). More symptoms are shown on the lexico-semantic level. People with AD use a higher number of closed function words such as conjunctions, determiners, pronouns and prepositions (Croisile, et al., 1996; Sajjadi et al., 2012; Drummond et al., 2015; Ahmed et al., 2013; Jarrold, et al., 2014) as well as a greater number of high-frequency words (Kempler et al., 1987; Kavé & Levy, 2003) and indefinite terms (Feyereisen et al., 2010; Visch-Brink et al., 2009; Lai, 2014). AD patients also tend to overuse deictic language (Nicholas et al., 1985; March et al., 2006) and make more frequent semantic and lexical errors (Kempler et al., 1987; Kavé & Levy, 2003). Furthermore, they have word finding difficulties (Croisile et al., 1996; Forbes-McKay & Venneri, 2005; Forbes-McKay et al., 2013; Ash et al., 2007; de Lira et al., 2011) and produce more revisions (Forbes-McKay & Venneri, 2005; Forbes-McKay et al., 2013), repetitions (Nicholas et al., 1985; Visch-Brink et al., 2009; Sajjadi et al., 2012) and neologisms (Fraser et al., 2016).

On the morpho-syntactic level, people with AD make more inflectional errors compared to people with no brain damage (Altmann et al., 2001; Cuetoset al., 2007; Sajjadi et al., 2012). The syntactic process is usually reported to be unimpaired (Kavé & Levy, 2003; Forbes-McKay & Venneri, 2005) but some studies report a simplified syntax with short sentences and reduced syntax (Ash et al., 2007; de Lira et al., 2011; Sajjadi et al., 2012). The discourse and pragmatic levels are impaired in AD (Carlomagno et al., 2005; Sajjadi et al., 2012; Ahmed et al., 2013; Lai, 2014; Drummond et al., 2015), especially referential and temporal cohesion (Nicholas et al., 1985; Ripich et al., 2000; Altmann et al., 2001; Dijkstra et al., 2004; Drummond, et al., 2015; Lai, 2014). This might be due to the lower number of mentioned key concepts and the addition of more irrelevant and implausible details (Bschor et al., 2001; Kavé & Levy, 2003; Ahmed et al., 2013). Symptoms, especially on the discourse and pragmatic level, are characteristic of AD and likely due to deficits in working memory.

The *logopenic variant of Primary Progressive Aphasia* (lvPPA) can be considered a variant of AD according to Knopman and Nestor (2017). lvPPA is characterized by a reduced speech rate together with a great number of filled pauses and revisions (Gorno-Tempini et al., 2008; Wilson et al., 2010; Ash et al., 2013). Core features of this presentation are word retrieval and sentence repetition deficits (Gorno-Tempini et al., 2011). It is assumed that phonological short-term memory impairment is the cognitive mechanism underlying most features of logopenic lvPPA (Gorno-Tempini et al., 2008). Therefore, while the repetition of short single words can be spared, the reproduction of sentences is characteristically impaired. This mechanism transfers to sentence comprehension, which is influenced more by the length and probability of the sentence than by its grammatical complexity (Gorno-Tempini et al., 2008).

1.3.2. Frontotemporal Dementia

Frontotemporal Dementia (FTD) affects adults usually in their fifties to sixties. Degeneration in the frontal and temporal lobes leads to a gradual progression of language deficits and behavioural change (Cardarelli et al., 2010). Data about picture description tasks are available for two forms of FTD: progressive non-fluent aphasia, also known as *non-fluent variant Primary Progressive Aphasia* (nfvPPA), and Semantic Dementia or more specifically *semantic variant Primary Progressive Aphasia* (svPPA).

The core symptoms of nfvPPA are agrammatism and effortful speech. People with nfvPPA often produce short, simple phrases and omit grammatical morphemes, for example, function words and inflexions (Gorno-Tempini et al., 2011). Their speech production is slow and effortful with a speech rate that might be up to only a third compared to healthy seniors (Ash et al., 2009; Gunawardena et al., 2010; Wilson et al., 2010; Rogalski et al., 2011). Inconsistent errors occur on the phonetic level such as deletions, substitutions, distortions, insertions or transpositions, of which people with nfvPPA are aware. Sentence comprehension can also be impaired and is clearly influenced by the grammatical complexity of the sentence (Gorno-Tempini et al., 2011).

svPPA is characterised by anomia and single-word comprehension deficits (Gorno-Tempini et al., 2011). While other language domains are relatively spared, naming abilities in svPPA are severely affected as well as single-word comprehension, especially for low-frequency words (Adlam et al. 2006). Surface dyslexia and dysgraphia are also present in patients with svPPA which means reading and writing of words with an atypical relationship between spelling and pronunciation are impaired. They tend to "regularise" such words. For example "pear" is read as "peer" (Wilson et al., 2010). Motor speech, repetition and syntax are spared. However, paragrammatic errors, such as substituting less appropriate closed-class words or inflexions, can occur (Meteyard & Patterson, 2009).

1.3.3. Language impairments in motor disorders associated with Aphasia

Few studies have been conducted to investigate connected speech in motor disorders including Parkinson's Disease (PD), Huntington's Disease (HD), Corticobasal Syndrome and Amyotrophic Lateral Sclerosis (ALS). The studies analysed phonetic and phonological, lexico-semantic, morphosyntactic and discourse-pragmatic features of language.

Parkinson's Disease is a chronic degenerative disorder affecting the central nervous system, mainly affecting the motor system (Parkinson's Foundation, 2020). As the disease slowly worsens non-motor symptoms can also arise (National Institute of Neurological Disorder and Stroke, 2020). Most common early symptoms are slowness of movement, rigidity, tremor and difficulties walking, though cognitive symptoms can also occur (Kalia &

Lang, 2015). In the advanced stages, Parkinson's disease dementia becomes common as well as other symptoms including sleep, emotional and sensory problems (Sveinbjornsdottir, 2016).

Although there is a lack of studies focusing on connected speech in people with PD the available evidence suggests that there are no significant differences between patients with PD and non-brain damaged people on the phonological, lexico-semantic and morphosyntactic levels. Perceived phonetic impairment is mostly due to alteration of prosody and duration of pauses (Rusz et al., 2011; Ash et al., 2012). A reduction in topic maintenance and local cohesion is reported on the discourse and pragmatic level (Ash et al., 2012). Overall, findings suggest that the language of people with PD is intact, with exceptions on the acoustic and phonetic levels. Changes in prosodic characteristics in PD are most likely due to the affected motor system making the voice less powerful and fine motor movements needed for speech production harder.

Huntington's disease is a mostly inherited neurodegenerative disorder (Illarioshkin et al., 2018) with a quite early onset between 30 and 50 years. Early symptoms are often characterised by subtle problems with mental abilities and mood often followed by an unsteady gait. In more advanced stages chorea, involuntary uncoordinated body movements become more apparent (Dayalu & Albin, 2015). As the disease advances coordinated movements become difficult which among others results in an inability to talk or apraxia (Caron et al., 2020) and mental abilities deteriorate into dementia (Frank, 2013).

Two studies have investigated speech deficits of people with HD (Murray & Lenz, 2001; Jensen et al., 2005). Language deficits are reported in the phonetic domain but not in the phonological domain. Although the morphological level seems unimpaired, on the syntactic level patients with HD produce a high rate of syntactic errors (Jensen et al., 2005) and a reduced number of well-formed sentences and utterances (Murray & Lenz, 2001). However, the number of dependent clauses is not reduced, hence syntactic complexity is not affected. On the discourse level, people with HD produce as many information units as non-brain-damaged people, except for action content units which show a deficit (Jensen et al., 2005). Symptoms in language production in HD are most likely caused by the underlying motor condition.

Corticobasal syndrome combines features of frontotemporal dementia (Constantinides et al., 2019) and a rare, progressive and atypical syndrome of Parkinson's (Parmera et al., 2016). Symptoms include frontal deficits also affecting speech, apraxia, myoclonus or rigidity as well as alien limb syndrome (Finger, 2016).

Only one study seems to research connected speech in CBS. Gross et al. (2010) focused on the discourse and pragmatic level which is significantly impaired when compared to people with no brain damage. The discourse in patients with CBS shows lower local and global coherence and is less accurate. They have difficulties identifying the overall point of a story as well as connecting events and maintaining the story theme (Gross et al., 2010). The previously described deficits cannot be accounted for by difficulties in naming or perceiving depicted elements or remembering story elements. They might be most likely due to the combined underlying frontal and motoric impairment. As a result, people with CBS have problems planning discourse leading to a lack of coherence and maintaining the theme of a conversation.

Amyotrophic lateral sclerosis (ALS), also referred to as motor neurone disease (MND), is a progressive neurodegenerative disorder that primarily affects motor neurons in the brain and spinal cord (NHS, 2021) and affects people before the age of 45 (Grad et al., 2017). Due to the degeneration of the upper and lower motor neurons people with ALS experience muscle weakness as an initial symptom leading to difficulties in general movement, swallowing and speech production (NIH National Institute on Neurological Disorders and Stroke, 2013). As the disorder progresses ALS spreads to other unaffected regions causing amongst other symptoms difficulties in thinking and behaviour or frontotemporal dementia (van Es et al., 2017). In the late stages of ALS pneumonia, most likely caused by dysphagia, or respiratory failure leads to death usually within 30 months after the beginning of the first symptoms (Kiernan et al., 2011).

Furthermore, amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are often seen as two diseases with distinct clinical and pathological features. However, there is a growing body of evidence that suggests significant overlap between these two diseases, with some cases showing both ALS and FTD symptoms. One of the most striking areas of overlap between ALS and FTD is language impairment. Both diseases can cause language difficulties that can present as dysarthria or dysphonia (Caselli et al., 1993). Additionally, ALS patients may experience problems with comprehension and processing of verbs (Bak & Hodges, 2004; Grossman et al., 2008), whereas FTD patients may have noun or object deficits (Bak & Hodges, 2003). Despite these differences, the language dysfunction in both diseases may be central in nature, and both can present with language deficits on neuropsychological testing (Bak & Abrahams, 2016). Furthermore, up to 50% of ALS patients have some form of cognitive impairment, which can include executive dysfunction and language deficits, with a smaller percentage having full-blown dementia syndrome (Goldstein & Abrahams, 2013).

Four studies have been conducted to analyse connected speech in people with ALS/MND focusing on phonetic, syntactic and discourse abilities. One study analysing phonetic and semantic measures concluded that it is impossible to distinguish people with ALS from people with no brain damage based on these language measures (Tsermentseli et al., 2015). In contrast, pragmatic and syntactic processes have been reported with a reduced number of words, shorter sentences and a reduced number of utterances when people with ALS were compared to healthy speakers (Ash, et al., 2014). Additionally, patients with ALS have trouble maintaining the theme of the story and connecting events (Ash, et al., 2014). The symptoms in the language of people with ALS are also a result of the affected frontal and motor domains in the brain leading to difficulties in general speech production but also coherence.

1.4. Difficulty of conducting a spoken picture description in people with aphasia, dysarthria and other neurological difficulties

It was presumed for a long time that writing and speaking are closely related and generally impaired in a qualitatively identical way. This was based on the assumption that writing is a sound-based strategy and is usually perceived as more challenging than speech (Goodglass & Hunter, 1970). However, objections to the hypothesis arose. Single case studies of people with Wernicke's aphasia (Hier & Mohr, 1977), aphasia in MS (Olmos-Lau et al., 1977) unspecified aphasia (Bub & Kertesz, 1982) and jargon aphasia (Robson et al., 1998) described cases with superior written naming over oral naming abilities. Writing skills were unusually well preserved with a superior awareness of errors which was not present in spoken language. The described cases can lead to the conclusion that written naming is a process not relying on phonological errors and that lexical information can be converted directly into graphemic code in written language. In contrast, the oral output needs access to the underlying phonology which means that reading and writing are mediated by different neural mechanisms (Bub & Kertesz, 1982). Basso et al. (1978) conducted a larger study assessing spoken and written skills in patients with left-sided brain damage. Three patients with anarthria showed common symptoms of disrupted speech output, no matter how elicited, but showed particularly good writing skills which were normal to minimally affected. Two other cases with Broca's aphasia also showed preserved writing skills, as well as two cases with fluent aphasia, who showed many symptoms in spoken language such as anomia, paraphasia and circumlocutions, but nearly unimpaired writing. However, a study by Graham et al. (2004) with fourteen patients with nfvPPA showed deficits in picture description tasks which were largely parallel in both modalities. Some patients showed even worse grammar abilities in telegram style in the written part. Basso et al. (1978) and Graham et al. (2004) studied people with different diagnoses. As shown in Rutter (2014) people with neurological disorders can perform very heterogeneously which also affects differences in spoken and written language production.

Notwithstanding, spoken picture descriptions cannot be conducted in every case, especially in patients with severe dysarthria, anarthria or aphonia. Writing is for these patients sometimes the only way of communication. Written picture descriptions are in these cases a suitable way to elicit language samples that might show a different picture of a patient's language production and comprehension. Symptoms that are caused by motoric dysfunctions could be better distinguished from cognitive functions in written tasks. A dysarthria for example would make it hard to assess whether a patient also has impairments in semantics or syntax.

Overall picture description tasks are a very well-suited diagnostic instrument for post-stroke aphasia and neurodegenerative diseases. They are used in research and standardised diagnostic batteries such as the BDAE. Different measures have been added to analyse and compare picture descriptions across different disorders and languages. However, there exists no common ground upon which picture stimulus should be used or how a picture description task should be analysed. Although writing is a preserved ability in some syndromes of post-stroke aphasia and neurodegenerative disorders, the assessment of written speech production is also quite neglected especially when it comes to picture description tasks.

To analyse the differences between spoken and written language further the following chapter will present common features and points of distinction between oral and written language. Understanding the similarities and differences of both language production modes will help form hypotheses and assumptions about how spoken and written language might present in healthy ageing adults and people with neurodegenerative diseases. The findings obtained will help build lists with principal symptoms upon which guidance for the evaluation of picture description tasks in both modes, speaking and writing, can be developed.

2. Spoken vs. written picture description

This chapter will discuss the advantages and limitations of spoken and written language respectively as well as the differences between spoken and written language from previous research. General differences between spoken and written language in the different language domains will be presented as well as possible reasons why spoken and written language differ.

Furthermore, two previous studies will be summarised that have influenced the testing method in the present thesis immensely.

2.1. Advantages and limitations of spoken picture descriptions

Spoken picture descriptions are one of the most sensitive tests for identifying subtle symptoms in early Alzheimer's dementia (AD; Bayles & Kaszniak, 1987; Forbes-McKay & Venneri, 2005; Mueller et al. 2018). Spoken language is also used more frequently in daily living for communication and is generally the centre of speech and language therapy.

Picture description tasks are usually assessed in the spoken version. For the person being assessed, this is a short and comparably easy task, depending on the symptoms, as they just need to speak. The administration is also simple as just the picture and an audio recorder are needed. However, voice recordings are becoming an increasingly sensitive topic regarding data protection. Audio files must be recorded and stored on secure servers together with their transcriptions because voice samples have a higher recognition value than for example, handwritten pieces. The need for transcriptions is another limitation of spoken picture description tasks. While the administration might be short for the patient, the diagnostician has to transcribe everything the patient expressed precisely after the patient has left, which is a time-consuming process. Trained transcribers need one hour to transcribe a file that is less than 15 minutes long. Slurred speech and background noises lead to even longer evaluation times and a more complicated analysing process (Worthy, 2021). Although there exists a wide variety of automatic spoken-to-written language conversion devices, they are prone to errors, especially if the recorded sample is spoken in strong dialect or accent or contains a lot of pauses or paraphasias or is generally inarticulate because of dysarthria. Artificial intelligence is still unable to understand human language error-free (McGowan Transcriptions UK, 2021).

However, a crucial point is that spoken language cannot be assessed in every person who might experience language problems. People with aphonia, severe dysarthria or even anarthria might not be able to describe pictures orally, as described in Chapter 1.4. Especially for this group of patients, written picture descriptions might be the best solution to obtain a more appropriate language sample according to their actual language abilities. In a patient with dysarthria for example, short sentences could be a symptom of the affected motor system, however, short sentences could also be caused by further neurological impairments. In these cases a written picture description would help to distinguish between symptoms that were caused due to deficits on the motor system or neurologically.

In our ageing society neurodegenerative disorders are becoming more and more prevalent. For example, today more than 50 million people are living with dementia worldwide with 10 million new cases of dementia diagnosed every year (World Health Organization, 2021). In 2050 this number is projected to triple to over 130 million per year. Dementia is just one example of neurodegenerative diseases. Neurological diseases are a leading cause of disability-adjusted life-years (the sum of years of life lost and years lived with disability; GBD 2016 Neurology Collaborators, 2019). Early diagnosis can help to improve the life quality of people with neurodegenerative disorders. However, healthcare systems worldwide are already struggling under financial pressure. Reducing costs has become a necessity. One possible way to reduce costs is to make the diagnosis of linguistic symptoms in possible neurodegenerative cases more effective and therefore cheaper. This could be achieved by using written picture description tasks.

The ultimate goal in the treatment of older people is that they can live as long as possible independently and communicate their needs. Thus, the speech and language therapist’s (SLT) primary goal is to determine the patient's communication abilities so that appropriate support can be offered. For this, either a test battery or a picture description task is commonly used. In a practice setting, SLTs usually have up to an hour to make a diagnosis. In hospitals, however, SLTs often have only as little as 10 minutes with the patient. Therefore, in the practice, an SLT can determine which domains of language are affected to which level, whereas an SLT in the hospital can often only estimate if the communication abilities of a patient are affected. SLTs in practices furthermore often have long waiting lists which will not improve with predicted higher future demand. The need for fast and effective diagnostic tools is therefore clear.

Written picture description tasks could help to make the diagnostic process more effective. They are easy to administer as the patient is simply asked to describe a picture by writing down their description on a piece of paper. This could be instructed by an SLT but also by a GP or even a relative/carer, prior to SLT assessment. The SLT would then analyse the written picture description and use this information in determining whether the patient needs therapeutic intervention or not. The written picture task can be administered immediately at the GP practice or at the admission to hospital. Patients not needing further diagnosis or treatment from an SLT can be immediately transferred to other specialists. This could be for example the case in patients with neurodegenerative disorders traditionally classified as non-language-dominant neurodegenerative diseases such as behavioural variant frontotemporal dementia or progressive supranuclear palsy syndrome.

There are potential further savings in transcription time. As described above, an experienced transcriber needs about 20 minutes to transcribe five minutes of audio that was recorded by a healthy person in a quiet environment. If instead of a spoken picture description a written picture description would be used for the assessment of symptoms in language production this time can be saved, leading to a significant cost reduction. The saved costs can then be used towards other tests or treatment.

However, costs can only be reduced if written picture description tasks are an adequate means of assessing people's language production. Therefore, advantages and disadvantages need to be assessed as well as practicability.

2.2. Advantages and limitations of written picture descriptions

Written picture descriptions are not yet a standard in the diagnostids of aphasia and neurodegenerative diseases but should be considered as written language is useful to assess in general. They are also comparably easy to assess, as only the picture, a piece of paper and a pen are needed. For the administered person, having to write a picture description might seem to be a more elaborate process, but written picture descriptions do not take necessarily more time than their spoken counterpart. For the diagnostician, the advantage is certainly the faster evaluation process, as the intermediate stage of transcribing will be omitted which would lead to financial benefits for healthcare systems worldwide as described above. According to Croisile et al. (1996), written picture descriptions can be sensitive to early symptoms of AD. Written communication is used less in daily communication and is usually more formal, although this factor might change in younger generations. In the last decade(s) short text messages have replaced oral communication increasingly (Lenhart, 2010; Bramley, 2015). Also, the form of writing has changed with the arising use of emoticons, which might influence future testing of language as well as therapy planning.

2.3. Literature review

Research on differences between spoken and written language in healthy individuals is relatively old (1920ies to 1980ies; for example, Dewey, 1923; Chafe, 1982). In contrast, only a few studies have been conducted to examine differences between spoken and written language in people with neurodegenerative diseases. No research has been done yet on the differences between spoken and written language in Primary Progressive Aphasia (PPA). In this review, I present major findings from empirical research about lexical, semantic and syntactic differences between spoken and written language in healthy individuals and individuals with aphasia or neurodegenerative diseases. I will also explain possible reasons for differences and discuss differences in disorders regarding written language. The chapter will focus on neurodegenerative diseases (a) as they have been neglected and (b) as a written picture description is particularly relevant for the diagnosis of neurodegenerative diseases.

2.3.1. General Differences between Spoken and Written Language

2.3.1.1. Word Count and Lexical Choice

Most studies conducted since 1920 used word counts as the primary method to distinguish between spoken and written language (Dewey, 1923; Horn, 1926; Voelker, 1942; Fossum, 1944; Drieman, 1962; Blankenship, 1962; Devito, 1965; Gibson, Gruner, Kibler, & Kelly, 1966; Devito, 1967; Gruner, Kibler, & Gibson, 1967). Although most studies differed regarding the used database, the method used to gather data and the particular purpose for word count, they all agree that lexical structure and word choice differ in English spoken and written language.

Horn (1926) and Voelker (1942) attempted to create a census of the most frequently used words in written and spoken language, respectively. However, Horn (1926) and Voelker (1942) did not bring together their findings, which prevented any further examination of the variations between spoken and written language. On the other hand, Fairbanks (1944) and Bachmann-Mann (1944) made a more significant contribution by examining the variations in spoken and written linguistic behaviour between schizophrenia patients and controls. Their study marked a turning point in the field by introducing the Type-Token-Ratio (TTR), which has become an indispensable tool in current comparative linguistic studies.

Driemann's (1962) study was a landmark in the field, as it expanded on the work of Fairbanks (1944) and Voelker (1942) by establishing a set of crucial assumptions for conducting comparative spoken and written language research. These assumptions, which include having identical topics, collecting data from the same subjects, ensuring identical circumstances of data collection, and using only complete oral and written communications, set the foundation for reliable and valid linguistic comparisons. Driemann's (1962) experiment involved instructing eight graduate students in psychology to evaluate the impact of two pictures on them, using both written and oral versions. The written descriptions were found to contain a more varied vocabulary, longer words, and more attributive adjectives while being shorter in word count. When compared with similar studies conducted in America, the Netherlands, and France, Driemann (1962) found that the results were consistent despite the differences in the experimental setup, subject choice, and measures used to differentiate between spoken and written samples. This consistency was attributed to the subjects' similar educational backgrounds and the use of the same measurement technique, the TTR. These findings hold great value for researchers, as they provide a solid basis for conducting future comparative spoken and written language studies.

DeVito's (1965, 1966,1967) series of studies conducted in the 1960s challenged the conventional tenor of the time by demonstrating that spoken language is not necessarily inferior to written language. Instead, spoken language is actually more concrete and figurative than written language and contains more finite verbs. These findings were later supported by Gruner et al. (1967), who conducted a comparative study comparing the similarities and differences between spoken and written language. Although written language contained a more varied vocabulary, the top 25 most frequently used words were similar in both spoken and written language samples. However, the distribution of these words varied depending on the modality of communication. The most common words included pronouns, demonstratives, and determiners, besides the conjunction "and".

Other researchers, such as O'Donnell (1974), have used word counts to examine language's syntax and discourse-related aspects. They found that the higher frequency of conjunctions in spoken language led to greater use of nominal clauses, while written language was characterized by a higher frequency of adjectival, adverbial, and interjected clauses. However, the change in modality was only one of many factors that influence lexical choice. For example, if participants were instructed to write in a formal style, they were found to prefer words with Latin etymology over Anglo-Saxon words (Levin et al., 1981). As a result, Latinate words tended to be used more frequently in written language, along with technical terms and word definitions. These choices were also influenced by factors such as the purpose and context of the communication, the topic, and the participant's linguistic background. Therefore, all of these variables must be controlled to truly understand the modality's impact on lexical choice.

2.3.1.2 Syntactic Structure

The study of the relationship between spoken and written language has been a subject of interest for several decades. In an attempt to better understand these differences, several approaches have been proposed and studied. TTR (Type-Token Ratio) has been found to be useful in calculating the proportional distribution of word classes but not effective in identifying higher-level differences in syntactico-semantic structures. To address this issue, Blankenship (1962) proposed a new approach, which focuses on the study of verbal expressions as the unit of analysis. This modification of Fries's (1952) grammatical system defines verbal expressions as a group of words that function in relation to a verb and is similar to a clause or a simple sentence. Despite these efforts, Blankenship failed to find significant variations in sentence length, which could be due to the choice of formal and planned speeches and publications as the data sample.

O’Donnel et al. (1967) took a different approach, combining Chomsky’s (1957, 1965) theories with Hunt’s (1965) minimal terminable syntactic unit (T-unit). By analysing spoken and written language samples of thirty children in different grade levels, they found that T-units were longer in spoken language in third graders and longer in written language in fifth and seventh graders. These findings suggest that syntactic complexity increases with advanced grade levels, however, the reasons for this transition and differences remain unclear.

Chafe (1982) proposed a unique approach, introducing idea units as a unit of analysis. Derived from spontaneous speech, idea units are defined as a set of syntactic structures bounded by a coherent unit of intonation and pauses. Chafe assumed that the pace of thought is reflected in both spoken and written language. By comparing written academic papers and informal spoken conversations of fourteen students, Chafe found that written language had a greater proportion of complex syntactic structures, such as complement and relative clauses, nominalizations, and more elaborate verb constructions. However, these results may be influenced by the differences in discourse context between informal dinner conversations and formal academic papers.

In conclusion, the consensus among studies is that written language tends to have more elaborate syntactic and semantic structures compared to spoken language (Woolbert, 1922; Borchers, 1936; Drieman, 1962; DeVito, 1964; DeVito J. A., 1966; DeVito J., 1966; Devito, 1967; O’Donnell, Griffin, & Norris, 1967; Huddleston, 1971; Poole & Field, 1976; Ochs, 1979; Chafe, 1982; Goody, Thought and Writing, 1980), with greater usage of subordinate structures (Harrell, 1957; Blankenship, 1962; O'Donnell, 1974; Ochs, 1979; Chafe, 1982), subject-predicate constructions (Blankenship, 1962; O'Donnell, 1974; Ochs, 1977; Ochs, 1979) declaratives, subjunctives, exclamations and interrogations (Portnoy, 1973; Ochs, 1979) passive verbs (Blankenship, 1962; O'Donnell, 1974; Ochs, 1977; Ochs, 1979), and definite articles, as well as a more deliberate approach to organizing ideas (Ochs, 1979; Rubin, 1980).

Furthermore, written language has a more sophisticated usage of grammatical structures such as gerunds, participles, adjectives, perfective auxiliaries, and modals as compared to spoken language, as established by several studies (Drieman, 1962; DeVito J., 1966; DeVito J. A., 1966; Devito, 1967; O'Donnell, 1974; Ochs, 1979; Chafe, 1982). This is due to eliminating redundancies, such as false starts, digressions, and repetitions, often present in spontaneous speech (Woolbert, 1922; Horowitz & Newman, 1964; O'Donnell, 1974; Chafe, 1982). Moreover, written language enables a more deliberate organization of ideas by utilizing expository language, such as propositions, topic sentences, and supporting evidence (Olson, 1977; Rubin, 1980). Written language is also more constrained compared to spoken language, requiring the production of complete ideas or information units and unambiguous communication (Woolbert, 1922; Borchers, 1936; Devito, 1965; Olson, 1977; Goody, 1980; Rubin, 1980). Additionally, prosody, a crucial aspect of spoken language, cannot be expressed in written language, but some of its functions can be replaced by graphic signs indicating syntactic relations (, ; :), pauses (, - ()), communicative intentions (statements (.), questions (?) and exclamations (!)) and emphasis (UPPER CASE, <u>underlining</u>, **bold** and *italics*). These findings are crucial for researchers and linguists studying written and spoken language differences.

The review of existing literature highlights several issues that must be addressed in order to validate findings on differences between spoken and written language. One of the main challenges is controlling data quality and comparability. Conclusions regarding differences between spoken and written language are often drawn from inconsistent sources, which raises questions about the validity of these conclusions. For example, when comparing formal speeches and published articles from the same individual, no significant differences were found (Blankenship, 1962), whereas when comparing informal dinner conversations and academic papers from the same individual, significant differences emerged (Chafe, 1982). This highlights the need to ensure that the comparisons being made are between spoken and written language, or between formal and informal discourse, rather than between individuals.

Another issue is the lack of clear definitions and measures of variables, particularly in the areas of syntax and sentence structure. Research in this area has been inconsistent, with findings often contradictory, due to different measurement methods, such as T-units versus idea units. Moreover, the tasks used in these studies are not always comparable, as they often involve different types of discourse, such as narrations, descriptions, explanations, protests, and arguments. This suggests that the underlying mechanisms behind surface behaviour may be ignored if researchers rely solely on quantitative counts.

Studies comparing spoken and written picture descriptions are rare, and the methodology used in these studies is often inconsistent. For example, one study (Rapp & Caramazza, 1997) found that the spoken picture descriptions of a single patient with aphasia contained few recognizable words, but still included syntactic elements and prosody, while the written picture descriptions consisted of content words, but lacked syntax. Another study (Drijbooms et al., 2017) found that written language was more lexically diverse but also shorter and that there were clear syntactic differences between spoken and written picture story descriptions. However, the target sample in this study consisted of children who are still developing their linguistic skills, making it difficult to generalize these findings to adults or to single-picture descriptions.

Machine learning models have also been developed to predict Alzheimer's disease from spoken and written language (Alkenani et al., 2021). However, these models were not able to compare spoken and written language from the same source. While spoken picture description samples were obtained from the Aphasia Bank, written blog articles were obtained from the Alzheimer's Disease Blog Corpus. This demonstrates the lack of comparable data sources for studying differences between spoken and written language, particularly in healthy adult populations, which would be necessary to examine pathological changes in language production.

2.3.2. Why Do Spoken and Written Language Differ?

2.3.2.1. Mode of Acquisition

The distinction between spoken and written language remains a topic of ongoing research and inquiry. There are several reasons for the differences between these two modes of language use. Firstly, the modes of acquisition of speech and writing differ. Speech is acquired naturally without formal instruction, while writing is a consciously learned skill typically acquired in school. This leads to differences in the modes of reception, production, and transmission, as well as differences in the structuring and organization of language. Writing, being a product of formal instruction, is a more standardized, authoritarian, and systematic procedure, and individuals are taught to pay attention to correctness, grammar, word choice, and organization. This is emphasized by teachers, exams, publishers, editors, and writing manuals, making the process of writing a more deliberate one that requires conscious analytical processes (Ong, 1980).

Secondly, writing is usually independent of context, whereas speech requires at least one person to listen. This means that the speaker and listener can rely on a common point of view and context, whereas writing is not always addressed to someone in particular, leading to abstraction from situational, temporal, and spatial boundaries (Goody & Watt, 1963; Greenfield, 1972). As a result, written language becomes more explicit and autonomous in the representation of meaning.

Finally, the mode of expression is another difference between spoken and written language. Spoken language is multi-modal and always includes linguistic, prosodic, contextual, and kinaesthetic cues that signal the meaning, whereas written language relies solely on the linguistic channel for the expression of intention. Although punctuation and other formatting options can convey similar intentions as prosody, the writer must rely on understanding the absent communication partner. Spoken language can express emotional, contextual, propositional, and culturally specific messages and signals, as well as an illocutionary force, while written language mainly expresses propositional messages, as it lacks non-propositional devices (Greenfield, 1972; Vachek, 1976; Olson, 1977). The formatting options in written language are more limited compared to oral language and rely on the correct interpretation of the communication partner, as misunderstandings cannot be immediately resolved. As a result, writers often resort to syntactic complexity and lexical elaboration in an attempt to overcome the absence of contextual information and prosodic properties (Gumperz et al., 1982).

2.3.2.2 Medium

The medium through which speech and writing are manifested distinguishes the two modalities. Speech is manifested phonically and conveyed through sound waves that are perceived by a listener, while writing is manifested graphically and transmitted through light waves that are perceived by a reader. The physical properties of speech result in the rapid fading of its sounds, making its reception dependent on the presence and proximity of both the sender and the receiver of a message. Conversely, written language is more enduring and can be easily transported, making it suitable for communication between partners separated by time and space. The extent to which speech and writing share phonological and semantic processes is a topic of ongoing debate among researchers. Some, such as Posner and Hanson (1980), posit that both codes share these processes as they both refer to non-physical aspects of a stimulus and require abstraction from language-related features. However, Hanson (1981) identified several features that are unique to each modality. The semantic content of spoken messages can be influenced by non-linguistic variations such as kinesic, pitch, or intonational contrast, which are difficult to express in writing and, therefore, writing is less dependent on modality-specific contrasts (Francis, 1958; DeVito, 1966; Vachek, 1976).

2.3.2.3. Situational context

The contrasting uses of spoken and written language are influenced by situational factors. While informal conversations are typically conducted through spoken language, formal contexts such as job applications, credit and tax forms, and social service requests require written language. However, with the growing use of text messaging services like Snapchat for bank transfers and credit applications, the line between formal written language and casual spoken language is becoming blurred. Despite this, spoken language is often considered insufficient for conveying complex information, which is a challenge in short text messages as well. Situational and functional differences play a significant role in shaping lexical, semantic, syntactic, and pragmatic choices in both modalities. Spoken language often focuses on interpersonal relationships, as evidenced by the prevalence of self-references, commands, and tag questions, while written language leverages lexical, syntactic, and graphic devices that are specific to the conventions of written expression (Gumperz et al., 1982).

2.3.2.4. Production Speed

One noteworthy difference between speech and writing lies in their production speed. The process of writing necessitates the use of a tool, along with deliberate coordination of cognitive and motor skills, resulting in a mechanical and artificial procedure (Ong, 1980, p. 199). In contrast, speech production is a natural process, enabling a faster rate of production. The mechanical aspect of writing, however, slows down its production speed, with research suggesting that the average speed for written English is only about one-tenth that of spoken language, which is estimated to be around 180 words per minute (Chafe, 1982). This contrast in production speed has been found to impact the expression of thought. The slower pace of writing allows for thoughts to develop ahead of expression, promoting the integration of multiple ideas into a cohesive linguistic unit. In spoken language, thoughts and expressions occur in a more concurrent manner (Horowitz & Newman, 1964; DeVito, 1966; Chafe, 1982).

2.3.2.5. Plannability and Permanence

How written and spoken language are expressed is subject to differences beyond mere speed. Writing is characterized by its correctability, permanence, and ability to reorganize thoughts and expressions in a way that is not feasible in speech. According to Martinet (1962), speech is ephemeral and fades quickly unless recorded, while writing leaves permanent marks and is reproducible. The durability of written language enables it to be reviewed and altered as needed, facilitating the reorganization and reconsideration of expressions (Goody, 1977). In comparison, spontaneous speech is less plannable than written language, leading to a higher type-token ratio in writing as the writer can modify their choice of words.

In unplanned discourse, speakers rely heavily on immediate context and morpho-syntactic structures acquired in the early stages of language development, whereas in planned discourse, late-emerging structures are more commonly employed. The rate of repeated and replaced lexical items is higher in unplanned discourse, and the content of social acts is also higher. Based on these characteristics, Ochs (1979) posits that differences between spoken and written language stem from syntactic and semantic variations between planned and unplanned discourse.

It is worth mentioning that, as well as being planned, written language is often subjected to further revision processes. If deleted, substituted, or corrected elements were examined, false starts, imperfections, and hesitations would become apparent, which are usually observed in speech. These blemishes are usually edited out in written language, which is not possible in speech due to its spatiotemporal restrictions.

2.4. Conclusion

Summarising, this review discussed how and why written and spoken language differ. It is assumed that spoken and written language makes use of the same semantic base and lexico-syntactic system. Variation was found in the distribution and choice of vocabulary and syntactic types mainly due to modality-specific constraints. Different strategies are used to express thematic cohesion in spoken and written language. While in speech certain kinaesthetic and prosodic cues are employed automatically to express intentions writers must carefully choose words and syntactic patterns to express similar intentions. This implies that differences between spoken and written language should not be investigated in a purely quantitative fashion and rather be analysed in a broader framework.

Concluding, the major conclusion is that comparative studies should include more factors influencing language as opposed to just collecting quantitative differentiation. Written language cannot offer exactly what spoken language can offer but conveys therefore other important aspects. In written language visuospatial characteristics could be used for example the spacing of words or lines or the regularity of the letters. It is suggested that people with PD write very small and cramped text which is also called micrographic (Kekatos, 2017). A visuospatial analysis of handwriting might detect this earlier. It is also suggested that people with AD show irregular formed letters coupled with trembles (Kekatos, 2017) which could potentially also be used for early diagnosis.

2.5. Comparison of spoken and written picture descriptions in primary progressive aphasia and healthy controls

Prior to the research project of this thesis, two relevant studies were conducted as part of MSc dissertations. A summary of these studies now follows, in order to put the current thesis into context. Furthermore, the measurement tools and testing procedure in these two studies will be described, as many of the same methods were used for the testing procedure of this thesis.

In 2014, Rutter tested ten patients with Primary Progressive Aphasia and Corticobasal Syndrome as well as fourteen controls. The patient and control groups both described the Cookie Theft Picture from the Boston Diagnostic Aphasia Examination, once in the classical oral form and once in writing. The study aimed to determine the differences between spoken and written picture description tasks and whether written picture description tasks could replace oral picture description in the diagnostic process of Primary Progressive Aphasia. However, the control group was not matched to the patient group regarding gender and age, factors that can influence language production significantly.

In 2018, I conducted a follow-up study replicating the study by Rutter (2014). I recruited controls that were a better match for the patient group of Rutter (2014). The testing procedure and methods for analysis were taken from Rutter (2014).

2.5.1 Testing procedure

Patient group

Rutter (2014) assessed and analysed spoken and written speech samples of 10 patients. During the testing session at the Anne Rowling Neurology Research Clinic the patients were given the Cookie Theft Picture and asked to describe what is happening in the picture, first orally and then in written form. The oral picture description was videotaped to be able to analyse gestures post-testing. In between the picture description tasks the patients completed a motor exam, lasting seven to ten minutes, which is not described in further detail.

Control Group

The 15 healthy controls of Rutter (2014), that were tested at the University of Edinburgh, performed the picture description tasks in the same order as the patients, which was oral first and written second. This was to keep the language samples more comparable between the patient and control group. The spoken description was audiotaped. Between both picture description tasks, the controls performed two short distractor tasks, the trail-making task which was derived from the Army Individual Test Battery (1994) and the digit symbol coding test, which is a part of the 3rd edition of the Wechsler Adult Intelligence Scale (WAIS-III, Wechsler, 1997). The distractor tasks took about five minutes to complete.

2.5.2. Transcription and coding process

Two speech and language therapists transcribed and scored the spoken and written speech samples independently in Rutter's (2014) study. The transcriptions and scorings were later compared. In case of disagreement, a repeated examination of the video or audio file was initiated followed by a discussion to solve the discrepancies. The transcription and scoring process followed a determined order.

For the oral task, first, the description was transcribed verbatim. Then gestures were identified, and the corresponding symbol was added. After that, the total time was noted. In the next step pauses, word-finding difficulties, and phonemic paraphasias were identified and marked with the corresponding respective symbol before the remaining symptoms were identified and marked accordingly. Then the total number of informational content units were counted and the number of informational content units for actions. As a next step, the number of main clauses, subordinate clauses and incomplete clauses was noted together with the count of syntactic errors. Finally, the number of words was counted.

The written sample was transcribed word-by-word first. Second, it was noted whether the sample showed capitalised writing and if the sample was written in print, cursive or a mix of both, as well as whether it was written in a text or list format. Then spelling errors, punctuation errors, word onset errors, unidentifiable words, and letters as well as crossed-out words and letters were identified and marked with the corresponding symbols before all remaining symptoms were identified and marked accordingly. In the next step, informational content units were counted, in total and for actions only, as well as the number of clauses, subordinate clauses and incomplete clauses. Then the syntactic errors were marked and counted. Finally, the total word count was noted.

The complete list of scoring symbols used by Rutter (2014) is provided in Table 2. The information units are presented in Table 3.

Table 2
Scoring Symbols used by Rutter (2014)

Symbol	Meaning
*	Prompt
J	Jargon
Sp	Spelling error
F	Filler
W	Word-finding difficulty
P	Pause
C	Conduite d’approche
S	Semantic substitution
Ph	Phonemic aphasia
R	Repetition
V	Revision
M	Metacognitive comment
Gs	Gestures, meaningful
MC	Main clause
SC	Subordinate clause
IC&	Incomplete clause
&	Grammatical error
$	Dysfluencies, not further specified
FO	Error in focus
IrrCon	Irrelevant content
/	Punctuation error
*	Crossed out word
%	Crossed out letters
$	Unidentifiable letter
^	Inappropriate gap in a word
“	Unidentifiable word
=	Word onset error

Table 3
Information Units

Category	Items
Subjects	The boy, the girl and the woman
Places	The kitchen and the exterior seen through the window
Objects	Cookie, jar, stool, sink, plate, dishcloth, water, window, cupboard, dishes and curtains
Actions	Boy taking or stealing, boy or stool falling, woman drying or washing dishes/plate, water overflowing or spilling, action performed by the girl, woman unconcerned by the overflowing, woman indifferent to children

2.5.3. Linguistic analysis

In the linguistic analysis, Rutter (2014) included a comprehensive exploration of five main domains: Global length, syntax, informational content, speech fluency, and writing disfluencies. The formulas Rutter (2014) used can be seen in Table 4. The variables and measurements used in both tasks are shown in Table 5. Variables and measurements used in the spoken part only are listed in Table 6 and variables and measurements used in the written part only are illustrated in Table 7.

Table 4
Formulas for indices and ratios

Type of ratio index	Formula
Syntactic complexity ratio 1 (SCR1)	(Number of subclauses/total number of clauses)*100
Syntactic complexity ratio 2 (SCR2)	(Number of subclauses/total word count)*100
Grammar error ratio (GRR)	(Number of grammar errors/total word count)*100
Incomplete clauses index (ICR)	(Number of incomplete clauses/total word count)*100
Conciseness ratio 1 (CON1)	(Total amount of information/total word count)*100
Conciseness ratio 2 (CON2)	(Total amount of information/total time spoken in seconds)*100
Silent pause ratio (PAR)	(Silent pauses/total time spoken in seconds)*100
Disfluency ratio (DISR)	(Number of disfluencies/total word count)*100
Semantic substitutions ratio (SEMR)	(Number of semantic substitutions/total word count)*100
Spelling error ratio (SPR)	Number of spelling errors/total word count)*100
Word onset error ratio (WOR)	(Number of word onset errors/total word count)*100
Jargon syllables ratio (JGR)	(Number of Jargon syllables/ total word count)*100

Table 5
Variables and measures used in both tasks

Measure	Explanation
Word count (WC)	• Excluded jargon syllables, unidentifiable words, letters and fillers • Contractions were counted as two words
Semantic paraphasias (SEM)	Substitution of one full word for another on the basis of a meaning relation between the two
Repetitions (RP)	Words or a phrase that were repeated in the exact same way
Revisions (RV)	Every kind of revised single words or phrases or modification
Word-finding difficulties (WF)	WF observations were based on our subjective impression, also regarding mimic and gestures, groping action, filler phrases, substitutions with non-content-words, circumlocutions and extended pauses
Main clauses (MC)	Each group of independent sentence, containing at least a subject and a verb
Subordinate clauses (SC)	Each group of words forming a dependent clause
Incomplete Clauses (IC)	Each clause in which either the subject, object or obligatory object was missing
Grammar errors (GR)	Every kind of erroneous use of grammatical rules, including incomplete sentences
Informational content (INF)	Places, subjects, objects and actions (see Table 3)
Inappropriate content (IRR)	Every kind of inappropriate, irrelevant or wrong content
Inappropriate use of discourse prominence	Every kind of referential communication, using the pronouns he/she without previous specification of the subject which is referred to
Metacognitive comments	Comments of the patient/participants about their own performance or abilities, or a phrase that indicated the end of their description
Others	Any other symptoms or relevant factors that did not fit into one of the specified categories, for instance perseveration

Table 6
Variables and measurements used in the spoken task only

Measure	Explanation
Time	Time in seconds from onset until the end of description
Gesture/pointing behaviour	Every kind of meaningful gesture or pointing action that is observed in patients who have a very poor speech production
Pauses	Subsequently perceived pauses between words within and/or between clauses defined as one second of silent break or a shorter but inappropriate silent break between words within a clause
Prompts	Every attempt of the experimenter to encourage the patient/participant to tell more when the outcome was poor
Jargon syllables	Every kind of senseless syllables or utterance that was not understandable to either of the scorers
Fillers	Every kind of non-silent pause between words or syllables filled with um, erm or variations of these
Conduite d’approche	Every sequence of phonemic approximations to a target word
Phonological Paraphasias	Substitution or addition of phonemes in a target word

Table 7
Variables and measurements used in the written task only

Measures	Explanation
Capitalised/regular writing	It was noted whether the patient wrote with capitalised letters or regular cursive or printed letters
Consistency in capitalised or regular writing	It was noted whether the patient was consistent or inconsistent in capitalised or regular writing
Text format	It was noted when the patient wrote in list format, rather than a text
Spelling errors	Every kind of spelling error, excluding wrongly capitalised
Word onset errors	every occasion of wrongly capitalised words, excluding cases when the patient was inconsistent in printing/writing in general
Punctuation errors	Every occasion of erroneous use of comma and full stop
Unidentifiable words	Every word that could not be identified by either of the raters
Unidentifiable letters	Every letter or group of letters that could not be identified by either of the scorers
Crossed out words	Every occasion of a crossed-out word, additionally counted as revision
Crossed out letters	Every occasion of a crossed-out letter, additionally counted as revision

The global length domain comprised the total word count (WC) excluding fillers, ellipses, word fragments and jargon syllables as well as the number of clauses (MC and SC).

The analysis of the syntax domain was adapted from suggestions of Goodglass and Kaplan (1983), Shewan (1988) and Croisile et al. (1997) to investigate grammar errors and syntactic complexity. The density of grammar errors (GR) and incomplete sentences (IC) was measured by counting these errors and dividing the result by the total word count, giving the grammar error ratio (GRR) and incomplete clause index (ICR). To analyse syntactic complexity, two different formulas were used. For the first ratio, the number of subordinate clauses was divided by the number of total clauses (SCR1). For the second syntactic complexity index, the number of subclauses was divided by the total word count (SCR2). In subsequent analysis both indexes of syntactic complexity correlated highly with each other ($r = 0.98$ for the spoken picture description and $r = 0.96$ for the written picture description task), so for later comparisons only SCR1 was used.

The *informational content* domain was analysed following studies by Croisile et al. (1996), Groves-Wright et al. (2004) and Ahmed et al. (2013). To calculate a conciseness ratio, as presented in Table 3 above, 23 information units (INF), divided into four categories, objects, subjects, places and actions, were counted. Two different conciseness ratios were determined. For the first conciseness ratio (CON1) the total number of content units was divided by the total number of words, similar to studies by Hier et al. (1985) and Cooper (1990). The second conciseness ratio (CON2) was calculated by dividing the total amount of information units by the total time, following Brookshire and Nicholas (1994). As the correlation between both ratios was high ($r = 0.93$) Rutter (2014) used only CON1 in her further analyses. In addition, actions were investigated separately, as it has been frequently reported that people with aphasia (Bastiaanse & Jonkers, 1998), PSP (Bak et al. 2006) and MND (Bak & Chandran, 2012) have difficulties expressing verbs and actions. Rutter (2014) also included pointing behaviour and meaningful gestures in the informational content and calculated an irrelevant content ratio by counting irrelevant/inappropriate content units and dividing them by the total word count.

In the analysis of the domain *speech fluency and lexical aspects,* Rutter (2014) included a variety of variables: semantic paraphasias (SEM), word finding difficulties (WF), repetitions (RP), revisions (RV), conduites d'approche (CA), jargon syllables (JG), filler (F) and silent pauses (PA). She also noted that she only analysed disfluencies, rather than distinguishing between natural disfluencies and pathological dysfluencies of non-fluent speech (Wingate, 1989). Fillers, conduits d'approches, repetitions and revisions were pooled in one general oral disfluency variable (DIS_{OR}). For the written task, semantic substitutions, repetitions, revisions, crossed-out words and letters as well as unidentifiable words and letters were added to a written disfluency variable (DIS_{WR}). Additionally, separate ratios for jargon syllables (JGR) and semantic substitutions (SEMR) were determined by dividing their number by the total amount of words. Furthermore, spelling errors (SP) and punctuation errors (PCT) were counted as well as word onset errors, which were classified as a feature of spelling errors and counted separately. Lastly, Rutter (2014) also examined whether the writing was capitalised or in regular print or cursive.

2.5.4. Testing of better matched controls

The aim of my MSc dissertation in 2018 was to collect better control data for the comparison of spoken and written picture description tasks in patients with rare neurodegenerative diseases, addressing the limitations of the study of Rutter (2014). I supplemented the previous data collected by Rutter (2014) with a new set of control participants. The newly recruited control group was counterbalanced with the already existing patient group recruited by Rutter (2014).

I recruited 10 unimpaired control participants with the Volunteer Panel of the University of Edinburgh. All participants were native English speakers with no vision or hearing impairment. The control participants completed the same testing procedure as described by Rutter (2014) in the labs of the Psychology building of the University of Edinburgh (see chapter 2.5.3.).

The demographic information and the years of education were matched to the initial patient group. Participants were chosen to be as similar as possible regarding age, gender and years of education to their patient counterparts to create a control group matching the patient group as close as possible.

2.5.5. Results and Conclusions of previous studies

The results of the studies of Rutter (2014) and Lingscheid (2018) revealed that the control group produced shorter descriptions in the written task compared with the spoken task. However, in the domains of syntax, phonology and semantics, the control group revealed no significant differences between the oral and written picture description task. The patient group showed a heterogeneous pattern. Whereas some patients performed better in the oral picture description task, others revealed better results in the written picture description task. A general pattern could not be derived from these results, though they suggest that for some patients, a written assessment may provide a more accurate picture of their true communicative abilities.

As an overall conclusion of the findings in both studies, it can be said that the choice between oral or written versions depends on the clinician's objective. The oral task should be the first choice if a more detailed assessment of, for example, type of grammatical errors is needed, as in most cases this task provides more output that can be analysed. Furthermore, the oral task also works well as a tool for the evaluation of general communication abilities apart from pure language production. This includes pointing behaviour, gestures and also pragmatic skills. However, if the clinician's objective is a quick first evaluation of whether a patient has a language impairment, the written administration would be the method of choice as it is quicker to analyse and more reliable as an indicator. In addition, the studies also demonstrated that when testing for deficits in language function using picture description tasks, a combination of clinical expertise and statistical significance is important to distinguish as accurately as possible between healthy and impaired language.

The main limitation of the presented studies was the small sample size. Additionally, the range of symptoms was too generalised and did not include behavioural symptoms as well as distinctions between different types of phonetic and semantic symptoms. These limitations will be addressed in this thesis, in which I investigated differences between spoken and written picture description in a much larger group of healthy individuals. Furthermore, the influence of basic language functions, cognitive functions, administration mode and picture stimulus were investigated and will be presented in the present thesis.

3. Comparison of Spoken and Written Picture Descriptions in Healthy Adults

To address the limitations of both MSc studies described in chapter 2 the present study increased the sample size to 30 participants and extended the list of symptoms. Furthermore, the number of picture description tasks and picture stimuli were increased as well as the number of distractor tasks.

A new comparison of spoken and written picture descriptions in healthy adults was performed, to address the limitations of the studies presented in chapter 2.4.

The sample size was increased to 30 participants to produce a more robust data set displaying natural differences in healthy adults. Younger adults were included in the study as certain neurodegenerative diseases that can affect language can emerge early in life particularly in frontotemporal dementia (Ratnavalli et al., 2002), but also in Alzheimer's dementia (Koedam et al., 2010), Parkinson's disease (Wickremaratchi et al., 2009) and genetic diseases like Huntington disease (Maat-Kievit et al., 2002).

Furthermore, a second, non-familiar picture, the Indian Street Scene was added to the picture description task to test the influence of familiarity on the outcome of picture description tasks. This topic will be addressed in more depth in chapters 7 and 8.

Additionally, the first list of measures was extended to gain a more diverse picture of healthy patterns and differences in spoken and written picture description tasks. Amongst others, different measures were added to the general category as well as to the categories of syntax, lexicon and phonology.

Finally, the number of distractor tasks was increased and diversified to control for other language variables or cognitive functions that can influence the performance of picture description tasks.

3.1. Participants

A group of 30 participants was recruited with the Edinburgh University Volunteer Panel. The demographic information of the participants is presented in Figure 13.

The group consisted of 19 females and 11 male participants. The average age was 66.67 (*SD* = 7.86) with the youngest participant being 46 years old and the oldest participant having reached an age of 82. The educational level of the participants was well distributed with most participants holding a bachelor's degree (*n* = 9) or some post-graduate education (*n* = 8).

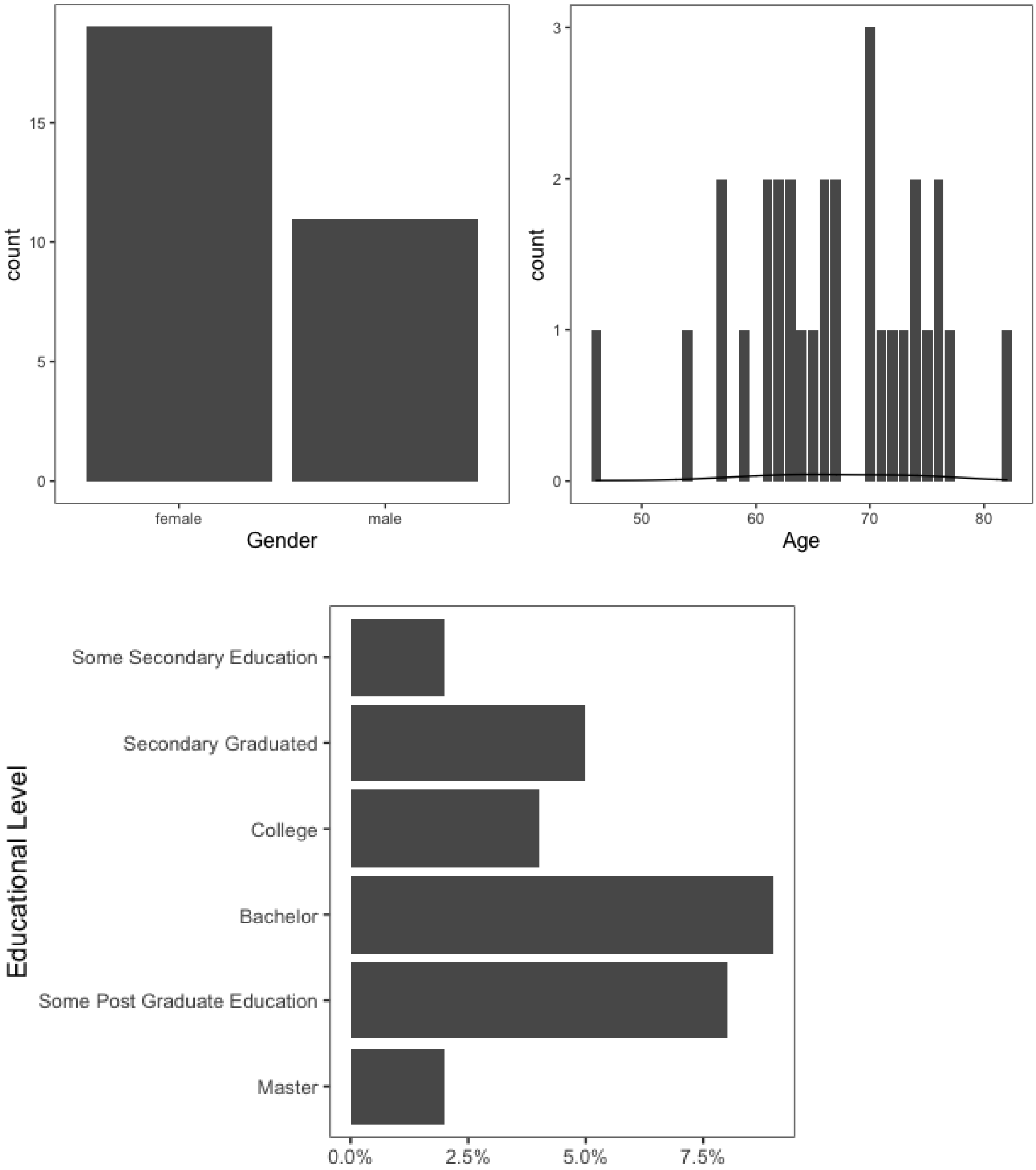


Figure 13: Demographic information of participants

The participants came to the University of Edinburgh Psychology building for two testing sessions in total, with a break of seven days in between both testing sessions. Each testing session lasted from 45 minutes to one hour, depending on how fast participants finished their tasks. Figure 18 shows a flowchart of the task order in both in-person testing sessions.

3.2. Testing Procedure

Before the testing session, the participants were asked to fill out a questionnaire about their school experience, especially how they felt when learning reading and writing, as well as their current reading behaviour. The Dyslexia Questionnaire can be found in Appendix B. If a participant had forgotten to fill in the questionnaire, they were handed a questionnaire and asked to complete the questionnaire before the next testing session and bring it along with them the following testing session. Should they have forgotten to bring the questionnaire to the second testing session, they would have been asked to complete the questionnaire at the end of the second testing session. This was mainly to avoid priming the written tasks, though none of the participants forgot to bring the questionnaire to the second testing session. Participants were asked to complete the questionnaire, firstly to identify potential dyslexia (which was not an exclusion factor of the study) and secondly to assess the participant's attitude toward writing. Participants that have a more positive attitude towards writing or positive school memories might enjoy writing more in their written picture description task whereas participants with a more negative attitude might have a lower word count (NHS, 2018).

Upon arrival, participants were randomly distributed into 4 different groups. Participants were presented with the pictures in a different order depending on their group. Also, the task type of spoken or written picture description varied between groups. At the end of both testing sessions, all participants have described both pictures once in spoken and once in handwritten form, resulting in a total of 4 picture descriptions.

For the oral task participants were asked to describe "What is happening in the picture?". Their description was audio recorded and later transcribed by a trained speech and language therapist following the guidelines of the Aachener Aphasie Test (Huber et al., 1983) for the transcription of spontaneous speech. The samples were transcribed using the software PRAAT (Boersma & Weenink, 2021). The software can automatically section a language sample in speech and non-speech sections. This made it possible to calculate silent pauses in an automated process, to make transcriptions more exact and comparable.

For the written picture description task, participants were asked to write down on a piece of paper "What is happening in the picture?". They were advised to write in full sentences. In contrast, participants were not advised to speak in sentences in the spoken picture description task. In the previous MSc studies, it was experienced that participants speak in complete sentences in the oral task. However, several participants wrote in bullet points in the written task. The participants were advised to write in full sentences so that the language samples were comparable, both with other written samples in this study and to spoken language tasks.

In between the picture description tasks, the participants completed 3 tests in each session, resulting in 6 tests. The tests were used in two different ways. First, as distractors between picture descriptions, participants forget what they have spoken or written. Secondly, the tests assessed different language domains or neurological abilities that can influence the participants' performance in the picture description task. Furthermore, some of the distractor tests are commonly used clinical tests. Results of these tests can be compared to other studies and norm data to identify how the population of this study performs compared to the general population, which can provide an orientation of the grade of generalisability for other results of this study. The distractor tests were ordered in a way so that tests that are assessed orally were followed by a test that is assessed in written form or respectively reversed.

The first test was the **Edinburgh Cognitive and Behavioural Amyotrophic Lateral Sclerosis Screen** (ECAS; Abrahams et al., 2021; Niven et al., 2015). Only parts of the ECAS were used to assess specific abilities. These were *word generating, theory of mind and visual perception*. The number of words generated could influence participants' general naming and word-finding ability, leading to a higher density of words per minute or more diversity in words used. Theory of mind could influence how a participant interprets actions performed from and between characters. Visual perception can have a general influence on the perception of pictures. Furthermore, the ECAS was developed to assess Amyotrophic Lateral Sclerosis also known as Motor Neuron Disease, a disorder in which motor symptoms make oral language assessment almost impossible. Therefore, a written version of the ECAS was developed to give patients an assessment opportunity. The ECAS is one of the only tests providing an additional written assessment option, making it even more suited for the presented study design and aim.

Next, the participants completed the **Northwestern Anagram Test** (Thompson et al., 2011; Figure 14). The NAT tests grammar. In this testing session, the short version was used, consisting of 10 questions, which participants had to generate from pre-existing words they had to put in the correct order. The NAT was chosen as it is a short test for syntax. As described in more detail in chapter 2.3. no study has compared connected language production and results of a test evaluating grammar abilities in a designated single test. Furthermore, the NAT was developed to test action word processing which is often affected in patients with PPA.

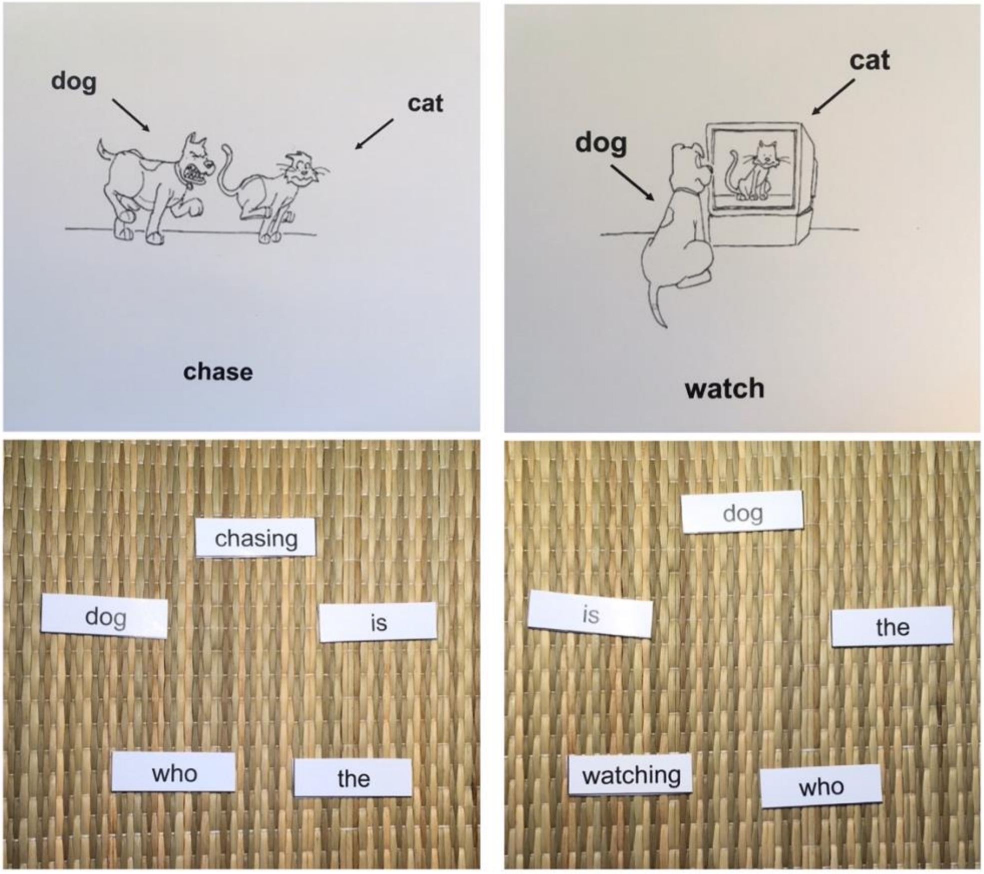


Figure 14: Northwestern Anagram Test Examples, “Who is the dog chasing?” (left), “Who is the dog watching?” (right)

Finally, the last distractor test in the first testing session was **Raven’s Progressive Matrices** (Raven, 1938; Figure 15). Raven’s matrices measure intelligence nonverbally by completing a logical pattern of pictures. Intelligence or analytical abilities can also influence how a person perceives and interprets a picture. Furthermore, intelligence can also influence general language abilities. The test was limited to 10 minutes. For the analysis, correctly solved puzzles within the time frame were counted.

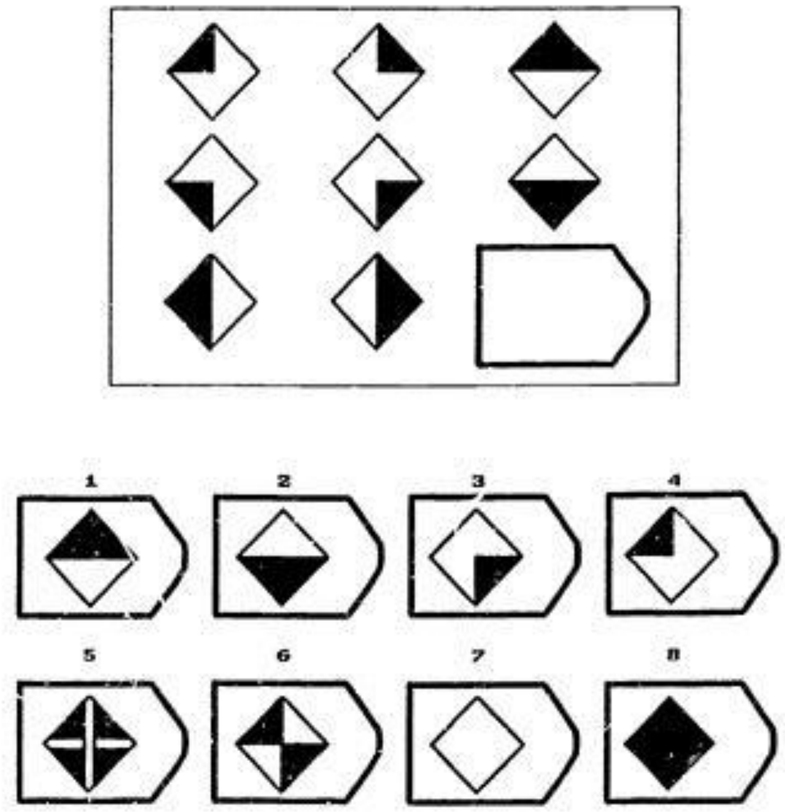


Figure 15: Raven's Progressive Matrices

In the second testing session, which took place a week later, the first distractor task the participants completed was the **Sorting Test** (Delis et al., 2001; Figure 16). In this test, participants had to form groups with shapes that have words and patterns on them. They do this twice in total with two different sets. This test had several different goals for this study. First, it assesses the participant's grade of abstraction. Second, as the participants had to describe why they sorted something a certain way, conclusions can be made on their general description and reasoning ability regarding language. Thirdly, it distracted them from the pictures.

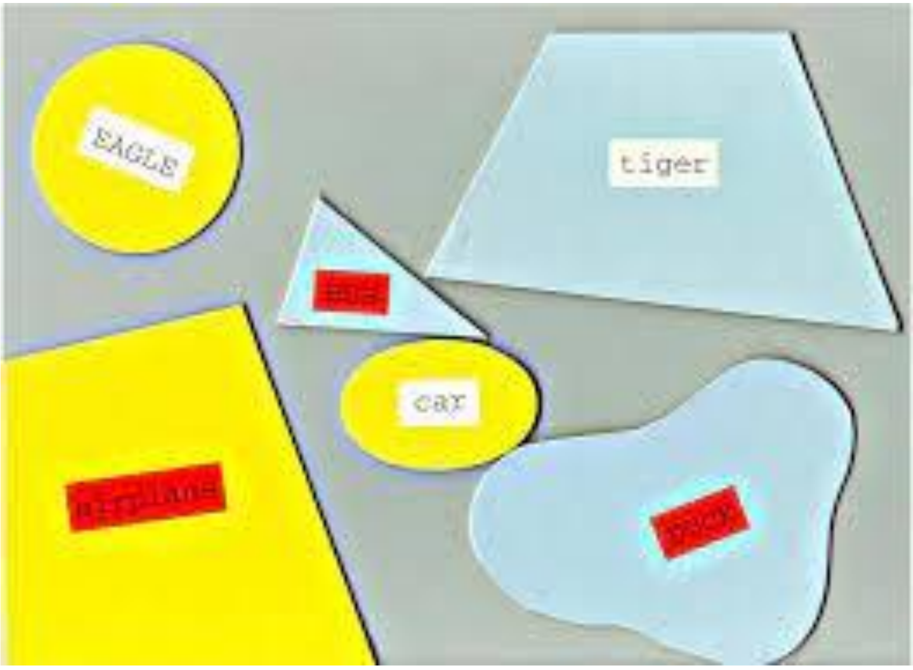


Figure 16: Sorting Test

Thereafter, the participants copied the **Rey Figure** (picture example Figure 17). This was not intended to be a memory task. It was intended to show in how much detail a participant looks at pictures, as some details need to be copied exactly.

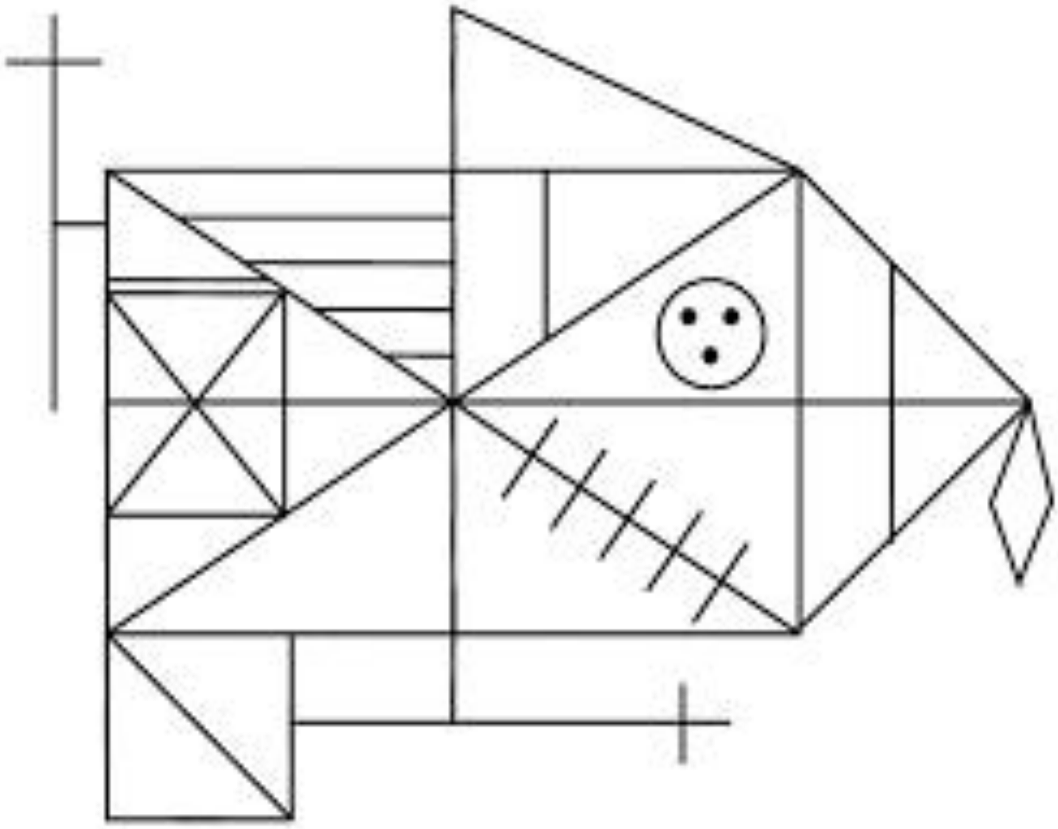


Figure 17: Rey Figure

The last distractor task was the **Graded Naming Test** (GNT; Warrington & McKenna, 1980). The GNT is a naming test that has, differently from other most commonly used naming tests, no ceiling effect. This was especially important in the present group of participants who were healthy adults from the age of 40. The GNT also provides an estimation of the size and retrieval of the participant's lexicon. Again, participants who show a better score in the GNT might use more diverse words to describe the picture and might also have fewer word-finding problems, a higher number of words per minute and might make fewer pauses in their description as they have better access to their lexicon.

Recruitment
Fill out Dyslexia Questionnaire

↓

Groups			
• Group 1 • 1st Testing • BCT spoken • Picture D written • 2nd Testing • Picture D spoken • BCT written	• Group 2 • 1st Testing • BCT written • Picture D spoken • 2nd Testing • BCT spoken • Picture D written	• Group 3 • 1st Testing • Picture D spoken • BCT written • 2nd testing • Picture D written • BCT spoken	• Group 4 • 1st Testing • Picture D written • BCT spoken • 2nd Testing • BCT written • Picture D spoken

↓

1st Testing				
Picture Description	ECAS	NAT	Raven's progressive matrices	Picture Description

↓

2nd Testing a week later				
Picture Description	Sorting Test	Rey Figure	GNT	Picture Description

Figure 18: Flowchart of task order for participants

3.3. Scoring and Analysis of Linguistic Variables

The obtained data were analysed by me. For the picture description task, the procedure presented in sections 2.5.3. and 2.5.4. was used. This makes the results more comparable to the different studies. Deviating from the previously described study is the number of symptoms analysed. Certain symptoms were added to the analysis to get a more diverse picture or pattern of symptoms. An overview of all symptoms in the different domains is presented in Tables 8-12. It is not expected that these will lead to a significant difference when analysing data of participants that do not have a specific language disorder, though the previous studies have shown that the distinction between different types of phonological and semantic paraphasias as well as disfluencies is important to get a more differentiated picture about symptom patterns in different types of neurological language disorders. An overview of the symptoms and their definition that were included in the linguistic analysis is presented in chapter 3.2.2.

Table 8
Spontaneous language and communicative behaviour

Observation	Description	Modality	Coding
Melodic line/intonation	Anomalies of sentence melody or intonation, unusual emphasis or stress on words	spoken	(un)impaired
Phrase length/uninterrupted run of words	Average length of sentences or utterance	Spoken and written	(Number of Clauses / Number of Words)
Sticking to topic	Is the participant sticking to the topic of the picture or deviating to different topics?	Spoken and written	(un)impaired
Adaptation to communication partner	Is the participant talking with an appropriate speech tempo? Is the participant using an appropriate level of vocabulary?	spoken	(un)impaired
Clarity	Unstructured utterances, illogical sequences	Spoken and written	(un)impaired
Coherence	Is the description systematic and logical?	Spoken and written	(un)impaired
Cohesion	Impairment of connection of phrases and sentences	Spoken and written	(un)impaired
Information content (Cookie Theft Picture only)	Count of correct Information Units according to BDAE (Kertesz, 1982)	Spoken and written	Count information units
Relevance	Relevant content, content that is directly relevant to the picture presented	Spoken and written Spoken and written	(un)impaired
Intrusion	Inappropriate repetition of previous content	Spoken and written	Count
Mutism	Failure to speak	spoken	(not) present
Logorrhoea	Uninhibited, exuberant speech production, even closed questions are answered excessively, it is hard to interrupt the participant	spoken	(not) present
Total number of words	Total word count	Spoken and written	Count

Total number of clauses	Number of complete clauses	Spoken and written	Count
Informational content per words (Cookie Theft Picture only)	Number of information units per 100 words	Spoken and written	(Information Units / Number of Words) *100
Information content for actions per words (Cookie Theft Picture only)	Number of information units of the action category per 100 words	Spoken and written	(Information Units of Actions / Number of Words) *100
Irrelevant content per words (Cookie Theft Picture only)	Number of irrelevant information units per 100 words	Spoken and written	(Irrelevant Content / Number of Words) *100

Table 9
Formulaic language

Observation	Description	Modality	Coding
Echolalia	Repetition of utterances of the communication partner, with or without changes in word order or wording	spoken	(not) present
Perseveration	Unintentional or inappropriate repetition of a word or clause that has been activated before although the context requires a new reaction, semantic and prosodic or morphosyntactic features can indicate a perseveration	Spoken and written	(not) present
Empty phrases	Overly used flowers of speech or cliches	Spoken and written	(not) present
Stereotypia	A word or empty phrase that is correct regarding context though repetitive and rigid	Spoken and written	(not) present
Automatisms	A word or empty phrase that is rigid, repeated constantly and neither lexically nor syntactically appropriate regarding context as well as against the intention of the communication partner	spoken	(not) present
Recurring utterances	Automatisms that consist solely of syllables, words or phrases that are strung together in a fluent way	spoken	(not) present

Table 10
Semantics

Observation	Description	Modality	Coding
Stagnation in fluency	Interrupted speech flow, speech tempo under 90 words per minute, average phrase length under five words	Spoken and written	(not) present
Interjections	Inappropriate insertion of inappropriate words	spoken	Count
Repetitions	Inappropriate repetition of words	Spoken and written	Count
Stereotypes	A word or empty phrase that is correct regarding context though repetitive and rigid	Spoken and written	Count
Discontinuation of sentences	A discontinued clause	Spoken and written	(not) present
Close semantic paraphasia	Incorrect occurrence of a word of the standard language that is moderately close to the meaning of the target word	Spoken and written	Count
Wide semantic paraphasia	Incorrect occurrence of a word of the standard language that deviates largely of the target word	Spoken and written	Count
Semantic conduite d'approche	Gradual semantic approach to the target word, usually as part of a self-correction process	Spoken and written	Count
Semantic conduite d'écart	Gradual semantic drift from the target word, usually as part of a self-correction process	Spoken and written	Count
Semantic neologism	Word composition that is not used in standard language (lashes -> eye feathers)	Spoken and written	Count
Semantic jargon	Pointless sequence of semantic paraphasias or neologisms and stereotypes in fluent speech production	Spoken and written	Count
Revisions	Form of self-correction	Spoken and written	Count
Semantic substitution	Replacement of a word with another word of the standard language	Spoken and written	Count
Not identifiable words	Words that cannot be comprehended or read	Spoken and written	Count
Crossed out letters	Letters that were crossed out by the participant	written	Count
Crossed out words	Words that were crossed out by the participant	Written	Count

Disfluency rate per word count	Any form of long pauses, repetitions, semantic substitutions, revisions crossed out letters or words as well as unidentifiable words	spoken	(Disfluencies / Number of Words) *100
Jargon syllable ratio	Rate of semantic and phonological jargon syllables per 100 words	Spoken and written	(Jargon Syllables / Number of Words) *100
Semantic substitution ratio	Rate of semantic substitutions per 100 words	Spoken and written	Semantic Substitutions / Number of Words) *100
Punctuation	Any wrong punctuation or missing punctuation	written	Count
Spelling	Any form of spelling errors	written	Count
Word onset errors	False capitalisation of words	written	Count
Regular print or cursive (variation?)	Form of writing used by the participant	written	Coded as print, cursive or mix

Table 11
Phonology

Observation	Description	Modality	Coding
Phonematic elision	A produced word is missing one or more sounds or letters (in any position)	Spoken and written	Count
Phonematic addition	One or more sounds or letters were added to a word (in any position)	Spoken and written	Count
Phonematic substitution	One or more sounds or letters were exchanged with another sound or letter in a word	Spoken and written	Count
Phonematic metathesis	Sounds or letters in a word were mixed up without changing, exchanging or missing a sound or letter	Spoken and written	Count
Phonematic conduite d'approche	Gradual phonological approach to the target word, usually as part of a self-correction process	Spoken and written	Count
Phonematic conduite d'écart	Gradual phonological drift from the target word, usually as part of a self-correction process	Spoken and written	Count
Phonematic neologism	Change of the sound of a word in a way it cannot be recognised anymore	Spoken and written	Count
Phonematic jargon	Pointless sequence of phonematic paraphasias (elision, addition substitution or metathesis) or neologisms in fluent speech production	Spoken and written	Count

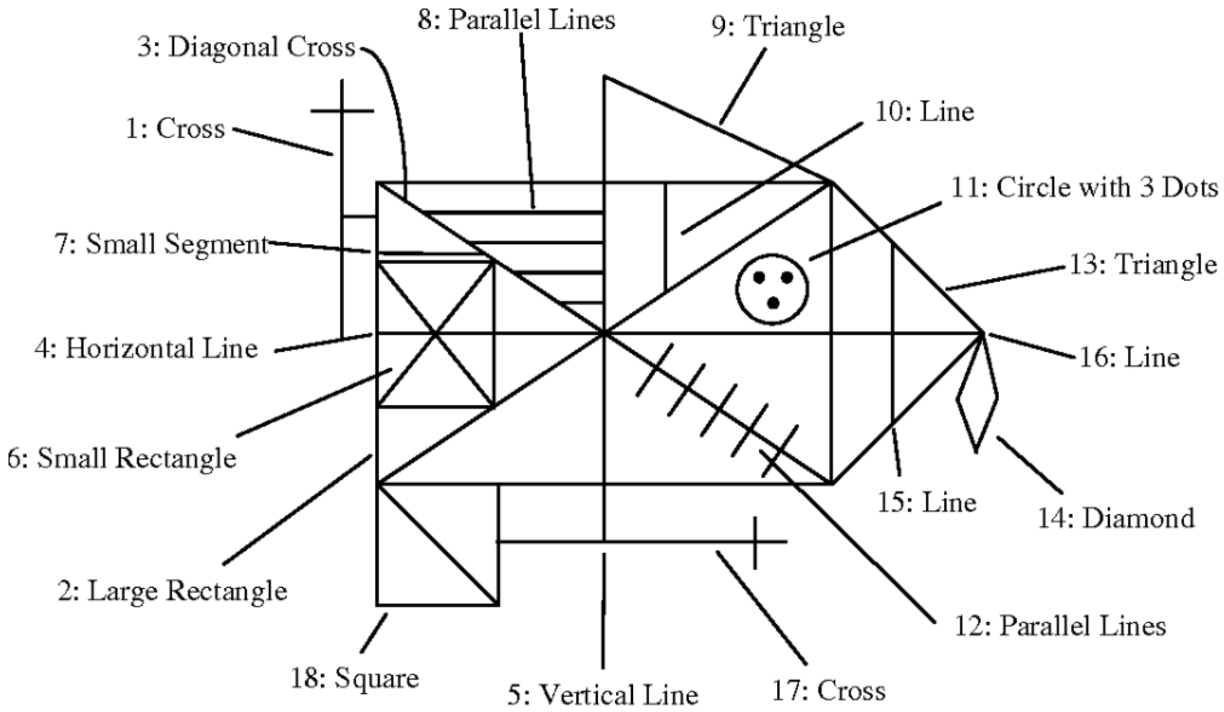
Table 12
Syntax

Observation	Description	Modality	Coding
Unimpaired	Clinically normal amount of syntax errors	Spoken and written	(not) present
Agrammatism	One or word sentences, short and easy or incomplete sentences, aborted sentences, missing function words or word inflection	Spoken and written	(not) present
Paragrammatism	Long and complex sentences, interleaved sentences, duplication of sentences, wrong word infelction	Spoken and written	(not) present

3.4. Scoring and Analysing of Distractor Tasks

The standardised tests were scored and analysed according to the regular testing procedure. This was the case for the respective parts of the ECAS, the short version of the NAT, the sorting test of Delis Kaplan and the GNT.

Some tests deviated from their standardised procedure as described above. As the entire standardised testing procedure of Raven's Progressive Matrices would have taken too long, especially regarding the fact that this test was intended to be a distractor task only controlling IQ, the test was shortened to 10 minutes, in which the participants solved as many puzzles as they could. This method was used following Creed et al. (1999). For the analysis correctly solved puzzles were counted, resulting in a number score. The Rey Figure also deviated from its original standardised testing procedure. As the Rey Figure task was not intended to test memory, participants copied the picture just once, while the picture was present in front of them. For scoring each detail/unit of the line drawing that was included and placed correctly was scored with two points. Figure 19 presents the different details with their points. Participants could score a total of 36 points.



Unit Correct	Placed Properly	2
	Placed Poorly	1
Unit Distorted, incomplete but recognisable	Placed Properly	1
	Placed Poorly	**1/2**

Figure 19: The Rey Osterrieth Scoring System

3.5. Statistical Analysis

The statistical analysis was done using three different statistic programmes: JASP (JASP Team, 2022), JAMOVI (The jamovi project, 2021) and R (R Core Team, 2016).

3.5.1. General Linguistic Variables

3.5.1.1. Number of Words

The first variable that was analysed regarding differences between the spoken and written picture description task was the total number of words. The paired samples t-test revealed that this difference was significant (t (29) = 7.57, p <.01). The results are visualised in Figure 20.

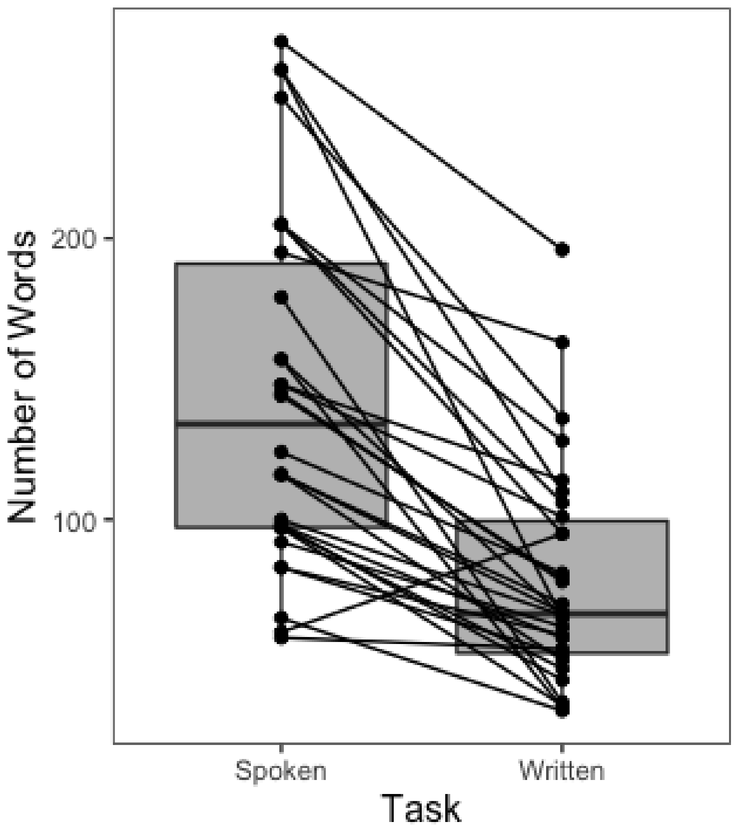


Figure 20: Boxplot results with individual results for each participant for Number of Words

3.5.1.2. Information Units

As described in chapter 3.2.2. the pictures can be described with a certain number of information units. Information units can be organised into four categories: subjects, places, objects and actions. Table 3 presents the 22 items in the different categories as in the BDAE (Kertesz, 1982). For the present analysis, the standardised version was used as described in chapter 3.2.2.

The t-test for differences in informational content for both tasks revealed no significant differences between spoken and written tasks (t (29) = .72, p = .48). For informational content in the action category, the Wilcoxon signed-ranks test indicated no significant difference between the spoken and written picture description task (z = 1.10, p = .135). The results are presented in Figure 21.

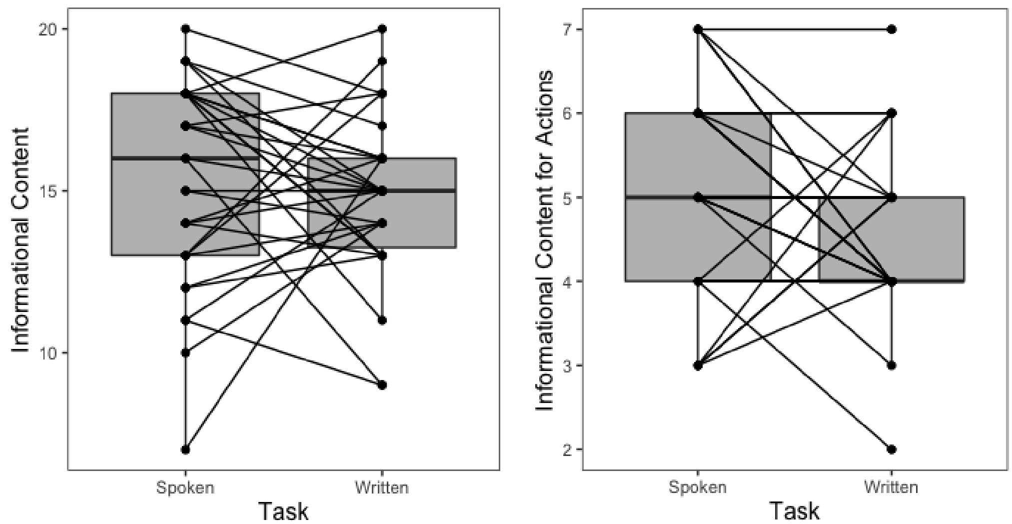


Figure 21: Boxplot results with individual results for each participant for Informational Content and Informational Content for Actions

3.5.1.3. Conciseness ratio

The conciseness ratio is calculated by dividing the number of information units by the number of words, multiplied by 100. It, therefore, represents the percentage of information units in a text. Through calculating and comparing the conciseness ratio it is possible to compare the density of information units of different pictures with an unequal number of total information units.

The participants used more information units per words in the written task (M = 22.1, SD = 8.4) compared to the spoken picture description task (M = 12.2, SD = 4.9).

The Wilcoxon test revealed a significant difference between the spoken and written tasks (z = 7.57, p <.001). The results can be seen in Figure 22.

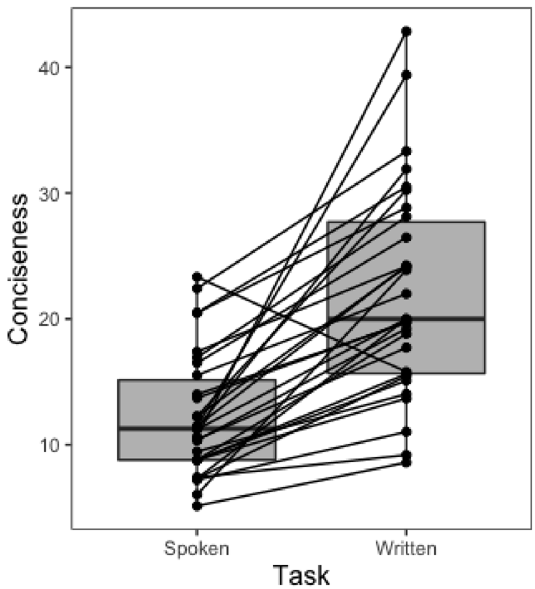


Figure 22: Boxplot results with individual results for each participant for Conciseness

3.5.2. Syntax

3.5.2.1. Number of (Main-)Clauses

The participants used a lower number of clauses to describe the Cookie Theft Picture in the written task (M = 5.8, SD = 3.2) compared to the spoken task (M = 9.8, SD = 4.5).

The paired samples t-test revealed a significant difference between the number of clauses in the spoken and written task (t (29) = 5.30, p < .001). Figure 23 visualises the results.

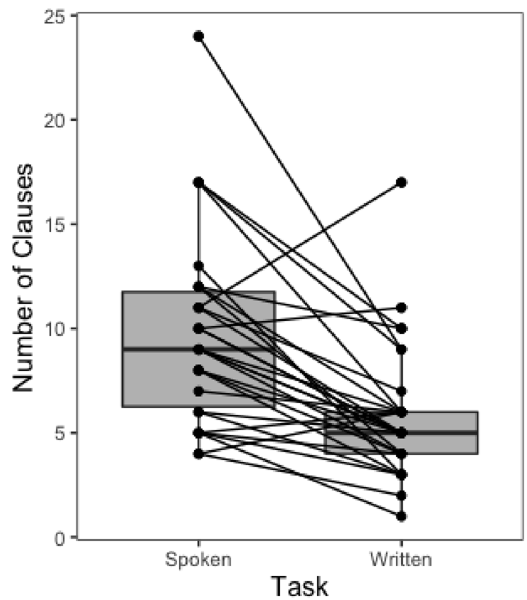


Figure 23: Boxplot results for with individual results for each participant for Number of Clauses

3.5.2.2. Number of Subclauses

Like the number of clauses, participants used more subclauses in the spoken picture description task (*M* = 3.0, *SD* = 2.3) compared to the written task (*M* = 2.4, *SD* = 1.4).

The paired Wilcoxon test revealed that the difference in the number of subclauses used in both tasks was not significant (*z* = 196.50, *p* = .598). Figure 24 depicts the results.

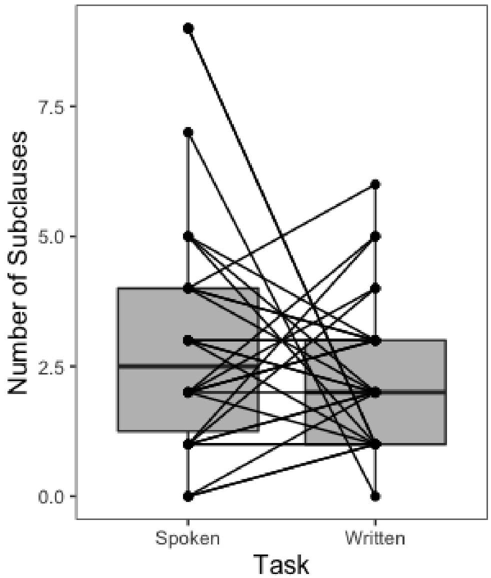


Figure 24: Boxplot results with individual results for each participant for Number of Subclauses

3.5.2.3. Number of Incomplete Clauses

The number of incomplete clauses can provide information about the complexity of a task. This effect has been reported in L2 language performance (Michel, 2011; Frear & Bitchener, 2015).

The number of incomplete clauses is slightly higher in the spoken picture description task (*M* = 2.3, *SD* = 0.5) compared to the written task (*M* = 1.3, *SD* = 0.4). Two participants produced an incomplete clause in the oral task whereas only one participant produced an incomplete clause in the written task. One participant produced two incomplete clauses in each task respectively.

Hence, the Wilcoxon test revealed no significant difference between the oral and written tasks (*z* = 28.5, *p* = .492), as the number of produced incomplete clauses is extremely low with most participants producing no incomplete clauses.

3.5.4.4. Syntactic Complexity

The syntactic complexity was calculated by dividing the number of subclauses by the total number of clauses multiplied by 100. Therefore, it presents the percentage of subclauses a participant used in their description.

Participants produced about 72 percent more subclauses per main clauses in the written task (M = 51.9, SD = 54.2) than in the oral task (M = 37.8, SD = 31.6). One participant even used only one main clause with three subclauses, resulting in a score of 300.

However, the Wilcoxon test revealed that the difference in syntactic complexity between both tasks was not significant (z = 152.00, p = .160), which might be due to the outstanding performance of the participant described previously. The results are displayed in Figure 26.

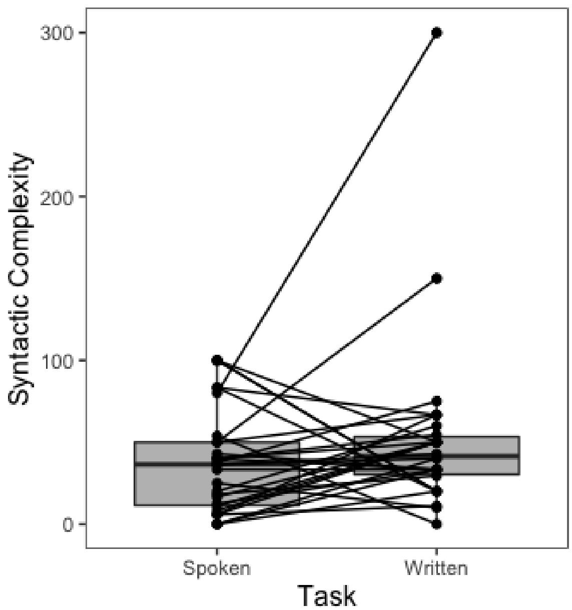


Figure 25: Boxplot results with individual results for each participant for Syntactic Complexity

3.5.4.5. Syntactic Errors

Finally, participants only rarely produced syntactic errors. On average participants produced the same number of syntactic errors in the oral and written picture description task. The difference between the spoken and written task was not significant (z = 104, p = .73).

The syntactic error ratio was on average the same in both the spoken and written picture description task. As a result, the Wilcoxon test revealed no significant difference between both tasks. Results are presented in Figure 27.

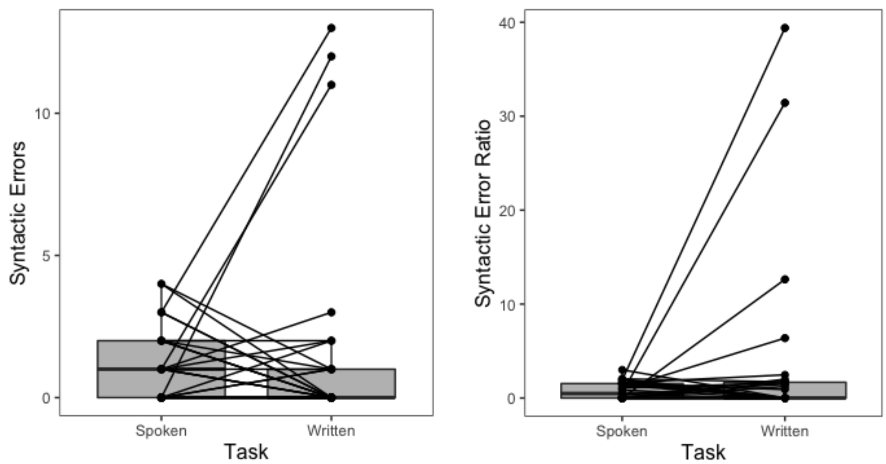


Figure 26: Boxplots results with individual results for each participant for Syntactic Errors and Syntactic Error Ratio

3.5.3. Semantics

Semantics in this study encloses the lexical diversity participants presented in their picture description samples and the correct use of words (lexical errors) as well as the certainty with which participants used words in their descriptions (disfluency rate).

3.5.3.1. Lexical Diversity

For lexical diversity, the Type-Token-Ratio (TTR) was calculated, as it is a widely used and stable measure of lexical variation in spoken and written texts. In addition, the Measure of Textual Lexical Diversity (MTLD) was calculated as it takes lexical redundancy into account (Scott, 2013). This is especially important for this study as the spoken and written picture description samples are not equal in length as described in chapter 3.5.1.

Participants had a higher TTR in the written task (*M* = 67.3, *SD* = 11.9) compared to the oral task (*M* = 59.5, *SD* = 8.2) of the Cookie theft Picture. Exact values can be taken from Figure 28. The paired samples t-test revealed a significant difference in the TTR (t (29) = -3.68, p <.001) between the spoken and written picture description tasks.

However, the MTLD presents a different picture. The participants had a similar MTLD score in both, the spoken (*M* = 51.1, *SD*, 11.2) and written (*M* = 55.3, *SD* = 32.8) picture description task. The Wilcoxon test revealed no significant difference (z = 247.00, p = .778) between both tasks. Figure 28 shows the results.

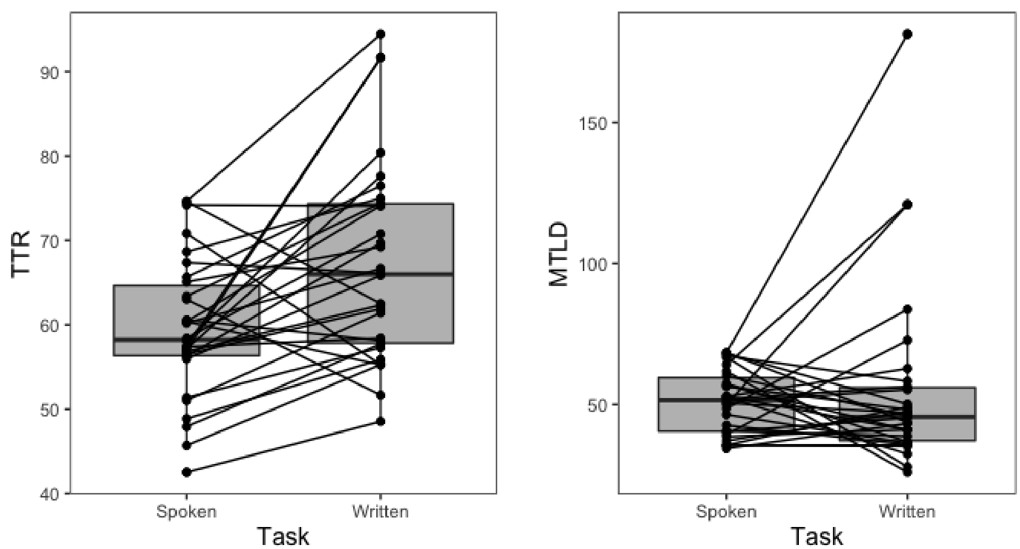


Figure 27: T-test results with individual results for each participant for Type-Token-Ratio and Measure of Textual Lexical Diversity

3.5.3.2. Lexical Errors

In this study lexical errors are defined as the deviation of the meaning of the target word (Llach, 2011). This includes close and wide semantic paraphasias, semantic conduite d’approche and d’écart, semantic neologism, semantic jargon, semantic repetitions and semantic substitutions. In addition, a jargon syllable ratio and semantic substitution ratio were calculated and analysed. A definition of all lexical errors can be found in Table 10.

The participants did not produce any semantic paraphasias, semantic conduits d’approches or d’écarts, semantic neologisms, semantic jargon syllables or semantic repetitions. The only lexical error produced by participants was semantic substitution, which means they exchanged a word with a word that exists in English though has a different meaning. However, this error occurred very rarely. Furthermore, the Wilcoxon test revealed no significant difference between both tasks ($z = 2$, $p = .773$).

As the participants did not produce any jargon syllables only the semantic substitution ratio was analysed. Again, the values are slightly higher in the written task of the picture description task. The Wilcoxon test revealed once more no significant differences between both tasks ($z = 2$, $p = .789$).

3.5.3.3. Disfluencies

As described in chapter 3.2.4. a detailed analysis of disfluencies would go beyond the scope of this study, in this study disfluencies are defined as the uncertainty of word choice in participants. About 10% of daily produced utterances contain natural disfluencies that are mostly made up of repetitions, revisions, filled and unfilled pauses as well as other hesitations. These are mainly due to underlying problems in the formulation and planning of upcoming speech (Schnadt, 2009).

Additionally, to compare the results of this study to the studies of Rutter (2014) and Lingscheid (2018) described in chapter 2 the disfluency ratio was calculated. The disfluency ratio comprises the sum of conduites d'approche, repetitions, revisions and fillers divided by the word count and multiplied by 100.

The participants produced overall more disfluencies in the oral picture description task compared to the written picture description task. In the spoken task, the disfluencies produced by participants were by far the most filled pauses (M = 4.8, SD = 4.7) whereas no silent pauses were produced. In the written picture description task, the disfluency that was produced most by participants was spelling errors (M = 0.6, SD = 1.5). In the written task participants did not produce any revisions. The exact values of the presented and remaining variables are presented in Tables 13 and 14. The detailed analysis of disfluency types makes clear that spoken and written language is possibly prone to different types of disfluencies. While participants used filled pauses to win more time to think about what they would like to say next, in the written task there is no need to fill the empty gap as the production speed is lower and it is less obvious to just pause the language production for some time.

The disfluency ratio was significantly lower in the written task (t (29) = 3.03, p <.001). However, Figure 28 shows that this is not the case for every participant. Most participants produced fewer disfluencies per word in the written task, though about a third of participants had a higher disfluency ratio in the written task.

Table 13
Disfluencies in the spoken task

	Repetitions	Discontinuations	Revisions	Non-Identifiable Words	Filled Pauses	Silent Pauses
Valid	30	30	30	30	30	30
Missing	0	0	0	0	0	0
Mean	0.633	0.633	0.500	0.100	4.767	0.000
Median	0.000	0.000	0.000	0.000	4.000	0.000
Std. Deviation	1.273	0.964	0.900	0.305	4.739	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	5.000	3.000	3.000	1.000	17.000	0.000

Table 14
Disfluencies in the written task

	Repetitions	Discontinuations	Revisions	Non-Identifiable Words	Crossed-Out-Letter	Crossed-Out-Word	Punctuation Error	Spelling Error	Word Onset Error
Valid	30	30	30	30	30	30	30	30	30
Missing	0	0	0	0	0	0	0	0	0
Mean	0.000	0.033	0.000	0.167	0.200	0.300	0.633	0.633	0.200
Median	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Std. Deviation	0.000	0.183	0.000	0.592	0.664	0.651	1.564	1.564	0.664
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	1.000	0.000	3.000	3.000	3.000	7.000	7.000	3.000

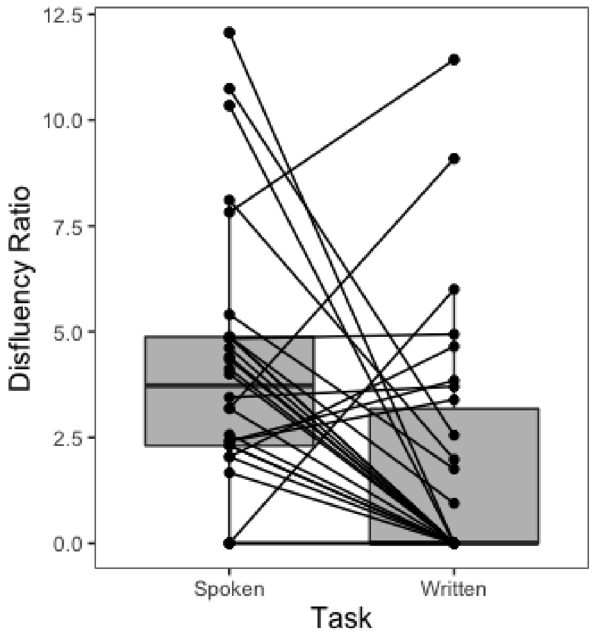


Figure 28: Boxplot results with individual results for each participant for the Disfluency Ratio

3.5.4. Phonology

Phonology regards the sound pattern of language and therefore focuses on one of the smallest units of language. On this level language is usually impaired in form of a deficit in language development, auditory perception and processing disorder or after a stroke or the onset of a neurodegenerative disease (Munson, 2001). Phonological symptoms include elisions, additions, substitutions and metathesis of one or more phonemes of a word as well as an approach or departing from the target word on the sound level. Furthermore, the phoneme of a word can be changed in a way that the target word cannot be recognised anymore. If the produced language consists mostly of phonological paraphasias and neologisms, phonological jargon is present (Huber et al., 2013).

As presented in Tables 14 and 15 phonological errors occurred very rarely in the spoken and written picture description tasks of the Cookie Theft picture. The only phonological error that was produced in the oral picture description task was the phonological conduit d'approche (*M* = 0.2, *SD* = 0.4). In the written task participants only produced phonological metathesis in their descriptions (*M* = 0.1, *SD* = 0.2).

The numbers were too small to reveal any significant differences.

Table 14
Phonological errors in spoken descriptions

	Phonological Elision	Phonological Addition	Phonological Substitution	Phonological Metathesis	Phonological Conduit D'approche	Phonological Conduit D'ecart	Phonological Neologism	Phonological Jargon
Valid	30	30	30	30	30	30	30	30
Missing	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.000	0.000	0.233	0.000	0.000	0.000
Std. Deviation	0.000	0.000	0.000	0.000	0.430	0.000	0.000	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000

Table 15
Phonological errors in written descriptions

	Phonological Elision	Phonological Addition	Phonological Substitution	Phonological Metathesis	Phonological Conduit D'approche	Phonological Conduit D'ecart	Phonological Neologism	Phonological Jargon
Valid	30	30	30	30	30	30	30	30
Missing	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.000	0.033	0.000	0.000	0.000	0.000
Std. Deviation	0.000	0.000	0.000	0.183	0.000	0.000	0.000	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000

3.6. Discussion

To summarise, 30 participants were recruited and completed a total of four picture description tasks as well as seven distractor tasks that will also deliver important information on how connected language and isolated language functions, as well as cognitive abilities, are connected in healthy adults. In this chapter, the oral and written description of the Cookie Theft Picture were analysed and compared.

The group consisted of more female than male participants. This might skew the results, however not to a great extent. Females have a slight linguistic advantage over males, but effect sizes are small and there is little variance in the normal population (Hyde & Linn, 1988; Leaper & Ayres, 2007; Newman et al., 2008; Barry et al., 2008; Voyer & Voyer, 2014). The results of a previous statistical analysis after the in-person testing session revealed that female and male participant did not produce a significant different number of words (spoken: t (29) = 0.24, p = .81; written: t (29) = -0.05, p = .96) or a significant difference in the MTLD (spoken: z = 107.00, p = .93; written z = 76.00, p = .23) Therefore, gender differences can be neglected if the focus lies on the general healthy population. Furthermore, in neurodegenerative diseases, no gender differences could be observed in linguistic symptoms (Hodges & Patterson, 2007; Laws et al., 2010; Grossman, 2010; Onyike & Diehl-Schmid, 2013). Poststroke aphasia has a higher prevalence in women, however, this effect is most likely an age-effect (Wallentin, 2018).

The age of the participants was well distributed across the sample with a peak of participants aged 70. Therefore, the sample can give a broad overview of the performance of healthy participants that are the same age as people that might be affected by neurodegenerative diseases or post-stroke aphasia. A more detailed description of the age at onset is given in chapter 1.3.

The educational level of participants was a little bit higher than compared to the general population. In the presented sample 63 per cent of the participants attained some form of post-secondary education compared to 49.4 per cent in the general population in the UK. Most participants, 30 per cent, attained a bachelor's degree compared to 25.5 per cent in the general population of the UK in 2021. Ten per cent of the participants held a master's degree compared to 12.9 per cent in the general population (Organisation for Economic Co-Operation and Development, 2021). The differences might be due to the selected group. All participants were enrolled with the volunteer panel of the University of Edinburgh Psychology Department. Participants might be more likely to have attended the University of Edinburgh, might be the partner of a University of Edinburgh alumni or the parent of a student at the University of Edinburgh. Therefore, some results need to be interpreted with caution and cannot be generalised to the general population in the UK.

In the picture description tasks, four main domains were analysed: General length and completeness, syntax, semantics and phonology. T-tests and Wilcoxon tests were conducted to analyse differences between spoken and written picture description tasks.

In general length, a significant difference between the spoken and written task was revealed for the number of words participants used to describe their pictures. These findings support previous data from Rutter (2014) and general differences between spoken and written language production described in more detail in chapter 2.3. No significant differences were found in the use of information units and information units for actions. This shows that although participants used significantly more words in the spoken picture description task, they still used the same amount of information to describe the Cookie Theft Picture in both tasks. Consequently, the conciseness ratio was significantly higher in the written task.

The results in the domain-general length and completeness were also reflected in the syntactic domain. As with the number of words, the participants used significantly more sentences and sub-clauses in the oral picture description task. However, no significant differences were found in the number of incomplete clauses, syntactic errors or syntactic complexity. The participants might be more focused on the written task leading to more concise picture descriptions.

In the domain of semantics, some significant differences between spoken and written picture descriptions were revealed. Although the Type Token Ratio (TTR) did reveal a significant difference between the spoken and written task the Measure of Textual Lexical Diversity (MTLD) revealed no significant differences. The TTR and MTLD present a different pattern upon which conclusions can hardly be drawn. The results of the TTR correspond to the differences found in previous literature as presented in chapter 2.3., whereas the MTLD revealed no significant differences between spoken and written language. However, in previous literature, only the TTR was used to test for differences in lexical diversity between spoken and written language. Furthermore, the TTR is still used in diagnostic practice. Regarding the results of this analysis, it might have been the wrong approach as written language is usually shorter. To gain a more conclusive picture, an analysis of TTR and MTLD with more data from spoken and written language samples describing the same topic will be needed. Lexical errors were only produced rarely by participants and no significant differences between oral and written picture descriptions were found. Healthy participants produce lexical errors very rarely. Lexical errors might occur more often in written samples, though according to the results not to a significant extent. Further research could investigate the number of lexical errors in a larger sample to create more robust data for healthy language. Compared to previous studies by Rutter (2014) and Lingscheid (2018) a low number of lexical errors might be a pattern of healthy language compared to patients with neurodegenerative diseases or aphasia. However, the group of participants was very small.

Disfluencies occurred significantly more in the oral picture description task. Rutter (2014) and Lingscheid (2018) also observed a tendency for a higher disfluency ratio in the spoken picture description task. This goes following the assumption that participants might feel the urge to fill the silence in the spoken picture description task with fillers when they are looking for a word or more information by using fillers such as “uhm”. Further analysis showed that fillers were mostly used at the beginning of sentences. In the written picture description task, there is no need to fill the silence. Therefore, spelling mistakes are more apparent. However, it is not clear whether spelling errors are a sign of uncertainty or if they might be caused by dyslexia or the increasing reliance on technical support. This question will be discussed in more depth in chapter 5.

In the domain of phonology, no analytical test could be conducted as the participants produced a too-small number of phonological errors. This is following the data of Rutter (2014) and Lingscheid (2018). According to the presented data, phonological errors occur very rarely in healthy adults and might therefore be the first sign of pathological processes.

3.6.1. Implications of spoken vs written administration in the diagnosis of PPA

The administration mode of picture description tasks can influence the diagnosis in people with PPA. Previously data of healthy participants were reviewed. In this section, possible implications for the diagnosis of people with PPA will be discussed.

The domain of general length would present differently in people with PPA. A core symptom of lvPPA and nfvPPA is a reduced speech rate or effortful speech. In svPPA the general length seems not to be affected. The findings of Henry et al. (2012) indicate that spoken language performance is strongly correlated with reading and spelling profiles in people with PPA. Therefore, people with lvPPA and nfvPPA might produce shorter spoken as well as written picture descriptions whereas the length of picture descriptions in svPPA might remain unchanged in both modalities.

In the healthy participants, the conciseness ratio was significantly higher in the written task. This observation might not uphold in lvPPA and svPPA. People with lvPPA have word retrieval deficits which might lower the number of information units in both modalities. svPPA would present with its core symptoms in the number of information units and conciseness ratio. As people with svPPA present with anomia and paragrammatism they would for example use the word child for both, the girl and the boy which could also be observed in both spoken and written picture descriptions according to Henry et al. (2012). On the other side, nfvPPA is characterised by short and simple phrases. Therefore, a higher conciseness ratio might be observed in both modalities as word retrieval is not affected.

Regarding the syntactic domain, nfvPPA can be distinguished best from lvPPA and svPPA by its short and simple sentences. People with nfvPPA will produce simple syntax structures in both modalities. Syntax production in nfvPPA is also affected by complexity. Therefore, a low syntactic complexity ratio will be found in people with nfvPPA. A symptom of svPPA is paragrammatism. Paragrammatism is a variable in the syntax category. Healthy adults as well as people with lvPPA and nfvPPA do not produce paragrammatic structures. Therefore, paragrammatism would be a symptom that will distinguish svPPA best from other forms of PPA. These effects could be observed in both spoken and written picture descriptions. In lvPPA sentence repetition is mostly affected which would not be tested in a picture description task. Therefore, people with lvPPA might not show any symptoms in the syntax category.

The lexical domain revealed mixed results in healthy participants for spoken and written picture descriptions. In PPA, the language of people with svPPA becomes empty which means that the lexicon gets reduced. People with svPPA would produce picture descriptions with a low TTR and MTLD in both modalities. lvPPA is characterised by lowered word retrieval which might also lead to a lowered TTR and MTLD. This might make it hard at first to distinguish svPPA from lvPPA. However, in this case the type of words used needs to be observed. In svPPA fewer unique words and more pronouns are found. People with lvPPA have problems retrieving single words which might lead to circumlocutions.

nfvPPA will not lead to a lowered TTR or MTLD as symptoms are more located on the syntactic and phonetic-phonological levels.

Phonetic and phonological errors were very rare in healthy participants in both modalities. However, the phonological level is well suited to distinguish different types of PPA. People with lvPPA would not be affected on the phonological level people with nfvPPA and svPPA can be distinguished by the type of errors. Phonetic and phonological errors are a symptom of nfvPPA and would distinguish people with nfvPPA from people with lvPPA. A symptom that is special to svPPA is surface dyslexia and dysgraphia expressed through the regularisation of words both in spoken and written language. Errors in nfvPPA are inconsistent and people with nfvPPA are aware of their errors while errors in svPPA follow a certain logic and are consistent. People with svPPA, however, seem not to be aware of their errors.

While disfluencies were rarely produced by healthy participants, disfluencies are a core symptom of lvPPA. As word retrieval is affected people with lvPPA have a reduced speech rate with filled pauses when they are looking for the word they want to use. Speech production in nfvPPA is also effortful and slow sometimes with pauses in the middle of sentences or words. People with svPPA, however, will not produce disfluencies. Although Henry et al. (2012) suggest that symptoms would be presented in both modalities, disfluencies might not be detected in the written administration mode as written language does not detect unfilled pauses and filled pauses as well as a reduced speech rate.

Overall, the oral administration of picture descriptions might be suited better for diagnosing people with lvPPA as core symptoms such as word retrieval deficits, reduced speech rate, filled pauses and revisions occur more in spoken language. Written picture description tasks, however, could be used well for diagnosing nfvPPA and svPPA. nfvPPA can be distinguished best from other subtypes of PPA by its inconsistent phonetic/phonological errors and short simple phrases. People with svPPA could be diagnosed through their surface dyslexia and dysgraphia and regularizations in the written administration.

3.7. Conclusion

The analysis revealed that the spoken picture description samples are in general longer and consist of more words as well as sentences. However, this does not apply to the information that is conveyed. Participants used on average the same number of information units in both tasks. As a result, the written picture descriptions were more concise than their spoken counterparts. Furthermore, the written language samples were also more complex in syntax and presented a higher lexical diversity. Written language is in general more complex and denser in information. This phenomenon was also described in previous research as presented in chapter 2.3. A reason for this observation might be, that written language has a lower production speed compared to spoken language. As a result, participants can focus more on the complexity and correctness of their written language production.

No significant differences were revealed between spoken and written language samples. This was also found in the results of previous studies on healthy participants (Rutter 2014; Lingscheid, 2018).

As a result, regarding data from healthy participants, spoken and written language samples could be used in the diagnosis of pathological changes in language. The results in chapters 3 and 4 shows that pathological language will include a higher number of syntactic, lexical and phonological errors. However, more research on patients with neurodegenerative diseases such as primary progressive aphasia is needed to confirm this hypothesis.

It is also not clear how far the performance on connected language production and especially picture description tasks is influenced by other isolated language functions and cognitive abilities. The next chapter examines the possible influence of isolated language functions and cognitive abilities on the spoken and written picture description task of the

Cookie Theft picture with the results of the tests that were presented as distractor tasks in this chapter.

For clinical work, it is important that in healthy individuals spoken and written picture descriptions vary mostly in their length, informational density and lexical diversity. However, phonological, semantic and syntactic symptoms might not change in both modalities. Further studies need to investigate this hypothesis. Should the hypothesis be confirmed it might mean that written picture descriptions might be sufficient for the diagnosis of language impairment in people with neurodegenerative diseases or post-stroke aphasia. In PPA for example written picture description could help distinguish the different variants. People with lvPPA would produce descriptions that are shorter and contain circumlocutions and revisions. Descriptions of people with nfvPPA would consist of short and simple sentences that might also contain irregular phonetic and phonological errors. The written sample of people with svPPA would be normal in length, however with a low TTR/MTLD and characteristic regularisations of words. If these assumptions would be confirmed in further studies, the use of written picture descriptions would lead to a significant cost reduction in the diagnosis process that could be used for more specific therapeutic interventions.

4. Relationship between picture description and other language and cognitive functions

After the previous chapter discussed the differences between oral and written administration of the Picture description task, this chapter will analyse the influence of other language and cognitive functions on the picture description task. Isolated language functions such as naming, syntax and possible underlying dyslexia can influence the performance on connected speech tasks such as a picture description task directly. Concept formation, visuospatial abilities, social cognition and intelligence can also have an impact on connected language production.

Thirty healthy participants completed seven language and cognitive tests as well as four picture description tasks in total. Correlation analyses were conducted to determine the influence of naming, syntax, possible dyslexia, concept formation, visuospatial abilities, social cognition and intelligence on oral and written picture description tasks of the Cookie Theft Picture and the Indian Street Scene.

The analyses revealed that isolated language functions and other cognitive functions have a limited influence on connected language production in written picture description tasks. However, the analysed group size was small, and some analyses were conducted with an exploratory method. Further research in larger groups of healthy adults but also adults with neurodegenerative diseases is needed to confirm the results of this study and to reveal the implications for healthy ageing and pathological mechanisms.

4.1. Factors that influence picture description tasks

Several factors can influence the outcome and symptomatic pattern of a picture description task. In this chapter, the focus lies on the influence of language and cognitive functions on connected language production in picture description tasks. Thirty healthy participants took part in an in-person testing session at the University of Edinburgh, as described in chapter 3. They completed four picture description tasks in total. The participants produced a spoken and a written picture description for the Cookie Theft Picture from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass et al., 2000) and an Indian Street Scene. Between the picture description tasks, the participants completed different tests assessing further language and cognitive functions, namely naming, syntax, possible dyslexia, concept formation, visuospatial abilities, social cognition and intelligence.

The following section will briefly review previous evidence on the influence each of the presented functions might have on picture description tasks and connected language production.

4.1.1. Naming

Naming can give an estimate of the size of a person's lexicon and their ability to access words. For this study, the Graded Naming Test (GNT; Warrington & McKenna, 1980) was chosen to assess naming as it has no ceiling effects in healthy participants. Some previous studies have investigated the correlation between naming and picture descriptions in patients with different types of aphasia whereas other studies evaluated differences in people with dementia. Although a neuroanatomical study suggests that naming and picture description tasks use different constructs in the temporal lobe and are therefore differently affected by the disease (Emerton et al., 2014), research in patients has provided mixed evidence on their relationship to aphasia.

Different patients with aphasia presented indeed with different symptom patterns. Patients with Broca's aphasia performed significantly better on the naming task compared to the picture description task whereas patients with Wernicke's performed significantly better on the picture description task (Williams & Canter, 1982; Emerton et al., 2014). No significant differences between both tasks were found in patients with amnestic or conduction aphasia (Williams & Canter, 1982; Emerton et al., 2014). However, other researchers could not confirm differences between syndromes. Patients with fluent and non-fluent aphasia did not perform significantly different in confrontation naming and picture description tasks in a study by Basso et al. (1990).

Naming abilities might however influence the informativeness and general length of connected language samples (Giles et al., 1996; Fergadiotis & Wright, 2016; Boucher et al., 2020). Naming abilities and informativeness of connected language tasks decreased in patients with Alzheimer's with increasing severity (Giles et al., 1996; Fergadiotis & Wright, 2016).

From previous research on people with dementia and aphasia, it can be concluded that the performance on naming tasks and picture description tasks might be correlated. A better-connected lexicon and better word retrieval might improve the communication of information and might also reduce semantic errors and disfluencies. However, the direction of the correlation is not always clear for different forms of language impairment. Data from healthy individuals on this relationship has not been reported. According to the presented studies naming abilities might be correlated with informativeness, length and lexical diversity of picture descriptions. Furthermore, participants with better access to their lexicon might produce fewer symptoms that could be interpreted as searching behaviour or self-corrections. Searching behaviour could be, for example, fillers, filled or unfilled pauses and repetitions. Self-corrections can be expressed by revisions or conduites d'approche as well as crossed-out letters and words in written language.

4.1.2. Syntax

The syntax is strongly interconnected with sentence production and therefore connected speech; however, no study has researched the influence of grammatical abilities on picture description tasks with separate designated tests yet. So far two studies have investigated the effect of ageing on syntactic structures in connected speech. Older adults are less efficient in forming syntactically complex representations which might be largely influenced by limitations in working memory (Salis, 2011). Furthermore, mean length of utterance, lexical-semantic processes and local coherence worsen with age, though age seemed to not affect phonological and syntactic abilities (Frau et al., 2021).

In the present study, participants completed the short version of the Northwestern Anagram Test. The test was chosen as it has a short duration and does not rely on memory or writing abilities, as the words for the sentence the participant has to build are already present on cards.

As so far, no studies have researched syntactic abilities and connected speech in isolated designated tests, and no assumptions can be made about whether syntax and measures of picture description tasks are correlated.

4.1.3. Dyslexia

Dyslexia can also influence the production of spoken and written speech. Usually, dyslexia is more related to written speech, however, symptoms can also influence oral connected language (Catts et al., 2005). People with dyslexia have difficulties spelling words and a lack of expression as well as general difficulties in planning and revising different types of texts. They can also confuse similar letters and words and often avoid reading and writing if possible. (NHS, 2018). Furthermore, some people with dyslexia might find it hard to remember for example PIN or telephone numbers or meet deadlines but can also often have a remarkably good memory which resulted as a compensation mechanism to avoid reading in school (NHS, 2018).

Research on dyslexia is usually concerned with dyslexia in childhood and its development. Symptoms usually persist into adulthood however, people with dyslexia often have learned strategies and compensation mechanisms that can help with written tasks (NHS, 2018). Research is mostly focused on the phonetic domain (Boets et al., 2013; Vandermosten et al., 2012), rather than on connected speech.

Before their first testing session, the participants of this study completed a questionnaire that investigated their reading behaviour and experiences during school time. The questionnaire is a shortened version of the Adult Reading History Questionnaire (ARHQ; Lefley & Pennington, 2000) and can be found in Appendix C. The ARHQ is a self-screening tool that measures the risk of reading disability. The questionnaire consisted of eight rating scale questions and two multiple choice questions, one of which asked for more details if someone in the family has or had problems with reading or spelling. A higher score indicates possible underlying dyslexia. The ARHQ has cut-off scores that are validated and reliable. However, the questionnaire of the present study is a shortened version and therefore the cut-off scores cannot be applied. Yet, the goal of the presented analysis is to estimate possible correlations between a higher score in the dyslexia questionnaire and the performance on picture description tasks. Therefore, a cut-off score was not necessary. Participants who report greater difficulty with reading and writing might use fewer words in the written picture description part, apart from producing potentially more phonetic errors. Furthermore, the family history concerning dyslexia was assessed.

The presented data shows that a higher score on the Dyslexia Questionnaire might be correlated with a higher number of different phonetic errors and less text length as well as less complex syntactic structures.

4.1.4. Concept Formation

Concept formation is the process in which a person abstracts a common idea from one or more examples and learns defining features that are characteristic of a class (for example those defining a dog) or necessary to identify members of a class of objects (square), relations (between) or actions (running) (Fine et al., 2009). Concepts are needed in language to organise the semantic lexicon. Concept formation skills help structure words into categories and subcategories. This not only applies to verbalized language but also to sign language (Furth, 1961).

Better concept formation and executive functioning skills are also connected with better therapy outcomes in anomic aphasia patients. Dignam et al. (2017) tested 34 adults with chronic aphasia who participated in Aphasia Language Impairment and Functioning Therapy. They found that scores of the sorting test from the Delis-Kaplan Executive Function System and anomia therapy outcomes correlated positively. There seems to be a connection between word encoding and categorisation. It is not clear whether better concept formation skills and therefore categorisation abilities lead to better word retrieval as well. The presented studies underline how closely concept formation and language are interconnected. However, the influence of concept formation abilities on connected language production and picture description tasks has not been assessed yet.

In this study, participants completed the sorting test from the Delis-Kaplan Executive Function System; D-KEFS (Delis et al., 2001). The D-KEFS sorting test is a reasoning and categorisation task. The participants have to build categories and explain their reasoning for the categories shortly. Therefore, the test can show how well participants can form categories and how flexible the categories are they already have built previously in life. As described above categories are an important part of languages. The number of categorisations the participants formed correctly was counted as the score they achieved.

Due to the presumed relationship between concept formation and language use, participants that have good general language skills may have higher scores on the sorting test. Furthermore, according to previous research, participants that have a lower score on the sorting test might have a higher number of semantic symptoms.

4.1.5. Visuospatial Abilities

Visuospatial abilities might influence picture description tasks. Visuospatial abilities likely have an impact on the perception and analysis of a picture which will subsequently influence language production. In addition, there is some evidence for a more direct relationship between visuospatial abilities and language abilities from previous research.

Visuospatial abilities seem to play a key role in language development. Children with specific language impairment had a significant deficit in both visuospatial storage and visuospatial central executive tasks (Vugs et al., 2013). Hence, development deficits in visuospatial abilities and language might be connected. However, it is not clear whether this relationship persists into adulthood.

Visuospatial abilities are also important for processing demonstratives ("where?", "this one"). It has been suggested that language processing relies on neural resources for perception, attention and extra-linguistic cognition (Rocca et al., 2020). Participants that have a lower score in the Rey Complex Figure task might produce more semantic paraphasias or grammar errors resulting from problems in processing demonstratives in particular.

Visuospatial processing was tested with the Rey Osterrieth Figure Test. Participants copied the image once. A point was given for each correct detail and another point was given for the correct placement. In total 36 points could have been reached. Appendix D presents the complex Figure with its points.

4.1.6. Social Cognition

Social cognition or Theory of Mind (ToM) is closely connected with language. Previous research suggests that social cognition is a distinguishing feature of language and communication. Language is a facilitator for social interaction (Seyfarth & Cheney, 2014) but is also necessary for producing ToM (Garfield et al., 2002; Marton et al., 2005). Furthermore, in adolescents, problems with language structure were a predictor variable for difficulties with social problem solving (Goodman & Stuhlmüller, 2013).

Given the connection between social cognition, ToM and language, ToM may be an important factor for generating and updating a mental model of the listener's knowledge of the picture and thus constructing a coherent and comprehensible description.

In this study, social cognition was tested with the social cognition subtest of the Edinburgh Cognitive and Behavioural ALS Screen (ECAS). According to previous research, syntactic and semantic symptoms might be correlated with social cognition.

4.1.7. Intelligence

Research on intelligence and language is usually focused on children with language impairment (Cole et al., 1992) or language learning (Genese, 1976). Furthermore, studies have tended to examine topics such as how far intelligence relies on verbal reasoning (Oller Jr, 1981) that sets humans apart from animals (Premack, 2004) or artificial intelligence (Badler, 1989). Studies on the relationship between intelligence and language production in adults are less common and discuss the influence of intelligence on the recovery of stroke (David & Skilbeck, 1984) or cross-validation of for example a reading test (Willshire et al., 1991). Studies investigating the interaction of language and intelligence are in addition almost 30 years old. The presented studies stated that intelligence and language have a fluctuating relationship. In some cases, the intelligence quotient correlated with the severity of for example aphasia (David & Skilbeck, 1984), however, a direct influence could not be confirmed.

This study uses the number of puzzles from Raven's Progressive Matrices (Raven, 2003) participants could solve in 10 minutes as a measure of intelligence which is correlated with performance in the picture description tasks. As the described studies do not provide a clear hypothesis regarding which components of language production may be influenced, an exploratory analysis was conducted.

Summarising, the seven presented language abilities and cognitive functions might be correlated with performance on picture description tasks in a variety of ways. The general aim of the analysis in this chapter is to explore the influence of language abilities and cognitive abilities on spoken and written picture description tasks. Therefore, the performance of participants in tests assessing different language and cognitive domains was compared to their performance in spoken and written picture description tasks. For some language and cognitive domains, previous studies reported data and tendencies. However, for some language domains, the direction of the correlation is not clear, or it was not reported which language domain was affected mainly. Furthermore, no study has analysed the influence of different language and cognitive functions on spoken and written language yet.

4.2. Analysis Method

In the following analysis, the data described in chapter 3 was used. The results of the different tests were analysed using the statistic programs R and JASP.

To halve the number of possible correlations and therefore reduce the likelihood of false positives, the scores of each variable of the description of the Cookie Theft Picture and the Indian Street Scene were combined. The spoken and written variables for each picture respectively were averaged to obtain one set of spoken and one set of written data. For example, the variable Number of Information Units: For each patient, the mean value of their Number of Information Units in the oral description of the Cookie theft Picture and the Indian Street Scene was calculated. The same was done for the written description of both pictures. These values were then used in the correlation analysis.

Correlation analyses were conducted to determine the relationship between picture description variables and other language functions as well as cognitive functions. For all correlation analyses, Spearman's Rank-Order Correlation was used as it is not influenced by outliers.

For each function, the correlations tested were guided by hypotheses from previous literature. For naming scores, previous literature reported significant relationships between naming abilities and connected language production for information units, length of utterance, lexical diversity and disfluencies. Hence, the analysis was focused on the correlation of the GNT with the spoken and written scores of Number of Words, Number of Information Units and Action Units, Disfluency Ratio, TTR and MTLD.

For the syntax domain, there were no published results about the relationship between single tests testing syntax and connected language. Therefore, the correlation analysis was exploratory and correlation analysis with the score in the NAT was tested for spoken and written Number of Words, Information Units, Conciseness Ration Number of Clauses, Syntactic Complexity Ratio, Syntactic Error Ratio, TTR, MTLD, Disfluency Ratio and Semantic Substitution ratio was conducted.

Previous studies on dyslexia and connected language suggest a correlation between dyslexia symptoms and the number of phonetic errors and shorter text length as well as less complex syntactic structures. As described in chapter 2. the participants did not produce a lot of phonological errors. Hence, no correlation analysis for phonological errors could be conducted. However, a correlation analysis between the dyslexia questionnaire score and variables Number of Words and Syntactic Complexity was performed for the oral and written task. According to previous research, concept formation skills might be related to semantic abilities in the connected language.

The participants produced a very low number of semantic errors though. Hence, a correlation analysis was only possible between the score in the Sorting Test and the oral and written score of the Semantic Substitution Ratio of the participants.

For visual-spatial abilities, significant correlations between visual-spatial abilities and semantics as well as syntax were revealed. Therefore, a correlation analysis between the score of the Rey Osterrieth Figure and the oral and written scores for the Semantic Substitution Ratio, Number of Clauses, Syntactic Complexity and Syntactic Error Ratio was conducted.

For Social Cognition previous research suggests a correlation with the semantic and syntactic domain. However, as described above, the participants produced semantic errors very rarely. Hence, a correlation between the score in the subtests of the ECAS and the spoken and written variables Semantic Substitution Ratio, Number of Clauses, Syntactic Complexity as well as Syntactic Error Ratio was performed.

As described above, domain intelligence was reported to be correlated with general good language skills. Therefore, the correlation analysis was exploratory for the domain of syntax. A correlation analysis was conducted between the score of the Standard Progressive Matrices and the variables for spoken and written Number of Words, Information Units, Conciseness Ratio, Number of Clauses, Syntactic Complexity Ratio, Syntactic Error Ratio, TTR, MTLD, Disfluency Ratio and Semantic Substitution ratio was conducted.

4.3. Results

The demographic variables for the participant group described in this chapter are presented in chapter 3.5. An overview of the mean scores the participants reached in the individual tests is presented in Table 16.

Table 16
Mean scores participants achieved in the individual Tests

	Naming Graded Naming Test	**Syntax Northwestern Anagram Test**	**Dyslexia Questionnaire**	**Concept Formation Sorting Test**	**Visuospatial Abilities Rey Figure**	**Social Cognition ECAS**	**Intelligence Raven's Progressive Matrices**
Valid	30	30	30	30	30	30	30
Missing	0	0	0	0	0	0	0
Mean	26.833	9.133	6.600	11.767	35.733	23.500	34.867
Std. Deviation	2.534	1.548	2.699	1.977	0.691	1.548	6.061
Minimum	18	5	2	7	33	17	24
Maximum (Participant)	30	10	13	15	36	24	46
Maximum (Possible)	30	10	32	16	36	24	60

4.3.1. Naming

A significant negative correlation was revealed between the GNT and the Number of Information Units for Actions in the written task (*rs* (28) = -.49, *p* = .006) and is visualized in Figure 29. However, no significant correlations could be found for other variables. A detailed overview of detailed values can be found in Tables 17 and 18.

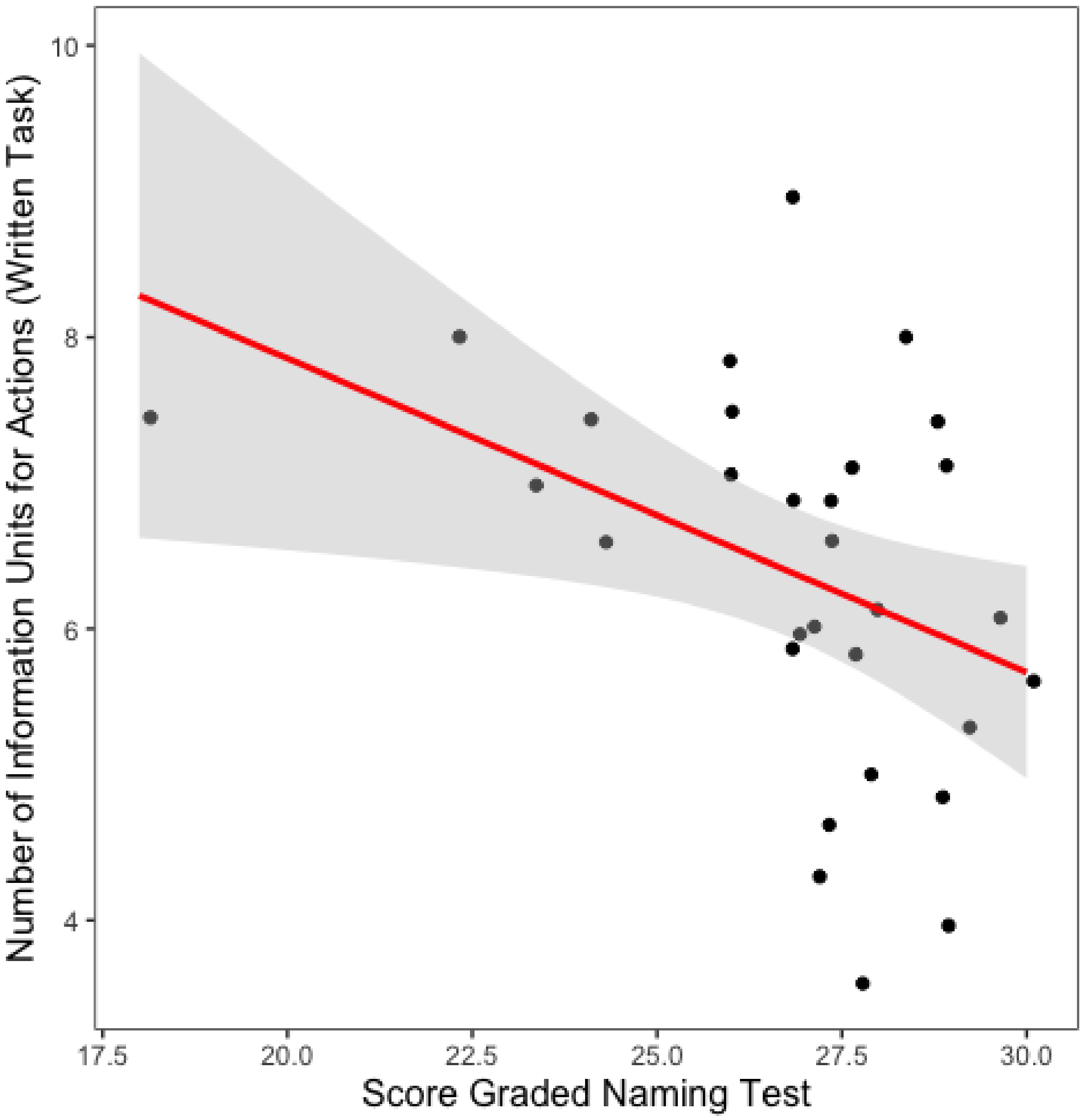


Figure 29: Correlation between the Score in the Graded Naming task and the Number of Information Units for Actions in the written task

4.3.2. Grammar

The correlation analysis for the domain Grammar was exploratory. The NAT correlated significantly negatively with the written task variables Conciseness Ratio (*rs* (28) = -.38, *p* = .037), Semantic Error Ratio (*rs* (28) = -.36, *p* = .048), TTR (*rs* (28) = -.66, *p* = .001), MTLD (*rs* (28) = -.54, *p* = .002) and marginally with the written Number of Words (*rs* (28) =

.33, *p* = .075), Number of Clauses (*rs* (28) = .26, *p* = .168) and Disfluency Ratio (*rs* (28) = -.35, *p* = .058). The significant correlations are presented in Figure 30.

No significant correlations were revealed for any oral variables and other written variables. Detailed values are presented in Tables 17 and 18.

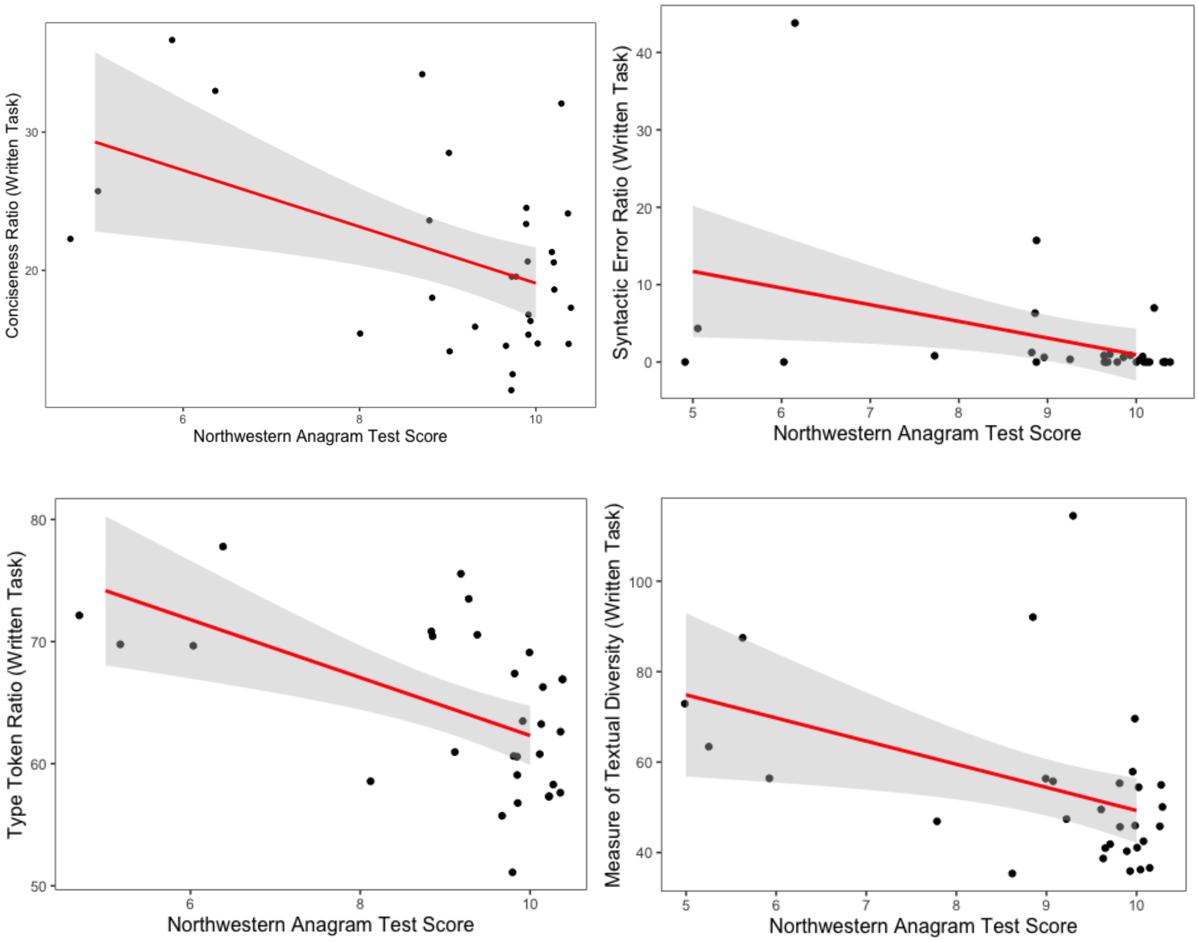


Figure 30: Correlation between the Northwestern Anagram Test and the Conciseness Ratio (upper left), Syntactic Error Ratio in the written task (upper right), the Type Token Ratio in the written task (bottom left) and the Measure of Textual Diversity in the written task (bottom right)

4.3.3. Dyslexia

Some participants revealed a higher score in comparison to the group on the Dyslexia Questionnaire as well as a possible history of dyslexia in the family. However, as described previously in chapter 3.5.4. the participants did not produce a lot of phonetic errors to be analysed statistically. A correlation analysis could only be conducted for text length and syntactic complexity. The score of the Dyslexia questionnaire did not correlate significantly with the Number of Words or Syntactic Complexity in the spoken and written task. Detailed values can be taken from Tables 17 and 18.

4.3.4. Concept formation

An analysis of correlations between concept formation skills and the semantic substitution ratio was not possible as participants produced a too low number of semantic errors.

4.3.5. Visuospatial Abilities

Semantic errors were very low in occurrence in the group of participants in this study. Therefore, only one correlation between visuospatial abilities and the Semantic Substitution Ratio could be conducted. No significant correlation was revealed between the score in the Rey Figure Test and oral and written Semantic Substitution Ratio. Furthermore, no significant correlations were found in the syntactic domain. Detailed values are presented in Tables 17 and 18.

4.3.6. Social Cognition

No significant correlation was revealed between the ECAS and oral and written Number of Clauses or Syntactic Complexity. Marginal effects were revealed for the oral and written Semantic Substitution Ratio. Detailed values can be found in Tables 17 and 18.

4.3.7. Intelligence

The correlation analysis for the domain Intelligence was exploratory. Therefore, a correlation analysis between Standard Progressive Matrices and different oral and written language variables was conducted. However, no significant correlations were revealed. Detailed values can be taken from Tables 17 and 18.

Table 17

Correlation matrix oral task (top: p-value, bottom: corelation coefficient (*rs*), significant values marked)

	Naming	Syntax	Dyslexia	Concept Formation	Visuospatial Abilities	Social Cognition	Intelligence
Number of Words	.199 -.24	.179 .25	.384 .16				.986 -.01
Number of Information Units	.976 -.01	.970 .01					.119 -.29
Number of Information Units for Actions	.289 .20						
Conciseness Ratio		.068 -.34					.886 -.03
Number of Clauses		.171 .26			.655 -.09	.924 .02	.686 .08
Syntactic Complexity		.488 -.13	.985 .01		.368 -.17	.981 .01	.536 -.12
Syntactic Error Ratio		.250 -.22			.609 .10	.084 -.32	.979 .01
TTR	.110 .30	.230 -.20					.576 .11
MTLD	.777 -.05	.723 -.07					.984 .01
Semantic Substitution Ratio		.894 -.03					.531 -.12
Disfluency Ratio	.824 .04	.841 .04					.799 .05

Table 18

Correlation matrix written task (top: p-value, bottom: correlation coefficient (*rs*), significant values marked)

	Naming	Syntax	Dyslexia	Concept Formation	Visuospatial Abilities	Social Cognition	Intelligence
Number of Words	.458 -.1	.075 .33	.393 .16				.172 -.26
Number of Information Units	.342 .18	.906 .02					.293 -.20
Number of Information Units for Actions	.006 -.49						
Conciseness Ratio		.037 -.38					.064 .34
Number of Clauses		.168 .26			.425 -.15	.237 .22	.237 -.22
Syntactic Complexity		.729 -.06	.159 .01		.840 .04	.735 .06	.833 -.04
Syntactic Error Ratio		.048 -.36			.670 .08	.087 -.32	.826 -.04
TTR	.376 .17	.001 -.66					.860 .03
MTLD	.548 -.11	.002 -.54					.873 .03
Disfluency Ratio	.504 -.13	.058 -.35					.238 .22

4.4. Discussion

This study aimed to assess the influence of different language and cognitive functions on picture description tasks. Naming, syntactic abilities, dyslexia, concept formation skills, visuospatial abilities, social cognition and intelligence might influence picture description tasks.

Previous studies have found correlations between the naming abilities of patients or healthy participants and the informativeness of connected speech samples (Giles et al., 1996; Fergadiotis & Wright, 2016; Boucher et al., 2020). This study found only one significant correlation with GNT scores: Naming and Number of Information Units for Actions in the written picture description task. The correlation went in an unexpected direction. Participants with higher naming scores produced fewer action units. It is not clear why participants with greater vocabulary produced less action content. One possibility might be that participants with higher GNT scores prioritised describing objects of the presented scene and spent less time reporting actions. It might also be possible that participants with better naming scores do not have to describe words and use more adjectives or adverbs. "The mother is cleaning the dishes (action unit) and she is looking out of the window (action unit) that's why she does not notice the sink is overflowing (action unit)" becomes "The mother is washing the dishes (action unit) oblivious (adjective) to the overflowing (adjective) sink." In this example, two action units were replaced by two adjective constructions. The presented group is a group of healthy participants. Naming correlated only with informativeness for actions and only in the written task. However, the effect was reversed, which could be explained by participants with greater vocabulary using more complex constructions that replace longer action units. It is likely that naming abilities might have a greater influence in patients with dementia and aphasia leading to more pronounced symptoms and therefore correlations. Further studies researching the connection between naming abilities and connected language production might focus on comparing healthy data to pathological data so that the line between physiological ageing and pathological mechanisms becomes clearer.

The correlation analysis for the syntactic abilities of the participants and their picture descriptions was exploratory. To get an estimation of possible correlations representative variables of each domain were included in the correlation analysis and included: Number of Words (general length), Number of Information Units, Conciseness, Number of Clauses, Syntactic Complexity, Syntactic Error Ratio, TTR, MTLD and Disfluency Ratio. Phonological and Semantic symptoms could not be included in the analysis as participants produced these types of errors too rarely. The NAT was correlated with four variables in the written task only: Conciseness Ratio, Syntactic Error Ratio, TTR and MTLD. Participants that had a higher score in the NAT and therefore had better syntactic abilities had a lower Syntactic Error Ratio. This suggests that poor performance in neuropsychological tests of syntax is a useful indicator for syntactic errors in connected language. However, this effect occurred in the written task only. Written language production is more challenging and therefore places greater demands on construction processes. Furthermore, the NAT is a test of written sentence construction and might hence reveal greater correlations with written language production. Interestingly, participants with a better grammatical understanding produced less lexically diverse and concise picture descriptions. Due to the exploratory analysis and the number of variables assessed, this might be a false positive. Participants with better grammatical skills had a lower conciseness and lexical diversity. However, the effect was revealed for Conciseness Ratio, TTR and MTLD. A connection between syntax and conciseness as well as lexical diversity is not widely discussed in previous literature. A new domain, Phraseology is discussed (Paquot, 2017). However, Phraseology sees syntax and lexical diversity as one concept that is usually assessed and evaluated with the same mark. According to Paquot (2017), participants with good syntactic skills would also produce more lexically diverse language. However, the results of the present study suggest the opposite: Participants with high syntactic abilities tend to be less lexically sophisticated and concise. As there are no previous studies that tested syntax in an isolated test and connected language further studies are needed to test the hypothesis generated in this study.

We also considered correlations with symptoms of dyslexia. Dyslexia research focusses foremost on childhood development. However, dyslexia persists also into adulthood. In most cases, adults with dyslexia have learned compensation strategies that can mask dyslexic symptoms. According to previous research, one would expect a higher number of phonetic errors, less lengthy texts and simpler sentence structure in participants with a higher dyslexia questionnaire score and therefore a possible underlying dyslexia (NHS, 2018). Symptoms would also be expected to be more present in the written picture description task. Phonetic errors could not be included in the analysis because the participants produced a very low number of phonetic errors. However, the results revealed no significant correlations between the Dyslexia Questionnaire Score and the Number of Words (general length) or syntactic errors. It might be possible that adults have acquired compensation strategies that help them to perform as well as adults without underlying dyslexia. However, it should also be noted we did not select participants based on a diagnosis of dyslexia and therefore that any dyslexic signs revealed by the questionnaire are likely to be mild. Future studies might find a connection between dyslexia and oral and written language production. Reliable results can only be obtained when participants with underlying dyslexia and without dyslexia are compared.

Visuospatial abilities and therefore the perception of a picture can influence a picture description task. Previous research suggests an influence of visuospatial abilities on the semantic and phonetic domain. The influence of gender is also discussed. As described above phonological and semantic errors could not be included in the analysis as the participants produced a too low number of phonetic errors, and no significant correlations were found for the other measures tested. Future studies might focus on researching the influence of visuospatial abilities in a larger participant group that might also consist of adult participants with language-related disorders.

Social cognition is an important aspect of language and especially communication. Previous studies found a connection with general language performance but also possible influences on the semantic and syntactic domain which was analysed in this study. As described in chapter 3.5.3., the participants did produce semantic errors very rarely.

Therefore, only the syntactic domain was included in this correlation analysis. No significant correlations between social cognition and syntax were revealed. Further studies are needed to either replicate findings from previous studies or results from this study. A larger sample size of healthy participants and participants with a language impairment or impairment in social cognition might give more insight into the interaction of social cognition and connected language.

Intelligence and language are often discussed as language being a part of human intelligence. Previous studies do not divide into different language domains and their interaction; hence the analysis of intelligence and language was exploratory. As for the syntax domain, the correlation analysis included the Number of Words (general length), Number of Information Units, Conciseness Number of Clauses, Syntactic Complexity, Syntactic Error Ratio, TTR, MTLD and Disfluency. No significant correlations were revealed. A reason for no significant correlations might be the small group of participants that did not consist of participants that might have an above or below-average IQ. Future studies might focus on a higher diversity of participants when testing the interaction of intelligence and connected language.

Overall, future studies are needed to either accept or reject the hypothesis built through the analysis described above. A key objective for future studies should be to find a more diverse participant pool of both healthy individuals and adults with language disorders. Furthermore, written connected language should be included in correlation analysis as, according to the results discussed above, correlations might be more pronounced in the written task. Originally the study aimed to collect data from 100 participants, however, due to the Covid-19 pandemic, only 30 participants could be tested with the presented test battery. One hundred participants would have afforded a much greater power to detect relationships. Future studies should therefore also increase the sample size to produce more powerful data for correlation analyses.

Although correlations were observed in the written task only, correlations were also only found with tests assessing language. This means, at least in healthy adults, picture description tasks likely only assess language and might not be measurably affected by other cognitive functions such as concept formation, visuospatial abilities, social cognition and intelligence. Therefore, picture descriptions could be regarded as a language-specific test. Furthermore, the lack of correlations, especially in expected directions, in the language domains but also in the cognitive domains might highlight that an analysis of connected language is still needed as a key part of clinical assessment, since the measurement of isolated skills and abilities is not always a good predictor of how people use language in connected speech tasks.

4.5. Conclusion

This chapter discussed the influence of different language functions as well as cognitive functions on oral and written picture description tasks. Thirty healthy participants took part in an in-person testing session and completed seven tests assessing different language domains and cognitive abilities. Furthermore, the participants described the Cookie Theft Picture and the Indian Street Scene, each in an oral and written version, resulting in four picture descriptions in total.

According to the results of the correlation analysis, cognitive test scores seem not to have a major influence on the performance of healthy individuals on picture description tasks. Isolated language functions might influence the performance in written picture description tasks in healthy individuals. Furthermore, the results suggest that picture descriptions are a specific assessment of language skills as no correlations with cognitive functions were observed. However, the presented analysis makes clear that the performance in tests assessing single abilities and functions cannot predict the performance of an individual in connected language tasks such as picture description tasks, at least for the sample size reported.

The analysis of the influence of syntax and intelligence was exploratory and therefore tested a variety of variables. Further research could focus on the influence of these two functions on connected language. Additionally, also factors such as naming, possible underlying dyslexia, concept formation, visuospatial abilities and social cognition need to be tested in a larger group of healthy adults but also adults with a language disorder or cognitive impairment. This approach would deliver more robust data and insights into physiological functioning, healthy ageing and its distinction from pathological patterns.

Overall, different language and cognitive functions did not have a great impact on picture description tasks. This means that picture descriptions could be a “pure” assessment of language. However, data from people with neurodegenerative diseases and from a larger sample is needed to support the conclusions drawn in this study.

5. Comparison of different modes of administration

Video-call and online testing have gained popularity in recent years, not only due to technical advancement but also due to the Covid-19 pandemic. Few studies have compared in-person and video-call diagnoses of language disorders in adults, typically revealing no significant results. Automated online tests for the diagnosis of language-related disorders have not been researched yet.

The following chapter discusses the influence of three different administration modes, which is in-person, video-call and automated online testing, on different linguistic domains in oral and written picture description tasks of the Cookie Theft Picture. As in previous studies, participants of the in-person and video-call testing groups did not reveal significant differences in the oral picture description task. Differences in the written task were mainly influenced by the writing mode, which was handwriting versus typing, but also by the administration mode. Participants in the video-call testing group used a chat style pattern in their written picture descriptions resulting mainly in more incomplete sentences and syntactical errors. However, participants who typed their written picture description could also make use of spelling and grammar checks as well as text editing, which may have masked errors, though there is no direct evidence for this in this study. Participants in the online testing condition produced significantly longer and more lexically diverse picture descriptions in both the oral and written picture description task compared to both the in-person testing group and video-call testing group. This effect might be due to the feeling of being less observed leading to less pressure to respond quickly.

Regarding the results, I conclude that in-person testing might still be the method of choice in the diagnosis of language disorders. Video-call testing might present a reliable remote testing method for oral picture description tasks. Automated online testing might *present an option for remote testing to gain a first impression of an individual's language abilities.*

In the wake of increasing technologization in the medical field regarding e-doctor appointments, but also due to the covid-19 pandemic, methods of remote testing have moved to the centre of attention in speech and language therapy. There are various positive effects and new possibilities that remote testing methods can put forward. Patients living in remote locations that do not have, for example, year-round access or are hard to reach can still be attended to online, on the phone, or in case of a good internet connection even over video-call. In addition, patients with anxiety disorders, who find it very stressful to leave the house or let people into their house, have easier access to healthcare without risking an exacerbation of their condition. And, as exemplified by the covid-19 pandemic, people who are shielding or who are immunocompromised may need to access healthcare without in-person interaction. Remote testing and consultation also expand the opportunity for patients to be consulted by experts in their field in cases where a diagnosis or treatment needs special attention with minimal travel and accommodation costs.

As remote testing is quite a new phenomenon there are not many studies concerning differences between in-person and online testing methods. The number of studies looking into methods of online or remote testing in healthcare has increased since March 2020. Most studies are concerned with patient monitoring (Chee et al., 2020), orthopaedic diagnosis of, for example, shoulder (Pinnamaneni et al., 2021) or knee pain (Lamplot et al., 2021) and the application of Covid-19 test kits (El-Tholoth et al., 2020). Only a few studies have researched the application of psychometric tests yet such as the Minnesota Multiphasic Personality Inventory (Corey & Ben-Porath, 2020) or the Test for Rating Emotions in Speech (Mentzel et al., 2020). These studies mostly use video calls as the chosen method to communicate with patients online for diagnostic purposes. In the case of patient monitoring, online applications are partly used like several types of pill reminders or the iWander app that allows monitoring of people with dementia (Sposaro et al., 2010). No study yet has investigated the possibility of using autonomous online tests of language ability.

Studies using video-call testing emphasise the fact that classical and standardised in-person testing methods cannot be "translated" one to one. Objects that need to be manipulated by the patient, for example, the Sorting Test from Delis-Kaplan (Delis et al., 2001) cannot be translated into an online environment without programming online tests. In their guide for tele-practice, Pearson (2021) advises professionals to adhere to the regulations and legal requirements from federal, state and local authorities, which means that testing material cannot be duplicated without permission, including sending testing material to patients. While tests using only visual stimuli requiring a verbal response can be easily adapted to tele-practice, it is stated that tests that require the patient to use manipulatives (Tower Test and Sorting Test) or draw on record forms (Trail Making Test) need a facilitator that is with the patient, for example, a caregiver (Pearson, 2021). No one strategy will fit equally well with all tests and methods. Therefore, it is necessary to develop and assess the remote applicability of tests individually and compare the obtained results with testing in an in-person testing setting.

So far, seven studies have compared forms of tele-assessment with in-person testing conditions in the diagnosis of speech and language disorders in adults. Duffy et al. (1997) compared results from eight people with acquired speech and language disorders (of no further detailed origin) using the National Aeronautics and Space Administration-Launched Advanced Communications Technology Satellite, videotaped samples and telemedicine evaluations from Mayo Clinic practices. They completed the Environment and Examination Protocol, an oral mechanism examination, the motor speech examination and a language examination including following simple and complex verbal commands, picture identification, picture naming, sentence repetition, word definitions, proverb explanations, narrative description of a depicted scene, oral spelling, reading aloud and answering questions about text material that has been read. The evaluations were reliable, and diagnoses were consistent with medical diagnosis and lesion localisation. The patients reported high satisfaction (Duffy et al., 1997). Newer studies were not conducted until 2005 when PCs and video calls became more common and popular. Brennan et al. (2005) compared the performance of 40 patients with brain injuries on the Cinderella Story Retelling procedure in a within-subject design. The patients were tested face to face and via video call with no significant differences between both conditions. The patient feedback was also positive with the majority of patients stating they would use videoconferencing for testing again. Georgeadis et al. (2010) repeated the study 5 years later, coming to the same result and conclusion. The first standardised test that was used to compare video-call and in-person assessments was the Boston Diagnostic Aphasia Examination. Palsbo (2007) compared the performance of 24 post-stroke patients in the subtests of speech comprehension, expression and motor speech in a between-subject design. No significant differences between administration groups were found with the conclusion that assessing a patient’s functional communication via video call is equivalent to an in-person assessment. The first standardised tests that were used for a comparison of tele assessment and in-person testing in a complete set were the short form of the Boston Diagnostic Aphasia Examination and the Boston Naming Test (Goodglass et al., 2000). Theodoros et al. (2008) assessed 32 patients with post-stroke or traumatic brain injury aphasia in a between-subject design. No significant differences between types of administration were revealed. The patients with the video-call condition also reported high overall satisfaction.

Considering picture description tasks in particular, despite the short form of the Boston Diagnostic Aphasia Examination including a picture description task, Turkstra et al. (2012) were the first to report the influence of the administration mode on the variables Number of Words, Content-Units and Type Token Ratio in Conversation, Story Generation, Procedural Description and Picture Description. Twenty patients with traumatic brain injury were tested in a between-subject design using the Mediated Discourse Elicitation Protocol with the pictures of the aphasia bank. Turkstra et al. (2012) also found no significant differences between in-person testing and video-call assessment. However, the patients of the telehealth group produced on average more words than the in-person group. The patient feedback was also positive. The most recent study comparing videoconference administration and in-person testing is from Dekhtyar et al. (2020). This study also included writing and drawing tasks. Twenty adults with chronic acquired aphasia completed the Western Aphasia Battery – Revised (Kertesz, 2006) in a within-subject design. Dekhtyar et al. (2020) found no major significant differences between the face-to-face and video testing conditions. Discrepancies were in the normal range and underlined the importance of clinical judgment. However, they pointed out that not all tests can be assessed in the same way as they can be assessed in person. Some subtests underwent modifications to accommodate interaction by computer. All stimuli were scanned and uploaded as PDF files that were later shared with the participants via screen share. Written or drawn answers from the patients were screenshotted and saved for scoring. Actions the participant has to do with their foot were adjusted to be done with the finger. Furthermore, Dekhtyar et al. (2020) reported that all patients needed help from their caregivers while testing.

In all presented studies, video conference tools were the leading method used for tele-assessment in speech and language therapy. The maximal sample size was 40 in mixed designs, however, about two-thirds of the studies used a between-subject design. All studies report no significant differences, however, almost all studies only compared the syndrome classification of lesion localisation. All studies report high patient satisfaction. Video-call assessment was found to be more time efficient with an increased attentiveness to stimuli by the patients. A faster administration time may be less straining on the patient's focus. Video-call testing also reduces the cost of travel and therefore treatment. Professional services are also more accessible to people who would otherwise be unable to receive services for example due to geographical locale. The more immediate availability of speech and language therapy services might also improve attendance and adherence to therapeutic diagnostic and treatment protocol. However, some limitations are present when testing via video call. Connection difficulties, audio and video delay as well as the reduced quality of visual stimuli and cues compared to in-person assessment might reduce the validity of video-call testing results or interrupt the testing procedure. Furthermore, the privacy of sensitive patient data cannot be guaranteed in video calls as the data exchange is not 100% safe and can be hacked. Another barrier to video-call testing might be that not all patients have access to the equipment needed for video-call testing or further tele-therapy and it is not clear whether the needed equipment can be provided or whether the patients can be reimbursed. Finally, it seems that for remote testing a third person needs to be present in the room to assist with upcoming problems without cueing the patient.

All in all, video-call testing presents a good opportunity to improve accessibility to testing either to get a professional assessment, to reach isolated regions or to uphold social distancing if necessary. However, further studies are needed. First larger sample sizes, as well as replications of studies, are needed to establish the generalisability of findings. A more detailed analysis of changes in different language domains according to administration mode is necessary. Furthermore, the possibility of automated online testing, without a real-time interaction with the assessor, needs to be evaluated, either in form of a (pre-)screening or as a non-congruent option. Only one study has assessed the influence of the administration mode on written tasks (Dijkstra et al., 2004). As discussed in previous chapters, written samples can have a high value in diagnostics and should therefore be analysed in further studies and with more participants. In addition, differences between handwriting and typing have not been researched in the assessment of speech and language disorders, although as described in more detail in chapter 2.2., in tele-testing but also in-person testing of participants with motor problems typing might be the only option.

The following chapter compares different modes of administration of language-based tests in healthy adults aged over 40 years. Therefore, the data reported in chapter 3, which was collected in an in-person condition before the start of the pandemic, will be compared with data that was collected via video call and through automated online testing. The picture description task was the same in all groups. Participants completed a series of psychometric tests, which were also the same as the in-person protocol if applicable or adapted to the online environment. For example, the Raven's Progressive Matrices test was replaced with a comparable test that was developed for online testing. Changes that were made to the procedure described in chapter 3 are explained in more detail in sections 6.1.2. and 6.1.3.

As previous studies comparing video conference assessment and in-person testing have found no significant differences there is no expected direction of whether participants might for example produce more or fewer words in a video call or automated online testing setting.

Another point of comparison in this study was the comparison of handwritten and typed administration. As discussed above, visual stimuli with an oral response can be translated easily into an online environment. Handwritten responses are harder to translate into video-call testing and almost impossible in an automated online testing environment. Consequently, all participants in the online group provided typed written picture descriptions. In the video-call testing assessment, some participants experienced difficulties typing their written responses while seeing the picture stimulus at the same time. This was the case in half of the video-call testing sessions. Participants who could not type their responses wrote their picture description on a piece of paper and mailed a scan or photo of their description after the testing session. The relationship between handwriting and typing will be discussed in more detail in section 5.4.

5.1. Testing procedures

5.1.1. In-Person-Testing

The testing procedure of the In-Person-Testing condition was described in detail in Chapter 3.2. To summarise, before arriving at the University of Edinburgh, participants filled out the dyslexia questionnaire, assessing their attitude towards school and reading behaviour as well as general data about gender, age and educational level. Upon arrival the participants were randomly distributed into four different groups, which determined in which order they were presented with the picture description stimuli (Cookie Theft Picture or Indian Street Scene) and task (spoken or written). In this chapter only the data of the Cookie Theft Picture will be analysed. Participants took part in two testing sessions which were one week apart. In each testing session they described a picture at the beginning and at the end of the testing session. In between both picture description tasks they completed three different tests, which were always in the same order.

The first test in the first testing session was the Edinburgh Cognitive and Behavioural ALS Screen (ECAS; (Abrahams, Bak, & Newton, 2021), followed by the Northwestern Anagram Test (NAT; (Thompson, Weintraub, & Mesulam, 2011) and Raven's Progressive Matrices (Raven, Raven's Progressive Matrices (1938): sets A, B, C, D, E, 1938). In the second testing session, participants first completed the Sorting Test (Delis, Kaplan, & Kramer, Delis-Kaplan Executive Function System, 2001), then they copied the Rey Complex Figure Test (Osterrieth, 1944) and the last test was the Graded Naming Test (GNT; (Warrington & McKenna, 1980). The reason for choosing these tests specifically and their order is explained in chapter 3.2. An overview of the testing procedure can also be found in chapter 3.2.

5.1.2. Video-Call-Testing

Thirty participants were tested via skype and zoom. All participants were recruited with the help of the Edinburgh University Volunteer Panel. The video-call testing session aimed to create a testing procedure that comes as close to the in-person testing procedure as possible. However, some changes had to be made to adapt the in-person procedure to the video chat condition. For example, some tests needed to be altered or had to be taken out of the testing procedure as they were not feasible for video chatting. Furthermore, the video-call testing session was reduced to only one testing session per participant to reduce the risk of losing complete data from two individual testing sessions due to bad connection in the second testing and to ensure sufficient recruitment, as the video chat was stressful to set up for some participants, especially in the beginning of the lockdown period. Each testing session lasted from 45 minutes to one hour. The procedure is presented in Figure 31 as a flow chart and the tests and their adaptation are explained below.

First, the participants described a picture in a randomised order for task and picture stimuli as described in the in-person testing condition (chapter 3.2.).

After the first picture description task, the participants completed the video chat version of the ECAS (Gray, 2020), but not the entire test. Participants completed the subtests fluency for letters “S” and “T”, visuospatial dot counting, cube counting and number location and social cognition. The ECAS was followed by the second picture description task.

After the second picture description task, the participants completed the short version of the Hagen Matrices Test (HMT-S; Heydasch et al., 2020). The HMT-S was chosen to replace Raven’s Progressive Matrices as it provides an online version and a short version that is validated and reliable. Participants had to complete logical patterns that are based on Raven’s Progressive Matrices. The HMT-S takes in general 10 minutes to complete. Examples of the pattern can be found in Appendix E. After the completion of the HMT-S, participants described the third picture.

The third picture description was followed by the GNT, which was completed as described in Chapter 3.2.

When the GNT was finished, the participants performed their last picture description task, followed by copying the Rey Figure. The participants copied the image of the Rey figure that was presented on the screen on a piece of paper. Once they were finished, they held up the paper in the camera of which a screenshot was then taken.

We initially planned for participants to produce written picture descriptions by handwriting on a paper and then hold this up to their camera to be recorded as described by Dekhtyar et al. (2020). Unfortunately, in pilot testing procedures it became clear that the written samples could not always be photographed and accurately deciphered when taken as a screenshot in the video chat. As a consequence, a pilot test was run with 10 adults who described either the Cookie Theft Picture or the Indian Street Scene once written by hand and once typed, one week apart, to analyse whether there was a difference in the picture description tasks regarding the number of words, grammar and spelling. No significant difference was found in this pilot. Therefore, in the main study, we allowed either handwritten or typed responses. When possible, participants wrote their descriptions in the chat box. Participants who could not use the chat box, because the picture stimulus would disappear or they couldn’t find it, wrote their description on a blank piece of paper and sent the picture description after the testing session by mail to the researcher.

In comparisons of three administration groups, the handwritten and typing responses of the video-call group will be analysed as one group in these first analyses. However, in section 5.4. the video-call group was divided into a handwriting and typing group to perform a direct comparison of both writing modes.

Recruitment
Fill out Dyslexia Questionnaire

↓

Randomised Testing Group			
Group 1 Cookie Theft Picture *spoken* Indian Street Scene *written* Indian Street Scene *spoken* Cookie Theft Picture *written*	Group 2 Cookie Theft Picture *written* Indian Street Scene *spoken* Cookie Theft Picture *spoken* Indian Street Scene *written*	Group 3 Indian Street Scene *spoken* Cookie Theft Picture *written* Indian Street Scene *written* Cookie Theft Picture *spoken*	Group 4 Indian Street Scene *written* Cookie Theft Picture *spoken* Cookie Theft Picture *written* Indian Street Scene *spoken*

↓

Testing
Picture Description
ECAS (fluency, visuospatial, social cognition)
Picture Description
HTM-S (Progressive Matrices)
Picture Description
Graded Naming Test
Picture Description
Rey Figure

Figure 31: Flowchart of the Testing Procedure for the Video-Call Testing Group

5.1.3. Online Testing

Forty participants were tested online on the platform FindingFive (FindingFive Team, 2019). All participants were recruited with the Volunteer Panel of the University of Edinburgh. The platform FindingFive was chosen as it provides the possibility to audio-record the participants' spoken responses. The online testing session was also aimed to be as close to the in-person testing session as possible, however, as with the video-call testing condition, not all tests could be performed in the online testing condition. Therefore, as in the video-call condition, the NAT and the sorting test could not be performed in the online condition. Furthermore, it was not possible to assess the Rey Figure task online.

To take part in the testing session, the participants followed a link in their invitation e-mail. The participants could only participate once, and they couldn't take breaks during the testing session. As presented in Figure 32, when the participants started, they were randomly assigned to one of four testing groups, as described in the previous administration groups. Screenshots of the picture description tasks and the complementary tests are presented in Figure 33.

First participants described their first picture. To do so, in the oral condition the participants clicked on the recording button to start the recording. They could only record their description once. As soon as they stopped the recording by clicking "stop", they were led to the next test. In the written condition the participants typed their descriptions in a box directly underneath the picture.

After their first picture description task, participants completed the ECAS. For the fluency tasks, the recordings started automatically once the participants proceeded from the introduction to the test and confirmed that they have understood the instruction. The recording stopped automatically after a minute and participants were led to the next screen. The visuospatial tasks and tasks for social cognition were answered by writing the number or word in the box underneath the picture.

Next, the participants completed their second picture description task, which was followed by the HTM-S. The HTM-S was created as an online test. To answer, the participants clicked on the piece that would complete the pattern logically.

Then the participants described their third picture after which they completed the GNT. The GNT was recorded again. Once participants have read the instructions and clicked to go ahead, the recording started automatically while the picture was presented and finished as soon as the participants moved to the next picture.

As soon as the GNT was finished, the participants described the fourth and last picture. At the end of the testing session, the participants completed the Dyslexia Questionnaire. The Dyslexia Questionnaire was put at the end in this administration group so that it did not prime the participants' answers in the picture description tasks.

Recruitment
Invitation to FindingFive via Link

↓

Randomised Testing Group			
Group 1 Cookie Theft Picture *spoken* Indian Street Scene *written* Indian Street Scene *spoken* Cookie Theft Picture *written*	Group 2 Cookie Theft Picture *written* Indian Street Scene *spoken* Cookie Theft Picture *spoken* Indian Street Scene *written*	Group 3 Indian Street Scene *spoken* Cookie Theft Picture *written* Indian Street Scene *written* Cookie Theft Picture s*poken*	Group 4 Indian Street Scene *written* Cookie Theft Picture *spoken* Cookie Theft Picture *written* Indian Street Scene *spoken*

↓

Testing Procedure
Picture Description
ECAS (fluency, visuospatial, social cognition)
Picture Description
HTM-S (Progressive Matrices)
Picture Description
Graded Naming Test
Picture Description
Dyslexia Questionnaire

Figure 32: Flowchart of the Testing Procedure for the Online Testing Group

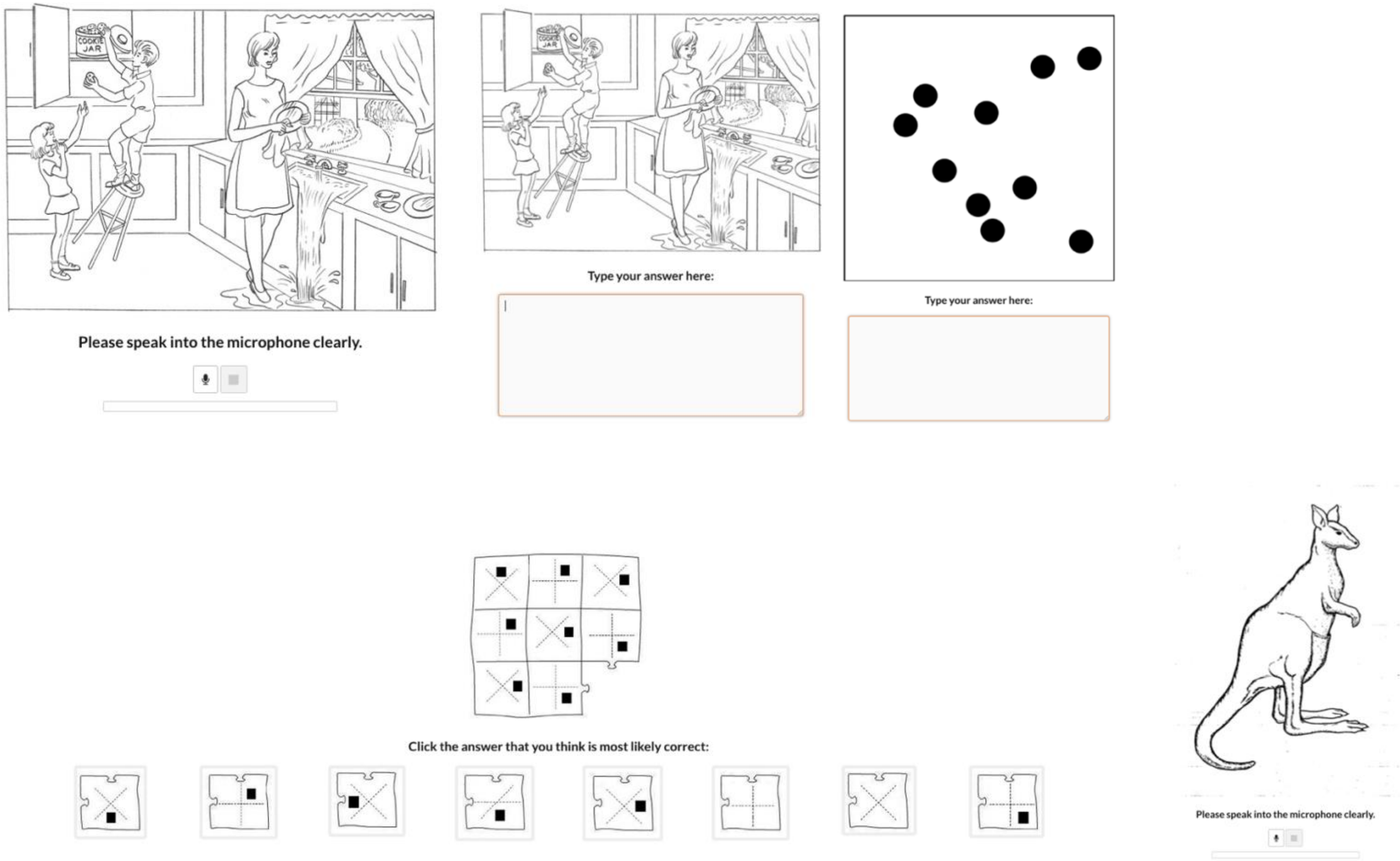


Figure 33: Screenshots of the picture description tasks and the complimentary tests, top left oral task of the Cookie Theft Picture Description Task, top centre written task of the Cookie Theft Picture Description Task, top right ECAS, bottom left HMT-S, bottom left GNT

5.2. Comparison of administration mode groups

Both, the in-person testing group and the video-call group contained 30 participants. The online testing group consisted of 40 participants. A demographic overview of each group is presented in Table 19. The groups were well matched for gender, age and education as well as their performance on ECAS, GNT and the Dyslexia Questionnaire. A chi-square test of independence indicated that the proportion of gender did not differ significantly between administration groups [X^2 (2, N = 100) = 0.01, p = .996]. Participants were aged between 46 and 82 and the mean age did not vary across groups [F (2, 97) = .009, p = .991]. Most of the participants held a bachelor's degree or had some postgraduate education. The distribution of educational levels did not differ significantly between administration groups [X^2 (10, N = 100) = 3.34, p = .972].

Table 19

Overview of demographic variables for the video-call and online testing group

	In-Person	**Video-Call**	**Online**
N	30	30	40
Gender (f/m)	19/11	19/11	25/15
Age (*M*/*SD*)	66.67/7.86	66.67/7.79	66.45/7.80
Education (in %)			
• Some Secondary	6.67	0.00	2.50
• Secondary Graduated	16.67	13.33	12.50
• College	13.33	13.33	12.5
• Bachelor	30.00	36.67	30.00
• Some Post Graduate	26.67	30.00	32.50
• Masters	6.67	6.67	10.00
ECAS (*M/SD*)	23.50/1.55	22.13/1.28	22.03/1.41
Graded Naming Test (*M/SD*)	26.83/2.53	26.17/1.51	25.93/1.25
Dyslexia Questionnaire (*M/SD*)	6.60/2.70	7.07/3.15	7.10/3.31

5.3. Linguistic Analysis

The following section presents the analysis of different linguistic variables and the influence of the administration mode. Oral and written picture descriptions of the Cookie Theft Picture from all administration groups were transcribed as described in chapter 3.2. Speech markers were identified and calculated for each participant. A description of the different speech markers can be found in chapter 3.2. The speech markers were analysed in a 2 x 3 ANOVA, respectively, with Task (Spoken vs. Written) as a within-subjects factor and mode of administration (in-person, video-call, online) as a between-subjects factor. Where the effect of administration mode or its interaction with the task was significant, pairwise post-hoc t-tests with Bonferroni correction were conducted to compare the different administration modes. For the general analysis R and JASP were used.

5.3.1. General length and grade of detail

The number of words each group produced in each task is presented in Figure 34. As described in chapters 2 and 3 the number of words a participant used to describe the depicted scene is the first indicator of general length.

The 2-way ANOVA revealed a significant main effect for Task (F [1, 194] = 48.98, p <.001) and Administration Group (F [2, 194] = 11.581, p <.001), with a higher word count in the oral picture description task compared to the written picture descriptions. However, the interaction effect was not significant (F [2, 194] = .42, p = .66).

Post-hoc pairwise comparisons revealed that the online group produced more words than both the in-person group ($t = 3.13$, $p.adj < .001$) and the video-call group ($t = 4.64$, $p.adj < .001$). There was no difference between the in-person and video-call groups.

The number of information units the participants used is presented in Figure 34. The Cookie Theft Picture can be described with 22 information units in total. The number of information units a participant used to describe the picture can imply how complete a picture description is.

The 2-way ANOVA revealed no significant effect of Administration Group (F [2, 194] = .74, $p = .48$) or Task (F [1, 194] = 1.09, $p = .30$). The interaction effect also not significant (F [2, 194] = 1.05, $p = .35$).

The conciseness ratio of each participant is shown in Figure 34. It reflects how many information units a participant produced depending on the length of their description in per cent ((Number of Information Units/Number of Words)*100).

The two-way analysis of variance yielded a main effect for Administration Group (F [2, 194] = 149.32, $p <.001$) and Task (F[1,194] = 180.99, $p <.001$), with a significant interaction effect (F [2, 194] = 111.14, $p <.001$). Written picture descriptions were significantly more concise than the oral picture descriptions of the Cookie Theft Picture.

Post-hoc analyses of simple main effects revealed that the video-call group was significantly more concise than both the in-person group ($t = -17.9$, $p.adj < .001$) and the online group ($t = -21.5$, $p.adj < .001$) in the written task. There was no difference between the in-person and online testing groups in the written task. Furthermore, there was a significant difference between the spoken and written for the in-person testing ($t = -2.87$, *p.adj* = .005) group and the video-call testing group (t = -19.8, *p.adj* < .001). Both groups had a higher conciseness ratio in the written task.

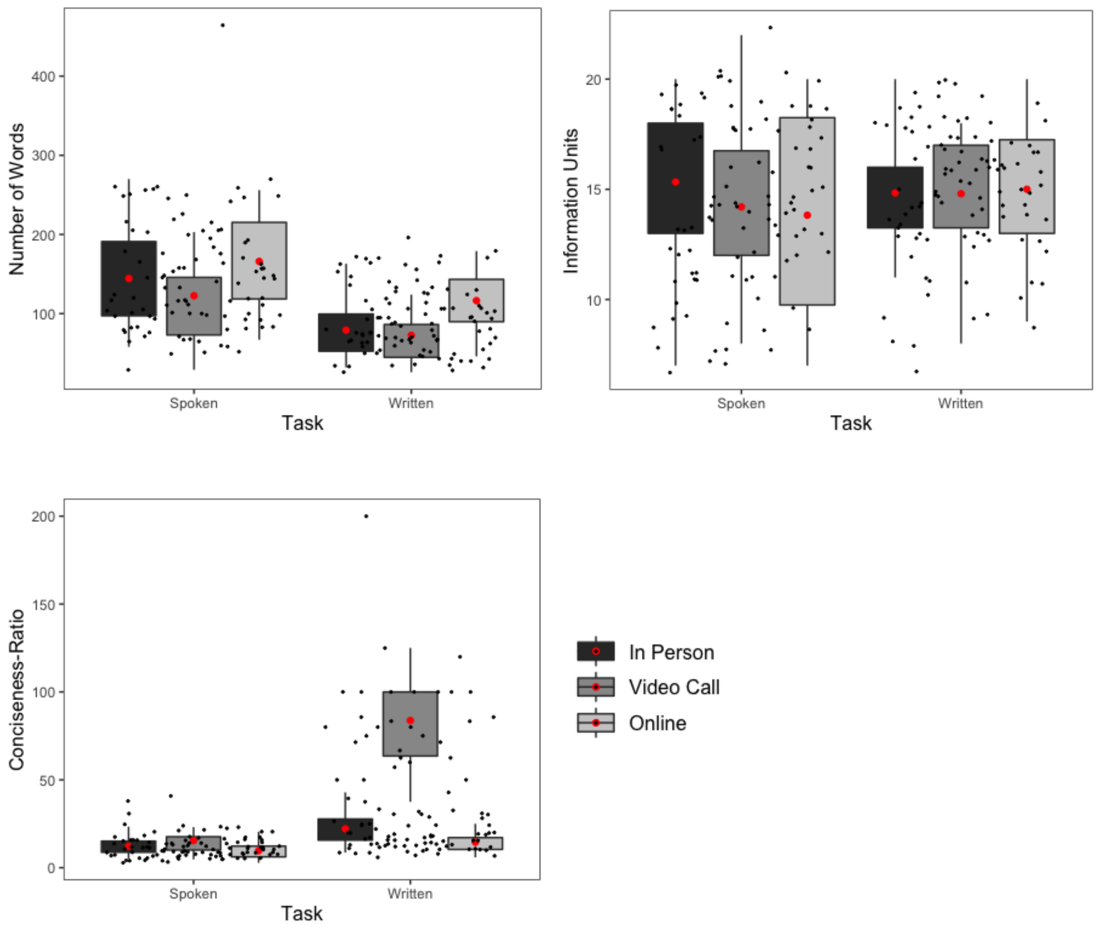


Figure 34: Number of words Number (upper left), Number of Information Units (upper right) and Conciseness Ratio (bottom left) of each group in each task

5.3.2. Syntax

The number of main clauses a participant used is another measure for the general length of a picture description task. Figure 35 presents an overview of the number of clauses of each participant in both tasks.

The analysis of variance revealed a significant effect of Administration Group (F [2, 194] = 23.23, p <.001) and Task (F [1, 194] = 21.93, p <.001) on the number of clauses a participant used to describe the Cookie Theft Picture. However, there was no significant interaction effect (F [2,194] = .76, p = .48). The participants used significantly more clauses in the written task compared to the oral picture description task.

Post-hoc pairwise comparisons revealed that the online group produced more clauses than both the in-person group (t = -5.78, *p.adj* < .001) and the video-call group (t = -5.74, *p.adj* < .001) in both tasks. There was no difference between the in-person and video-call groups.

The number of subclauses a participant used is the first measure of the syntactic complexity. The number of subclauses produced by each participant is presented in Figure 35.

The analysis of variance revealed a significant effect of Administration Group (F [2,194] = 23.21, p <.001) and Task (F [1,194] = 8.05, p = .005) with participants producing in more subclauses in the oral picture description task. The interaction effect of Administration Group and Task though was not significant (F [2,194] = .64, p = .53).

In post-hoc pairwise comparisons the online group produced significantly more subclauses than both the in-person group (t = -4.39, *p.adj* < .001) and the video-call group (t = -6.59, *p.adj* < .001) in both tasks. There was no significant difference revealed between the in-person and video-call groups.

The number of incomplete clauses participants produced in each task are presented in Figure 35.

The 2-way ANOVA revealed a significant effect of Administration Group (F [2, 194] = 4.04, p = .019) and Task (F [1, 194] = 8.45, p = .004) with participants producing more incomplete sentences in the written task. Furthermore, a significant interaction of Administration Group and Task (F [2, 194] = 3.12, p = .047) was revealed.

A post-hoc analysis of pairwise comparisons per task revealed a significant difference between the in-person testing group and the video-call testing group in the written task (t = -3.54, *p.adj* < .001). The video-call testing group produced significantly more incomplete sentences in the written task compared to the in-person testing group. No significant difference was revealed between the in-person group and the online testing condition as well as the video testing group and online testing group in the written task. Furthermore, no significant differences between all administration groups were revealed in the oral task.

The syntactic complexity ratio is the percentage of subclauses compared to the main clauses a participant used to describe the Cookie Theft Picture. Figure 35 illustrates the syntactic complexity ratio of each participant.

The analysis of variance revealed a significant effect of Administration Group (F [2,194] = 3.42, p = .035), however no significant effect for Task (F [1,194] = .03, p = .873) and neither a significant interaction (F [2,194] = 1.45, p = .238).

Post-hoc pairwise comparisons revealed that the video-call group had a lower syntactic complexity ratio than both the in-person group (t = -2.60, $p.adj$ = .010) and the online testing group (t = 3.73, $p.adj$ = .019). There was no difference between the in-person testing and the online testing group.

The syntactic error ratio reflects the number of syntactic errors produced per 100 words. Figure 35 presents the syntactic error ratio of each participant in each task.

The analysis of variance shows a significant effect of Task (F [1,194] = 11.44, p = .001). Participants had a significantly higher syntactic error ratio in the written task. However, no significant effect for Administration Group (F [2,194] = 1.16, p = .317) and no significant interaction of Administration Group and Task (F [2,194] = .09, p = .914) was revealed.

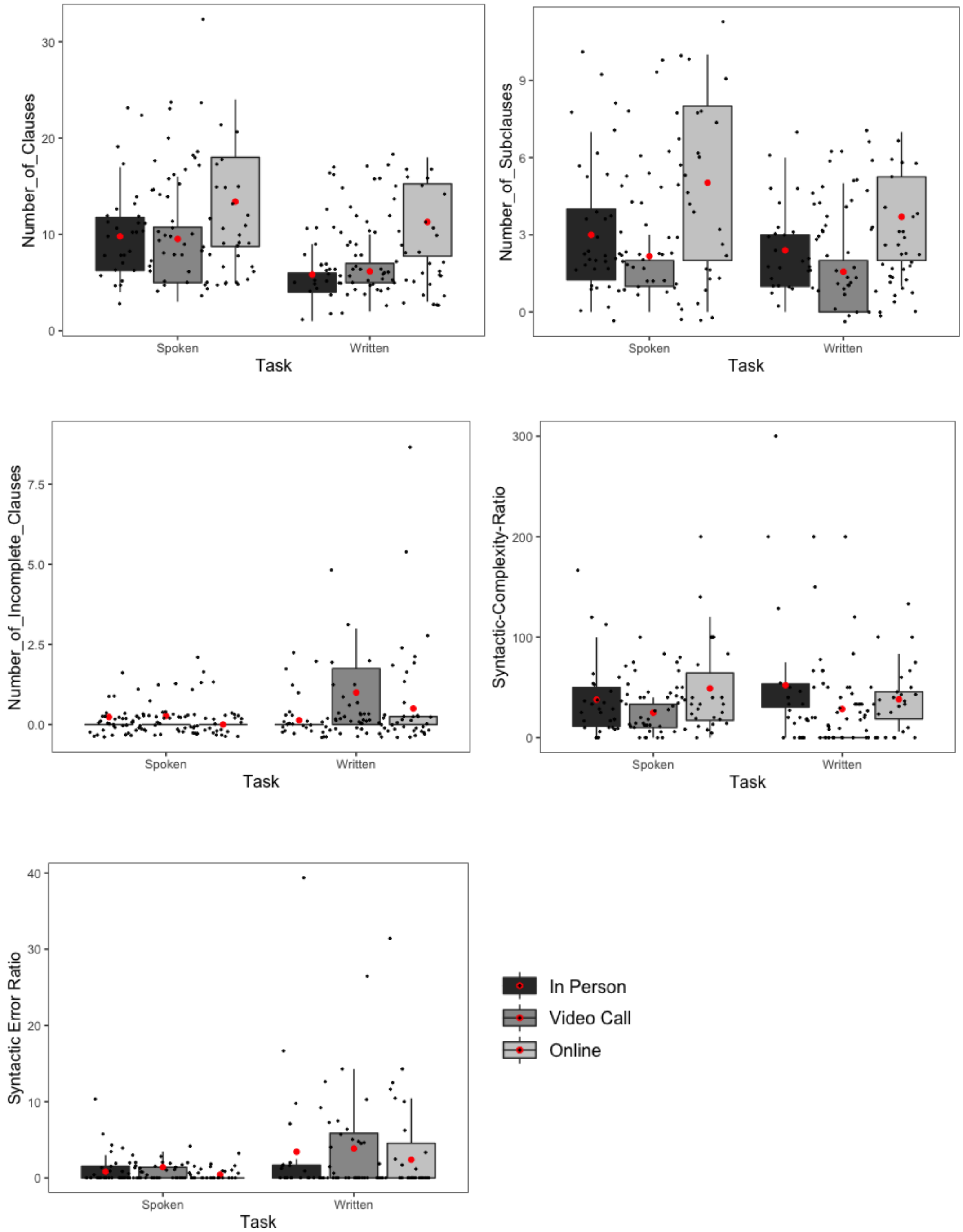


Figure 35: Number of clauses (upper left), Number of subclauses (upper right), Number of incomplete clauses (middle left), Syntactic-Complexity-Ratio (middle right) and Syntactic Error Ratio (bottom left) for each group for each task

5.3.3. Semantics

As presented in section 2.5.3. semantics in this study comprise the lexical capacity of a person and their ability to use it. Therefore, the lexical diversity of the participants' language samples was analysed as well as the number of lexical errors and disfluencies they produced, as these can be a sign of uncertainty in word choice.

For lexical diversity, two types of measures can be used. Commonly used in literature is the Type Token Ratio (TTR). However, the TTR does not take the length of a language sample into account. The Measure of Lexical Diversity (MTLD) also factors in the length of a language sample in its calculation. As discussed above, the language samples of the different tasks and across the different administration modes differ to a certain extent in their length. Hence, the MTLD will deliver a result that will be more precise to compare. Nonetheless, the TTR will also be analysed to make the results of this study more comparable to other studies. The TTR and MTLD of each participant are presented in Figure 36.

The 2-way ANOVA revealed no significant effect of Task (F [1, 194] = .48, p = .491) and only a tendency for Administration Group (F [2, 194] = 2.88, p = .059) on the TTR. However, the interaction effect of Administration Group and Task was significant (F [2, 194] = 10.84, p <.001).

Post-hoc analyses of simple main effects revealed that the online group had a significantly higher TTR than both the in-person group (t = -7.26, *p.adj* < .001) and the video-call testing group (t = -3.90, p.adj < .001) in the oral task. However, there were no significant differences in the written task.

For the MTLD, the analysis of variance revealed a significant influence of both, Administration Group (F [2, 194] = 14.61, p <.001) and Task (F [1, 194] = 55.02, p <.001) as well as a significant interaction effect (F [2, 194] = 13.373, p <.001).

A post-hoc analysis of simple main effects revealed that online group had a significantly higher MTLD than both the in-person group (t = -3.87, *p.adj* < .001) and the video-call testing group (t = -4.67, *p.adj* < .001) in the written task. Furthermore, there was a tendency toward a significant difference between the in-person testing group and the video-call testing condition (t = -2.36, *p.adj* < .001), with the video-call testing group having a higher MTLD compared to the in-person testing group in the written task. However, there were no significant differences in the oral task.

Lexical errors comprise mostly semantic errors. This study this includes close and wide semantic paraphasias, semantic conduite d'approche and d'écart, semantic neologism, semantic jargon, semantic repetitions and semantic substitutions. Semantic errors occurred very rarely in participants of all modes, both the spoken and written picture description. The exact values of individual semantic errors can be found in Appendix F.

The only semantic error that occurred in the oral task were semantic substitutions and these were the only semantic error that occurred in both tasks. Therefore, the semantic substitution ratio (SSR) was calculated and analysed. Figure 36 presents the SSR of each participant.

The analysis of variance revealed a significant effect of Administration Group (F [2, 194] = 4.65, p = .011) as well as Task (F [1, 194] = 4.22, p = .041) and a significant interaction effect of Task and Administration Group (F [2,194] = 4.53, p = .012).

Post-hoc analyses of simple main effects revealed that the video-call testing group had a significantly higher SSR than both the in-person group ($t = -3.62$, $p.adj < .001$) and the online testing group ($t = 3.87$, $p.adj < .001$) in the written task. However, there were no significant differences in the oral task.

As presented in chapter 3.5.1. disfluencies are to a certain degree normal for healthy language. As the nature of disfluencies differs in the oral and written task, the different types of disfluencies will first be analysed according to their average prevalence. For the comparison of disfluencies in the spoken and written picture description task, a disfluency ratio was calculated and analysed.

Disfluencies that can occur in spoken language and were included in this study are filled and silent pauses, repetitions, revisions, discontinued sentences and not identifiable words. In the oral task, the disfluency type that was observed the most across the different modes of administration was filled pauses. Filled pauses occurred almost eight times more often than any other type of disfluency in the oral task. Other types of disfluencies occurred on average very rarely with no generalisable pattern.

In the written task disfluencies also occurred very rarely. Disfluencies that can occur in written language and were included in this study are, as in oral language, repetitions, revisions, discontinued sentences and not identifiable words. Furthermore, disfluencies that can occur in written language and were analysed are crossed-out letters and words, punctuation errors, spelling errors and word onset errors. An overview of all types of symptoms and errors for oral and written language as well as their definition can also be found in chapter 3.3.

The 2-way ANOVA revealed a significant effect Administration Group (F [2,194] = 11.79, p <.001) but not for Task (F [1,194] = .01, p = .932) on the disfluency ratio. Furthermore, the interaction effect of Task and Administration Group was significant (F [2,194] = 15.14, p <.001).

Post-hoc analyses of simple main effects revealed that the video-call testing group had a significantly higher disfluency ratio than both the in-person group (t = -5.75, $p.adj$ < .001) and the online testing group (t = 6.74, $p.adj$ < .001) in the written task. However, there were no significant differences in the oral task. Figure 36 visualises the presented values.

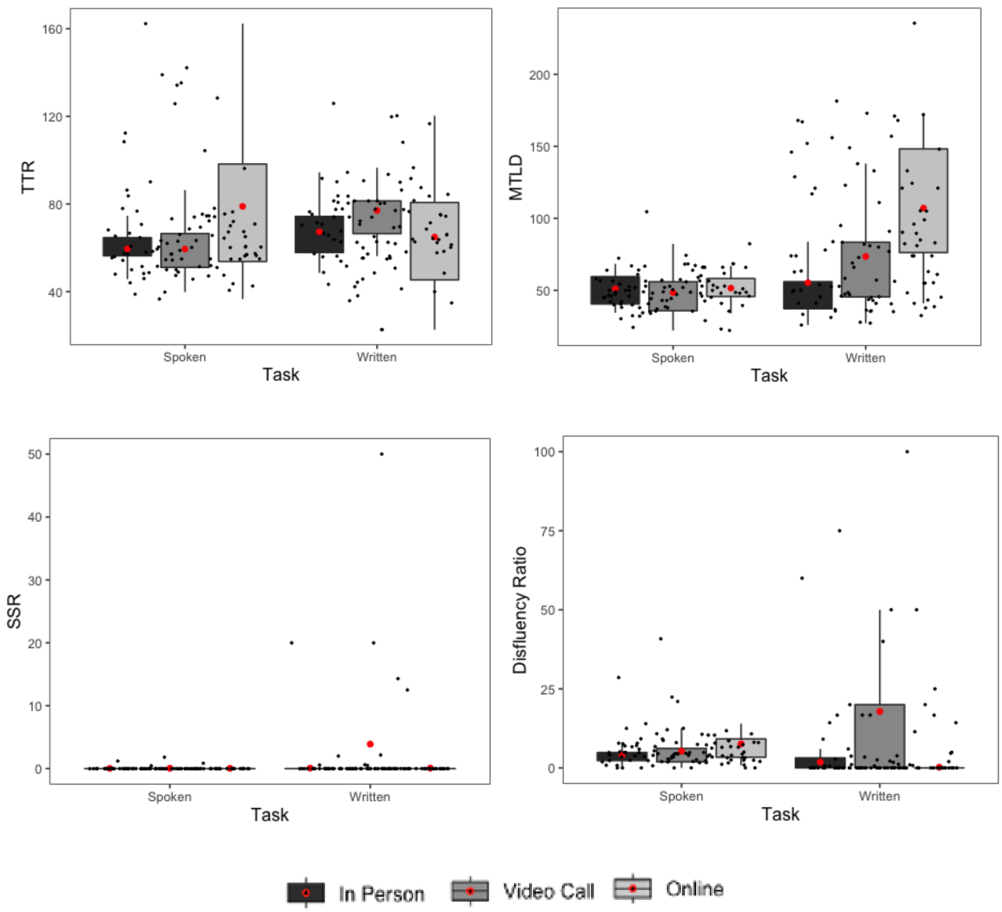


Figure 36: Type-Token-Ratio (upper left), Measure of Textual Lexical Diversity (upper right), Sematic Substitution ratio (bottom left) and Disfluency Ratio (bottom right) for each group for each task

5.3.4. Phonology

Overall, phonological errors occurred very rarely in both tasks across all modes of administration, so no statistical analysis was undertaken.

In the oral task participants only produced phonological conduites d'approches and conduites d'écarts. In the written sample participants produced phonological additions, substitutions, metathesis, phonological conduites d'approches and conduites d'écarts as well as phonological neologisms. Exact values are presented in Appendix G.

Summarising, the administration mode and the task affected different linguistic variables in the picture description of the Cookie Theft Picture. Table 20 presents the differences in an overview. The online testing group produced significantly longer picture descriptions regarding the number of words and sentences, including main and subclauses, compared to the in-person testing group and the video-call testing group. Furthermore, the oral picture descriptions were longer in each administration group.

The administration mode and task had no significant effect on the number of information units the participants used in their descriptions. Phonological errors occurred in all groups very rarely so no further analyses could be conducted.

Regarding lexical diversity, two different measures were analysed with different results. The online testing group had a higher TTR in the oral task compared to both the in-person testing group and the video-call group whereas no effect was revealed in the written task. By contrast, the administration mode had a significant effect on the MTLD in the written task. The online testing group had a higher MTLD in the written task compared to both the in-person testing group and the video-testing group. Additionally, the video-call testing group had a higher MTLD than the in-person testing group.

The video-call testing group also revealed an interesting pattern regarding different linguistic variables in the written task. The descriptions of the participants in the video-call testing group were more concise compared to the in-person testing group and online testing group but included also more incomplete sentences than the other administration groups. The syntactic complexity ratio was also lower in the video-call testing group compared to the in-person and online testing group, while the semantic substitution ratio, as well as disfluency ratio, was higher in the video-call testing condition compared to the other administration modes. A possible explanation for this phenomenon could be the characteristics of the video-call administration mode which will be discussed in greater detail in the following section.

Table 20 presents an overview of the differences between the administration groups.

Table 20

Overview of differences between the In-person, video-call and online testing group

	In-person >* Video	In-person < Online	Video < In-person	Video < Online	Online < In-person	Online > Video
Number of Words					S&W*	S&W
Conciseness Ratio			W	W		
Number of Clauses					S&W	S&W
Number of Subclauses					S&W	S&W
Number of Incomplete Clauses			W			
Syntactic Complexity Ratio	S&W					S&W
TTR					S	S
MTLD	W				W	W
Semantic Substitution Ratio			W	W		
Disfluency Ratio			W	W		

**< and > = Direction of interaction. S = Spoken task. W = Written task. S&W = Both tasks (no interaction).*

5.4. Analysis of Differences between Handwriting and Typing in the Video-Call Condition

Facing rapid technological advancement, handwriting has become more and more replaced by typing in work and school settings as well as in private conversations and daily routines. Offices require the use of electronic documents, and students are encouraged or even required to write assignments and notes on computers and laptops, in private life people prefer mail and texts over handwritten letters and even classic handwritten shopping lists are replaced by handy apps that even sort the groceries in the order they're usually found in the supermarket (Chemin, 2014).

Typing has become an increasingly important aspect of our daily lives, which raises the question of whether it should be included in the diagnostic process of language-oriented tests that still commonly use handwriting. During the current pandemic online diagnostic and testing have become more crucial to protect vulnerable individuals. Typed texts or words are sometimes the only sources of written speech samples that can be obtained from a patient.

However, typewriting cannot be equated to handwriting. The motoric specification is different and typed text can also be easier altered compared to handwritten texts, as further discussed below. Furthermore, in almost every typing program also online and in apps, the typist can make use of spelling corrections and suggested words, which should help reduce errors and make typing even quicker and more effective.

A lot of studies investigated the effect of typing on written language acquisition (Longcamp et al., 2005; Longcamp et al., 2006) and second language acquisition (Lyu et al., 2021) as well as on information retention (Mueller & Oppenheimer, 2014) and essay grading in students (Mogey et al., 2010). Fewer studies evaluated general differences between handwriting and typing in healthy adults, older and younger adults and students, as well as differences in inexperienced typers.

The first variable that can be compared easily is text length. Mueller and Oppenheimer (2014) compared handwritten and typed lecture notes of students. They found that typed notes were usually longer than handwritten notes. However, this might not apply to assignments or descriptions intended for another person. Lee (2020) analysed submissions of Japanese EFL students that were either typed on a smartphone or handwritten with the result that handwritten submissions were significantly longer than those that were composed on a smartphone. It appears that the task might influence differences in the length of written and typed texts but also typing experience can have an influence on text length. Aberšek et al. (2018) compared handwritten and typed scientific texts of sixth grade students that were divided into three groups according to their basic computer skills. Students with good or moderate computer skills wrote on average more words in the typing condition whereas students with poor computer skills produced more words in the handwriting condition. Kalman et al. (2015) compared typing in younger and older adults with the result that older adults type fewer words across different tasks compared to younger adults who are more proficient in typing.

Another variable that can be compared easily across different groups and tasks is writing speed. Comparing solely the task type handwriting versus typing it could be found that typing is usually faster than handwriting for both experienced typists (Mueller & Oppenheimer, 2014) and inexperienced typists using a two-finger typing method (Brown, 1988). However, it was also found that older adults take longer to complete typing tasks compared to younger adults that are more used to typing (Kalman et al., 2015). Interestingly, it was found that students in years 5 and 6 of school show a correlation between handwriting and typing speed, which means that generally written language production proficiency influences writing speed regardless of the mode used. However, typing speed was still consistently faster (Connelly et al., 2010).

The quality of the written piece is also influenced by the mode of either handwriting or typing but also the task and proficiency of the writer. When taking notes, students in the typing condition used a simpler sentence structure and showed a general lower semantic level compared to students taking handwritten notes (Mueller & Oppenheimer, 2014). Conelly et al. (2010) also found that the compositional quality was higher in the handwritten pieces. However, this difference might also be due to the writing task. In writing tasks like essays, that can be corrected and re-structured or worked over, typed pieces were often better structured (Mogey et al., 2012) as it gives the writer more flexibility when working on a computer. It is also assumed that typing allows writers to employ more cognitive resources on text quality rather than text production with the result that typists can devote more attention to ideation, syntactic and semantic monitoring as well as pragmatic awareness (Christensen, 2004). Shibata and Omura (2018) contradict this assumption and revealed in their study that when note taking handwriting has a strong advantage for keeping information regardless of an individual's typing skill, which is also supported by the study of Mueller and Oppenheimer (2014) and Morehead et al. (2019), as it interferes less with other cognitive functions. However, the flexibility of typed text regarding the ability to be re-structured easily as well as being easier to correct gives the typist the possibility to focus more on text structure and quality of writing.

It is also important to note that age and typing proficiency influence the extent to which typists correct and alter their written pieces. In the study of Kalman et al. (2015)

texts from older adults showed more errors as older adults made fewer alterations. Compared to younger adults they spent the most time typing and only 30% of the time correcting their texts compared to younger adults who spent 50% of their time producing and correcting their piece of writing. Hence, the age of the person producing a typed text has always to be considered. Older adults might find typing more difficult overall due to declines in motor skills, timing and sequencing (Krampe, 2002) as well as a decrease in processing speed (Salthouse, 1996). Errors in the typed language of older adults might also occur more often due to difficulties in spelling with age (MacKay et al., 1999; Shafto, 2010).

Summarising, it is not clear whether typed texts are generally longer or shorter than handwritten texts as text length is also influenced by the task type. Studies suggest that typing is in general faster than handwriting regardless of typing skill. The text quality though depends on the proficiency and age of the typist as well as on the task. Consequently, for diagnostic tasks, specific data of adults of different ages have to be analysed to be able to assume whether the typed text of an individual deviates from the norm and whether the difference is clinically relevant. Therefore, data from older individuals is crucially needed as present research compares mostly differences between handwriting and typing in younger individuals, such as students and children.

The following analyses present the differences between handwritten and typed samples of written picture descriptions of the Cookie Theft Picture which were obtained from the video-call testing group described in section 5.1.2. Due to technical differences and difficulties during the video chat session, 15 participants wrote handwritten picture descriptions that were scanned after the video chat session and sent to the researcher. A further 15 participants wrote their picture descriptions in the chat window of the video-call software. The participants were not randomly assigned to the handwriting or typing conditions. The groups were formed due to technical problems some participants experienced. Therefore, some differences that might be revealed might be influenced by other factors, such as the technical proficiency of participants. Furthermore, the presented group is very small with just 15 participants in each group. The results of the analyses can therefore be only a first impression that can provide first data for further research in larger groups. The following sections report the similarities and differences between handwritten and typed language samples that occurred in the description of the Cookie Theft Picture.

The analysis was carried out with JASP and R. Handwritten samples and typed picture descriptions were compared using t-tests. If the assumptions for a t-test were violated Wilcoxon tests were performed.

5.4.1. General Length and Grade of Detail

Handwritten and typed picture descriptions did not vary significantly in number of words (t (28) = .34, p = .73) or information units (t (28) = 1.65, p = .110) used. Furthermore, they did not vary significantly in conciseness ratio (t (28) = .96, p = .344). The values of each category can be found in Figure 50.

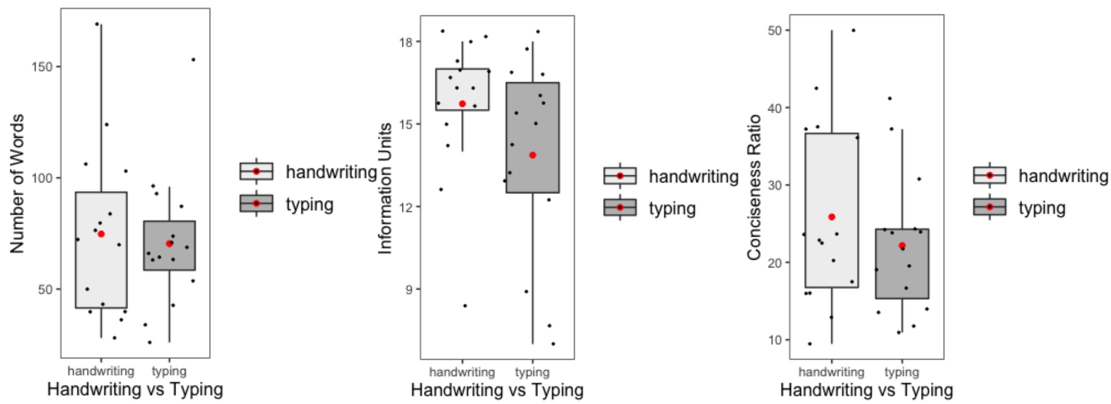


Figure 34: Differences between handwriting and typing for Number of Words (left), Number of Information Units (centre) and Conciseness Ratio (right)

5.4.2. Syntax

The participants in the handwriting and typing group did not use a significantly different number of main- (t (28) = 1.10, p = .280) or subclauses in their picture description tasks (z = 74.00, p = .103). However, incomplete clauses were only produced by participants in the typing condition (z = 30.00, p <.001). Furthermore, the participants in the typing group had a higher syntactic complexity ratio compared to the participants in the handwriting group (z = 58, p = .022). In contrast, the handwriting group produced more syntactic errors in their picture descriptions than the typing group (z = 155, p = .043). The exact values for each participant are presented in Figure 51.

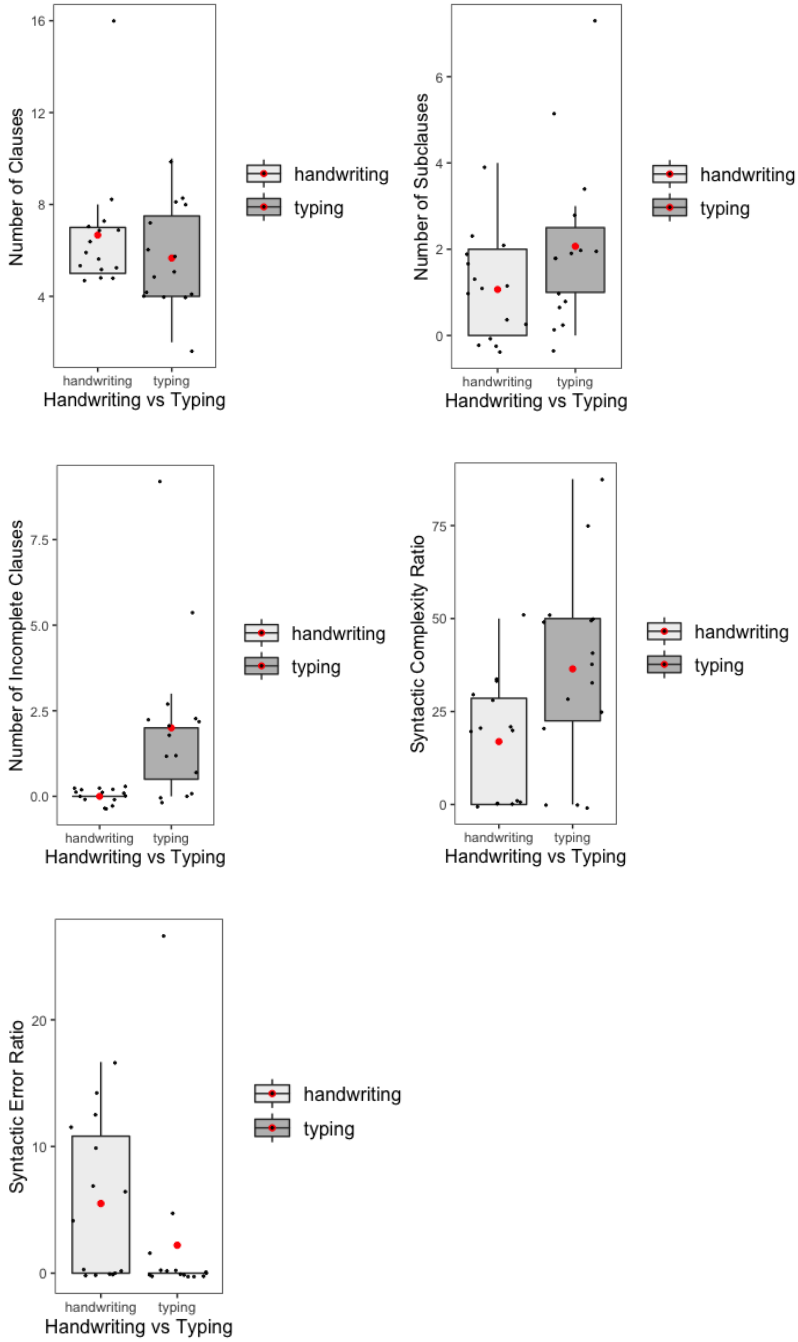


Figure 35: Differences between handwriting and typing for Number of Clauses (top left), Number of Subclauses (top right), Number of Incomplete Clauses (centre left), Syntactic Complexity (centre right) and Syntactic Error Ratio (bottom left)

5.4.3. Semantics

As the number of words did not differ significantly between both groups, for the analysis of lexical diversity only the TTR was analysed. Handwritten and typed picture descriptions did not differ significantly (t (28) = -.26, p = .801).

Lexical errors occurred very rarely. An overview of lexical errors can be found in Appendix H.

Disfluencies also occurred rarely in the picture description samples of the participants in the Video-Call testing group. The Wilcoxon test revealed no significant difference between both conditions regarding the overall disfluency ratio (z = 145, p = .102).

However, several types of disfluencies occurred in the handwriting group only whereas other types were observed in both groups or the typing condition only. Disfluencies that were observed in the handwriting condition only were not identifiable words, crossed out letters and crossed out words as well as spelling errors. Discontinued sentences were only observed in the typed picture description samples. Disfluencies that were observed in both groups were errors of punctuation and word onset. There were too few errors for each type of statistical analysis, but exact values of the disfluencies are presented in Appendix I.

The exact values of the TTR and disfluency ratio are shown in Figure 52.

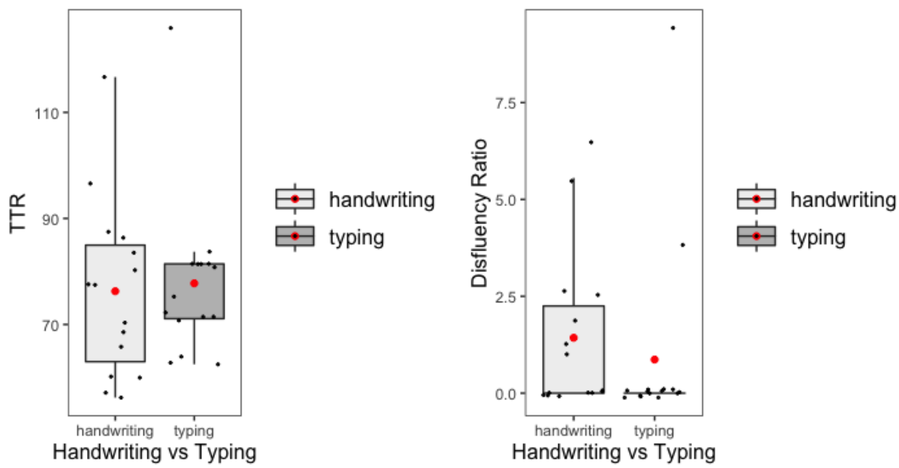


Figure 36: Differences between handwriting and typing for Type-Token Ratio (left) and Disfluency Ratio (right)

5.4.4. Phonology

In the video-call group, no phonological errors were observed, regardless of the writing condition.

Summarising, the picture descriptions of the handwriting and typing groups did not differ significantly regarding general length and grade of detail. Participants also used a comparable number of main- and subclauses in their descriptions of the Cookie Theft Picture. However, participants of the typing group produced more incomplete sentences and greater syntactic complexity, whereas participants of the handwriting group had a higher number of syntactic errors. No significant differences were found between both groups regarding lexical diversity and disfluencies. Lexical and phonological errors were very rare in both groups and therefore not analysed.

The presented results revealed a few major differences between the handwriting and typing conditions. Therefore, the decision to combine both conditions in the main analysis might not have affected the results of the comparison of the three administration modes. However, the group size was small. Larger group studies are needed to determine whether handwritten and typed picture descriptions might differ more subtly.

5.5. Comparability of Administration Modes

Methods of remote testing have become more popular in recent years, in part due to technical advancement but also driven by contact restrictions with the recent Covid-19 pandemic. Studies that researched the use of remote testing in the diagnosis of language disorders have mostly focussed on tele-testing methods such as video-call testing. In this study, a total of 100 healthy individuals completed spoken and written picture description tasks in three different administration modes. Thirty participants were tested in an in-person setting, 30 participants took part in video-call testing and the remaining 40 participants completed automated online testing. Two major differences were found. First, the participants in the online group produced more detailed descriptions of the spoken and written task compared to the other administration modes. Second, in the written task the video-call testing group produced more concise picture descriptions which however were syntactically and lexically less complex and contained more incomplete sentences. The implications of these results are considered below.

The comparability of the different administration modes can be measured in different ways. This study was mostly concerned with differences in length, complexity, and number of errors in the fields of phonology, semantics, and syntax.

However, the first distinguishing difference between the administration modes was that not all tests can be translated to the different modes equally. This is not only true for the distractor tasks such as the sorting test or Rey figure but also the picture description tasks themselves. Not all participants in the video-call testing group could make use of the chat box so they had to write down their descriptions on a piece of paper that was scanned after the testing session and mailed to the researcher. This led to a small subset of grouping (video-call testing group → handwriting versus typing) which therefore could be analysed separately resulting in a comparison of handwritten and typed language.

The analysis of handwritten and typed language revealed no significant differences in the length or complexity of the picture descriptions. These results were in line with previous research stating that the quality of handwritten or typed pieces does not differ among experienced typists (Connelly et al., 2010). However, this study revealed significant differences in the number of spelling errors and syntactic errors. Participants that were typing their picture description tasks could make use of automatic spellcheckers, unlike the participants who were handwriting their descriptions. Due to their technical advantage participants in the typing condition produced significantly fewer to no spelling errors and grammatical errors.

Besides the effect of technical advantage, interpersonal communication might influence picture description tasks in the different modes of administration. Comparing the different modes of administration, the participants of the online testing group produced significantly longer descriptions. The participants of the in-person mode and in the video-call condition might have felt observed as the assessor was present during their testing session. In contrast, participants in the online condition sat perhaps alone in front of the PC. The presence of the assessor in the in-person testing condition and the video-call testing group might have put pressure on the participants to complete the tasks more quickly, reducing the detail of their responses. This hypothesis is supported by the results of the comparison between handwriting and typing. No significant difference was found in the number of words in this comparison suggesting it was not the response mode itself that determined the length of responses but rather the demands of being in a live interaction with another person who was “waiting” for the participant to finish.

The feeling of being rushed or observed did not influence the “completeness” of the picture description tasks. Across all administration groups, the participants showed no significant difference in the number of information units they used. Thus, the mode of administration does not influence the completeness of the picture description task. However, participants of the online testing group had a higher lexical diversity when writing compared to the other two conditions. The higher score of lexical diversity might have been influenced by the number of words. However, the MTLD takes the number of words into account. Thus, it is more likely that participants of the online testing group have taken more time for their descriptions as they were not observed giving them more time and opportunity to produce lexically more diverse texts.

Only some statistical tests could be conducted in the domain of phonology, syntax, and semantics as the number of errors was too low. However, specific differences in the different modes of administration were found. These differences are partly based on the mode of writing. The participants of the online testing group revealed few or no spelling errors or other types of disfluencies in the written condition. The in-person testing group revealed a few more disfluencies, however not to a significant extent. The video-call group however revealed significantly more disfluencies compared to the online testing group, which may be an effect of the chat style pattern the participants in the video-call condition used. Some effects were present for the video-call testing group in the written task that was not seen in spoken production. When writing, the video-call testing group had higher conciseness and syntactic complexity ratios than the other groups but produced also more incomplete clauses, syntactic errors and disfluencies. One potential explanation for this observation is that participants in the video-call testing group wrote more in a chat-style pattern. Although they were instructed to write in full sentences as the participants from the in-person and online testing group, the participants in the video-call testing group frequently wrote very short, partly incomplete sentences. This behaviour might have been triggered by the characteristics of the text chat interface in the video-call setting. The use of spellchecking software might also have an effect, however, not to a significant extent in this group.

Summarising, the differences between the different modes of administration in healthy participants might be the result of the difference between handwriting and typing. Typists have an advantage as they can use automatic spell and grammar checks. Second, differences between the administration modes might have resulted from the feeling of being observed, which affects the written picture description more. The results of the present study comply with previous findings of Duffy et al. (1997), Brennan et al. (2005), Palsbo (2007), Theodoros et al. (2008), Georgeadis et al. (2010), Turkstraet al. (2012) and Dekhtyar et al. (2020) who found no significant differences between in-person and video-call administration. However, the found differences between handwritten and typed picture descriptions as well as face-to-face testing (in-person and video-call) need to be considered when choosing a remote administration method for the diagnosis of language disorders.

5.5.1. Implication for diagnosing PPA

The previous section discussed how the administration mode influenced the performance in picture description tasks of healthy participants. In this section I will discuss possible implications for the diagnosis of PPA.

For people with PPA the difference in the possibility of using a spellchecker in video-call or online testing settings might mask symptoms. In nfvPPA inconsistent phonetic and phonological errors are an important symptom for differential diagnostics besides short and simple phrases. People with svPPA have characteristic surface dyslexia and dysgraphia which might be masked through autocorrection of spellcheckers. For lvPPA the technical advantage of spellcheckers might not mask symptoms. People with lvPPA have word retrieval deficits and a reduced speech rate as core symptoms. These would not be automatically corrected.

In the online testing setting participants might have perceived less time pressure or the feeling of being observed and someone is waiting for their answer. The influence of stress and time pressure has not been researched in PPA yet. However, data from people with aphasia suggests that in people who are aware of their linguistic deficit language production can lead to anxiety. In testing situations or daily communication, they anticipate errors which can feel like a threat, induce anxiety and worsen symptoms (Cahana-Amitay et al., 2011). People with nfvPPA and svPPA are aware of their errors.

In a face-to-face diagnostic setting, for example, in-person or via video chat, people with nfvPPA or svPPA might feel pressure and anxiety leading to a worsening of their symptoms. In online testing settings they might feel less pressured and perform better. Research of the relationship between stress and Alzheimer's revealed that stress can exacerbate symptoms (Justice, 2018). Following these findings, it is possible that word retrieval might worsen in people with lvPPA in stressful testing situations especially if they feel time pressure in face-to-face testing settings.

As the number of information units was not influenced by the administration mode, it might also not have an impact on people with PPA. Differences in healthy participants were only found for lexical diversity in the written online task. This might have an effect on people with PPA. In lvPPA word retrieval might be improved as the testing procedure would allow more time and less stress. Therefore, picture descriptions might be lexically less impaired, and symptoms can be masked. The factor of time and stress might however not mask symptoms in nfvPPA and svPPA. nfvPPA has no lexical deficits and the core symptom of svPPA anomia which will not improve by more time given. People with svPPA will still have a low TTR/MTLD as described in more detail in chapter 3.6.

People with PPA will likely still show deficits in the domain of phonology, syntax, and semantics. lvPPA is characterised by a reduced speech rate with filled pauses and revisions which would be characterised as disfluencies. Short and simple phrases together with inconsistent phonetic and phonological errors are characteristic of nfvPPA. Therefore, people with nfvPPA might show a reduced syntactic complexity and phonological errors independent from the administration mode. As described above the symptoms might be less pronounced in lvPPA and svPPA in the online testing setting. The symptoms of svPPA might be masked, however, as phonological deficits produced because of the surface dyslexia and dysgraphia will be autocorrected in any condition that involves typing.

In people with PPA certain symptoms might be masked depending on the administration mode. Online testing might take pressure from people with lvPPA in the testing setting which might improve their word retrieval. The reduced speech rate would not be influenced in oral administration. Revisions might be autocorrected in typed written tasks. In people with nfvPPA short and simple phrases would still be present in the different administration modes as well as effortful and slowed speech in oral tasks. In typed written tasks spellcheckers might mask inconsistent phonetic and phonological errors. Online testing might also in take pressure from people with nvfPPA in the testing setting which might lead to a slight improvement of symptoms and can mask severity. This would not be the case in people with svPPA as they are usually not aware of their deficits. Therefore, anomia and paragrammatism will still be present across different administration modes. Surface dyslexia and dysgraphia might however be autocorrected in typed written tasks.

5.6. Is online testing a feasible method?

The presented data make clear that there are many similarities but also differences between the different modes of administration. Tests cannot be translated one to one in an online environment, however, online assessment as opposed to traditional in-person testing provides some advantages.

Implications for diagnosis are mixed. While the spoken video-call condition did not differ significantly from the in-person testing condition and could be used interchangeably, automated online testing results must be evaluated with caution. Healthy individuals produce longer descriptions that are lexically more diverse. However, the generalisation from healthy individuals to patients might be limited. Patients might be disadvantaged by online testing as they might need more general help or support to operate the technology from a caregiver. They might also rely on more feedback compared to healthy individuals and might need more detailed instructions or might find it harder to follow instructions to "self-administer" tests. In addition characteristic symptoms from PPA for example could be masked. Further research needs to establish in how far this might translate to patients with aphasia or neurodegenerative diseases and thus cover symptoms concerning general length, syntax and semantics. Looking ahead, future generations might get used to communication that does not receive immediate feedback, and this may lead to less significant differences between face-to-face communication and isolated communication settings.

All in all, the administration mode has a higher influence on different language domains in the written task compared to the oral task. The difference can be traced back to the differences between handwriting and typing. Participants that type their written response to the picture description task can make use of automatic spelling and grammar corrections. Furthermore, the feeling of being observed, as in the written task, might have led to similar differences in the written task between face-to-face testing and automated online testing.

In the end, in-person testing seems still to be the method of choice when it comes to language diagnosis as it is not only the language that is tested. It is much more the general communicative behaviour of a person that is assessed. An online assessment cuts out a lot of natural communication behaviour. Online methods might be a viable option for a first impression of the language abilities of a person if certain functions such as spelling checks are controlled however as of now, they cannot replace in-person testing to its full extent. As technology is changing rapidly, technical advancement such as broad availability of track pads might change online administration in future. This will make future comparisons necessary.

6. Picture Description Across Time and Space

The following chapter presents and describes pictures used to elicit connected language samples in research and diagnostics. Furthermore, gaps in knowledge were identified which will be partly assessed, analysed, and discussed with collected empirical data in Chapter 7.

Pictures are a widely used medium in the assessment of language disorders. There exist several measures for a "good" picture that is well suited to elicit connected language. A good picture should have a clear focussed and well-defined context. It should contain describable pictorial themes and events from different thematic sections that are linked and easily identifiable as well as comprehensible. It should be possible to describe the picture with vocabulary acquired early in life. A "good" picture also shows a familiar scene containing aspects of a person, time and place. Descriptions of the pictures should be predictable and hence comparable across different individuals. However, not all pictures that are widely used meet the requirements for a "good" picture. The most widely used pictures are the Cookie Theft Picture and the Picnic Scene. These pictures depict a scene that may not be familiar to people from all cultures.

It can be concluded that picture stimuli cannot be used without being adapted to the cultural background of the target population and need to be translated like language stimuli.

Pictures are a widely used medium to elicit connected language. However, not every picture is equally suitable for the assessment of language. First, a "good" picture has a clear focus (Giles et al., 1996) with a structured and well-defined context (March et al., 2006) and easily identifiable events (March et al., 2006) and is easily comprehensible (Forbes-McKay & Venneri, 2005). Also, it contains describable pictorial themes (Forbes-McKay & Venneri, 2005) from different thematic sections which are possibly linked (March et al., 2006) and may be described with simple vocabulary acquired early in life (Bschor et al., 2001). Furthermore, aspects of person, time, place and action should be depicted (Giles et al., 1996) in familiar surroundings for most people (Bschor et al., 2001). Finally, the produced output resulting from its description is predictable (Kavé & Levy , 2003) and therefore comparable among different individuals.

The most widely cited picture stimuli in literature are the “Cookie Theft” picture from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1983) and the “Picnic Scene” picture of the Western Aphasia Battery (WAB; Turkstra, 1982). Other studies used Norman Rockwell pictures (Murray, 2010; Tomoeda et al., 1996; Bayles et al., 1993), the “Cat Rescue” (Nicholas & Brookshire, 1993) or the “Circus Scene” (from the Hopkins Corpus; Tippett et al., 2017). Some studies added different pictures like the “Bus Stop Scene” to the “Cookie Theft” picture (Cleland & Pickering, 2006; Forbes-McKay & Venneri, 2005). Shimada et al. (1998) created and used their picture, a living room scene (see page 263).

Over time pictures for picture description tasks have undergone certain changes which will be discussed in the following sections.

6.1. Adaptation to technology

One feature that changed over time was the resolution of pictures. Thanks to advancing technology certain actions and details are now clearer and more visible than they were in earlier versions. One example of this process is the picture of the picnic scene from the WAB (see Figures 37 and 38) which was adapted in 2006 from its original in 1982. While the roles of the depicted persons and their actions have remained the same the picture became more detailed. The lines got finer and some aspects and angles changed.

According to the criteria for a “good” picture, the focus of the picture has changed positively. The depicted characters are now turned more towards the person viewing the picture. This could make it easier for people to describe them as the faces of the characters are easier to see and interpret. It became also catchier to the human eye as faces and face recognition have an especially high priority for information processing in our brain (Johnson et al., 2008). The structure and context have not changed and neither has the thematic content of the picture, but due to the higher resolution, events have become easier to identify and are, therefore, more comprehensible. Also due to the higher resolution and the shift in perspective of certain events the different thematic sections have become easier to identify and are, hence, more describable. The required vocabulary needed to describe the picture should be acquired early in childhood as it depicts a lot of familiar scenes such as drinking, reading, mum/woman, dad/man, house, and lake. In both pictures, aspects of person, action, time and place are depicted and have not changed in the revision of the picture. Concerning the criterion of being a familiar scene to everyone, both pictures fulfil this only partly. While the pictures depict a familiar scene to north Americans it might be less familiar for people from Europe, let alone Asia or Africa. This point will be discussed further in section 4.4. Both pictures also fulfil the last criterion, that is, descriptions of the picture are predictable and consistent among different individuals.

Unfortunately, there exists no information on why the picture was changed. Was it indeed an adaptation to technology or were there reasons apart from improving the resolution of the picture? As the pictorial themes, actions and characters have not changed drastically better resolution might have been one of the main reasons. A higher resolution is also harder to copy as finer lines would disappear when copies are taken from copies. Copying testing kits is common practice in healthcare institutions. Tests and pictures are often copied to be able to hold several testing sessions at once in larger institutions or for educational purposes. Testing kits are expensive; therefore, institutions usually buy only one set and make copies. The picture does not seem to look more modern as the scene and clothing seem still quite conservative and the roles of different characters have not changed. An example of how a picture is adapted to social norms is the revision of the Cookie Theft picture, which will be discussed in the following section.

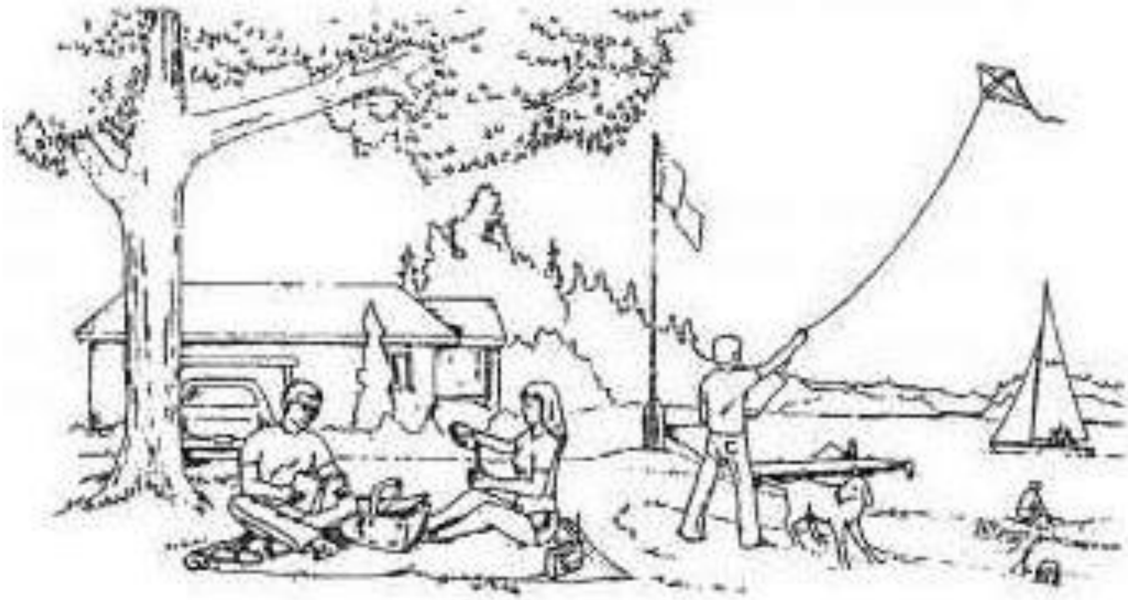


Figure 37: the “Picnic Scene” picture of the Western Aphasia Battery (WAB; Turkstra, 1982)

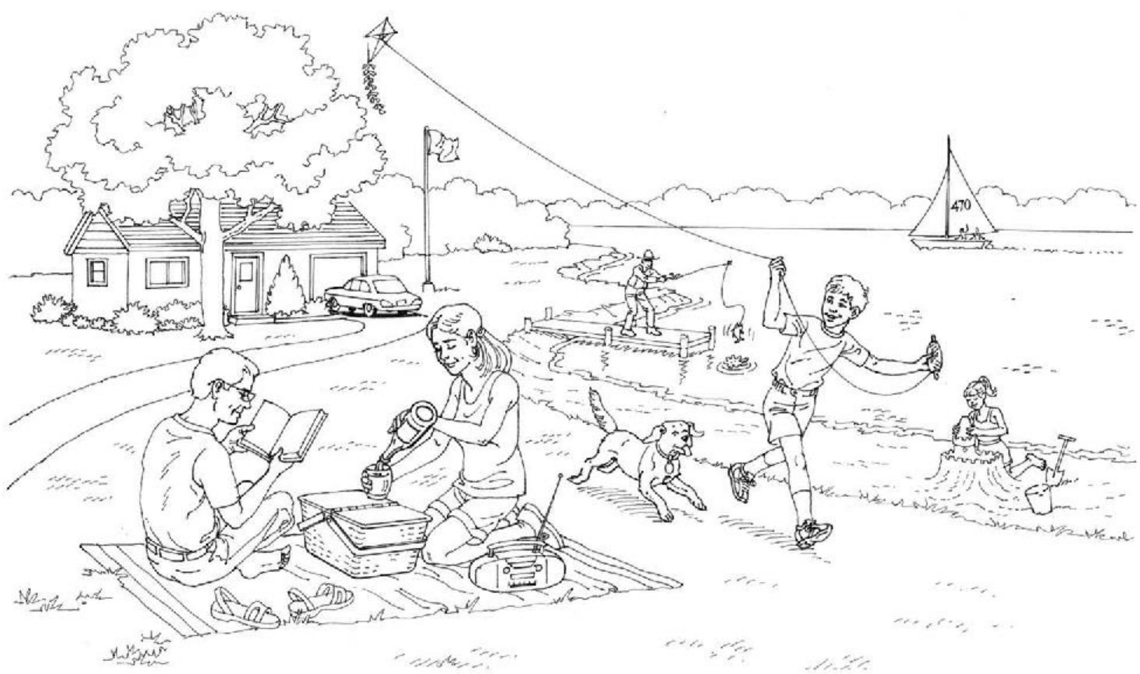


Figure 38: “Picnic Scene” picture of the Western Aphasia Battery – Revised (WAB-R; Kertesz, 2006)

6.2. Adapting to Changing Social Norms

The Cookie Theft picture from the BDAE was recently updated by Berube et al. (2019; original see Figure 39, new versions see Figures 40 and 41). The picture underwent some major changes. The most obvious change is the added colour, which is quite uncommon for testing material. Colours can change the focus of a picture. This was also reported for the updated version by Berube et al. (2019). According to the feedback, Berube et al. (2019) changed the colour in an adaptation so that all actions depicted stand out uniformly. Another point of concern is that the added colour takes away a lot of contrast. The sharp outer lines have vanished, which can make it hard for people with impaired vision to recognise any content in the picture.

Regarding the content, some characters and actions remained unmodified, whereas other actions and characters were added to the picture. Due to the better resolution, it is now better visible that the girl is eating a cookie. In the older picture from 1972, it was not clear what the girl was doing exactly. Some people described her as giggling, some described her as being shocked, whereas others stated she was eating a cookie. In the first updated picture, the girl is depicted with blond hair and the boy has brown hair. In the adapted version (Figure 57) the girl is of mixed heritage whereas the boy looks more of European heritage. Compared to 1972 the children are also producing a bit more mess. Cookies are falling on the floor and a tiny dog, which was added to the picture, is eating the cookies from the floor.

Another new feature is a man doing the dishes. He is standing at a still overflowing sink that has now more dishes in it. The dishes on the counter have vanished. The colour of the man has changed in the adaptation as well to a slightly darker tone.

An aspect that has probably changed most in the picture is the view through the window. The peaceful garden has vanished and was replaced by a scene that looks like it is placed more in a small town with modern, square-shaped houses. There is still a tiny garden with a fence. A woman is busy on the phone, not noticing that she is mowing the flower bed instead of the lawn. Behind her is a cat chasing three birds that are flying away.

According to the criteria for a good picture for a picture description task, the updated picture delivers still a clear context. Some new context and themes were added to the picture. In the first update, the focus was drawn more to the right-hand side of the picture due to the colouring though this issue was solved in the adaptation. The two or three main events are easy to identify in all versions of the picture and they are easy to comprehend. The themes of all pictures are also very describable and derived from different thematic sections (stealing/falling, cleaning/overflowing, phoning/mowing). Only the fact that the girl is eating the cookie has become clearer in the updated pictures. The vocabulary needed to describe the picture is also rather simple and acquired fairly early in life as a household scene is depicted. In all pictures, the aspect of person, action, time and place is also clearly depicted and describable. The expected descriptions of the picture will also be expected to be predictable and consistent among different individuals.

However, the depicted scene is associated and rooted in Western culture. The picture seems less north American compared to the Picnic Scene however this picture would still not represent a scene that is familiar to people from Asia or Africa. At this point, it can be concluded that pictures and revised pictures from standardized test batteries are not fulfilling the point of being familiar with people from different cultural backgrounds. The produced output might not be consistent across individuals from different cultural groups.

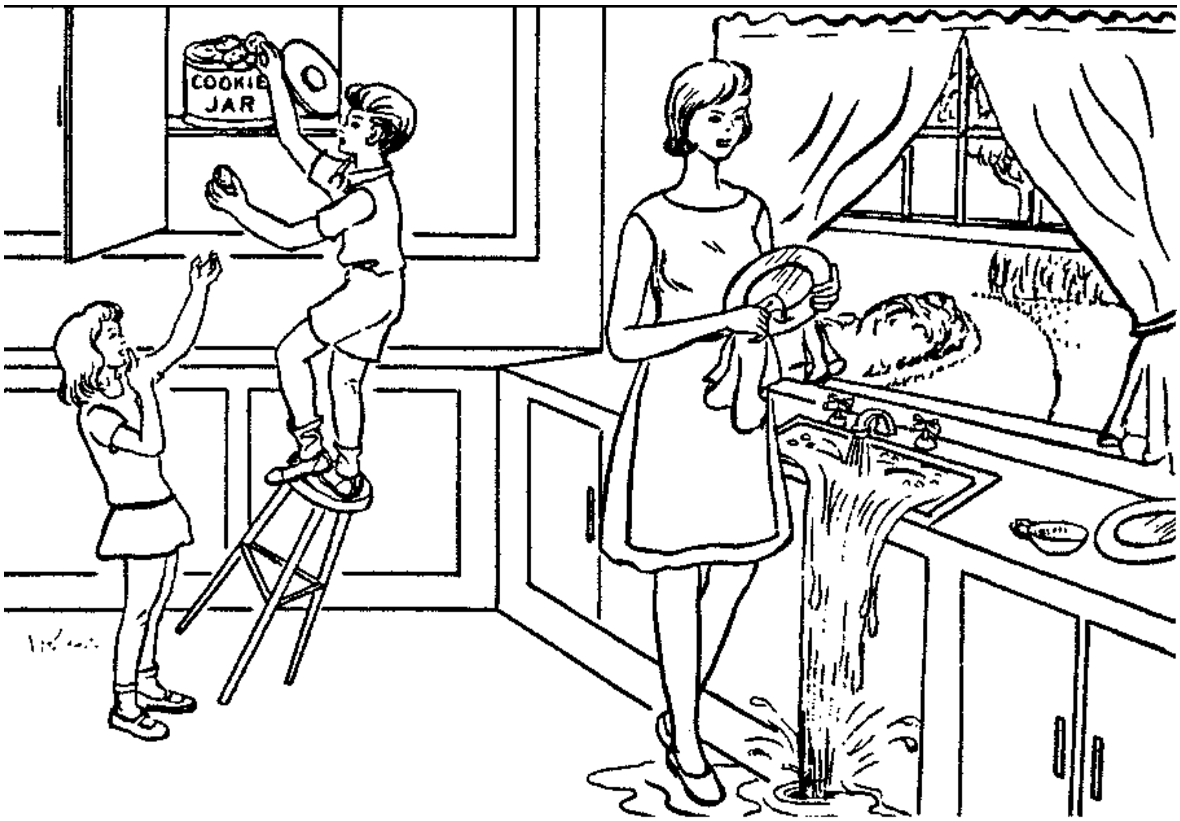


Figure 39: Cookie Theft Picture from the BDAE-3 (Goodglass , Kaplan, & Barressi, 2000)

Figure 40: Updated version of the Cookie Theft Picture (Berube, et al., 2019)

Figure 41: Latest version of the Cookie Theft Picture (Berube, et al., 2019)

6.3. Adaptation to Different Countries and Cultures

The most widely cited picture stimuli in literature are the “Cookie Theft” picture from the Boston Diagnostic Aphasia Examination (BDAE; Goodglass & Kaplan, 1983) and the “Picnic Scene” picture of the Western Aphasia Battery (WAB; Turkstra, 1982). Other studies used Norman Rockwell pictures (see Figure 42; Murray, 2010; Tomoeda et al., 1996; Bayles et al., 1993), the “Cat Rescue” (Nicholas & Brookshire, 1993; see Figure 43) or the “Circus Scene” (see Figure 43; from the Hopkins Corpus; Tippett et al., 2017). Some studies added different pictures like the “Bus Stop Scene” (see Figure 44) to the “Cookie Theft” picture (Cleland & Pickering, 2006; Forbes-McKay & Venneri, 2005). Shimada et al. (1998) created and used their picture, a living room scene, which is displayed in Figure 44.

Figure 42: Norman Rockwell Pictures, The Runaway (left) and Easter Morning (right)

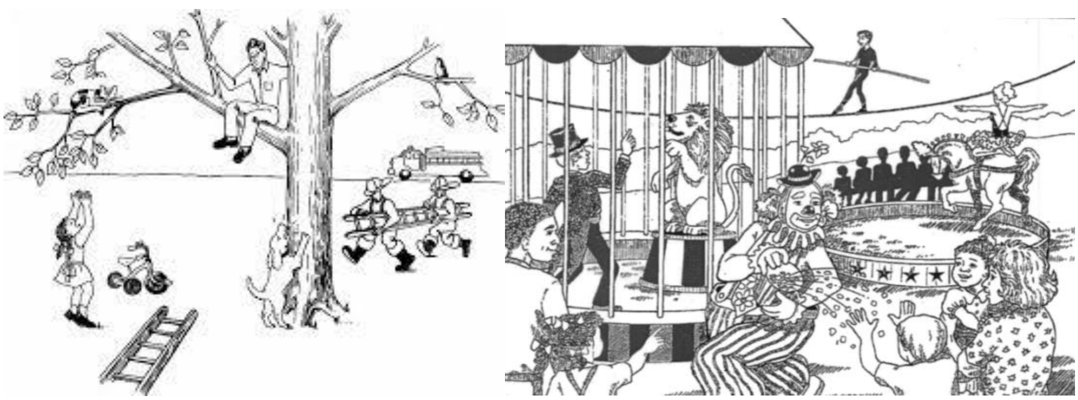


Figure 43: Cat Rescue Scene (left) Circus Scene (right)

Figure 44: Bus Stop Scene (left) and Living Room Scene (right)

All pictures mentioned above fulfil the majority of the above-listed criteria except for the point of being familiar to most people. They depict scenes that are more familiar to American or European populations and not relatable for people not living in non-western countries. Almost no literature or research exists about picture description tasks with “non-western” scenery. Few data have been obtained using western picture material from standardised tests for Indians (Kaur et al., 2017) and Japanese (Shimada et al., 1998). Experience from researchers testing non-western groups has shown that people from different cultural backgrounds can have problems in recognising certain depicted objects although they know and see a certain item daily. One example is the egg in a cup (see Figure 45). People from Guam, see and eat eggs daily but they do not recognise the classical western breakfast egg in a cup as presented in the Pyramids and Palm Trees Test (Bak & Hodges, 2003; T. Bak, personal conversation, October 17, 2018). This leads to the question of whether people who have to describe pictures that are not familiar to them are doing “bad” in tests only because they are not familiar with the scene.

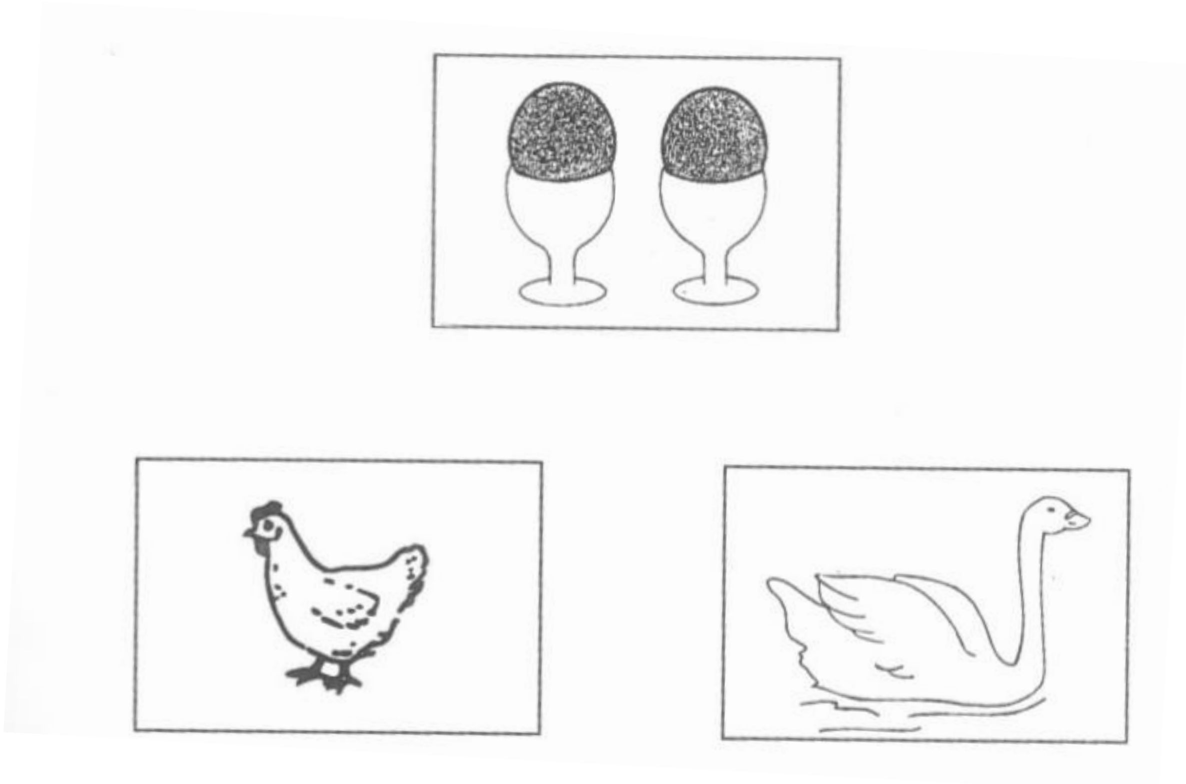


Figure 45: Picture from Pyramids and Palm Trees Test

6.4. Adaptation of Pictures to different Cultures

Nonetheless, there exists picture material that can be categorised as “non-western”. Pauranik (personal communication, June 3, 2020) and Iyer (personal communication, June 3, 2020) have obtained unpublished data for Indians using their own or altered picture material. They have collected data for healthy controls across different age groups and data for different neurological disorders such as dementia and aphasia. However, there is no data about a comparison of western and “non-western” picture material.

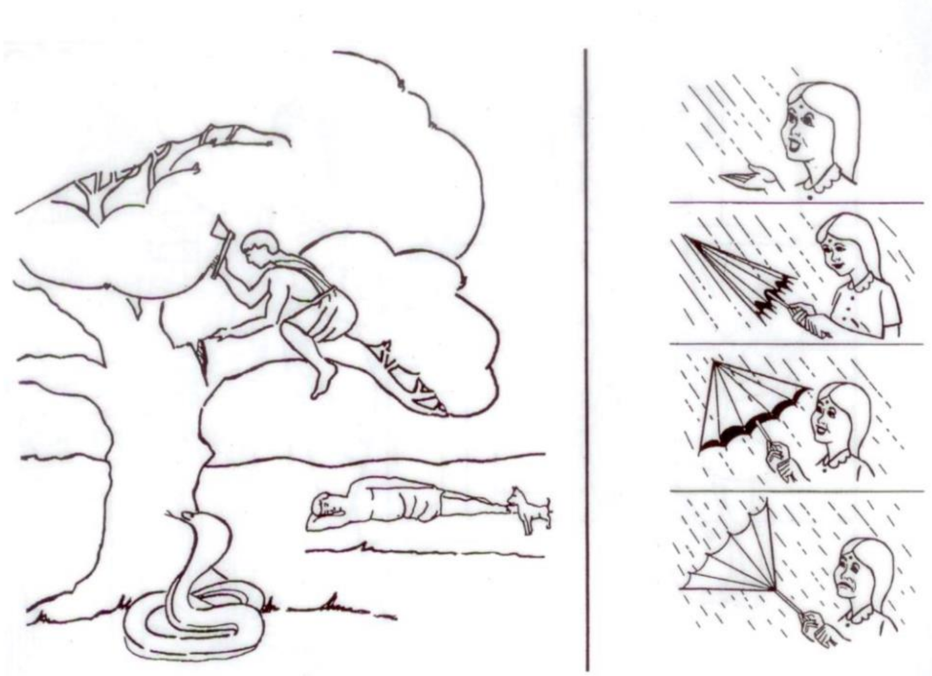


Figure 46: Unique Indian picture material used by Pauranik, static scene on the left, picture sequence on the right

According to the criteria for a good picture the static scene (see Figure 46) from Pauranik fulfils several points. First, the picture has a clear focus, an outdoor scene, and it also depicts a defined context with three points of action that could lead to other situations that the person describing the picture could surmise. The events are easy to identify: In the focus is a man who is cutting off the branch of a tree he is sitting on with an axe; directly underneath him is a cobra snake and further in the background of the picture is a man sleeping on the grass with a dog standing next to him. The observer of the picture can assume that the man cutting the branch of the tree will fall from the tree, right next to the dangerous-looking snake. The observer might also assume that the man sleeping in the background might be woken up by the noise the crash and shock will produce. In general, the scene is comprehensible and could be described with vocabulary acquired early in life. The picture also contains different pictorial themes and thematic sections that are still linked if the person describing the picture describes events that are going to happen in the future. Aspects of person, place, and actions are depicted clearly, and the time can be assumed as possibly mid-day as the person in the background might take a midday nap due to heat (which can be assumed as the people depicted are dressed in light clothes). The surroundings seem familiar to people living in a hot, maybe subtropic climate where snakes are commonly found. Finally, the output produced will be predictable to a certain degree as it cannot be assumed automatically that the observer will describe future events of the picture, and the output will also likely be comparable across different subjects describing the picture.

There also exists “non-western” picture material that has been adapted from for example the BDAE (Goodglass et al., 2000), WAB (Kertesz, 2006) and Frenchay Aphasia Screening Test (Enderby et al., 1986). Iyer et al. (2020) have developed a neurocognitive test battery, the Indian Council of Medical Research - Neurocognitive Tool Box (ICMR-NTB), to standardise the diagnosis of dementia and MCI in India. They have adapted and translated several standardised neurological testing materials for example the Trail Making Test, the Picture Naming Test, and the Frenchay Aphasia Screening Test. It is not clear yet if more culturally familiar picture material elicits more or better language samples. Iyer et al. (2020) only describe that Indians do not recognise the picnic basket in the Picnic Scene from the WAB.

A form of adaptation of pictures to different cultures is taking already existing pictures and changing specific details like for example clothing to the respective culture. Two examples with rather small changes are the adaptation of the Cookie Theft Picture (see Figure 47) and the Living Room Scene by Pauranik (see Figure 48). In the Cookie Theft Picture the clothes of the mother have been changed from a dress with an apron, often described by participants as a classical fifties outfit, to a sari. The woman now also has long hair that is braided instead of a short bob hairstyle. As the picture has not changed otherwise, the points meeting or missing the criteria of a good picture still apply as discussed above.

The living room scene underwent more changes and appears updated. According to the criteria for a good picture the living room scene has a clear focus with a well-defined and structured context. The depicted events are easy to identify and comprehensible. Different thematic themes are depicted like watching TV, knitting and being on the phone, all linked by the location which is the living room. The pictorial themes can also be described with simple vocabulary acquired early in life. The picture also depicts aspects of persons, actions, place and time in a familiar surrounding for most people, again mostly more for people from western nations. Also, as with most pictures, the produced output resulting from a description of this picture is predictable and consistent among different individuals.

Compared to the "classic" picture, the adaptation of the living room scene by Pauranik (A. Pauranik, email, July 5, 2019) seems to contain more detail. There are more pictures on the wall and also more detail that can be seen through the window or the opening the boy comes into the room through.

Figure 47: Adaptation from the Cookie Theft Picture from Pauranik

Figure 48: Adaptation from Shimada by Pauranik

Also, the following adaptations of existing pictures by Pauranik seem to depict more detail concerning actions and persons. It is not clear what an optimal number of items in a picture for a picture description might be or which grade of detail is needed. Further studies would be needed to address this question. The style of the pictures is also more comical, though this is presumably due to the drawing style of the artist.

The adaptation of the picnic/garden scene (see Figure 49) shows one more person and at least three more animals as well as more activities. Based on the criteria for a good picture the picture is clear and has a well-defined and structured focus of playing in the park. The different actions and scenes are easy to identify and comprehend and fulfil the point of depicting different events being linked in one location. The actions of the characters can be described with easy vocabulary acquired early in life and aspects of place, person and action are easily identifiable. The surroundings of the park could be described as being more appropriate and more familiar to a wider population as a large private garden. Though the aspect of time is not further specified, the scene can be set sometime during the day. As with almost all of the presented pictures the produced output from a description of this picture is predictable and consistent.

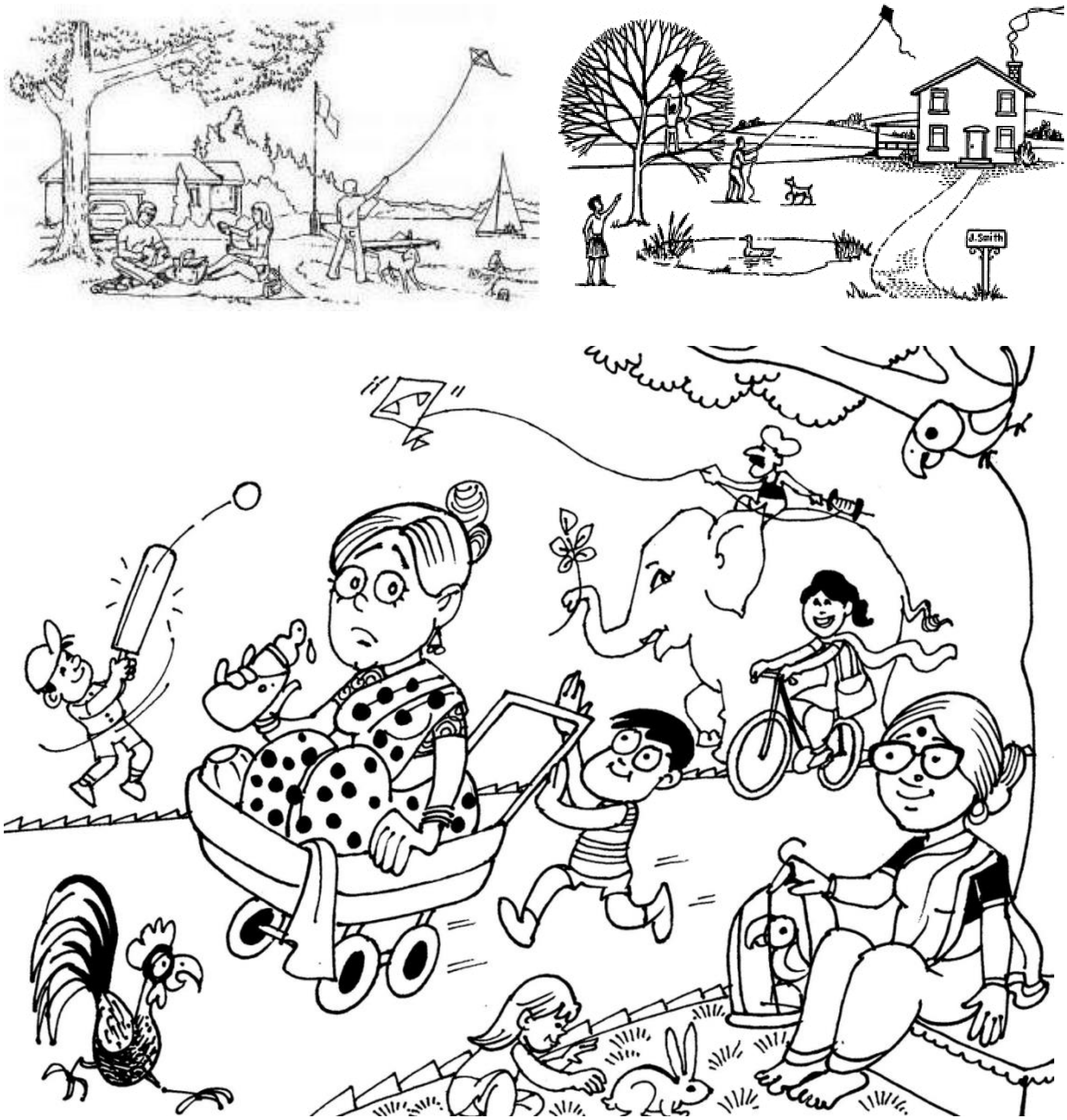


Figure 49: Adaptation from the Picnic/Garden Scene by Pauranik

The river scene has been adapted twice: once by Pauranik and once by Paplikar et al. (2020), who collected data to adapt and validate the Frenchay Aphasia Screening Test (FAST) to the Indian context.

The “western” river scene that was adapted by Pauranik shows a man in the front walking his dog. Following the criteria of a good picture, the river scene has a clear focus with a structured and well-defined focus. The different actions and scenes are easy to identify and comprehend though it is not always describable with vocabulary acquired early in life, for example, kayaking. The events are linked by location. Furthermore, aspects of person and action are easy to identify, although the exact time is not clear. It could be on a weekend in the afternoon. Finally, the produced output of this picture will be predictable and comparable across different individuals.

The adaptation of the street scene by Pauranik (see Figure 50) has again more detail which might depict the reality of life in India more accurately: more people, more animals, more action on the street. Therefore, the picture might be a better reflection of the visual reality of Indian life.

As per the criteria, a good picture is structured, well-defined and has a clear focus. Although there have been more actions and events added, the different aspects of the picture are still easy to identify and comprehend. Aspects of place, person and action are clear and identifiable. All aspects are linked by the location. The scene can be described with vocabulary acquired early in life. However, again the time in the picture is not entirely clear. It appears to be during the day as people seem to be working. According to the sun standing lower over the hills, it could be the morning or afternoon. Descriptions of the picture are likely to produce output that is comparable across different subjects and consistent.

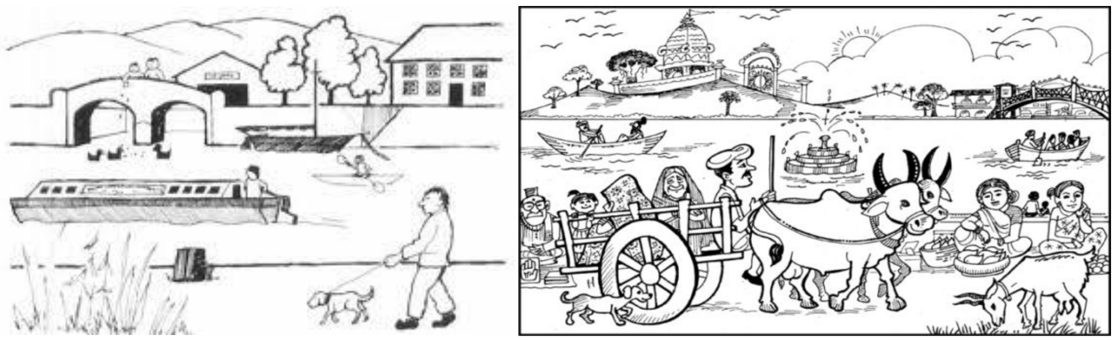


Figure 50: Adaptation from the FAST River Scene by Pauranik

The river scene that was adapted by Paplikar et al. (2020; see Figure 51) has a slightly different angle than the scene from which Pauranik adapted his picture. The fulfilment of the criteria for a good picture mostly corresponds to the aspects described in the previous “western” river scene that was adapted by Pauranik.

The adapted scene from Paplikar et al. (2020) depicts the river from the same angle. According to Paplikar et al. (2020), the dog was replaced by a goat as this was deemed more representative of a rural scene in India, compared to a man walking a dog. Paplikar et al. (2020) also replaced the boat that is tied to the shore near the bridge with a ship sailing behind the bridge as canoes are not commonly found in rural India. The kayaker in the front was replaced by a man rowing a small boat; however, Paplikar et al. (2020) do not expand further on this change. Additionally, Paplikar et al. (2020) explain in their article that all three men were redrawn to look more Indian in facial features, skin tone and clothing.

Concerning the criteria for a good picture, the adaptation of Paplikar et al. (2020) complies mostly with the fulfilment of the criteria for a good picture of the “western” river scene. Furthermore, the elimination of the canoe/kayak makes it easier for the viewer to describe the picture as it is now completely describable with vocabulary acquired early in life.

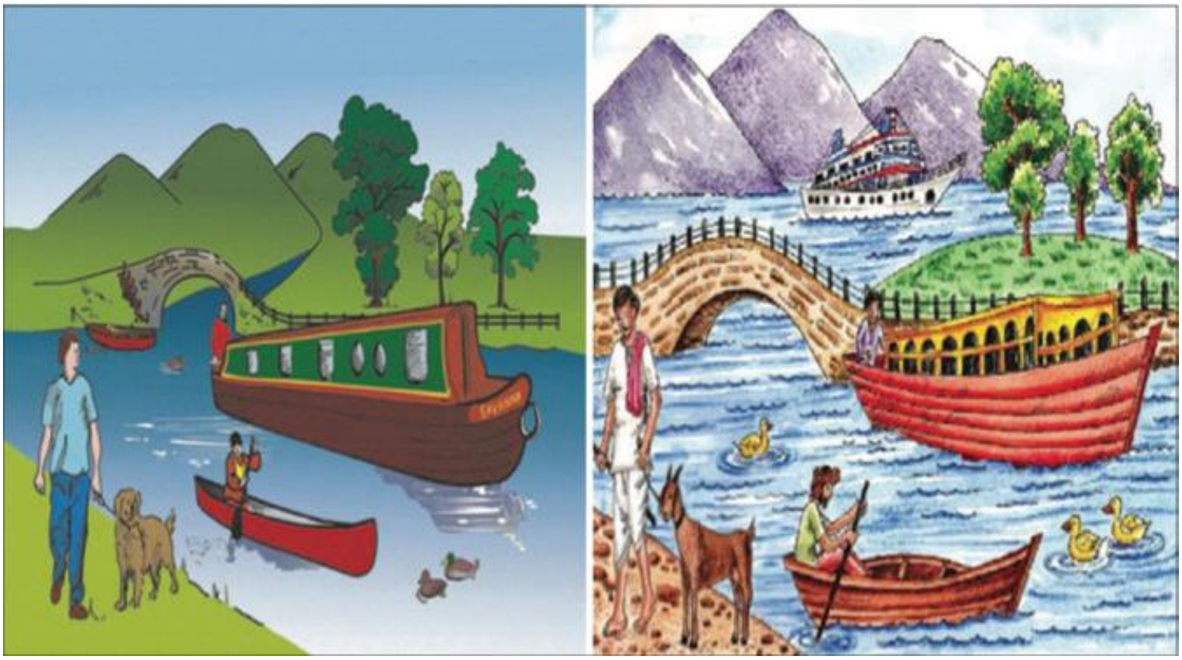


Figure 51: Adaptation from the FAST River Scene by Paplikar et al. (2020)

Pauranik also adapted an Italian version of the river scene to the Indian context (see Figure 52). Bonelli et al. (2015) adapted the river scene culturally to the Italian context by creating the country scene. According to Bonelli et al. (2015), the language domains and test structure were unchanged, however, test instructions were changed to meet the new linguistic and spatial features of the picture stimuli. Data from aphasic people and controls have been obtained for the I-FAST, however, it is not further discussed how participants interpreted the picture. Furthermore, participants did not describe both pictures, so there is no data on whether the cultural adaptation produces better language output.

Concerning the criteria of a good picture, as the river scene, the country scene has a clear focus, is structured and well defined. The different aspects of the picture are linked by their location and are easy to identify and comprehend. Aspects of persons, actions, and places are easy to identify, however, the time is not completely clear. It is during the daytime, however, there are no points of reference whether it is in the morning, at noon, in the afternoon or the early evening. The entire scene can be described with simple words acquired early in life. As with most pictures, descriptions of this picture are likely to be predictable and comparable across different individuals.

Under the terms of a good picture, the street scene by Pauranik contains a greater number of actions and detail compared to the country scene by Bonelli et al. (2015). Though the picture still has a clear focus, the different actions are well structured and connected, partly with each other and partly by the location. Most of the situations depicted can be described with vocabulary that is acquired early in life, however, it is not clear whether the concept of a beggar is acquired early in life or later. It mainly depends on where the person describing the picture was raised. In general, there are more beggars in cities compared to the countryside. Also, for example in bible stories, children come across the concept of a beggar in theory by the story of Saint Martin. However, a beggar can be a common picture in daily life. Further, the picture contains information about persons, actions, time, and place. A description of the picture will likely produce comparable and predictable output from different individuals.

Figure 52: Adaptation from the I-Fast Country Scene by Pauranik

6.5. Can pictures be copied, or do they need to be adapted?

Overall, this chapter makes clear that pictures are not a universally comprehensible stimulus. Many aspects influence how different people, raised in different environments and cultures interpret pictures and the way they approach them. A large number of pictures have been adapted to newer technology or different cultures. For some pictures data were collected, however, pictures have not been compared directly regarding the output they produce. Furthermore, the fact that pictures get adapted makes clear that picture stimuli from tests cannot just be copy-pasted. Pictures cannot just be imported. They need to be translated, like their language counterparts, to different cultures.

After reviewing different pictures from different sources and the current collected criteria for a good picture, many pictures met at least partly the criteria for a good picture. Recapturing a good picture should be clear focussed with a well-defined context. It should contain describable pictorial themes and events from different thematic sections that are linked and easily identifiable as well as comprehensible. It should be possible to describe the picture with vocabulary acquired early in life. Furthermore, it should show a familiar scene containing aspects of a person, time and place. Descriptions of the pictures should be predictable and hence comparable across different individuals. The comparison made clear that this definition is still quite loose, and some points might be more important than others. After analysing the different pictures, I would define a "good" picture a bit differently. First, I agree with the statement that a picture should have a clear focus. However, the scope of the picture description should also be clear. The first version of the Cookie Theft Picture for example has a view outside the window, though many participants were unsure whether this would be part of the picture description or not (see Chapter 3). Some participants included the window in the picture description, some did not. The new version included actions outside the window which makes clearer now that the window and its view need to be included in the description. The Cookie Theft Picture also contains another improvement that I would include in my criteria: Actions are depicted more clearly. Amongst participants there was also disagreement about what the girl is exactly doing. Some interpreted her as shocked or gasping and some said she was laughing. The new version of the Cookie Theft Picture makes clear that the girl is eating cookies.

Besides the actions and also characters being depicted clearer another important factor is the number of actions. The number of actions and characters is important as a picture description tests the semantic domain. A picture therefore should contain enough actions and characters that can be described. Ideally, with regard to patients that might have for example svPPA, these characters and actions should be from the same category but still distinguishable. One example would be the boy and the girl from the Cookie Theft Picture which might both be referred to as child by a person with svPPA.

Therefore, the number of actions and characters does not have to be exceedingly large. 3-4 characters and 2-3 actions might be sufficient. However, the number of actions and characters should be confirmed by further research which would be outside the scope of this study.

Another point that could distort the focus of a picture can be colours. The new version of the Cookie Theft picture for example needed to be changed as the brighter colours in the garden took the focus away from the children's actions. Furthermore, elderly people often have problems with colour vision which would be disadvantageous for diagnosing people with neurodegenerative diseases. Therefore, it might be better if pictures are kept in black and white. I also agree with the point that a picture should show a familiar scene that can be described with vocabulary acquired early in childhood. However, I do not agree with the aspect of time, as this detail is almost never included in descriptions and does also not deliver any further information about a possible disorder. Cultural adaptation is important for the factor of familiarity and, as Palikar et al. (2020) describe, it is sometimes little details that need being changed, for example when in the boat scene the dog was changed for a goat as these are more commonly owned in India. As picture descriptions are frequently used for the diagnosis of neurodegenerative diseases, it is also important that possible confusion about the scene due to being unfamiliar are avoided.

To summarise, , a good picture has a clear focus and scope that shows 3-4 characters and 2-3 actions. The words needed for the description should stem from the same category but can be distinguished, like for example girl and boy. A good picture is also adapted to the cultural background of the patient's early childhood. Lastly, a good picture is kept in black and white to avoid guidance of focus.

The following chapter tests empirically how different adaptations of pictures produce different results and shows further the need for culturally adapted picture stimuli.

7. Influence of the grade of familiarity of a picture on spoken and written picture description tasks

The following chapter explores the influence of the familiarity of a picture stimulus on different language domains.

A group of 100 participants (30 in in-person, 30 in video-call and 40 in online testing conditions) described a familiar scene, the Cookie Theft Picture, and an unfamiliar picture, the Indian Street Scene, each once spoken and once written.

The analysis of variance revealed a possible effect on the completeness of the picture descriptions. The participants tended to misinterpret scenes depicted in the unfamiliar picture stimulus. This might cause misdiagnosis of patients with symptoms in the semantic domain.

However, further analysis of the influence of the familiarity of a chosen picture stimulus on patients with neurodegenerative diseases affecting the language domain is needed.

7.1. Introduction

As described in the previous chapter familiarity with a picture in a picture description task can have an impact on the outcome of the performance of a patient and their diagnosis. Pictures are not universally comprehended. The cultural differences both, between the presented picture stimulus and the person describing the picture, can influence the description of a picture. However, the influence of the familiarity of a picture has not yet been investigated.

In the presented study a group of British/Scottish participants described an American and an Indian picture to test the influence of familiarity on different linguistic domains. The American picture was the Cookie Theft Picture which is widely used in the diagnostic of aphasia as part of the BDAE. The Indian Picture was the Street Scene derived from Pauranik. The Indian Street Scene was chosen as it uses comparable actions from the Cookie Theft Picture such as stealing and falling but also comparable subjects like the boy and the woman. Participants of three different administration groups (first reported in Chapter 5) described the pictures once spoken and once in writing, resulting in a total of four picture descriptions which will be compared in this chapter.

7.2. Picture Stimuli

As mentioned in the previous chapter, the American picture that is classed as more familiar to the group of participants was the Cookie Theft Picture which is described in detail in section 5.2. It presents a classical family situation that could happen also in European households and can be described using 22 information units. The picture that is categorised as less familiar to the presented group of participants was the Indian Street Scene which is also described in more detail in section 5.3. The Indian Street Scene presents a more exotic Scene to the European eye. Indeed, the participants in the present study commented a lot on the foreign clothing and buildings present in the scene. In addition, the monkey is an animal which is usually not encountered on European streets. The Cookie Theft picture has established protocols for scoring the number of information units produced in the description. No such published protocol exists for the

Indian Street Scene, so a new list of information units was developed and applied in this study, a difference that will be discussed further and in more detail in section 7.5.

7.3. Testing Procedure

One hundred participants, recruited with the Volunteer Panel of the University of Edinburgh School of Philosophy, Psychology and Language Sciences were recruited and took part in in-person testing, video-call testing or online testing which is described in section 3.2. in more detail. The exact composition of the participant samples can also be found in section 3.1. The participants described the Cookie Theft Picture and the Indian Street Scene once orally and once written respectively in a counterbalanced order. The picture descriptions were obtained in two testing sessions in the in-person testing condition and one testing session in the video-call and online testing session. The participants described two pictures per testing session with distractor tasks in between both picture descriptions. A more detailed description of the different testing procedures can be found in chapter 3.

7.4. Analysis

To analyse the effect of the familiarity of a picture on different linguistic variables of a picture description the picture descriptions were transcribed and rated according to the scheme presented in Chapters 2 and 3.

For the statistical analysis, the programmes R and JASP were used. First, a three-way mixed ANOVA was conducted to determine the effects of Stimulus (Cookie Theft vs. Indian Street Scene), Task (Spoken vs. Written) and Administration Group (In-person vs. Video vs. Online) on the different language domains. If either the interaction between group and stimulus or the three-way interaction was significant a two-way repeated measures ANOVA was conducted as a post-hoc analysis for Stimulus x Task in each group with alpha .0167. In groups with a significant interaction between Stimulus and Task, only a test of the Stimulus effect was performed for the spoken and written task separately with a Bonferroni corrected alpha of .025.

The focus of the analysis was on the effect of Stimulus. Therefore, post-hoc tests for the task have been excluded from the analysis as the influence of the Administration Task on different linguistic domains was discussed in chapter 3 in more depth. The influence of the administration group was analysed and discussed in chapter 5. Hence, the effect of the administration method will only be addressed briefly in the analysis.

7.5. General Length and Conciseness

As described in Chapters 3, 4 and 5 the number of words a participant used is a first impression of the length and complexity of a picture or task.

The three-way mixed ANOVA revealed a significant effect of Stimulus [F (1, 97) = 39.05, p = .006], Administration Group [F (2, 97) = 5.33, p = .001] and Task [F (1, 97) = 111.96, p < .001] as well as a significant interaction of Stimulus and Task [F (1, 97) = 4.29, p = .041] , Administration Group and Task [F (2, 97) = 6.51, p = .002] and a significant three-way interaction of Stimulus, Task and Administration Group [F (2, 97) = 8.35, p < .001].

In the post-hoc analysis, the two-way repeated measures ANOVA examining the effects of Stimulus and Task in each group revealed no significant interaction between

Stimulus and Task in the in-person testing group [F (1, 29) = .99, p = .33] as well as in the online testing group [F (1, 39) = .05, p = .83]. However, there was a separate significant effect of Stimulus (in-person: F (1, 29) = 15.91, p < .001; online: F (1, 39) = 11.56, p = .002) and Task (in-person: F (1, 29) = 33.78, p < .001; online: F (1, 39) = 34.10, p < .001) in both groups. The conducted ANOVA revealed a significant interaction effect in the video-call testing group [F (1, 29) = 10.46, p = .003] as well as a separate significant effect of Stimulus [F (1, 29) = 12.17, p = .002] and Task [F (1, 29) = 42.10, p < .001].

The pairwise comparison between the Stimuli in the different tasks revealed a significant difference for Number of Words in the video-call testing group in both the spoken [t (29) = -3.47, p = .002] and written task [t (29) = -5.68, p < .001]. In both tasks, the participants used more words in their description of the Indian Street Scene.

The results suggest that the stimulus might influence the Number of Words participants used in their descriptions. The participants used more words in their description of the Indian Street Scene. All values are depicted in Figure 53.

Figure 53: Effect of Stimulus, Task and Administration Group on the Number of Words produced

The pictures can be described with a certain number of information units as described in section 2.5.2. Information units can be organised into four categories: subjects, places, objects and actions. Table 20 presents the 22 possible information units for the Cookie Theft Picture, in the different categories as used in the BDAE (Goodglass et al., 2000). For the present analysis, the standardised method for scoring information units was used, as described in section 2.5.3.

There exists no description for why the present 22 items of the Cookie Theft Picture were chosen. It is not clear whether the list of items is based on a statistical analysis of which items were used by test participants or if the list of items was generated more intuitively. For the Indian Street Scene, an analogous set of information units did not exist before the start of this study. A list of items in the same categories as the Cookie Theft Picture was generated for the Indian Street Scene picture. First, it was planned to generate a list based on statistically most frequently named items in each category when participants described the Indian Street Scene. However, the majority of the participants did not interpret the situation surrounding the man with the car correctly. Participants were most concerned with the side the car might be driven on the street or on which side the steering wheel is to describe if the man is standing on the driver's or passenger's side. They did not recognise that the man stepped out of the car to see what is going on in the street/on the bridge he wanted to drive over. Furthermore, the street was often interpreted as being a bench. A statistical approach based on the frequency of answers would have not represented the content of the picture correctly. Hence a list was generated that consists of 36 items presented in the picture. Table 22 presents the items in their categories respectively. The choice of items followed the pattern of information units of the Cookie Theft Picture.

Table 21
Information Units Cookie Theft Picture

Category	Items
Subjects	the boy, the girl and the woman
Places	the kitchen and the exterior seen through the window
Objects	cookie, jar, stool, sink, plate, dishcloth, water, window, cupboard, dishes, and curtains
Actions	boy taking or stealing, boy or stool falling, woman drying or washing dishes/plate, water overflowing or spilling, action performed by the girl, woman unconcerned by the overflowing, woman indifferent to the children

Table 22
Information Units Indian Street Scene

Category	Items
Subjects	The man getting out of the car, the woman buying bananas, the man selling bananas, the monkey, the man walking/reading the newspaper, the bird/s, the beggar, the boy and the dog
Places	The street, the pavement, the wall
Objects	Car, bag, bananas, (news)paper, bowl, coins, (man)hole, manhole cover, clock tower (cricket)bat, ball
Actions	Man getting out of the car, observing the scene on the blocked street; woman buying bananas; man selling bananas; man sitting on the pavement; monkey sitting on the wall; monkey stealing bananas; man reading newspaper; man walking; man about to fall; beggar lying on the pavement; man begging/ holding out begging bowl; boy chasing dog; dog stealing ball; bird flying; birds pecking seed from ground

As the number of information units differs between both pictures only the conciseness ratio was calculated and analysed rather than the raw number of information units. The conciseness ratio measures the relation of information units to the number of words. The more information units a participant used per word count the more concise their picture description is. Conciseness is one measurement of complexity besides the ratio of main clauses to subclauses or TTR/MTLD. When the focus lies on different functions the conciseness of a text might be reduced the higher the complexity of a task is. An example could be the higher demand of having to interpret an unfamiliar picture.

The mixed three-way ANOVA revealed a significant effect of Stimulus [F (1, 97) = 17.68, $p < .001$], Task [F (1, 97) = 64.86, $p < .001$] and Administration Group [F (2, 97) = 17.55, $p < .001$] as well as a significant interaction of Task and Administration

Group [*F* (2, 97) = 12.09, *p* < .001] and a significant three way interaction [*F* (2, 388) = 8.01, *p* < .001].

The post-hoc analysis of the two-way repeated measures ANOVA examining the effects of Stimulus and Task for each group revealed no significant interaction between Stimulus and Task in the online-testing group [*F* (1, 39) = .72, *p* = .40]. However, there was a separate significant effect of Stimulus [*F* (1, 39) = 22.33, *p* < .001] and Task [*F* (1, 39) = 10.07, *p* = .003]. The conducted ANOVA revealed a significant interaction effect in the in-person testing group [*F* (1, 29) = 7.41, *p* = .011] and the video-call testing group [*F* (1, 29) = 5.31, *p* = .029] as well as a separate significant effect of Stimulus (video-call: *F* (1, 29) = 6.62, *p* = .015) and Task [in-person: *F* (1, 29) = 22.94, *p* < .001; video-call: *F* (1, 29) = 30.95, *p* < .001) in both groups.

The pairwise comparison between the Stimuli in the different tasks in the in-person testing group revealed a significant difference of the Conciseness Ratio on both the spoken [*t* (29) = -6.79, *p* < .001] and written task [*t* (29) = -5.04, *p* < .001]. The participants had a higher conciseness ratio in the Indian Street Scene. In the videocall-testing group, the pairwise comparison between the Stimuli in the different tasks revealed a significant difference in the Conciseness Ratio in both the spoken [*t* (29) = -7.62, *p* < .001] and written task [*t* (29) = -6.63, *p* < .001] with a higher conciseness ratio in the Indian Street Scene in both tasks. Figure 54 presents all values.

The cultural origin of a stimulus might influence the conciseness of a picture description task. The participants had a higher conciseness ratio in their descriptions of the Indian Street Scene.

Figure 54: Effect of Stimulus, Task and Administration Group on the Conciseness Ratio

A further interesting variable was analysed to determine the influence of the cultural origin of the picture stimulus on the completeness of the picture description: The Percentage of Possible Information Units Produced. For the computation of this variable for the Cookie Theft Picture, the Number of Information Units participants used was divided by 22 (the number of all possible information units that could have been produced in the Cookie Theft Picture) and multiplicated by 100. For the Indian Street Scene, the Number of Information Units was divided by 36 (the number of all possible information units that could have been produced in the Indian Street Scene) and multiplied by 100.

Nevertheless, the three-way ANOVA did not reveal any significant effects. There was no significant effect of Stimulus [F (1, 97) = .73, p = .394], Task [F (1, 97) = .01, p = .927] or Administration Group [F(2, 97) = .03, p = .968], as well as no significant interaction of Administration Group and Stimulus [F (2, 97) = 1.08, p = .340] Task and Administration Group [F (2, 97) = .88, p = .420] or Stimulus and Task [F (1, 97) = 1.49, p = .225] or Stimulus, Task and Administration Group [F (2, 97) = .31, p = .732]. Figure 55 visualises all values.

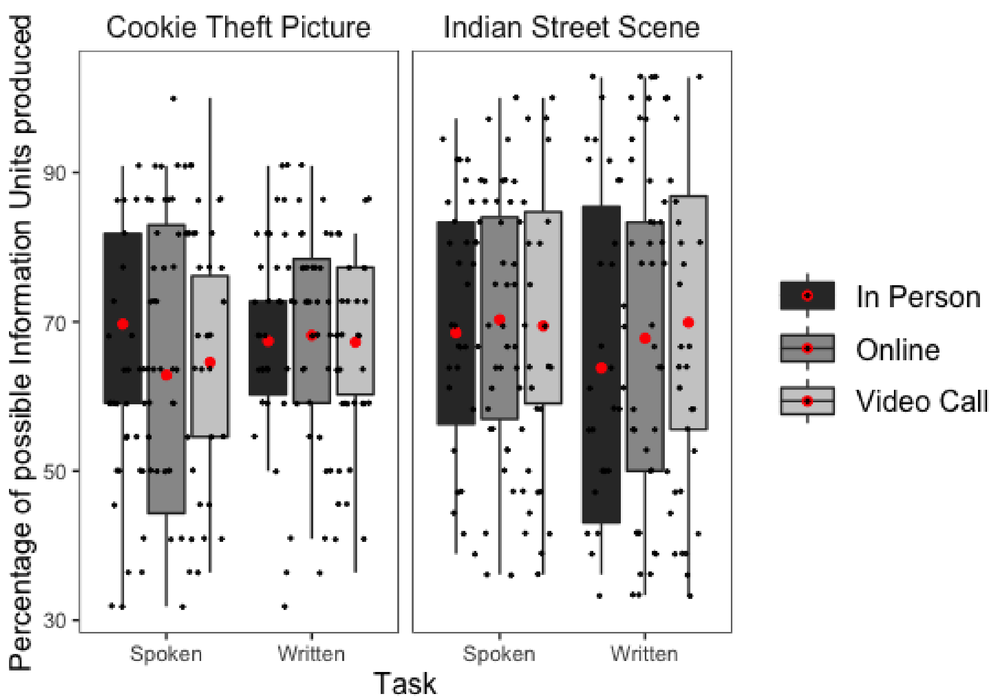


Figure 55: Effect of Stimulus, Task and Administration Group on the Percentage of possible Information Units used

7.6. Syntax

The number of clauses a participant used is a first indicator for the general length as well as syntactic complexity of a picture description. As with number of words the participants used in average more sentences to describe the Indian Street Scene (spoken: M = 16.1, SD = 4.4; written: M = 9.8, SD = 3.0) than the Cookie Theft Picture (spoken: M = 11.2, SD = 5.7; written: M = 8.1, SD = 4.4) in both tasks. However, the number of subclauses the participants used in their description of the Cookie Theft Picture (spoken: M = 3.56, SD = 2.9; written: M = 2.7, SD = 1.9) and the Indian Street Scene (spoken: M = 4.4, SD = 3.27; written: M = 2.6, SD = 2.0) did not differ to a great extend in both pictures and for both tasks. The participants produced very rarely incomplete clauses (Cookie Theft Picture spoken: M = 0.2, SD = 0.4; Cookie Theft Picture written: M = 0.5, SD = 1.3; Indian Picture spoken: M = 0.3, SD = 0.7; Indian Picture written: M = 0.3, SD = 0.9).

To evaluate the influence of the picture stimulus but also the task and administration group on the syntactic domain in consideration of the different lengths of the picture descriptions the syntactic complexity ratio and the syntactic error ratio were analysed.

The three-way mixed ANOVA of the syntactic complexity ratio revealed a significant effect of Stimulus [F (1, 97) = 5.60, p = .020] with the Syntactic Complexity Ratio being in average higher in the description of the Cookie Theft Picture (Cookie Theft Picture spoken: M = 38.4, SD = 37.1; Cookie Theft Picture written: M = 38.9, SD = 38.2; Indian Picture spoken: M = 33.5, SD = 28.0; Indian Picture written: M = 29.8, SD = 30.9). Furthermore, a significant interaction effect of Stimulus and Administration Group

[F (2, 97) = 5.36, p = .006] as well as Task and Administration Group [F (2, 97) = 3.80, p = .026] was observed.

In the post-hoc analysis of the two-way ANOVA for each Administration Group a significant effect of Stimulus was revealed in the in-person testing group only [F (1, 29) = 11.06, p = .002] and a significant effect of Task only in the online testing group [F (1, 39) = 6.37, p = .016]. There were no significant interaction effects (in-person: F (1, 29) = 2.40, p = .132; video-call: F (1, 29) = 1.14, p = .294; online: F (1, 156) = .05, p = .818). The results are displayed in Figure 56.

The results suggest that the Stimulus might influence the Syntactic complexity in the in-person testing condition. The Syntactic Complexity is greater in the descriptions of the Cookie Theft Picture. In the online testing group, the effect of the Task might be larger as discussed in more detail in section 5.3.2.

Figure 56: Effect of Stimulus, Task and Administration Group on the Syntactic Complexity Ratio.

The syntactic error ratio was too low across all participant groups, stimuli and tasks to be analysed or evaluated. Appendix J gives an overview of the exact values from the syntactic error ratio.

The three-way ANOVA shows that Stimulus, Task and Administration Group have mixed effects on the syntactic complexity. It can be concluded that syntactic complexity in a picture description task is not greatly affected by the choice of picture stimulus.

7.7. Semantics

Although previous studies used the Type Token Ratio to analyse the lexical diversity of a language sample for this study the Measure of Textual Lexical Diversity was chosen as it considers the length of a language sample.

In average the descriptions of the Cookie Theft Picture were lexically more diverse compared to the picture descriptions of the Indian Street Scene (Cookie Theft Picture spoken: $M = 50.5$, $SD = 13.1$; Cookie Theft Picture written: $M = 81.3$, $SD = 45.1$; Indian Picture spoken: $M = 44.4$, $SD = 7.4$; Indian Picture written: $M = 55.9$, $SD = 25.5$).

The three-way analysis of variance revealed a significant effect of Stimulus [F (1, 97) = 46.89, $p < .001$], Task [F (1, 97) = 62.55, $p < .001$] and Administration Group [F (2, 97) = 8.05, $p = .001$]. Furthermore, a significant interaction effect of Stimulus and Administration Group [F (2, 97) = 21.10, $p < .001$], Stimulus and Task [F(1, 97) = 16.91, $p = .001$] as well as Administration Group and Task [F (2, 97) = 13.00, $p < .001$], and Stimulus, Task and Administration Group [F (2, 97) = 16.62, $p < .001$] was observed.

In the post-hoc analysis the two-way ANOVA of Stimulus x Task for each group revealed a significant effect of Stimulus only on the in-person-testing group [F (1, 29) = 5.22, $p = .030$] and Task only for the video-call testing group [F (1, 29) = 16.61, $p < .001$]. The two- way ANOVA for the online testing group revealed a significant effect of Stimulus [F (1, 39) = 83.33, $p < .001$] and Task [F (1, 39) = 84.76, $p < .001$], as well as a significant interaction of Stimulus and Task [F (1, 39) = 48.06, $p < .001$].

A pairwise comparison for the online testing group in each task revealed a significant difference for both the spoken [t (39) = 5.38, $p < .001$] and written task [t (39)

= 8.24, $p < .001$] with participants having a higher MTLD in their descriptions of the Cookie Theft picture. All results are presented in Figure 57.

The presented results suggest that the Stimulus affects the lexical diversity of a picture description task. The lexical diversity was higher in the descriptions of the Cookie Theft Picture. This effect was most pronounced in the written task of the online testing group.

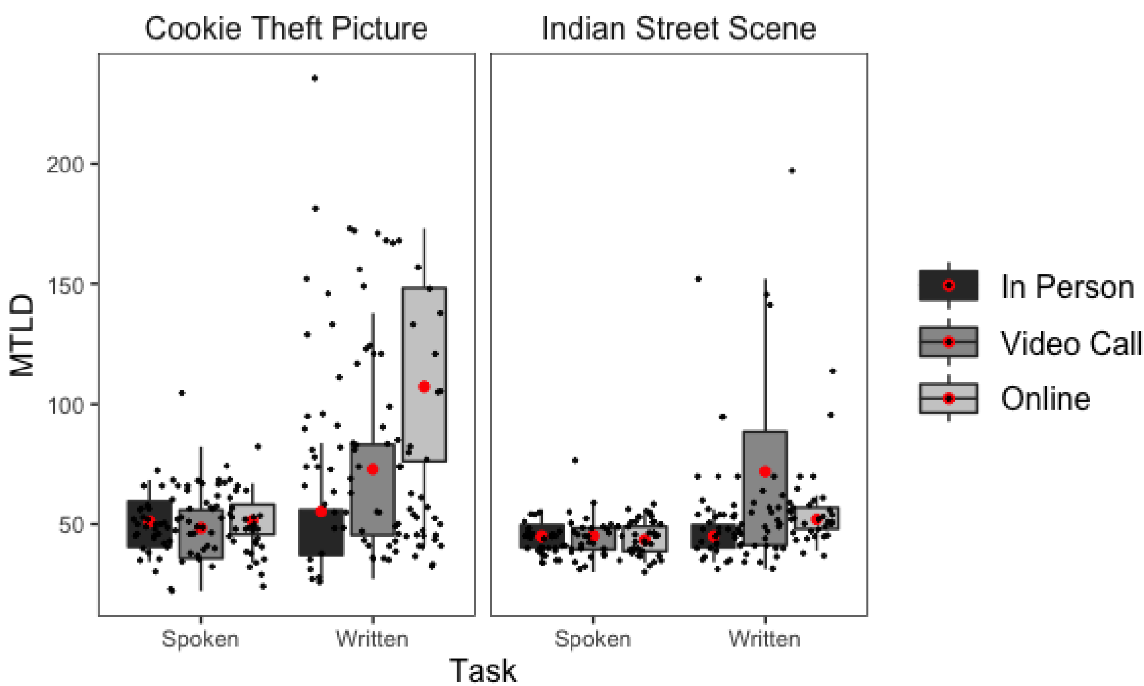


Figure 57: Effect of Stimulus, Task and Administration Group on Measure of Textual Lexical Diversity

Lexical Errors occurred very rarely across all tasks and stimuli in the description of the Cookie Theft Picture and the Indian Street Scene. Therefore, an analysis of the means was not possible and would have led to inconclusive results. The exact values of the Semantic Substitution Ratio can be taken from Appendix K.

As described in chapter 2. a detailed analysis of disfluencies would go beyond the scope of this study. To explore potential differences in this factor, disfluencies were coded at a coarse level. In this study, disfluencies are defined as the uncertainty of word choice in participants. About 10% of daily produced utterances contain natural disfluencies which are mostly due to underlying problems in the formulation and planning of upcoming speech (Schnadt, 2009).

To analyse the percentage of disfluencies occurring in the descriptions of both stimuli and also to compare the results of this study to the studies described in chapter 2 the disfluency ratio was calculated. The disfluency ratio comprises the sum of disfluencies divided by the word count and multiplied by 100.

In the oral task silent and filled pauses, repetitions, revisions and not identifiable words were counted as disfluencies. Repetitions, revisions, discontinued sentences, not identifiable words, crossed-out letters and words, punctuation errors, spelling errors as well as word onset errors were marked as disfluencies in the written task.

On average, the participants had an equally high disfluency ratio in both of their descriptions of the Cookie Theft Picture (Cookie Theft Picture spoken: $M = 5.9$, $SD = 5.9$; Cookie Theft Picture written: $M = 6.0$, $SD = 16.1$). However, in the description of the Indian Street Scene the disfluency ratio differed greatly between both tasks with participants producing fewer disfluencies in the written task (Indian Picture spoken: $M = 5.4$, $SD = 5.2$; Indian Picture written: $M = 0.5$, $SD = 1.3$).

The three-way ANOVA revealed a significant effect of Stimulus [F (1, 97) = 17.84, $p < .001$], Task [F (1, 97) = 4.54, $p = .036$] and Administration Group [F (2, 97) = 5.98, $p = .004$]. Furthermore, a significant interaction effect of Stimulus and Administration Group [F (2, 97) = 18.14, $p < .001$], Stimulus and Task [F (1, 97) = 12.08, $p = .001$] as well as Administration Group and Task [F (2, 97) = 21.22, $p < .001$], and Stimulus, Task and Administration Group [F (2, 97) = 8.02, $p = .001$] was observed.

The two-way ANOVA of Stimulus x Task for each group revealed only a significant effect of Task in the in-person testing group [F (1, 29) = 20.76, $p < .001$] and the online testing group [F (1, 39) = 93.43, $p < .001$]. For the video-testing group a significant effect of Stimulus [F (1, 29) = 17.32, $p < .001$] and Task [F (1, 29) = 4.61, $p = .040$] was observed, as well as a significant interaction effect of Stimulus and Task [F (1, 29) = 8.79, $p = .006$].

The pairwise comparison for the video-call testing group in each task revealed a significant difference for the written task [t (39) = 2.57, $p = .014$] with participants having a higher Disfluency Ratio in their descriptions of the Cookie Theft picture. There was no significant difference in the oral condition [t (39) = -.30, $p = .763$]. Figure 58 presents all values.

According to the results, the task might have a greater influence on disfluencies than the cultural origin of the picture stimulus. However, in the video-call testing group, the Stimulus might affect the Disfluency Ratio in the written task.

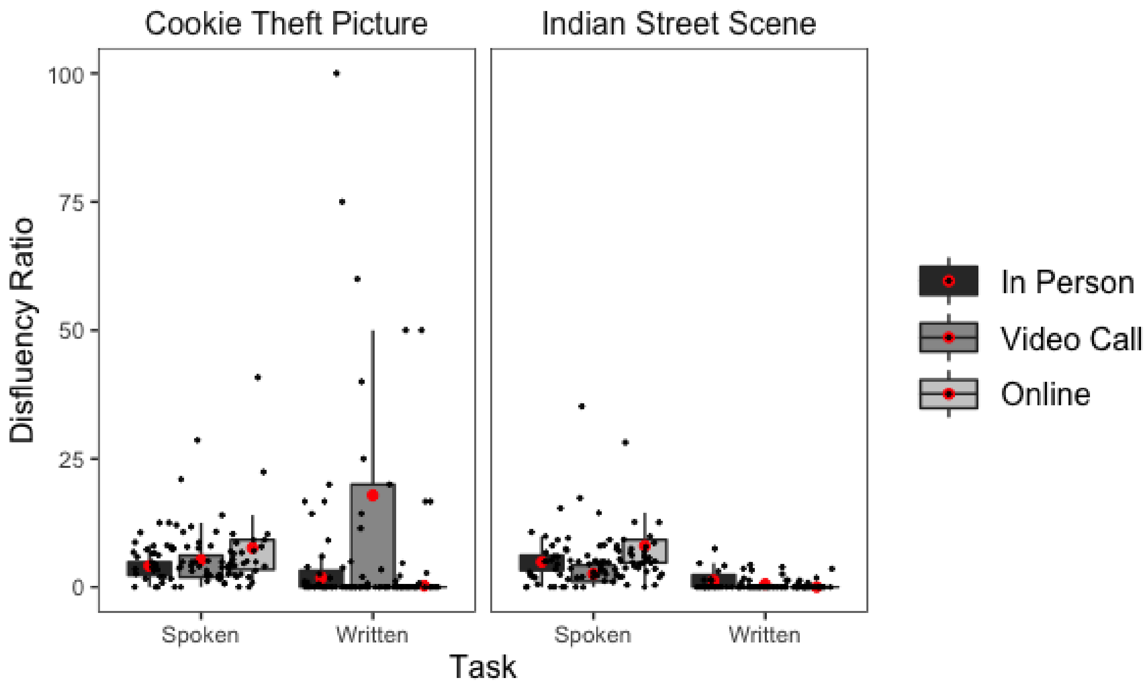


Figure 58: Effect of Stimulus, Task and Administration Group on the Disfluency Ratio

7.8. Phonology

In previous chapters, the domain of phonology and types of phonological errors were described. Phonological errors are errors on the last step of productive language and are usually observed in children or adults with neurological deficits.

Phonological errors occurred very rarely. The numbers were too small to reveal any significant differences between tasks or stimuli. Therefore, an analysis of variance was not possible and the exact values are presented in Appendix L.

7.9. Discussion

As discussed at the beginning of this chapter, the influence of the cultural origin and familiarity of a picture stimulus on the outcome of a picture description task has not been addressed in research yet. Across different countries, different picture stimuli are used. Some pictures have been adapted to cultural differences. However, is this necessary and does it make a difference to the quality and content of picture descriptions?

A group of British native speakers described two different pictures once in an oral picture description task and once in a written picture description task. During data collection, it was observed that participants frequently commented on the cultural origin of the picture. One of the pictures was the Cookie Theft Picture showing a familiar scene to the European eye despite its American origin. Participants merely used different words such as biscuit or tap to describe the scene. Some participants commented that this scene must be in the US as the word Cookies is written on the jar. The unfamiliar picture, the participants described, was the Indian Street Scene, which was chosen based on the shared vocabulary such as falling and stealing as well as boy and woman. In the testing sessions, it was observed that participants used more meta comments such as "this must be in India the way the people are dressed" or "it must be in a hot country as the man lying on the pavement is wearing almost no clothes". This suggests that the participants were sensitive to and aware of its familiarity. The purpose of this chapter was to ascertain whether the type of picture used had quantifiable effects on the language produced during the description.

The chosen picture stimulus might influence the number of words the participants used to describe both pictures. However, the Indian Street Scene has a higher number of information units that need to be included in the picture descriptions which might lead to more lengthy descriptions. The participants need more words to cover all characters, actions and locations in the Indian Street Scene. Therefore, more information units are needed, that were created to quantify how far the familiarity of the picture stimulus might influence aspects of quality such as completeness and conciseness of a picture description task.

The conciseness ratio differed significantly between both picture stimuli. The descriptions of the Indian Street Scene were more concise compared to the Cookie Theft Picture. This means that the participants included more information units per word in their descriptions of the Indian Street Scene. The picture descriptions were also relatively complete, as the participants included proportionally the same possible number of information units in their descriptions of the different picture stimuli. Familiarity might not influence the completeness of a picture description task. The length of the picture description is rather influenced by the number of information units that need to be covered. The Indian Street Scene contains more information units. Consequently, the participants used more words to describe the Indian Street Scene. However, the conciseness ratio increased. This suggests that the participants did not increase the number of words evenly. It is not clear whether the familiarity of the depicted scene or the number of information units influenced the picture description task. Contrary to the observation described in section 7.5. the misinterpretation and expressed meta-comments in the descriptions of the Indian Street Scene did not influence the conciseness or completeness of the description of the unfamiliar Indian Street Scene negatively. A plausible reason for this observation might be the participants using simpler language to describe the Indian Street Scene leading to a higher conciseness (Gustafsson, 2016).

One way to measure the complexity of language is the syntactic complexity ratio. The picture stimulus influenced the syntactic complexity of the in-person testing group.

The participants used more complex syntactic structures to describe the Cookie Theft Picture than the Indian Street Scene. However, the Task and the Mode of Administration might have had more influence on the video-call testing condition and the online-testing group. This is described in more detail in chapters 4 and 6.

Another measure of the quality of a picture description task is lexical diversity. The MTLD was higher in the descriptions of the Cookie Theft Picture compared to the Indian Street Scene. This effect was most pronounced in the written task of the online-testing group. This might have been a combined effect of picture stimulus and the writing mechanism of typing. Typing, as described in section 5.4., gave the participants the possibility to edit their text. The participants could focus more on editing their written description when describing the Cookie Theft Picture as the scene was more familiar to them. Compared to the Cookie Theft Picture the written/typed description of the Indian Street Scene demanded more focus to interpret the scene which led to a lower MTLD.

Disfluencies, which are a symptom of impaired lexical choice, were not equally influenced by the chosen picture stimulus. The chosen picture description task still has the greatest influence on the number of disfluencies a participant produced as silent or filled pauses that are characteristic of oral language does not occur in written language. This effect is described in chapter 3 in more depth. However, the participants of the video-call testing group had a significantly higher disfluency ratio in the written task of the Cookie Theft Picture than in the Indian Street Scene. As described in section 5.4. the participants of the video-chat group used a chat-style pattern to describe the Cookie Theft Picture in the written task. However, the data suggests that this pattern is not present in the written description of the Indian Street scene. This indicates that the participants did not resort to a chat-style pattern when they were describing a more complex or unfamiliar picture stimulus.

As described in previous chapters the phonological domain is almost not affected by the choice of the stimulus of the task nor by the mode of administration.

7.9.1. Implications of picture stimulus choice for diagnosing PPA

The chosen picture stimulus for picture description tasks can have an influence on the picture descriptions produced. In the previous section the influence of the chosen picture stimulus on healthy participants was discussed. The following section will discuss the possible influences on the diagnosis of PPA.

As presented above, the Indian street scene contained more information units. An increased number of information units could also be observed as presented in Chapter 6. A higher number of information units that need to be covered might influence the performance of people with PPA. People with nfvPPA might struggle more as more content needs to be covered and speech production is already effortful for them. For the convenience of the patient, it would be better to use a familiar shorter scene as it would already give enough information for the diagnostician. The length of the picture description might also lead to more difficulties for people with lvPPA as they might struggle with remembering what they already said. Concerning the length of the scene, people with svPPA might struggle least, as their symptoms concern mostly the lexical domain. Furthermore, in people with PPA the conciseness ratio might be overall lower as in lvPPA the word retrieval is impaired and svPPA presents with anomia. An unfamiliar scene might enhance these symptoms further. People with nfvPPA use simple and short phrases which might enhance the conciseness ratio for both the familiar and unfamiliar scenes.

The domain of syntax might also be influenced by an unfamiliar scene. Reduced syntactic complexity is the primary symptom of nfvPPA. The unfamiliar scene might lead to even shorter and simpler phrases, however further studies testing the influence of the familiarity of the picture stimulus on patients with nfvPPA are needed to support this assumption. In lvPPA a reduced syntactic complexity might lead to a misdiagnosis with nfvPPA in the worst case if only a picture description task is used for testing. In less severe cases the diagnosis might be more complicated. The syntactic complexity in svPPA might be less influenced by the familiarity of the picture as in these cases language production is not impaired on the syntactic level.

In healthy participants the lexical diversity was lower in their descriptions of the unfamiliar Indian Street Scene. It might be possible that this effect might be more pronounced in people with a lvPPA and svPPA. They have a lower lexical diversity compared to controls with no neurodegenerative disorders (Patterson & MacDonald, 2016). Furthermore, the unfamiliar scene of the Indian picture might require the use of lower frequency terms which might enhance word retrieval deficits. People with nfvPPA are likely less affected. However, in combination with the influence of the familiarity on the syntactic complexity the distinction between lvPPA and nfvPPA could become blurred. Both cases would present with short simple sentences and lowered lexical diversity.

Disfluencies are the main symptom of lvPPA and mostly produced when people with lvPPA are experiencing word retrieval problems. Disfluencies might therefore occur more in descriptions of unfamiliar scenes. People with svPPA are usually not aware of their lexical deficit and seem fluent. It is also not likely that the familiarity would affect the production disfluencies in nfvPPA as patients with nfvPPA present primarily with symptoms in the syntactic domain and effortful slowed language production.

Healthy participants did not produce phonetic or phonological errors. This was independent from the chosen picture stimulus. However, inconsistent phonological errors occur in nfvPPA. People with nfvPPA might produce more phonological errors when describing an unfamiliar scene as they might experience a higher demand. It would be less likely that the familiarity of the picture stimulus would have an effect on lvPPA and svPPA.

Overall, the familiarity of a picture could influence the diagnosis of PPA to an extent at which it will be hard to distinguish between certain subtypes of PPA. People with nfvPPA would likely struggle with an increased number of information units, like in the Indian picture in this study. They might produce even shorter sentences leading to a higher conciseness ratio. However, the combination of the influence of the familiarity of the picture and the characteristically reduced syntactic complexity with the already low fluency might blur the distinction to lvPPA. In lvPPA, the syntactic complexity would also be reduced, and word retrieval difficulties and disfluencies might be enhanced. Furthermore, people with lvPPA might struggle with the length of the picture description of unfamiliar scenes as these can be longer as presented in Chapter 6. The conciseness ratio might also be lower, due to the characteristic anomia which might be enhanced by the unfamiliar scene. This could blur the distinction to svPPA. People with svPPA could also present with a lower conciseness ratio and word retrieval difficulties. However, the familiarity of the picture would not have an influence on syntax, fluency or phonetic and phonological errors in svPPA.

7.10. Conclusion

Summarising, familiarity with a picture might affect the complexity of a picture description task. The participants used less complex syntactic structures to describe the unfamiliar Indian Street Scene. They also used lexically less diverse language in their descriptions of the more unfamiliar picture stimulus. Patients might find it harder to describe scenes that are not part of their daily life which might lead to misdiagnosis with for example semantic dementia. The distinction between nfvPPA might be more complicated. Wrongly diagnosed patients might not be treated correctly or overtreated, straining the budget in healthcare.

According to the results of this study, a good picture should present a culturally familiar scene to avoid possible complications for diagnosis. Even small differences between the cultural origin of the picture and the participants (US and UK) led to meta comments which can be a sign of confusion and make the task even harder for patients. The results also showed that an increased number of information units does not necessarily lead to a clearer understanding of the picture, at least in healthy participants. This could be different in patients and needs to be addressed in further research. Therefore, a good picture does not need to show a great number of information units as discussed in Chapter 6. In the present study the familiarity of the picture had the greatest influence in the syntactic and lexical domain. The leading symptoms in nfvPPA and lvPPA could therefore become blurred in consequence of culturally unfamiliar scenes.

This supports the importance of the previously discussed point for a good picture for a picture description task: A good picture should show a familiar scene from early life and should be matched to the patient’s cultural background.

Although the grade of familiarity of the picture stimulus had only a minor effect on other language features such as number of words, conciseness and phonology it might still be necessary to translate also cultural differences of picture stimuli when translating neurological tests. However, the necessity can only be proven entirely with data from patients with neurodegenerative diseases. Therefore, further studies researching the effect of the chosen picture stimulus on people with a language impairment are needed.

8. Conclusion

This chapter aims to summarise the key research findings to the research questions and research goals. It will also discuss the value and contribution to research as well as for clinical use. Limitations of the study will also be reviewed and opportunities for future research will be proposed.

8.1. Introduction

The thesis aimed to generate data on spoken and written picture description tasks in healthy adults aged 40 and older. Furthermore, this thesis aimed to explore which factors can influence picture description tasks in healthy adults. The data of healthy participants can be used in future studies as a baseline to which data of patients with aphasia or neurodegenerative disorders can be compared.

With a mix of literature reviews and empirical data the present thesis aimed to discuss four main questions:

- What are the differences between spoken and written picture description tasks in healthy adults and would a written administration come to a comparable result to the traditional spoken task? (Chapters 1-3)
- Do specific language functions or cognitive abilities influence the performance in spoken and written picture description tasks of healthy adults? (Chapter 4)
- How does the administration mode influence the performance of healthy adults on spoken and written picture description tasks? (Chapter 5)

- In how far does the cultural origin of a picture stimulus influence performance on spoken and written picture description tasks in healthy adults? (Chapters 6 and 7)

8.2. Overall findings and contributions to the field

8.2.1. Differences between spoken and written picture description tasks

Previous literature and widespread clinical practice can give the impression that picture description tasks are a well-suited diagnostic instrument for the diagnosis of post-stroke aphasia and neurodegenerative forms of aphasia (see sections 1.1. and 1.3). However, as discussed in Chapter 1, the comparability and reliability of the task are limited as there are multiple different standardised diagnostic procedures tailored for different purposes for picture descriptions, which makes it hard to compare results and data from different studies or clinics.

Moreover, in some cases, oral picture descriptions cannot be obtained from patients in cases of severe apraxia in which written language is the only preserved means of communication (e.g., due to motor impairment affecting speech production). In these cases, the assessment of written language could be of crucial importance. However, the assessment of spoken language has long been the default. The need for a picture description using written language found a practical necessity in the assessment of language abilities in connection with other disorders such as in people with motor neurone disease. An illustration of this could be Bak et al. (2001) where a written picture description task was the only way to obtain a language sample.

However, the assessment of spoken picture description tasks is also very time-consuming. The common procedure of picture description tasks requires the patient to describe a picture orally. After the assessment, the diagnostician has to transcribe the picture description before it can be analysed. A written assessment would be, therefore, more efficient. Furthermore, there are multiple different standardised processes tailored for different purposes for the analysis of picture description tasks making research and diagnostic less comparable. Previous studies report mostly data from adults with vascular and neurodegenerative aphasia. However, data from healthy adults is very rare and often only assessed as control or samples were taken from different settings such as scientific work and dinner conversations. Despite this, in general, in clinical work physiological processes are studied before pathological processes can be derived and explained.

This thesis compared, therefore, spoken and written picture descriptions in healthy adults to generate normative data. According to our knowledge, this is the first piece of work analysing the differences between spoken and written picture descriptions systematically. To address previous research gaps the number of possible linguistic symptoms that also occur in healthy language was increased to gain a more differentiated picture. Spoken and written picture descriptions were also taken from the same picture, making the samples more comparable. Differences were more likely due to the language production modality rather than the topic or communication setting.

Previous research has revealed that spoken and written language differs in many important aspects. Not only the modality but also patterns of use, styles, prestige, and production speed lead to significant differences such as written language being shorter and more complex. Previous research has been inconsistent as different measures were used. Furthermore, the oral and written samples compared in previous studies were sometimes obtained in different situations, such as comparisons between scientific essays and dinner conversations. Therefore, an analysis of spoken and written language samples of a comparable task was necessary to compare differences between spoken and written language to find differences that are specific for each modality. Furthermore, this thesis introduced a more diverse and extended set of measurements to get a more nuanced impression of the differences between spoken and written picture description tasks. According to previous research in spoken and written language the same semantic base and lexico-syntactic systems are used. Variations are mostly expected in the choice of vocabulary and syntactic types due to modality-specific constraints.

In the presented MSc studies of Chapter 2 patients were affected very heterogeneously across both modalities which made building a general pattern difficult. The healthy control group produced fewer words in the written task as well as fewer main clauses but more complex syntactic structures. The number of grammatical errors decreased in the written sample while the conciseness ratio increased. These results are in line with previous theories. Writing has a slower production speed compared to spoken language leading to shorter language samples that can be syntactically more complex.

The data in this thesis replicate and confirm previous data and hypotheses. The study described in Chapter 3 had an increased number of participants and symptoms which included for example two measures of lexical diversity. Spoken picture descriptions were longer but less dense in information as well as less syntactically complex and lexically diverse. No differences were found in the number of syntactic, semantic or phonological errors. The results confirm previous hypotheses of spoken and written language using the same semantic base and lexico-syntactic system with expected variations in word choice and syntactic types. The presented results suggest that differences might be mostly due to differences in the production speed of spoken and written language. However, it is not clear whether this difference lies in the level of language planning or implementation. Previous literature suggests that the permanence and intention of written language might influence the production of written language already on the planning level. Spoken picture descriptions on the other hand might be planned and produced almost simultaneously. In both cases, further research is needed to either confirm or reject these hypotheses.

The collected data is the first base for data from unimpaired people on differences between spoken and written language. Errors in syntax, phonology and semantics were very low in both modalities. Hence the errors in these linguistic domains might be an indicator of pathological processes and could be detected in both spoken and written assessments of patients.

For clinical work, the data on differences between spoken and written picture description tasks suggest that the choice between oral and written administration depends on the clinician's objective but is also a necessity. Oral picture description tasks provide more output (Number of Words) that can be analysed. Furthermore, communicative behaviour such as gestures can be analysed which is not possible in written language samples. However, written language samples can be used as an indicator for a first evaluation that is fast and efficient.

8.2.2. Influence of other language functions and cognitive abilities

Specific language functions and cognitive abilities might influence connected language production and therefore picture description tasks. This thesis explored correlations between key symptoms in picture description tasks and naming abilities, grammar, possible underlying dyslexia, concept formation skills, visuospatial abilities, social cognition, and intelligence. Previous studies revealed a correlation between connected language production tasks and naming abilities, possible underlying dyslexia, concept formation skills, visuospatial abilities as well as social cognition. Hypotheses were derived from previous studies and were tested in this thesis. For a correlation between isolated grammar tests or intelligence and connected language tasks, no specific previous data was found. Hence, an exploratory analysis of correlations between isolated grammar and intelligence with key symptoms in picture description tasks was conducted.

Correlations between specific language functions and the performance in written picture description tasks were revealed, however, cognitive abilities might not have a significant influence on picture description tasks. A paradox negative correlation was found between naming in the Graded Naming Test (GNT) and the Number of Information Units in the written task. Participants that could name more words in the GNT produced a lower number of information units in their picture descriptions. The direction of the correlation was unpredicted by previous literature. The GNT has no ceiling effects (Cambridge Cognition Ltd, 2022) reducing the possibility of skewed results. However, the GNT only tests the naming of objects. It might be possible that participants with a high score in the GNT were focused more on describing objects and did not feel the need to describe objects with actions. However, the participants produced fewer information units including actions. A replication study is needed to investigate the possible link between naming abilities and the number of information units. Should the effects not be replicable they might be a spurious correlation. A further negative correlation was revealed between the Northwestern-Anagram Test (NAT) and the Syntactic Error Ratio in the written task. As expected, participants with better grammatical abilities produced fewer syntactic errors. The NAT score was furthermore correlated negatively with the lexical diversity measures Type Token Ratio (TTR) and Measure of Lexical Diversity (MTLD) in the written task. These correlations suggest that participants that have good grammatical skills might be lexically more sophisticated.

All in all, no single ability or function predicts the performance of a healthy person on connected language tasks such as picture description tasks. Significant correlations were revealed in the written task only and the direction of correlations was not always as predicted. A reason for the low number of significant correlations might be the small sample size and ceiling effects in tests in a healthy sample. These limitations will be discussed further in section 8.3. Previous research on healthy participants supports the assumption that properties of discourse in picture description tasks including the Cookie Theft Picture can be separable from other language task scores (Alyahya et al., 2020).

From the results, it can be concluded for practical purposes that isolated language functions and cognitive abilities cannot predict performance on picture description and connected language tasks. This implies for clinical work that isolated (e.g., naming and processing of passive) as well as connected language functions (e.g., picture description tasks and spontaneous speech samples) need to be included in the assessment before a specific treatment plan can be set up.

8.2.3. Influence of administration mode

In the last years, technical advancements made video-call and online testing possible, and it became a necessary tool in the Covid-19 pandemic. However, only seven studies have compared data from in-person and video-call testing since 1997 in adults, generally revealing no significant differences. Furthermore, the studies were conducted in small groups. Detailed linguistic symptoms have not been compared but rather a syndrome classification and lesion location which might be important for a diagnosis. However, it is not sufficient for setting up a treatment plan tailored to the patient's needs. In addition, video-call testing and online testing settings have limitations in the choice of testing methods. Tests that need manipulation of objects (testee interacting with objects for example in the Delis-Kaplan Sorting Test; Delis et al. 2001) or require participants to draw, write by hand or do movements with extremities other than their hands are difficult to translate into a video-call or online testing setting. Due to this lack, handwritten language samples are complicated to obtain which raises the question of whether handwritten and typed language samples are comparable.

To address the presented problems and research gaps the present study compared three administration modes (in-person testing, video-call testing and online testing) and the influence of writing mode (handwriting vs. typing). 30 participants completed the in-person testing condition, 30 participants the video-call testing condition and 40 participants completed the online testing session.

In the oral picture description task, no significant differences were revealed between the in-person testing group and the video-call testing group. In contrast, participants in the online testing group produced significantly longer and lexically more diverse picture descriptions compared to the other groups in both the oral and written task.

The written task revealed further significant differences. Most of these might be the result of differences between handwriting and typing which were analysed in more depth. The participants in the video-call testing group used a chat-style pattern in their written picture description tasks that was characterised by short and often incomplete sentences. Furthermore, participants who typed their picture description task, which was half of the video-call testing group and all participants of the online testing group, could make use of spelling and grammar checking software, which might have masked errors. However, this study could not control for this and hence cannot give evidence for this hypothesis. Moreover, it raises the question of how far too much control over the technical devices of the person being tested influences the tester and testee relationship negatively. Therefore, as the use of spelling and grammar checking software would be very difficult to control in clinical practice, the presented results remain ecologically valid. However, as described in more detail in section 8.3. further research is needed to evaluate the influence of spelling and grammar checking software on online assessment of connected language tasks.

The results suggest that in-person testing might still be the best administration method for connected language tasks as it was the method that was least interrupted by technical problems and the pattern of the results was mostly comparable to previous data on differences between spoken and written language. Furthermore, for most participants, the in-person testing format was the most natural communication setting as most testing situations still take place in face-to-face situations. However, video-call testing might be a reliable substitute if in-person testing is not feasible.

Automated online testing without direct interaction between diagnostician and patient might be an option to gain a first impression in a remote setting. However, more commonly available technical advancements (e.g., tablets for drawing, tools that can turn off spelling and grammar checking software easily) might be necessary to obtain reliable results. In the used online format, a lot of observational data was not available after the testing sessions. Some participants might have had problems understanding the tasks; however, it is not clear as there was no feedback available. Only one participant wrote a mail after the testing session giving feedback on how he found it hard to navigate through some of the online testing parts. This participant was also the only one who started the testing session and gave up, as he wrote in his mail, to start the testing session at a later point when he had more time to start and finish the testing session. The results revealed that participants might have felt the need to explain more of the pictures as there was no person present that could give immediate feedback (e.g., nodding, non-verbal feedback and encouraging sounds). Participants in the in-person and video-call testing conditions asked for/demanded feedback by asking for example: “Is this enough?” or “Did I forget anything?” and using meta-comments: “I think I have covered it all.” or “I’m not sure if I have described everything.”. Additionally, some tests could not be performed in the online testing environment such as tasks involving drawing or manipulating objects. There are programs available today that might help include these tasks in an online testing environment, however, they are expensive or not widely available. Drawing might be more common on tablets now; however, it is still quite novel, particularly for the older population (Kosowicz & MacPherson, 2016). Therefore, it might be challenging to implement testing procedures focussing on neurodegenerative diseases.

Furthermore, a comparison of administration modes was performed with healthy adults. Individuals with aphasia or neurodegenerative diseases might be dependent on help from a caregiver or family member which might influence the test results. In common with previous studies, it was observed in the present study that participants relied often on help from either caretakers or family members to complete the testing session. This behaviour was not quantified, and it is not clear what effect it may have had on results. Therefore, future research should include measurements for received help of caretakers or family members present and how far this might influence the performance of the participant/patient.

8.2.4. Influence of the cultural origin of picture stimuli

The main part of a picture description task is the picture stimulus. A good picture for a picture description task should fulfil several criteria to be well-suited: a clear focus, aspects of time, person and place, a well-defined context, describable themes and events that are easily identifiable and comprehensible, and contains vocabulary from early childhood. In addition, the description must be predictable to be comparable across individuals and the scene should be familiar (see the introduction of Chapter 6). The point of familiarity is not always fulfilled for pictures. As described in section 5.4. the culture and region a person is raised in influences their perception and interpretation of a picture and might hence influence their picture description. Over time, pictures have been adapted to changing technology, social norms as well as other countries and cultures. However, to our knowledge, the influence of different cultural origins and therefore the grade of familiarity with a person have never been compared directly.

Numerous pictures exist that are used for picture description tasks in clinical language assessment. However, there are huge differences in what pictures depict and not all fulfil the criteria for a good picture as described above. It is not always described why a certain picture was chosen even for widely used diagnostic batteries. The main concern is that most pictures do not fulfil the criterion of depicting a scene that is familiar to people from different cultural backgrounds. However, pictures are used in different cultures without being adapted to the cultural background which might lead to misdiagnosis. Nevertheless, unfamiliar pictures could have the function of making the test more difficult and hence more sensitive. For diagnostic purposes it is therefore important to take possible cultural influences into account when interpreting the results.

To obtain an impression of which influence the familiarity of a depicted scene has on picture description tasks 100 healthy participants have been compared. The participants described the Cookie Theft Picture and the Indian Street Scene, both once oral and once written, resulting in four picture descriptions in total. A comparison of all picture descriptions in all administration modes revealed that the British participants misinterpreted the Indian Street Scene often. The descriptions of the Indian picture had a lower informational density and were less concise. Furthermore, a higher number of meta comments was observed in the testing sessions when participants described the Indian Street scene. However, no significant impact on syntax or lexical diversity was revealed.

It can be concluded that the familiarity of a depicted scene in a picture description task might influence the completeness of a picture description task.

The results of the literature review and the collected empirical data suggest that pictures are not universally comprehensible. Pictures for picture description tasks need to be translated and adapted to the cultural background of the country in which it will be administered as it is more widespread for linguistic stimuli. For linguistic stimuli, there already exists more awareness and appreciation for cultural differences that need to be transferred to visual stimuli such as pictures.

8.3. Limitations and Recommendations

One aspect that reduces generalisability for the entire thesis is that the sample sizes were relatively small. Furthermore, the sample was recruited from Edinburgh, the city with the highest number of graduates in Scotland. Compared to the rest of the United Kingdom, in Edinburgh more than 41% (Census, 2021) of the population holds a degree while only 22% (Office for National Statistics, 2017) in the rest of the UK hold a degree. The participants might have produced more sophisticated picture descriptions with fewer errors and better informativeness. Due to the small sample size, data about the healthy ageing process could not be collected. Instead, data from a broad age range was gathered and analysed. To address the limitation of the small highly educated participant sample, future research should increase the sample size and recruit participants from more diverse educational backgrounds. Furthermore, the number of participants in different age groups needs to be increased to investigate healthy ageing processes.

However, the concept of generalisability must be looked at more closely. Research results but also diagnostic results have to be compared in the background of the sample they were collected with. For example, the Body Mass Index was created by Adolph Quetelet in the 1830ies. He measured the weight and height of Scottish conscripts. He assumed that the mathematical mean of a population was the ideal (Eknoyan, 2007). Consequently, people deviating from the "ideal" were labelled as overweight or underweight. The BMI is still widely used today and a diagnostic tool. However, the BMI is and was initially a way of measuring populations, not individuals and was designed for statistics and not for measuring individual health. The BMI for example overestimates the obesity in African Americans (The Endocrine Society, 2009), illustrating the importance of interpreting results in the context of the population they were obtained from.

Throughout the entire thesis, all data were collected, analysed and interpreted by only one trained Speech and Language therapist. Therefore, reliability in the assessment and linguistic analysis might be biased. Linguistic analyses were carried out according to the spontaneous language guideline from the Aachener Aphasie Test (AAT; Huber et al., 1983) in which I was trained as a professional speech and language therapist. Statistical analyses were carried out and interpreted using several sources and programmes. However, research biases cannot be excluded entirely. With greater resources, it would have been ideal to have a subset of the language data coded by a second SLT to confirm reliability in coding. Future studies should follow a type of blind study design. The researchers testing the participants should not be the same person analysing the data as well as coding and interpreting the data. Additionally, at least two different researchers should annotate the language samples to address possible differences that can be discussed so that more reliable data can be analysed.

When the interaction of single language functions and cognitive abilities with connected language was analysed, ceiling effects of tests, such as the NAT and ECAS, especially in healthy individuals, might have led to fewer significant correlations that could not confirm results and hypotheses from previous studies. It was also unlikely to get sufficient data from people with dyslexia in the presented sample as the sample size was comparatively small. This led to non-significant results in the correlation analysis. Some analyses revealed an unexpected negative correlation. To obtain more conclusive correlations in the interaction of single language functions and cognitive abilities with connected language, future research should focus on a choice of tests without ceiling effects which are especially important when testing healthy participants. Future studies also need to confirm the results of explorative analyses in this study. Additionally, a replication study for unexpected directions of correlations in this study is needed to either confirm or reject the results of this study.

Due to the Covid-19 pandemic, it was not possible to get access to patient data. Consequently, several hypotheses could not be tested regarding the relationship between healthy and pathological language performance. In the case of differences between spoken and written picture description tasks, it could not be studied whether phonological, syntactic and semantic errors occur solely or more often in patients with neurodegenerative diseases that affect the linguistic domain. Furthermore, patients might have correlations between isolated language functions or cognitive abilities that might influence the performance of picture description tasks. Motor constraints might for example influence the performance in naming tasks as well as word production in the picture description task.

Future research should also include patients with neurodegenerative disorders affecting the linguistic domain as they are less explored in than stroke aphasia. The studies could focus on testing different hypotheses of this thesis, amongst others, whether phonological, semantic and syntactic errors occur solely or more often in patients with neurodegenerative diseases that affect the language domain. Lastly, patient data is necessary as an influence of isolated language functions or cognitive abilities on picture description tasks might be more likely in people with neurodegenerative diseases.

In the comparison of the different administration modes, the subdivision in a handwriting group and typing group decreased the sample size further, as described above, decreasing generalisability. Differences between handwriting and typing might become even more important in future with technology being increasingly widely used. To gain a more differentiated picture of how people use technology the sample size needs to be increased and recruited from a more diverse background. Participants need to be recruited from different educational backgrounds and socioeconomic statuses. Furthermore, the age of the participants and their habit of using electronic devices needs to be evaluated and included in the analysis.

There were also some unavoidable differences between different administration modes in terms of timing and order of tasks. The in-person testing group attended two separate testing sessions which were 7 days apart and lasted 45 minutes to one hour each. The video-call testing group and the online testing group took part in only a single testing session that lasted 30 minutes to one hour. It is not clear whether this might have led to changes in the performance of the participants. The order of the pictures (Cookie Theft Picture and Indian Street Scene) and description modes (oral or written administration) was fully counterbalanced to avoid order effects. The order of the tests between the picture description tasks was not randomised, however. In the in-person testing sessions, the order of the pictures and description modes did not influence the length of the picture descriptions or their informativeness. As the picture descriptions did not differ significantly in length a fatigue effect can be excluded additionally. The video-call testing sessions and the online testing were also shorter as not all tests could be performed in the virtual setting. Tests that need the manipulation of objects, drawing or handwriting could not be always performed. Some tests might have been translated in the online testing environment with higher programming expertise or better technical devices. However, due to time limits and cost constraints, this was not feasible in this thesis. Future studies should focus on planning more comparable testing sessions, as the differences in the structure of the testing session(s) might have influenced the comparability of the different administration modes. Testing sessions in future research should use the same tests that are executable in all administration modes. Furthermore, the duration of the testing session needs to be comparable. To obtain more conclusive correlations in the interaction of single language functions and cognitive abilities with connected language, future research should focus on a choice of tests without ceiling effects which are especially important when testing healthy participants. Future studies also need to confirm the results of explorative analyses in this study. Additionally, a replication study for unexpected directions of correlations in this study needed to either confirm or reject the results of this study.

The influence of the cultural background of a picture stimulus is not entirely generalisable. I originally planned to compare the UK sample of participants with a group of participants from India. However, a sample with participants from India as well as patients could not be tested due to time constraints and the Covid-19 pandemic. Data from participants from India could have revealed a possible double dissociation which would have supported the hypothesis that picture stimuli need to be translated and adapted to the cultural background as well as their linguistic counterparts. To confirm the results and hypotheses of the influence of the cultural origin of picture stimuli on picture description tasks a replication of the study with participants from India is necessary.

In the statistical analysis, meta-comments could not be quantified. Hence the observation of an increased number of meta-comments produced by the participants in the picture description task of the Indian Street Scene cannot be assessed formally. A post-hoc questionnaire assessing the participants' impression and perception of the unfamiliar Indian Street Scene might have added further valuable data. However, an additional questionnaire might have lengthened the comparably long testing session which might have led to a further decrease in participants and sample size. To allow a statistical analysis of the increased number of meta-comments a meta-comment variable needs to be added to the initial annotation process. Furthermore, a post-hoc questionnaire testing the perception of the unfamiliar depicted scene could be added in future studies to analyse the impression of the participants.

8.4. Closing summary

This thesis aimed to generate data of spoken and written picture description tasks from healthy adults and to determine which factors might influence their produced picture descriptions. Therefore, four main components were tested, analysed and interpreted separately: differences between spoken and written picture descriptions, influence of isolated language functions and cognitive abilities, the influence of the administration mode and influence of the cultural origin of a picture stimulus.

The results suggest that spoken picture descriptions are longer, consisting of more words and sentences, whereas written descriptions are more concise, syntactic complex and lexically diverse. Additionally, healthy adults rarely produce phonetic, syntactic and semantic errors. Single language functions and cognitive abilities might not predict the performance of healthy adults in picture description tasks, though the sample size precludes drawing a strong conclusion about this. Healthy adults perform also differently in remote testing conditions. While the participants performed similarly in the spoken picture description task of the in-person testing and video-call testing condition, the participants in the online testing group produced longer descriptions in both the spoken and written picture description task. Differences in the written task were likely caused by the writing mode. Typed descriptions contained fewer errors which were likely due to automated spell check software. Furthermore, participants in the video-call testing session used a chat-like writing pattern which might be due to the nature of the video-call setting. Lastly, participants tend to misinterpret unfamiliar scenes depicted in pictures with a different cultural background.

This thesis introduced an expanded list of measures, or more closely linguistic features, which allowed a more varied analysis. Furthermore, the number of participants was expanded, and healthy data were collected. Spoken and written language was compared using the same language production task which improved the method used in previous studies comparing spoken and written language. Therefore, differences are most likely caused by differences between both modes and not the task. This thesis was also the first study that compared automated online testing as a form of linguistic assessment for diagnosis and compared the performance of participants in different language domains in three administration modes. Previous research only compared the general diagnosis or lateralization of the lesion. Finally, picture stimuli that are used for diagnostic and picture description tasks were gathered and compared. The influence of the cultural origin of a picture stimulus was investigated for the first time.

Overall, spoken and written picture descriptions do not differ significantly in linguistic errors (phonetic, syntax and semantic) that are observed in and produced by patients with post-stroke aphasia or neurodegenerative diseases affecting the language domain. In the assessment of language isolated and connected language functions need to be included in the diagnostic process as the performance of individuals can be heterogeneous. A remote administration of certain diagnostic tests is possible. However, when choosing a remote administration mode, certain differences need to be considered as the performance of an individual is influenced by the technical device used. Finally, picture stimulus and the cultural background of an individual need to be coordinated as the interpretation and performance on a task might be compromised which might lead to misdiagnosis.

References

Abrahams, S., Bak, T., & Newton, J. (2021). *Edinburgh Cognitive and Behavioural ALS Screen (ECAS)*. The University of Edinburgh.

Abuom, T. O., & Bastiaanse, R. (2012). Characteristics of Swahili–English bilingual agrammatic spontaneous speech and the consequences for understanding agrammatic aphasia. *Journal of Neurolinguistics*, *25*(4), 276–293. https://doi.org/10.1016/j.jneuroling.2012.02.003

Adlam, A. L. R., Patterson, K., Rogers, T. T., Nestor, P. J., Salmond, C. H., Acosta-Cabronero, J., & Hodges, J. R. (2006). Semantic dementia and fluent primary progressive aphasia: two sides of the same coin? Brain, 129(11), 3066–3080. https://doi.org/10.1093/brain/awl285

Ahmed, S., de Jager, C. A., Haigh, A. M., & Garrard, P. (2013). Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer's disease. *Neuropsychology*, *27*(1), 79–85. https://doi.org/10.1037/a0031288

Albert, M. L. (1989). The role of perseveration in language disorders. *Journal of Neurolinguistics*, *4*(3–4), 471–478. https://doi.org/10.1016/0911-6044(89)90035-3

Alkenani, A. H., Li, Y., Xu, Y., & Zhang, Q. (2021). Predicting alzheimer's disease from spoken and written language using Fusion-Based stacked generalization. *Journal of Biomedical Informatics*, *118*, 103803. https://doi.org/10.1016/j.jbi.2021.103803

Allibrio, G., Gori, M. A., Signorini, G., & Luzzatti, C. (2009). Un esame del linguaggio per la diagnosi dei deficit afasici al letto del malato. *Giornale Di Psicologia*, *3*(1), 7–22.

Altmann, L. J. P., Kempler, D., & Andersen, E. S. (2001). Speech errors in alzheimer's disease: Reevaluating morphosyntactic perseveration. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1069–1082. https://doi.org/10.1044/1092-4388(2001/085)

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2020). A unified model of post-stroke language deficits including discourse production and their neural correlates. *Brain*, *143*(5), 1541–1554. https://doi.org/10.1093/brain/awaa074

American Speech-Language-Hearing Association. (2021). *Aphasia*. https://www.asha.org/practice-portal/clinical-topics/aphasia/

Antonucci, S. M. (2009). Use of semantic feature analysis in group aphasia treatment. *Aphasiology*, *23*(7–8), 854–866. https://doi.org/10.1080/02687030802634405

Ardila, A., & Rosselli, M. (1993). Language deviations in aphasia: A frequency analysis. *Brain and Language*, *44*(2), 165–180. https://doi.org/10.1006/brln.1993.1011

Arning, L., & Nguyen, H. P. (2021). Huntington disease update: New insights into the role of repeat instability in disease pathogenesis. *Medizinische Genetik*, *33*(4), 293–300. https://doi.org/10.1515/medgen-2021-2101

Ash, S., Evans, E., O'Shea, J., Powers, J., Boller, A., Weinberg, D., Haley, J., McMillan, C., Irwin, D. J., Rascovsky, K., & Grossman, M. (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology*, *81*(4), 329–336. https://doi.org/10.1212/wnl.0b013e31829c5d0e

Ash, S., McMillan, C., Gross, R. G., Cook, P., Gunawardena, D., Morgan, B., Boller, A., Siderowf, A., & Grossman, M. (2012). Impairments of speech fluency in Lewy body spectrum disorder. *Brain and Language*, *120*(3), 290–302. https://doi.org/10.1016/j.bandl.2011.09.004

Ash, S., Menaged, A., Olm, C., McMillan, C. T., Boller, A., Irwin, D. J., McCluskey, L., Elman, L., & Grossman, M. (2014). Narrative discourse deficits in amyotrophic lateral sclerosis. *Neurology*, *83*(6), 520–528. https://doi.org/10.1212/wnl.0000000000000670

Ash, S., Moore, P., Antani, S., McCawley, G., Work, M., & Grossman, M. (2006). Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology*, *66*(9), 1405–1413. https://doi.org/10.1212/01.wnl.0000210435.72614.38

Badler, N. I. (1989). Artificial intelligence, natural language, and simulation for human animation. In *State-of-the-art in computer animation* (pp. 19–31). Springer.

Bak, T. H., & Abrahams, S. (2016). The FTD-ALS spectrum. *Cambridge University Press eBooks*, 68–81. https://doi.org/10.1017/cbo9781316091586.009

Bak, T. H., & Hodges, J. R. (2003). Kissing and dancing—a test to distinguish the lexical and conceptual contributions to noun/verb and action/object dissociation. Preliminary results in patients with frontotemporal dementia. *Journal of Neurolinguistics*, *16*(2–3), 169–181. https://doi.org/10.1016/s0911-6044(02)00011-8

Bak, T. H., & Hodges, J. R. (2004). The effects of motor neurone disease on language: Further evidence. *Brain and Language*, *89*(2), 354–361. https://doi.org/10.1016/s0093-934x(03)00357-2

Barry, D., Bates, M. E., & Labouvie, E. (2008). FAS and CFL forms of verbal fluency differ in difficulty: A meta-analytic study. *Applied Neuropsychology*, *15*(2), 97–106. https://doi.org/10.1080/09084280802083863

Bayles, K. A. (1987). *Communication and cognition in normal aging and dementia*. Pro Ed.

Bayles, K. A. (1990). Language and Parkinson Disease. *Alzheimer Disease & Associated Disorders*, *4*(3), 171–180. https://doi.org/10.1097/00002093-199040300-00005

Ben-David, B. M., Mentzel, M., Icht, M., Gilad, M., Dor, Y. I., Ben-David, S., Carl, M., & Shakuf, V. (2020). Challenges and opportunities for telehealth assessment during COVID-19: iT-RES, adapting a remote version of the test for rating emotions in speech. *International Journal of Audiology*, *60*(5), 319–321. https://doi.org/10.1080/14992027.2020.1833255

Berube, S., Nonnemacher, J., Demsky, C., Glenn, S., Saxena, S., Wright, A., Tippett, D. C., & Hillis, A. E. (2019). Stealing cookies in the Twenty-First century: Measures of spoken narrative in healthy versus speakers with aphasia. *American Journal of Speech-Language Pathology*, *28*(1S), 321–329. https://doi.org/10.1044/2018_ajslp-17-0131

Biss, K. (2021). *Fees | Kate Biss - Speech and Language Therapist*. Kate Biss Speech and Language Therapy. https://katebiss-speechtherapist.co.uk/fees

Blankenship, J. (1962). A linguistic analysis of oral and written style. *Quarterly Journal of Speech*, *48*(4), 419–422. https://doi.org/10.1080/00335636209382571

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer*. Praat: Doing Phonetics by Computer. http://www.praat.org/

Boets, B., op de Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., Bulthé, J., Sunaert, S., Wouters, J., & Ghesquière, P. (2013). Intact but less accessible phonetic representations in adults with dyslexia. *Science*, *342*(6163), 1251–1254. https://doi.org/10.1126/science.1244333

Borchers, G. (1936). An approach to the problem of oral style. *Quarterly Journal of Speech*, *22*(1), 114–117. https://doi.org/10.1080/00335633609380170

Boroditsky, L., & Gaby, A. (2010). Remembrances of times east. *Psychological Science*, *21*(11), 1635–1639. https://doi.org/10.1177/0956797610386621

Boucher, J., Marcotte, K., Brisebois, A., Courson, M., Houzé, B., Desautels, A., Léonard, C., Rochon, E., & Brambati, S. M. (2020). Word-finding in confrontation naming and picture descriptions produced by individuals with early post-stroke aphasia. *The Clinical Neuropsychologist*, 1–16. https://doi.org/10.1080/13854046.2020.1817563

Boyle, M. (2004). Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology*, *13*(3), 236–249. https://doi.org/10.1044/1058-0360(2004/025)

Boyle, M. (2015). Stability of Word-Retrieval errors with the AphasiaBank stimuli. *American Journal of Speech-Language Pathology*, *24*(4). https://doi.org/10.1044/2015_ajslp-14-0152

Bramley, D. (2015, February 10). *Report: Americans spending nearly five hours a day on smartphones*. Cision PR Newswire. https://www.prnewswire.com/news-releases/report-americans-spending-nearly-five-hours-a-day-on-smartphones-300033422.html

Brown, C. M. L. (1988). Comparison of typing and handwriting in “Two-Finger typists.” *Proceedings of the Human Factors Society Annual Meeting*, *32*(5), 381–385. https://doi.org/10.1177/154193128803200533

Bschor, T., Kühl, K. P., & Reischies, F. M. (2001a). Spontaneous speech of patients with dementia of the alzheimer type and mild cognitive impairment. *International Psychogeriatrics*, *13*(3), 289–298. https://doi.org/10.1017/s1041610201007682

Bschor, T., Kühl, K. P., & Reischies, F. M. (2001b). Spontaneous Speech of Patients With Dementia of the Alzheimer Type and Mild Cognitive Impairment. *International Psychogeriatrics*, *13*(3), 289–298. https://doi.org/10.1017/s1041610201007682

Bub, D., & Kertesz, A. (1982). Evidence for lexicographic processing in a patient with preserved written OVver oral single word naming. *Brain*, *105*(4), 697–717. https://doi.org/10.1093/brain/105.4.697

Burns, A., & Iliffe, S. (2009). Alzheimer’s disease. *BMJ*, *338*(feb05 1), b158. https://doi.org/10.1136/bmj.b158

Cahana-Amitay, D., Albert, M. L., Pyun, S. B., Westwood, A., Jenkins, T., Wolford, S., & Finley, M. (2011). Language as a stressor in aphasia. *Aphasiology*, *25*(5), 593–614. https://doi.org/10.1080/02687038.2010.541469

Cambridge Cognition Ltd. (2022). *Graded Naming Test (GNT)*. Cambridge Cognition. https://www.cambridgecognition.com/cantab/cognitive-tests/graded-naming-test-gnt/

Cameron, R. M., Wambaugh, J. L., & Mauszycki, S. C. (2010). Individual variability on discourse measures over repeated sampling times in persons with aphasia. *Aphasiology*, *24*(6–8), 671–684. https://doi.org/10.1080/02687030903443813

Capilouto, G. J., Wright, H. H., & Wagovich, S. A. (2006). Reliability of main event measurement in the discourse of individuals with aphasia. *Aphasiology*, *20*(2–4), 205–216. https://doi.org/10.1080/02687030500473122

Cardarelli, R., Kertesz, A., & Knebl, J. A. (2010). Frontotemporal dementia and primary progressive aphasia: A review. *American Family Physician*, *82*(11), 1372–1377. https://doi.org/10.2147/ndt.s38821

Carlomango, S., Santoro, A., Menditti, A., Pandolfi, M., & Marini, A. (2005). Referential communication in alzheimer's type dementia. *Cortex*, *41*(4), 520–534. https://doi.org/10.1016/s0010-9452(08)70192-8

Caron, N. S. (2020). *Huntington disease* (G. Wright & M. R. Hayden, Eds.). Gene Reviews.

Caselli, R. J., Windebank, A. J., Petersen, R. C., Komori, T., Parisi, J. E., Okazaki, H., Kokmen, E., Iverson, R., Dinapoli, R. P., Graff-Radford, N. R., & Stein, S. (1993). Rapidly progressive aphasic dementia and motor neuron disease. *Annals of Neurology*, *33*(2), 200–207. https://doi.org/10.1002/ana.410330210

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are Specific Language Impairment and Dyslexia Distinct Disorders? *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378–1396. https://doi.org/10.1044/1092-4388(2005/096)

Chafe, W. L. (1982). Integration and Involvement in Speaking, Writing, and Oral Literature. In D. Tannen (Ed.), *Oral and Written Discourse* (pp. 1–21). Norwood.

Chee, D., Nice, R., Hamilton, B., Jones, E., Hawkins, S., Redstone, C., Cairnes, V., Pohl, K., Kennedy, N. A., Ahmad, T., Goodhand, J., & McDonald, T. J. (2020). Patient-Led remote intracapillary pharmacokinetic sampling (fingerpricks): Validation and acceptability for home therapeutic drug monitoring in patients with ibd in the COVID-19 pandemic. *SSRN Electronic Journal*, 1–23. https://doi.org/10.2139/ssrn.3729267

Chemin, A. (2014, February 14). *Handwriting vs typing: Is the pen still mightier than the keyboard?* The Guardian. https://www.theguardian.com/science/2014/dec/16/cognitive-benefits-handwriting-decline-typing

Christensen, C. A. (2004). Relationship between orthographic-motor integration and computer use for the production of creative and well-structured written text. *British Journal of Educational Psychology*, *74*(4), 551–564. https://doi.org/10.1348/0007099042376373

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences*, *102*(35), 12629–12633. https://doi.org/10.1073/pnas.0506162102

Ciurli, P., & Basso, A. (1996). *Esame del Linguaggio-II*. Organizzazioni Speciali.

Cleland, A., & Pickering, M. (2006). Do writing and speaking employ the same syntactic representations? *Journal of Memory and Language*, *54*(2), 185–198. https://doi.org/10.1016/j.jml.2005.10.003

Cole, K. N., Dale, P. S., & Mills, P. E. (1992). Stability of the intelligence quotient-language quotient relation: Is discrepancy modeling based on a myth? *American Journal of Mental Retardation, 97*(2), 131–143.

Connelly, V., Gee, D., & Walsh, E. (2007). A comparison of keyboarded and handwritten compositions and the relationship with transcription speed. *British Journal of Educational Psychology*, *77*(2), 479–492. https://doi.org/10.1348/000709906x116768

Conroy, P., Sage, K., & Ralph, M. L. (2009). Improved vocabulary production after naming therapy in aphasia: can gains in picture naming generalise to connected speech? *International Journal of Language & Communication Disorders*, *44*(6), 1036–1062. https://doi.org/10.1080/13682820802585975

Constantinides, V. C., Paraskevas, G. P., Paraskevas, P. G., Stefanis, L., & Kapaki, E. (2019). Corticobasal degeneration and corticobasal syndrome: A review. *Clinical Parkinsonism & Related Disorders*, *1*, 66–71. https://doi.org/10.1016/j.prdoa.2019.08.005

Corey, D. M., & Ben-Porath, Y. S. (2020). Practical guidance on the use of the MMPI instruments in remote psychological testing. *Professional Psychology: Research and Practice*, *51*(3), 199–204. https://doi.org/10.1037/pro0000329

costhelper health. (2021). *How much does speech therapy cost?* Speech Therapy Cost. https://health.costhelper.com/speech-therapist.html

Creed, P. A., Wiener, K. K., & Creed, P. (1999). Use of the shipley institute of living scale and the raven’s standard progressive matrices with unemployed populations. *Journal of Applied Health Behaviou*, *1*(1), 22–26.

Croisile, B., Ska, B., Brabant, M. J., Duchene, A., Lepage, Y., Aimard, G., & Trillet, M. (1996). Comparative study of oral and written picture description in patients with alzheimer's disease. *Brain and Language*, *53*(1), 1–19. https://doi.org/10.1006/brln.1996.0033

Cuetos, F., Arango-Lasprilla, J. C., Uribe, C., Valencia, C., & Lopera, F. (2007). Linguistic changes in verbal expression: A preclinical marker of Alzheimer's disease. *Journal of the International Neuropsychological Society*, *13*(03). https://doi.org/10.1017/s1355617707070609

Dalton, S. G. H., & 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*(1S), 293–320. https://doi.org/10.1044/2018_ajslp-17-0166

David, R. M., & Skilbeck, C. E. (1984). Raven IQ and language recovery following stroke. *Journal of Clinical Neuropsychology*, *6*(3), 302–308. https://doi.org/10.1080/01688638408401220

de Lira, J. O., Ortiz, K. Z., Campanha, A. C., Bertolucci, P. H. F., & Minett, T. S. C. (2010). Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *International Psychogeriatrics*, *23*(3), 404–412. https://doi.org/10.1017/s1041610210001092

Delis, D. C., Kaplan, E., & Kaplan, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. Pearson.

DeVito, J. A. (1966a). Psychogrammatical factors in oral and written discourse by skilled communicators. *Speech Monographs*, *33*(1), 73–76. https://doi.org/10.1080/03637756609375483

DeVito, J. A. (1966b). The encoding of speech and writing. *The Speech Teacher*, *15*(1), 55–60. https://doi.org/10.1080/03634526609377493

DeVito, J. A. (1967). Levels of abstraction in spoken and written language. *Journal of Communication*, *17*(4), 354–361. https://doi.org/10.1111/j.1460-2466.1967.tb01193.x

Dewey, G. (1923). *Relativ frequency of english speech sounds*. Harvard University Press.

Dignam, J., Copland, D. A., O'Brien, K., Burfein, P., Khan, A., & Rodriguez, A. D. (2017). Influence of Cognitive Ability on Therapy Outcomes for Anomia in Adults With Chronic Poststroke Aphasia. *Journal of Speech Language and Hearing Research*, *60*(2), 406–421. https://doi.org/10.1044/2016_jslhr-l-15-0384

Dijkstra, K., Bourgeois, M. S., Allen, R. S., & Burgio, L. D. (2004). Conversational coherence: Discourse analysis of older adults with and without dementia. *Journal of Neurolinguistics*, *17*(4), 263–283. https://doi.org/10.1016/s0911-6044(03)00048-4

Drieman, G. (1962). Differences between written and spoken language. *Acta Psychologica*, *20*, 78–100. https://doi.org/10.1016/0001-6918(62)90009-4

Drijbooms, E., Groen, M. A., & Verhoeven, L. (2016). Children's use of evaluative devices in spoken and written narratives. *Journal of Child Language*, *44*(4), 767–794. https://doi.org/10.1017/s0305000916000234

Drummond, C., Coutinho, G., Fonseca, R. P., Assunção, N., Teldeschi, A., de Oliveira-Souza, R., Moll, J., Tovar-Moll, F., & Mattos, P. (2015). Deficits in narrative discourse elicited by visual stimuli are already present in patients with mild cognitive impairment. *Frontiers in Aging Neuroscience*, *97*(7). https://doi.org/10.3389/fnagi.2015.00096

Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N. C., Galasko, D., Gauthier, S., Hampel, H., Jicha, G. A., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., . . . Scheltens, P. (2010). Revising the definition of alzheimer's disease: A new lexicon. *The Lancet Neurology*, *9*(11), 1118–1127. https://doi.org/10.1016/s1474-4422(10)70223-4

Duffy, J. R., Werven, G. W., & Aronson, A. E. (1997). Telemedicine and the diagnosis of speech and language disorders. *Mayo Clinic Proceedings*, *72*(12), 1116–1122. https://doi.org/10.4065/72.12.1116

Eknoyan, G. (2007). Adolphe Quetelet (1796 1874) the average man and indices of obesity. *Nephrology Dialysis Transplantation*, *23*(1), 47–51. https://doi.org/10.1093/ndt/gfm517

elemy. (2021, June 8). *How much does speech therapy cost in 2022? (averages)*. The Elemy Learning Studio. https://www.elemy.com/studio/speech-therapy/costs

El-Tholotha, M., Bau, H. H., & Song, J. (2020). A single and Two-Stage, Closed-Tube, molecular test for the 2019 novel coronavirus (COVID-19) at home, clinic, and points of entry. *ChemRxiv : The Preprint Server for Chemistry*. https://doi.org/10.26434/chemrxiv.11860137

Emerton, B. C., Gansler, D. A., Sandberg, E. H., & Jerram, M. (2013). Functional anatomic dissociation of description and picture naming in the left temporal lobe. *Brain Imaging and Behavior*, *8*(4), 570–578. https://doi.org/10.1007/s11682-013-9281-9

Enderby, P. M., Wood, V. A., Wade, D. T., & Hewer, R. L. (1986). The frenchay aphasia screening test: A short, simple test for aphasia appropriate for non-specialists. *International Rehabilitation Medicine*, *8*(4), 166–170. https://doi.org/10.3109/03790798709166209

Fergadiotis, G., & Wright, H. H. (2011). Lexical diversity for adults with and without aphasia across discourse elicitation tasks. *Aphasiology*, *25*(11), 1414–1430. https://doi.org/10.1080/02687038.2011.603898

Fergadiotis, G., & Wright, H. H. (2015). Modelling confrontation naming and discourse performance in aphasia. *Aphasiology*, *30*(4), 364–380. https://doi.org/10.1080/02687038.2015.1067288

Fergadiotis, G., Wright, H. H., & West, T. M. (2013). Measuring lexical diversity in narrative discourse of people with aphasia. *American Journal of Speech-Language Pathology*, *22*(2). https://doi.org/10.1044/1058-0360(2013/12-0083)

Feyereisen, P., Berrewaerts, J., & Hupet, M. (2007). Pragmatic skills in the early stages of alzheimer's disease: An analysis by means of a referential communication task. *International Journal of Language & Communication Disorders*, *42*(1), 1–17. https://doi.org/10.1080/13682820600624216

FindingFive Team. (2019). *FindingFive: A web platform for creating, running, and managing your studies in one place*. FindingFive Corporation (nonprofit).

Fine, E. M., Delis, D. C., Dean, D., Beckman, V., Miller, B. L., Rosen, H. J., & Kramer, J. H. (2009). Left frontal lobe contributions to concept formation: A quantitative MRI study of performance on the Delis–Kaplan Executive Function System Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *31*(5), 624–631. https://doi.org/10.1080/13803390802419017

Finger, E. C. (2016). Frontotemporal dementias. *Continuum: Lifelong Learning in Neurology*, *22*(2), 464–489. https://doi.org/10.1212/con.0000000000000300

Forbes-McKay, K. E., & Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer’s disease with a picture description task. *Neurological Sciences*, *26*(4), 243–254. https://doi.org/10.1007/s10072-005-0467-9

Forbes-McKay, K., Shanks, M. F., & Venneri, A. (2013). Profiling spontaneous speech decline in alzheimer’s disease: A longitudinal study. *Acta Neuropsychiatrica*, *25*(6), 320–327. https://doi.org/10.1017/neu.2013.16

Fossum, E. C. (1944). An analysis of the dynamic vocabulary of junior college students∗. *Speech Monographs*, *11*(1), 88–96. https://doi.org/10.1080/03637754409390092

Francis, W. N. (1958). *The structure of american english*. The Ronald Press.

Frank, S. (2013). Treatment of Huntington’s Disease. *Neurotherapeutics*, *11*(1), 153–160. https://doi.org/10.1007/s13311-013-0244-z

Fraser, K. C., Meltzer, J. A., Graham, N. L., Leonard, C., Hirst, G., Black, S. E., & Rochon, E. (2014). Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. *Cortex*, *55*, 43–60. https://doi.org/10.1016/j.cortex.2012.12.006

Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2015). Linguistic features identify alzheimer’s disease in narrative speech. *Journal of Alzheimer’s Disease*, *49*(2), 407–422. https://doi.org/10.3233/jad-150520

Frau, F., Hilviu, D., Prola, A., Gabbatore, I., Bosco, F. M., Marini, A., & Pavia, I. U. (2021). *Stability and decline of narrative abilities in healthy ageing*. Researchgate.Net. https://www.researchgate.net/profile/Federico-Frau/publication/353130288_Stability_and_decline_of_narrative_abilities_in_healthy_ageing/links/60e8826c0fbf460db8f5c0ec/Stability-and-decline-of-narrative-abilities-in-healthy-ageing.pdf

Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing*, *30*, 45–57. https://doi.org/10.1016/j.jslw.2015.08.009

Furth, H. G. (1961). The influence of language on the development of concept formation in deaf children. *The Journal of Abnormal and Social Psychology*, *63*(2), 386–389. https://doi.org/10.1037/h0049189

Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social cognition, language acquisition and the development of the theory of mind. *Mind and Language*, *16*(5), 494–541. https://doi.org/10.1111/1468-0017.00180

Garrard, P., & Forsyth, R. (2010). Abnormal discourse in semantic dementia: A data-driven approach. *Neurocase*, *16*(6), 520–528. https://doi.org/10.1080/13554791003785901

Garrard, P., Haigh, A. M., & de Jager, C. (2011). Techniques for transcribers: assessing and improving consistency in transcripts of spoken language. *Literary and Linguistic Computing*, *26*(4), 389–405. https://doi.org/10.1093/llc/fqr018

Genesee, F. (1976). The role of intelligence in second language learning. *Language Learning*, *26*(2), 267–280. https://doi.org/10.1111/j.1467-1770.1976.tb00277.x

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain*, *88*(3), 585. https://doi.org/10.1093/brain/88.3.585

Gibson, J. W., Gruner, C. R., Kibler, R. J., & Kelly, F. J. (1966). A quantitative examination of differences and similarities in written and spoken messages. *Speech Monographs*, *33*(4), 444–451. https://doi.org/10.1080/03637756609375510

Giles, E., Patterson, K., & Hodges, J. R. (1996). Performance on the Boston Cookie theft picture description task in patients with early dementia of the Alzheimer's type: Missing information. *Aphasiology*, *10*(4), 395–408. https://doi.org/10.1080/02687039608248419

Goldstein, L. H., & Abrahams, S. (2013). Changes in cognition and behaviour in amyotrophic lateral sclerosis: nature of impairment and implications for assessment. *Lancet Neurology*, *12*(4), 368–380. https://doi.org/10.1016/s1474-4422(13)70026-7

Goodglass, H., Baressi, B., & Kaplan, E. (1983). *The Boston Diagnostic Aphasia Examination*. Lippincoot and Wilkins.

Goodglass, H., & Hunter, M. (1970). A linguistic comparison of speech and writing in two types of aphasia. *Journal of Communication Disorders*, *3*(1), 28–35. https://doi.org/10.1016/0021-9924(70)90029-8

Goodglass, H., Kaplan, E., & Baressi, B. (2000). *Boston Diagnostic Aphasia Examination* (Third Edition). Pearson.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, *5*(1), 173–184. https://doi.org/10.1111/tops.12007

Goody, J. (1977). *The domestication of the savage mind*. Cambridge University Press.

Goody, J. (1980). Thought and writing. In E. Gellner (Ed.), *Soviet and Western Anthropology* (pp. 119–113). Columbia University Press.

Goody, J., & Watt, I. (1963). The Consequences of Literacy. *Comparative Studies in Society and History*, *5*(3), 304–345. https://doi.org/10.1017/s0010417500001730

Gorno-Tempini, M. L., Brambati, S. M., Ginex, V., Ogar, J., Dronkers, N. F., Marcone, A., Perani, D., Garibotto, V., Cappa, S. F., & Miller, B. L. (2008). The logopenic/phonological variant of primary progressive aphasia. *Neurology*, *71*(16), 1227–1234. https://doi.org/10.1212/01.wnl.0000320506.79811.da

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., Ogar, J. M., Rohrer, J. D., Black, S., Boeve, B. F., Manes, F., Dronkers, N. F., Vandenberghe, R., Rascovsky, K., Patterson, K., Miller, B. L., Knopman, D. S., Hodges, J. R., Mesulam, M. M., & Grossman, M. (2011). Classification of primary progressive aphasia and its variants. Neurology, 76(11), 1006–1014. https://doi.org/10.1212/wnl.0b013e31821103e6

Grad, L. I., Rouleau, G. A., Ravits, J., & Cashman, N. R. (2016). Clinical spectrum of amyotrophic lateral sclerosis (ALS). *Cold Spring Harbor Perspectives in Medicine*, *7*(8), 1–15. https://doi.org/10.1101/cshperspect.a024117

Graham, N. L., Patterson, K., & Hodges, J. R. (2004). When more yields less: Speaking and writing deficits in nonfluent progressive aphasia. *Neurocase*, *10*(2), 141–155. https://doi.org/10.1080/13554790409609945

Gray, D. (2020, November 10). *Remote ECAS now available – ECAS*. Remote ECAS. https://ecas.psy.ed.ac.uk/remote-ecas-now-available/

Greenfield, P. M. (1972). Oral or written language: The consequences for cognitive development in africa, the united states and england. *Language and Speech*, *15*(2), 169–178. https://doi.org/10.1177/002383097201500207

Gross, R. G., Ash, S., McMillan, C. T., Gunawardena, D., Powers, C., Libon, D. J., Moore, P., Liang, T. W., & Grossman, M. (2010). Impaired information integration contributes to communication difficulty in corticobasal syndrome. *Cognitive and Behavioral Neurology*, *23*(1), 1–7. https://doi.org/10.1097/wnn.0b013e3181c5e2f8

Grossman, M. (2010). Primary progressive aphasia: clinicopathological correlations. *Nature Reviews Neurology*, *6*(2), 88–97. https://doi.org/10.1038/nrneurol.2009.216

Grossman, M., Anderson, C. S., Khan, A., Avants, B. B., Elman, L., & McCluskey, L. (2008). Impaired action knowledge in amyotrophic lateral sclerosis. *Neurology*, *71*(18), 1396–1401. https://doi.org/10.1212/01.wnl.0000319701.50168.8c

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. https://doi.org/10.1016/j.jcomdis.2003.08.004

Gruner, C. R., Kibler, R. J., & Gibson, J. W. (1967). A quantitative analysis of selected characteristics of oral and written vocabularies. *Journal of Communication*, *17*(2), 152–158. https://doi.org/10.1111/j.1460-2466.1967.tb01172.x

Guiora, A. Z. (1983). Language and Concept Formation: A Cross- Lingual Analysis. *Behavior Science Research*, *18*(3), 228–256. https://doi.org/10.1177/106939718301800304

Gumperz, J. J., Kaltman, H., & O'Connor, C. (1982). The transition to literacy. In D. Tannen (Ed.), *Oral and written discourse* (p. N/A). Pranger.

Gunawardena, D., Ash, S., McMillan, C., Avants, B., Gee, J., & Grossman, M. (2010). Why are patients with progressive nonfluent aphasia nonfluent? *Neurology*, *75*(7), 588–594. https://doi.org/10.1212/wnl.0b013e3181ed9c7d

Gustafsson, P. (2016). *Namig a world of things - an investigation of the development of noun-phrases in written expository discourse, from adolescents to expert writers* (Master Thesis). https://www.lunduniversity.lu.se/lup/publication/8895490

Hanson, V. L. (1981). Processing of written and spoken words: Evidence for common coding. *Memory & Cognition*, *9*(1), 93–100. https://doi.org/10.3758/bf03196954

Harrell, L. E. (1957). A comparison of the development of oral and written language in School-Age children. *Monographs of the Society for Research in Child Development*, *22*(3), 1. https://doi.org/10.2307/1165494

Head, H. (1926). *Aphasia and kindred disorders*. Cambridge University Press.

Henry, M. L., Beeson, P. M., Alexander, G. E., & Rapcsak, S. Z. (2012). Written Language Impairments in Primary Progressive Aphasia: A Reflection of Damage to Central Semantic and Phonological Processes. *Journal of Cognitive Neuroscience*, *24*(2), 261–275. https://doi.org/10.1162/jocn_a_00153

Heydasch, T., Haubrich, J., & Renner, K.-H. (2020). The short version of the hagen matrices test (HMT-S). *Methods, Data, Analyses*, *7*, 183–205. https://doi.org/10.12758/mda.2013.011

Hier, D., & Mohr, J. P. (1977). Incongruous oral and written naming: Evidence for a subdivision of the syndrome of wernicke's aphasia. *Brain and Language*, *4*(1), 115–126. https://doi.org/10.1016/0093-934x(77)90010-4

Hodges, J. R., & Patterson, K. (2007). Semantic dementia: a unique clinicopathological syndrome. *The Lancet Neurology*, *6*(11), 1004–1014. https://doi.org/10.1016/s1474-4422(07)70266-1

Hoffmann, I., Nemeth, D., Dye, C. D., Pákáski, M., Irinyi, T., & Kálmán, J. (2009a). Temporal parameters of spontaneous speech in Alzheimer's disease. *International Journal of Speech-Language Pathology*, *12*(1), 29–34. https://doi.org/10.3109/17549500903137256

Hoffmann, I., Nemeth, D., Dye, C. D., Pákáski, M., Irinyi, T., & Kálmán, J. (2009b). Temporal parameters of spontaneous speech in alzheimer's disease. *International Journal of Speech-Language Pathology*, *12*(1), 29–34. https://doi.org/10.3109/17549500903137256

Horn, E. (1926). *A basic writing vocabulary: 10,000 words most commonly used in writing* (4th ed.). University of Iowa Monographs in Education.

Horowitz, M. W., & Newman, J. B. (1964). Spoken and written expression: An experimental analysis. *The Journal of Abnormal and Social Psychology*, *68*(6), 640–647. https://doi.org/10.1037/h0048589

Howard, D., Swinburn, K., & Porter, G. (2010). Putting the CAT out: What the Comprehensive Aphasia Test has to offer. *Aphasiology*, *24*(1), 56–74. https://doi.org/10.1080/02687030802453202

Huber, W., Poeck, K., & Springer, L. (2013). *Klinik und Rehabilitation der Aphasie*. Georg Thieme Verlag.

Huber, W., Poeck, K., & Weniger, D. (1983). *Aachener Aphasie-Test*. Hogrefe Verlag.

Huddleston, R. D. (1971). *The sentence in written english: A syntactic study based on an analysis of scientific texts (cambridge studies in linguistics, series number 3)* (1st ed.). Cambridge University Press.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*(1), 53–69. https://doi.org/10.1037/0033-2909.104.1.53

Illarioshkin, S. N., Klyushnikov, S. A., Vigont, V. A., Seliverstov, Y. A., & Kaznacheyeva, E. V. (2018). Molecular pathogenesis in huntington's disease. *Biochemistry (Moscow)*, *83*(9), 1030–1039. https://doi.org/10.1134/s0006297918090043

Imada, T., Carlson, S. M., & Itakura, S. (2012). East-West cultural differences in context-sensitivity are evident in early childhood. *Developmental Science*, *16*(2), 198–208. https://doi.org/10.1111/desc.12016

jamovi (1.6). (2021). [Computer Software]. The jamovi project. https://www.jamovi.org

Jarrold, W., Meitner, B., Vergryi, D., Richey, C., Gorno-Tempini, M. L., & Ogar, J. (2014). Aided diagnosis of dementia type through Computer-Based analysis of spontaneous speech. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality* (pp. 27–37). Association for Computational Linguistics.

Jarvis, S. (2013). Capturing the diversity in lexical diversity. *Language Learning*, *63*, 87–106. https://doi.org/10.1111/j.1467-9922.2012.00739.x

JASP Team. (2022). *JASP* (0.16.3) [Computer software]. https://jasp-stats.org/

J.C.R. (1938). *Raven's Progressive Matrices (1938): sets A, B, C, D, E*. Australian Council for Educational Research.

Jensen, A. M., Chenery, H. J., & Copland, D. A. (2006). A comparison of picture description abilities in individuals with vascular subcortical lesions and Huntington's Disease. *Journal of Communication Disorders*, *39*(1), 62–77. https://doi.org/10.1016/j.jcomdis.2005.07.001

Ji, L. J., Zhang, Z., & Nisbett, R. E. (2004). Is it culture or is it language? Examination of language effects in Cross-Cultural research on Categorization. *Journal of Personality and Social Psychology*, *87*(1), 57–65. https://doi.org/10.1037/0022-3514.87.1.57

Johnson, M. H., Grossmann, T., & Farroni, T. (2008). The social cognitive neuroscience of infancy: Illuminating the early development of social brain functions. *Advances in Child Development and Behavior*, *36*, 331–372. https://doi.org/10.1016/s0065-2407(08)00008-6

Justice, N. J. (2018). The relationship between stress and Alzheimer's disease. *Neurobiology of Stress*, *8*, 127–133. https://doi.org/10.1016/j.ynstr.2018.04.002

Kalia, L. V., & Lang, A. E. (2015). Parkinson's disease. *The Lancet*, *386*(9996), 896–912. https://doi.org/10.1016/s0140-6736(14)61393-3

Kalman, Y., Kavé, G., & Umanski, D. (2015). Writing in a digital world: Self-Correction while typing in younger and older adults. *International Journal of Environmental Research and Public Health*, *12*(10), 12723–12734. https://doi.org/10.3390/ijerph121012723

Kambanaros, M. (2009). Group effects of instrumentality and name relation on action naming in bilingual anomic aphasia. *Brain and Language*, *110*(1), 29–37. https://doi.org/10.1016/j.bandl.2009.01.004

Kaur, H., Bajpai, S., Pershad, D., Sreenivas, V., & Nehra, A. (2017). Development and standardization of indian aphasia battery. *Journal of Mental Health and Human Behaviour*, *22*(2), 116–122. https://doi.org/10.4103/jmhhb.jmhhb_45_16

Kavé, G., & Levy, Y. (2003). Morphology in picture descriptions provided by persons with alzheimer's disease. *Journal of Speech, Language, and Hearing Research*, *46*(2), 341–352. https://doi.org/10.1044/1092-4388(2003/027)

Kekatos, M. (2017, 31. Januar). *What your handwriting says about YOUR health*. Mail Online. https://www.dailymail.co.uk/health/article-4176286/What-handwriting-says-health.html

Kempler, D., Curtiss, S., & Jackson, C. (1987). Syntactic preservation in alzheimer's disease. *Journal of Speech, Language, and Hearing Research*, *30*(3), 343–350. https://doi.org/10.1044/jshr.3003.343

Kertesz, A. (1982). *The Western Aphasia Battery* (W. Knitsch, Ed.). Grune and Stratton.

Kertesz, A. (2006). *Western Aphasia Battery Revised (WAB-R)*. Pearson.

Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., Burrell, J. R., & Zoing, M. C. (2011). Amyotrophic lateral sclerosis. *The Lancet*, *377*(9769), 942–955. https://doi.org/10.1016/s0140-6736(10)61156-7

Koedam, E. L., Lauffer, V., van der Vlies, A. E., van der Flier, W. M., Scheltens, P., & Pijnenburg, Y. A. (2010). Early-Versus Late-Onset alzheimer's disease: More than age alone. *Journal of Alzheimer's Disease*, *19*(4), 1401–1408. https://doi.org/10.3233/jad-2010-1337

Kosowicz, M., & MacPherson, S. E. (2016). Improving multitasking assessment in healthy older adults using a prop-based version of the Breakfast task. *Applied Neuropsychology: Adult*, *24*(3), 252–263. https://doi.org/10.1080/23279095.2015.1136310

Krampe, R. T. (2002). Aging, expertise and fine motor movement. *Neuroscience & Biobehavioral Reviews*, *26*(7), 769–776. https://doi.org/10.1016/s0149-7634(02)00064-7

Kroll, B. (1977). Ways CommunicatorsEncode propositions in spoken and written english. In T. Bennett-Kastor & E. Ochs (Eds.), *Discourse Across Time and Space* (pp. 69–108). Department of Linguistics, University of Southern California.

Kuwabara, M., & Smith, L. B. (2016). Cultural differences in visual object recognition in 3-year-old children. *Journal of Experimental Child Psychology*, *147*, 22–38. https://doi.org/10.1016/j.jecp.2016.02.006

Lai, Y. H. (2014). Discourse features of Chinese-Speaking seniors with and without alzheimer's disease. *Language and Linguistics*, *15*(3), 411–434. https://doi.org/10.1177/1606822x14520665

Lamplot, J. D., Pinnamaneni, S., Swensen-Buza, S., Lawton, C. D., Dines, J. S., Nawabi, D. H., Young, W. K., Rodeo, S. A., & Taylor, S. A. (2021). The knee examination for video telemedicine encounters. *HSS Journal®: The Musculoskeletal Journal of Hospital for Special Surgery*, *17*(1), 80–84. https://doi.org/10.1177/1556331620975039

Laws, K. R., Duncan, A., & Gale, T. M. (2010). 'Normal' semantic–phonemic fluency discrepancy in alzheimer's disease? A meta-analytic study. *Cortex*, *46*(5), 595–601. https://doi.org/10.1016/j.cortex.2009.04.009

Leaper, C., & Smith, T. E. (2004). A Meta-Analytic review of gender variations in children's language use: Talkativeness, affiliative speech, and assertive Speech. *Developmental Psychology*, *40*(6), 993–1027. https://doi.org/10.1037/0012-1649.40.6.993

Lenhart, A. (2010, August 27). *Cell phones and american adults*. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2010/09/02/cell-phones-and-american-adults/

Levin, H., Long, S., & Schaffer, C. A. (1981). The formality of the latinate lexicon in english. *Language and Speech*, *24*(2), 161–171. https://doi.org/10.1177/002383098102400204

Llach, M. (2011). *Lexical errors and accuracy in foreign language writing*. Multilingual Matters.

Longcamp, M., Boucard, C., Gilhodes, J. C., & Velay, J. L. (2006). Remembering the orientation of newly learned characters depends on the associated writing knowledge: A comparison between handwriting and typing. *Human Movement Science*, *25*(4–5), 646–656. https://doi.org/10.1016/j.humov.2006.07.007

Longcamp, M., Zerbato-Poudou, M. T., & Velay, J. L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychologica*, *119*(1), 67–79. https://doi.org/10.1016/j.actpsy.2004.10.019

Lyu, B., Lai, C., Lin, C. H., & Gong, Y. (2021). Comparison studies of typing and handwriting in Chinese language learning: A synthetic review. *International Journal of Educational Research*, *106*, 101740. https://doi.org/10.1016/j.ijer.2021.101740

Maat-Kievit, A., Losekoot, M., Zwinderman, K., Vlis, M. V. V. D., Belfroid, R., Lopez, F., van Ommen, G. J., Breuning, M., & Roos, R. (2002). Predictability of age at onset in huntington disease in the dutch population. *Medicine*, *81*(4), 251–259. https://doi.org/10.1097/00005792-200207000-00001

MacKay, D. G., Abrams, L., & Pedroza, M. J. (1999). Aging on the input versus output side: Theoretical implications of age-linked asymmetries between detecting versus retrieving orthographic information. *Psychology and Aging*, *14*(1), 3–17. https://doi.org/10.1037/0882-7974.14.1.3

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology*, *25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85–104. https://doi.org/10.1191/0265532202lt221oa

March, E. G., Wales, R., & Pattison, P. (2006). The uses of nouns and deixis in discourse production in Alzheimer's disease. *Journal of Neurolinguistics*, *19*(4), 311–340. https://doi.org/10.1016/j.jneuroling.2006.01.001

Marie, P. (1906). La troisieme circonvolution frontale gauche ne joue aucun role special dans la fonction de langage. *La Semaine Médicale*, *26*, 241–247.

Marton, K., Abramoff, B., & Rosenzweig, S. (2005). Social cognition and language in children with specific language impairment (SLI). *Journal of Communication Disorders*, *38*(2), 143–162. https://doi.org/10.1016/j.jcomdis.2004.06.003

Massingham, L. (2021, May 31). *Assessment and therapy costs*. Independent Speech and Language Therapy Services Ltd. https://www.islts.co.uk/pricing/

McGowan Transcriptions UK. (2021). *The limitations of machine learning for transcription*. McGowan Transcriptions. https://www.mcgowantranscriptions.co.uk/machine-learning-ai-transcription/

McKenna, P., & Warrington, E. K. (1983). *Graded naming test*. NFER-Nelson.

Meteyard, L., & Patterson, K. (2009). The relation between content and structure in language production: An analysis of speech errors in semantic dementia. *Brain and Language*, *110*(3), 121–134. https://doi.org/10.1016/j.bandl.2009.03.007

Michel, M. C. (2011). Effects of task complexity and interaction on L2 performance. In P. Robinson (Ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 141–173). John Benjamins Publishing Company.

Mogey, N., Cowan, J., Paterson, J., & Purcell, M. (2012). Students' choices between typing and handwriting in examinations. *Active Learning in Higher Education, 13*(2), 117–128. https://doi.org/10.1177/1469787412441297

Mogey, N., Paterson, J., Burk, J., & Purcell, M. (2010). Typing compared with handwriting for essay examinations at university: letting the students choose. *ALT-J*, *18*(1), 29–47. https://doi.org/10.1080/09687761003657580

Mortensen, L. (2005). Written discourse and acquired brain impairment: Evaluation of structural and semantic features of personal letters from a Systemic Functional Linguistic perspective. *Clinical Linguistics & Phonetics*, *19*(3), 227–247. https://doi.org/10.1080/02699200410001698652

Mueller, K. D., Hermann, B., Mecollari, J., & Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology*, *40*(9), 917–939. https://doi.org/10.1080/13803395.2018.1446513

Mueller, P. A., & Oppenheimer, D. M. (2014). The Pen Is Mightier Than the Keyboard. *Psychological Science*, *25*(6), 1159–1168. https://doi.org/10.1177/0956797614524581

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, *44*(4), 778–792. https://doi.org/10.1044/1092-4388(2001/061)

Murray, L. L. (2010). Distinguishing clinical depression from early alzheimer’s disease in elderly people: Can narrative analysis help? *Aphasiology*, *24*(6–8), 928–939. https://doi.org/10.1080/02687030903422460

Murray, L. L., & Lenz, L. P. (2001). Productive syntax abilities in huntington’s and parkinson’s diseases. *Brain and Cognition*, *46*(1–2), 213–219. https://doi.org/10.1016/s0278-2626(01)80069-5

N. Ripich, Brian D. Carpenter, Elai, D. (2000). Conversational cohesion patterns in men and women with alzheimer’s disease: A longitudinal study. *International Journal of Language & Communication Disorders*, *35*(1), 49–64. https://doi.org/10.1080/136828200247241

National Aphasia Association. (2021, January 8). *Living with Aphasia: UK Connect*. https://www.aphasia.org/stories/living-with-aphasia-uk-connect/

National Institute of Neurological Disorder and Stroke. (2020, September 24). *Parkinson's Disease | National Institute of Neurological Disorders and Stroke*. https://www.ninds.nih.gov/health-information/disorders/parkinsons-disease

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*(3), 211–236. https://doi.org/10.1080/01638530802073712

NHS. (2018, March 10). *Dyslexia*. Nhs.Uk. https://www.nhs.uk/conditions/dyslexia/

NHS. (2021, July 1). *Motor neurone disease*. Nhs.Uk. https://www.nhs.uk/conditions/motor-neurone-disease/

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research*, *36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Nicholas, M., Obler, L. K., Albert, M. L., & Helm-Estabrooks, N. (1985). Empty speech in alzheimer's disease and fluent aphasia. *Journal of Speech, Language, and Hearing Research*, *28*(3), 405–410. https://doi.org/10.1044/jshr.2803.405

NIH National Institute on Neurological Disorders and Stroke. (2013). *Amyotrophic lateral sclerosis (ALS) fact sheet | national institute of neurological disorders and stroke*. Amytrophic Lateral Sclerosis (ALS) Fact Sheet. https://www.ninds.nih.gov/health-information/patient-caregiver-education/fact-sheets/amyotrophic-lateral-sclerosis-als-fact-sheet

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*(10), 467–473. https://doi.org/10.1016/j.tics.2005.08.004

Ochs, E. (1977). Planned and unplanned discourse. In *Discourse Accross Time and Space* (pp. 51–80). Brill.

O'Donnell, R. C. (1974). Syntactic Differences between Speech and Writing. *American Speech*, *49*(1/2), 102. https://doi.org/10.2307/3087922

O'Donnell, R. C., Griffin, W. J., & Norris, R. C. (1967). A transformational analysis of oral and written grammatical structures in the language of children in grades three, five, and seven. *The Journal of Educational Research*, *61*(1), 35–39. https://doi.org/10.1080/00220671.1967.10883574

Oller, J. W. (1981). Language as intelligence? *Language Learning*, *31*(2), 465–492. https://doi.org/10.1111/j.1467-1770.1981.tb01394.x

Olmos-Lau, N., Ginsberg, M. D., & Geller, J. B. (1977). Aphasia in multiple sclerosis. *Neurology*, *27*(7), 623. https://doi.org/10.1212/wnl.27.7.623

Olson, D. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review*, *47*(3), 257–281. https://doi.org/10.17763/haer.47.3.8840364413869005

Omani Collison, G. (1974). Concept Formation in a Second Language: A Study of Ghanaian School Children. *Harvard Educational Review*, *44*(3), 441–457. https://doi.org/10.17763/haer.44.3.d70567743hm20170

Ong, W. J. (1978). Literacy and orality in our times. *ADE Bulletin*, 1–7. https://doi.org/10.1632/ade.58.1

Onyike, C. U., & Diehl-Schmid, J. (2013). The epidemiology of frontotemporal dementia. *International Review of Psychiatry*, *25*(2), 130–137. https://doi.org/10.3109/09540261.2013.776523

Organisation for Economic Co-Operation and Development. (2021). *Educational attainment and labour-force status*. OECD. https://stats.oecd.org/Index.aspx?DataSetCode=EAG_NEAC

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe; contribution à l'étude de la perception et de la mémoire. *Archives de Psychologie*, *30*, 206–356.

Paplikar, A., Iyer, G. K., Varghese, F., Alladi, S., Pauranik, A., Mekala, S., Kaul, S., Sharma, M., Dhaliwal, R. S., Saroja, A. O., Dharamkar, S., Dutt, A., Divyaraj, G., Ghosh, A., Kandukuri, R., Mathew, R., Menon, R. N., Narayanan, J., Nehra, A., . . . Varma, R. (2020). A Screening Tool to Detect Stroke Aphasia: Adaptation of Frenchay Aphasia Screening Test (FAST) to the Indian Context. *Annals of Indian Academy of Neurology*, *23*(Suppl 2), S143–S148. https://doi.org/10.4103/aian.aian_499_20

Paquot, M. (2017). The phraseological dimension in interlanguage complexity research. *Second Language Research*, *35*(1), 121–145. https://doi.org/10.1177/0267658317694221

Parkinson's Foundation. (2020). *Understanding Parkinson's*. https://www.parkinson.org/understanding-parkinsons

Parmera, J. B., Rodriguez, R. D., Studart Neto, A., Nitrini, R., & Brucki, S. M. D. (2016). Corticobasal syndrome: A diagnostic conundrum. *Dementia & Neuropsychologia*, *10*(4), 267–275. https://doi.org/10.1590/s1980-5764-2016dn1004003

Patterson, K., Graham, N., Ralph, M. A. L., & Hodges, J. (2006). Progressive non-fluent aphasia is not a progressive form of non-fluent (post-stroke) aphasia. *Aphasiology*, *20*(9), 1018–1034. https://doi.org/10.1080/02687030600739463

Patterson, K., & MacDonald, M. C. (2016). Sweet nothings: Narrative speech in semantic dementia. In S. Andrews (Ed.), *From Inkmarks to Ideas* (pp. 299–317). Taylor & Francis.

Pearson. (2021). *Telepractice and the D-KEFS*. Pearsonassessments.Com. https://www.pearsonassessments.com/professional-assessments/digital-solutions/telepractice/telepractice-and-the-d-kefs.html

Pinnamaneni, S., Lamplot, J. D., Rodeo, S. A., Swensen-Buza, S., Lawton, C. D., Dines, J. S., Young, W. K., & Taylor, S. A. (2021). The Virtual Shoulder Physical Exam. *HSS Journal®: The Musculoskeletal Journal of Hospital for Special Surgery*, *17*(1), 59–64. https://doi.org/10.1177/1556331620975033

Poole, M. E., & Field, T. (1976). A comparison of oral and written code elaboration. *Language and Speech*, *19*(4), 305–312. https://doi.org/10.1177/002383097601900401

Portnoy, S. (1973). A comparison of oral and written Behavior. In R.S.F. & K. Salzinger (Eds.), *Studies in Verbal Behavior* (p. N/A). Amsterdam University Press.

Posner, M., & Hanson, V. L. (1980). *Analyzing spoken and written language*. Erlbaum.

Premack, D. (2004). Is language the key to human intelligence? *Science*, *303*(5656), 318–320. https://doi.org/10.1126/science.1093993

R Core Team. (2016). *R* [A Language and Environment for Statistical Computing].

Ramsay, C. B., Nicholas, M., Au, R., Obler, L. K., & Albert, M. L. (1999). Verb Naming in Normal Aging. *Applied Neuropsychology*, *6*(2), 57–67. https://doi.org/10.1207/s15324826an0602_1

Rapp, B., & Caramazza, A. (1997). The Modality-Specific organization of grammatical categories: Evidence from impaired spoken and written sentence production. *Brain and Language*, *56*(2), 248–286. https://doi.org/10.1006/brln.1997.1735

Ratnavalli, E., Brayne, C., Dawson, K., & Hodges, J. R. (2002). The prevalence of frontotemporal dementia. *Neurology*, *58*(11), 1615–1621. https://doi.org/10.1212/wnl.58.11.1615

Raven, J. C. (2003). *Raven's Progressive Matrices*. Harcourt Assessment.

Robson Tim Pring Jane Marshall Sara, J. (1998). Written communication in undifferentiated jargon aphasia: A therapy study. *International Journal of Language & Communication Disorders*, *33*(3), 305–328. https://doi.org/10.1080/136828298247767

Rocca, R., Coventry, K. R., Tylén, K., Staib, M., Lund, T. E., & Wallentin, M. (2020). Language beyond the language system: Dorsal visuospatial pathways support processing of demonstratives and spatial language during naturalistic fast fMRI. *NeuroImage*, *216*, 116128. https://doi.org/10.1016/j.neuroimage.2019.116128

Rogalski, E., Cobia, D., Harrison, T. M., Wieneke, C., Thompson, C. K., Weintraub, S., & Mesulam, M. M. (2011). Anatomy of language impairments in primary progressive aphasia. *Journal of Neuroscience*, *31*(9), 3344–3350. https://doi.org/10.1523/jneurosci.5544-10.2011

Rubin, A. (1980). A theoretical taxonomy of the differences between oral and written Language. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical Issues in Reading Comprehension: Perspectives From Cognitive Psychology, Linguistics, Artifical Intelligence, and Education* (pp. 411–437). Routleg.

Rusz, J., Cmejla, R., Ruzickova, H., & Ruzicka, E. (2011). Quantitative acoustic measurements for characterization of speech and voice disorders in early untreated Parkinson's disease. *The Journal of the Acoustical Society of America*, *129*(1), 350–367. https://doi.org/10.1121/1.3514381

Sajjadi, S. A., Patterson, K., Tomek, M., & Nestor, P. J. (2012). Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. *Aphasiology*, *26*(6), 847–866. https://doi.org/10.1080/02687038.2012.654933

Salis, C. (2010). Understanding of auditory discourse in older adults: The effects of syntax and working memory. *Aphasiology*, *25*(4), 529–539. https://doi.org/10.1080/02687038.2010.527998

Salthouse, T. A. (1996a). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295x.103.3.403

Salthouse, T. A. (1996b). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428. https://doi.org/10.1037/0033-295x.103.3.403

Savage, S. A., Piguet, O., & Hodges, J. R. (2015). "Knowing What You Don't Know": Language Insight in Semantic Dementia. *Journal of Alzheimer's Disease*, *46*(1), 187–198. https://doi.org/10.3233/jad-142703

Schnadt, M. J., Corley, M., & King, S. (2009). *Lexical Influences on Disfluency Production*. Amsterdam University Press.

Semenza, C., & Cipolotti, L. (1989). *Neuropsicologia con carta e matita*. CLEUP.

Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current Opinion in Neurobiology*, *28*, 5–9. https://doi.org/10.1016/j.conb.2014.04.003

Shafto, M. A. (2010). Orthographic error monitoring in old age: Lexical and sublexical availability during perception and production. *Psychology and Aging*, *25*(4), 991–1001. https://doi.org/10.1037/a0020117

Shimada, M., Meguro, K., Yamazaki, H., Horikawa, A., Hayasaka, C., Yamaguchi, S., Yamaguchi, K., Katsuyama, N., Nakano, M., & Yamadori, A. (1998). Impaired verbal description ability assessed by the Picture Description Task in Alzheimer's disease. *Archives of Gerontology and Geriatrics*, *27*(1), 57–65. https://doi.org/10.1016/s0167-4943(98)00099-5

Slobin, D. I. (1991). Aphasia in turkish: Speech production in broca's and wernicke's patients. *Brain and Language*, *41*(2), 149–164. https://doi.org/10.1016/0093-934x(91)90150-y

speakmore. (2021, March 5). *Speech & language therapy services and fees*. Speakmore, Independent Speech & Language Therapy. https://speakmore.co.uk/speech-language-therapy-services/

Sposaro, F., Danielson, J., & Tyson, G. (2010). *iWander: An android application for dementia patients*. 3875–3878.

Sung, J. E., DeDe, G., & Lee, S. E. (2016). Cross-Linguistic differences in a Picture-Description task between korean- and English-Speaking individuals with aphasia. *American Journal of Speech-Language Pathology*, *25*(4S), 813–822. https://doi.org/10.1044/2016_ajslp-15-0140

Sveinbjornsdottir, S. (2016). The clinical symptoms of parkinson's disease. *Journal of Neurochemistry, 139*, 318–324. https://doi.org/10.1111/jnc.13691

Swinburn, K., Porter, G., & Howard, D. (2004). *Comprehensive Aphasia Test*. Psychology Press.

Szatloczki, G., Hoffmann, I., Vincze, V., Kalman, J., & Pakaski, M. (2015). Speaking in alzheimer's disease, is that an early sign? Importance of changes in language abilities in alzheimer's disease. *Frontiers in Aging Neuroscience, 7*. https://doi.org/10.3389/fnagi.2015.00195

The Endocrine Society. (2009). *Widely used body fat measurements overestimate fatness in African-Americans, study finds*. ScienceDaily. https://www.sciencedaily.com/releases/2009/06/090611142407.htm

Thompson, C. K., Weintraub, S., & Mesulam, M.-M. (2011). *Northwestern Anagram Test*. Northwestern University.

Tippett, D. C., Thompson, C. B., Demsky, C., Sebastian, R., Wright, A., & Hillis, A. E. (2017). Differentiating between subtypes of primary progressive aphasia and mild cognitive impairment on a modified version of the frontal behavioral inventory. *PLOS ONE*, *12*(8), e0183212. https://doi.org/10.1371/journal.pone.0183212

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children.

Journal of Speech, Language, and Hearing Research, *40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

Tomoeda, C. K., Bayles, K. A., Trosset, M. W., Azuma, T., & McGeagh, A. (1996). Cross-Sectional analysis of alzheimer disease effects on oral discourse in a picture description task. *Alzheimer Disease & Associated Disorders*, *10*(4), 204–215. https://doi.org/10.1097/00002093-199601040-00006

Tsermentseli, S., Leigh, P. N., Taylor, L. J., Radunovic, A., Catani, M., & Goldstein, L. H. (2015). Syntactic processing as a marker for cognitive impairment in amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *17*(1–2), 69–76. https://doi.org/10.3109/21678421.2015.1071397

Ulatowska, H., Streit Olness, G., Wertz, R., Samson, A., Keebler, M., & Goins, K. (2003). Relationship between discourse and western aphasia battery performance in african americans with aphasia. *Aphasiology*, *17*(5), 511–521. https://doi.org/10.1080/0268703034400102

Unsworth, S. J., Sears, C. R., & Pexman, P. M. (2005). Cultural influences on categorization processes. *Journal of Cross-Cultural Psychology*, *36*(6), 662–688. https://doi.org/10.1177/0022022105280509

Vachek, J. (1976). *Selected writings in english and general linguistics*. De Gruyter Mouton.

van Es, M. A., Hardiman, O., Chio, A., Al-Chalabi, A., Pasterkamp, R. J., Veldink, J. H., & van den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *The Lancet*, *390*(10107), 2084–2098. https://doi.org/10.1016/s0140-6736(17)31287-4

Vandermosten, M., Boets, B., Poelmans, H., Sunaert, S., Wouters, J., & Ghesquiere, P. (2012). A tractography study in dyslexia: neuroanatomic correlates of orthographic, phonological and speech processing. *Brain, 135*(3), 935–948. https://doi.org/10.1093/brain/awr363

Visch-Brink, E. G., Rhee-Temme, W., Rietveld, T., Krulder, J. W. M., Harskamp, F., & Cammen, T. J. M. (2009). Improvement of spontaneous speech in early stage Alzheimer's disease with rivastigmine. *The Journal of Nutrition, Health and Aging*, *13*(1), 34–38. https://doi.org/10.1007/s12603-009-0007-z

Voelker, C. H. (1942). The one-thousand most frequent spoken-words. *Quarterly Journal of Speech*, *28*(2), 189–197. https://doi.org/10.1080/00335634209380756

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities*, *34*(9), 2586–2597. https://doi.org/10.1016/j.ridd.2013.05.014

Wallentin, M. (2018). Sex differences in post-stroke aphasia rates are caused by age. A meta-analysis and database query. *PLOS ONE*, *13*(12), e0209571. https://doi.org/10.1371/journal.pone.0209571

Wambaugh, J. L., & Ferguson, M. (2007). Application of semantic feature analysis to retrieval of action names in aphasia. *The Journal of Rehabilitation Research and Development*, *44*(3), 381. https://doi.org/10.1682/jrrd.2006.05.0038

Warrington, E., & McKenna, P. (1980). *Graded Naming Test*. Cambridge Cognition.

Wepman, J. M., & Jones, L. V. (1961). *The language modalities test for aphasia*. Education Industry Service.

Wickremaratchi, M. M., Perera, D., O'Loghlen, C., Sastry, D., Morgan, E., Jones, A., Edwards, P., Robertson, N. P., Butler, C., Morris, H. R., & Ben-Shlomo, Y. (2009). Prevalence and age of onset of Parkinson's disease in Cardiff: a community based cross sectional study and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *80*(7), 805–807. https://doi.org/10.1136/jnnp.2008.162222

Williams, S. E., & Canter, G. J. (1982). The influence of situational context on naming performance in aphasic syndromes. *Brain and Language*, *17*(1), 92–106. https://doi.org/10.1016/0093-934x(82)90007-4

Willshire, D., Kinsella, G., & Prior, M. (1991). Estimating WAIS-R IQ from the national adult reading test: A cross-validation. *Journal of Clinical and Experimental Neuropsychology*, *13*(2), 204–216. https://doi.org/10.1080/01688639108401038

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., Miller, B. L., & Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain*, *133*(7), 2069–2088. https://doi.org/10.1093/brain/awq129

Woolbert, C. H. (1922). Speaking and writing—A study of differences. *Quarterly Journal of Speech*, *8*(3), 271–285. https://doi.org/10.1080/00335632209379390

World Health Organization. (2021, September 2). *Dementia*. Dementia. https://www.who.int/en/news-room/fact-sheets/detail/dementia

Worthy, B. (2021). *Transcribing an hour of audio: How long does it really take?* GMR Transcription Services. https://www.gmrtranscription.com/blog/transcribing-an-hour-of-audio

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, *45*(1), 27–36. https://doi.org/10.1044/jshd.4501.27

Appendices

Appendix A: Individual Performance of Patients (Rutter, 2014)

Scoring table for oral speech of the patients

	Case 2	Case 4	Case 5	Case 6	Case 7	Case 8
*Time total in seconds	106	76	87	71	112	
Word count (excl jargon)	211	119	130	119	109	
*Gesture+	0	0	1	0	0	
*pauses in sec	14	25	18	24	45	
*WF Difficulties	7	15	4	4	8	
Prompts	0	0	0	0	0	
Jargon syllables	8	5	0	0	0	
Fillers	16	14	12	3	16	
Conduit d'approches	1	6	0	0	2	
Semantic substitutions	2	6	1	2	1	
*Phonemic paraphasias	0	0	0	0	1	
Repititions	6	7	3	0	8	
Revisions	2	0	0	3	0	
Metacog. Comments	4	1	3	3	0	
Main clauses	15	11	10	11	8	
Subordinate clauses	5	2	2	2	3	
Incomplete clauses	6	2	1	2	1	
Syntactic errors (total)	7	4	2	5	1	
content places	1	0	1	2	2	
content actions	5	5	6	6	5	
content subjects	3	3	3	3	2	
content objects	8	4	5	4	3	
Incorrect irrelevant content	6	2	3	2	1	
Focus	0	1	0	0	0	

	Case 10	Case 15	Case 16	Case 20
*Time total in seconds		43	70	174
Word count		0	75	24
*Gesture+		11	5	0
*pauses in seconds		0	27	109
*WF Difficulties		8	7	2
Prompts		0	3	4
Jargon syllables		26	0	8
Fillers		0	1	7
Conduit d'approches		0	0	0
Semantic substitutions		0	0	1
*Phonemic paraphasias		0	0	0
Repetitions		0	1	0
Revisions		0	0	0
Metacog. Comments		0	8	0
Main clauses		0	2	2
Subordinate clauses		0	0	0
Incomplete clauses		0	9	3
Syntactic errors (total)		0	9	3
content places		0	1	0
content actions		3	2	2
content subjects		2	0 + 2(GS)	0
content objects		3	1 +1(GS)	3
Inappropriate or irrelevant content		0	1	0
Focus		0	1	0

Scoring table for written speech of the patients

	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 10
Word count	17	7	47	71	8	20	10
Print (1) or write (2)	2	2	1	2	2	2	2
Consistent (1) inconsistent (2) in printing	1	1	1	1	2	1	1
List (1) or text (2) format	1	1	2	1	1	2	2
spelling errors	4	2	0	4	4	0	3
wordonset	0	1	0	5	0	1	1
Punctuation errors	1	1	0	3	0	2	3
Unidentifiable words	1	0	0	1	1	0	0
Unidentifiable letters	0	1	0	1	0	0	0
Crossed out words	0	1	0	2	0	0	2
Crossed out letters	0	6	0	3	0	2	0
Gaps within word	1	0	0	0	0	0	0
Semantic substitution	1	0	1	3	0	1	1
Repetitions	0	1	0	1	0	0	0
Revisions	0	0	0	3	0	1	1
Metacog. Comments	0	0	0	0	0	0	0
Main clauses	3	0	5	5	0	5	1
Subordinate clauses	0	0	1	3	0	0	0
Incomplete clauses	0	2	0	9	8	0	2
Syntactic errors (total)	3	2	0	17	8	14	4
content places	0	0	0	2	1	0	0
content actions	1	0	5	4	0	5	2
content subjects	2	0	3	3	3	3	1
content objects	3	0	7	8	3	5	3
Inappropriate or irrelevant content	0	1	0	1	0	0	0
Focus	0	0	0	0	0	0	0
other	0	0	0	0			

	Case 15	Case 16	Case 20
Word count	14	6	0
Print (1) or write (2)	1	2	2
Consistent (1) inconsistent (2) in printing	1	1	1
List (1) or text (2) format	1	1	1
spelling errors	0	1	0
Word onset	0	1	0
Punctuation errors	4	4	
Unidentifiable words	0	0	1
Unidentifiable letters	0	0	0
Crossed out words	0	0	0
Crossed out letters	0	0	0
Gaps within word	0	0	0
Semantic substitution	0	0	0
Repetitions	0	0	0
Revisions	0	0	0
Metacog. Comments	0	0	0
Main clauses	0	0	0
Subordinate clauses	0	0	0
Incomplete clauses	5	4	0
Syntactic errors (total)	5	4	0
content places	1	1	0
content actions	0	0	0
content subjects	3	2	0
content objects	7	2	0
Inappropriate or irrelevant content	1	0	0
Focus	0	0	0
Other	No verbs, Clusters nouns to sentences	0	0

Appendix B: Dyslexia Questionnaire

1. Which of the following most nearly describes your attitude toward school when you were a child:

Loved school; favourite activity				Hated school; tried to get out of going
0	1	2	3	4

2. How much difficulty did you have learning to read in elementary school?

None				A great deal
0	1	2	3	4

3. Did you ever reverse the order of letters or numbers when you were a child?

No				A great deal
0	1	2	3	4

4. How much difficulty did you have learning to spell in elementary school?

No				A great deal
0	1	2	3	4

5. Did you experience difficulties in secondary school or higher schools in English classes?

No; enjoyed and did well		Some		None
0	1	2	3	4

6. How many books (including eBooks) do you read for pleasure each year?

More than 10	6-10	2-5	1-2	None
0	1	2	3	4

7. Do you read newspapers or magazines?

Every day	Once a week	Once in a while	Rarely	Never
0	1	2	3	4

8. How many hours per day do you spend reading in the internet (computer, laptop, tablet, mobile)?

More than 4	2-4	1-2	Less than 1	None
0	1	2	3	4

*****Check the most appropriate answer for each of the following questions*****

9. Does anybody else in your family has or ever had a problem with reading or spelling?

- ❍ Yes
- ❍ No
- ❍ Not sure

If yes, please give details: ____

10. What is the highest educational level that you have attained?

- ❍ Secondary school, did not graduate (or equivalent)
- ❍ Secondary school graduate (or equivalent)
- ❍ Institute (or equivalent)
- ❍ College (or equivalent)
- ❍ University bachelor’s degree (or equivalent)
- ❍ University master’s degree (or equivalent)
- ❍ Some postgraduate education, no advanced degrees
- ❍ PhD or more advanced degrees

Appendix C: ARHQ

A Revised Adult Dyslexia Checklist

By M.D. Vinegrad

Please tick Yes or No to each question. Don't miss any questions out. If in doubt, tick which ever feels like the truer answer.

	YES	NO
1) Do you find difficulty in telling left from right?	☐	☐
2) Is map reading or finding your way to a strange place confusing?	☐	☐
3) Do you dislike reading aloud?	☐	☐
4) Do you take longer than you should to read a page of a book?	☐	☐
5) Do you find it difficult to remember the sense of what you have read?	☐	☐
6) Do you dislike reading long books?	☐	☐
7) Is your spelling poor?	☐	☐
8) Is your writing difficult to read?	☐	☐
9) Do you get confused if you have to speak in public?	☐	☐
10) Do you find it difficult to take messages on the telephone and pass them on correctly?	☐	☐
11) When you have to say a long word, do you sometime find it difficult to get all the sounds in the right order?	☐	☐
12) Do you find it difficult to do sums in your head without using your fingers or paper?	☐	☐
13) When using the telephone, do you tend to get the numbers mixed up when you dial?	☐	☐
14) Do you find it difficult to say the months of the year forwards in a fluent manner?	☐	☐
15) Do you find it difficult to say the months of the year backwards?	☐	☐
16) Do you mix up dates and times and miss appointments?	☐	☐
17) When writing cheques, do you frequently find yourself making mistakes?	☐	☐
18) Do you find forms difficult and confusing?	☐	☐
19) Do you mix up bus numbers like 95 and 59?	☐	☐
20) Did you find it hard to learn your multiplication tables at school?	☐	☐

This questionnaire was completed by 679 adults divided into the following groups:-

Students enrolled on BA & BSc degree courses	57%
Students enrolled on Certificated and Access courses	7%
A-Level students	15%
Non-student adults	21%

The age range was 18 to 68 and the proportion of men and women was approximately equal.

The sample contained 32 known dyslexic individuals. A statistical procedure involving discriminant function analysis, a factor analysis and a series of t-tests was carried out. The objective was to identify those items on the questionnaire that best discriminated between dyslexic and non-dyslexic individuals.

The 12 best items in order of importance are:-

Order	Item	Order	Item
1	Q17	7	Q19
2	Q13	8	Q14
3	Q7	9	Q20
4	Q16	10	Q4
5	Q18	11	Q1
6	Q10	12	Q11

If the majority of these items are ticked, this is a strong indication of dyslexia.

On the questionnaire as a whole (i.e. all 20 items):-

60% of people gave no more than four 'yes' responses
90% of people gave no more than eight 'yes' responses

Nine or more 'yes' responses on the questionnaire, as a whole is therefore a powerful indicator of a difficulty. The items ticked should be compared with the 'best twelve' shown above.

For further information contact: Dr Michael Vinegrad, 0208 699 9545
Reference: A Revised Adult Dyslexia Checklist. Educare, No. 48, March 1994

Appendix D: Rey-Figure Scoring Overview

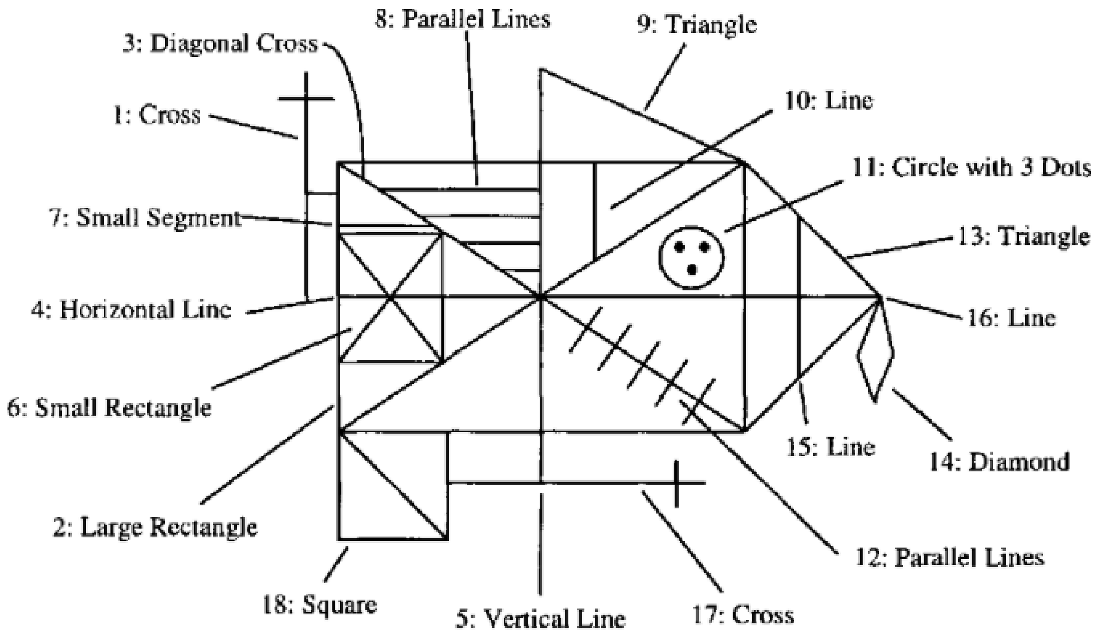


Figure 2. The Osterrieth Scoring System

Unit Correct	Placed Properly	2
	Placed Poorly	1
Unit Distorted, incomplete but recognisable	Placed Properly	1
	Placed Poorly	1/2
Absent or Unrecognisable		0

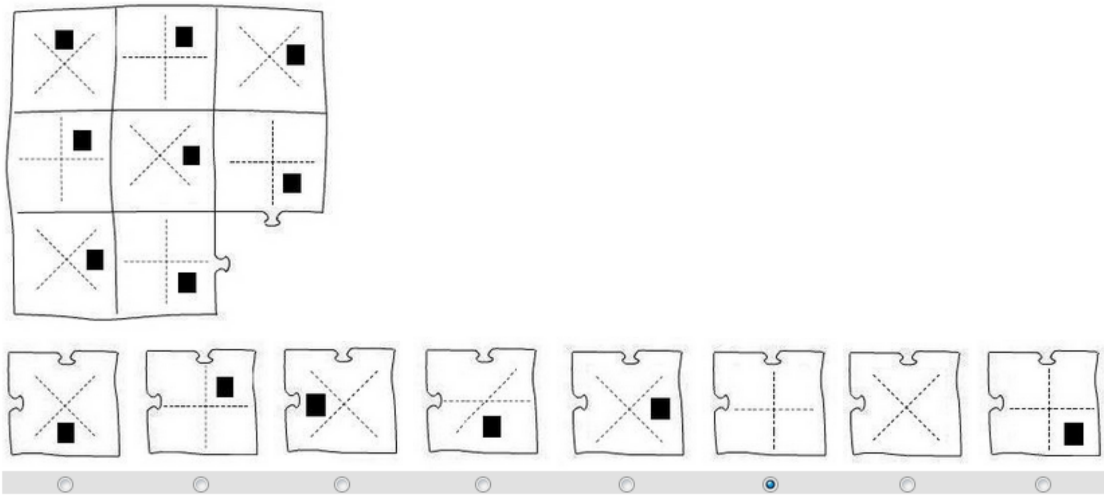
Appendix E: HMT-S

Dear participant,

this test is about finding rules in abstract patterns and to complete them in a logical way. Each task shows an incomplete jigsaw puzzle. The patterns you will see follow rules which may apply to a row, a column or to a diagonal. They may apply to the figure as a whole or to parts of it only. They may involve addition, subtraction, the alignment of figures or single components. Only one of the eight pieces given is the correct one required to complete the design.

It is your task to select the piece which completes the jigsaw puzzle. Each task needs to be completed within 2:00 minutes.

First sample
Which piece is the one required to complete the design?
(Click on the correct piece and then on "Continue")

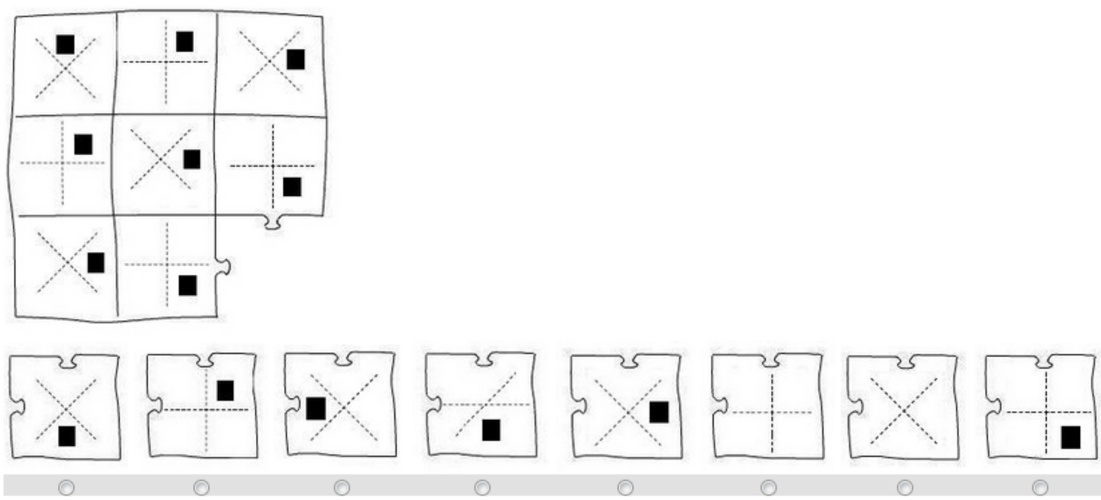


First example

Solution
The first piece is the only correct one.

The rule which defines the puzzle is: The dashed lines in the diagonal always form an X, i.e. the missing piece needs to have a dashed X. The second rule is that the position of the black rectangle systematically varies: It moves clockwise in each column and row, i.e. it has to be situated in the center of the lower part in the missing piece.

Click "Weiter/Continue" to go to the second sample task.



Appendix F: Semantic Errors

Cookie Theft Written

	Semantic Paraphasia			Semantic Conduit D'Approche			Semantic Conduit D'Écart		
	In-person	Video-call	Online	In-person	Video-call	Online	In-person	Video-call	Online
Mean	0.000	0.000	0.025	0.000	0.000	0.025	0.000	0.000	0.050
Std. Deviation	0.000	0.000	0.158	0.000	0.000	0.158	0.000	0.000	0.221
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000

Appendix G: Phonological Errors

Cookie Theft Spoken

	Phonological Conduit D‘Approche			Phonological Conduit D‘Écart		
	In-person	Video-call	Online	In-person	Video-call	Online
Mean	0.233	0.200	0.125	0.000	0.033	0.000
Std. Deviation	0.430	0.407	0.335	0.000	0.183	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	1.000	1.000	1.000	0.000	1.000	0.000

Cookie Theft Written

	Phonological Additions			Phonological Substitutions			Phonological Metatheses		
	In-person	Video-call	Online	In-person	Video-call	Online	In-person	Video-call	Online
Mean	0.000	0.000	0.100	0.000	0.000	0.025	0.033	0.000	0.025
Std. Deviation	0.000	0.000	0.304	0.000	0.000	0.158	0.183	0.000	0.158
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	1.000	0.000	0.000	1.000	1.000	0.000	1.000

Cookie Theft Written

	Phonological Conduit D'Approche			Phonological Conduit D'Écart			Phonological Neologism		
	In-person	Video-call	Online	In-person	Video-call	Online	In-person	Video-call	Online
Mean	0.000	0.000	0.050	0.000	0.000	0.050	0.000	0.000	0.050
Std. Deviation	0.000	0.000	0.221	0.000	0.000	0.221	0.000	0.000	0.221
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000

Cookie Theft Written

	Phonological Jargon		
	In-person	Video-call	Online
Mean	0.000	0.000	0.025
Std. Deviation	0.000	0.000	0.158
Minimum	0.000	0.000	0.000
Maximum	0.000	0.000	1.000

Appendix H: Lexical Errors

Cookie Theft Written

	Semantic Substitutions	
	handwriting	typing
Valid	15	15
Missing	0	0
Mean	0.267	0.000
Std. Deviation	0.458	0.000
Minimum	0.000	0.000
Maximum	1.000	0.000

Appendix I: Disfluencies

Cookie Theft Written

	Discontinued Clauses		Not Identifiable Word(s)		Crossed Out Letter(s)		Crossed Out Word(s)	
	handwriting	typing	handwriting	typing	handwriting	typing	handwriting	typing
Valid	15	15	15	15	15	15	15	15
Missing	0	0	0	0	0	0	0	0
Mean	0.000	0.067	0.200	0.000	0.267	0.000	0.267	0.000
Std. Deviation	0.000	0.258	0.414	0.000	0.799	0.000	0.799	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	1.000	1.000	0.000	3.000	0.000	3.000	0.000

Cookie Theft Written

	Punctuation Error		Spelling Error		Word Onset Error	
	handwriting	typing	handwriting	typing	handwriting	typing
Valid	15	15	15	15	15	15
Missing	0	0	0	0	0	0
Mean	0.267	0.333	1.267	0.000	0.200	0.200
Std. Deviation	0.458	0.816	2.052	0.000	0.561	0.775
Minimum	0.000	0.000	0.000	0.000	0.000	0.000

Cookie Theft Written

	Punctuation Error		Spelling Error		Word Onset Error	
	handwriting	typing	handwriting	typing	handwriting	typing
Maximum	1.000	3.000	7.000	0.000	2.000	3.000

Appendix J: Syntactic Error Ratio

Syntactic Error Ratio

	Cookie Theft Spoken			Cookie Theft Written			Indian Picture Spoken			Indian Picture Written		
	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.000	0.000	0.000	0.025	1.247	0.630	0.884	0.000	0.000	0.000
Std. Deviation	0.000	0.000	0.000	0.000	0.000	0.158	1.553	0.872	1.729	0.000	0.000	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	0.000	0.000	0.000	1.000	5.970	3.922	6.667	0.000	0.000	0.000

Appendix K: Semantic Substitution Ratio

Semantic Substitution Ratio

	Cookie Theft Spoken			Cookie Theft Written			Indian Picture Spoken			Indian Picture Written		
	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.029	0.060	0.042	0.084	3.893	0.084	0.055	0.000	0.011	0.017	0.059	0.000
Std. Deviation	0.157	0.329	0.202	0.374	10.514	0.364	0.229	0.000	0.073	0.094	0.225	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.862	1.802	1.205	2.000	50.000	2.174	1.170	0.000	0.459	0.515	0.893	0.000

Appendix L: Phonological Errors

Cookie Theft Picture Spoken

	Phonological Elision			**Phonological Addition**			**Phonological Substitution**			**Phonological Metathesis**		
	In Person	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Std. Deviation	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Cookie Theft Picture Spoken

	Phonological Conduit D'Approche			Phonological Conduit D'Écart			Phonological Neologism			Phonological Jargon		
	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.233	0.200	0.125	0.000	0.033	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Std. Deviation	0.430	0.407	0.335	0.000	0.183	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	1.000	1.000	1.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Cookie Theft Picture Written

	Phonological Elision			Phonological Addition			Phonological Substitution			Phonological Metathesis		
	In Person	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.025	0.000	0.000	0.100	0.000	0.000	0.025	0.033	0.000	0.025
Std. Deviation	0.000	0.000	0.158	0.000	0.000	0.304	0.000	0.000	0.158	0.183	0.000	0.158
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	1.000	0.000	1.000

Cookie Theft Picture Written

	Phonological Conduit D'Approche			Phonological Conduit D'Écart			Phonological Neologism			Phonological Jargon		
	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.050	0.000	0.000	0.050	0.000	0.000	0.050	0.000	0.000	0.025
Std. Deviation	0.000	0.000	0.221	0.000	0.000	0.221	0.000	0.000	0.221	0.000	0.000	0.158
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000

Indian Picture Spoken

	Phonological Elision			Phonological Addition			Phonological Substitution			Phonological Metathesis		
	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.125	0.000	0.000	0.025	0.000	0.033	0.025	0.000	0.000	0.050
Std. Deviation	0.000	0.000	0.335	0.000	0.000	0.158	0.000	0.183	0.158	0.000	0.000	0.221
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	1.000	0.000	0.000	1.000	0.000	1.000	1.000	0.000	0.000	1.000

Indian Picture Spoken

	Phonological Conduit D'Approche			Phonological Conduit D'Écart			Phonological Neologism			Phonological Jargon		
	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.067	0.100	0.050	0.000	0.033	0.025	0.000	0.000	0.050	0.000	0.000	0.000
Std. Deviation	0.254	0.403	0.221	0.000	0.183	0.158	0.000	0.000	0.221	0.000	0.000	0.000
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	1.000	2.000	1.000	0.000	1.000	1.000	0.000	0.000	1.000	0.000	0.000	0.000

Indian Picture Written

	Phonological Elision			Phonological Addition			Phonological Substitution			Phonological Metathesis		
	In Person	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**	**In Person**	**Video-call**	**Online**
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.100	0.000	0.000	0.033	0.000	0.075	0.033	0.000	0.075	0.000	0.000	0.050
Std. Deviation	0.403	0.000	0.000	0.183	0.000	0.267	0.183	0.000	0.267	0.000	0.000	0.221
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	2.000	0.000	0.000	1.000	0.000	1.000	1.000	0.000	1.000	0.000	0.000	1.000

Indian Picture Written

	Phonological Conduit D'Approche			Phonological Conduit D'Écart			Phonological Neologism			Phonological Jargon		
	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online	In Person	Video-call	Online
Valid	30	30	40	30	30	40	30	30	40	30	30	40
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.000	0.075	0.000	0.000	0.025
Std. Deviation	0.000	0.000	0.335	0.000	0.000	0.000	0.000	0.000	0.267	0.000	0.000	0.158
Minimum	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Maximum	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	1.000