Improving cognitive testing in the context of dementia: Tools for an earlier and better-characterized diagnosis of Alzheimer´s dementia and Primary Progressive Aphasia.

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

genehmigt durch die Fakultät für Naturwissenschaften

der Otto-von-Guericke-Universität Magdeburg

von M.Sc. Ornella Billette

geb. am 03.05.1989 in Reims (Frankreich)

Gutachter: Prof. Dr. Peter Nestor

Prof. Dr. Asmus Vogel

eingereicht in 01.06.2018

verteidigt am 12.10.2018

To all people facing a disease that cannot be cured.

Acknowledgement

This PhD would not have been possible without the help of many people and here is my chance to thank them.

Firstly, I would like to express my deep gratitude to all study participants and their families. It has been a pleasure and an incredible enrichment on both the human and the professional side. Although many of them went through the hardest moments of their lives, they kept smiling and gave their time to help our research.

I want to thank my supervisor Prof. Peter Nestor for his support throughout the PhD. He always took the time to answer what I called “a little question” but was in fact a collection of not-so-little questions. I cannot thank him enough for introducing me to the field of Primary Progressive Aphasia and giving me the opportunity to immerse in clinical research. It was a pleasure to do my Phd with him, not only for his incredible clinical knowledge but also for the culinary discussions.

Special thanks goes to Prof. Zsolt Cséfalvay who collected data in Slovakia, and allowed us to internationalize even more our project.

I am grateful to Florian Kamm for computerizing the Graded Object-Naming Task and the Virtual City Task and introducing me to Blender and Psychopy.

I am indebted to many colleagues in the DZNE for helping me in numerous aspects of my PhD. Urte Schneider, Franziska Schulze and Christin Russ and their extensive knowledge of all patients in the memory clinic have been indispensable. Karen Müller-Zabel did not only guide me through all paperwork but was also always present for anything one could need. My German would not be what it is without my fellow PhD students Daniel Preiß and Elisabeth Tute. I can´t thank all of them enough for their help and the many laughter we had over the years.

Thank you to Julia Große-Schwiep, Julia Andreev and Ann-Katrin Domke who were student assistants in our research group and helped to collect data, and brought good mood every day at the office.

Last but not last, my family and friends have been precious not only during my PhD but during my whole studies. Sophie, whom I actually met on day one of the psychology bachelor, has been an incredible support and although 1250 kilometers separate us, she

lived every second of this PhD with me. Merci to my mum who always encouraged me to follow my intuition and study what would make me happy. Danke to my parents in law for baby-sitting and giving me the precious time that allowed me to finish writing my thesis. A special thanks to Poppy who took time to answer some of my questions on English grammar and whose wonderful British accent always makes my day. Thank you to my husband Felix for his never-ending and soothing support. He never doubted me and gave me all the encouragement I needed when I needed it. Finally, thank you to my son, Émile, for being the funniest baby and keeping the mood up in the last months of my PhD.

Abstract

To date, trials to delay, or prevent cognitive decline in neurodegenerative diseases such as Alzheimer's (AD) remained ineffective. It is now believed that they should be conducted in the earliest stages of the disease. Tools to track early cognitive decline are, however, lacking. Moreover, the correspondence between the clinical profile and the underlying pathology is not systematic. Primary Progressive Aphasia (PPA) for example, presents different clinical profiles, only probabilistically associated with a certain pathology.

This work aimed at developing an international test battery, the DZNE-Cog, to track early cognitive decline. Further aims were to develop tools to ease PPA subtyping as well as to examine the clinical profile of PPA patients with confirmed amyloid pathology.

Firstly, two subtests of the DZNE-Cog have been developed: the Graded Object-Naming Task (GONT) and the Virtual City Task (VCT) that investigate naming abilities and topographical memory. The GONT involves naming objects of graded difficulty while the VCT requires memorizing a route driven in a virtual environment. The GONT was piloted with 27 German and 63 Slovak healthy elderly (HC). It was then administered cross-sectionally to 78 HC, 17 patients with Subjective Cognitive Impairment (SCI), 19 patients with Mild Cognitive Impairment (MCI), 13 AD and 26 PPA and longitudinally to 41 HC, 10 AD, 15 MCI and 15 SCI. The VCT was administered cross-sectionally to 20 HC, 10 AD, 11 MCI and 4 PPA and longitudinally to 69 HC, 15 SCI, 16 MCI and 3 AD. Performance in both tests were compared to the Boston Naming Test and the recall of the Rey-Osterrieth Complex figure.

Secondly, Agrammatism being a central feature for subtyping PPA, the Make A Sentence Test (MAST) and SEntence Comprehension Test (SECT) were developed and administered to 41 PPA patients, 21 AD and 30 HC. Furthermore, mistake patterns in verb inflection in German were explored in 9 patients with the semantic variant of PPA (SvPPA), 4 PPA patients with a mixed clinical profile and confirmed amyloid pathology (Aß+PPA) and 12 AD. Finally, repetition of words and sentences, verbal and spatial spans, semantics and grammar were assessed in 11 Aß+PPA, 9 SvPPA, 6 patients with the non-fluent variant of PPA (NfvPPA), and 28 HC.

The GONT and VCT showed superiority over the gold standard in their domain on cross-sectional comparisons. Both tests displayed graded difficulty, international applicability and tracked slight cognitive change over time. The sample size and amount of longitudinal data being, however, restricted, future work will have to confirm the results.

The MAST and SECT displayed good abilities to detect agrammatism in NfvPPA and PPA with a mixed clinical profile. Over-regularizations of regular verbs seemed specific to SvPPA and could therefore work as an analogue to surface dyslexia in language with high grapheme-phoneme correspondence. Finally, repetition of words and sentences were not of use for the differential diagnosis of NfvPPA and Aß+PPA.

Zusammenfassung

Therapeutische Interventionen, die versucht haben, den Abbau kognitiver Fähigkeiten im Zusammenhang mit neurodegenerativen Erkrankungen wie Morbus Alzheimer (AD) zu verlangsamen, zu verzögern oder zu verhindern, sind bis jetzt erfolglos geblieben. Dieser Mißerfolg führte zu der Annahme, dass die Interventionsversuche möglicherweise an einem zu fortgeschrittenem Zeitpunkt im Krankheitsverlauf ansetzten, an welchem zerebrale Schäden bereits nicht mehr aufhaltbar bzw. zu beheben waren. Daraus ergab sich der Ansatz, therapeutische Studien möglichst frühzeitig schon im pre-symptomatischen Stadium durchzuführen. Neuropsychologische Tests, welche frühe kognitive Veränderungen aufzeigen, fehlen allerdings. Ein Ziel der vorliegenden Arbeit liegt in der Entwicklung einer international anwendbaren neuropsychologischen Testbatterie, der DZNE-Cog, die den frühen kognitiven Abbau verfolgt.

Hinzu kommt, dass die Assoziation zwischen klinischem Profil und der zugrundeliegenden Pathologie bei weitem nicht systematisch ist. Ein Beispiel hierfür ist die Primäre Progressive Aphasie (PPA), eine Demenzform, in welcher in erster Instanz die Sprache beeinträchtigt wird. Diese Beeinträchtigung kann sich in verschiedenen klinischen Profilen äußeren, die nur probabilistisch mit bestimmten zugrundeliegenden Pathologien verbunden sind. Aus diesem Grund ist eine Verbesserung der derzeitigen neuropsychologischen Standardtests zur PPA-Subtypisierung durch die Entwicklung neuer Tests eine zweite Zielstellung der vorliegenden Arbeit. Des weiteren wurde das klinische Profil von PPA-Patienten untersucht, bei welchen eine zugrunde liegende Amyloid Pathologie bestätigt wurde.

Im ersten Schritt wurden zwei Subtests der DZNE-Cog entwickelt und vorgestellt: der Graded Object-Naming Task (GONT), welcher die Benennungsfähigkeiten untersucht ,und der Virtual City Task (VCT), welcher das topografische Gedächtnis prüft. Im GONT sollen Probanden Bilder von Objekten verschiedener Schwierigkeitsstufen benennen, während der VCT das Erlernen einer Route zum Ziel hat, die in einer virtuellen Umgebung abgefahren werden soll.

Die Pilotierung involvierte die Administration des GONT in einer Stichprobe von 27 deutschen und 42 slowakischen, gesunden älteren Probanden (HC), um die internationale Anwendbarkeit des Tests zu prüfen. Der finalisierte GONT wurde sowohl querschnittlich mit 78 gesunden älteren Probanden (HC), 17 Patienten mit subjektiver kognitiver Störung (SCI), 19 Patienten mit leichter kognitiver Störung (MCI), 13 AD und 26 PPA, als auch längsschnittlich mit 41 HC, 10 AD, 15 MCI und 15 SCI durchgeführt. Der VCT wurde sowohl querschnittlich mit 20 HC, 10 AD, 11 MCI und 4 PPA als auch längsschnittlich mit 69 HC, 15 SCI, 16 MCI und 3 AD durchgeführt. Die Leistungen in beiden Tests wurden mit den maßstäblichen Tests in den jeweiligen Testungsgebieten, dem Boston Naming Test und dem Abrufen der Rey-Osterrieth Complex Figure, verglichen.

Da Agrammatismus ein zentrales Merkmal für die Subtypisierung von PPA darstellt, aber schnelle standardisierte Tests zur Objektivierung derzeit fehlen, wurden im zweiten Schritt der Make A Sentence Test (MAST) und der Sentence Comprehension Test (SECT) entwickelt. Diese Tests untersuchen jeweils produktive und rezeptive Grammatikfähigkeiten und wurden mit 41 PPA-Patienten aller klinischer Subtypen sowie 21 AD und 30 HC durchgeführt.

Zudem wurden Verbkonjugationen und die zugehörigen Fehlermuster im Perfekt und Präteritum bei 9 Patienten mit der semantischen Variante der PPA und einer ausgeschlossenen Amyloid Pathologie (SvPPA), bei 4 PPA-Patienten mit gemischtem klinischen Pofil und bestätigter zugrundeliegender Amyloid Pathologie (Aß+PPA) und bei 12 AD untersucht. Schließlich wurde das Nachsprechen von Wörtern und Sätzen, die visuo-räumliche und verbale Spanne sowie die semantischen und grammatikalischen Fähigkeiten bei 11 Aß+PPA, 9 SvPPA, 6 Patienten mit der nicht-flüssigen Variante der PPA (NfvPPA) und 28 HC untersucht.

Die Ergebnisse zeigten, dass der GONT und der VCT als kurze Tests mit graduierter Schwierigkeit charakterisiert werden können, die sowohl quer- als auch längsschnittlich besser als die derzeitig maßstäblichen neuropsychologischen Tests des jeweiligen Fähigkeitsgebiets geeignet sind. Außerdem können der GONT und der VCT milde Veränderungen der kognitiven Leistung im Laufe der Zeit verfolgen. Letztlich konnte die internationale Anwendbarkeit des GONT bestätigt werden. Die Größe der Stichprobe und die Anzahl der längsschnittlichen Daten waren dennoch begrenzt, weshalb weitere Untersuchungen zur Bestätigung der Ergebnisse durchgeführt werden sollten.

Die neuentwickelten Tests MAST und SECT haben ihre gute Eignung zur Prüfung von Agrammatismus, besonders bei NfvPPA und PPA-Patienten, mit einem gemischten Krankheitsbild zeigen können. Darüber hinaus konnten wir bestätigen, dass korrekte Verbkonjugation in SvPPA durch Häufigkeit und Regelmäßigkeit des Verbs beeinflusst wird. Es wurde weiterhin belegt, dass Überregulierung unregelmäßiger Verben spezifisch für SvPPA ist, während sich bei Aß+PPA eine Reihe verschiedener anderer Fehler zeigen. Demzufolge könnte in Sprachen mit hoher Graphem-Phonem-Korrespondenz, in welchen Oberflächendyslexie schwer nachweisbar ist, die Überregulierung unregelmäßiger Verben in SvPPA als analoges Symptom genutzt werden.

Abschließend und entgegen der derzeitigen Annahmen zeigte sich, dass das Nachsprechen von Wörtern und Sätzen für die Differenzialdiagnostik von NfvPPA und Aß+PPA nicht geeignet ist.

Table of content

Acronyms

A4:	Anti-Amyloid treatment in Asymptomatic Alzheimer´s Disease
AAT:	Aachener Aphasie Test
ADAS-Cog:	Alzheimer's Disease Assessment Scale-cognitive subscale
Aß+PPA:	Amyloid related Primary Progressive Aphasia
ACE-R:	Addenbrooke´s Cognitive Examination Revised
AD:	Alzheimer´s Disease
ANOVA:	ANalysis Of VAriance
BNT:	Boston Naming Test
CCT:	Camel & Cactus Test
CERAD:	Consortium to Establish a Registry for Alzheimer's Disease
CSF:	Cerebrospinal Fluid
df:	Degree of Freedom
DIF:	Differential Item Functioning
DIAN:	Dominantly Inherited Alzheimer Network
DO80:	Dénomination Orale d'Images (Test)
fMRI:	Functional Magnetic Resonance Imaging
FTD:	Fronto-Temporal Dementia
FTLD:	Fronto-Temporal Lobar Degeneration
GONT:	Graded Object-Naming Task
HC:	Healthy Control
MCI:	Mild Cognitive Impairment
MRI:	Magnetic Resonance Imaging
NfvPPA:	Non-Fluent Variant of Primary Progressive Aphasia
LvPPA:	Logogenic Variant of Primary Progressive Aphasia
MAST:	Make A Sentence Test
MMSE:	Mini-Mental State Examination
MOCA:	MOntreal Cognitive Assessment
MPPA:	Mixed Primary Progressive Aphasia
N.A:	Non-available
NAT:	Northwestern Anagram Test
N.S:	Non Significant
PET:	Positron Emission Tomography
PNT:	Philadelphia Naming Test
PPA:	Primary Progressive Aphasia
R&P:	Repeat and Point test
ROCF:	Rey-Osterrieth Complex Figure
SCI:	Subjective Cognitive Impairment
SD:	Standard Deviation
SvPPA:	Semantic Variant of Primary Progressive Aphasia
SECT A:	SEntence Comprehension Test Auditory version
SECT V:	SEntence Comprehension Test Visual version
TROG:	Test for Reception Of Grammar
VCT:	Virtual City Task

List of figures

List of tables

Chapter 1.

Introduction

"*Nature has only a single path and that path is run but once, and to each stage of existence has been allotted its own appropriate quality; so that the weakness of childhood, the impetuosity of youth, the seriousness of middle life, the maturity of old age — each bears some of Nature's fruit, which must be garnered in its own season."*

Cicero, de Senectute

Etymologically the word "dementia" comes from the classical Latin "dementis", meaning "without intellect, without spirit" (privative prefix "de" and substantive "mens"). Clinically, the definition of the concept of dementia has largely evolved since its first use. It was limited and defined during the 19th century when the word "dementia" was used to describe the end point of insanity, referring to any behaviour that was out of the normal range regardless of its expression, origin or reversibility. In the last version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013) dementia falls under the rubric of major neurocognitive disorder. It is

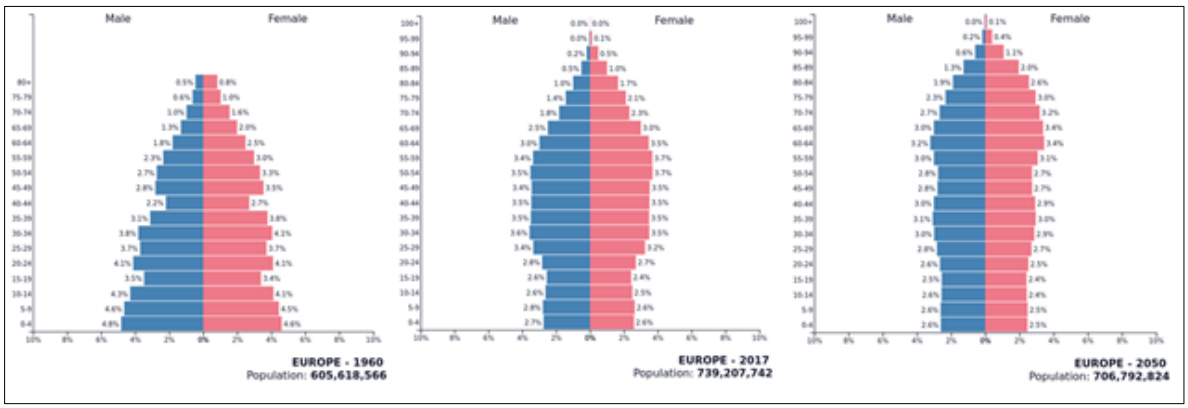
defined by a significant decline in one or more cognitive domains (memory, language or executive function for example) that interfere with independence in activities of daily living. Moreover, it is not linked to a state of delirium and is not better explained by another mental disorder.

Various neurodegenerative diseases, to date incurable, can cause dementia with various clinical profiles. Alzheimer´s disease is the most common dementia-causing pathology. It is responsible for the most prevalent dementia profile, namely dementia of Alzheimer´s type where memory and orientation are impaired first of all. In Germany, 2/3 of the 1.5 million demented elderly people suffer from dementia of Alzheimer type and 40 000 new cases are expected each year (Statistisches Bundesamt, 2015). Alzheimer´s disease can also lead to different rare dementia types like Posterior Cortical Atrophy or Primary Progressive Aphasia where visuo-spatial abilities and language, respectively, are impaired first of all. Different underlying pathologies like a tauopathy or a TDP-43 proteinopathy in turn, can also cause Primary Progressive Aphasia. The imperfect correspondence between the clinical profile and the underlying pathology, illustrated by the example of Primary Progressive Aphasia, is a major throwback for both clinicians and researchers in the field of dementia.

Increasing age is the most important risk factor for dementia and is, unlike most other identified factors (for example cardio-vascular risks like obesity or inactivity) not avoidable. In Europe, the percentage of people with a diagnosis of dementia of Alzheimer´s type increases from 1.6% in 65-69 year olds to 15.6% in 80-84 year olds (Eurocode, 2006). The increasing life expectancy will lead to a reversal of the age pyramid in western societies (see figure 1.1) with a growing number of people over 65. Although studies have reported a relative decline in the incidence of dementia in age specific groups (Manton, Gu, & Ukraintseva, 2005; Matthews et al., 2013), predictions foresee an increase in the absolute number of dementia cases due to the mentioned raise in the number of elderly. Dementia causes a loss of autonomy, often leading to being placed in a nursing home, which in turn is a massive societal burden. Therefore, in the last decades, both at the national and international level, research in the dementia field was made a financial, infrastructural and sociological priority. In 2007, for example, France declared Alzheimer's disease the nation's greatest research need of that year.

Figure 1-1 Predicted age pyramid in Europe

Source: https://www.populationpyramid.net/europe/



In spite of the great interest directed toward a better understanding of the so-called "epidemic of the century", many unanswered questions remain. Research on preventive and therapeutic solutions has been, to date, unfruitful. Risk and protective factors need to be more clearly identified. A better understanding of the correspondence between the clinical profile and the underlying disease is needed. Tools for an early and accurate diagnosis of the dementia type and associated underlying disease are required.

This work intends to make its contribution in the worldwide fight against dementia by presenting a tool for tracking early cognitive changes in the elderly and investigating the clinical profile of a rarer dementia syndrome: Primary Progressive Aphasia.

1.1 Organization of the thesis

The first part of this thesis deals with the early detection of cognitive impairment. The development of two subtests of the DZNE-Cog, a neuropsychological test battery to track changes from normal cognition to early cognitive impairment in single subjects over time is presented.

The second part deals with a rare dementia syndrome, namely Primary Progressive Aphasia (PPA). New neuropsychological tools aiming at better describing and easing the subtyping of the different variants of PPA are presented and the clinical profile of PPA patients with a confirmed underlying amyloid pathology is discussed.

Chapter 2.

Early detection of cognitive impairment

2.1 Introduction

Dementia and its many economical, psychological and logistical consequences are challenging for aging societies. Presently, therapeutic studies to slow, delay, or even, prevent cognitive decline in neurodegenerative diseases such as Alzheimer's disease (AD) have been ineffective. Although pharmaceutical trials managed to reverse part of the main biological consequences of the disease (amyloid deposition) and improve cognitive impairment in the animal model (Morgan et al., 2000; Yin et al., 2016), so far all phase III trials failed to stop cognitive decline in the long term (Holmes et al., 2008; Rafii & Aisen, 2015). In the past, most trials have, however, focused on participants with clinically established dementia. Their failure has led to the hypothesis, in both the research field and pharmaceutical industries, that they were perhaps conducted too late in the course of the illness, when cerebral damage is so extended in the brain that it can no longer be reversed or stopped. Indeed, in most cases, by the time a clinical diagnosis of AD is given the neuronal loss is already extensively advanced (Price et al., 2001).

This hypothesis has led to the view that neurodegenerative diseases should be treated at the earliest symptomatic stages, or at best, even before the first symptoms have developed, in the so-called pre-symptomatic stage (Dubois et al., 2007). The typical biological features of AD, namely the extracellular deposition of amyloid plaques as well as the increase of intracellular neurofibrillary tangles, start settling many years before the first symptoms of the disease appear, or at least, before they can be revealed by the neuropsychological tools at our disposition (Price & Morris, 1999; Stomrud, Hansson, Blennow, Minthon, & Londos, 2007).

Several studies try to implement the idea of an early therapeutic intervention. The study of the Dominantly Inherited Alzheimer Network (DIAN), for example, follows cohorts with

familial AD (www.dian-info.org). Carriers of genetic mutations causing AD as well as their relatives are enrolled. The risk of inheriting the gene being of 50%, the study follows both carriers and non-carrier relatives, the later serving as a control group. Targeting carriers years before the age of onset of the affected family member allows testing the preventive value of therapeutic compounds, mostly anti-amyloid, at the earliest stages or even before the disease manifests clinically.

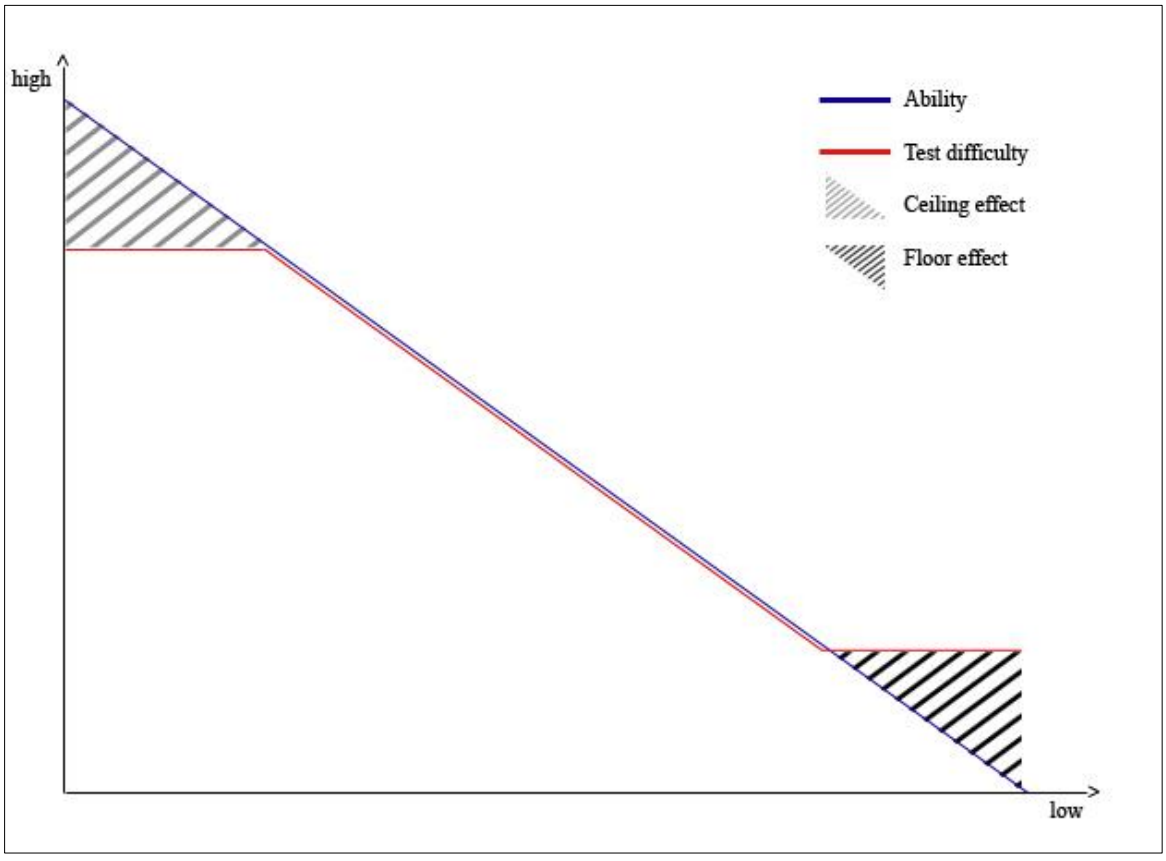
The Anti-Amyloid Treatment in Asymptomatic Alzheimer's study ("A4 study") targets older individual (65-85 year olds) with normal cognitive abilities that present a higher load of amyloid in the brain compared to the average population (https://a4study.org). The aggregation of amyloid in the brain being one of the hallmarks of AD, these participants are considered at high risk of developing the disease (Jack et al., 2010) and thus might be a desirable target group for therapeutic trials.

While starting therapeutic trials earlier in the course of the disease seems a good strategy, it raises several logistical problems. Therapeutic trials in the earliest stages of the disease require not only identifying individuals at higher risk of developing AD but also clinical endpoints that prove efficacy of treatment. Official guidelines recommend that the efficacy of a treatment is documented not only through functional but also cognitive improvement (European Medicines Agency, 2016; U.S. Department of Health and Human Services Food and Drug Administration, CDER & CBER, 2018). Current neuropsychological tools available in clinical settings and the research field are, however, insensitive to the earliest changes in cognitive abilities. This is particularly true in people with a high level of education where the first impairments are detected later in the course of the disease. Indeed, at an equal level of neurofibrillary tangles and amyloid plaques in the brain, more educated patients have more preserved cognitive abilities on formal neuropsychological examination (Rentz et al., 2016). This elongation of the average delay between the settling of the first brain damage and the appearance of the first cognitive symptoms in more educated elderly is often attributed to the so-called "cognitive and brain reserve". A higher level of education and activity would result in broader knowledge and better interconnection between brain structures. In early stages, it would compensate to a certain extent the neuronal loss and the associated cognitive decline (Stern, 2012). This idea highlights one of the biggest challenges in the neuropsychological field: detecting early cognitive deterioration in highly educated elderly. Most existing neuropsychological tools suffer from ceiling effects in healthy highly educated people; masking the decline in abilities longer than in less educated

elderly (Tuokko, Garrett, McDowell, Silverberg, & Kristjansson, 2003). There is, therefore, a need for developing neuropsychological tests that are devoid of ceiling effects in healthy elders. Such tests might even challenge the concept of cognitive reserve. Indeed, the delay in the detection of the first cognitive impairment in highly educated elderly might be a bias due to the lack of appropriate tools. Furthermore, many neuropsychological tests suffer from floor effects in patients, making it difficult in moderate and more severe dementia to describe progression. Tracking change in more advance dementia is desirable to better characterize the evolution of the disease and particularly to develop psychosocial assistance strategies in nursing house or at home. The concepts of ceiling and floor effects are depicted in figure 2.1.

Figure 2-1 Depiction of floor and ceiling effects

Ceiling effect appears when, upon a certain level of ability, all participants obtain the maximum score. Example: A test requires memorizing a maximum of five items and all participant that can memorize five or more items will obtain the maximum score. The test will not allow differentiating between participants that can memorize five, six or more items. Floor effect appears when, under a certain level of ability, all participants obtain the minimum score. Example: A test requires memorizing a minimum of three items and all participant that cannot memorize three items will obtain the minimum score. The test will not allow differentiating between participants that can memorize two, one or no item.



In a clinical trial aiming at slowing cognitive decline, accurate and objective quantification of the impact of an intervention is needed. To date, one of the most used cognitive measures in phase III clinical trials is the Alzheimer´s Disease Assessment´s Scale-Cognitive subscale (ADAS-Cog) (W. G. Rosen, Mohs, & Davis, 1984) that was created to longitudinally track the severity of cognitive impairments in AD. It was, however, developed at a time when clinical trials were conducted in established dementia. With the shift toward trials in the earliest stages of AD, ADAS-Cog has shown some limits and lacks sensitivity to track the transition from normal aging to early stage dementia (Karin et al., 2014; Posner et al., 2013). In a study using item response theory (IRT) the inability of ADAS-cog to discriminate between different levels of impairment in mild AD was shown (Benge, Balsis, Geraci, Massman, & Doody, 2009). The authors concluded that ADAS-Cog is more adapted to moderate and established dementia than for the early detection of cognitive impairment. Noteworthy, is that the memory subtest is the best to predict a disease in the milder stages. Furthermore, its sensitivity to mild impairment can be improved with the inclusion of a delayed recall subtest (Lowe, Balsis, Benge, & Doody, 2015) or by using the 13-items version of the test with both delayed recall and digit cancelation subtests (Mohs et al., 1997).

Another reported weakness of the scale is that a decline in the score does not always correspond to a significant change on the individual clinical level (Rockwood, Fay, Gorman, Carver, & Graham, 2007). The accepted 4-points change criterion in ADAS-Cog for meaningful clinical decline did not always detect real cognitive decline in single subjects (Rockwood, Fay, & Gorman, 2010).

2.2 Aim

The study aimed at developing a neuropsychological test battery, the DZNE-Cog, to accurately and sensitively track change from normal cognition to early stage dementia. It involved developing tests of graded difficulty that cover the main cognitive areas: memory, language, visuo-perceptual ability and attention/executive functions. The tests had to be devoid of ceiling effects in HC and floor effects in patients. In praxis, all HC should not get the maximum score and all patients the minimum score on the test. Time being a major constraint in both clinical and research settings, the battery should not take more than 30 minutes to administer. Furthermore, with many therapeutic trials being led internationally, we aimed at creating a battery that would be as free as possible of cultural references. The battery had to be simple to administer so that researchers and clinicians with diverse range of experience could implement it. Fixed quotation criteria had to be developed to facilitate standardized administration and replication of findings. Finally, the battery was planned to be fully computer-based with an easy interface to navigate between the four tests components and facilitate data capture.

In this thesis, preliminary results on two of the four planned tests are presented and discussed: the language and the memory assessments.

2.3 Inclusion criteria

In chapters 3 and 4, healthy controls (HC) were former participants of studies carried out in the institute or were recruited thanks to events like “the long night of science” and advertisements in local newspapers. Participants were included as HC when they had no history of major neurological or psychiatric disorder, performed normally on general neuropsychological assessment and scored a minimum of 27 points on the Mini-Mental State Examination (MMSE) (Folstein, Folstein, & McHugh, 1975).

Patients diagnosed with Mild Cognitive Impairment (MCI), early AD and participants with Subjective Cognitive Impairment (SCI) were recruited in the local memory clinic. Inclusion criteria specific to patients were: an MMSE higher than 18 and deficits that were not so extended that the participants could not understand the tasks. Diagnoses were made by a neurologist following the international guidelines for the diagnosis of AD (McKhann et al., 2011) and MCI (Albert et al., 2011). The label of SCI implied complaints regarding cognitive decline that motivated a consultation with a neurologist with no objective cognitive decline in the neuropsychological and neurological examinations (Jessen et al., 2014).

2.4 Ethical approval

Ethical approval for both the study on tracking early cognitive changes (chapters 3, 4 and 5) and the study on Primary Progressive Aphasia (chapters 8 and 9) were obtained from the ethic committee of the Otto-von-Guericke-University in Magdeburg. The ethical approval for the study on Primary Progressive Aphasia presented in chapter 7 was obtained from Cambridgeshire 3 Research Ethics Committee. Written informed consent was obtained for all participants, or when needed their next of kin.

Chapter 3.

Development of an international Graded Object-Naming Task

3.1 Introduction

Anomia, the inability to recall words, is not the most prominent feature in the classical presentation of Alzheimer´s disease (AD) where episodic memory is the first reported impairment, both by patients and their caregivers (McKhann et al., 2011; Ryu, Lee, Kim, & Lee, 2016). Decline in naming abilities, however, eventually appears progressively in the course of the disease along with impairment of semantic knowledge (Huff, Corkin, & Growdown, 1986; Lin et al., 2014; Salmon & Bondi, 2009; Silagi, Bertolucci, & Ortiz, 2015). Decline in language abilities is reported as early as in MCI (Petersen et al., 1999) and mild AD (Balthazar, Martinelle, Cendes, & Damasceno, 2007). Possible discrepancies in the reported stage of occurrence of naming disorders might be partly explained by the different methods and diagnostic threshold used in different studies as well as the occurrence of ceiling effect in numerous short naming tasks.

3.1.1 Ceiling effects

Thanks to their rapid and simple administration as well as their high sensitivity to language impairments, picture naming tasks are widely used in clinical settings. A myriad of naming tasks exists in different languages : the *Boston Naming task (BNT)* including 60 object line drawings (Kaplan, Goodglass, & Weintraub, 1983), the *Dénomination Orale d'Images (DO80)* including 80 line drawings (Deloche, Hannequin, et al, 1997), the *Philadelphia Naming Test (PNT)* including 175 line drawings (Roach et al., 1996), the naming subtests of well-known batteries like the *Aachener Aphasie Test (AAT)* including 10 line drawings of objects, actions and colours (Walter, Poeck, Weniger, & Willmes, 1983), or from the *Western Aphasia Battery* including 12 items of colours, tools and actions (Kertesz, 1982).

Subtests from larger language batteries, however, are mostly very short and simple. They are, therefore, not sensitive enough for detecting early cognitive decline. Healthy participants are also often expected to perform close to ceiling in longer naming tests involving a broader range of difficulty, like the aforementioned BNT, PNT or DO80. Most existing tests seem inappropriate for early detection of cognitive decline especially in population with a higher level of education (Deloche et al., 1996; Katsumata et al., 2015; Roach, Schwartz, Martin, Grewal, & Brecher, 1996). Moreover, most of them were developed for the evaluation of naming abilities in sudden brain injuries caused by stroke or head trauma where simple tests are often believed to be sufficient to show impairment, and not specifically for tracking subtle cognitive decline.

To tackle the ceiling effects seen in most picture naming tests the *Graded Naming Test* (GNT) was developed (McKenna & Warrington, 1983). It is made of 30 line drawings whose namability ranged from 100% to around 15 % in HC. A longitudinal study including AD patients and participants with SCI found significant changes in GNT performance after 8 months in AD but not in SCI (Chamberlain et al., 2011). The GNT seem to be more adapted to track progression of advanced disease than early cognitive changes. The amount of conversion from SCI to AD was, however, low in this lapse of time, mitigating this hypothesis. Indeed, the SCI sample might have included many people who will not convert to dementia. A study associating the performance obtained in the GNT and in the Cambridge Neuropsychological Test Automated Battery (CANTAB) Paired Associated Learning (PAL) revealed, however, a high predictive value of the combined two tests for the diagnosis of probable AD (Blackwell et al., 2004). Furthermore, in an analysis of the contribution of the different subtests of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) test battery (Moms et al., 1989), confrontation naming was found to be the only subtest in the non-memory domain that would increase the discriminative power between healthy elderly and mild AD (Welsh, Butters, Hughes, Mohs, & Heyman, 1992). These results highlight the crucial importance of including naming tests in batteries aiming at detecting early cognitive impairment in neurodegenerative diseases.

3.1.2 International value

Language heterogeneity is a recurring problem in international therapeutic trials where it is desirable to use comparable outcomes. Picture denomination tasks are typically developed and validated in a single language. They are later translated or adapted in further languages

either in their full version or by choosing a subset of items from the original pool. A simple translation does, however, not always yield a similar level of difficulty in different languages. This leads to biased assessment preventing direct comparison (Puente & Puente, 2009).

The BNT (Kaplan, Goodglass, & Weintraub, 1983) is one of the most used confrontation naming tests, not only in the English speaking world for which it was developed, but also in its internationally adapted versions. It was translated in numerous languages and shortened with the aim of reducing time cost while keeping a high sensitivity to cognitive decline (Katsumata et al., 2015). In a sample of monolingual Spanish and English HC and patients, comparing performance in the original version of the BNT and its Spanish translation, the authors found that 27 out of 60 items displayed Differential Item Functioning (DIF) (Jahn et al., 2013). This indicates that for almost half of the items, at an equal level of naming ability, participants of the two language samples would obtain different scores in the BNT. These results underline the weakness of translating a naming task that was constructed in another language.

In Germany, the AAT (Huber, Poeck, Weniger & Willmes, 1983) is the gold standard for language assessment. It was also adapted to a few other languages like English (Miller, Willmes, & De Bleser, 2010), Thai (Pracharitpukdee, Phanthumchinda, Huber, & Willmes, 1998) or Italian (De Bleser et al., 1986). The battery is, however, time consuming and directed at brain trauma or stroke related aphasia, where language impairment is sudden, and can recover in the first months after onset. It is therefore not suitable in the context of neurodegenerative diseases.

3.2 Aim

Cognitive profiles in neurodegenerative diseases being very heterogeneous, batteries of neuropsychological tests tend to be more sensitive to early cognitive decline than single test. Therefore, we attempted to create a Graded Object-Naming Task, as part of a larger test battery, that would be short and easy to administer. It should track early changes in naming ability in a single subject over time and therefore be devoid of floor and particularly ceiling effects. Moreover, it should be usable in international settings and serve as an outcome measure in therapeutic trials.

In chapter 3.3 the creation of a preliminary item set for the Graded Object-Naming Task (GONT), its administration to a sample of HC German elderly as well as the reduction of its size are presented. In chapter 3.4 the internationality of the created item set is investigated by comparing the performance obtained by German and Slovak healthy elderly. Chapter 3.5 investigates the reproducibility of the results by comparing the performance obtained by two samples of German healthy elderly in the GONT. Chapter 3.6 presents the final item choice as well as the implementation of a stop and return criteria for administration. Chapter 3.7 explores the test validity and compares the performance obtained cross-sectionally on the GONT and Boston Naming Task (BNT). Finally, the longitudinal data obtained on the final version of the GONT and the BNT are presented in chapter 3.8

3.3 Test construction and piloting

In this chapter, the creation of an item set for the international Graded Object-Naming Task (GONT) as well as its preliminary piloting on German healthy elderly (HC) are presented.

3.3.1 Method

3.3.1.1 Item set preparation

The GONT is based on original photographs of diverse objects found in households, zoos, music instrument shops and a technical museum. Pictures were taken in full colours in a canonical view to increase ecological validity compared to line drawings or cartoon like pictures. A first set of 209 photographs was compiled to cover a range of difficulty going from very easy to very hard to name for HC of different age and education level.

Several objects might have multiple names used by local population. The highest frequency word used by the majority of the people might not be the most correct word to qualify the object. Therefore, when setting the valid answer, we intentionally chose the most correct but not obviously the most used name for an object.

To overcome the known limitations existing in international comparisons of naming performance, exemplars were deliberately chosen to be as free of cultural references as possible. Therefore, items that are highly culturally or regionally biased, such as food and plants, were avoided. Certain categories of animal were, however, included as they were assumed to be known from most target population through zoo visits and television programs.

3.3.1.2 Participants

From July 2013 to February 2014, the 209 compiled photographs were administered to 54 locally recruited HC (see section 2.3 for further information).

3.3.1.3 Administration

In the same session, participants were administered a German adaptation of the Mini-Mental State Examination (MMSE) (Folstein et al., 1975) as well as the newly created GONT. Pictures were presented sequentially in the centre of a computer screen using the software Psychopy (Peirce, 2007). The participant was asked to name precisely the depicted object. When he/she provided a superordinate (e.g. „Monkey" for "Capuchin monkey"), he/she was asked to provide a more precise answer. When the participant provided a name

that is commonly used for the object, but that is the not the most correct name, he/she was asked to provide another answer. In German, an “Hourglass” can also be called an “Egg clock”, for example, in this case the participant was asked to think of another name and the answer was recorded as correct when he/she came up with “Hourglass”. The examiner went on to the next picture after the participant answered or said he/she could not name the presented object. At this stage, each given answer was recorded and later compared to the expected answer. Screening the recorded answers allowed identifying ambiguous pictures that yield visual mistakes. An answer was only considered correct when matching the determined valid answer. To simplify quotation and reduce possible inter-rater variability we recorded superordinate or synonyms as incorrect.

3.3.1.4 Statistical considerations

Item responses were dichotomous, with one representing success and zero failure to name the object correctly. The percentage of namability for each item was calculated with the following formula: *((number of HC who could name the item /total number of HC)*100).* The first item set reduction was done qualitatively by removing items subject to visual mistakes or of poorer photographic quality. The Rasch dichotomous model was applied to the remaining data with the help of the software R in version 3.1.1 (R core Team, 2014) with the eRm package in version 0.15-5 (Mair, Hatzinger, & Maier., 2015). The model allows calculating the probability of each participant to be successful on an item given his ability level and the item difficulty. The summed Item Information Plot that displays the amount of information provided by the test as a whole given a certain ability level was computed. Descriptive data were analysed using SPSS version 21.0 (IBM corp., Armonk, NY). The item response being dichotomous, the reliability coefficient was calculated using the Kuder-Richardson-20 test. Normality was assessed with the Shapiro-Wilk test adapted for small sample size and the homogeneity of the variances was assessed using Levene´s test. Given the normality of the data, and the equality of variances a 1-way ANOVA was run to compare the total score on the final item set between male and female.

3.3.2 Results

3.3.2.1 Demographics

Four participants had to be excluded because of an MMSE score below 27. The age, education level, gender and MMSE scores of the 50 remaining participants are displayed in table 3.1. The HC sample covered a wide range of education level and age at assessment. Women represented 2/3 of the sample.

Table 3-1 Demographics for the included healthy elders

Abbreviations: F = female, M = male.

	Healthy controls **Mean (SD)** ***[range]***
Education level (years)	13.7 (2.3) *[10-18]*
Age at assessment (years)	67.8 (6) *[50-82]*
Gender	34 F; 16 M
MMSE (Max. 30)	29 (0.85) *[27-30]*

3.3.2.2 Qualitative item set reduction

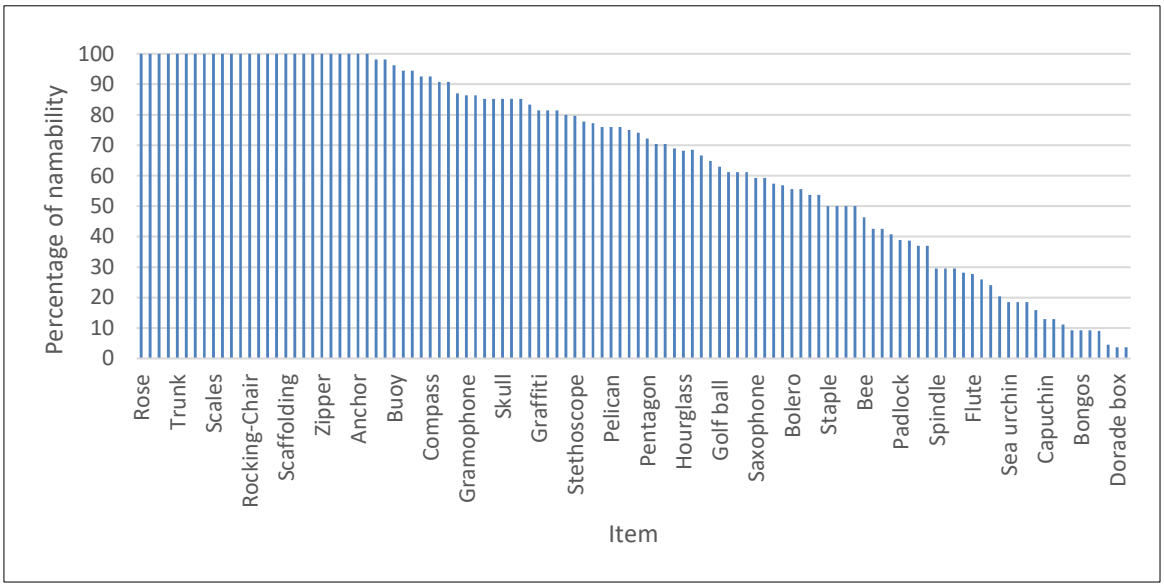
After piloting, items were qualitatively excluded: 4 items could not be named by anyone (for example “trilobite”), 9 items had ambiguous names in German (for example “folding rule”) and 26 items gave raise to unexpected perceptual mistakes or had lower quality (for example: a balcony was often taken for a letterbox.)

3.3.2.3 Namability

After exclusion, 170 items remained. Based on the results obtained in the elderly sample, the percentage of namability for each remaining items was calculated. Fifty-nine items that were redundant (same namability level) were further excluded. Hundred eleven pictures covering the whole range of difficulty in HC (100 to 2.5 % of namability) were selected (see figure 3.1). Twenty-six items could be named by 100 % of HC. Although, these items had no variance and did not provide any information in healthy elders they were included in the item set. Indeed, they were assumed to show a range of difficulty in anomic patients suffering, for example, from AD or Primary Progressive Aphasia.

Figure 3-1 Difficulty range of the 111 selected items

*Namability was computed with the following formula: ((number of HC who could name the item /total number of HC)*100).*



3.3.2.4 Face validity

Naming is a particularly narrow concept. Many confrontation naming tasks are based on the same procedure. The GONT of the DZNE-Cog seemed to stay in line and display good face validity. This was confirmed in chapter 3.7 of this thesis by comparing the results to another confrontation-naming task.

3.3.2.5 Internal consistency

The 111 remaining items displayed a very high reliability coefficient $\alpha = 0.999$. Using the split half method, the reliability coefficients were still very high $\alpha_a = 0.997$ and $\alpha_b = 0.997$. Item-total correlation ranged from acceptable 0.633 to very good 0.999. These high indicators are not surprising at this stage of the test development, considering the large amount of items and the narrowness of the concept of naming.

3.3.2.6 Rasch analysis

Twenty-six items were removed from the Rasch analysis because they had no variance. Reasons for keeping them in the final set are discussed in chapter 3.7.

Results of the Andersen LR-test confirmed that the data fit the Rasch model, predicting that the performance obtained on the item set reflects the underlying naming ability of a subject (LR value: 63.5, Chi-square df: 75, p-value = 0.8). The item-specific Wald-test was non-significant for all items. This indicates that there was no differential item functioning in the

item set: no sub-group in the sample would be advantaged, because of the nature of the item, given that they have the same level of abilities.

The plotted joint Item Characteristic Curve (figure 3.2) pictures the fact that the chosen items cover a full range of difficulty, particularly in the highest level of naming ability. This is also confirmed by the Information Function plot (figure 3.3), which highlights that the GONT can capture performance at best in the highest level of the latent trait, namely the naming ability. The underrepresentation of items that cover the lowest level is explainable by the fact that 26 items named by all controls had to be removed from the analyses because of a lack of variance. Their ability to distinguish between patients, those participants with the lowest naming abilities (the latent trait) is presented in chapter 3.7.

3.3.2.7 Gender comparison

Considering the high proportion of female in our sample, we compared the performance between genders, to discard a possible bias. When summing the total amount of correct answer on the final data set, Levene´s test ($p>0.7$) and the Shapiro-Wilk test ($p>0.5$) confirmed respectively that the distribution among gender was normal with equality of the variance. No significant difference was found between males and females' total score ($F=2$; $p=.016$) confirming the results of the item-specific Wald-test mentioned in section 3.3.2.6.

Figure 3-2 Item Characteristic Curve plot for the 111 selected items

Each coloured line represents the probability to name correctly an item given a certain level of naming ability. The colour legend was left out for legibility purpose. The figure shows that the item set is able to discriminate at best between participants with higher ability (from 0 onwards on the x-axis). Indeed, at this level items have a wide range of probability to be named correctly.

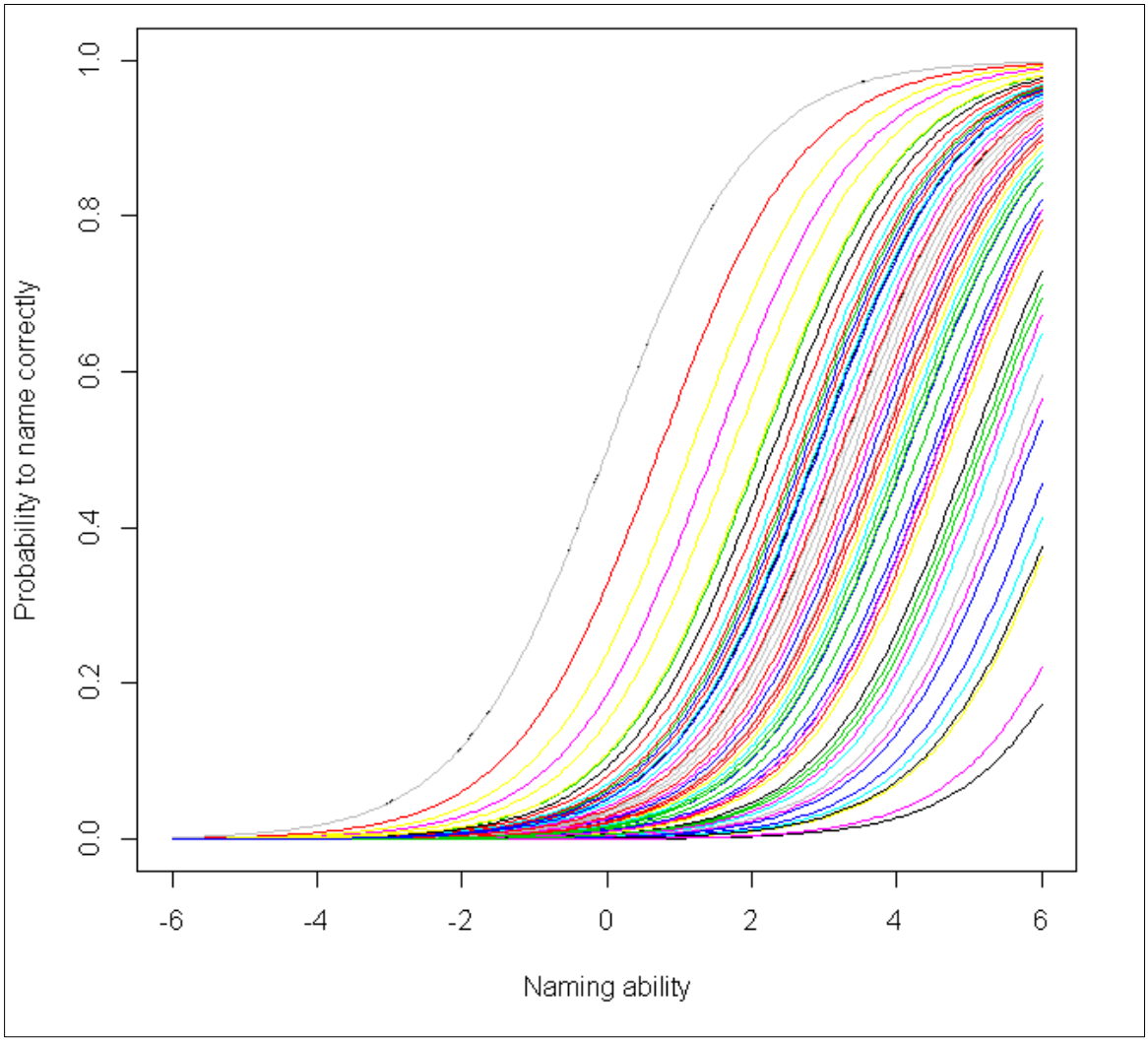
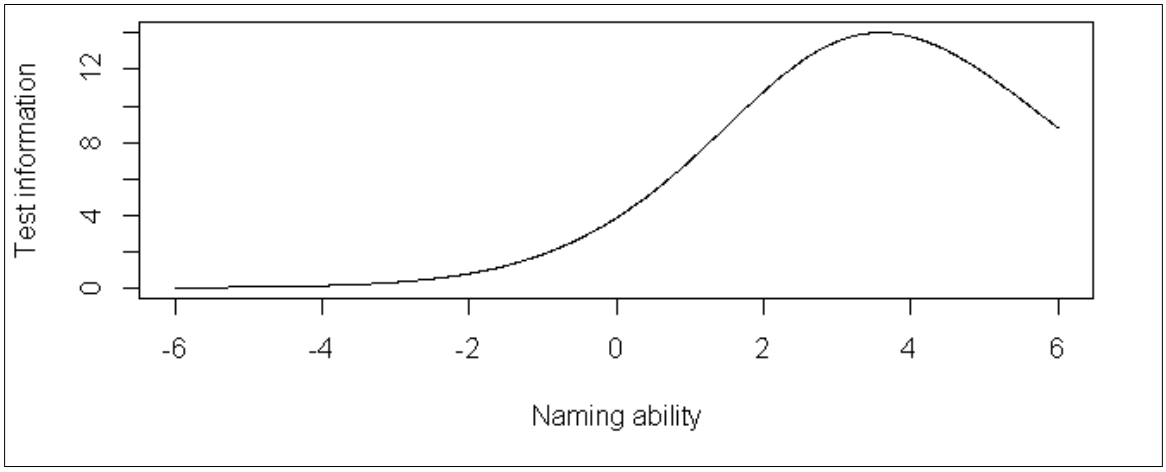


Figure 3-3 Test Information Function plot for the 111 selected items

The black line represents the level of information obtained with the admnistration of the item set for a given naming ability. The figure confirms that in the chosen item set, the maximum information is obtained in the highest level of naming ability (from 0 to 6).The lack of information in the lowest level of ability is explained by the exlusion of the easiest items from this analysis.



3.3.3 Discussion

The piloting study revealed that a set of 111 items could cover the full range of difficulty in German HC of various education levels with a high level of reliability. The classical ceiling effect in HC observed in many confrontation naming tasks was avoided in our item set, indeed a maximum information level was reached in the higher abilities level. Further analysis with patients is presented in chapter 3.7 and confirms that the test can cover lower level of abilities and avoid floor effect.

The item set was deliberately kept large at this stage of the study. It was finalized based on the results obtained in the cross-sectional analysis as well as after comparing the results in HC in German and a further language (see chapter 3.4). Items that do not work in another language were excluded.

An answer was only considered correct when matching the determined valid answer. This implied rejecting synonyms, super-ordinates and words that are used colloquially even if they are the most common names for an object. This strict quotation criterion possibly undermined participant´s performance. It was, however, chosen to improve inter-rater reliability. The final aim being the comparison of performance on the single subject level overtime, we assume it would not be a problem to record the answer as wrong if a person gave a similar, but not expected answer. Indeed, in the absence of disease, we expected a

participant to be consistent and give the same answer at each assessment. Score variations would, therefore, most likely be linked to a change in ability and not the quotation method.

As mentioned, our sample was composed of 2/3 of women. Studies have sometimes shown advantages of males over females in confrontation-naming task like the BNT for example (Hall, Vo, Johnson, Wiechmann, & O'Bryant, 2012; Lansing, Ivnik, Munro Cullum, & Randolph, 1999). While these discrepancies were possibly caused by the presence of gender biased items in the BNT, the results seem to show that our item set avoids this pitfall. Indeed, no significant differences between the performance of males and females on our preliminary reduced item set could be found. Moreover, the test aiming at showing change in a single subject the existence of a bias would be of lesser importance. Further steps of the study would gain, however, at having a similar proportion of men and women.

To summarize, it seems that 111 items can cover a wide range of difficulty in healthy German elders. Further analysis presented in this work assessed the internationality of the items as well as their reproducibility in German healthy controls. Moreover, a shortening of the item set occurred to reduce administration time. Indeed, the final item set should be based on items that are international, reliable and able to track change over time.

3.4 International comparison

As mentioned earlier in this work, we aimed at creating a Graded Object-Naming Task (GONT) that could be used in international settings and yield comparable results regardless the language in which it is administered. This would be of considerable interest, particularly for international therapeutic trials where cognitive outcome measures need to be comparable across study sites. Most existing confrontation naming tasks were developed in one language and later translated in further languages. Direct translation is, however, often impossible or subject to cultural or regional bias, making the different versions difficult to compare (Puente & Puente, 2009). In creating the GONT, we intended specifically to choose items that would be devoid of cultural bias and would give comparable results in different languages, and by extension countries.

To assess the applicability of the test in another language and culture, the previously reduced item set was administered to Slovak healthy elderly and compared to a demographically similar sample of German healthy elderly (HC). The aim was to identify items that would, in spite of the early international conception, not work outside of Germany and obtain a feedback for the internationalisation of the test.

3.4.1 Method

3.4.1.1 Slovak sample

Sixty-three Slovak HC were administered the item set reduced to 111 pictures. Participants were over 50, had no major neurological or psychiatric disorder, and were native Slovak speakers. Visual acuity and hearing abilities were normal or corrected to normal. During the same session the MOntreal Cognitive Assessment (MOCA) (Nasreddine et al., 2005) was administered to all participants. Twenty-one participants had to be excluded because of a MOCA score inferior to 26.

3.4.1.2 German sample

Twenty-seven German HC were administered the reduced item set (see section 2.3 for inclusion criteria). In the same session the German versions of the MOCA (Nasreddine et al., 2005) and the MMSE (Folstein et al., 1975) were administered.

3.4.1.3 Administration

The chosen 111 pictures were ranked in order of difficulty, using the percentage of namability obtained from the piloting study (see chapter 3.3). The test was implemented in

the Psychopy software (Peirce, 2007) and pictures were presented sequentially on a computer screen. Participants were given the following instructions:

„ In this test you will be shown some pictures of objects. Your task is to name the object as precisely as possible”

He/she was shown the training item representing a spoon and fork, with an arrow pointing at the spoon (see figure 3.4). If the participant said “cutlery”, he or she was asked to be more precise. In the case where a participant gave a wrong answer, the examiner provided the correct one. When the participant produced the correct answer, he/she was given a positive feedback and in absence of further question, the test started.

Figure 3-4 Training item in the Graded Object Naming Task

This example was used to draw the participant´s attention on the fact that in presence of an arrow, the part of the object it was pointing at needed to be named

During the test if the participant gave two answers (e.g. “It is a camel or a dromedary” for “camel”) the correct answer was recorded. If the participant provided a superordinate, for instance if he/she said “bird” for the target item “pelican”, the examiner asked him to be more precise. The subject’s response was only recorded as correct if he produced “pelican”. The same applies for synonyms. For instance, the test contains a picture of a “weather vane”; in English, this item can also be called a “weather cock”. If the subject answered “weather cock”, again, he/she would be prompted to produce another name, and the answer would only be scored as correct if they then said “weather vane”.

At this stage, no time limit for giving the answer was set up. The examiner moved onto the next picture when the participant gave a correct answer or said “I do not know”.

3.4.1.4 Statistical considerations

Descriptive statistics were run using SPSS version 21 (IBM corp., Armonk, NY). The item responses were dichotomous, with one representing success and zero failure to name the object correctly. The total score on the 111 items was computed by summing the amount of correct answers. The normality of the distribution for the demographics was assessed with the Shapiro-Wilk test adapted for small sample size and the homogeneity of the variances was assessed using Levene´s test. Given the non-normality of the data, a Mann-Whitney test was run to compare the demographics between groups.

The Shapiro-Wilk test was used to assess the normality of the distribution for the percentage of correct answers obtained on the test in both the German and Slovak samples. Given the normality of the data a one-way-Anova was run to compare the two groups.

The two samples were screened for Differential Item Functioning (DIF). It identifies items that would yield different results for participants of the different language groups although they have the same latent level of naming ability. DIF was analysed using the Angoff´s Delta Plot, also known as transformed item difficulties approach, following the method described in 2014 by Magis & Facon. The analysis was run in the software R version 3.1.1 (R Core Team, 2014) with the deltaPlotR package version 1.5 (Magis & Facon, 2014). This method is particularly adapted for small sample size. No item purification process (IPP) was used to reduce the impact of items that are flagged DIF on the total test score. Indeed, it was shown that in small samples they do not provide more information than the standard approximation of the threshold (Magis & Facon, 2013).

3.4.2 Results

3.4.2.1 Demographics

The descriptive statistics for the German and Slovak samples are reported in table 3.2. No significant difference was found between groups for education level, age or MOCA score. A larger proportion of female was present in both samples.

Table 3-2 Demographics of the German and Slovak healthy controls

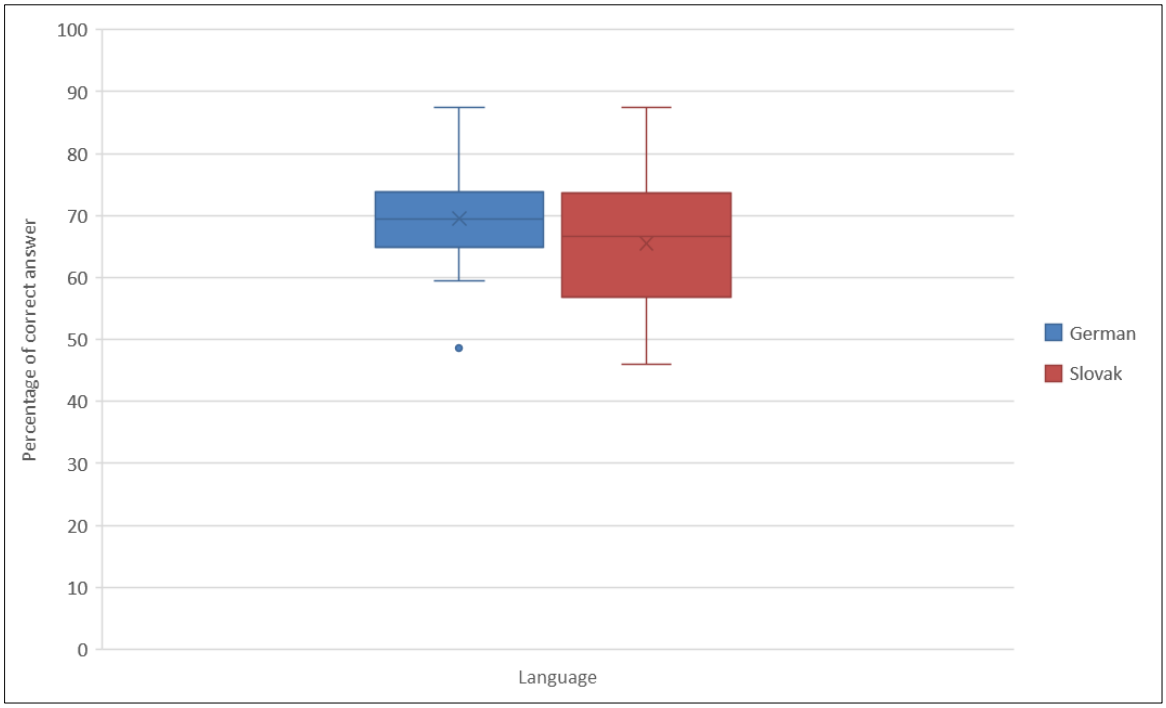
Abbreviations: n.a = non-available, n.s = non-significant, F = female, M = male.

	German Sample Mean (SD)	**Slovak Sample Mean (SD)**	**Significance *P-value***
Education level (years)	15.2 (1,8)	14.1 (3.5)	*n.s*
Age at assessment (years)	60.9 (6.8)	61.6 (8.6)	*n.s*
Gender	21 F; 6 M	27 F; 15 M	*n.s*
MOCA (Max. 30)	28 (1.2)	28.4 (1.5)	*n.s*
MMSE (Max. 30)	28.7 (1.3)	n.a	*n.a*

3.4.2.2 Performance

The percentage of correct answer for both German and the Slovak elders are displayed on figure 3.5. German elderly controls reached on average 69.6% of correct answer (SD: 8.4) while Slovak elderly controls had on average 65.5 % of correct answer (SD: 9.8). No significant difference between the two samples was found on the total score in the GONT (F: 3.1; p=0.08).

Figure 3-5 Comparison of the total score on the Graded Object Naming Task in two different language groups

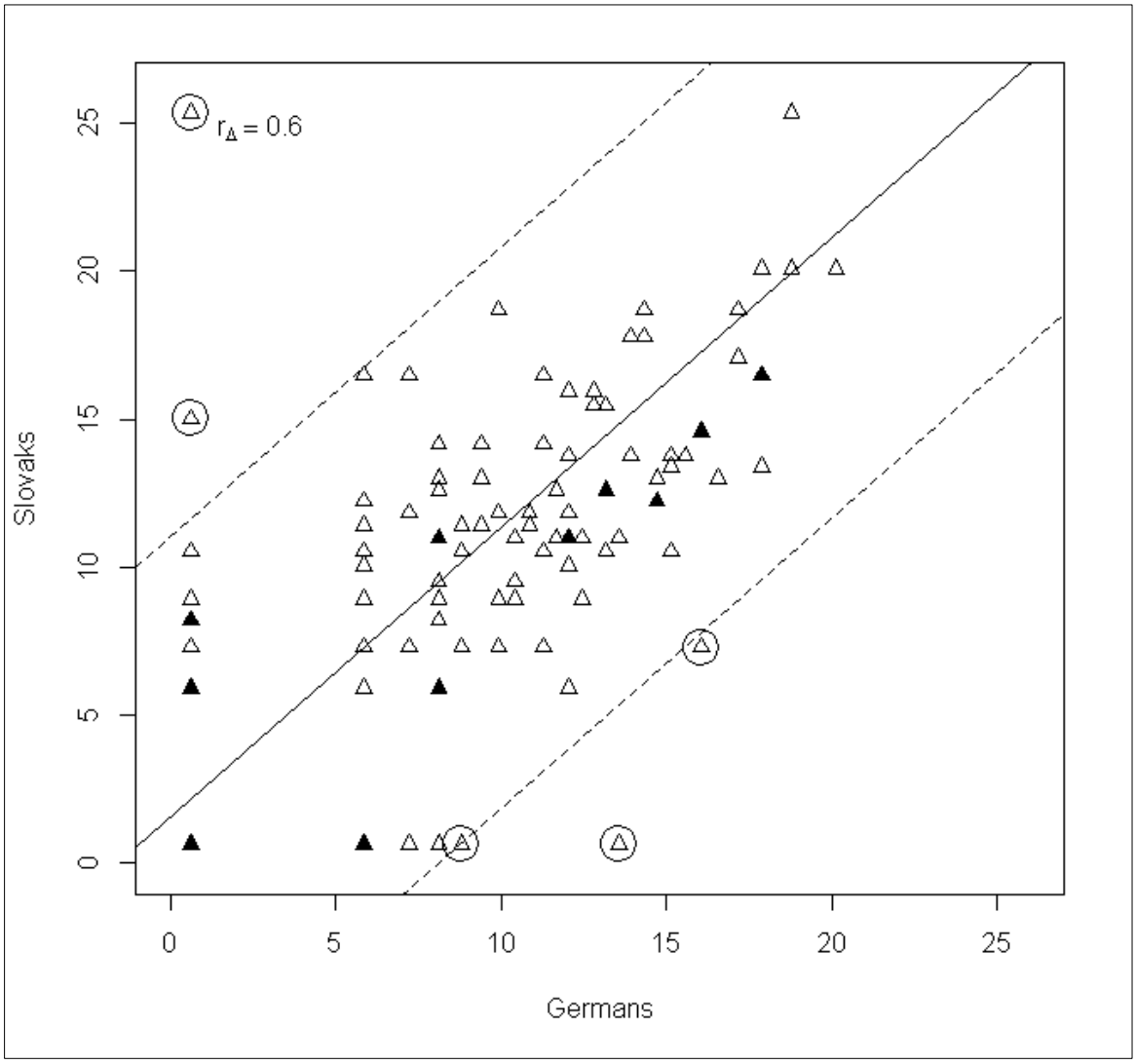


3.4.2.3 Differential Item Functioning

To avoid including items that would not work in an international context the data pool was screened for DIF. The analysis pointed out five of 111 items (4.5%) that displayed DIF: spoke, cannon, camel, chisel and warthog. These items would yield different performance in participants of the different language groups that have, however, the same level of abilities. The diagonal plot that illustrates the results of the DIF analysis is displayed in figure 3.6. The correlation between the Delta scores was of 0.6.

Figure 3-6 Diagonal plot illustrating the results of the DIF analysis among German and Slovak healthy elderly

Each triangle represents the delta score of an item. Items that are outside of the detection threshold, denoted by the dotted line are flagged with DIF. These are encircled for a quicker identification. Full black triangles are drawn onto existing Delta scores wherein multiple items are located.



3.4.2.4 Internal consistency reliability

After exclusion of the items flagged as DIF, the test displayed high reliability coefficients in the Slovak $\alpha = 0.88$ and German samples $\alpha = 0.84$.

3.4.3 Discussion

The GONT as part of the DZNE-Cog, was developed to be applicable in an international context. The results seem to indicate that the GONT works similarly in different languages. Indeed, after analysis only five out 111 items displayed DIF. DIF is in these items explained by the fact that in one of the two languages, participants systematically used a more common but less exact word. Therefore, the flagged items were removed from the item pool. Noteworthy, is that the correlation between the Delta scores, on which the DIF analysis is based, was only of 0.6. This could have undermined the detection of further items with DIF.

The conception of the preliminary item set was discussed with German, French, Spanish and English native speakers to avoid exemplars that were culturally biased. Five objects did, however, show high differences in namability in two European languages. It underlines both the effect of the language and cultural background on the namability level of particular objects. Testing in further languages might reveal that other exemplars are not nameable or in very different proportions in a particular language. Although this seems inevitable, a further version of the test might profit from an even broader international cooperation in the early conception of an item set. Moreover, further work will have to extend the comparison of the performance on the GONT in other languages. It would be especially interesting to compare the results obtained in languages other than Indo-European like Japanese or Chinese. Indeed, a higher similarity in performance would be expected in languages sharing the same origins than in those from different language families.

A possible weakness of our study is that most participants originated from the same region. Regional dialects being numerous in Germany, but also in other countries, it might be interesting to have a sample that is more representative of the national population. Further studies, with larger sample size could aim at including participants nationally spread to account for regional differences.

3.5 Comparing two German samples

Previous work on the Graded Object Naming Task (GONT) confirmed its ability to display graded difficulty in healthy elderly and its applicability in two different languages.
The aim of the GONT, as part of a broader cognitive battery, being to track change at the single subject level, it implies having a task that is reproducible; hence, that yields consistent results in similar populations. In this chapter, the reproducibility of the GONT was assessed by comparing the results obtained in two similar German samples.

3.5.1 Method

3.5.1.1 Participants

The first and second samples of German elderly are described respectively in chapters 3.3 and 3.4.

3.5.1.2 Statistical considerations

Descriptive statistics were run using SPSS version 21 (IBM corp., Armonk, NY). The total score on the GONT was based on the sum of the correctly named items, the maximal score being 111. Normality was assessed with the Shapiro-Wilk test adapted for small sample size and the homogeneity of the variances was assessed using Levene´s test. The demographics being non-normally distributed, a Mann-Whitney test was run to compare the two German groups. The percentage of correct answers being normally distributed a one-way ANOVA was run to compare the groups.

Differential Item Functioning (DIF) using the Angoff´s Delta Plot was run using the software R version 3.1.1 (R Core Team, 2014) with the deltaPlotR package version 1.5 (Magis & Facon, 2014) (For further information see section 3.4.1.4).

3.5.2 Results

3.5.2.1 Demographics

Descriptive statistics for the two German samples are reported in table 3.3. No significant difference was found in the MMSE score. The second German sample had, however, a significantly higher education level and was significantly younger. A larger proportion of female was present in both samples.

Table 3-3 Demographics and neuropsychology in two German HC samples

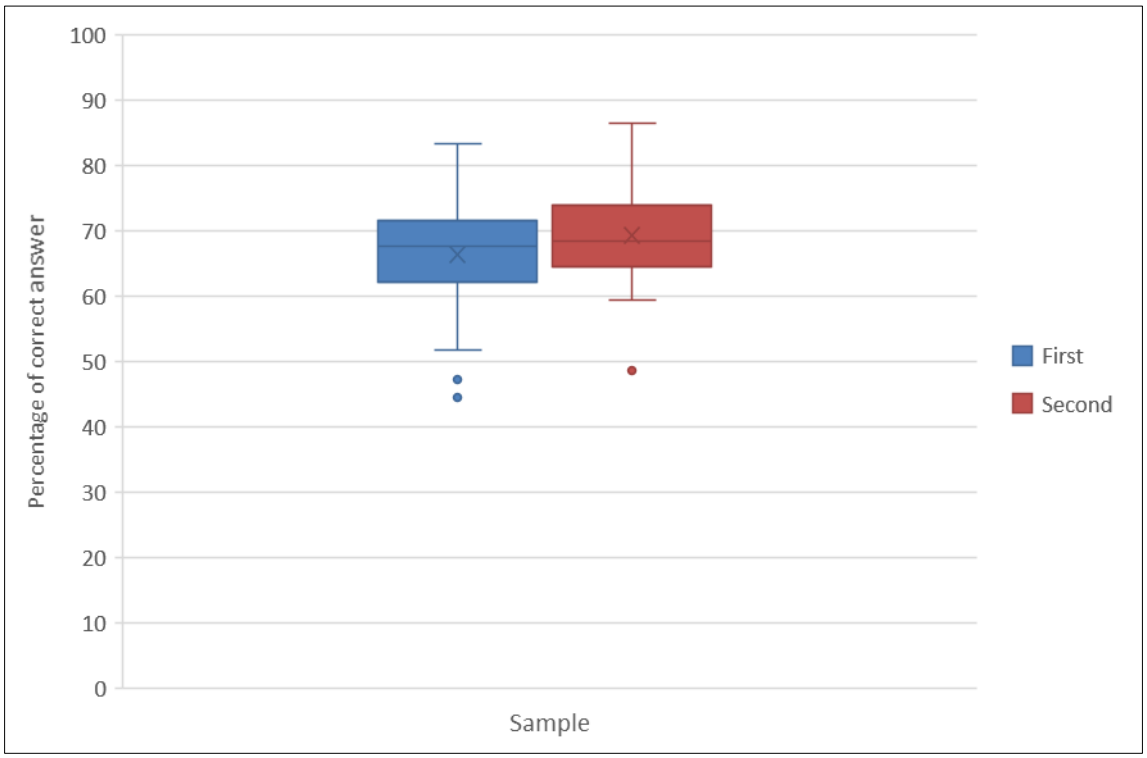
Abbreviations: n.a = non-available, n.s = non-significant, F = female, M = male.

	First German sample Mean (SD)	**Second German sample Mean (SD)**	**Significance *P-value***
Education level (years)	13.7 (2.3)	15 (1.9)	U (1, 77) =0.46, *p= 0.02*
Age at assessment (years)	67.8 (6)	60.9 (6.8)	U (1, 77) =0 34, *p <0.001*
Gender	34 F; 16 M	21 F; 6 M	*n.s*
MOCA (Max. 30)	n.a	28 (1.2)	n.a
MMSE (Max. 30)	29 (.85)	28.7 (1.3)	U (1, 77) = 615.5, *p = 0.5*

3.5.2.2 Performance

The percentage of correct answer on the task for both German HC samples is displayed on figure 3.7. Participants reached on average 66.5% of correct answers (SD: 8) in the first sample and 69.3% (SD: 8.2) in the second sample. As observable on figure 3.7 the difference did not reach significance (F: 2.1; p=0.16).

Figure 3-7 Comparison of the total score on the Graded Object Naming Task in two German healthy elderly samples

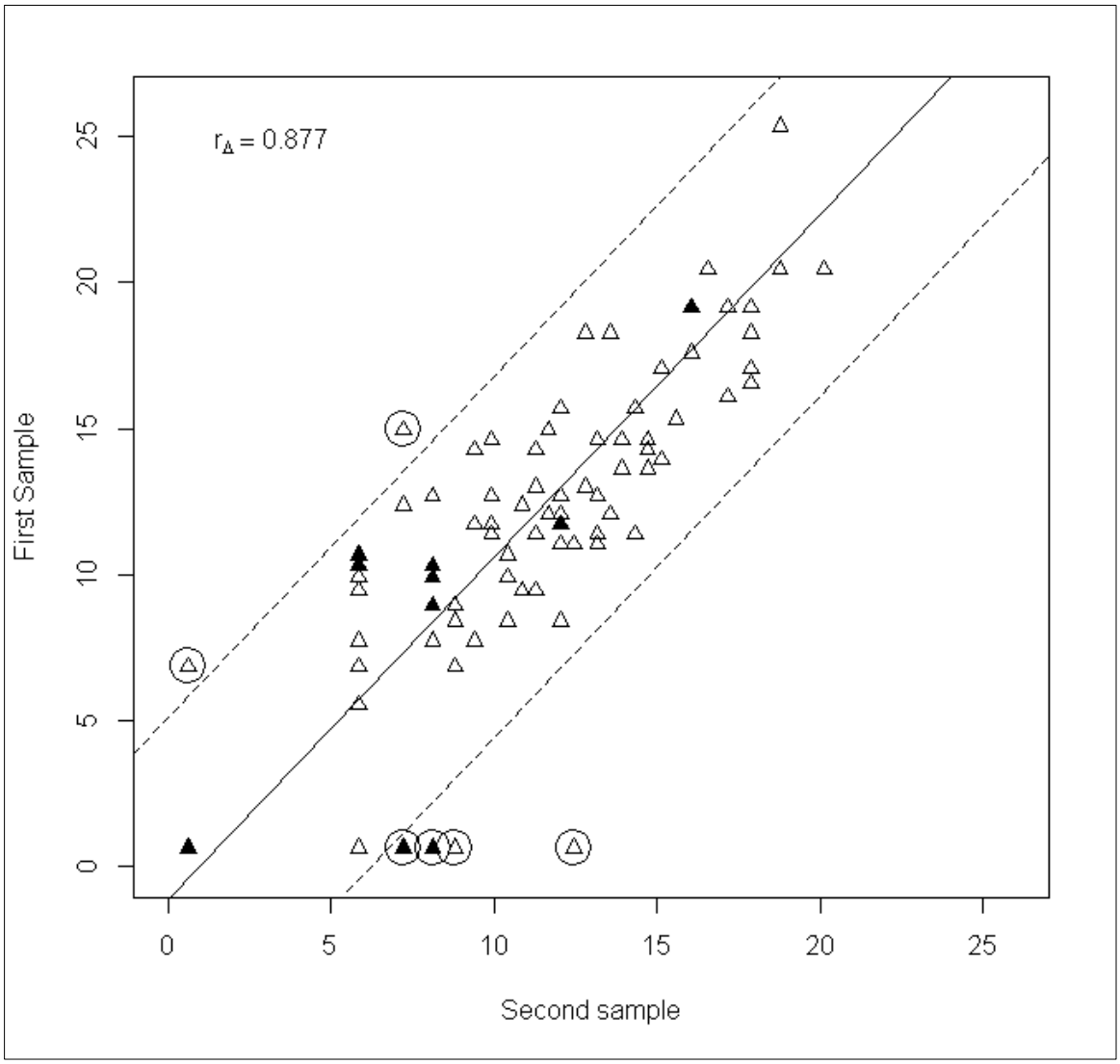


3.5.2.3 Differential Item Functioning

The performance in the item set in two comparable German samples was screened for DIF. The analysis supporting no missing data, it was run on a subset of 32 HC elderly from the first German sample. The analysis pointed out eight of 111 items (7.2%) that displayed DIF: saucer, violin, chimney, anchor, microscope, compass, green house and bee. These items would yield different performance in participants of the different samples that have, however, the same level of abilities. The diagonal plot that illustrates the results of the DIF analysis is displayed in figure 3.8. The correlation between the Delta scores was of 0.877.

Figure 3-8 Diagonal plot illustrating the results of the DIF analysis among two German samples

Each triangle represents the delta score of an item. Items that are outside of the detection threshold, denoted by the dotted line are flagged with DIF. These are encircled for a quicker identification. Full black triangles are drawn onto existing Delta scores wherein multiple items are located.



3.5.3 Discussion

Analyses ran in this chapter seem to indicate a high reproducibility of the results obtained in the GONT in two comparable samples of German healthy elders.

After analysis, eight items had to be excluded because they produced different results in the two German samples. Those differences can be explained by the fact that several items (saucer, violin, chimney, sole, green house and microscope) were named by all or almost all HC in the first sample, but not in the second sample. To the contrary, “compass” was named by 100% of the HC in the second sample but not in the first sample. Finally, in the

first sample bee was often mistaken for a wasp whereas this was not the case anymore in the second sample.

The slight differences obtained on specific items could be explained by a significant difference in the total education level in the two samples. Most items that were flagged by the DIF analysis, were, however, named by a smaller proportion of HC elderly in the second sample than in the first sample. The latter having a slightly lower education level and education level having a positive impact on naming abilities (Connor, Spiro III, Obler, & Albert, 2004; Welch, Doineau, Johnson, & King, 1996), this interpretation seems unlikely. The differences observed might be due to unavoidable sampling variations. A possible interpretation would be that the eight items that show DIF might be less stable and for this reason should not be included in the final item set.

Noteworthy, is that the two groups were administered item sets of different lengths. Indeed, the first sample had to name 209 pictures whereas the second sample was administered the shortened version composed of 111 items. The difference in the test length might have had an impact on the performance, biasing the analysis. Indeed, there seem to be a slight but non-significant advantage in the score for the second sample, which was administered a shorter item set. Naming being a robust ability, it is, however, unlikely that the item set length could have such an impact on the total score.

In summary, it seems that 98 items from the original item set have a high reproducible value in similar samples of healthy elders. This is of crucial importance when using the task longitudinally to track early changes on a single subject.

3.6 Final item set reduction and implementation of stop criterion

In previous chapters, the ability of the newly developed Graded Object-Naming Task (GONT) to cover a whole range of difficulty in healthy elderly was demonstrated. Based on the results obtained, the item set could be firstly reduced from 209 to 111 items. The international value of the remaining 111 items was confirmed by comparing a sample of German and Slovak healthy elderly. Following the analysis, only five items had to be excluded. These would yield different results in participants of different language groups, although they have comparable level of ability. Finally, the reproducibility of the results was endorsed by comparing the performance obtained in two samples of German healthy elderly. Only eight further items had to be removed because they would yield results that are not comparable across two similar samples. To summarize, 98 items showed high international value, reproducibility and covered a whole range of difficulty.

The GONT being part of a battery that should be quick to administer, and time being one of the major constraint in clinical routine, we aimed at maximizing the administration time. For this purpose, we propose to reduce further the item set to create three sets of difficulty that are administered in accordance with the participant´s level. Furthermore, we intended to implement a stop criterion to shorten administration time and reduce frustration in participant while preserving a high information quality. The results of these analyses are presented in this chapter.

3.6.1 Method

3.6.1.1 Participants

Forty-nine patients (15 AD, 19 MCI and 15 SCI) as well as 85 HC were administered the 98 items version of the GONT (see section 2.3 for inclusion criteria).

3.6.1.2 Final item set reduction

The administration of the 98 selected items took up to 8 minutes. In order to minimize administration time, we aimed at obtaining three sets of 25 items of easy, intermediate and hard levels with the best pictures of our item set. At this stage, a qualitative approach was adopted to reduce the item set one last time. Picture of lesser quality or that had an ambiguous angle were removed.

3.6.1.3 Implementation of a stop criterion

The percentage of namability for each item was calculated with the following formula: *((number of HC who could name the item /total number of HC)*100).* After rank ordering the items from very easy to very difficult, the probability of naming a harder item after making *n* number of mistakes was calculated in all participants. A stop criterion that could be used in clinical settings was determined based on these results.

The following formula was used:

Number of time a participant had n wrong answer in a row before getting a right answer/Number of time a participant had n number of wrong answer in a row.

3.6.1.4 Implementation of a return criterion

To save time in clinical settings, the easiest item of the intermediate difficulty item set would be the first administered item. When a participant manages to name a certain number of items in this set, the points from the easy item set would be automatically given, and the administration would go on until the stop criterion is reached. When a participant does not manage to name a certain number of items in the intermediate set, the easy item set would be administered until the stop criterion is reached, or until the 25 items have been administered. To determine the best return criterion, we computed the probability of making *n* number of mistakes in the first 12 items of the second item set in HC.

The following formula was used:

*(Number of time a HC had n wrong answer in the 12 first items of the intermediate item set/Number of HC)*100*

3.6.1.5 Internal consistency reliability

Using SPSS version 21 (IBM corp., Armonk, NY) Cronbach's Alpha for the shortened version of the GONT was computed.

3.6.2 Results

3.6.2.1 Participants

The demographics as well as the scores on the MMSE (Folstein et al., 1975) and a 30-item version of the BNT (Merten, 2004) are presented in table 3.4.

Table 3-4 Demographics and neuropsychology

Abbreviations: AD = Alzheimer´s disease, MCI= Mild Cognitive Impairment, SCI= Subjective Cognitive Impairment, HC = healthy control, F = female, M = male.

	AD Mean (SD)	**MCI Mean (SD)**	**SCI Mean (SD)**	**HC Mean (SD)**
Education level (years)	9 (1.7)	9.5 (1.7)	10.6 (1.6)	11.2 (1.9)
Age (years)	74.3 (6.6)	73.05 (6.6)	68.7 (8.4)	66.6 (7.8)
Gender	8 F, 7M	8 F, 11 M	8 F, 7 M	53 F,32 M
MMSE (max.30)	23.3 (3.3)	27.3 (1.6)	28.8 (1.1)	28.6 (1.3)
BNT (max.30)	21.5 (5.9)	25.1 (4.9)	27.9 (2.1)	27.4 (2.6)

3.6.2.2 Final item set reduction

Twenty-two items were removed because of their ambiguity. Pictures were considered ambiguous when the examiner systematically had to point at the part to be named, or the picture had an equivocal angle possibly biasing the namability. One item was removed because it had no variance, even in patients. The remaining 75 items were separated in three item sets of different difficulty level (see figures 3.9, 3.10 and 3.11). The easy set consists of 25 items that can be named by 100 to 90% of healthy elderly, whereas the intermediate set consists of 25 items that can be named by 90 to 59% of healthy elderly. Finally, the hardest item set comprised 25 items that can be named by 57 to 7% of healthy elderly. See appendix I for a list of all retained items.

Figure 3-9 Difficulty range for the easy item set of the Graded Object Naming Task

SCI being mostly a very heterogeneous group that can include patients with presymptomatic dementia but also healthy participants, patients included only AD and MCI. The namability is the percentage of HC or patients that could name the given item. The bars represent the standard error of the mean.

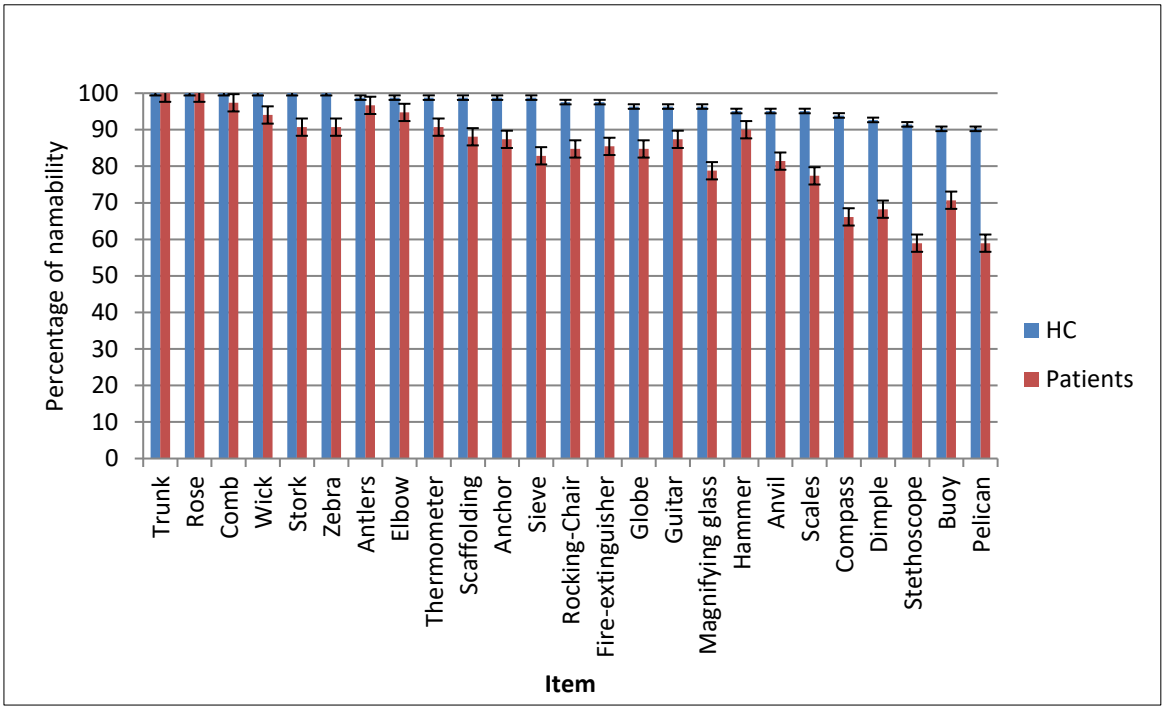


Figure 3-10 Difficulty range for the intermediate item set of the Graded Object Naming Task

SCI being mostly a very heterogeneous group that can include patients with presymptomatic dementia but also healthy participants, patients included only AD and MCI. The namability is the percentage of HC or patients that could name the given item.

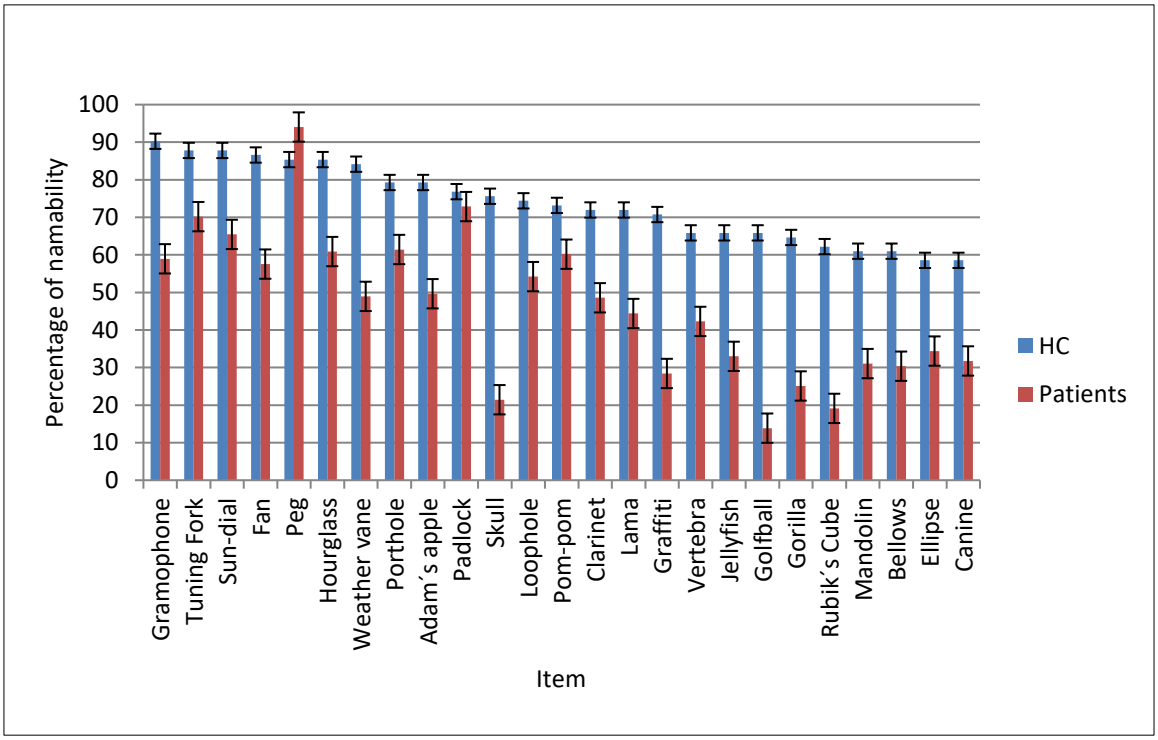
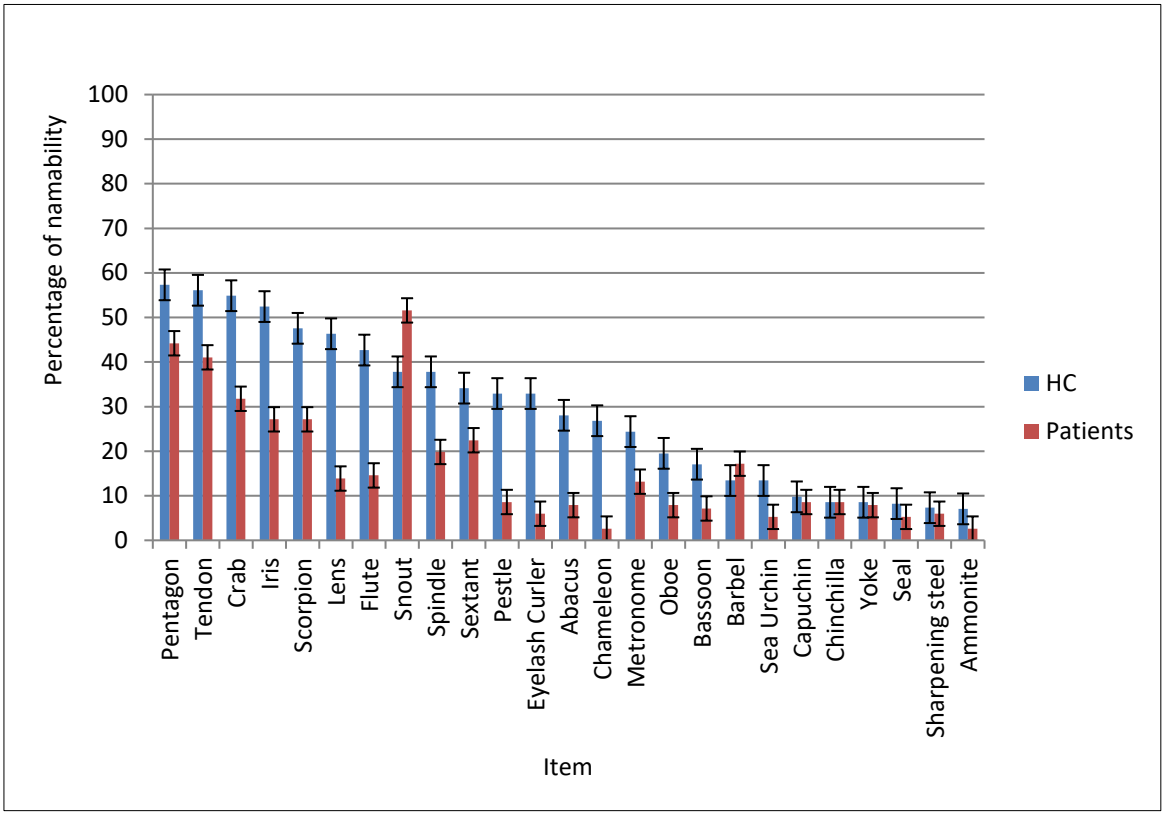


Figure 3-11 Difficulty range for the hard item set of the Graded Object Naming Task

SCI being mostly a very heterogeneous group that can include patients with presymptomatic dementia but also healthy participants, patients included only AD and MCI. The namability is the percentage of HC or patients that could name the given item. The bars represent the standard error of the mean.

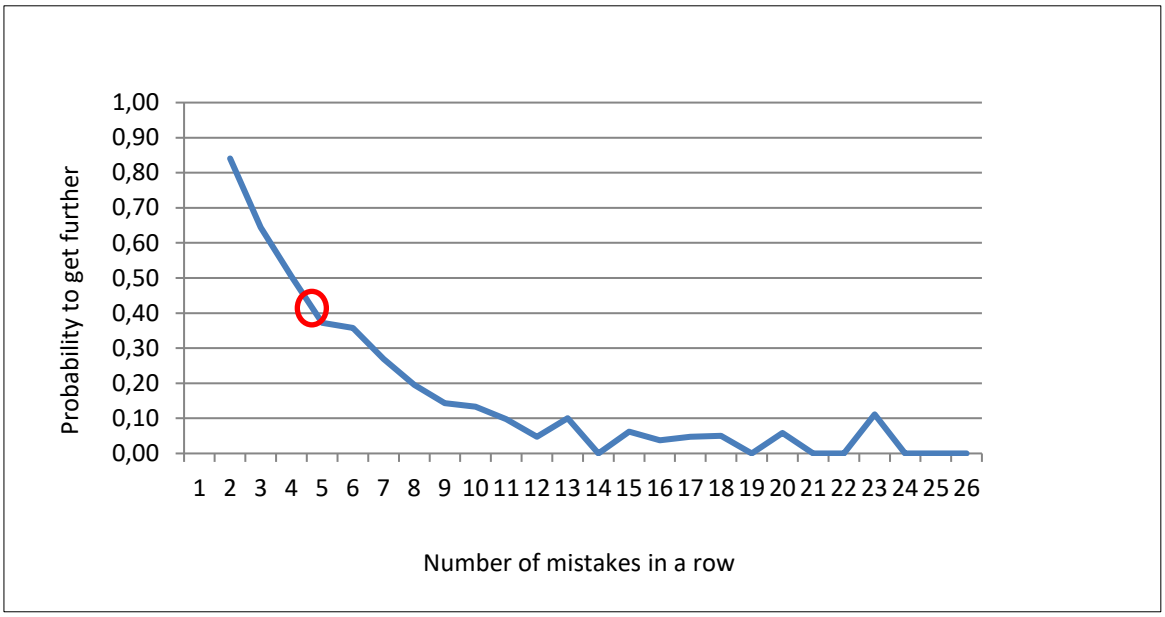


3.6.2.3 Stop criterion

The probability of naming correctly an item after making four mistakes in a row in the task was only of 0.41 (see figure 3.12). More than one out of two participants did not name any harder item correctly after making four mistakes in a row. For this reason we consider 4 mistakes as being a reasonable stop crieria for the GONT.

Figure 3-12 Justification of the stop criterion

The probability a participant has, to name correctly a harder item (y-axis) is represented in function of the number of mistakes he/she made in a row (x-axis). The red circle represents the chosen stop criterion: after four mistakes in a row, a participant has only a probability of 41% to name correctly a harder item than the last one he/she failed to name.



3.6.2.4 Return criterion

The results showed that 95% of HC make less than four mistakes on the first 12 items of the second item set (intermediate level). For this reason, we advise that if a participant makes four mistakes or more on the 12 first items administered, the first item set (easy level) is administered. Otherwise, the rest of the test would be administered, continuing with item 13 of the intermediate item set. All items of the easy item set would be considered as correctly named. The test would then end after 4 consecutive mistakes.

3.6.2.5 Internal consistency reliability

The shortened version of the test displayed a high reliability coefficients $\alpha = 0.92$.

3.6.3 Discussion

In clinical routine, time is one of the major constraints. For this reason, newly developed tests should always aim at giving the maximum information about the participant´s ability in a minimum of time. Therefore, we decided to shorten the test a last time, rank-order the items and create three sets of 25 items that would be of easy, intermediate and harder difficulty. The 75 selected items displayed a high reliability coefficient indicating a very good internal consistency for the final version of the task.

Analysis showed that 95% of healthy controls name a minimum of nine out of the 12 first items of the second item set. Moreover, in a sample of 134 participants (patients and healthy controls) the probability of getting a right answer after four consecutive mistakes is of 41%. Based on these results, a participant coming for the first time in the memory clinic would be given the easiest item of the intermediate item set. In the case where the participant makes four mistakes or more on the 12 first items of intermediate difficulty, the easy item set would be administered and the test would end after four consecutive mistakes or after the 25 easy items have been administered. When participants make less than four mistakes in the 12 first items of the intermediate item set, they are automatically allocated the points for the easy item set and the test continues until four consecutive mistakes are reached.

Items constituting the first item set being named from 100 to 90 % of healthy controls, the automatic allocation of the points to participant that obtain the expected performance on the first half of the second item set can artificially improve their performance. Indeed, in 10 % of the cases a healthy participant would not be able to name the last item of the easy item set. The point would, however, be given. The most crucial information to track change in subject´s performance is, however, located at the end of his performance spectrum. Punctual errors in the first item set would not be particularly informative. For this reason, and considering the time gain arising from using this administration method, the minimal loss of information is not seen here as a major caveat.

To summarize the final item choice, the creation of three item sets of different difficulty levels as well as the implementation of both a stop and return criteria should allow reducing administration time considerably while obtaining a maximum of information about participant´s ability. It should ease the implementation of the GONT in routine clinical examination to track early changes in lexical abilities.

3.7 Cross-sectional administration of the Graded Object-Naming Task

Primary Progressive Aphasia (PPA) is a neurodegenerative syndrome characterized, at first, by an isolated progressive language impairment (Mesulam, 1982). The profile of impairment can vary, giving rise to different PPA subtypes. Latest diagnostic criteria recognize the non-fluent variant of PPA (NfvPPA) that is characterized by agrammatism in language production and/or apraxia of speech; the semantic variant of PPA (SvPPA) characterized by impaired confrontation naming and single-word comprehension; as well as the logopenic variant of PPA (LvPPA) characterized by word finding difficulties both in spontaneous speech and naming, and impaired repetition of sentences and phrases (Gorno-Tempini et al., 2011). The description of LvPPA is, however, considered too narrow to describe the full scope of impairment in patients that have neither NfvPPA nor SvPPA and often present an underlying AD (Sajjadi, Patterson, & Nestor, 2014). Therefore, the label of mixed PPA (mPPA) emerged in the literature (Mesulam, Wieneke, Thompson, Rogalski, & Weintraub, 2012).

In this chapter, the cross-sectional administration of the firstly reduced item set to patients with NfvPPA, SvPPA and mixed PPA patients with confirmed amyloid pathology is presented. The task was also administered to HC, SCI, patients with MCI and early AD. We compared the performance across groups in the Graded Object-Naming Task (GONT), and contrasted it with the results obtained in the Boston Naming task (BNT) (Merten, 2004), a gold standard in the examination of lexical abilities. Confrontation naming and object knowledge being the central deficit in SvPPA, we expected this group to have the lowest performance, followed by the mixed PPA, early AD and MCI. SCI being mostly a very heterogeneous group that can include patients with presymptomatic dementia but also healthy participants, we had no prior assumption on whether the SCI group would perform different from HC.

3.7.1 Method

3.7.1.1 Participants

Patients with PPA were recruited in the local memory clinic as well as in other memory clinics in Germany thanks to collaborations. They were diagnosed with PPA when they fit the ground criteria: an inaugural language impairment that is at first isolated, progressive and not better explained by any other psychiatric or neurologic disorder (Mesulam, 1982).

They were further subtyped based on the neuropsychological examination, the clinical interview, the Magnetic Resonance Imaging (MRI) and amyloid PET reports following the latest diagnostic criteria (Gorno-Tempini et al., 2011). Many patients with a positive amyloid PET, that indicates an underlying AD, did not fulfil the criteria for any of the proposed variants or fulfilled more than one variant. These mixed PPA with disease confirmation are referred to as Aß+PPA.

In this section the results of all participants of the longitudinal study (presented in chapter 3.8) at baseline (78 HC, 17 SCI, 19 MCI, 13 AD) as well as the results obtained by 26 PPA (12 Aß+PPA, 5 NfvPPA and 9 SvPPA) were included. Seven supplementary HC were excluded because of an MMSE lower than 27.

3.7.1.2 Material

3.7.1.2.1 26 items

Given the extent of naming deficits in the PPA cohort, especially for Aß+PPA and SvPPA, we mainly administered the 26 items that were named by 100% of the HC in the piloting study (see chapter 3.3). To note, only 14 of these items were retained in the final version of the test.

3.7.1.2.2 75 items

The total score on the GONT was based on the sum of the correctly named items, the maximal score being 75. For patients with anomia (SvPPA and Aß+PPA) the test was often abandoned before the end (after 26 items). The items being rank-ordered, those that were not administered were automatically considered as incorrectly named.

3.7.1.3 Statistical considerations

Descriptive statistics were run using SPSS version 21 (IBM corp., Armonk, NY).

Normality was assessed using the Shapiro-Wilk test, adapted for small sample size and the homogeneity of the variances was assessed using Levene´s test. Given the non-homogeneity of variance and the non-normal distribution of the percentage of correct answers in the full test, in the 26 items with 100% of namability and in the BNT, a Kruskal-Wallis test was run to compare the performance across groups.

Although naming is a narrow concept, we compared the results obtained in the GONT to performance obtained in a German short version of the BNT (Merten, 2004), a standard confrontation naming task. We included in the analysis the seven HC that had a low MMSE

and had therefore a sample of 160 participants. Given the non-normality of the data distribution, Spearman´s rho correlation was computed.

3.7.2 Results

3.7.2.1 Cross-sectional comparison

3.7.2.1.1 26 items

The descriptive statistics for the 26 items with a namability of 100% are summed up in table 3.5 and figure 3.13.

There was a significant difference on the score obtained on the 26 items across groups. Post-hoc analysis revealed that the SvPPA and Aß+PPA were significantly worse than all the other groups. SvPPA were significantly worse than Aß+PPA (see table 3.5).

3.7.2.1.2 75 items

The descriptive statistics for the percentage of correctly named items on the total test are presented in table 3. 5 and figure 3.14.

There was a significant difference on the percentage of success across groups. Post-hoc analysis revealed that SvPPA, Aß+PPA and AD were significantly impaired compared to HC. Only a trend toward a significant difference between MCI and HC was found (p=0.062) (see table 3.5).

Figure 3-13 Performance obtained on the 26 items with a 100% namability

The bars represent the standard error of the mean.

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, NfvPPA = Non-fluent variant of Primary Progressive Aphasia, AD = Alzheimer´s disease, Aß+PPA = Amyloid related Primary Progressive Aphasia, SvPPA= Semantic variant of Primary Progressive Aphasia.

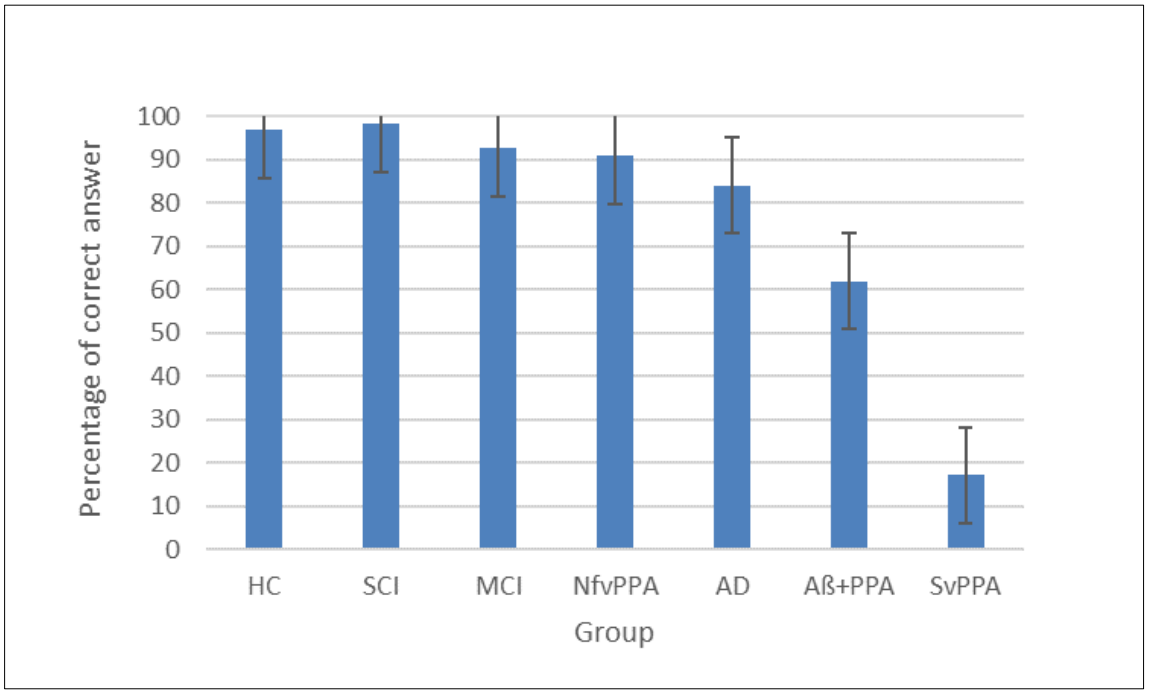
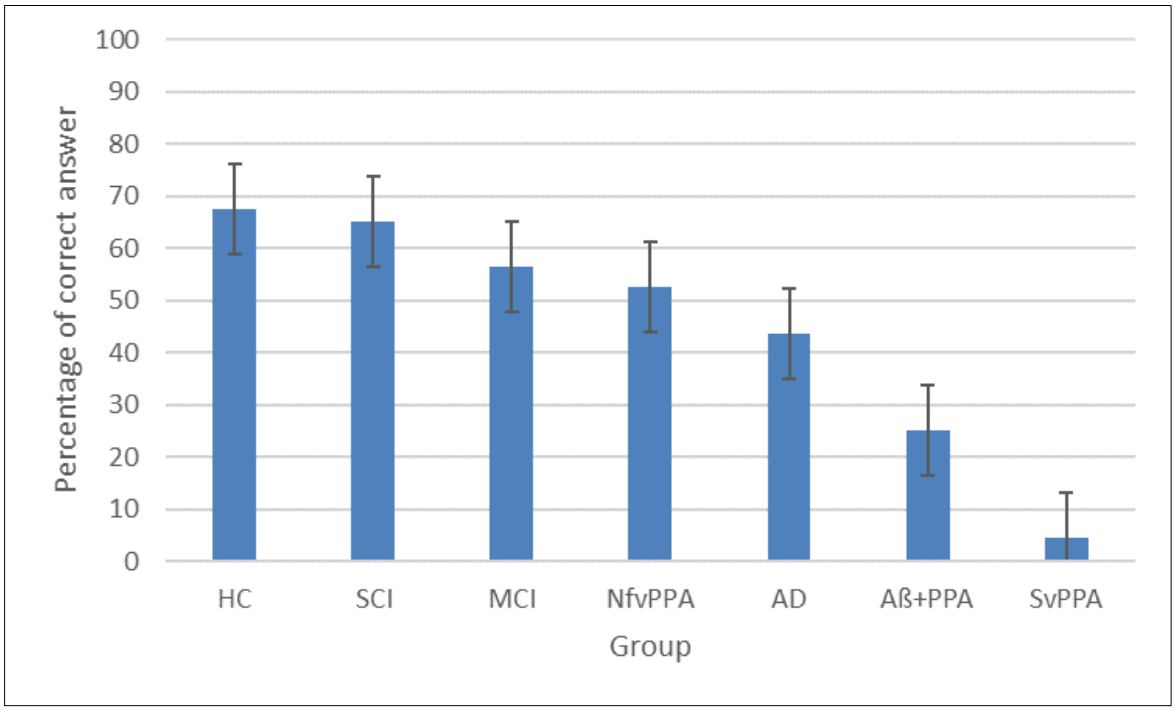


Figure 3-14 Performance in the total test (75 items)

For patients with SvPPA and Aß+PPA the test was often abandoned before the end. Items being rank ordered by difficulty level, those that could not be administered were considered as not successfully named. The bars represent the standard error of the mean.

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, NfvPPA = Non-fluent variant of Primary Progressive Aphasia, AD = Alzheimer´s disease, Aß+PPA = Amyloid related Primary Progressive Aphasia, SvPPA= Semantic variant of Primary Progressive Aphasia.



3.7.2.2 Construct validity

The correlation between the performance obtained in the widely used BNT and the percentage of success in the newly developed GONT was high and significant (Spearman´s Rho: 0.8, $p<0.001$).

3.7.2.3 Boston Naming Test

The descriptive statistics for the percentage of correctly named items on the BNT are presented in table 3.5 and figure 3.15.

There was a significant difference on the percentage of success across groups. Post-hoc analysis revealed that only SvPPA and Aß+PPA were significantly impaired compared to controls (see table 3.5).

Figure 3-15 Performance in the BNT

The bars represent the standard error of the mean.

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, NfvPPA = Non-fluent variant of Primary Progressive Aphasia, AD = Alzheimer´s disease, Aß+PPA = Amyloid related Primary Progressive Aphasia, SvPPA= Semantic variant of Primary Progressive Aphasia.

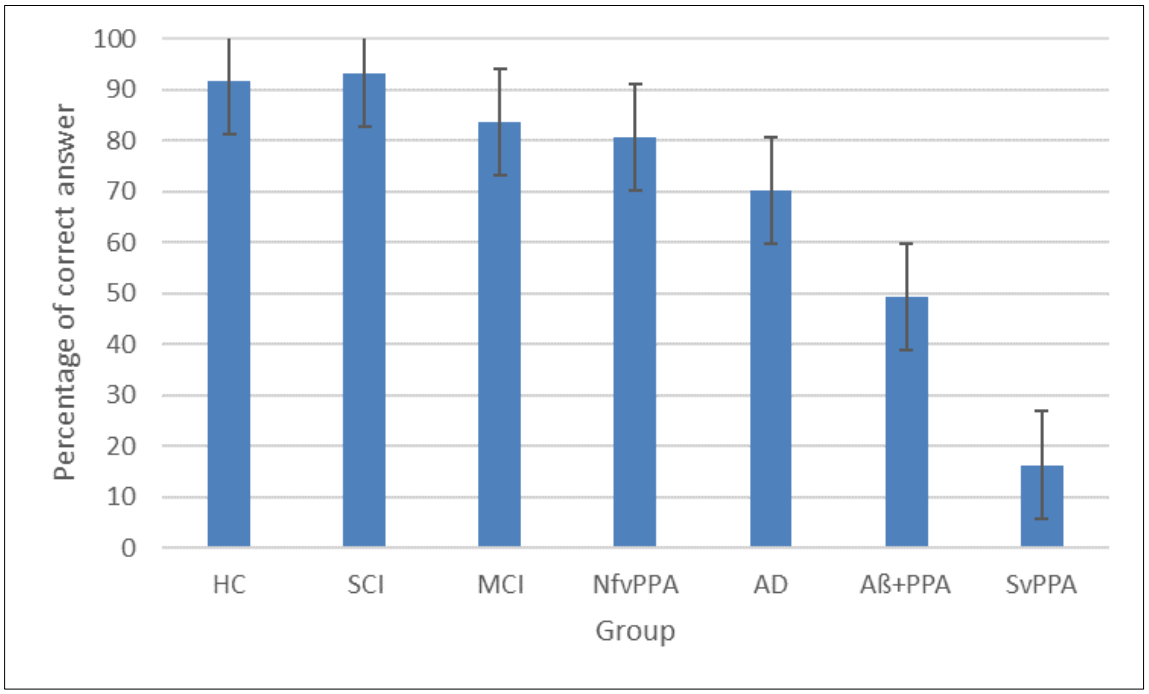


Table 3-5 Performance in the Graded Object Naming Task

** p<.05 compared to the control population, [A]p<.05 compared to SCI, [B]p<.05 compared to MCI, [C]p<.05 compared to NfvPPA, [D]p<.05 compared AD, [E]p<.05 compared to Aß+PPA, [F]p<.05 compared to SvPPA. Bold italics alphabets indicate significance survives Bonferroni correction for multiple comparisons.*

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, NfvPPA = Non-fluent variant of Primary Progressive Aphasia, AD = Alzheimer´s disease, Aß+PPA = Amyloid related Primary Progressive Aphasia, SvPPA= Semantic variant of Primary Progressive Aphasia.

	HC Mean (SD)	**SCI Mean (SD)**	**MCI Mean (SD)**	**NfvPPA Mean (SD)**	**AD Mean (SD)**	**Aß+PPA Mean (SD)**	**SvPPA Mean (SD)**	**Significance *P-value***
Percentage of success (26 items)	96.8^{E}***F*** (2.8)	98.2^{E}***F*** (.7)	92.5^{E}***F*** (12.1)	90.8***F*** (14.5)	84***F*** (18.8)	61.9****A***^{B}***F*** (25.6)	17.1****ABCD***^{E} (21.2)	$X^2(6, 153) = 68.4$, p<0.001
Percentage of success (75 items)	67.6***DEF*** (9.5)	65.2***DEF*** (8.2)	56.5***EF*** (14)	52.5^{F} (14.8)	43.6****A***^{E}***F*** (12.6)	25.2****AB***^{DF} (14.2)	4.5****AB***^{C}***D***^{E} (5.5)	$X^2(6, 153) = 82.7$, p<0.001
Percentage of success in BNT (30 items)	91.7***EF*** (7.6)	93.1***EF*** (6.6)	83.7^{E}***F*** (16.4)	80.7***F*** (17.4)	70.3***F*** (21.1)	49.4****A***^{BF} (23.4)	16.3****ABCD***^{E} (15.7)	*$X^2(6, 153) = 69.9$, p<0.001*

3.7.3 Discussion

3.7.3.1 Cross-sectional comparison

As expected, SvPPA and Aß+PPA had the worst performance in the 26 items with a 100% namability. NfvPPA, AD, MCI and SCI had normal performance on this set of items. While these items are not very informative in early AD, MCI or SCI because they perform close to ceiling, in PPA patients they allow to quantify the naming impairment in the different subtypes. As reported in the literature NfvPPA are mostly preserved in confrontation naming while SvPPA are severely impaired (Gorno-Tempini et al., 2011). Aß+PPA show a lighter impairment that is significantly worse than NfvPPA and significantly less important than in SvPPA. Although PPA patient were of limited number (especially the NfvPPA), we see a potential in this set of highly nameable items for using with patients that are highly anomic.

In the total test, only AD, Aß+PPA and SvPPA were significantly impaired compared to controls. In MCI, there was only a trend toward a significant difference to HC. While, the primary aim of the GONT was not the comparison between different groups of patients on a single time point, the results seem to show that our test is superior to the BNT in separating patients groups. Indeed, in our sample, no significant difference or trend toward significance was found between HC and AD or MCI in the BNT. This might be due to a light ceiling effect, which is commonly reported in short confrontation-naming tasks. It seems that the 30 items BNT version used in this cohort had the same cross-sectional potential than our set of highly nameable items. In both tasks, only Aß+PPA and SvPPA were impaired compared to control. Only a shorten version of the BNT was used and the full version might have yield better cross-sectional abilities. This is, however, unlikely as the 30-item-version is highly correlated with the full version (Merten, 2004).

Noteworthy is that on average HC only named correctly 68% of items in the GONT. The test could display a hidden ceiling effect, in that no one can name the most difficult items. This hypothesis can, however, be excluded as we know from previous analysis (see section 3.6) that the hardest item of the test could be named by 7% of a sample of 85 HC. This result underlines the ability of the test to capture the full scope of performance, even in participant with very good lexical abilities. This is crucial when trying to track change in participants with high level of education.

3.7.3.2 Construct validity

Although, naming is a robust and straightforward concept, we looked into the construct validity of our task and found a high and significant correlation between the world-widely used BNT and the presented GONT. Moreover, the cross-sectional analysis being run on patients with PPA, known for presenting with naming impairment (especially for Aß+PPA and SvPPA) and showing a significant impairment in the test, it underlines a satisfying construct validity.

To summarize, it seems that the GONT offers some potential for testing patients with profound language impairment like PPA. Moreover, the lack of ceiling effect, often reported in classical confrontation naming task, makes it particularly adapted for the examination of highly educated participants.

3.8 Longitudinal assessment of the Graded Object Naming Task

One of the primary aims of the study was to construct a graded picture-naming test that would be able to track change from normal cognition to early dementia. This implies having a test that would be sensitive to slight changes in abilities in a single person. The good reliability as well as the high reproducibility of the task has been previously demonstrated. The longitudinal administration of the firstly reduced item set to HC, participants with SCI, MCI and early AD is presented in this chapter. The results in the Graded Object-Naming Task (GONT) were compared to the evolution of the general cognitive level described by the MMSE, as well as with the performance in a German short form of the BNT (Merten, 2004).

3.8.1 Method

3.8.1.1 Participants

Forty-one HC and 40 patients (10 AD, 15 MCI and 15 SCI) participated to a minimum of two time points in the longitudinal study (see section 2.3 for further information). Sixty-two participants (29 HC, 15 SCI, 12 MCI and 6 AD) were tested on a minimum of three time points.

3.8.1.2 Progressor group versus stable group

For the longitudinal study, participants were subdivided into progressor and stable groups. A participant was included in the progressor group, regardless the diagnostic, when he or she lost a minimum of 0.5 (or in a second variant 1 Z-score) in the MMSE (Folstein et al., 1975) between baseline and the last available visit. The status (progressor vs stable) was based on the last available visit to account for participants that have a nonlinear evolution (e.g. Improvement due to practice effect followed by worsening due to cognitive decline). The prime aim of the test being to track change in single subject and not differential diagnosis, the distinction between progressor and stable was favoured over the diagnosis status. This also allowed accounting for participant that entered the study as HC, declined slightly in the test interval, but not enough to fulfil the criteria for MCI or AD

Using the 0.5 Z-score loss criterion 54 stable and 27 progressor participants were administered the test at least twice, and 42 stable and 20 progressor participants at least three time. Using the 1 Z-score loss criterion, 65 stable and 16 progressor participants were

administered the test at least twice, and 50 stable and 12 progressor participants at least three times.

3.8.1.3 Neuropsychological assessment

In the longitudinal part of the study, patients and HC were administered, in a single testing session of approximatively 90 minutes, a battery of standard neuropsychological tests listed in table 3.6. In the same session, the novel neuropsychological test from the DZNE-Cog was administered. Finally, two in-house informant questionnaires on activities of daily living and behaviour were, when possible, given to the next of kin. Testing session was repeated longitudinally with standard time lapse of 6 months (μ=6.5; SD=0.6).

3.8.1.4 Statistical considerations

Normality was assessed with the Kolmogorov-Smirnov test and the homogeneity of the variances was assessed using Levene´s test. Given the non-equality of the variance and the non-normal distribution for demographics and performance on the neuropsychological tests, a Kruskal-Wallis test was run to compare the different groups. The gender division was assessed with a Pearson Chi-Square test. The Dunnett´s T3 test was used for post-hoc comparisons.

The distribution of the percentage of correct answer on the GONT being normal in HC, a Pearson correlation between baseline and the second visit was run to assess the test-retest reliability.

R version 3.1.1 (R Core team, 2014) and lme4 version 1.1.13 (Bates, Maechler, Bolker & Walker, 2015) was used to perform a linear mixed effects analysis on the relationship between the performance in the GONT, the time point and the status of cognitive stability (progressor versus stable). An interaction of the time point and the status of cognitive ability was entered as a fixed effect into the model. Intercepts for participants and time lapse between two time points were entered as random effects. P-values were obtained by likelihood ratio test of the full model with the questioned effects, against the model without the effects and interaction questioned. The same procedure was applied to the performance obtained in the BNT.

Table 3-6 List of standard neuropsychological tests used in the longitudinal study

Cognitive Domain	Test used
Global functioning	- MMSE - ADAS-Cog
Attention & Executive functions	- Subtest digit symbol substitution
Fluency	- Category fluency (animals) and lexical fluency (S) [2 minutes] - Lexical fluency (K) [1 minute]
Non-verbal memory	- Immediate and delayed recalls of the Rey-Osterrieth complex figure (ROCF)(Osterrieth, 1944; Rey, 1941)
Verbal memory	- Digit span
Visuo-construction	- Copy of the ROCF
Confrontation naming	- BNT-30 (Merten, 2004)
Questionnaires	- ADL questionnaire for caregiver - Behavioural questionnaire for caregiver
Depression	- GDS

3.8.2 Results

3.8.2.1 Demographics and general neuropsychology

Table 3.7 presents the demographics and MMSE results for the HC and patients involved in the longitudinal study. As expected AD had the lowest MMSE score, followed by MCI. SCI and HC did not differ significantly. AD had a significantly lower level of education than HC.

Table 3-7 Demographics and MMSE at baseline

** p<.05 compared to the control population, [a]p<.05 compared to SCI, [b]p<.05 compared to MCI, [c]p<.05 compared to AD. Bold italics alphabets indicate significance survives Bonferroni correction for multiple comparisons.*

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, AD = Alzheimer´s disease, F= Female, M= Male, n.s = non-significant.

	HC Mean (SD)	**SCI Mean (SD)**	**MCI Mean (SD)**	**AD Mean (SD)**	**Significance *P-value***
Age (years)	68.2 (6.9)	68.3 (8)	72.5 (7.4)	75.6 (7)	$X^2(3, 81) = 9.8$, *p=0.02*
Education (years)	15.1 (3.1)^{C}	14.6 (3.3)	13.6 (2.9)	11.8 (2.8)*	F (3, 81) =9.7, *p=0.02*
Gender	27F, 14 M	8 F, 7 M	6 F, 9 M	5 F,5 M	*n.s*
MMSE (Max.30)	28.7 (1.2)^{BC}	28.8 (1.1) ^{BC}	27 (1.7)^{*AC}	22.8 (3.3)^{*AB}	$X^2(3, 81) = 35$, *p<0.001*
N	41	15	15	10	

3.8.2.2 Test-retest reliability

The correlation between the percentage of correct answer in the GONT obtained at baseline (65.6, SD: 10.5) and at the second visit (67.6, SD 10.5) in HC (N=41) was high r=0.86 and highly significant p<0.001, indicating a good rest-retest reliability in the task.

3.8.2.3 Descriptive change

3.8.2.3.1 Two-time points

Table 3.8 presents the demographics and table 3.9 the general neuropsychology performance for the subject in both the stable and progressor groups. Figure 3.16 displays the evolution over two time points in the MMSE, BNT and GONT. The figure showed that in all tests, the progressor group have a lower baseline performance. In the BNT and GONT, both the progressor and stable groups seem to improve minimally on the second time point.

Table 3-8 Demographics of all participants over two time points

Abbreviations: F= Female, M= Male.

Demography				
Criterion for progression	-0.5 Z-score		-1 Z-score	
Group	**Stable Mean (SD)**	**Progressor Mean (SD)**	**Stable Mean (SD)**	**Progressor Mean (SD)**
Age at baseline (years)	68.7 (7.7)	72.3 (6.9)	69 (7.4)	73.5 (7.4)
Education (years)	14.3 (3.3)	14.4 (3.2)	14.2 (3.2)	14.8 (3.4)
Gender	32 F, 22 M	14 F, 13 M	37 F, 28 M	9 F, 7 M
Sub-groups				
N	54	27	65	16
AD	4	6	5	5
MCI	10	5	13	2
SCI	11	4	13	2
HC	29	12	33	8

Figure 3-16 Performance of progressor and stable groups over two time points

The bars represent the standard error of the mean.

A.1 represents changes in the MMSE score over time in the 0.5 Z-score loss classification

A.2 represents changes in the MMSE score over time in the 1 Z-score loss classification

B.1 represents changes in the BNT score over time in the 0.5 Z-score loss classification

B.2 represents changes in the BNT score over time in the 1 Z-score loss classification

C.1 represents changes in the GONT score over time in the 0.5 Z-score loss classification

C.2 represents changes in the GONT score over time in the 1 Z-score loss classification

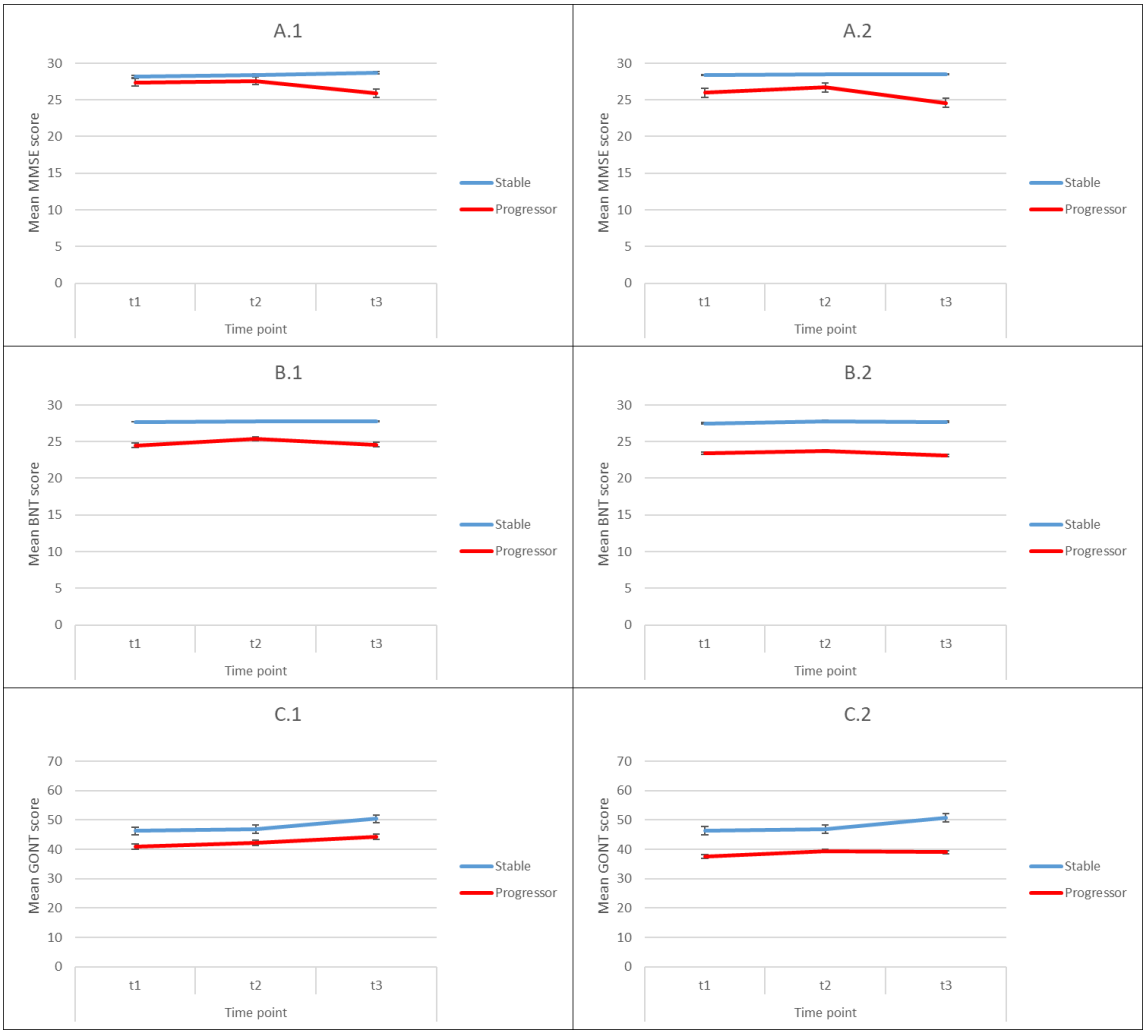


Table 3-9 Neuropsychology of all participants over two time points

For logistical reasons, it was not possible to administer all neuropsychological tests to all participants. Therefore, there is variation in the number of participant across tests.

Criterion for progression	**-0.5 Z-score**				**-1 Z-score**			
Group	**Stable Mean (SD) [N]**		**Progressor Mean (SD) [N]**		**Stable Mean (SD) [N]**		**Progressor Mean (SD) [N]**	
Time point	T1 [54]	T2 [54]	T1 [27]	T2 [27]	T1 [65]	T2 [65]	T1 [16]	T2 [16]
MMSE (Max. 30)	28 (1.8) [54]	28.3 (1.9) [54]	27 (3.6) [27]	26.7 (4) [27]	28.2 (1.8) [65]	28.3 (1.9) [64]	25.7 (2.6) [16]	25.6 (4.6) [16]
BNT (Max. 30)	27.5 (2.3) [53]	27.6 (2) [53]	24 (6.5) [27]	24.4 (6.3) [27]	27.1 (3.2) [64]	27.2 (3) [63]	23.3 (7) [16]	23.2 (6.6) [16]
GONT (Max. 75)	45.6 (8.1) [53]	46.8 (46.8) [53]	39.4 (14.6) [27]	40.2 (15) [27]	44.9 (8.9) [65]	45.8 (10.9) [65]	37.4 (15.8) [16]	38.6 (16.7) [16]
ROCF Copy (Max. 36)	32 (3.4) [35]	32.6 (3.3) [53]	29.3 (6.7) [17]	28.4 (8.6) [27]	32 (3.3) [41]	32.4 (3.4) [63]	27.7 (7.7) [11]	26.2 (10.2) [16]
ROCF immediate Recall (Max. 36)	16.4 (7.7) [35]	17.4 (7.4) [53]	13.5 (8) [17]	15.9 (8.5) [27]	16.1 (7.4) [41]	17.5 (7.2) [63]	12.9 (9.5) [11]	13.9 (9.3) [16]
ROCF delayed Recall (Max. 36)	15.8 (7.3) [35]	17 (7) [53]	12.8 (7.7) [17]	14.8 (8.5) [27]	15.7 (7) [41]	17 (6.9) [63]	11.6 (9) [11]	12.7 (9.2) [16]
ADAS-Cog (Max. 70)	8.3 (3.6) [36]	8.2 (4.4) [53]	9.5 (7) [15]	11.7 (7.9) [26]	8.2 (3.4) [42]	8.3 (4.4) [62]	10.9 (8.8) [9]	13.7 (9.1) [16]
GDS (Max.15)	0.7 (0.8) [33]	1.4 (2.1) [49]	0.9 (1.2) [15]	1.4 (1.9) [23]	0.6 (0.8) [39]	1.4 (2) [57]	1.2 (1.4) [9]	1.6 (1.5) [14]

3.8.2.3.2 Three time points

Table 3.10 presents the demographics and table 3.11 the general neuropsychology performance for the subject in both the stable and progressor groups. Figure 3.17 displays the evolution of performance in the MMSE, BNT and GONT over three time points. In the BNT, both classification criteria produce similar results: both group stay relatively stable. In the GONT, the 1 Z score classification seem to be the most sensitive to separate the progressor group from the stable group. Indeed, the stable group seem to improve slightly while the progressor group stays relatively stable.

Table 3-10 Demographics of all participants over three time points

Abbreviations: F= Female, M= Male.

Demography				
Criterion for progression	-0.5 Z-score		-1 Z-score	
Group	**Stable Mean (SD)**	**Progressor Mean (SD)**	**Stable Mean (SD)**	**Progressor Mean (SD)**
Age at baseline (years)	68.6 (7.9)	73.2 (7.1)	68.7 (7.6)	74.4 (7.9)
Education (years)	14.6 (3.2)	14.1 (3.4)	14.4 (3.2)	14.6 (3.7)
Gender	25 F, 17 M	10 F, 10 M	29 F, 21 M	6 F, 6 M
Sub-groups				
N	42	20	50	12
AD	3	3	3	3
MCI	8	4	10	2
SCI	11	4	13	2
HC	20	9	24	5

Figure 3-17 Performance of progressor and stable groups over three time points

The bars represent the standard error of the mean.

A.1 represents changes in the MMSE score over time in the 0.5 Z-score loss classification

A.2 represents changes in the MMSE score over time in the 1 Z-score loss classification

B.1 represents changes in the BNT score over time in the 0.5 Z-score loss classification

B.2 represents changes in the BNT score over time in the 1 Z-score loss classification

C.1 represents changes in the GONT score over time in the 0.5 Z-score loss classification

C.2 represents changes in the GONT score over time in the 1 Z-score loss classification

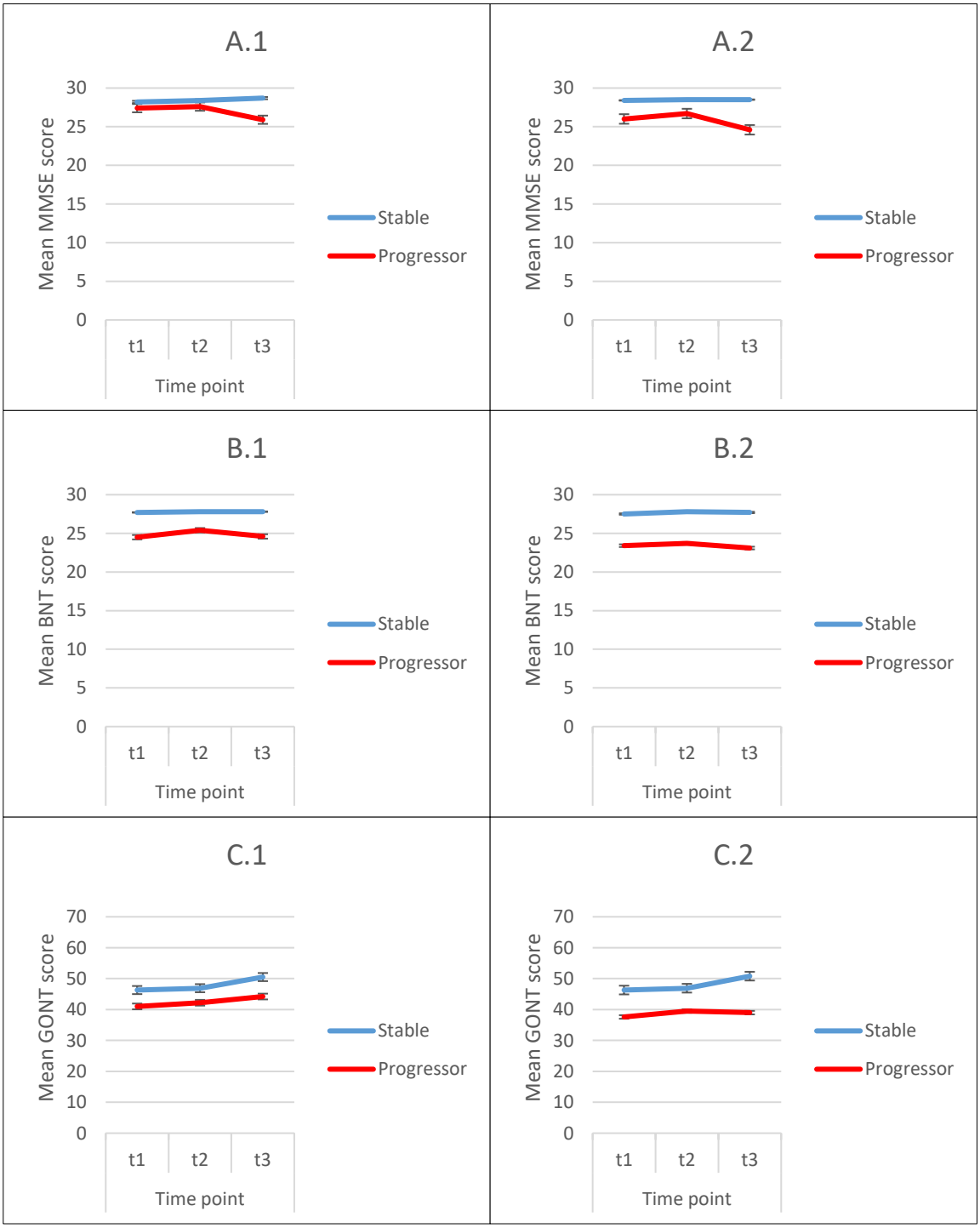


Table 3-11 Neuropsychology over three time points

For logistical reasons, it was not possible to administer all neuropsychological tests to all participants. Therefore, there is variation in the number of participant across tests.

Criterion for progression	**-0.5 Z-score**						**-1 Z-score**					
Group	**Stable Mean (SD) [N]**			**Progressor Mean (SD) [N]**			**Stable Mean (SD) [N]**			**Progressor Mean (SD) [N]**		
Time point	T1 [42]	T2 [42]	T3 [42]	T1 [20]	T2 [20]	T3 [20]	T1 [50]	T2 [50]	T3 [50]	T1 [12]	T2 [12]	T3 [12]
MMSE (Max. 30)	28.2 (1.8) [42]	28.4 (1.9) [42]	28.7 (1.6) [42]	27.4 (3.3) [20]	27.6 (2.9) [20]	25.9 (3.8) [20]	28.4 (1.7) [50]	28.5 (1.8) [50]	28.5 (1.6) [50]	26 (3.6) [12]	26.7 (3.4) [12]	24.6 (1.6) [12]
BNT (Max. 30)	27.7 (2.3) [41]	27.8 (1.9) [41]	27.8 (1.8) [42]	24.5 (5.6) [20]	25.4 (5) [20]	24.6 (5.9) [20]	27.5 (2.6) [49]	27.8 (1.9) [49]	27.7 (2.1) [50]	23.4 (6.5) [12]	23.7 (5.8) [12]	23.1 (6.8) [12]
GONT (Max. 75)	46.3 (7.9) [42]	46.9 (11.2) [42]	50.5 (8.4) [42]	41 (13.2) [20]	42.2 (12.2) [20]	44.2 (14.7) [20]	46.3 (7.7) [50]	46.9 (10.4) [50]	50.8 (8.1) [50]	37.6 (15.4) [12]	39.5 (14.8) [12]	39 (16.4) [12]
ROCF Copy (Max. 36)	31.9 (3.2) [26]	32.6 (3.3) [41]	32.6 (3) [40]	28.6 (7.6) [12]	29 (6.8) [20]	28.2 (7.5) [19]	31.8 (3.1) [30]	32.3 (3.4) [49]	32.2 (3.4) [48]	27.4 (9.1) [8]	27.6 (8) [12]	26.8 (8.9) [11]
ROCF immediate Recall (Max. 36)	16.3 (8.1) [26]	17.7 (7.2) [41]	20 (6.7) [40]	13.4 (8.5) [12]	16.8 (7.9) [20]	18.2 (8.2) [19]	15.9 (7.8) [30]	18 (6.8) [49]	20 (6.5) [48]	13.3 (10) [8]	15.3 (9.2) [12]	16.9 (9.5) [11]
ROCF delayed Recall (Max. 36)	15.9 (7.7) [26]	17.2 (7.2) [41]	19.3 (7.4) [40]	12.6 (7.9) [12]	15.8 (7.6) [20]	17.1 (8.2) [19]	15.7 (7.2) [30]	17.3 (6.8) [49]	19.2 (7) [47]	11.9 (9.4) [8]	14.3 (9.1) [12]	15.9 (9.9) [11]
ADAS-Cog (Max. 70)	8.3 (4.1) [27]	7.7 (4.1) [41]	6.8 (3.7) [41]	10.5 (7.7) [11]	10.7 (6.7) [20]	9.5 (6.6) [19]	8.1 (3.8) [31]	7.8 (4.2) [48]	7.1 (3.8) [49]	12.6 (9.2) [7]	12.2 (7.3) [12]	10.2 (8.2) [11]
GDS (Max.15)	0.6 (0.8) [24]	1.3 (2.1) [39]	1.2 (2.1) [41]	0.7 (0.9) [10]	1.7 (1.6) [17]	1.2 (1.5) [16]	0.6 (0.7) [28]	1.3 (2) [46]	1.2 (2) [48]	1 (1.1) [6]	1.8 (1.6) [10]	1.1 (1.5) [9]

3.8.2.4 Statistical change

3.8.2.4.1 2 time points

In the classification of progressor versus stable participants on the base of a 0.5 Z-score loss between baseline and the last available visit, there was only a significant effect of the group on the performance in the GONT (p=0.006) and the BNT (p=0.001). No effect of the time point (p=0.09) or interaction effect between the group and the time point (p=0.12) could be found in the GONT. Similarly, no effect of the time point (p=0.56) or interaction effect (p=0.57) could be found in BNT.

When using the classification based on a 1 Z-score loss between baseline and the last available visit, there was again only a significant effect of the group on the GONT (p=0.007) and the BNT (p=0.002). No significant effect of the time point (p=0.21) or an interaction effect between the group and the time point (p=0.14) could be found in the GONT. Similarly, no significant effect of the time point (p=0.05) or an interaction effect between the group and the time point (p=0.42) could be found in the BNT.

3.8.2.4.2 3 time points

In the classification based on a 0.5 Z-score loss between baseline and the last available visit, there was only a significant effect of the time point (p=0.03) on the performance in the GONT. No effect of the group (p=0.13) or interaction effect (p=0.27) could be found. In BNT there was no significant effect of the time point (p=0.92), the group (p=0.08) or interaction effect of both factors (p=0.98).

In the classification based on a 1 Z-score loss between baseline and the last available visit, there was a significant effect of the time point (p=0.005), the group (p=0.004) as well as an interaction effect between both factors (p=0.03) on the performance in the GONT. In the BNT no significant effect of the time point (p=0.6) or an interaction effect between the group and the time point (p=0.34) could be found.

3.8.3 Discussion

Previous chapters presented the ability of the newly developed GONT to cover a wide range of difficulty while avoiding ceiling effects in HC as well as floor effects in patients. The internationality of the item set as well as its reproducibility in two comparable samples was confirmed. Finally, its ability to distinguish better than the BNT between patient groups and HC was shown in a cross-sectional analysis. Good construct validity and internal

consistency were reported. One of the main aim being to track change in participant over time, we explored the results obtained longitudinally in the GONT and compared them with performance in the BNT.

There was a significant positive correlation in HC between the performance obtained in the task at baseline and at the second visit. It indicates good test-retest reliability, which is essential to ensure that changes in performance are due to change in ability, and not an intrinsic property of the task. Noteworthy is that, some of the HC belonged to the progressor group. Test-retest reliability being run on the full group of HC, it might have led to a slight under estimation of the test-retest reliability. Indeed, performance on the GONT from HC who were included in the progressor group might have been slightly lower at retest. At the group level, the score on the MMSE seemed, however, to stay stable from baseline to the second visit.

The analysis ran on two time points (around 6 month of interval) did not show any significant ability of either the GONT or the BNT to track change in participants that got worse in the MMSE. Naming not being the first and most prominent symptom in AD or MCI, this time lapse was possibly too short for a significant drop in naming abilities, both when taking in account people who lost 0.5 and 1 Z-score in the MMSE between baseline and the last available visit. A larger drop in the MMSE score might have led to more significant results but would have required a larger sample. Indeed, only few people in the available sample had a larger drop in the MMSE score in this time lapse. When analysing the results over three time points (around 1 year of interval), although the sample size was considerably reduced, there was a significant interaction effect of both the group and the time point in the GONT. It indicates its ability to distinguish, after a year, participant that had a drop in the general cognitive level (indicated by a drop in MMSE score). This was not the case for the BNT where no such effect could be found in both the -0.5 and the -1 Z-score classifications. Noteworthy is that the effects in the GONT reached significance only when in the -1 Z-score classification. A loss of 0.5 Z-score was equivalent to the loss of around one point in the MMSE score in our sample. This might have been fortuitous and led to the inclusion of many stable participants in the progressor group. Past studies considered a drop of two to four points reliable (Hensel, Angermeyer, & Riedel-Heller, 2007). This was, however, derived from the results of a cohort exclusively composed of participants who did not develop dementia in the 1.5-year interval of test-retest. Other authors reported a drop of 5 points over 2 years in MMSE as a reliable indicator of cognitive

decline (Mitruschina & Satz, 1991). This cohort did include participants who developed various neurological disorders during the test-retest interval. In our sample, to be included in the progressor group in the classification based on a loss of 1 Z-score minimum, a participant needed a drop of 2 points. This classification seemed more sensitive than the one based on a drop of 0.5 Z-score, although still lower than the reliable drop reported in past studies. The aim of the GONT, in the context of a full neuropsychological battery, being to track the earliest cognitive changes even before the stage where diagnoses are currently made, it seemed reasonable to lower the threshold compared to past studies. Moreover, in past studies the reliable change was computed mostly on the base of healthy participants, who did not receive a diagnosis in the test-retest interval. It, however, did not exclude the possibility that some were in an early phase of cognitive decline, and thus led to an over-estimation of what can be considered a normal MMSE drop. Our study might suffer the inverse bias, in that we might have included in the progressor group participants whose MMSE drop was not reliably linked to cognitive decline but to situational or emotional reasons.

It is important to mention that the classification was made on the difference between the baseline and the last available visit. This varied from participant to participant from two to five visits. It is possible that several participants that had just two visits were included in the group of stable although they would have reached a loss of minimum 0.5 Z-score if they had further visits. This might have reduced the power of our analysis for both the GONT and the compared BNT. Further work on the task would profit from an increased number of visits for all participants.

A further limitation of this study is that the GONT did not have a parallel version that could have minimized practice effects. Creating graded object-naming tasks of perfect equal difficulty being laborious, we propose to administer the GONT as the third test of the planned battery to minimize both primacy and recency effects. Another limitation was the low sample size, particularly for the progressor group. Moreover, the influence of education level on the sensitivity of the test could not be addressed longitudinally because of the reduced sample size. Further work with a larger sample size, might be needed to explore the ability of the GONT to show reliable decline in participants with high level of lexical abilities. It might be of interest to compute the reliable change index in groups of different education level and gender. Finally, the task aiming at being international, it might be worthy to assess the longitudinal value of the item set in another language.

The choice of the MMSE as the only marker of cognitive stability might have a limited sensitivity. Further work, with larger sample might gain at using several markers to assess cognitive decline. The results of caregiver questionnaire on the activity of daily living as well as possible biomarkers like an Amyloid PET might increase the sensitive allocation of participant to a group of progressor or stable.

Despite the limitations and considering the results, we are confident that the test can be administered longitudinally and that the possible changes observed would be representative enough of the participant´s ability level. Moreover, the results seem to show an advantage of the GONT over the BNT in tracking change over time.

3.9 Conclusion

We achieved the creation of an object naming task of graded difficulty. It is devoid of both floor effects, as shown in SvPPA that are highly anomic, and ceiling effects, as shown by the least nameable items of the task (named by 7% of the HC). As intended, the task is fully computerized on the Psychopy software. It is also short and easy to administer. It has clear criteria for both administration and quotation that reduce inter-rater variability.

The item set has shown its applicability in two different languages: German and Slovak. Although, testing in further languages is needed to confirm the findings, the results are encouraging and support the idea that the task can be used in different cultural background.

Finally, the GONT was developed in a view of tracking change over time. The results of the longitudinal study, although restricted to a small sample size and only three time points, seem to show the ability of the task to identify a drop in participants whose general cognitive level (shown by the MMSE) has declined. Its ability to track change seemed superior to the BNT that is a gold standard in the assessment of naming capacity.

In summary, the preliminary results obtained on the newly developed GONT task are promising. Future work is needed to strengthen the findings and explore its psychometric properties in more depth.

Chapter 4.

Assessing topographical memory with the help of virtual navigation

4.1 Introduction

Subjective memory complaint is very common in elderly with or without objective cognitive deficits in standardized neuropsychological examinations (Jonker, Geerlings, & Schmand, 2000). Memory complaint is often a generic complaint from patients and their relatives to refer to word finding difficulties, but also temporal or spatial disorientation. Many caregivers report noticing the first signs of cognitive decline in unusual environment, for example in holiday, where the relative had trouble to recognize places or find his way back. Indeed, spatial disorientation, first in new and with progression of the disease in known environments, is an early symptom of Alzheimer´s disease (AD) (Allison, Fagan, Morris, & Head, 2016; Pai & Jacobs, 2004; Pengas et al., 2010). It results in people getting lost, wandering away from home and putting themselves in danger (Hope et al., 2001).

The difficulties in topographical memory, the memory that allows us to navigate in our environment, are measurable both in ecological neuropsychological testing and corroborating informant questionnaires (Guariglia, 2009). In the four mountains test, a test of places recognition from different viewpoints, topographical short-term memory was found worse in AD and patients with amnestic mild cognitive impairment (MCI) than in Frontotemporal Lobar Degeneration (FTLD) or healthy controls (Bird et al., 2010). Virtual route learning, tested by the Virtual Route-Leaning Test (VRLT) seemed, however, to be even more sensitive and specific than the four mountains test to distinguish between both MCI and AD, and patients with semantic dementia or healthy controls. Indeed, the VRLT had a high sensitivity (95%) and specificity (94%) to differentiate AD patients and HC (Pengas et al., 2010). A study in preclinical AD found impairment in wayfinding only, route

learning becoming impaired only in early symptomatic AD (Allison et al., 2016). The authors suggested that wayfinding might be sensitive to the earliest cognitive decline of AD. In this study participants were, however, allocated to the preclinical AD group on the base of the Aß42 load in the cerebrospinal fluid (CSF). Aß42 load in the CSF does, however, not yield a 100% predictive value for the development of AD (Forlenza et al., 2015). The group of clinically normal elderly without preclinical AD (negative CSF) and with preclinical AD (positive CSF) were perhaps slightly biased by both type I and II errors, attenuating the reported sensitivity of the route-learning task.

The existence of a topographical memory network involving the right hippocampus, right caudate nucleus, right inferior parietal lobule as well as medial parietal region bilaterally was described using fMRI (Maguire, 1998) and reinforces the idea that the assessment of topographical memory might be sensitive to the earliest stages of AD. Indeed, using multivariate statistics, a recent study found a similar network involving the retrosplenial and lateral parietal cortices as well as right medio-dorsal thalamus and right caudate nucleus that greatly overlaps the territory of atrophy in early AD (Pengas et al., 2012).

It is admitted that decline in the episodic memory performance is one of the first symptoms in early AD (Almkvist & Bäckman, 1993; Yau et al., 2015). It is, even already observable at the MCI stage (Petersen et al., 1999). Moreover, Alzheimer´s disease is related to pathology in both hippocampal and mediotemporal regions (Braak & Braak, 1991) which are known to be involved in episodic memory. Therefore, episodic memory assessment is a cornerstone of neuropsychological examination when early dementia is suspected.

Presently, memory evaluation relies mostly on verbal memory tests like word list learning. It is true in test batteries aiming at diagnosing AD, like the Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog) (W. G. Rosen et al., 1984) with a free recall of 10 words and recognition of 12 words, but also in single tests recommended for the diagnosis of dementia. The Rey Auditory Verbal Learning Test (Rey, 1941) or the California Verbal Learning Test (Delis, Kramer, Kaplan, & Ober, 1987) are, for example, widely used for the evaluation of memory decline in early dementia and both rely on free recall of word lists. In the broadly used CERAD-test (Consortium to Establish a Registry for Alzheimer's Disease) (Moms et al., 1989) to assess clinical symptoms in AD, the delayed recall of 10 words is even the most effective subtest to discriminate between HC and mild demented patients, and even more in moderate and severe dementia (Welsh, Butters, Hughes, Mohs,

& Heyman, 1991; Welsh et al., 1992). In the MOntreal Cognitive Assessment (MOCA) (Nasreddine et al., 2005), delayed recall is also part of the subtests that discriminate best between HC and MCI, but also MCI and AD (Cecato, Martinelli, Izbicki, Yassuda, & Aprahamian, 2016).

Verbal memory tests can often differentiate well between HC and elderly with MCI or early AD. They, however, lack internationality as they heavily rely on language. Translation or adaptation in other languages may display different level of difficulty. Moreover, parallel version of the same level of difficulty using word lists is difficult to achieve and might bias longitudinal assessment of individual in the earliest stages of dementia.

Therefore, the creation of a graded topographical memory task would be useful, as part of the DZNE-Cog, to avoid the mentioned limitations in memory assessment.

4.2 Aim

Considering the promising predictive value of topographical memory tests in the early diagnosis of Alzheimer's disease, we aimed at developing a virtual route-learning task to track early cognitive changes in single subject overtime. The task should be fully computerized, as well as short and easy to administer. Particular care was given to avoid ceiling effects, particularly in HC with a high level of education, and floor effects in patients. The task was part of a larger battery for the early detection of cognitive decline: the DZNE-Cog. It was developed to be used internationally, therefore, emphasis was put on creating an intercultural test.

In chapter 4.3 the creation of a preliminary item set for the Virtual City Task (VCT) and, its administration to a sample of German healthy elderly are presented. In chapter 4.4 improvements to the item set and its piloting with German healthy elderly are described. Chapter 4.5 introduces the final item set selected for the task and chapter 4.6 explores the cross-sectional administration of the task. Finally, the longitudinal data obtained on the final version of the virtual-city task and the recall of Rey-Osterrieth Complex Figure (ROCF) are presented in chapter 4.7.

4.3 Test construction and piloting of the Virtual City Task

In this chapter, the creation of an item set for the topographical memory subtest of the DZNE-Cog is described. The results of its piloting on healthy German elderly are presented and its limitations discussed.

4.3.1 Method

4.3.1.1 Material

Five different virtual environments in three dimensions were created. Streets of different widths (4, 8 and 12 meters or narrow, medium and large), buildings of different heights (2 to 5 storeys) and lengths (8, 12, 16 meters) with different façades were used. Textures showing recognizable words ("Bakery" or "Cinema" for example) were avoided and generic buildings providing no lexical support for orientation were privileged. The same building shape or façade could come up more than once in the same environment. To keep the scenery as simple as possible no car, person, advertisement or outdoor furniture was added to the scene (see figure 4.1).

In the five created virtual cities, 15 routes thought to cover a wide range of difficulty were created (see figure 4.2 for an example). The trials differed in length, number of possible decision and angle of the starting point. The navigation was under first person perspective and the eye-level was set at one meter sixty high.

The models were generated in the open-source 3D modelling software Blender (version 2.67b, 2013) and implemented in the game engine Panda 3D (Carnegie Mellon, University, 2010).

4.3.1.2 Administration

Each trial started with a learning phase where a pre-recorded route was shown. The participant was then automatically brought back to the starting point for the reproducing phase. He/she was asked to virtually redrive the seen route using a joystick. Each session started with training. The training route contained no decision to make but many curves for the participants to get familiar with the joystick.

Before the test started, participants were informed that a happy smiley face would appear when they reach the goal and a sad smiley face when they made a wrong decision (see

figure 4.3). Up to five tries to navigate the route successfully were given. The learning phase was repeated after each failure to reach the goal.

Figure 4-1 Screenshot of a scene in the Virtual City Task

Figure 4-2 Bird´s eye view of a scene plan

The green circle depicts the departure and the red arrow the route to reproduce.

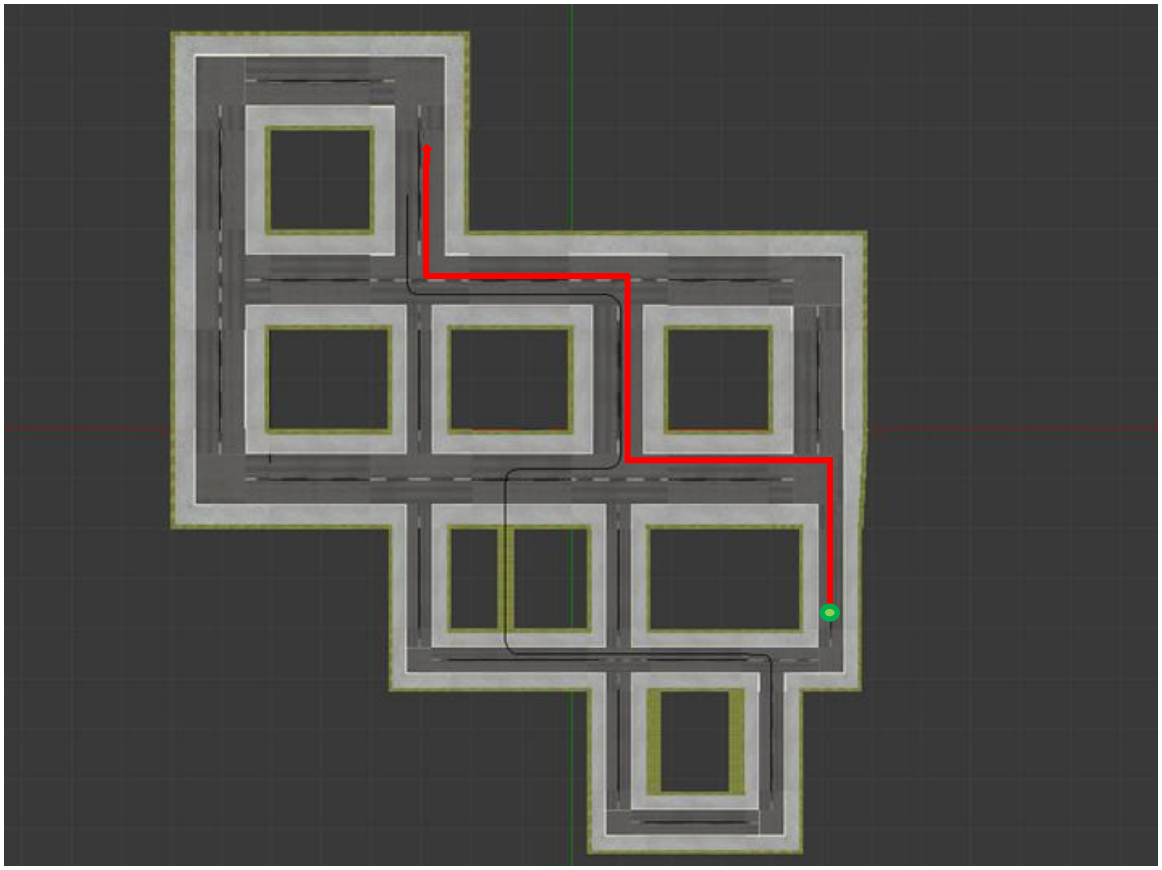
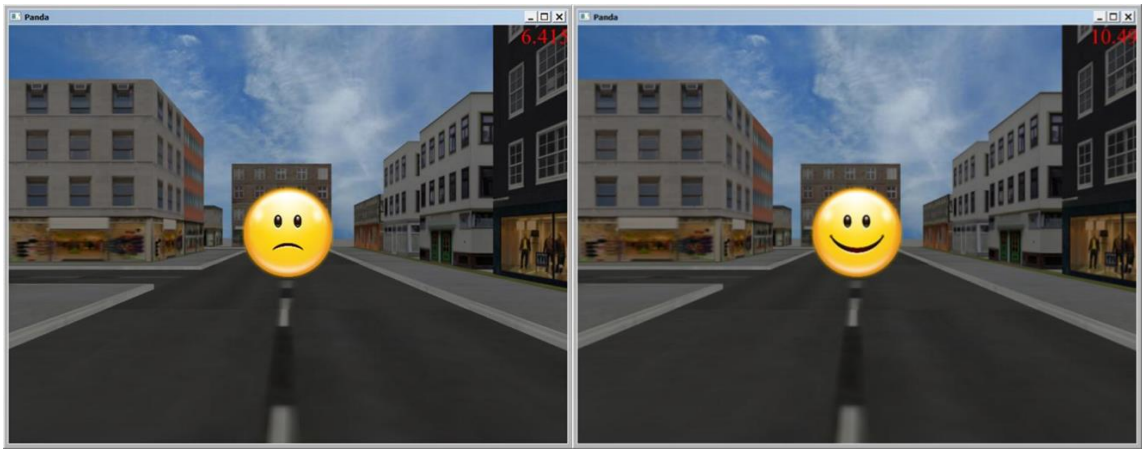


Figure 4-3 Screenshot of the given feedback



4.3.1.3 Participants

Twenty-one healthy elderly were administered the 15 created routes (See section 2.3 for inclusion criteria).

4.3.2 Results

4.3.2.1 Participants

The demographics and MMSE results for the 21 HC are displayed in table 4.1.

Table 4-1 Demographics and MMSE performance

Abbreviations: F= Female, M= Male.

	Mean (SD)
MMSE (Max.30)	28.6 (1.3)
Gender	11 F, 10 M
Age (Years)	63.7 (6.5)
Education (Years)	15.2 (2.6)

4.3.2.2 Limitations

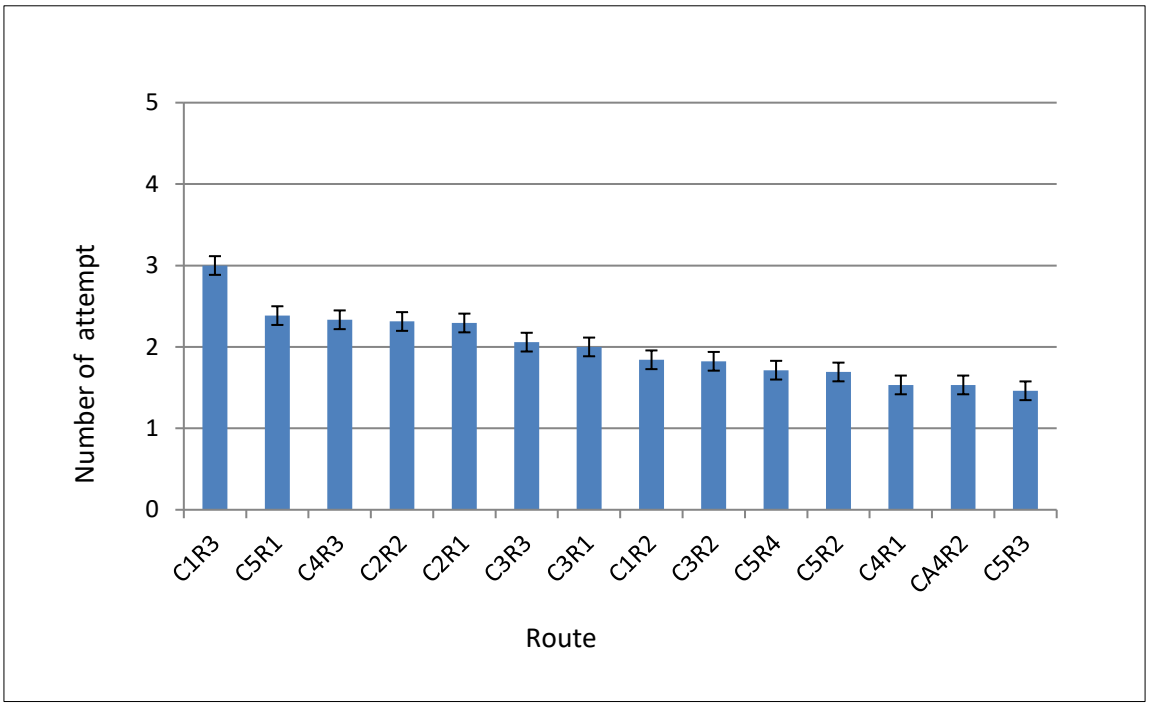
During piloting a large amount of people suffered from motion sickness (N=6/21). They reported feeling dizzy and nauseous after a few tries and the test had to be abandoned. Participants often attributed motion sickness to the navigation speed and a too crowded scenery.

Piloting also revealed difficulties in steering the joystick especially in narrow streets and curves. Many participants were not able to follow the route without the help of the examiner.

The average number of tries needed by HC to complete the route is displayed in figure 4.4. All routes could be achieved by healthy participants in three tries or less. Moreover, figure 4.4 shows that a certain degree of graded difficulty was achieved. The difficulty level did, however, seem a bit too high. Indeed, in seven routes, HC needed more than two tries on average to reach the goal successfully.

Figure 4-4 Difficulty range in healthy elders

The routes are ranked order by difficulty level. The names of the routes are shortened as follows C1R1 = City 1 Route 1. The bars represent the standard error of the mean.



4.3.3 Discussion

Piloting of the Virtual City Task (VCT) in healthy elderly uncovered several limitations in the created material.

A large amount of participants reported feeling sick during administration and had to abandon the test which caused a great loss of data. The reported symptoms (nausea and sudden warm sensation) are known manifestations of motion sickness that are due to the discrepancy between visual and sensory information (Gianaros, Muth, Mordkoff, Levine, & Stern, 2001) The artificial way in which the routes were shown in the learning phase was assumed to increase this effect. Indeed, the route trajectory was drawn with straight lines in the modelling software resulting in artificially angled curves. This might have enhanced the discrepancy between the sensory (sitting still) and visual information (jerky movements). Moreover, participants suffering from motion sickness mentioned that the environment was too crowded. The abundant number of buildings in the foreground led to a large and quick change of visual information on the screen while the scene was previewed. Finally, the navigation speed was reported as being too quick, possibly increasing symptomatic manifestations.

Motion sickness could be reduced by achieving a more natural way to preview the route in the learning phase, simplifying the environment and reducing the navigation speed.

Another limitation exposed by piloting was the large amount of participants that had trouble to steer the joystick, especially in narrow and curvy streets. These difficulties might have had an impact on concentration, and caused frustration possibly biasing memory performance. Reducing the joystick responsiveness (and the corresponding navigation speed) as well as excluding trials using the narrowest roads (4 meters large) should increase comfort in using the joystick.

Although piloting revealed a good feasibility as well as an acceptable degree of graded difficulty, a large amount of routes was too difficult for HC. In half of the trials, they needed more than two tries to successfully reproduce the route. This led to a lengthy administration time (from 40 to 60 minutes) creating fatigue and frustration and possibly affecting the test reliability. This effect might even be larger in people with early cognitive decline, the target of the DZNE-Cog. To reduce administration time and increase test reliability, the hardest routes should be excluded. Moreover, reducing the maximal number of given attempt to reach the goal should reduce administration time.

To summarize, preliminary piloting uncovered several major problems in the test material that needed to be tackle before pursuing further testing.

4.4 Piloting a revised material for the Virtual City Task

Preliminary piloting of the material for the Virtual City Task (VCT) uncovered major issues. Participants suffered from motion sickness and had difficulty to steer the joystick. Moreover, the administration time was long and the material contained unnecessary difficult items. Based on these findings, the material was improved before embarking in a second piloting phase. The VCT asking intensive concentration and administration time being too long, shortening the test seemed a paramount aim. The piloting of the modified item set and its shortening are presented in this chapter.

4.4.1 Method

4.4.1.1 Material and administration

Thirteen routes including eight new routes in two new environments were administered to the participants. Only medium and large routes (8 and 12 meters large) were included in the new environments and the scenery was kept simple (see figure 4.5). Existing environments were simplified by moving most buildings in the background and routes using the narrowest streets were excluded. The navigation speed and responsiveness of the joystick were reduced. Finally, the route preview in the learning phase was pre-recorded per hand to achieve a more natural driving feeling.

The administration procedure was the same than in the feasibility study (see chapter 4.3). At this stage, participants were given up to three tries to reproduce the route successfully.

Participants were awarded 3 points when they redrove the route successfully on the first attempt, 2 points on the second attempt, 1 point on the third attempt. No point was allocated when the participant did not redrive the route successfully after three tries.

Figure 4-5 Picture of a refined environment

4.4.1.2 Participants

Sixty-eight participants (43 HC, 5 AD, 11 MCI, 9 SCI) were administered the modified VCT (see section 2.3 for inclusion criteria).

4.4.1.3 Statistical considerations

To reduce the item-pool a stepwise deletion method based on the internal consistency alpha coefficient was used. Considering the aim of the test, namely the longitudinal tracking of cognitive change, this method seemed adapted. Indeed, internal consistency is of critical importance for tools used longitudinally. It ensures that changes in score reflect real changes in ability and not a lack of reliability.

4.4.2 Results

4.4.2.1 Participants

Thirteen participants had incomplete data set because they abandoned the test before completing all routes. This was especially recurrent in patients (8/13). The test being, at this stage, administered in a randomized manner, uncompleted routes were not always the hardest routes and could not automatically be quoted as “failed”. Therefore, incomplete datasets were removed from the analysis. Demographics and MMSE scores in participants with full datasets are displayed in table 4.2.

Table 4-2 Demographics of all participants with a full data set

Differences between the number of participant and the number of full dataset are explained by a high amount of participants that abandoned the test before completion.

Abbreviations: AD = Alzheimer´s disease, MCI= Mild Cognitive Impairment, SCI= Subjective Cognitive Impairment, HC = healthy control, F= Female, M= Male.

	AD	**MCI** **Mean (SD)**	**SCI** **Mean (SD)**	**HC** **Mean (SD)**
N participant	5	11	9	43
N full data set	1	9	7	38
Age (Years)	75	66.2 (9.2)	71.3 (6.7)	67.2 (7.1)
Education (Years)	17	13.6 (2.1)	14.7 (2.1)	13.9 (2.1)
Gender (Female/Male)	0 F, 1 M	3 F, 6 M	2 F, 5 M	21 F, 16 M
MMSE (Max.30)	24	27.9 (1)	28.6 (1.1)	28.7 (1.2)

4.4.2.2 Motion sickness and joystick use

The six participants who experienced motion sickness on the first piloting phase were invited a second time. None of them had complaints of dizziness or felt nauseous with the new version of the VCT. Steering problems observed in the first piloting session seemed significantly reduced and driving comfort increased.

4.4.2.3 Shortening of the test

The high amount of participants who abandoned the test before the end (almost 20%) indicated that the test was too long and demanding, especially for patients. For this reason, we aimed at reducing the item pool by keeping those showing the best internal consistency. After deletion, internal consistency improved (α 0.789) (see table 4.3).

Table 4-3 Step-wise improvement of the internal consistency

The names of the routes are shortened as follows C1R1 = City 1 Route 1.

Full item set (13 routes) : Cronbach´s α : 0.757		
Deleted Item	Item-Total Correlation	Cronbach´s α without deleted item
C9R3	0.215	0.759
C8R1	0.030	0.771
C7R2	0.030	0.787
C7R1	0.272	0.789
Reduced item set (9 routes) : Cronbach´s α : 0.789		

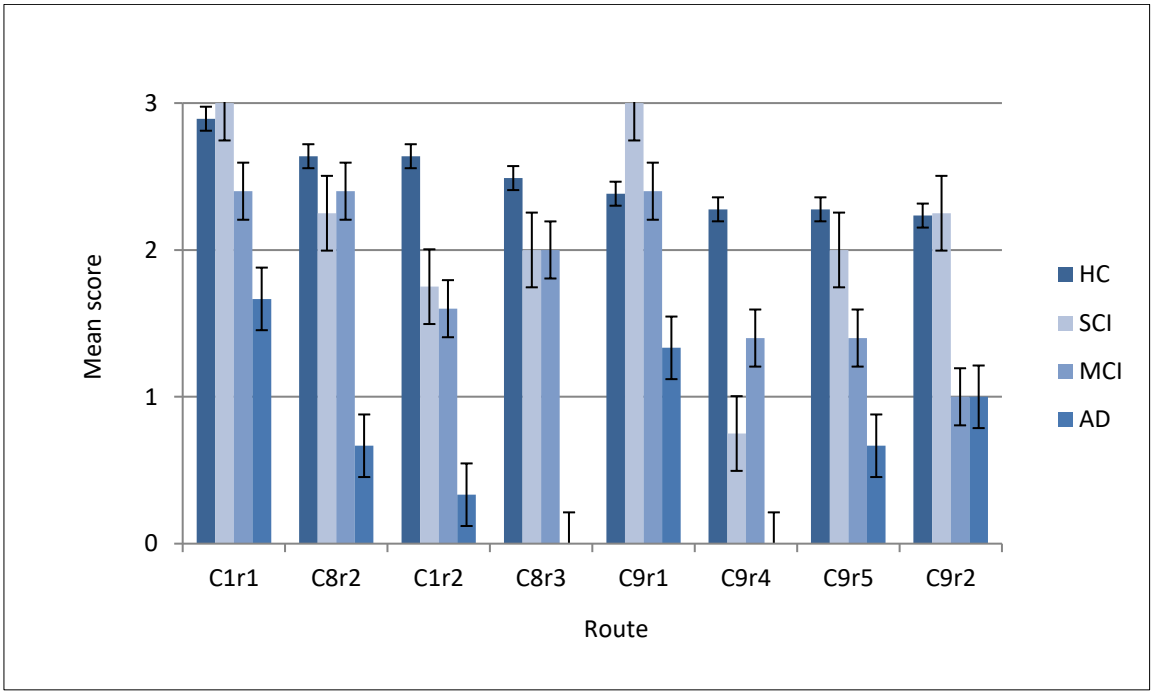
4.4.2.4 Difficulty range

As shown on figure 4.6, the shortened item set preserves graded difficulty in both HC and patients. Ceiling effect in HC and floor effect in more advanced patients (AD) were avoided. The difficulty level seemed improved: most routes were completed by HC in one to two tries on average. As expected AD had the worst performance, followed by MCI compared to HC.

Figure 4-6 Difficulty range of the 9-route Virtual City Task

The routes were ranked order by difficulty level. The name of the routes is shortened as follows C1R1 = City 1 Route 1. The bars represent the standard error of the mean.

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, AD = Alzheimer´s disease.



4.4.3 Discussion

Based on the results obtained in the preliminary piloting, the environment of the VCT, the navigation speed and the recording of the learning phase were modified. No participants suffering from motion sickness in the preliminary piloting experienced it in the modified virtual environment and no participant of the new sample set reported feeling sick. Joystick use was improved thanks to the exclusion of environments with narrow streets as well as the reduction of the navigation speed in the reproduction phase. Although the modifications of the item pool successfully tackled major issues uncovered by the piloting study, a large proportion of participants, especially patients, abandoned the test before completing all routes.

At this stage, the task comprised 13 routes and administration could take up to 60 minutes. The length of administration was assumed to be a reason for the large amount of participants who could not complete the task. Therefore, the item set was reduced to nine routes. Shortening of the item pool lowered the unnecessary high level of difficulty identified in the preliminary piloting while preserving a satisfying graded difficulty in both HC and patients.

Although, the less reliable routes were excluded, internal consistency coefficient only slightly improved. The fatigability and frustration generated by the length of the test might decrease in participant administered with the shortened version of the test, improving the reliability of the test. Administration time for nine routes could, however, take up to 40 minutes. While this length seemed desirable in preliminary piloting of a single test, this is too long to be included in a test battery. For this reason, the item set should be further shortened.

To summarize, the improved material allowed to successfully tackle several problems observed in the piloting phase: joystick use, motion sickness and difficulty level. A primary shortening of the task allowed reducing administration time while preserving graded difficulty. A shorter version of the test should, however, be created to be included in the DZNE-Cog.

4.5 Final item set for the Virtual City Task

Previous work on the Virtual City Task (VCT) allowed improving the material to obtain graded difficulty and avoid floor and ceiling effects. With a first reduction of the item set, the administration time could be reduced from 60 to 40 minutes. The task being part of a larger battery, a reduced administration time is a necessity for its applicability in memory clinics. Memory tests being demanding, it is important to keep the administration time short to avoid fatigue interfering with the performance and reducing reliability of the results. Therefore, a final reduction of the item set is presented in this chapter.

4.5.1 Method

4.5.1.1 Participants

Sixty-eight participants (48 HC, 7 AD, 8 MCI, 5 SCI) were administered the nine routes of the VCT (see section 2.3 for further information).

Administration procedure and scoring method are the same than in chapter 4.4.

4.5.1.2 Statistical considerations

The shortening of the item pool to reduce administration time was done in two steps. Firstly, the item-total correlation was calculated. In classical test theory, a good item-total correlation indicates that all items measure the same construct. Routes that had a lower item-total correlation were discarded.

Secondly, obtaining graded difficulty being a primary aim of the test, items of different levels of difficulty, that would discriminate best patients from HC were qualitatively selected.

4.5.2 Results

4.5.2.1 Demographics

Demographics and general neuropsychological scores are displayed in table 4.4.

Table 4-4 Demographics and neuropsychology

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, AD = Alzheimer´s disease, F= Female, M= Male.

	HC Mean (SD)	**SCI Mean (SD)**	**MCI Mean (SD)**	**AD Mean (SD)**
Age (years)	65 (8.3)	74.6 (3.8)	76.2 (5.5)	73.1 (5.5)
Education (years)	14.8 (2.5)	13.7 (2.5)	12.9 (3.1)	12.2 (1.3)
Gender	3 F, 4 M	4 F, 4 M	4 F, 1 M	30 F,18 M
MMSE (Max.30)	28.4 (1.4)	28.4 (0.6)	26.7 (3.3)	24.4 (1.5)
BNT (Max. 30)	27.6 (2.5)	26.6 (1.5)	23.3 (6.5)	24 (4.1)
ROCF Copy (Max. 36)	32.9 (2.7)	32.4 (2.7)	32.4 (3.8)	22.1 (11.6)
ROCF immediate recall (Max. 36)	17.7 (6)	17.2 (4.7)	15.5 (8)	3.9 (2.1)
ROCF delayed recall (Max. 36)	17.4 (5.9)	16.4 (4.1)	13.6 (9.6)	3 (2.3)
GDS (Max. 15)	1.1 (1.4)	3 (4.6)	2 (2)	0.3 (0.6)
MOCA (Max.30)	26.7 (2.5)	26.3 (1.1)	18.8 (2.6)	16 (3.9)

4.5.2.2 Shortening to 5 routes

Ten datasets (1 HC, 4 AD, 3 MCI, 2 SCI) were excluded from the analysis because of missing data. The test being, at this stage, administered in a randomized manner, uncompleted routes were not always the hardest routes and could not automatically be quoted as “failed”. Therefore, incomplete datasets were removed from the analysis

All items displayed a good item-total correlation (see table 4.5). This indicates that all routes seemed to measure the same latent construct.

Table 4-5 Item-total correlation for the 9 routes

	Item-Total Correlation
C1R1	0.615
C1R2	0.693
C7R3	0.637
C8R2	0.643
C8R3	0.600
C9R1	0.324
C9R2	0.528
C9R4	0.656
C9R5	0.660

Figure 4.7 displays the average score of both HC and patients on all of the nine routes. Based on the results obtained by HC and patients, the item set was qualitatively reduced to maintain a good graded difficulty while choosing routes that can differentiate patients from HC. Figure 4.8 displays the average score of HC and patients in the five retained routes. Routes C9R1, C9R5 and C9R4 have similar results in HC but seem to show a range of difficulty in patients and were therefore included. The reduced item set displayed a satisfying reliability coefficient (α: 0.773).

Figure 4-7 Difficulty range of the 9-route Virtual City Task

The bars represent the standard error of the mean.

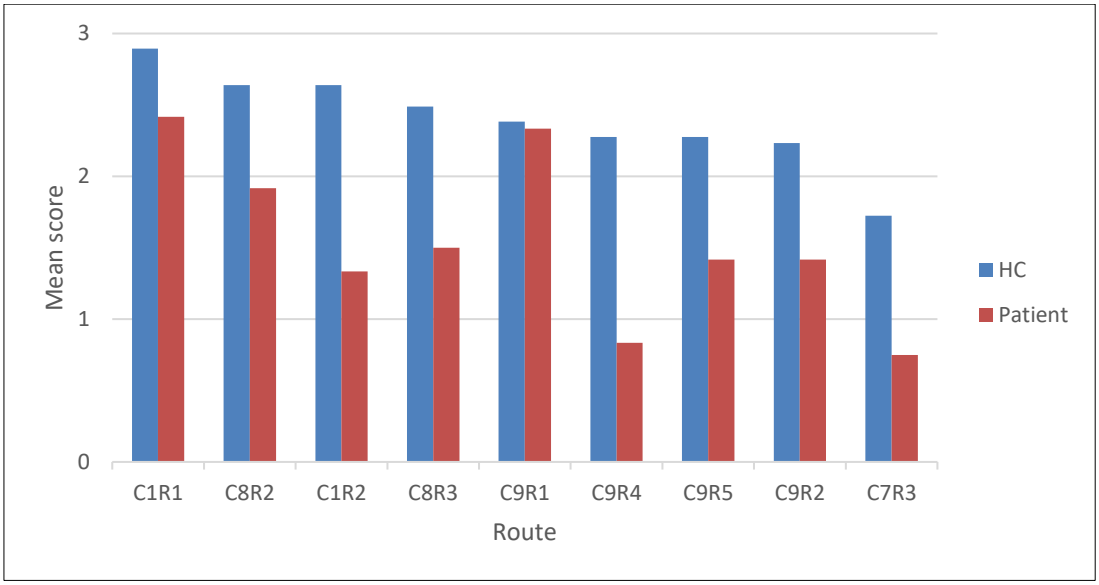
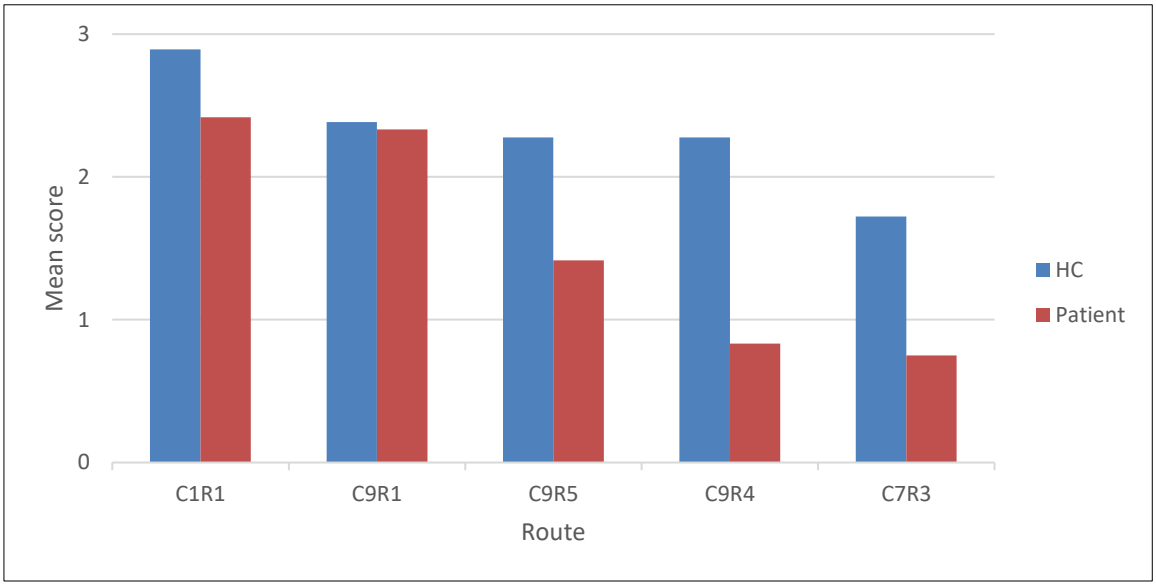


Figure 4-8 Difficulty range in the 5 retained routes

The bars represent the standard error of the mean.



4.5.3 Discussion

Administration time is a major constraint in the clinical field where the VCT, as part of a broader battery, will be used. Moreover, the task requiring sustained attention and involving high memory load, it is important to keep it short to avoid fatigability and frustration to interfere with the reliability of the measure.

Although previous shortening of the VCT lowered administration time from 60 to 40 minutes by reducing the item set from 13 to nine routes, 14.7% of the participants, in majority patients, interrupted the test before the end. As in previous piloting, participants abandoned the task because of fatigue and frustration. Despite, the slight reduction in abandon rate (from 20% to 14.7% by reducing from 13 to nine routes), the nine-route-version of the task was too long and demanding. Therefore, a final shortening of the item set from nine to five routes was operated.

The five-route-version of the task allowed reducing administration time to an average of 18 minutes. Although, 18 minutes is still too long to be included in a larger battery, a procedure inspired from Adaptive Testing would allow administering only the items that fit the participant´s level and therefore reduce administration time while maximizing the obtained information.

The five chosen routes display an acceptable but slightly low internal consistency coefficient for using longitudinally. It presents, however, a good gradual difficulty that is

advantageous for tracking early cognitive changes in participants with different premorbid cognitive abilities. The internal consistency coefficient was, however, generated from data obtained in the nine-route-version of the task. As mentioned before, the length of the test possibly biased the data because it induced fatigue and frustration. Therefore, internal consistency might improve when the five-route-version of the task is administered.

To summarize, a final shortening to five routes enabled reducing the administration time while preserving the test´s properties (graded difficulty, absence of floor and ceiling effects, separation of patients and HC).

4.6 Cross-sectional comparison

In this chapter, the cross-sectional administration of the Virtual City Task (VCT) to healthy elderly (HC) as well as patients with AD, MCI, and semantic variant of primary progressive aphasia (SvPPA) is presented. The performance obtained across groups in the VCT is contrasted with those obtained on the recall of the Rey-Osterrieth complex figure (ROCF) (Rey, 1941), a gold standard for the evaluation of non-verbal memory in memory clinics. Both AD and MCI are expected to be impaired compared to HC while SvPPA are expected to be preserved on the VCT. Indeed, the latter are thought to be, at least in the first years, preserved on topographical memory test (Pengas et al., 2010).

Repeated assessments are ubiquitous in memory clinics, where the presence of a neurodegenerative process has to be determined. An improvement in participant´s score in neuropsychological tests can be found on repeated assessments, which is not attributable to an improvement in cognitive abilities. It is mostly due to familiarity with the test material and/or the development of test-taking strategies. These so-called practice effects can mask early mild cognitive decline and thus, lowers the test sensitivity to early dementia stages (Salthouse & Tucker-Drob, 2008). Cognitive tests, and memory tests especially, are often subject to practice effects (Benedict & Zgaljardic, 1998; Crawford, Stewart, & Moore, 1989). The primary aim of the VCT, as part of a larger battery, being to track early cognitive changes, we aimed at creating a parallel version of the test that would be of equal difficulty to lower practice effects in repeated assessment. One of the main caveats when creating an alternate version is to obtain a similar level of difficulty. Indeed, many parallel versions are known to have a different level of difficulty than the original version (Lebedeva, Huang, & Koski, 2016) lowering their sensitivity to slight changes in retest situation. The creation of a parallel version as well as its administration to a sample of healthy elderly are presented in this chapter.

4.6.1 Method

4.6.1.1 Cross-sectional comparison

4.6.1.1.1 Participants

Forty-five participants (20 HC, 10 AD, 11 MCI and 4 SvPPA) were administered the 5-routes version of VCT presented in chapter 4.5 (see section 2.3 for further information on inclusion criteria). Six AD participants were excluded because of incomplete results. The

test being, at this stage, administered in a randomized manner, uncompleted routes were not always the hardest routes and could not automatically be quoted as “failed”. Therefore, incomplete datasets were removed from the analysis.

4.6.1.2 Parallel version

4.6.1.2.1 Participants

Thirty healthy German participants over 50 years were included (see section 2.3 for further information on inclusion criteria).

4.6.1.2.2 Material

To create a parallel version that has the same level of difficulty but that lowers practice effect in retest, the same environments were used in the parallel version but all buildings were replaced. Thereby, the participant drove the exact same route but the environment looked different. We assumed it would prevent him/her from remembering the route.

4.6.1.2.3 Administration

Test-retest occurred at one week of interval. Ten participants were administered the original version at the first visit, and the parallel version on the second visit whereas ten other participants were at first administered the parallel version followed on the second visit by the original version. In order to quantify practice effects when using the same version of the VCT, ten further participants were administered the original version on both visits.

4.6.1.3 Statistical considerations

Descriptive statistics were run using SPSS version 21 (IBM corp., Armonk, NY).

Given the small sample size Kruskal-Wallis test was run to compare the demographics and the performance in neuropsychological tests across groups. Dunnett´s T3 adapted for samples with unequal variance was used for post-hoc comparisons.

Cronbach´s alpha was computed to quantify internal consistency. Spearman´s correlation was used to compare the results obtained in the VCT and the immediate and delayed recall of the ROCF as well as to assess test-retest reliability of the VCT.

4.6.2 Results

4.6.2.1 Cross-sectional comparison

4.6.2.1.1 Demographics and neuropsychology

Demographics and performance in MMSE for all patients and HC are presented in table 4.6. Noteworthy is that AD participants who did not perform the whole test, had on average a lower MMSE score than those who did (18.7 against 26).

Table 4-6 Demographics and neuropsychology

** p<.05 compared to the control population, [A]p<.05 compared to MCI, [B]p<.05 compared to AD, [C]p<.05 compared to SvPPA. Bold italics alphabets indicate significance survives Bonferroni correction for multiple comparisons.*

Abbreviations: HC = healthy control, MCI= Mild Cognitive Impairment, AD = Alzheimer´s disease, SvPPA= Semantic variant of Primary Progressive Aphasia, F= Female, M= Male, n.s = non-significant.

	HC Mean (SD)	**MCI Mean (SD)**	**AD Mean (SD)**	**SvPPA Mean (SD)**	**Significance *P-value***
Age (years)	71.9 (3.1)	72.5 (5.2)	71.7 (10.2)	65.5 (3.1)	*n.s*
Education Level (years)	15 (2.4)***AB***	10 (1.8)*	9 (1.1)*C	15.5 (2.4)B	*$X^2(3, 39) = 25.9$ p<0.001*
Gender	9 F, 11 M	5 F, 6 M	3 F, 1 M	2 F, 2 M	*n.s*
MMSE (Max. 30)	28.6 (1.7)B	26 (2.2)	26 (1.4)*	16.7 (9.3)	*$X^2(3, 39) = 18.6$ p<0.001*
Score VCT (Max. 15)	11.1 (2.9)A***B***	7.3 (3.4)*C	4.2 (0.5)****C***	11.7 (0.5)A***B***	*$X^2(3, 39) = 17$ P=0.001*
Immediate recall ROCF (Max. 36)	19.7 (7)AB	8.7 (5.9)*	8.1 (3.5)*	14 (2.8)	*$X^2(3, 37) = 18.5$ p<0.001*
Delayed recall ROCF (Max. 36)	19.5 (7)AB	8.3 (5.1)*	9.4 (3.5)*	11.5 (6.4)	*$X^2(3, 37) = 17.9$ p<0.001*

4.6.2.1.2 Internal consistency

The five routes displayed a good internal consistency coefficient $\alpha = 0.8$.

4.6.2.1.3 Performance

As presented in table 4.6 both MCI and AD had significantly lower performance in the VCT than HC and SvPPA. No significant difference was found between MCI and AD. The AD sample size was, however, restricted. SvPPA performed normally.

As displayed in table 4.6 there was a significant difference in the total score obtained in both the immediate and the delayed recall of the ROCF between HC and both MCI and AD. Again, no significant difference was found between MCI and AD. SvPPA performed normally.

The performance obtained in the VCT significantly correlated with both the immediate and delayed recalls of the ROCF (respectively Spearman´s roh: 0.73, $p<0.001$ and Spearman´s roh: 0.71, $p<0.001$).

4.6.2.2 Parallel version

4.6.2.2.1 Demographics and neuropsychology

Demographics and neuropsychology for HC administered with the original and parallel versions are displayed in table 4.7. No significant difference was found between groups for age, education level or MMSE and ROCF scores.

Table 4-7 Demographics and neuropsychology

Participants who had the original version at baseline, and the parallel version the second visit as well as participants who had the parallel version at baseline and the original version one the second visit were combined in one group and presented in the column "parallel version".

Abbreviations: n.s = non-significant.

	Same Version Mean (SD)	**Parallel Version Mean (SD)**	**Significance *P-value***
N	10	20	*n.s*
Age (years)	71.8 (3)	71.2 (3)	*n.s*
Education (years)	15.1 (2.1)	14.4 (2.5)	*n.s*
MMSE (Max. 30)	28.7 (1.4)	28.8 (1.6)	*n.s*
Copy ROCF (Max. 36)	34.5 (1.6)	33.9 (1.6)	*n.s*
Immediate recall ROCF (Max. 36)	20.7 (6.2)	21.1 (6.9)	*n.s*
Delayed recall ROCF (Max. 36)	20.2 (6.5)	20.7 (6.6)	*n.s*
T1 VCT (Max. 15)	11.9 (2.6)	10.7 (2.9)	*n.s*
T2 VCT (Max. 15)	12.5 (1.3)	11.2 (2.6)	*n.s*

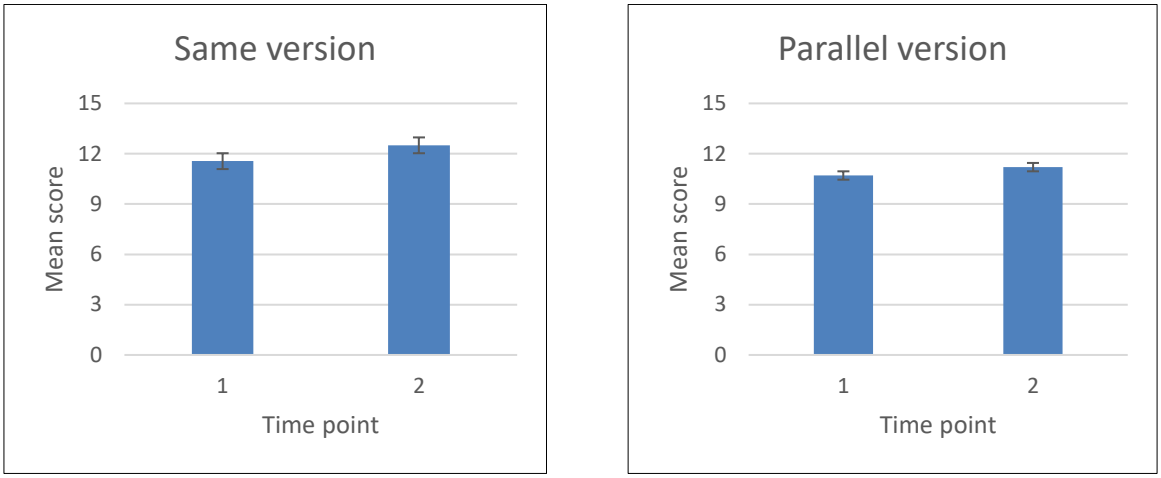
4.6.2.2.2 Performance

Participants had on average 10.7 points (SD: 2.9) on the first visit and 11.2 points (SD: 2.6) on the second visit when given parallel version whereas participants given the same version on both time points had respectively 11.9 (SD: 2.6) and 12.5 points (SD: 1.3) (see table 4.7).

No difference in practice effect was found in the two groups. Participants improved of 0.6 points when administered the same version and 0.5 points when given the parallel version at the second visit (see figure 4.9).

Figure 4-9 Performance of healthy elderly in the Virtual City Task on two different sessions

The bars represent the standard error of the mean.



4.6.2.2.3 Test-retest reliability

The administration of both the same (Spearman´s roh: 0.85) and the parallel versions (Spearman´s roh: 0.76) displays a good test-retest reliability.

4.6.3 Discussion

4.6.3.1 Cross-sectional comparison

In memory clinics, memory performance is routinely assessed with the recall of a word list on the verbal modality or a figure on the visual modality. Topographical memory is usually not tested and is mostly only briefly mentioned during anamnesis. We compared the results obtained in a routine non-verbal memory test, the recalls of the ROCF, with the results obtained in a test of topographical memory in three groups of patients and healthy elders.

As expected AD and MCI were impaired compared to HC and SvPPA in the VCT. These results are in keeping with previous findings showing that the assessment of topographical memory allows differentiating AD and MCI from SvPPA (Bird et al., 2010; Pengas et al., 2010). The results obtained were similar for the recalls of the ROCF: MCI and AD were impaired compared to HC and SvPPA performed normally. The VCT only was able to differentiate significantly between SvPPA and the other groups of patients. In both recalls of the ROCF, no significant difference between SvPPA and MCI or AD was found. Although the recalls of the ROCF and the VCT were both designed to test memory in the non-verbal modality, the latter targets especially topographical memory. Both tests are therefore not fully comparable. Performance in both tests is, nonetheless, similar in our sample and correlate significantly. This indicates that they do triangulate a core memory

deficit. In a meta-analysis comparing several routine neuropsychological tests, the delayed recall of the ROCF, although not sufficient for accurate diagnosis, seemed to be one of the best test to discriminate between AD and fronto-temporal dementia (all types mingled) (Hutchinson & Mathias, 2007). The VCT, however, seem to be more sensitive for the differential diagnosis of SvPPA versus AD spectrum. The patient groups being very restricted further work should include more participants to confirm this trend.

The small amount of AD participants included in the analysis is mostly due to the fact that many patients did not perform the whole test because they were too impaired. The high rate of abandon was likely due the randomized administration order of the routes. A patient could start the test with a hard route leading to frustration and fatigue, and eventually to abandoning the task. The test, as part of a broader battery, was developed to track early cognitive changes and does not target primarily patients in advanced stages but it would be interesting to obtain a test that can be given to a larger spectrum of impairment. Future work should abandon the randomized administration order and present the routes ranked order by difficulty. Furthermore, future version of the battery might profit from implementing computerized adaptive testing to minimize administration time while maximizing obtained information.

Finally, the VCT, although giving similar results in our sample than the ROCF, offers the advantage of being free of both intra and inter-rater variability. Indeed, inter-rater reliability particularly was shown in various range in the ROCF (Carr & Lincoln, 1988; Liberman & Stewart, 1994). Both the administration and the rating of the VCT are fully computerized, maximising standardization across different raters and centres.

4.6.3.2 Parallel version

In the context of dementia screening, repeated measurements are recurrent to monitor neurodegenerative processes. Therefore, we intended to create an alternate version of the VCT that would help minimize practice effects on retest. The results show that the task displays good test-retest reliability, both when using the same and alternate versions. This is of considerable importance when used repeatedly. As expected, since the routes stayed unchanged, both versions have a comparable level of difficulty.

A slight practice effect is, however, present both when the same version or the alternate version is administered twice. Therefore, the alternate version does not seem to tackle the issue of practice effect. Practice effects is a recurrent problem in the neuropsychological

field that is often present even when using parallel versions (Beglinger et al., 2005). The sample being small, and rather homogeneous, it might be interesting to include participants with cognitive impairment to extend the findings.

The development of further alternate versions might benefit from finer remodelling of the virtual environment. Practice effects might be reduced more effectively not only by replacing the facades of the buildings, but also by changing the ground, the sky and lightening. Noteworthy is that both administrations of the test were led at an interval of a maximum of 10 days for logistical reasons. In memory clinics, patients are mostly tested every 6 to 12 months, reducing the impact of practice effects. Further analysis should be led at a pace that is more similar to clinical routine.

In summary, the VCT seems to replicate findings showing that topographical memory is able to differentiate between the AD spectrum and SvPPA. Although not directly equivalent, it offers some benefits over the ROCF and could be implemented in clinical routine. Further work should improve the alternate versions and allow the administration to moderately demented patients.

4.7 Longitudinal assessment of the Virtual City Task

The Virtual City Task (VCT) was created to be part of a neuropsychological battery that would track changes from normal cognition to early dementia in a single subject. This implies obtaining a test that would be sensitive to changes in cognitive abilities over time. Previous work allowed obtaining an item set of graded difficulty, short to administer (18 minutes) and that offers benefits to differentiate SvPPA from AD and MCI over the recalls of the Rey-Osterrieth Complex Figure (ROCF), the gold standard in non-verbal episodic memory testing. In this section, the results of the longitudinal administration of the VCT to healthy elderly but also AD, MCI and SCI are presented. Performance obtained are compared to those obtained in the immediate and delayed recalls of the ROCF.

4.7.1 Method

4.7.1.1 Participants

One hundred three participants (69 HC, 15 SCI, 16 MCI and 3 AD) were administered the VCT on two different sessions (see section 2.3 for inclusion criteria). Twelve participants (2 AD, 5 MCI, 1 SCI and 4 HC) had incomplete datasets because they abandoned before completing all routes. The routes being administered in a randomized manner, uncompleted routes were not always the hardest routes and could not automatically be quoted as “failed”. Therefore, incomplete datasets were removed from the analysis

4.7.1.2 Progressor vs Stable groups

As for the longitudinal analysis of the results obtained in the Graded Object Naming Task, participants were allocated to a group of progressor or stable participants (see section 3.8 for further information). This allowed to capture the performance of participants enrolled as HC but that showed cognitive decline in the retest. A participant, regardless the diagnostic status, who lost 0.5 or in a second variant 1 Z-score between the last visit and the baseline in the MMSE (Folstein et al., 1975) was included in the progressor group The main aim of the test being to track subtle changes in individuals and not differential diagnosis, the distinction between progressor and stable participants was chosen over the diagnostic status. Our sample was composed of 64 stable and 27 progressor participants when using the 0.5 Z-score loss criterion and 78 stable and 13 progressor participants when using the 1 Z-score loss criterion.

4.7.1.3 Neuropsychological assessment

In a single testing session of approximatively 90 minutes, Patients and HC were administered a battery of standard neuropsychological tests listed in table 3.6 in section 3.8.1. The VCT was administered during the same session. The average time lapse between the two testing sessions was of 6.8 months (SD: 1.8).

4.7.1.4 Statistical considerations

Normality was assessed with the Kolmogorov-Smirnov test and the homogeneity of the variances was assessed using Levene´s test. Given the non-equality of the variance and the non-normality of the distribution of the demographics and performance in the neuropsychological tests, a Kruskal-Wallis test was run to compare the different groups. The gender division was assessed with a Pearson Chi-Square test. The Dunnett´s T3 test was used for post-hoc comparisons.

R version 3.1.1 (R Core Team, 2014) and lme4 version 1.1.13 (Bates et al., 2015) were used to perform a linear mixed effects analysis of the relationship between the performance in the VCT, the time point and the status of cognitive stability (see section 3.8.1.2 on progressor vs. stable). An interaction of the time point and the status of cognitive ability was entered as a fixed effect into the model. Intercepts for participants and time lapse between two time points were entered as random effects. P-values were obtained by likelihood ratio tests of the full model with the questioned effects, against the model without the effects or interaction questioned. The same procedure was applied to the performance obtained in the immediate and delayed recalls of the ROCF.

4.7.2 Results

4.7.2.1 Demographics and general neuropsychology

Table 4.8 presents the demographics and MMSE results for the HC and patients involved in the longitudinal study. No significant difference could be found between the different groups in age, education level or MMSE. Only one AD participant, however, was left after excluding incomplete datasets.

Table 4-8 Demographics and MMSE performance at baseline

Abbreviations: HC = healthy control, SCI= Subjective Cognitive Impairment, MCI= Mild Cognitive Impairment, AD = Alzheimer´s disease, F= Female, M= Male, n.s = non-significant.

	HC Mean (SD)	**SCI Mean (SD)**	**MCI Mean (SD)**	**AD Mean (SD)**	**Significance *P-value***
Age (years)	66.4 (8.1)	69.4 (8.6)	71.4 (7.2)	65	n.s
Education (years)	11.2 (2.1)	10.9 (1.9)	10.2 (1.7)	10	n.s
Gender	41 F, 24 M	8 F, 6 M	2 F, 9 M	1 F, 0 M	X^2 (3, 91) = 8.5, p=0.037
MMSE (Max. 30)	28.5 (1.4)	28.8 (1)	28.3 (1.3)	25	n.s
N	65	14	11	1	

Table 4.9 presents the demographics for the participants based on the 0.5 and 1 Z-score loss criteria. In the 0.5 Z-score loss criterion, progressor participants had a significantly higher level of education than stable participants. No other comparison between groups was significant.

Table 4-9 Demographics

Abbreviations: n.s = non-significant.

	Demography					
Criterion for progression	-0,5 Z-score			-1 Z-score		
Group	**Stable Mean (SD)**	**Progressor Mean (SD)**	**Sig. *P-value***	**Stable Mean (SD)**	**Progressor Mean (SD)**	**Sig. *P-value***
Age at baseline (years)	66.3 (8.8)	70.2 (5.7)	n.s	66.9 (8.4)	70.7 (5.3)	n.s
Education (years)	10.7 (1.8)	11.9 (2.3)	X^2 (1, 91) = 5.3, *p=0.021*	11 (2)	11.7 (2.3)	n.s
Gender	41 F, 23 M	11 F, 16 M	n.s	46 F, 32 M	6 F, 7 M	n.s
	Sub-groups					
N	64	27		78	13	
AD	0	1		0	1	
MCI	3	8		10	1	
SCI	12	2		12	2	
HC	49	16		56	9	

4.7.2.2 Change in Virtual City Score

Figure 4.10 shows the change in the VCT and the MMSE score over two time points in our sample. In both classifications, stable participants seem to slightly improve on the second visit in the MMSE whereas progressor participants slightly declined. In the VCT, stable participants seem to improve in both classifications whereas progressor participants seem to stay relatively stable over time.

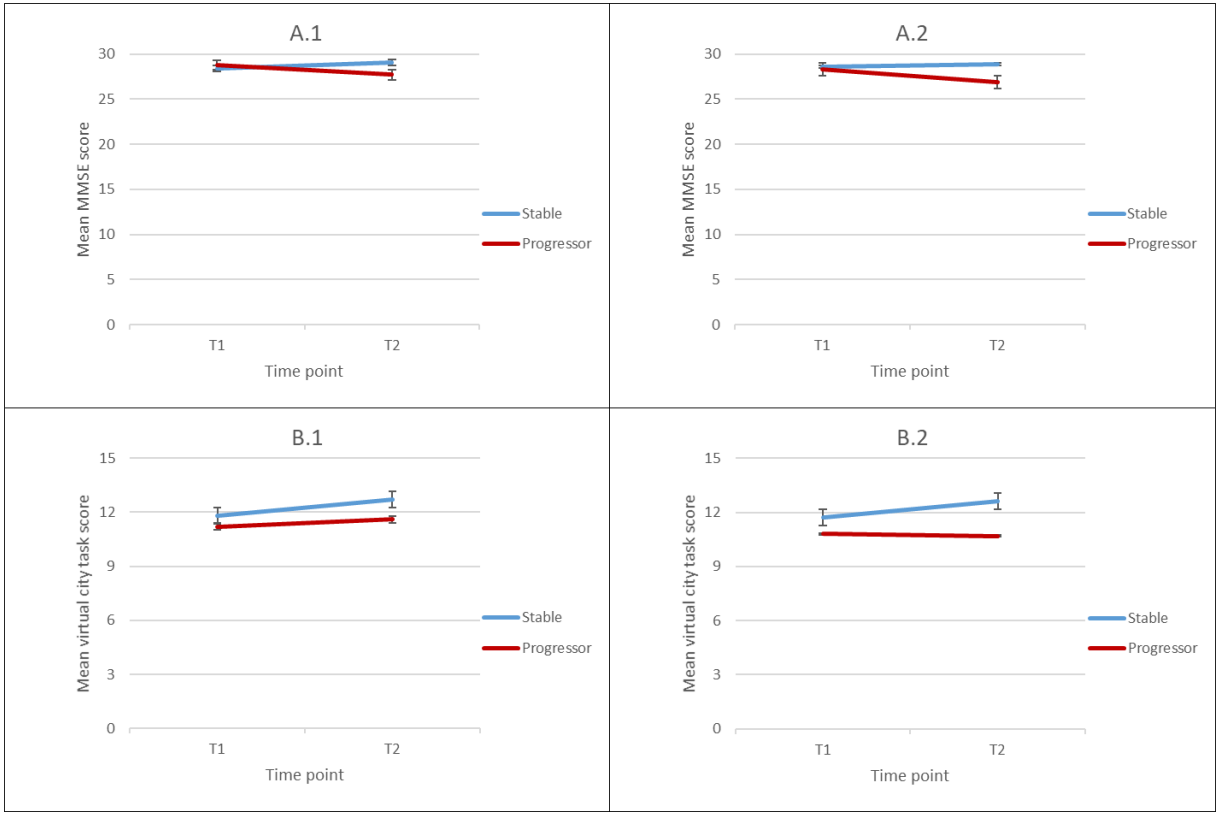
Figure 4-10 Performance of progressor and stable groups in the MMSE and VCT over time

A.1 represents changes in the MMSE score over time in the 0.5 Z-score loss classification

A.2 represents changes in the MMSE score over time in the 1 Z-score loss classification

B.1 represents changes in the VCT score over time in the 0.5 Z-score loss classification

B.2 represents changes in the VCT score over time in the 1 Z-score loss classification



In the classification of progressor versus stable participants on the base of a 0.5 Z-score loss between baseline and the last available visit, there was only a significant effect of the time point (p=0.03) on the performance in the VCT over two time points. No effect of the group (p=0.11) or interaction effect between the group and the time point (p=0.37) could be found.

When using the classification based on a 1 Z-score loss between baseline and the last available visit, there was a significant effect of the group (p=0.009) and the time point (p=0.01) on the performance in the VCT. Only a trend toward a significant interaction effect between the group and the time point (p=0.09) was found.

4.7.2.3 Change in ROCF

Figure 4.11 shows the change in the score of both the immediate and the delayed recalls of the ROCF over two time points in our sample. In both classification criteria and both groups, the score in the immediate and delayed recall of the ROCF improved at the second visit. The baseline score, however, was lower in the progressor group in both recalls and both classifications.

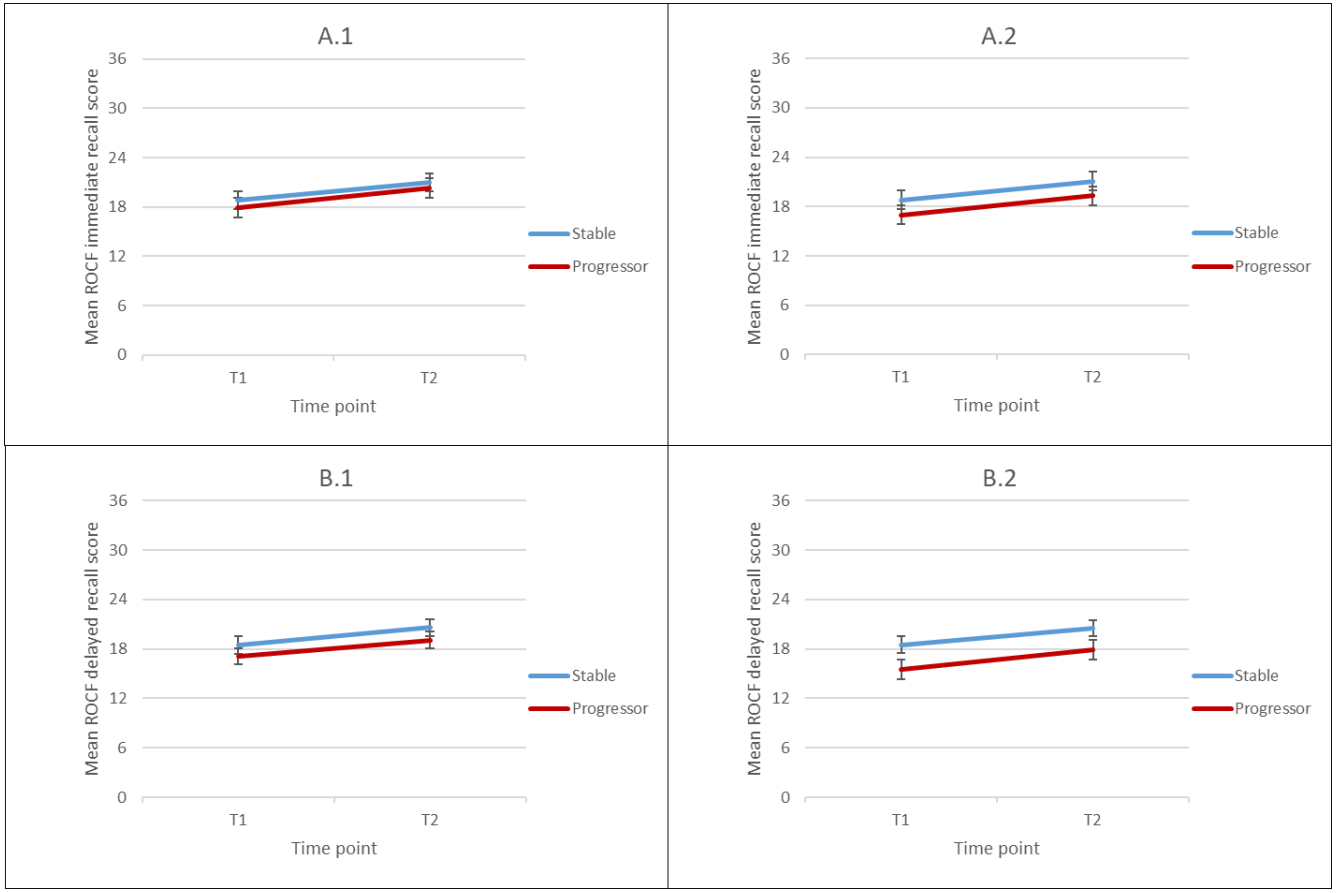
Figure 4-11 Performance of progressor and stable groups in the ROCF over time

A.1 represents changes in the ROCF immediate recall score over time in the 0.5 Z-score loss classification

A.2 represents changes in the ROCF immediate recall score over time in the 1 Z-score loss classification

B.1 represents changes in the ROCF delayed recall score over time in the 0.5 Z-score loss classification

B.2 represents changes in the ROCF delayed recall score over time in the 1 Z-score loss classification



4.7.2.3.1 ROCF immediate recall

In the classification of progressor versus stable participants on the base of a 0.5 Z-score loss between baseline and the last available visit, there was only a significant effect of the time point (p=0.048) on the performance in the immediate recall of the ROCF over two time points. No effect of the group (p=0.84) or interaction effect between the group and the time point (p=0.88) could be found.

When using the classification based on a 1 Z-score loss between baseline and the last available visit, there was again only a significant effect of time point (p=0.048) on the performance in the immediate recall of the ROCF. No significant effect of the group (p=0.58) or interaction effect between the group and the time point (p=0.97) could be found.

4.7.2.3.2 ROCF delayed recall

In the classification of progressor versus stable participants on the base of a 0.5 Z-score loss between baseline and the last available visit, there was only a significant effect of the time point (p=0.03) on the performance in the delayed recall of the ROCF over two time points. No effect of the group (p=0.47) or interaction effect between the group and the time point (p=0.92) could be found.

When using the classification based on a 1 Z-score loss between baseline and the last available visit, there was again only a significant effect of time point (p=0.03) on the performance in the delayed recall of the ROCF. No significant effect of the group (p=0.17) or interaction effect between the group and the time point (p=0.67) could be found.

4.7.3 Discussion

This work aimed at proofing the ability of the VCT to track cognitive decline over time and comparing it with a gold standard for the evaluation of nonverbal episodic memory, namely the recalls of the ROCF.

When comparing the results over two visits, only the VCT showed a trend toward a significant interaction effect between the stability status and the time point when using the 1 Z-score loss in MMSE classification. As mentioned in section 3.8.3, a loss of 0.5 Z score in MMSE corresponds to a loss of a bit more than one point in 6 months, which might be fortuitous. This less strict classification criterion might have led to the inclusion of stable participants in the progressor group, reducing the power of the analysis. Although the significance threshold for an interaction effect between the time point and the group in the

stricter classification was not reached, the results are encouraging. Indeed, the observable trend might be confirmed with a supplementary time point: only 6 months were perhaps too short to capture a slight cognitive decline. With the Graded Object-Naming Task, using the same method, a significant change in performance could only be seen when using three time points.

In the VCT, the stable group showed a score improvement on the second visit. This is most likely due the practice effect already mentioned in section 4.6.3. The score on the second assessment seem at the group level unchanged for progressor participants. In this group, practice effects might have masked a slight decline. Another hypothesis is that progressor participants did not profit from repeated assessment as much as stable participants. Some authors found, for example, that when tested repeatedly on the same day, MCI showed an attenuated practice effect compared to HC (Darby, Maruff, Collie, & McStephen, 2002). The absence of practice effect is an important cue when trying to differentiate between healthy and impaired elders.

A significant effect of the group was only found in the VCT when using the 1 Z-score loss criterion. When using the stricter classification criterion, there was a difference between stable and progressor at baseline. This difference was less marked in the 0.5 Z-score loss criterion and in the recalls of the ROCF. This might confirm the superiority of the VCT over the ROCF for cross-sectional separation of healthy elderly and their impaired counterparts.

A major limitation of this study relies in both the limited number of participants and visits. Due to logistical reasons, participants included in the longitudinal analysis were tested in parallel with the shortening of the city. In other words, some participants were administered nine routes at the first visit and only five at the second visit. This might have affected largely the results obtained as both visits were not fully comparable Moreover, the number of participants declining in the MMSE at the second visit and included in the progressor group was small, especially when using the 1 Z-score loss criterion. These analyses being part of a preliminary work on the test, the conclusions should be confirmed with a sample administered exclusively with the shortened version on all visits. Moreover, administering the routes rank-ordered by difficulty might reduce the number of dataset that have to be excluded from the analysis. The non-completion of a route could be quoted as failed and therefore all data could be analysed.

Finally, the use of the joystick being possibly stressful for some elderly, it might be profitable in further longitudinal analysis to use the second visit as baseline. Indeed, it might have an impact on their performance on the first time point that would decrease the power of the analysis.

To summarize, the results show a trend of the VCT to track cognitive change over time. The restricted sample size, number of visit as well as the parallel shortening of the item set are, however, major caveats that might have underestimate the value of the task. Further work is needed to further explore the psychometric qualities of the VCT.

4.8 Conclusion

The preliminary work on the Virtual City Task (VCT) showed the ability of five routes to cover a wide range of difficulty in both healthy and impaired elderly. Its superiority over the immediate and delayed recalls of the ROCF in differentiating AD and MCI patients from SvPPA and healthy elderly has been confirmed. Moreover, its full-computerized administration avoids inter-rater divergence reported in the ROCF, the current gold standard in the evaluation of non-verbal episodic memory.

Step-wise reduction of the item set allowed obtaining a test that would be short to administer, although too long to be part of a larger battery. Further work on the VCT should implement computerized adaptive testing.

Longitudinal analyses showed a trend toward an ability to show change over time that was not found in the ROCF. Important logistic limitations (small sample size, only two time points) might, however, have undermined the results and further work is needed to confirm this trend. Further longitudinal analyses would gain at improving the existing parallel version of the test to reduce practice effects.

Finally, although the test was constructed with a view to be usable on different international population, this was not directly confirmed. At the time of writing, the test is administered in Australia, and the collected data should allow confirming the international applicability of the VCT.

Chapter 5.

Conclusions and future directions

Two tests of the DZNE-Cog, a battery designed to track accurately and sensitively early cognitive changes have been presented: the Graded Object Naming Task (GONT) for language assessment and the Virtual City Task (VCT) for memory assessment.

As intended, both tests are short and easy to administer. Floor and ceiling effects were avoided and inter-rater divergence tackled thanks to a fully computerized administration and strict quotation criteria. A cross-sectional superiority over the gold standard in the domain, respectively the BNT and the ROCF was confirmed. Finally, both tests are internationally usable, although this would need to be verified in a non-German sample for the VCT.

Longitudinal analyses on the tasks have shown qualities in tracking change in early cognitive decline. The GONT was able to significantly track cognitive decline in elderly that had a drop in MMSE performance over three time points. Only a trend for sensitive tracking of cognitive decline was found in the VCT, albeit important logistic limitations might have undermined the results obtained. Indeed, only two time points were available and the sample size was limited.

Further work with more participants and time points should corroborate the trends obtained in the single tests. Moreover, larger dataset would allow investigating the psychometric properties of the tests in more depth with the mean of Item Response Theory. Furthermore, the complete battery in its final form (including the attention and visuo-perceptual tests) should be administered to a new sample of healthy controls and patients, to confirm its ability to show change in the four main domains of cognition. The obtained data would allow creating a composite score usable to appreciate change in cognitive abilities. With larger samples, it would be interesting to implement computerized adaptive testing where

participants are given items adapted to their capacities. Indeed, it would maximize the information obtained in a minimized amount of time.

To summarize, the preliminary work on the language and memory tasks of the DZNE-Cog yield encouraging results. Major limitations have, however, been pointed out and further work is needed to improve the tasks and confirm their qualities.

Chapter 6.

Investigating Primary Progressive Aphasia

6.1 Introduction

Aphasia is a disturbance of the expressive and/or receptive language with great implications for patients´ life quality (Lee, Lee, Choi, & Pyun, 2015). Model of acquired aphasia, and diagnostic tools based on these models, often rely on stroke related aphasia characterized by a sudden onset and an impairment profile depending on the concerned vascular territory (Yourganov, Smith, Fridriksson, & Rorden, 2015). Stroke aphasia goes typically through a recuperation and stabilisation phase, leaving patients with more or less sequelae in the different language domains (El Hachioui et al., 2013). Language impairment is evenly a major feature occurring in the course of many neurodegenerative diseases like the dementia of Alzheimer´s type (Weintraub, Wicklund, & Salmon, 2012). Unlike in stroke related aphasia, language in neurodegenerative diseases progressively deteriorates and can lead in the most advanced cases to jargon aphasia (Caffarra et al., 2013; Rohrer, Rossor, & Warren, 2009) or even mutism (Caso et al., 2014; Gorno-Tempini et al., 2006).

Language impairment can also be the presenting feature of dementia in which case it is Primary Progressive Aphasia (PPA). In PPA language impairment occurs in a context of a relative sparing, initially, of other cognitive abilities and autonomy (Mesulam, 1982). Since its first description in 1982 (Mesulam, 1982), many recommendations were proposed in order to classify the different variants of PPA. A first distinction was made between fluent and non-fluent PPA. Fluent PPA was further referred to as semantic dementia (Hodges, Patterson, Oxbury, & Funnell, 1992; Snowden, Goulding, & Neary, 1989) whereas non-fluent PPA, first described in Mesulam´s seminal paper, was referred to as agrammatic PPA (Grossman, 2012; Neary, 1998). Both non-fluent/agrammatic PPA and semantic dementia

can be presentations of fronto-temporal lobar degeneration (FTLD) (Knibb, Xuereb, Patterson, & Hodges, 2006; Neary, 1998). PPA can also be an atypical presentation of AD in which case it can either fall under the umbrella of the logopenic (LvPPA) (Gorno-Tempini et al., 2011) or mixed variants (mPPA) (Mesulam et al., 2009; Mesulam et al., 2012; Sajjadi, Patterson, Arnold, Watson, & Nestor, 2012).

Fluent progressive aphasia, semantic dementia or the semantic variant of PPA (SvPPA) is to date the best describe variant of PPA. It is characterized by impairment in confrontation naming, object knowledge and single word comprehension, possible surface dyslexia or dysgraphia as well as spared repetition and speech production (Gorno-Tempini et al., 2011). Patients with SvPPA often present additional neuropsychiatric features (H. J. Rosen et al., 2006): lack of empathy (Hutchings, Hodges, Piguet, Kumfor, & Boutoleau-Bretonniere, 2015), emotion recognition deficits (Irish, Kumfor, Hodges, & Piguet, 2013; Kamminga et al., 2015), sweet food preference (Ahmed et al., 2014) or/and perseveration and stereotypic behaviour (Harris et al., 2016). It is linked to a pronounced asymmetrical atrophy of the anterior temporal lobe, and is mostly due to an underlying TDP-43 proteinopathy, although some patients can have a Tauopathy or Alzheimer´s disease (Davies et al., 2005; Deramecourt et al., 2010; Hodges et al., 2010; Spinelli et al., 2017). The disease onset is mostly before 65, but was also reported later in diverging proportion (Hodges et al., 2010). SvPPA is more rarely associated with a genetic mutation than other subtypes of the FTLD spectrum (Goldman et al., 2005; Seelaar et al., 2008).

The non-fluent variant of primary progressive aphasia (NfvPPA) is characterized by agrammatism in language production, effortful halting speech with inconsistent speech sound errors and distortions, and can be accompanied by impaired comprehension of complex sentence, spared single word comprehension and object knowledge (Gorno-Tempini et al., 2011). It is associated with a left anterior insular and inferior frontal atrophy (Gorno-Tempini et al., 2004; Gorno-Tempini et al., 2006) and occurs more often in the context of a Tauopathy (Knibb et al., 2006; Mesulam et al., 2008; Spinelli et al., 2017) especially in presence of apraxia of speech (Josephs et al., 2006).

The neuropsychological profile of LvPPA, most recently described, is also currently the most questioned. Patients fulfilling the criteria for LvPPA have impaired single word retrieval, impairment in repetition of sentences and phrases, phonologic errors in speech and naming, with spared single words comprehension and object knowledge, spared motor

speech and absence of frank agrammatism (Gorno-Tempini et al., 2011). This subtype is mostly associated with left posterior temporo-parietal atrophy and an underlying AD (Botha et al., 2015; Harris et al., 2013; Leyton et al., 2011; Rabinovici et al., 2008; Spinelli et al., 2017). Some authors argue that the clinical profile of LvPPA is too narrow to grasp the range of deficits present in patients that have neither NfvPPA nor SvPPA. Therefore, they proposed the concept of mPPA because of the presence of semantic, agrammatic and repetition deficits (Mesulam et al., 2009; Sajjadi et al., 2014).

In PPA the correspondence between the clinical profile and the underlying pathology is far from systematic and there are many exceptions and overlaps. Moreover, even in the classic presentations, the subtyping of PPA patient is very difficult in clinical routine where only a limited time for exploring neuropsychological functions is available This is even truer in Germany, because most recent developments in language testing for PPA are published in the English language.

6.2 Aim

This study aimed at developing tools that can be used for better characterizing PPA patients as well as to investigate the heterogeneity of their neuropsychological profile. Chapter 7 presents two new neuropsychological tools to detect agrammatism, a critical deficit for the subtyping of PPA patients. Chapter 8 explores verb inflection in PPA and the specific mistakes occurring in each subtype. Finally, chapter 9 investigates the clinical heterogeneity of PPA patients with a confirmed amyloid pathology.

6.3 Inclusion criteria

6.3.1 England

See section 7.3.1 for the specific recruitment criteria used in chapter 7.

6.3.2 Germany

Patients from chapters 8 and 9 were recruited from memory clinics in Magdeburg, Gottingen, Munich, Leipzig, Bonn and Hannover. AD patients were diagnosed based on the latest diagnostic criteria (McKhann et al., 2011). All the other patients fulfilled the root criteria for PPA: the presence of a language impairment that was primary, first isolated, progressive, not better explained by another neurologic or psychiatric cause and not too impaired that they could not go through the different tests. Caregiver questionnaires were collected to ensure that the language impairment was the starting deficit. The patients were further subtyped using the latest diagnostic recommendations (Gorno-Tempini et al., 2011) and based on the neuropsychological examination, the interview, the MRI report and the result of the F18-Florbetaben amyloid PET. A group of PPA patients with a mixed profile and whose cognitive impairment did not fit NfvPPA or SvPPA profiles and exceed those described for LvPPA was identified. These patients all had a positive amyloid PET and are, therefore, referred to as Aß+PPA. Noteworthy is that part of the PPA patients presented in chapters 8 and 9 are the same than in chapter 3.7.

Healthy controls were spouses of patients, or recruited through advertisements published in the local newspapers. They had no major neurological or psychiatric disorders, were native German speakers over 50 and performed normally on the general neuropsychological screening that included the Mini-Mental State Evaluation (MMSE) (Folstein et al., 1975), a 30-items version of the Boston Naming Test (BNT) (Merten, 2004), the 15-items Geriatric Depression Scale (GDS) (Sheikh & Yesavage, 1986), the subtests animal fluency and K fluency in 1 minute from the Regensburger Wortflüssigkeits-Test (RWT) (Aschenbrenner, Tucha, & Lange, 2000) as well as the digit span subtest from the German adaptation of the Wechsler Intelligence Scale III (Wechsler, 1997a). Control participants were the same than in the longitudinal study presented in section 3.8 and 4.7. This allowed screening a possible degradation in cognitive abilities in the months following the study. Indeed, knowing that most neuropsychological tests are insensitive to early decline, part of the control group could have been in an undetected preclinical stage at the time of testing.

Written informed consent was obtained from the participants and, where appropriate, their next of kin.

6.3.2.1 Classical neuropsychological assessment

All participants underwent a standard neuropsychological assessment summed up in table 6.1.

Table 6-1 List of the neuropsychological tests used in PPA

Cognitive Domain	Test used
Global functioning level	- MMSE - ADL questionnaire for caregiver - Behavioural questionnaire for caregiver
Attention & Executive functions	- Digit symbol substitution
Fluency	- Category fluency - Lexical fluency
Non-verbal memory	- Immediate and delayed recalls of the ROCF - Corsi blocks
Verbal memory	- Digit span
Visuo-construction	- Copy of the ROCF
Semantic knowledge	- Repeat & Point
Confrontation naming	- BNT-30
Praxia	- Bucco-facial, meaningful and non-meaningful mono- and bimanual gestures.
Arithmetic	- Additions, subtractions and divisions
Depression	- GDS

Chapter 7.

SECT and MAST: assessing agrammatism in primary progressive aphasia

7.1 Introduction

PPA encompasses multiple clinical profiles caused by different underlying pathologies. Many diagnostic criteria were published in recent years (Gorno-Tempini et al., 2004; Gorno-Tempini et al., 2011; Knibb, Woollams, Hodges, & Patterson, 2009; Mesulam et al., 2009; Neary et al., 1998; Thompson et al., 2012). They present agrammatism as a crucial symptom for the subtyping of PPA variants. The exact definition of "agrammatism" and how to objectify its presence are, however, rarely precisely described.

Recommendations for PPA subtyping emphasize making a distinction between lexico-semantic and grammatical/syntactic abilities. While lexico-semantic capacities are easily assessed with tests of single word comprehension, naming and associative knowledge, tests to quantify the impairment of grammatical abilities and fluency remain limited. The gold standard is the quantitative analysis of connected speech samples (Ash et al., 2009; Knibb et al., 2009; Rogalski et al., 2011; Sajjadi, Patterson, Arnold, et al., 2012; Sajjadi, Patterson, Tomek, & Nestor, 2012a, 2012b; Wilson et al., 2010). It is, however, time consuming and requires complicated scoring methods making it hardly suitable for clinical settings. In spite of a few published recommendations on scoring methods, the determination of the presence/absence of agrammatism or non-fluency depends generally on the clinician's opinion. This leads to a lack of standardization across centres.

A few tests were developed to test grammatical abilities in a standardized manner. The Northwestern Anagram Test (NAT) investigates the accuracy of syntax in sentence production (Weintraub et al., 2009). In this test the participant is invited to build, with the

help of word-cards, meaningful canonical and non-canonical sentences to describe a scene presented on a picture. The NAT gives a valuable insight in certain aspects of syntactic abilities but does not directly assess patients‘ abilities to inflect verbs or use appropriate closed class words to build grammatically correct sentences. Moreover, the ceiling effect reported in healthy participants (Weintraub et al., 2009) is a major hurdle for the early diagnosis of PPA, which is characterized by a slow and insidious onset.

The Northwestern Assessment of Verbs and Sentences (NAVS) (Cho-Reyes & Thompson, 2012) explores the production and comprehension of both verbs with different number and optionality of arguments and canonical and non-canonical sentences. In 2013, Thompson & al. showed that the sentence comprehension subtest, the sentence production priming subtest of the NAVS as well as the Northwestern Assessment of Verbs Inflection (NAVI) can discriminate between logopenic and agrammatic variants of PPA. Noteworthy is, that the assignment to the agrammatic or logopenic PPA sub-groups only occurred on the base of the word comprehension level and the ability to produce sentences in the NAT. This indicates principally that the NAT and the NAVS display similar results.

The Test for Reception of Grammar (TROG) (Bishop, 1989), initially developed to appraise the development of children´s grammatical abilities, is often used in PPA (Knibb et al., 2009; Sajjadi, et al., 2012a, 2012b). The participant has to choose a picture out of four alternatives that matches a sentence spoken by the examiner. Different syntactic structures of graded difficulty are present in the test. Despite the development of adult norms in the revised Test for Reception of Grammar (TROG-2) (Bishop, 2003), it still suffers from ceiling effect. Thus, making it unsuitable for early diagnosis of progressive language disorders like PPA.

7.2 Aim

The preservation or degradation of grammatical abilities being of primary importance for subtyping PPA variants, it is imperative to develop more standardized and clinically suitable testing methods. Ceiling effects should be avoided to increase the sensitivity to slow and insidious onset. In this study, we describe two new short neuropsychological tests that were specifically developed to assess grammatical abilities in PPA patients. The "SEntence Comprehension Test" (SECT) aims at assessing comprehension of grammatically complex sentences and the "Make A Sentence Test" (MAST) appraises grammatical abilities in speech production. Connected speech samples of 150 words were collected from patients. These were analysed and compared to the scores obtained in the new tests to assess how they relates to patients' spontaneous speech—and concurrently examine ecological validity.

7.3 Method

7.3.1 Participants

Forty-one patients with PPA participated in the study. Fifteen patients fulfilled the criteria for SvPPA and thirteen patients for NfvPPA. Thirteen remaining patients did not meet the criteria for any of the three described subtypes. They had a mixed impairment profile and were therefore labelled mixed PPA (mPPA). These patients possibly match what is classified as LvPPA in other studies. Previous work on this cohort (Sajjadi, Patterson, Arnold, et al., 2012) showed that, despite not fully meeting the cognitive criteria proposed recently, these patients have the same group-level atrophy pattern than the one described in LvPPA.

Patients were recruited to the Cambridge Longitudinal Study of Primary Progressive Aphasia from the Memory Disorders clinic held at Addenbrooke´s Hospital, University of Cambridge, UK. The study was approved by the Regional Ethics Committee. Thirty healthy controls were spouses of patients with matching age and education level. Written informed consent was obtained from the participants and, where appropriate, their next of kin.

PPA patients were subtyped by cognitive neurologists based on connected speech, "bedside" testing including object naming, word definition and repetition of words as well as on the results of the magnetic resonance imaging. Twenty-one patients who met the criteria for the amnestic-presentation of probable AD proposed by McKhann & al (2011) were recruited. Control participants had no cognitive impairment, no diagnosed neurological or psychiatric illness and performed normally on the Addenbrooke´s Cognitive Examination-revised (ACE-R) (Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006) and Mini-Mental State Examination (MMSE) (Folstein, Folstein, & Mchugh, 1975). Non-native English speakers or participants who were too impaired to provide a reliable connected speech sample were excluded.

7.3.2 Connected speech sample

Connected speech sample were recorded during a face-to-face semi-structured interview. This method was proven superior in obtaining a great variety of syntactic structures compared to more constraint picture description (Sajjadi, et al., 2012a). Based on previous work showing that 150 words are sufficient to grasp language impairment in PPA (Sajjadi, et al., 2012a), only samples with a minimum of 150 words were retained. Speech samples

were analysed as followed: the number of hesitations and the percentage of abandoned and elliptical sentences assessed fluency, the rate was measured by the number of words per minute, spontaneity was derived from the number of questions the examiner had to ask to obtain at least 150 words, the percentage of complex units including embedded, passive, relative, conditional and complex sentences were used to reflect syntactic complexity, and grammatical errors like verb agreement errors, closed-class word errors and miscellaneous grammatical errors not falling into the first two categories (e.g. gender mistakes) were analysed. For a more detailed description of the markers, see Sajjadi, et al., 2012a.

7.3.3 General neuropsychological assessment

All participants completed a standard neuropsychological battery and provided a connected speech sample on the same day. The battery comprised two measures of global functioning: the MMSE (Folstein et al., 1975) and ACE-R (Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006), as well as the forward and backward digit span subtests from the Wechsler Memory Scale (Wechsler, 1997b), a shortened version of the TROG (Bishop, 1989) with the 28 items that correlated most highly with the whole test performance (unpublished data from our research group comprising blocks G, H, K, M, N, P and S), the NAT (Weintraub et al., 2009), the Camel and Cactus Test of semantic abilities (CCT) (Bozeat, Lambon Ralph, Patterson, Garrard, & Hodges, 2000), the 64-item naming subtest of the Cambridge semantic memory battery (Adlam, Patterson, Bozeat, & Hodges, 2010), the letter and category fluency (from the ACE-R), the cube analysis from the Visual and Object Space Perception Battery (VOSP) (Warrington & James, 1991) and the copy and delayed recall of the Rey-Osterrieth complex figure (Osterrieth, 1944; Rey, 1941)

7.3.4 Novel neuropsychological tests

The lists of items for both tests are available in appendix II and III.

7.3.4.1 SECT V & A

The SECT tested the comprehension of sentences of varying syntactic complexity. In two different sessions, the test was administered in both auditory (SECT A) and visual (SECT V) versions to assess the impact of the auditory-verbal working memory capacity in online sentence comprehension. Eight embedded, eight passive and eight comparative sentences were presented intermixed to the participant. Half of the sentences contained nine words, the other half 12 words. Word-frequency (11-315 per million words) was matched among the different sentences using the British National Corpus Database

(www.natcorp.oc.ac.uk). A short question following the target sentence, presented either written in the visual version or spoken in the auditory version, assessed comprehension. Only a single word response was expected in both versions.

The test started with the following instructions “In this test you will be presented with a number of sentences each followed by a question. Please listen/read carefully and answer the question. I will not tell you if your answers are correct or incorrect”. For each syntactic category a practice item and feedback were given. A point was given for a correct answer, even in the case of an immediate auto-correction or hesitation.

7.3.4.2 Make A Sentence Test

The MAST assessed the ability to generate appropriate grammatical morphemes and inflections to construct sentences with varying complexity level.

Fifteen trials with four content words, including at least one verb, were designed. The participant had to use all four words to generate a meaningful sentence. Words were matched for word-frequency based upon the British National Corpus Database (www.natcorp.oc.ac.uk) (12-695 per million words) and tested using ANOVA to ensure differences were not significant across syntactic structure categories. Three blocks of five different syntactic structures (active, passive or interrogative sentences, embedded or relative clauses) were presented intermixed. To obtain the expected syntactic structure, words were given in a specific order that the participant was not allowed to change while producing a sentence. Some trial did allow more than one possible correct answer. For instance, the word string “Emma - bake - pie - party” could be correctly answered with “Emma baked a pie for the party”, “Emma will bake a pie for the party”, “Emma will bake a pie at the party”, “Emma likes to bake pies for parties”, etc.

The test started with the following instruction “In this test you are asked to make up a meaningful and sensible sentence from the words provided. It is important that you use all the words, in the order in which they are given. You may alter the form of the word to fit your sentence, for instance you might change “sit” to “sitting”, “was” to “were” or “see” to “seen”, and you may add your own words. Here are some practice items to help you to get started”. For each sentence category, a practice item and feedback were given to ensure that the participant fully understood the task.

Clinical judgement was used to determine if the participant could read and understand the content words. A set of line drawings illustrating the given words was at the examiner´s disposition when needed. Two points were given for a correct syntactic structure in a single sentence, one point for the target sentence with a single inflection or preposition error. No points were awarded when the participant failed to provide a correct answer or if it contained more than one grammatical error.

7.3.5 Statistical considerations

Demographics, performance in the different neuropsychological tests as well as the linguistic measures extracted from the connected speech samples were analysed using SPSS version 21.0 (IBM corp., Armonk, NY). For non-normally distributed scores, the Kruskal-Wallis test was used to compare means, otherwise a one-way ANOVA was used. Dunnett´s T3 post-hoc tests were used for non-normally distributed scores and the Scheffé post-hoc test was used for normally distributed scores in samples with an unequal number of subjects. Standard linear regressions were performed between the Z scores obtained by all patients as a group using, respectively, the SECT A, the SECT V and the MAST as the dependent variable, and the Z scores of the 7 measures described previously and obtained from the connected speech data as independent variables.

7.4 Results

7.4.1 Demographics and general neuropsychology

Demographics for all participants are displayed in table 7.1. No statistical difference was found ($p > 0.05$) between the five groups for the age at assessment or education level and between the four patient groups for disease duration.

Table 7-1 Demographics

Abbreviations: NfvPPA = Non-fluent variant of primary progressive aphasia, SvPPA = semantic variant of primary progressive aphasia, mPPA = mixed variant of primary progressive aphasia, AD = Alzheimer´s disease, HC = healthy control, n.a = non-available, F = female, M = male.

	NfvPPA Mean [Range]	**SvPPA Mean [Range]**	**mPPA Mean [Range]**	**AD Mean [Range]**	**HC Mean [Range]**
Age at onset (years)	65.7 [49-75]	64.6 [58-73]	69.5 [63-80]	63.7 [51-77]	n.a
Age at assessment (years)	68.8 [53-77]	68.7 [61-79]	73.1 [66-83]	68.4 [60-79]	67.8 [51-80]
Disease duration (months)	36.5 [18-60]	50 [24-78]	41.5 [24-72]	55.1 [24-168]	n.a
Education (years)	12.6 [10-20]	13.9 [10-19]	10.9 [9-18]	12.5 [10-19]	12.8 [10-20]
Gender	8 F, 5 M	8 F, 7 M	9 F, 4 M	10 F , 11 M	15 F, 15 M

Table 7.2 summarizes the results obtained in the general neuropsychological assessment. Patients were significantly worse than controls in the majority of the tests. The mPPA group had the worst performance in a number of tests. Exceptions were the CCT, confrontation naming and category fluency where the SvPPA group was worse, and the VOSP and copy of the ROCF where the NfvPPA group was more impaired. The AD group performed significantly worse at the recalls of the ROCF.

Table 7-2 General neuropsychology

p<0.05 compared to the control population, a p<0.05 compared to NfvPPA, b p<0.05 compared to SvPPA, c p<0.05 compared to mPPA, d p<0.05 compared to AD , **Bold: significance survives Bonferroni correction for multiple comparisons*

The number of participants for each tests differ from 72 to 92 accros tests because all the neuropsychology could not be administered to every healthy participant

	NfvPPA Mean (SD)	**SvPPA Mean (SD)**	**mPPA Mean (SD)**	**AD Mean (SD)**	**HC Mean (SD)**	**Significance *(P-value)***
MMSE (Max. 30)	22.2 (5.2)*****	23.4 (2.8)*****	20.5 (6.6)*****	22.7 (3.2)*****	29.1 (1.2)***abcd***	X^2 (4, 92) = 56.1, *p<0.001*
ACE-R (Max. 100)	67.3 (15.7)*****	54.1 (12.8)*****	51.5 (22.2)*****	65.2 (11.6)*****	94.3 (4.1)***abcd***	X^2 (4, 92) = 62.9, *p<0.000*
Forward digit span	4.7 (1.2)b	6.3 (1.2)a***c***	4 (1)****bd***	5.86 (1.2)***c***	6.2 (1.2)***c***	X^2 (4, 72) = 26.1, *p<0.001*
Backward digit span	3.2 (.8)-b	4.7 (1.2)a***c***	2.8 (0.6)***b**d	3.8 (1.3)c	4.5 (1.1)c	X^2 (4, 72) = 26.2, *p<0.000*
Letter fluency	5 (3)*****	5.9 (3.1)*****	4.8 (2.2)*****d	9.4 (5.9)*****c	14.4 (4.6)***abc***d	F (4, 91) = 19.8, *p<0.001*
Category fluency	8.9 (4.6)*****b	4.3 (2.6)*****a***d***	7.9 (3.7)*****	9.1 (2.9)****b***	19.7 (4.2)**abcd**	F (4, 91) = 57.7, *p<0.000*
Modified TROG (Max. 28)	23.5 (4.3)	26.9 (1.1)*c	19.9 (6)*****b	25.5 (3.1)*****	28 (0.0)b***c***d	X^2 (4, 73) = 30.9, *p<0.001*
NAT (Max. 10)	5.2 (3.2)*****	7.2 (2.5)*****c	3.5 (2.9)*****b	6.1 (2.5)*****	9.5 (0.6)***a***b***cd***	X^2 (4, 89) = 45.5, *p<0.001*
CCT (Max. 64)	50.1 (8.6)****b***	33.3 (11.3)****a***d	44.3 (13.2)*	47.7 (9.5)*****b	58.8 (2.7)a***b***c***d***	X^2 (4, 92) = 52.4, *p<0.001*
64 item naming	60.5 (1.44)***b***c	22.3 (16.4)****acd***	46.2 (14.9)*a***b***	55.9 (7)****b***	62.3 (1.9)***b***c***d***	X^2 (4, 92) = 57.1, *p<0.001*
VOSP (Cube analysis) (Max. 10)	6.9 (2.8)b	9.9 (.26)ac***d***	7 (2.8)b	7.2 (3.1)***b***	9.3 (1.5)	X^2 (4, 91) = 23.5, *p<0.001*
ROCF copy (Max. 36)	24.6 (8.1)*b	33.3 (2.3)ad	26.5 (14.7)	23.8 (12.8)*b	34.2 (1.6)ad	X^2 (4, 91) = 28.3, *p<0.001*
ROCF delayed recall (Max. 36)	8.8 (7.9)*	10.8 (8.5)*	6.5 (6.3)*****	4.1 (4.1)*****	18.3 (5.2)ab**cd**	X^2 (4, 91) = 38.1, *p<0.001*

7.4.2 Novel tests

7.4.2.1 Descriptive statistics

Performance of all participants in both the MAST and SECT are displayed in table 7.3.

Table 7-3 Performance in the novel tests

p<0.05 compared to the control population, a p<0.05 compared to nfvPPA, b p<0.05 compared to svPPA, c p<0.05 compared to mPPA, d p<0.05 compared to AD.* ***Bold**: significance survives Bonferroni correction for multiple comparisons.*

The number of participants for each tests differ from 69 to 92 accros tests because all the neuropsychology could not be administered to every healthy participant.

Novel Tests	**NfvPPA Mean (SD) [Range]**	**SvPPA Mean (SD) [Range]**	**mPPA Mean (SD) [Range]**	**AD Mean (SD) [Range]**	**HC Mean (SD) [Range]**	**Significance *(P-value)***
SECT A (Max. 24)	13.8 (4.5)*****c [13]	18.5 (4.5)****c*** [15]	8 (4.9)*****a***bd*** [18]	18.3 (5.5)****c*** [23]	23 (1.2)***a***b**c***d*** [4]	X^2 (4, 87) = 52.46, *p<0.001*
SECT V (Max. 24)	16.6 (5.1)***** [17]	19.3 (4.1)* [12]	14.6 (4.9)***** [14]	19.3 (4.01)***** [16]	23.7 (0.6)***a***b***cd*** [2]	X^2 (4, 86) = 44.13, *p<0.001*
SECT A Corrected for working memory impairment	18.3 (3.4)c [9.9]	19 (6.8)c [25.8]	11.6 (5.2)abd* [18.6]	19.8 (6.4)c [34.1]	23.4 (3.6)c [9.92]	X^2 (4, 69) = 21.88, *p<0.001*
MAST (Max. 30)	12.6 (8.3)***** [24]	13.8 (8.5)***** [29]	7.3 (9.6)***** [26]	15.2 (9.7)***** [30]	24.9 (4.5)***abcd*** [17]	X^2 (4, 92) = 36.63, *p<0.001*
MAST Corrected for semantic impairment	11.22 (7.28)* [20]	18.43 (9.68) [34,6]	8.75 (10.7)* [26.3]	14.65 (9.04) [26,7]	19.09 (3.57)ac [15.2]	X^2 (4, 92) = 14.49, *p<0.01*

7.4.2.2 Sentence comprehension

Healthy participants performed near ceiling on SECT V but SECT A had a wider score range. Patients performed significantly worse than controls on both versions. mPPA had the worst performance in both versions of the test, followed by NfvPPA (see table 7.3).

Participants performed better in the SECT V compared to the SECT A. This was likely due to the higher verbal working memory demand for the auditory version. A partial correlation—controlling for MMSE score—was run between the performance in the SECT A and V and the performance in forward digit span, a test of verbal working memory. The NfvPPA (Spearman´s rho: 0.76, p=0.004), mPPA (Spearman´s rho: 0.87, p<0.001) and AD (Spearman´s rho: 0.48, p=0.03) groups showed a significant correlation between the SECT A and forward digit span whereas no significant correlation was found for healthy controls (Spearman´s rho: 0.59, p=0.2) or SvPPA (Spearman´s rho: -0.14, p=0.7). A significant correlation was only found between performance in the SECT V and forward digit span in mPPA (Spearman´s rho: 0.78, p=0.003). Moreover, SECT A and V correlated respectively with the NAT (Spearman´s rho: 0.75, p<0.001; Spearman´s rho: 0.78, p<0.001) and TROG (Spearman´s rho: 0.72, p<0.001; Spearman´s rho: 0.73, p<0.001).

7.4.2.3 Make A Sentence Test

Control participants were off ceiling on the MAST. Patients performed significantly worse than the control group (see table 7.3). No significant difference in performance was found between the different groups of patients. Performance in the MAST correlated significantly with the NAT (Spearman´s rho: 0.77, p<0.001) and the TROG (Spearman´s rho: 0.67, p<0.001).

Test-retest data were not available but the MAST showed a high internal consistency on a single test point (Cronbach´s Alpha: 0.912).

7.4.2.4 Correction factors

We assumed that the impaired performance of the SvPPA group on the MAST originated from a deficit in semantic rather than grammatical abilities. Difficulties in comprehension of the content words were possibly a hurdle to the generation of grammatically correct sentences. This idea was reinforced by a significant correlation between scores on the CCT (a semantic association task) and the MAST in the SvPPA group (Spearman´s rho: 0.55, p=0.03) but not in the other groups (NfvPPA: Spearman´s rho: -0.17, p=0.6; mPPA: Spearman´s rho: -0.04, p=0.9; AD: Spearman´s rho: 0.24, p=0.3). To adjust for the impact of semantic deficits, a correction factor was applied to the MAST performance in that each MAST score was scaled to calibrate each participant´s CCT score to the grand mean CCT score. Performance in the corrected MAST are displayed in table 7.3. When corrected for semantic impairment, SvPPA and AD group performance on the MAST no longer differed

significantly from healthy controls. NfvPPA and mPPA remained, however, significantly impaired.

Similarly, the idea that the better performance in the SECT V over the SECT A was due to greater working memory demand in the former, was deepened by applying a correction factor for working memory abilities. The performance in the forward digit span was used to correct for the working memory deficits in the SECT A. Scores in SECT A and SECT V differ significantly in the NfvPPA group (p=0.04), mPPA group (p=0.002) and controls (p=0.01). After correction these significant differences all disappeared with the exception of mPPA (p=0.02).

7.4.2.5 Regression analysis

7.4.2.5.1 Sentence comprehension

The unstandardised regression coefficient (B) and intercept, the standardised regression coefficient (beta) and the collinearity statistics for SECT A and V, and MAST are summarized in table 7.4.

On SECT A, *R* (regression) was significantly different from zero $F(7, 62) = 15.7$, $p<0.00$. Four independent variables from the connected speech samples contributed significantly to the prediction of performance in the SECT A : grammatical errors (beta = -0.39); elliptical (beta = -0.36) and abandoned sentences (beta = -0.48); and the speech rate (beta = 0.29). Altogether 62.4 % of the variability in performance on the SECT A can be predicted knowing the scores of the 7 connected speech variables.

Table 7-4 Linear regression of the connected speech sample cues

Abbreviations: B= b weight, Std. Error = standard error, Sig. = significance, VIF = variance inflation factors., Error : verb agreement errors, closed-class word errors and miscellaneous grammatical errors not falling into the first two categories (e.g. gender mistakes), Hesitation : number of hesitations, Rate : words per minute, Complex unit : percentage of passive, embedded, relative, conditional and complex sentences, Elliptical : percentage of elliptical sentences, Question : number of questions the examiner had to ask to get the expected sample length and Abandoned : number of abandonned sentences.

Model		**Unstandardized Coefficients**		**Standardized Coefficients**			**Collinearity Statistics**	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Constant	SECT A	-0.59	0.5		-1.19	0.24		
	SECT V	-0.24	0.37		-0.66	0.51		
	MAST	-0.62	0.27		-2.27	0.03		
Z error	SECT A	-1.11	0.26	-0.39	-4.34	0.0	0.74	1.34
	SECT V	-0.53	0.19	-0.29	-2.82	0.01	0.74	1.35
	MAST	-0.43	0.15	-0.31	-2.86	0.01	0.76	1.32
Z hesitation	SECT A	0.25	0.25	0.12	1.0	0.32	.043	2.35
	SECT V	0.1	0.18	0.07	0.53	0.6	0.43	2.35
	MAST	0.20	0.15	-0.2	1.38	0.17	0.43	2.33
Z rate	SECT A	0.86	0.39	0.29	2.17	0.03	0.34	2.97
	SECT V	0.22	0.29	0.12	0.76	0.45	0.34	2.97
	MAST	0.22	0.22	-0.16	0.98	0.33	0.35	2.84

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
Z complex unit	SECT A	-0.51	0.55	-0.11	-0.91	0.37	0.4	2.5
	SECT V	-0.12	0.41	-0.04	-0.29	0.77	0.4	2.50
	MAST	-0.25	0.3	-0.13	-0.86	0.39	0.43	2.31
Z elliptical	SECT A	-1.04	0.4	-0.36	-2.62	0.01	0.33	3.03
	SECT V	-0.74	0.3	-0.39	-2.51	0.02	0.32	3.09
	MAST	-0.57	0.23	-0.41	-2.47	0.02	0.34	2.98
Z question	SECT A	0.16	0.11	0.18	1.48	0.14	0.43	2.35
	SECT V	-0.00	0.08	-.00	-0.00	1.0	0.42	2.41
	MAST	0.02	0.06	0.04	0.27	0.79	0.44	2.28
Z abandoned	SECT A	-1.49	0.32	-0.48	-4.62	0.0	0.56	1.79
	SECT V	-0.85	0.24	-0.42	-3.58	0.001	0.57	1.77
	MAST	-.066	0.19	-0.44	-3.47	0.001	0.59	1.69

R was significantly different from zero F(7, 61) =10.61, p<0.001 on SECT V. Three independent variables from the connected speech samples contributed significantly to the prediction of SECT V performance: grammatical errors (beta = -0.29); elliptical (beta = -0.39) and abandoned sentences (beta = -0.42). Altogether 52.4 % of the variability in performance on the SECT V can be predicted knowing the scores of the 7 connected speech variables.

7.4.2.5.2 Make A Sentence Test

For the MAST, *R* was significantly different from zero F(7, 67) =7.572, p<0.001.

Three independent variables from the connected speech samples contributed significantly to the prediction of performance in the MAST : grammatical errors (beta = -0.31), plus elliptical (beta = -0.41) and abandoned sentences (beta = -0.44). Altogether 40.7 % of the variability in performance on the MAST can be predicted knowing the scores of the 7 connected speech variables.

7.5 Discussion

Published criteria for PPA underline the crucial importance of speech fluency and grammatical abilities (Gorno-Tempini et al., 2004; Gorno-Tempini et al., 2011; Knibb et al., 2009; Mesulam et al., 2009; Neary et al., 1998; Thompson et al., 1997; Thompson et al., 2012). This is particularly true for the differential diagnosis of NfvPPA versus mPPA.

We developed the SECT (in visual and auditory versions) and the MAST to provide brief grammatical tests sensitive to the problematic of neurodegenerative diseases. Performance in both tests were representative of fluency and grammatical abilities of PPA patients in spontaneous speech. This was particularly highlighted by the regression analysis showing a significant relationship of all three tests to the production of grammatical errors during normal conversation. Noteworthy, is that the SECT A and the MAST especially, displayed a range of performance without ceiling effects, implying their suitability for the evaluation of early stage PPA.

In comparison with controls, all patients were impaired on the SECT A and V. All participants had slightly lower performance in the auditory version of the test. Both versions using the same stimuli set, we assume that the lower performance in the auditory version was explained by a greater working memory demand. In the visual version, the stimuli are laid before the subject whereas in the auditory version he/she has to remember both the sentence and the question to produce a response. This view was emphasized by a stronger correlation of the auditory version than the visual version with the digit span test. Moreover, once corrected for working memory abilities the differences in performance between the two versions disappeared in all group except mPPA. In this group, the correction only attenuated the difference in performance between the two versions. It seems that the auditory version of the SECT is more sensitive to change, which is of considerable interest for detecting early subtle deficits in PPA or even AD. The visual version of the SECT, however, seems to test grammatical comprehension in a purer manner as it is less influenced by working memory load. To summarize, the SECT, particularly in the visual presentation, seem to display good face validity as mPPA and NfvPPA groups showed the greatest impairment.

In the newly developed MAST, all patient groups were impaired relative to controls. When controlling for semantic impairment, the SvPPA had similar performance relative to

controls. A recent study found impairment in constrained tasks of grammatical abilities in SvPPA, though not as extended as NfvPPA (Cupit et al., 2016). The disappearance of the impairment in the MAST after controlling for semantic impairment might indicate that the source of the reported difficulties is a consequence of the damage to the semantic system. Indeed, producing a grammatically sound sentence appeals to a larger set of abilities than simply mastering the grammatical rules. Moreover, a few studies reported a simplification of the grammatical complexity in connected speech in SvPPA (Meteyard & Patterson, 2009; Meteyard, Quain, & Patterson, 2014). The difficulties observed in the SECT in SvPPA might also arise partly from semantic deficits.

As expected AD patients performed in the normal range on the corrected MAST. This confirms results from other studies showing that language impairment occurs in the course of the Alzheimer´s disease but is not a prominent presenting feature (Salmon & Bondi, 2009; Weintraub et al., 2012).

Although both developed for the assessment of grammatical ability in neurological diseases, the SECT and the NAT use different approaches. While in the SECT, the participant has to either answer a question about a sentence presented orally or visually, in the NAT the participant need to build sentences describing a picture by using word-cards. Thus, the NAT was developed for the assessment of grammatical production whereas the SECT addresses grammatical comprehension. It is, however, worth mentioning that grammatical production in the NAT arises from arranging already inflected word-cards to create a sound sentence. It is imaginable that performance in this test depends more on comprehending the word cards than having intact grammatical knowledge. This possible weakness was addressed in the MAST, in which the participant has to produce a sentence and inflect grammatical words. Although they rely on different aspects of grammatical abilities, performance seen in the SECT and NAT are similar in all group, and significantly correlated. This suggests that they measure a core grammatical deficit but the SECT seem to be more devoid of ceiling effects than the NAT, making it more suitable for early stages.

Finally, the results showed that NfvPPA patients performed in the normal range in the modified version of the TROG. Although this result might be related to a lack of statistical power due to the use of one-way ANOVA in small groups, a significant impairment in NfvPPA relative to control was found in the MAST and both versions of the SECT. This

suggests a higher sensitivity of the MAST and SECT to the grammatical deficits seen in PPA compared to the TROG. This test, however, was developed for another purpose.

Lack of ecological validity—a poor correspondence between the performance in a test and in real life—is a major critic in neuropsychological assessment. It is also true for tests of grammatical abilities. Indeed, proper grammar use relies on a broad set of abilities and while tests of grammatical abilities may have face validity, it remains difficult to know if they are sensitive to core grammatical deficits or simply to a more general impairment of working memory that would not appear in natural conversation. To address this challenge, we compared the performance obtained in the new tests to samples of patients´ connect speech. The analyses revealed that performance in the new tests could be partly predicted by the number of grammatical errors as well as the percentage of abandoned and elliptical sentences produced by patients in spontaneous speech. These findings highlight the validity of the new tests to assess both grammatical deficits (errors) and fluency impairments (elliptical and abandoned utterances) in spontaneous speech of PPA patients. Studies showed that fluency and grammatical abilities are depending on two different networks that are selectively impaired in different subtypes of PPA (Catani et al., 2013; Thompson et al., 2012). Deficits in grammatical abilities, however, are only one possible reason for reduced fluency. Indeed, fluency can also be impacted by word-finding difficulties. Rate (words/minutes), an indicator of fluency in spontaneous speech, significantly predicted the performance on the SECT A but not the understanding of identical, syntactically complex written sentences on the SECT V. This suggests that the capacity to sustain a normal rate in speech production as well as to understand syntactically complex sentences is an on-line process engaging verbal working memory and seemingly more sensitively revealed by the auditory version of the SECT. This idea is reinforced by the fact that SvPPA patients performed normally on tests of verbal working memory, had no extensive diminution of the rate of speech and like controls, performed similarly in both version of the test. To the contrary, NfvPPA and mPPA were impaired in verbal working memory tests compared to controls and had better performance on the SECT V than SECT A. This is in line with previous studies showing impairment of verbal working memory (Gorno-Tempini et al., 2008; Gorno-Tempini et al., 2004; Rohrer, Sauter, Scott, Rossor, & Warren, 2012) and a slow rate of speech in NfvPPA and mPPA (or LvPPA) (Gorno-Tempini et al., 2008; Gorno-Tempini et al., 2004; Knibb et al., 2009; Wilson et al., 2010).

The SECT and the MAST were developed with a view to avoid ceiling effects to identify early grammar impairment in disorders with insidious onset. Only the MAST seemed, however, to be devoid of ceiling effect. Another weakness of this study lies in that the regression analyses were run by collapsing all participants into a single group. Unfortunately, PPA sub-groups were too small to allow performing the regression in individual syndromic categories. Finally, for logistical reasons, test-retest reliability for the novel tests was not available.

In summary, the SECT A and V as well as the MAST were designed to assess grammatical ability and fluency in PPA. Performance in the tests were related to grammatical errors in connected speech. The novel tests seem therefore, to provide an objective measurement of grammatical abilities that is suitable in clinical settings. Further analysis is needed to describe the psychometric properties of the test.

Chapter 8.

Verb inflection in Primary Progressive Aphasia

8.1 Introduction

Inflectional morphology is the modification of a word´s root by typically adding an affix (i.e. prefix, suffix or infix) or changing the internal structure with a diphthong. Inflectional morphology can apply to nouns or verbs and provides information on tense, number, or gender. It is a complex linguistic process, seen as being at the crossroad of syntax, phonology and the lexicon (Spencer, 1994). Indeed, inflection provides information on the relationship between the inflected word and other components of the sentence and implies having knowledge of the syntactic properties of the language. Moreover, the process of affixation involves per se, phonological modification of a root. Finally, language is dynamic, evolves with time and influences from other languages, and can include irregularities. For example, the verb "buy" is not inflected "buyed" in the preterit but "bought". The way irregularities in languages are stored and processed was at the centre of many debates in the last decades. The dual-route processing view postulates that regular forms are generated from a set of grammatical rules whereas irregular forms are stored in a separate lexicon. Correct inflection of irregular verbs involves having access to that lexicon (Pinker, 1991). This model is supported by studies reporting a longer answer latency when asked to inflect irregular past tense than regular past-tense (Jaeger et al., 1996). These results were, however, nuanced by studies presenting regular and irregular verbs in blocks. Indeed, the repetition of regular verbs inside of a block might have caused priming effect reducing the answer latency for regular verbs only (Magen, 2014; Seidenberg & Hoeffner, 1998). Connectionist theoreticians favour the parallel distributed processing model in which regular and irregular forms are processed by a single system

using associative connections to decide whether a verb should be inflected in a regular or irregular manner (Rumelhart & McClelland, 1986). A third view proposes that inflection implies accessing a phonological rules system including large-scale rules for regular words and a set of more seldom rules that apply to irregular forms (Chomsky & Halle, 1968). These opposing views were the object of numerous studies, mostly in the English language and the debate persists.

8.1.1 Inflectional morphology in Primary Progressive Aphasia

Impairments in the application of grammatical rules inherent to language is of crucial important for the subtyping of Primary Progressive Aphasia (PPA). Its definition and screening lack, however, respectively of precision and standardization.

Inflectional morphology is a major aspect of grammar and was extensively studied in the different PPA subtypes. Inconsistent results were, however, obtained in verb inflection, a marker of agrammatism. In a semi-structured interview, no significant difference in the proportion of inflected verbs per token was found in a group of non-semantic PPA compared to healthy controls (HC) (Knibb, Woollams, Hodges, & Patterson, 2009). Similarly, in the picnic picture description task from the Western Aphasia Battery (WAB) (Kertesz, 1982), no significant difference in the proportion of inflected verbs in any of the PPA subtypes compared to HC could be found (Wilson et al., 2010). On the other hand, significant differences were found in the proportion of correctly inflected verbs in the non-fluent variant of PPA (NfvPPA) in different storytelling tasks (Ash et al., 2009; Thompson et al., 2012; Thompson et al., 2013). These findings, however, were not always replicated (Fraser et al., 2014). When directly comparing picture description and semi-structured interview in a PPA sample, mixed PPA (mPPA) made significantly more frequent verb inflection errors in the semi-structured interview while NfvPPA performed normally on both tasks (Sajjadi, et al., 2012b). Moreover, patients with the semantic variant of PPA (SvPPA) made more verb agreement errors than HC in the semi structured interview only (Sajjadi, et al., 2012a). Yet, those studies used methods based on free speech production in a semi-structured interview, in a picture description or in a storytelling task. In this context, patients might have restricted their expressive speech to simple syntaxes structures they master.

More consistent impairment patterns were obtained in task of constrained production of inflected verbs. Indeed, in the Northwestern Assessment of Verb Inflexion (NAVI); a

sentence completion task in which participants have to inflect transitive verbs in infinitive, progressive, present singular, present plural, past regular as well past irregular; NfvPPA were impaired in verb production (Thompson et al., 2013). In sentence completion task mixing verbs, pseudo verbs, nouns and pseudo nouns, with two levels of regularity, and frequency, SvPPA had poor performance on low frequency irregular words with a significant interaction effect of regularity and frequency (Wilson et al., 2014). Interestingly SvPPA showed significantly more over-regularization errors than the logopenic variant of PPA (LvPPA) but not NfvPPA. NfvPPA, in turn, were impaired in the inflection of pseudo-words, reflecting a possible difficulty to apply inflectional rules. LvPPA were impaired in all of the above-mentioned categories. In SvPPA, it seems that regularity and frequency plays a critical role in the successful inflection of verbs (Patterson, Lambon Ralph, Hodges, & McClelland, 2001). The expression and range of impairment in morphological inflection seems very heterogeneous not only between but also within PPA subtypes. Impairments in SvPPA seem to be mostly related to irregularity and low frequency whereas no clear systematic impairment in NfvPPA and LvPPA or mPPA is described. For review see Auclair-Ouellet, 2015.

8.1.2 Surface dyslexia in PPA

Over-regularization of irregular words when reading, the so-called surface dyslexia, is a hallmark of SvPPA. It is a relatively selective deficit in reading irregular words, which especially occurs when they have a low frequency. Surface dyslexics typically display an overreliance on the grapheme-phoneme correspondence, leading to over-regularizations such as reading “plaid” as “played”. Surface dyslexia is a known deficit in SvPPA (Hodges et al., 1992; Patterson & Hodges, 1992; Woollams, Ralph, Plaut, & Patterson, 2007) that is included as an additional diagnostic feature in the PPA latest diagnostic criteria (Gorno-Tempini et al., 2011). Its presence is, however, difficult to observe in languages with high grapheme-phoneme correspondence such as Italian or German where the expression of the impairment is subtler. Studies comparing German and English children showed that the accuracy of reading was higher in German, where orthography is more regular than in English and that German dyslexics suffer more likely from a reduced reading speed (Wimmer, 1993). In languages with low grapheme-phoneme correspondence, surface dyslexia is typically investigated by comparing the performance in reading lists of regular and irregular or exception words (Patterson & Hodges, 1992; Wilson et al., 2009). Given the relative regularity of the German language, such a procedure is not as straightforward.

Most irregular words in German are either loan words (44%) or have a slight change in vowel length (53%) (Ziegler, Perry, & Coltheart, 2000). Vowel length inconsistency being very subtle and loan words, not following the rule of the German language, detecting surface dyslexia can be challenging.

8.2 Aim

In a first step, we investigated verb inflection in a simple constraint task in SvPPA, PPA patients with a mixed impairment profile and a confirmed amyloid pathology and patients with typical Alzheimer´s Disease (AD). We expected an effect of regularity, modulated by frequency in SvPPA. In Aß+PPA we predicted a general impairment over the whole task.

As an alternative method to reveal the over-regularization effect that underpins surface dyslexia, we investigated the type of mistakes made in verb inflection in German with two past tense forms. Impairment in SvPPA that are specifically associated with regularization of irregular verbs are expected.

8.3 Method

8.3.1 Participants

Thirty-two HC and twenty-five patients (9 SvPPA, 12 amnestic AD, 4 Aß+PPA) participated in the study. Inclusion criteria are described in section 6.3.2.

8.3.2 Material

8.3.2.1 German morphology

The formation of the past participle; that forms together with the auxiliary "haben" or "sein" ("have" or "be") the perfect tense; is summarized in table 8.1. The formation of the preterit tense in German is summarized in table 8.2.

Table 8-1 Rules for inflection the German past participle

	Rule	**Infinitive**	**Preterit**
Regular Verbs			
Verbs ending with "ieren"	*Stem+t*	Kontrollieren	Kontrolliert
Most verbs	*Ge+stem+t*	Lernen	gelernt
Irregular Verbs			
Most verbs	*Ge+stem+en*	Sehen	gesehen
Mixed verbs			
Most verbs	*Ge+modified stem+t*	Bringen	gebracht

Table 8-2 Rules for inflecting the German preterit

	Regular verb "Lernen"	**Irregular verb "Sehen"**	**Mixed verb "Bringen"**
Rule	*Stem + regular ending*	*Modified stem + ending*	*Modified stem + regular ending*
Singular			
1. (Ich)	Ich lernte	Ich sah-	Ich brachte
2. (Du)	Du lerntest	Du sahst	Du brachtest
3. (Er/Sie/Es)	Er lernte	Er sah-	Er brachte
Plural			
1. (Wir)	Wir lernten	Wir sahen	Wir brachten
2. (Ihr)	Ihr lerntet	Ihr saht	Ihr brachtet
3. (Sie)	Sie lernten	Sie sahen	Sie brachten

8.3.2.2 Item set preparation

Ninety-six verbs, including 18 regular high frequency verbs, 18 irregular high frequency verbs, 18 regular low frequency verbs, 18 irregular low frequency verbs were compiled. Verb frequency was identified using the Celex Database. In a pilot study, 25 HC were given the compiled set of verbs. The verb was presented in the present form as an example, and the participant had to provide the past form of the verb in the German perfect and preterit. The different past tenses were chosen because of their distinctive frequency in German spoken language: the preterit is a low frequency tense whereas the perfect is a high frequency tense. After piloting, 6 regular high frequency verbs, 6 irregular high frequency verbs, 6 regular low frequency verbs, 4 irregular low frequency verbs were retained for the testing phase. We excluded verbs that could not be inflected by more than 75 percent of HC in both past tenses. For a list of the used verbs and their frequency, see appendix IV.

8.3.2.3 Administration

Verbs were included in simple sentences with an object like "it" or "him" when necessary. Participants were asked to read aloud, and fill in the blanks orally, see figure 8.1 for an example.

When necessary the auxiliary was given. It was, indeed. assumed that its inflection would be possible even in patients with advanced PPA because of its really high frequency. Participants were re-prompted when they misread the sentence, changed the pronoun or the verb that was given in the example sentence.

Figure 8-1 Item example from the verb inflection test

The parts in italics are those to be filled by the participant.

Example of a test item:

Du beginnst es

Gestern hast du es _____.

Gestern _____ du es.

Expected answer:

Du beginnst es

Gestern hast du es *begonnen*.

Gestern *begannst* du es.

English translation:

You begin it.

Yesterday you have begun it

Yesterday you began it.

8.3.2.4 Mistake classification

Table 8.3 summarizes the mistake types in verb inflection.

Table 8-3 Mistake classification

Category	Description	Example
Over-regularization	Regularized irregular verbs.	Gestern bin ich geschwimmt /schwimmte ich (Yesterday I have swimed/I swimed)
Semi-regularization	Partially regularized irregular verbs. A partial regularization occurs when the stem (a) or the ending (b) only are regularized.	(a) Gestern bin ich geschwimmen /schwimm ich (Yesterday I have swim/I swim) (b) Gestern bin ich geschwommt /schwammte ich (Yesterday I have swumed/I swamed)
Irregularization	Irregularized regular verbs.	Gestern habe ich gefolgen/ fielgte ich (Yesterday I have folluw/I folluw)
Present	Answers that are a repetition of the example, namely using the present tense. This was distinguished from a wrong tense use, as it was a default answer similar to a non-answer.	Gestern habe ich schwimme/schwimme ich (Yesterday I have swim/ I swim)
Non answer	"do not know" answers.	
Verb	Use of another verb than the one given.	Schwimmen -> Gestern habe ich gefolgt/folgte ich (to swim -> Yesterday I have followed/ I followed
Agreement	Verb agreement mistakes.	Just applicable in the preterit: Gestern warst ich (Yesterday I were)
Tense	Mistakes in the tense used (except present that is reflected in another category).	Gestern bin ich schwamm/geschwommen ich (Yesterday I have swam/ I swum)
Error	Further mistakes that do not fall into the previously described categories.	Gestern bin ich schwum/ schwum ich (Yesterday I have swom/ I swom)

Category	Description	Example
Other word	Other words given in lieu of the given verb.	Gestern habe ich ein Lasso/ein Lasso ich (Yesterday I have a lasso/I a lasso)
Infinitive	Answers that are the infinitive of the given verb.	Gestern bin ich schwimmen/ schwimmen ich (Yesterday I have to swim/ I to swim)

8.3.3 Statistical considerations

Demographics, performance in the neuropsychological tests and the mistake types were analysed using SPSS version 21.0 (IBM corp., Armonk, NY). A Kruskal-Wallis test was used to compare the demographics and the performance in general neuropsychology across groups. The Dunnett´s T3 adapted for small sample size was used for post-hoc comparisons.

The Cronbach´s alpha test was used to assess the reliability of the item set.

A repeated-measure Anova procedure was used to assess the effect of the tense (perfect or preterit), regularity (regular or irregular), and the frequency (high or low) on the performance in verb inflection and on the proportion of over-regularization within each group. The effect size ($\eta 2$ partial) and power are reported in the results section for each main and interaction effects. The number of low frequency irregular verbs being lower than in the other conditions, we ran the analysis with the proportion of correct answers or over-regularizations for each condition.

To compare the proportion of over-regularization errors to all other kinds of error an index of over-regularization was generated. This was computed by dividing the proportion of over-regularizations by the proportion of non-over-regularization errors on the task. In non-over-regularization errors, all errors except no answer and repetition of the present tense were included. Indeed, these two types of mistake did not reveal any task processing.

8.4 Results

8.4.1 Demographics and general neuropsychology

Demographics for patients and HC are summarised in table 8.4. No significant difference in disease duration or gender between groups was found. AD patients were significantly older than HC and Aß+PPA had a significantly lower number of school years.

Neuropsychological scores are summarized in table 8.5. All patients were worse than HC on the MMSE and BNT, measuring respectively the general level of cognitive functioning and naming abilities. SvPPA obtained the worst performance in the BNT, followed by Aß+PPA and AD.

Table 8-4 Demographics for all participants

** p<.05 compared to the control population, a p<.05 compared to SvPPA, b p<.05 compared to Aß+PPA and c p<.05 compared to AD*

	SvPPA Mean [Range]	**Aß+PPA Mean [Range]**	**AD Mean [Range]**	**HC Mean [Range]**	**Significance *(P-value)***
Age at assessment (years)	66.9 [59-78]	68 [60-78]	75.6* [65-89]	64.7^{C} [50-79]	$X^2(3, 57) = 13.8$, *p=0.003*
Disease duration (months)	56.7 [24-120]	28.5 [18-48]	46 [24-144]	n.a	$X^2(2, 25) = 4.2$, *p=0.123*
Education (years)	14.2 [11-18]	11.5* [11-12]	12.2 [8-17]	14.7^{B} [10-21]	$X^2(3, 57) = 12.4$, *p=0.006*
Gender	4 F, 5 M	3 F, 1 M	6 F, 6 M	19 F, 13 M	$X^2(3, 57) = 1.4$, *p=0.706*

Table 8-5 General neuropsychology

** p<.05 compared to the control population, [a] p<.05 compared to SvPPA, [b] p<.05 compared to Aß+PPA and [c] p<.05 compared to AD; bold italics signs indicate that the significance survived Bonferroni correction for multiple comparison. For logistical reasons, not all participants received all neuropsychological tests.*

	SvPPA Mean (SD)	**Aß+PPA Mean (SD)**	**AD Mean (SD)**	**HC Mean (SD)**	**Significance *(P-value)***
MMSE (Max.30)	20.8 (6.2)*	18.5 (6.1)	21.2 (3)*****	29.2 (.8)a***c***	$X^2(3, 57) = 40.3$, *p<.001*
BNT (Max. 30)	4.9 (4.7)****c***	15.7 (8.5)	20 (4.6)****a***	28.5 (1.1)***ac***	$X^2(3, 57) = 43.3$, *p<.001*
Forward digit span	5.6 (1)	3.7 (1)*	n.a	6.2 (.9)b	$X^2(2, 45) = 13.9$, *p<.001*
Backward digit span	4.2 (1)	2.5 (1.7)	n.a	4.6 (0.8)	$X^2(2, 45) = 8.3$, *p=0.016*
Letter fluency	5.3 (3.3)*	4.5 (3.5)*	n.a	12.3 (4)***A***B	$X^2(2, 40) = 18.3$, *p<0.001*
Category fluency	5.6 (3.5)*	9 (4.1)*****	n.a	23.2 (9.4)***AB***	$X^2(2, 45) = 25.7$, *p<.001*
Digit symbol substitution (Max. 19)	9.7 (1.7)*B	6.5 (1)*****A	n.a	12.2 (2.4)A***B***	$X^2(2, 45) = 15.7$, *p<0.001*
ROCF Copy (Max. 36)	33.3 (1.4)	25.6 (4.1)	n.a	33.7 (2)	$X^2(2, 45) = 11$, *p=0.004*
ROCF immediate recall (Max. 36)	13.8 (4.4)*	7.6 (3.7)*	n.a	19.4 (5.8)AB	$X^2(2, 45) = 15$, *p=0.001*
ROCF delayed recall (Max. 36)	11.7 (4.2)*	6.9 (2.8)*	n.a	19.4 (5.5)***A***B	$X^2(2, 45) = 18.1$, *p<0.001*
GDS (Max. 15)	2.5 (1.8)	1 (0.8)	0.9 (1)	0.5 (0.8)	$X^2(3, 49) = 13.4$, *p=0.004*

8.4.2 Item set

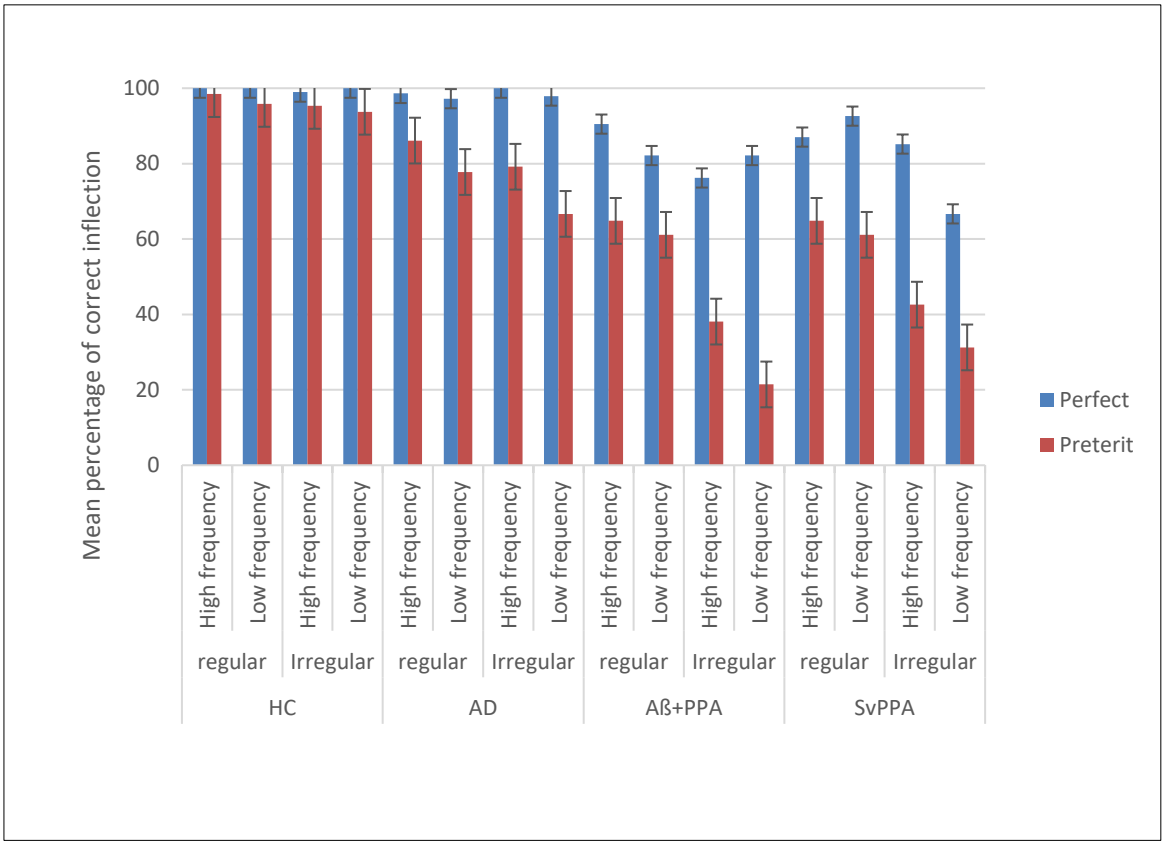
The full set of items displayed a very high reliability (α =.96). The item set displays a good reliability (α =0.82) in the perfect tense and a very good reliability (α =0.93) in the preterit.

8.4.3 Performance in verb inflection

Figure 8.2 shows that all participants had better performance in inflecting verbs in the perfect compared to the preterit and that, especially in SvPPA, success was regulated by the frequency and regularity of the verbs.

Figure 8-2 Performance in verb inflection for all participants

The bars represent the standard error of the mean.



There was a significant effect of the tense on performance in verb inflection in all groups: in HC ($F[1, 32] = 23.64$, $p<0.001$) with a medium effect size ($\eta 2$ partial = 0.43), and an observed power of 0.99; in AD ($F[1, 11] = 8.18$, $p=0.016$) with a medium effect size ($\eta 2$ partial = 0.43) and an observed power of 0.75; in Aß+PPA ($F[1, 4] = 66.05$, $p=0.004$) with a large effect size ($\eta 2$ partial = 0.96) and an observed power of 0.99 and finally in SvPPA ($F[1, 9] = 30.98$, $p=0.001$) with a large effect size ($\eta 2$ partial = 0.79) and an observed power of 0.99.

In SvPPA a further main effect of regularity (F [1, 9] = 9.78, p=0.014) with a medium effect size (η2 partial = 0.55) and an observed power of 0.78 and of verb frequency (F [1, 9] = 6.28, p=0.037) with a medium effect size (η2 partial = 0.44) and an observed power of 0.59 were found.

A significant interaction effect of the tense and the regularity (F [1, 12] = 6.48, p=0.027) with a low effect size (η2 partial = 0.37) and an observed power of 0.64 was only found in AD. A significant interaction effect of the tense and the frequency (F [1, 12] = 6.44, p=0.028) with a medium effect size (η2 partial = 0.37) and an observed power of 0.64 was, again, only found in AD.

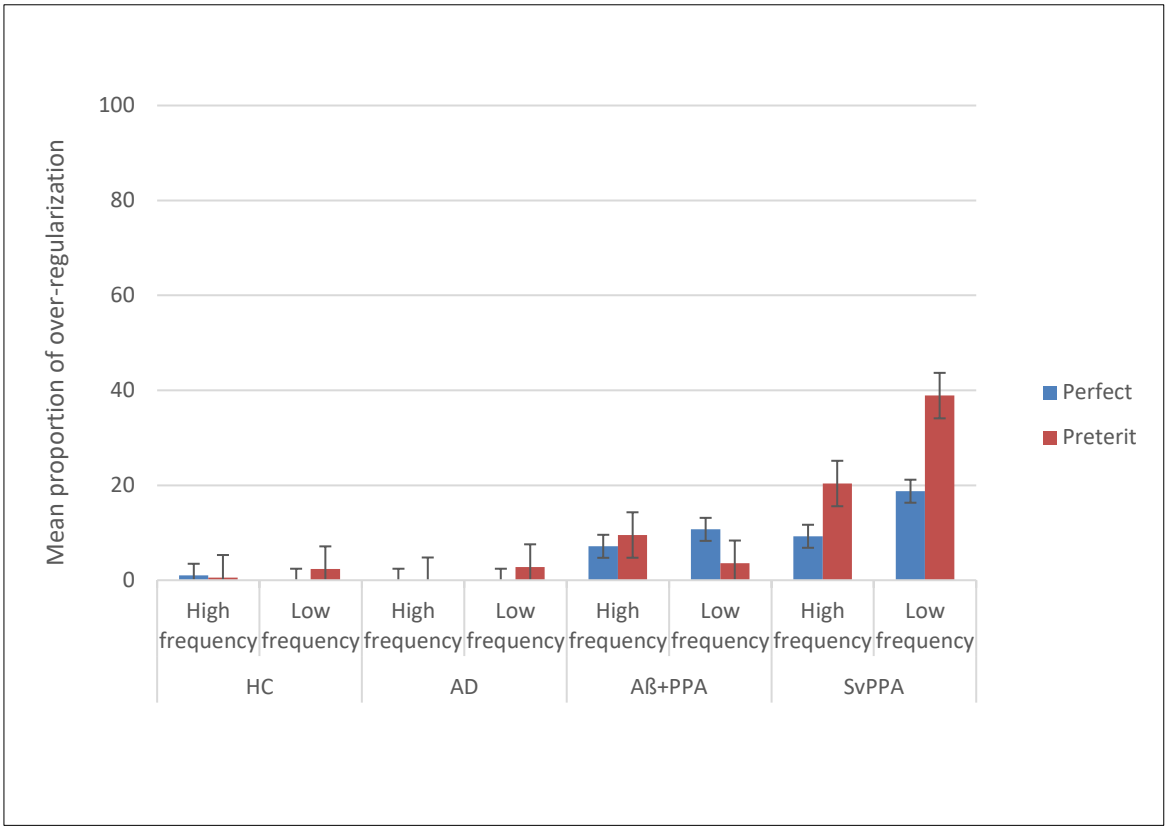
8.4.4 Over-regularization

Figure 8.3 shows that over-regularizations are more numerous in SvPPA followed by Aß+PPA. Moreover, they were more frequent in the low frequency verbs and the low frequency tense form (i.e. preterit).

When analysing the proportion of over-regularization, a significant main effect was only found in SvPPA for frequency (F [1, 9] = 8, p=.021) with a moderate effect size (η2 partial = 0.51) and an observed power of 0.71. A trend toward a significant main effect of the tense form (F [1, 9] = 3.87, p=0.085) with a low effect size (η2 partial = 0.33) and an observed power of 0.41 was, again, only found in SvPPA. No interaction effect reached significance.

Figure 8-3 Occurrence of over-regularization in all participants

The bars represent the standard error of the mean.



8.4.5 Mistake analysis

Aß+PPA and SvPPA were significantly worse than controls in the task. In all groups, mistakes were more numerous in the preterit [HC (mean=0.9, SD= 0.9), AD (mean=4.7, SD=5.7), Aß+PPA (mean=18.5, SD=1.9), SvPPA (mean=10.8, SD=7.5)] compared to the perfect [HC (mean=0.06, SD=0.2), AD (mean=0.3, SD=0.8), Aß+PPA (mean=4.7, SD=1.7), SvPPA (mean=3.4, SD=3.5)]. The four groups differed significantly in the difference between the amount of mistakes in preterit and in perfect (X^2 (3, 57) =29.9, p<0.001). The post-hoc analysis revealed that both Aß+PPA (p=0.02) and SvPPA (p=0.01) had a significantly higher discrepancy between both tense forms than HC.

The mistake types in the different groups is summarized in table 8.6. Interestingly, although Aß+PPA made significantly more mistakes than HC, those were not specifically of one particular type but spread over all categories. SvPPA, as expected, made significantly more over-regularizations than HC and AD. In contrast, Aß+PPA did not significantly differ from HC or AD in the amount of over-regularization.

Table 8-6 Mistake types

** p<.05 compared to the control population, [a] p<.05 compared to SvPPA, [b] p<.05 compared to Aß+PPA and [c] p<.05 compared to AD; bold italics signs indicate that the significance survived Bonferroni correction for multiple comparisons.*

	SvPPA Mean (SD)	**Aß+PPA Mean (SD)**	**AD Mean (SD)**	**HC Mean (SD)**	**Significance *(P-value)***
Total mistakes (max. 44)	14.2* (10.8)	23.2* (0.5)	5.1 (5.9)	0.9^{A}***B*** (1)	X^2 (3,57) = 30.8, *p<0.001*
Over-regularization	4.1^{*C} (3.3)	1.7 (1)	0.3^{A} (0.6)	0.2^{A} (0.6)	X^2 (3,57) = 36.9, *p<0.001*
Semi-regularization	0.9 (1.3)	1.5 (2.4)	0	0.1 (0.3)	X^2 (3,57) = 11.3, *p=0.010*
Irregularization	0.8 (1.4)	0.7 (1)	0	0	X^2 (3,57) = 19, *p=0.001*
Present tense	5 (8.6)	8.5^{*C} (2.4)	2^{B} (4.8)	0.1^{B} (0.2)	X^2 (3,57) = 23.6, *p<0.001*
Tense	0.9 (2.3)	1 (1.4)	0.1 (0.3)	0.2 (0.4)	n.s
Verb agreement	1 (1.3)	3.2 (2.6)	0.9 (1.3)	0.3 (0.5)	X^2 (3,57) = 13.7, *p=0.003*
Other word	0.7 (1)	2 (1.4)	1 (1.3)	0.03 (0.2)	X^2 (3,57) = 25.4, *p<0.001*
Other error	2.3 (2.2)	1.7 (1.3)	0.7 (0.9)	0.1 (0.3)	X^2 (3,57) = 24.2, *p<0.001*
No answer	0	2.7 (2.2)	0.5 (1.4)	0	X^2 (3,57) = 26.5, *p<0.001*
Infinitive	0.2 (0.4)	1.2 (1)	0.2 (0.6)	0	X^2 (3,57) = 20.5, *p<.001*

8.4.6 Ratio of over-regulation errors and other errors

The proportion of over-regularization and rule breaking errors (verb agreement, tense, other word, other error, infinitive, irregularization) were compared across the four groups. Aß+PPA made significantly more rule breaking errors than HC and AD, whereas SvPPA

made significantly more over-regularizations compared to HC and AD (see table 8.7). Figure 8.4 representing the ratio of both kind of mistakes, highlights the double dissociation between over-regularization errors (occurring in SvPPA principally) and rule breaking errors (occurring in Aß+PPA principally).

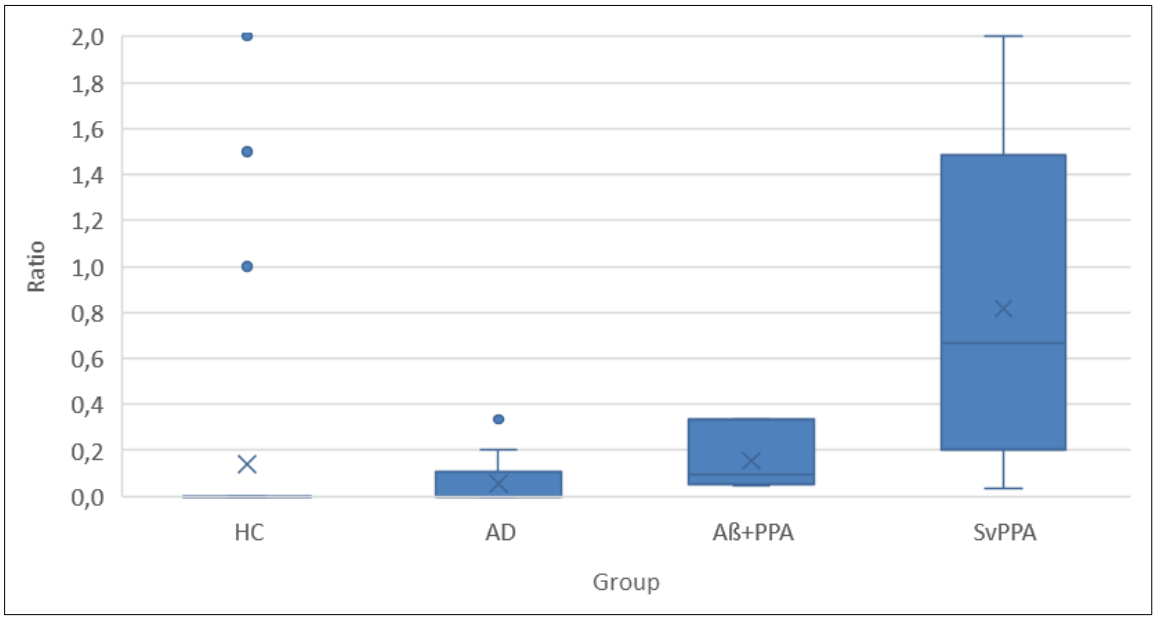
Table 8-7 Proportion of over-regularizations and other errors

** p<.05 compared to the control population, a p<.05 compared to SvPPA, b p<.05 compared to Aß+PPA and c p<.05 compared to AD; no significance survived Bonferroni correction for multiple comparisons.*

	SvPPA Mean (SD)	**Aß+PPA Mean (SD)**	**AD Mean (SD)**	**HC Mean (SD)**	**Significance (P-value)**
Proportion of over-regularizations	20.6% (16.5)^{*C}	8.7 % (4.8)	1.7 % (3.2)^{A}	0.9 % (3.2)^{A}	X^2 (3,57) = 36.9, *p<0.001*
Proportion of other errors	26.5% (25.9)	50% (4.9)^{*C}	11.7% (13.3)^{B}	1.5 % (2)^{B}	X^2 (3,57) = 28.3, *p<0.001*

Figure 8-4 Ratio of over-regularization/other mistakes

The ratio was computed by dividing the proportion of over-regularizations by the proportion of non-over-regularization errors on the task.



8.5 Discussion

The results of this pilot study confirmed the difficulty shown in diverse PPA subtypes in inflecting verbs in the past in German. The highest proportion of mistakes in verb inflection occurred in Aß+PPA, followed by SvPPA, while errors were rare in AD although their MMSE results implied a more advanced disease stage. A significant interaction effect of tense and regularity was, however, found in AD: they were significantly worse than controls on irregular verbs in the low frequency tense. This is in line with studies showing that AD are impaired compared to HC in inflecting irregular but not regular past participles in Italian (Colombo, Fonti, & Stracciari, 2009; Walenski, Sosta, Cappa, & Ullman, 2009) and English (Ullman et al., 1997).

In all groups, there was an effect of the tense. This was very likely due to the frequency differences of the two tense forms. Indeed, in German the preterit is almost exclusively used in written language, apart from very high frequency verbs like auxiliary and modal verbs.

Effects of the regularity and the verb frequency were only found in SvPPA. This is consistent with past studies showing that frequency and regularity have a crucial impact on correct verb inflection in SvPPA (Auclair-Ouellet, 2015; Cortese, Balota, Sergent-Marshall, Buckner, & Gold, 2006; Patterson et al., 2001).The interaction effect of verb frequency by regularity found in these studies could, however, not be replicated in SvPPA in our sample. This might be due to the lack of power due to the reduced sample size.

When analysing qualitatively the errors made by the different groups of participants it appeared that Aß+PPA made a wide range of mistakes (e.g. substituting another verb, verb agreement or tense errors for example) that implied rule breaking whereas errors in SvPPA proportionally involved more regularization of irregular verbs (homologue English example: I swim → I swimmed, I have swimmed). This confirms results from preivous studies showing a specific over-regularization tendency in SvPPA when inflecting verbs (Auclair-Ouellet, Macoir, Laforce, Bier, & Fossard, 2016) but also when reading exception words aloud (Patterson & Hodges, 1992; Woollams et al., 2007). As reported in a previous study over-regularizations in SvPPA are more present in low frequency verbs (Wilson et al., 2014).

SvPPA patients showed specific over-regularization mistakes that do not feature prominently in Aß+PPA although the latter make a range of other errors. This might indicate a broader impairment including not only semantic deficits but also grammatical and understanding deficits in Aß+PPA. Moreover, no significant effect of regularity or frequency on the global performance was found in Aß+PPA. This might reinforce the idea of a less specific impairment. The sample size was, however, small, especially for Aß+PPA, and may have reduced the statistical power of the analysis. Furthermore, the cohort included some outliers that might have distorted the statistical analysis. Given the reduced sample size, the piloting value of the study and the known heterogeneity existing in PPA patients, outliers were not excluded. Another limitation of the study was the unequal number of verbs across categories. Indeed, two verbs had to be excluded because they were too difficult to inflect in the preterit for healthy controls. This was, however, tackled by using the percentage of success per category.

A strength of our study is the pathological homogeneity of the Aß+PPA group. To our knowledge, this is the first study that investigated inflectional morphology in a group of PPA patients with confirmed amyloid pathology and mixed PPA profile.

To summarize, the effects of regularity and frequency in the inflection of past participle in SvPPA and the specific tendency to rely on rules and over-regularized irregular forms was replicated in German. Moreover, we argue that the specific over-regularization effect might offer potential as an analogue to surface dyslexia in languages with high grapheme-phoneme transparency like German. The results need to be confirmed in a larger sample where the inclusion of non-semantically impaired PPA patients like NfvPPA could be of interest.

Chapter 9.

Investigating Amyloid related PPA

9.1 Introduction

The criteria for the subtyping of Primary Progressive Aphasia (PPA) published in 2011 (Gorno-Tempini et al., 2011) present 3 different subtypes of PPA, the non-fluent variant of PPA (NfvPPA) characterized by impaired motor speech and/or agrammatism and the semantic variant of PPA (SvPPA) characterized by impaired confrontation naming and single-word comprehension, both extensively described in the past (Hodges et al., 1992; Neary, 1998; Snowden et al., 1989; Snowden, Neary, Mann, Goulding, & Testa, 1992), and the more recently pinpointed logopenic variant of PPA (LvPPA) (Gorno-Tempini et al., 2004). Impaired word finding and repetition of sentences and the presence of phonological errors with preserved motor speech and object knowledge characterize LvPPA.

Although some authors applied the diagnostic criteria in their cohorts with no unclassifiable cases (Rabinovici et al., 2008) most other groups reported a considerable number of unclassifiable cases in PPA cohorts: from 4.3%, (Leyton et al., 2011), 10% (Harris et al., 2013), 16.7 % (Villarejo-Galende et al., 2017), 31% (Wicklund et al., 2014) to 41.3% (Sajjadi, Patterson, Arnold, et al., 2012). Consequently, the published criteria are extensively challenged, especially concerning the neuropsychological profile of PPA patients that have neither NfvPPA nor SvPPA. Part of the unclassifiable PPA patients have a neuropsychological profile that goes beyond the criteria proposed for the three variants. They present with both word comprehension deficits and agrammatism. This led to the concept of mixed PPA (mPPA) (Mesulam et al., 2009; Mesulam et al., 2012; Sajjadi, Patterson, Arnold, et al., 2012).

While NfvPPA and SvPPA tends to be statistically more associated with pathology of the fronto-temporal lobar degeneration (FTLD) spectrum, respectively Tauopathy and TDP-43 proteinopathy, Alzheimer´s disease (AD) is the most reported underlying cause of LvPPA

(Botha et al., 2015; Harris et al., 2013; Leyton et al., 2011; Rabinovici et al., 2008; Spinelli et al., 2017) mPPA (Vandenberghe, 2016) and unclassifiable PPA cases (Villarejo-Galende et al., 2017). Some authors reported that LvPPA and mPPA do not only share the same underlying disease but also the same temporo-parietal pattern of atrophy (Sajjadi et al., 2014). They suggested that LvPPA and mPPA might form a single entity, which could be referred to as AD related PPA and might be too heterogeneous to be captured by the current criteria. Noteworthy is that no AD pathology confirmation was available for the cohort in this study, and that the hypothesis of underlying AD pathology in mPPA was based on the similarity to LvPPA (probabilistically more related to AD).

In the field of PPA, the main challenge remains the possibility to predict best the underlying pathology thanks to the clinical syndrome. Indeed, methods like Amyloid Positron Emission Tomography (PET) are still very expensive and not easily available for use in clinical routine. The possibility to subtype PPA patients more accurately is of critical importance not only once a therapy is available for one of the common underlying diseases but also to predict the development of the disease and hence give the best care to patients and their relatives.

9.2 Aim

AD being reported as the most common underlying cause of PPA (Villarejo-Galende et al., 2017) and in keeping with the hypothesis made by Sajjadi, et al (2014), in this study a clinically unbiased method was adopted to try to describe the heterogeneity of the cognitive profile of patients with confirmed amyloid pathology, the histopathological hallmark of AD. All patients fulfilling the ground criteria for PPA were included, and administered an extensive neuropsychological battery as well as an F18-Florbetaben Positron Emission Tomography (PET). The neuropsychological profile of the participant with a positive amyloid PET is investigated in this study.

9.3 Method

9.3.1 Participant

Twenty-six PPA patients were included in the study. Inclusion criteria are described in section 6.3.2. Patients went through a clinical 18-Florbetaben PET to confirm the presence of amyloid pathology in the brain, the hallmark of AD. The latest diagnostic recommendations (Gorno-Tempini et al., 2011) were applied to subtype the patients.

9.3.2 Neuropsychology

A large neuropsychological battery including the Repeat and Point (Hodges, Martinos, Woollams, Patterson, & Adlam, 2008), the Kaffee & Kuchen test (KKT) a test of semantic association of pictures based on the Camel & Cactus test (CCT) (Adlam et al., 2010) and the Pyramid and Palm trees Test (Howard & Patterson, 1992) developed for the German population (see appendix VI for the list of stimuli), a German adaptation of the SECT V (see chapter 7 and appendix II for original version and appendix V for the list of stimuli in German), the Graded Object-Naming Task (see chapter 3), a 30 items version of the BNT (Merten, 2004), a short in-house word and sentence repetition test (see table 9.1) as well as the memory span subtest from the WMS-III (Wechsler, 1997a) was administered

9.3.3 Repetition

A list of word and sentence-repetition was compiled and administered to all PPA patients as well as to 28 HC. Inclusion criteria for HC are described in section 6.3.2. A target was considered as successfully repeated when the participant produced the correct word/sentence with a correct prosody and without phonological distortion. Repetition was recorded and scored by two different raters. Consensus discussion resolved discrepancies in rating. See table 9.1 for the list of targets.

9.3.4 Working memory

Both digit and visuo-spatial spans forward and backward were administered to all participants. The longest string of digits or blocks successfully reproduced was considered as the span ability.

Table 9-1 List of items used in the repetition task

Words	Syllables count	Sentences	Syllables count
Schmetterling (Butterfly)	3	Das Wetter ist heute schön (The weather is nice today)	7
Beobachten (Observe)	4	Wir sind nach München gefahren (We drove to Munich)	8
Psychologische (Psychological)	5	Sie kommt nicht mit uns heute Abend (She does not come with us tonight)	9
Fußballweltmeisterschaft (Football World Cup)	6	Der Sohn hat einen Kuchen gebacken (The son baked a pie)	10

9.3.5 Statistical considerations

9.3.5.1 Neuropsychology

Patients were considered impaired on neuropsychological tests when they differed from HC at more than -1.5 Z-score.

9.3.5.2 Ratio

To differentiate between the different PPA subtypes´ profiles, a ratio of semantic and grammatical impairments was created. The results in the German version of the SECT V (presented in its original version in chapter 7) as well as the results of the KKT were Z–transformed. The Z–scores obtained in the KKT were subtracted from those obtained in the SECT V. The SECT V was chosen for this comparison over the MAST, a test of grammar production, because all participants could perform it. Some participants were too impaired to perform the MAST completely.

9.3.5.3 Repetition

The Generalized Estimating Equations (GEE) procedure available in SPSS version 21.0 (IBM corp., Armonk, NY) was used to assess the effect of syllables number on the repetition performance in HC and PPA patients. This procedure is adapted for correlated data obtained from repeated measures with a binary outcome measure (success or failure to repeat the target). The model was specified as binary logistic. The success or failure to

repeat an item was entered as the response. The group membership (patient or HC) and the number of syllables of the target (from three to six for words and from seven to ten for sentences) were entered as predictors. Analyses were run separately on words and sentences.

9.4 Results

9.4.1 Amyloid PET & Subtyping

Fifteen patients had a negative amyloid PET result: nine fulfilled the criteria for SvPPA and six for NfvPPA.

Eleven Patients had a positive amyloid PET result and are referred to as Aß+PPA. Two Aß+PPA did not present with repetition impairment and therefore did not fulfil the main diagnostic features for LvPPA. All patients presented with grammar deficits, as revealed by their low performance in both the MAST and SECT V and all except one presented with significant semantic deficits as revealed by their low performance in the KKT. Only nine patients fulfilled the main diagnostic features, of which only one fulfilled three fourths of the secondary diagnostic features and could fit the LvPPA criteria. The deficits of the eight remaining Aß+PPA were more extended than described for LvPPA (see table 9.2).

9.4.2 Ratio

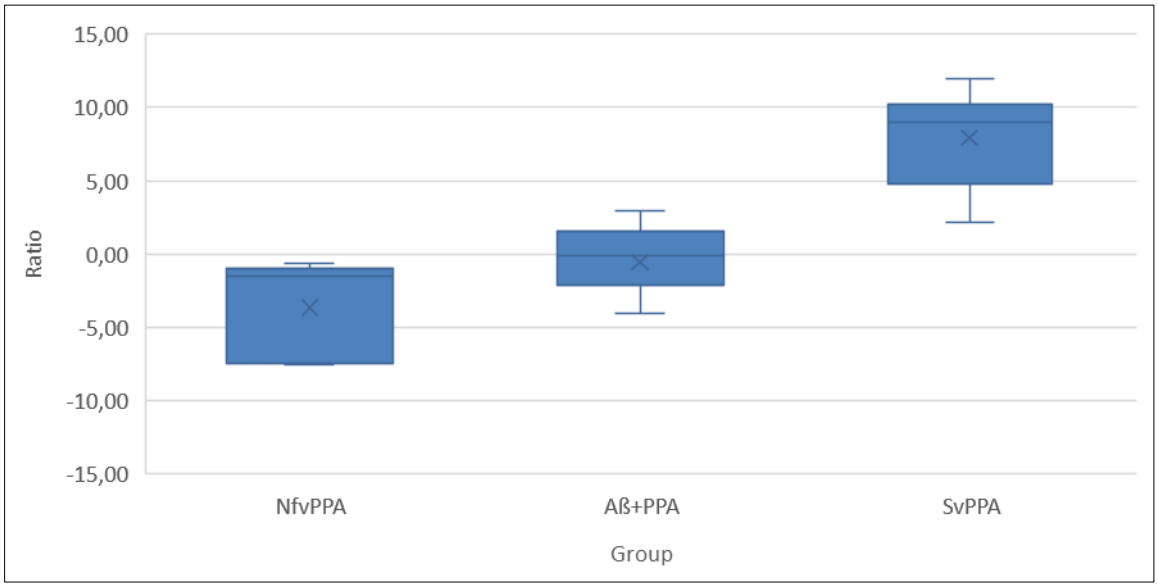
As depicted in figure 9.1, patients with a negative amyloid PET split in two opposite neuropsychological profiles where SvPPA have a positive ratio, indicating a better performance in the SECT V (comprehension of complex grammar) than in the KKT (semantic association) and NfvPPA present a negative ratio indicating a worse performance in the SECT V than the KKT. Aß+PPA spread over a null ratio that indicates similar performance in semantic association and comprehension of complex grammar. Noteworthy is that although the discrepancy between grammar and semantic was not as high as in SvPPA or NfvPPA, at the single subject level Aß+PPA could show profiles of either greater grammatical or semantic impairment

Table 9-2 Application of the LvPPA criteria to the Aß+PPA

	Main diagnostic features		**Other diagnostic features**					Fitting LvPPA criteria?
	Impaired single word retrieval?	Impaired repetition of sentences?	Phonological errors?	Spared single word comprehension and object knowledge?	Spared motor speech?	Absence of frank agrammatism?		
Test	BNT	Repetition of 5 sentences	Connected speech sample	KKT	Connected speech sample	German adaptation of the MAST	German adaptation of the SECT V	
Cut-off	>-1.5 Z score	< 4 correct	qualitative	>-1.5 Z score	qualitative	>-1.5 Z score	>-1.5 Z score	
Patient 1	Yes	Yes	Yes	No	Yes	No	No	**No**
Patient 2	Yes	Yes	Yes	No	Yes	No	No	**No**
Patient 3	Yes	Yes	Yes	No	Yes	No	No	**No**
Patient 4	Yes	Yes	Yes	No	Yes	No	No	**No**
Patient 5	Yes	No	Yes	No	Yes	No	No	**No**
Patient 6	Yes	No	Yes	No	Yes	No	No	**No**
Patient 7	Yes	Yes	Yes	Yes	Yes	No	No	**Yes**
Patient 8	Yes	Yes	Yes	No	Yes	No	No	**No**
Patient 9	Yes	Yes	Yes	No	No	No	No	**No**
Patient 10	Yes	Yes	Yes	No	Yes	No	No	**No**
Patient 11	Yes	Yes	Yes	No	Yes	No	No	**No**

Figure 9-1 Ratio of the performance in grammar compared to semantics in all patient groups

A ratio of "0" indicates an equivalent level in semantic knowledge and comprehension of grammatically challenging sentences. The ratio was computed by subtracting the Z–scores obtained in the KKT (semantic measure) from those obtained in the SECT V (grammar measure).



9.4.3 Repetition

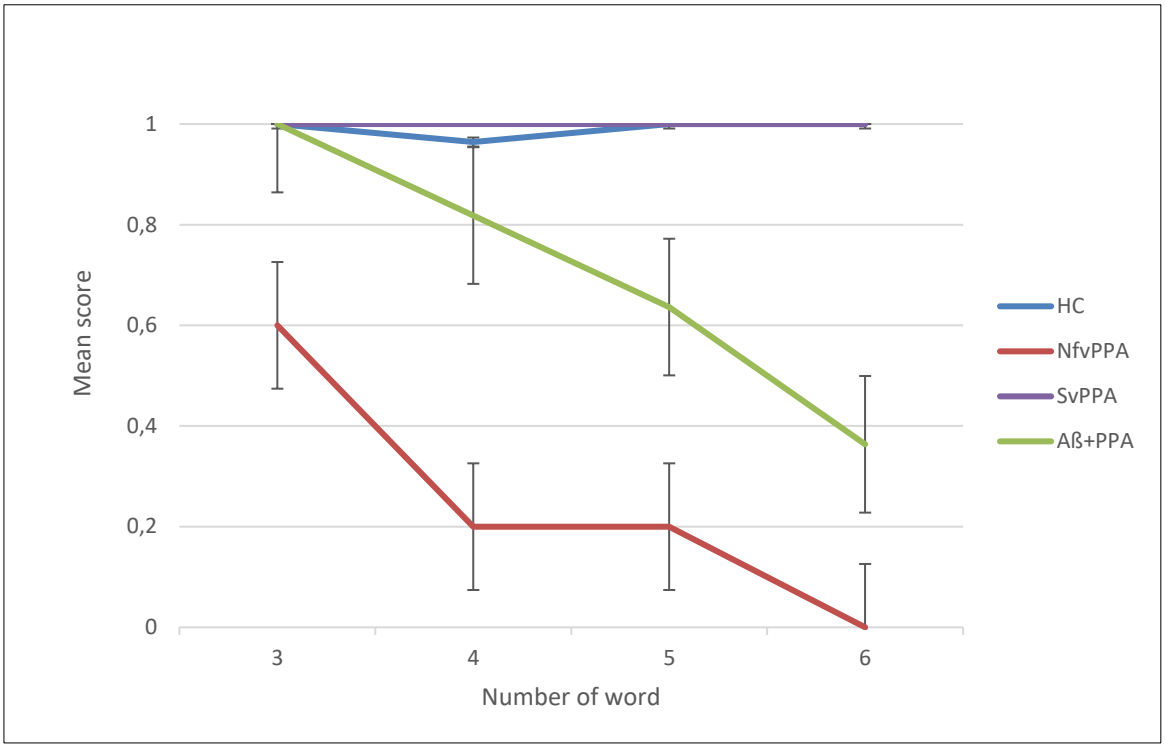
9.4.3.1 Words

When computing the total score obtained on the 4-word-repetition task, there was a significant difference in performance across groups (X^2 [3, 52] = 37.11, p<0.001). Post-hoc analysis revealed that both NfvPPA (p=0.02) and Aß+PPA (p=0.03) significantly differ from HC. This was not the case for SvPPA (p=0.89). No significant difference was found between Aß+PPA and NfvPPA (p=0.12). As observable on figure 9.2 both SvPPA and HC performed at ceiling, while performance decreased with syllable length in both Aß+PPA and NfvPPA.

As expected from figure 9.2 there was a significant main effect of the number of syllables on the repetition score (X^2 [1, 52] = 6.5, p=0.011). No main effect of the group membership appeared (X^2 [1, 52] = 0.08, p=0.772). There was, however, a significant interaction effect of the group membership and the number of syllables on the repetition score (X^2 [1, 52] = 27.22, p<0.001).

Figure 9-2 Difficulty range in the word repetition task

The bars represent the standard error of the mean.



9.4.3.2 Sentences

In the 4-sentence-repetition task, there was a significant difference of performance across diagnostic groups in the total score (X^2 [3, 52] = 33.03, p<0.001). Post-hoc analysis revealed that both NfvPPA (p=0.004) and Aß+PPA (p=0.002) significantly differ from HC. This was not the case for SvPPA (p=0.64). A significant difference was also found between Aß+PPA and NfvPPA (p=0.036). As displayed on figure 9.3 the performance in repetition of all participants (including HC) dropped as the number of syllables increased. NfvPPA performed at floor on almost all items. The longest sentence (10 syllables) seemed to provide better results in patients than the third sentence (9 syllables). This could be explained by the fact the longest sentence had one more word than the third sentence (see figure 9.4).

Analyses revealed a significant main effect of the number of syllables (X^2 [1, 52] = 2.9, p=0.022) but not of the group membership (X^2 [1, 52] = 5.2, p=0.088) on the repetition score. There was no significant interaction effect of the group membership and the number of syllables on the repetition score (X^2 [1, 52] = 0.73, p=.392).

Figure 9-3 Difficulty range in the sentence repetition task using the number of syllables

The bars represent the standard error of the mean.

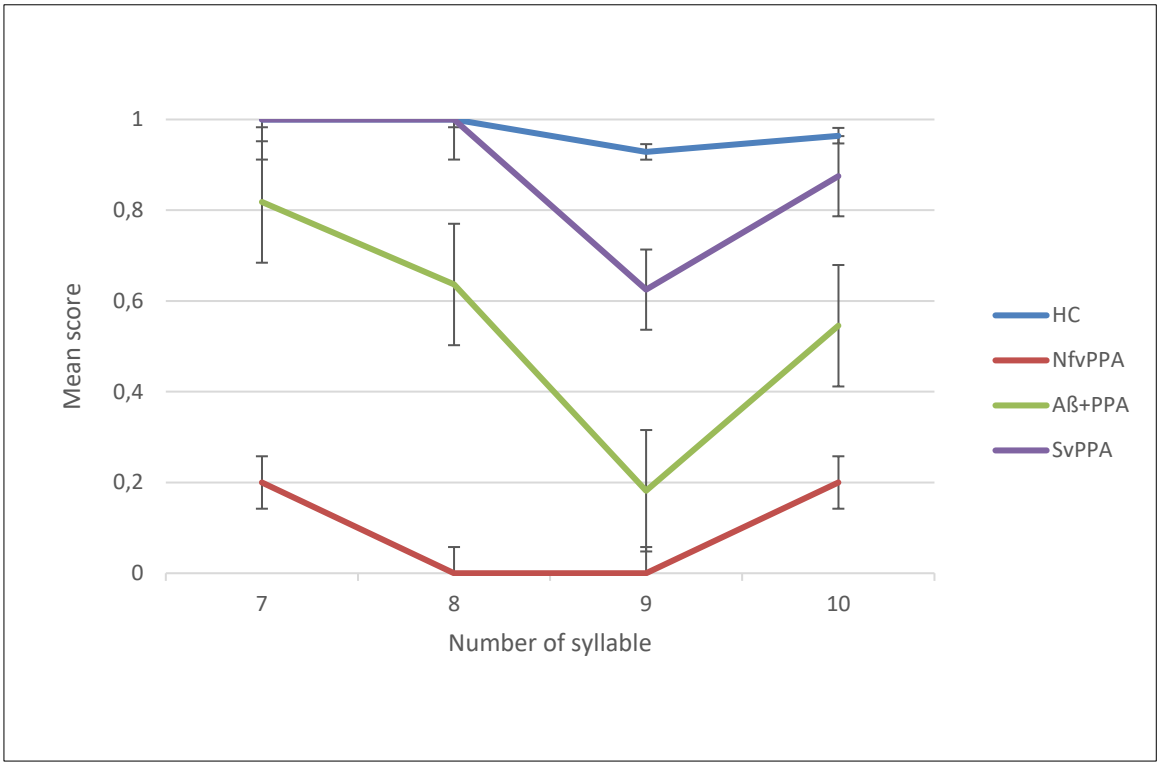
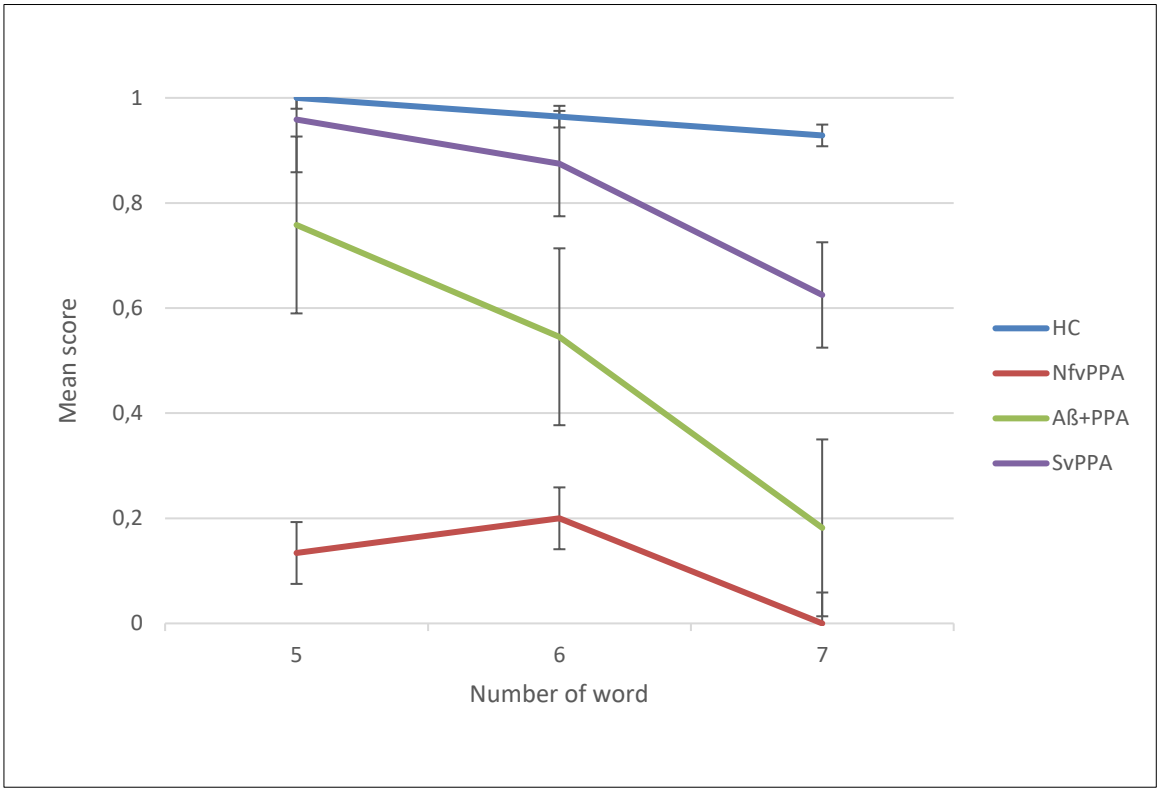


Figure 9-4 Difficulty range in the sentence repetition task using the number of words

The bars represent the standard error of the mean.



As displayed on table 9.3 there is perfect sensitivity of word and sentence repetition impairment in NfvPPA. Although Aß+PPA are often impaired on word and sentence repetition, some patients did perform normally and therefore the sensitivity of words and sentences repetition is lower. Impaired repetition of words and sentences is not specific of NfvPPA or Aß+PPA.

Table 9-3 Percentage of patients impaired in repetition compared to HC

The cut-offs are based on the results obtained in 28 HC. 99.96% of HC repeated correctly four words, and 99% repeated correctly four sentences.

	Words	**Sentences**	**Total**	**N**
Cut-off	<4	<4	<7	
NfvPPA	100%	100%	100%	6
Aß+PPA	72.7%	81.8%	81.8%	11
SvPPA	0%	37.5%	12.5%	8

9.4.4 Working Memory

The results at the group level in digit and visuo-spatial spans are displayed in table 9.4. Noteworthy is that Aß+PPA were impaired in both spans forward and backward compared to controls whereas NfvPPA displayed a dissociation between verbal and visuo-spatial span, with a significant impairment compared to controls only in digit span (both forward and backward) at the group level.

Similarly to the repetition of words and sentences, digit span forward displayed a perfect sensitivity to NfvPPA. Difficulties were, however, also present in a large amount of patients with Aß+PPA indicating a low specificity (see table 9.5)

Table 9-4 Performance in digit and spatial spans

** p<.05 compared to the control population, [a] p<.05 compared to NfvPPA, [b] p<.05 compared to SvPPA and [c] p<.05 compared to Aß+PPA; bold italics signs indicate that the significance survived Bonferroni correction for multiple comparisons.*

	HC Mean (SD)	**NfvPPA Mean (SD)**	**SvPPA Mean (SD)**	**Aß+PPA Mean (SD)**	**Significance *(P-value)***
Digit Span Forward	6.4***AB***C (0.9)	3.2******B (0.8)	5.2******A (0.5)	3,6* (1.7)	X^2 (3,53) = 32.3, *p<0.001*
Digit Span Backward	4.6AC (1)	2* (1.2)	4C (0.8)	1.9*B (1.6)	X^2 (3,53) = 26.4, *p<0.001*
Visuo-spatial Span Forward	5.4***C*** (0.9)	4 (1.4)	5.6 (1.4)	3.7* (0.7)	X^2 (3,53) = 17.6, *p=0.001*
Visuo-Spatial Span Backward	5.2C (0.9)	3.6 (1.5)	4.9 (1.5)	2.9* (1.5)	X^2 (3,53) = 15.9, *p=0.001*

Table 9-5 Percentage of patients impaired in spans compared to HC

	Digit span forward	**Digit span backward**	**Visuo-spatial span forward**	**Visuo-spatial span backward**	**N**
Cut-off	**<5**	**<3**	**<4**	**<4**	
NfvPPA	100%	60%	20%	40%	5
Aß+PPA	62,5%	50%	37,5%	62,5%	8
SvPPA	0%	0%	0%	25%	8

9.5 Discussion

In the past years, the idea that Alzheimer related PPA is more heterogeneous and extended than the described LvPPA has become more consensual (Sajjadi et al., 2014). Therefore, we chose to investigate PPA patients with confirmed amyloid pathology as a group. Consistent with other studies, only one patient in our cohort (3.8%) fulfilled the criteria for LvPPA. The other participants with confirmed Alzheimer pathology did not fulfil the criteria for one of the three proposed variants. Indeed, most of them presented a mixed profile with both grammar and semantic deficits. All patients were impaired in our tests of receptive and productive grammar and 10/11 patients had impaired semantic knowledge compared to healthy elderly. These deficits were possibly underestimated in past studies because of a lack of suitable test to investigate grammatical abilities in neurodegenerative diseases. Moreover, most non-verbal tests of semantic knowledge are suffering from ceiling effects in HC and are only sensitive to extended semantic deficits, like those seen in SvPPA.

9.5.1 Ratio

As shown in past studies, the current criteria for subtyping PPA do not always apply perfectly at the single subject level. Studies showed, for example, that to some extent SvPPA could also present grammar impairments (Cupit et al., 2016; Meteyard & Patterson, 2009; Meteyard et al., 2014). These deficits are not always explained by a longer disease duration but rather by the inherent properties of language that require intact abilities in a multitude of domains like grammar, semantics, lexical access and pronunciation for example, to speak normally. PPA subclassifications might gain in weighting the proportion of deficits in each domain rather than applying strictly the concept of impairment. In our cohort, when applying a ratio that compares the amount of difficulty in grammar and semantics, although some NfvPPA displayed semantic impairments and some SvPPA displayed grammar deficits compared to controls, the ratio of both impairments reveals a dichotomy where semantic deficits are more present in SvPPA and grammatical deficits in NfvPPA. In Aß+PPA the amount of difficulty in both domains is more balanced at the group level. At the single subject level, some patients display larger semantic impairment and others more deficits in grammar. A strict discrepancy between both domains is rarer than in the two other subtypes, justifying the label of “mixed PPA”.

9.5.2 Repetition and working memory

Investigations of phonological loop and working memory capacities have gained interest in the past years. Our results show that repetition deficits in Aß+PPA are not restricted to phrases and sentences; as specified in the current diagnostic recommendation (Gorno-Tempini et al., 2011); but already appear in single words. This is of considerable importance for the differential diagnosis with NfvPPA for which the impairment in the repetition of single words is not specific.

As expected, the length of words and sentences has an impact on the performance in repetition in Aß+PPA and NfvPPA. Noteworthy is that some SvPPA also display repetition deficits in longer sentences. This might suggest that semantic understanding of long stimuli works as a help for working memory retention. When the semantic system is disrupted, the retaining capacity might slightly decrease.

At the group level NfvPPA were only impaired in the verbal but not in the visuo-spatial span, whereas Aß+PPA were significantly impaired in both type of spans. Although there seem to be a relative preservation of visuo-spatial span in NfvPPA that is not found in Aß+PPA, this is not always true at the single subject level. Indeed, in our sample, one NfvPPA was impaired in forward digit span and two in backward digit span. The sample size was, however, very limited (five patients).

A previous study (Foxe, Irish, Hodges, & Piguet, 2013) reported an advantage in visuo-spatial span over verbal span in LvPPA. The authors argue that there is a specific short-term memory deficit due to an impaired phonological system in these patients. These results could, however, not be replicated in our study. Although only one patient fulfilled the criteria for LvPPA, we are confident that the LvPPA patients of this study and our Aß+PPA groups are comparable. Indeed, 100% of the patients who received an Amyloid PET in the study by Foxe et al tested positive.

Another recent study found in a cohort of AD-related PPA that 68% of the patients were impaired in sentence repetition and 90% in forward digit span, which was for the latter much higher than in the non-AD PPA (33%) (Giannini et al., 2017). The authors argue that the impairment of the phonological loop revealed by the impaired performance in the forward digit span is a central feature of Alzheimer related PPA, that could be used for disease prediction. Although we also found that a substantial number of Aß+PPA were impaired in both sentence repetition (around 82%) and forward digit span (around 63%)

our results do not indicate a specific impairment for Aß+PPA. Indeed, as mentioned earlier, NfvPPA are also impaired in high proportion. Moreover, in the study by Giannini et al, the AD-related group was directly compared to a non-AD-related PPA group composed of both SvPPA and NfvPPA. SvPPA being relatively preserved in both forward digit span and short sentence repetition to the contrary of the NfvPPA, mixing both group might have strongly biased the obtained results.

To summarize, the study confirms that the criteria for LvPPA are too narrow to allow a sensitive prediction of an underlying Alzheimer´s disease in the context of PPA. Patients with Aß+PPA have a heterogeneous profile of impairment and need to be tested sensitively in all domains of language: semantic, grammar, naming and repetition. Moreover, they present repetition deficit already at the single word level. Thus, repetition of words versus sentences does not seem valuable for the differential diagnosis with NfvPPA. A limitation to our study is the small sample size as well as the fact that not all participants performed all tests. This reduces the comparability across groups. Further work needs to include higher number of participants.

Chapter 10.

Concluding remarks

10.1 Contribution of the thesis

Dementia is currently a major public health challenge. For the most prevalent dementia causing disease, namely Alzheimer´s disease, only probabilistic diagnoses can be made in clinical settings. Technical and biological advances, like amyloid PET for example, being unavailable in most memory clinics, cognitive testing is essential for the diagnosis of most types of dementia syndromes. For this reason, this thesis intended to improve two current issues: tracking change from healthy aging to early cognitive impairment and better characterizing the syndrome of primary progressive aphasia, in which the correspondence between the clinical picture and the underlying disease is far from systematic.

The first part of this thesis aimed at improving early detection of cognitive decline with the means of a cognitive test battery. Two tests investigating memory and language abilities were presented. Both tests were able to show graded difficulty in order to track changes in population of diverse premorbid cognitive level. A superiority compared to the gold standard in their respective domain was found on cross-sectional comparisons in healthy elderly and patients with slight to moderate cognitive impairment. The international value of the tests was shown, especially in the language domain. Indeed, the material was administered in both German and Slovak. At the time of writing, data are collected in both Italian and Australian English to extend the findings. Finally, longitudinal observations showed the ability of both tests to track slight change over time. With the enhancement of the two presented tests based on the reported weaknesses (low sample size and too few longitudinal data) as well as the inclusion of the two remaining tests investigating attention and visuo-spatial abilities, we see a sweeping opportunity to move toward an earlier diagnosis of dementia. When therapeutics are available, tracking the earliest changes will be of crucial importance. To our knowledge, the DZNE-Cog is the first tool for tracking

early cognitive changes that has been developed to be usable in different cultural backgrounds and languages. This will allow the battery to be used for the evaluation of therapeutic trials, which are frequently led internationally.

The second part of this thesis focused on the syndrome of Primary Progressive Aphasia, for which clinical subtyping and correspondence with the underlying disease are particularly arduous. Chapter 7 presented two newly developed neuropsychological tools to address agrammatism. Agrammatism is a core deficit in both the non-fluent and the mixed variants of Primary Progressive Aphasia but short standardized diagnostic tools are still missing. Performance obtained in the tests were compared to those obtained in connected-speech sample to increase ecological validity. Results showed that they were able to satisfyingly detect grammar impairment in non-fluent and mixed PPA while avoiding ceiling effects in the healthy elderly. At the time of writing, data are collected for a German adaptation of both the SECT V and the MAST. Furthermore, in chapter 8 after confirming in a German sample the effect of regularity and frequency on verb inflection in the semantic variant of Primary progressive aphasia, we showed that over-regularization could work as an analogue to surface dyslexia in this group of patients. This is particularly valuable in languages with high grapheme-phoneme correspondence where surface dyslexia is hard to detect. In the concerned countries, inflectional morphology could be easily used in clinical settings. Finally, chapter 9 investigated Alzheimer related Primary Progressive Aphasia and showed that contrary to the current belief the repetition of words and sentences is not useful for the differential diagnosis with the non-fluent variant of Primary Progressive Aphasia. Moreover, the current criteria for the subtyping of Primary Progressive Aphasia need to be revised. The description of the logopenic variant is too restricted to capture patients that have neither the semantic nor the non-fluent variant of Primary Progressive Aphasia but often present with both grammar and semantic impairments. Future work to improve the diagnostic recommendation would benefit from confirming pathology with the help of amyloid PET or post-mortem brain biopsy.

In summary, this work intended to improve the contribution of cognitive testing in the field of dementia diagnosis. Earlier and more accurate diagnosis has important consequences on both patient and their family´s quality of life. Not only is it of major importance when treatments will be available but it also allows testing new treatments at a stage where the disease´s consequences might still be stoppable.

10.2 Future directions

The results obtained in the piloting of two of the planned tests for the DZNE-Cog offer a basis for further work. As mentioned in the previous section, both the Graded Object-Naming Task and the Virtual City Task will have to be improved based on the reported weaknesses. A larger sample size and longer longitudinal testing periods in different languages will be helpful. They will have to be administered along with the two remaining tests of the battery (the attention and visuo-perceptual tasks), to assess the quality of the battery as a whole. A full computerization of the battery will have to be achieved to ease standardized repeated administration by automatically storing the data. Further parallel versions of the Virtual City Task will have to be created to overcome known practice effects in memory testing. Finally, the full-computerized battery will have to be compared longitudinally to the ADAS-Cog, the gold standard in clinical routine. In the long term, the longitudinal administration of the DZNE-Cog to participants who underwent an Amyloid PET or accepted brain donation could bring insight into both healthy aging and degenerative trajectories.

In the field of Primary Progressive Aphasia, two major problems have been raised in this work: the lack of adapted neuropsychological tools in the grammar domain as well as the poor characterization of patients with neither the non-fluent nor the semantic variant of Primary Progressive Aphasia. Further work will have to be done to develop grammar tests that have good ecological validity and adapt them in other languages than English. As a first step, the newly developed tests (MAST and SECT) are currently administered in a German adaptation.

Further investigation is needed to better characterize patients with a mixed profile, which is, to date, not included in the diagnostic criteria. The publication of guidelines for the neuropsychological testing of PPA patients might help to precise the description of the different clinical profiles. Indeed, both in research and clinics, some aspects of language are often left untested and are therefore wrongly considered as preserved. There is a strong need of both improving specific tools for PPA diagnosis but also starting to enlarge the tested domains. Finally, disease prediction based on the clinical profile being a major hurdle in the field of PPA, future studies should as much as possible include disease confirmation by means of amyloid PET imaging or post-mortem brain biopsy.

References

Adlam, A.-L., Patterson, K., Bozeat, S., & Hodges, J. R. (2010). The Cambridge Semantic Memory Test Battery: Detection of semantic deficits in semantic dementia and Alzheimer's disease. *Neurocase, 16*(3), 193-207. doi:10.1080/13554790903405693.

Ahmed, R. M., Irish, M., Kam, J., van Keizerswaard, J., Bartley, L., Samaras, K., . . . Piguet, O. (2014). Quantifying the eating abnormalities in frontotemporal dementia. *JAMA Neurology, 71*(12), 1540-1546. doi:10.1001/jamaneurol.2014.1931.

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., . . . Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 270-279. doi:10.1016/j.jalz.2011.03.008.

Allison, S. L., Fagan, A. M., Morris, J. C., & Head, D. (2016). Spatial navigation in preclinical Alzheimer's disease. *Journal of Alzheimer's Disease, 52*(1), 77-90. doi:10.3233/JAD-150855.

Almkvist, O., & Bäckman, L. (1993). Progression in Alzheimer´s disease : Sequencing of neuropsychological decline. *International Journal of Geriatric Psychiatry, 8*(9), 755-763. doi:10.1002/gps.930080908

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Aschenbrenner, S., Tucha, O., & Lange, K. W. (2000). *Regensburger Wortflüssigkeits-Test* (RWT) [Assessment instrument]. Göttingen: Hogrefe Verlag.

Ash, S., Moore, P., Vesely, L., Gunawardena, D., McMillan, C., Anderson, C., . . . Grossman, M. (2009). Non-Fluent speech in frontotemporal lobar degeneration. *Journal of Neurolinguistics, 22*(4), 370-383. doi:10.1016/j.jneuroling.2008.12.001.

Auclair-Ouellet, N. (2015). Inflectional morphology in primary progressive aphasia and Alzheimer's disease: A systematic review. *Journal of neurolinguistics, 34*, 41-64. doi:10.1016/j.jneuroling.2014.12.002.

Auclair-Ouellet, N., Macoir, J., Laforce, R., Jr., Bier, N., & Fossard, M. (2016). Regularity and beyond: Impaired production and comprehension of inflectional morphology in semantic dementia. *Brain and language, 155-156*, 1-11. doi:10.1016/j.bandl.2016.02.002.

Balthazar, M. L .F., Martinelli, J. E., Cendes, F., & Damasceno, B. P. (2007). Lexical semantic memory in amnestic mild cognitive impairment and mild Alzheimer´s disease. *Arquivos de Neuro-Psiquiatria, 65*(3A), 619-622. doi:10.1590/S0004-282X2007000400014.

Bates, D.M., Maechler, M., & Bolker, B. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Beglinger, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J., . . . Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology, 20*(4), 517-529. doi:10.1016/j.acn.2004.12.003.

Benedict, R. H. B., & Zgaljardic, D. J. (1998). Practice effects during repeated administrations of memory tests with and without alternate forms. *Journal of Clinical and Experimental Neuropsychology, 20*(3), 339-352. doi:10.1076/jcen.20.3.339.822.

Benge, J. F., Balsis, S., Geraci, L., Massman, P. J., & Doody, R. S. (2009). How well do the ADAS-cog and its subscales measure cognitive dysfunction in Alzheimer's disease? *Dementia and Geriatric Cognitive Disorders, 28*(1), 63-69. doi:10.1159/000230709.

Bird, C. M., Chan, D., Hartley, T., Pijnenburg, Y. A., Rossor, M. N., & Burgess, N. (2010). Topographical short-term memory differentiates Alzheimer's disease from frontotemporal lobar degeneration. *Hippocampus, 20*(10), 1154-1169. doi:10.1002/hipo.20715.

Bishop, D. (1989). Test for Reception of Grammar (2nd.ed] [Assessment instrument]. Manchester: University of Manchester.

Bishop, D. (2003). Test for Reception of Grammar (Version 2). London: The PsychologicalCorporation.

Blackwell, A. D., Sahakian, B. J., Vesey, R., Semple, J. M., Robbins, T. W., & Hodges, J. R. (2004). Detecting dementia: Novel neuropsychological markers of preclinical Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 17*(1-2), 42-48. doi:10.1159/000074081.

Botha, H., Duffy, J. R., Whitwell, J. L., Strand, E. A., Machulda, M. M., Schwarz, C. G., . . . Josephs, K. A. (2015). Classification and clinicoradiologic features of primary progressive aphasia (PPA) and apraxia of speech. *Cortex, 69,* 220-236. doi:10.1016/j.cortex.2015.05.013.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P., & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia, 38*(9), 1207-1215. doi:10.1016/S0028-3932(00)00034-8.

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica, 82*(4), 239-259.

Caffarra, P., Gardini, S., Cappa, S., Dieci, F., Concari, L., Barocco, F., . . . Prati, G. D. (2013). Degenerative jargon aphasia: Unusual progression of logopenic/phonological progressive aphasia? *Behavioural Neurology, 26*(1-2), 89-93. doi:10.3233/BEN-2012-110218.

Carr, E. K., & Lincoln, N. B. (1988). Inter-rater reliability of the Rey figure copying test. *British Journal of Clinical Psychology, 27*(3), 267-268. doi:10.1111/j.2044-8260.1988.tb00786.x.

Caso, F., Mandelli, M. L., Henry, M., Gesierich, B., Bettcher, B .M., Ogar, J., . . . Gorno-Tempini, M. L. (2014). In vivo signatures of nonfluent/agrammatic primary progressive aphasia caused by FTLD pathology. *Neurology, 82*(3). doi:10.1212/WNL.0000000000000031.

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., & Rogalski, E. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. Brain, 136, 2619–2628. doi:10.1093/brain/awt163.

Cecato, J. F., Martinelli, J. E., Izbicki, R., Yassuda, M. S., & Aprahamian, I. (2016). A subtest analysis of the Montreal cognitive assessment (MoCA): Which subtests can best discriminate between healthy controls, mild cognitive impairment and Alzheimer's disease? *International Psychogeriatrics, 28*(5), 825-832. doi:10.1017/S1041610215001982.

Chamberlain, S. R., Blackwell, A. D., Nathan, P. J., Hammond, G., Robbins, T. W., Hodges, J. R., . . . Sahakian, B. J. (2011). Differential cognitive deterioration in dementia: A two year longitudinal study. *Journal of Alzheimer's disease : JAD, 24*(1), 125-136. doi:10.3233/JAD-2010-100450.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). Aphasiology, 26, 1250– 1277. doi:10.1080/02687038.2012.693584.

Colombo, L., Fonti, C., & Stracciari, A. (2009). Italian verb inflection in Alzheimer dementia. *Neuropsychologia, 47*(4), 1069-1078. doi:10.1016/j.neuropsychologia.2008.12.035.

Connor, L. T., Spiro, A., Obler, L. K., & Albert, M. L. (2004). Change in object naming ability during adulthood. The *Journal of Gerontology, 59B*(5), 203-209. doi:10.1093/geronb/59.5.P203.

Cortese, M. J., Balota, D. A., Sergent-Marshall, S. D., Buckner, R. L., & Gold, B. T. (2006). Consistency and regularity in past-tense verb generation in healthy ageing, Alzheimer's disease, and semantic dementia. *Cognitive Neuropsychology, 23*(6), 856-876. doi:10.1080/02643290500483124.

Crawford, J. R., Stewart, L. E., & Moore, J. W. (1989). Demonstration of savings on the AVLT and development of parallel form. *Journal of Clinical and Experimental Neuropsychology, 11*(6), 975-981. doi:10.1080/01688638908400950

Cupit, J., Leonard, C., Graham, N. L., Seixas Lima, B., Tang-Wai, D., Black, S. E., & Rochon, E. (2016). Analysing syntactic productions in semantic variant PPA and non-fluent variant PPA: How different are they? *Aphasiology, 31*(3) 1-26. doi:10.1080/02687038.2016.1180661.

Darby, D., Maruff, P., Collie, A., & McStephen, M. (2002). Mild cognitive impairement can be detected by multiple assessments in a single day. *Neurology, 59*(7), 1042-1046. doi:10.1212/WNL.59.7.1042

Davies, R. R., Hodges, J. R., Kril, J. J., Patterson, K., Halliday, G. M., & Xuereb, J. H. (2005). The pathological basis of semantic dementia. *Brain, 128*(Pt 9), 1984-1995. doi:10.1093/brain/awh582.

De Bleser, R., Denes, G.F., Luzzatti, C., Mazzucchi, A., Poeck, K., Spinnler, H., & Willmes, K. (1986). L´Aachner Aphasie Test (AAT) I. Problemi e soluzioni per una versione italiana del test e per uno studio crosslinguistico dei disturbi afasici. *Archivio di Psicologia, neurologia e Psichiatria, 47*(2), 209-237.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test: Adult Version* [Assessment instrument]. Manual. San Antonio, TX: Psychological Corporation.

Deloche, G., Hannequin, D., Dordain, M., Perrier, D., Pichard, B., Quint, S., . . . Cardebat, D. (1996). Picture confrontation oral naming: Performance differences between aphasics and normals. *Brain and Language, 53*(1), 105-120. doi:10.1006/brln.1996.0039.

Deloche, G. & Hannequin, D. (1997). DO 80 : Test de dénomination orale d'images [Assessment instrument]. Paris : ECPA.

Deramecourt, V., Lebert, F., Debachy, B., Mackowiak-Cordoliani, M. A., Bombois, S., Kerdraon, O., . . . Pasquier, F. (2010). Prediction of pathology in primary progressive language and speech disorders. *Neurology, 74*(1), 42-49. doi:10.1212/WNL.0b013e3181c7198e.

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., . . . Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *Lancet Neurology, 6*(8), 734-746. doi:10.1016/S1474-4422(07)70178-3.

El Hachioui, H., Lingsma, H. F., van de Sandt-Koenderman, W. M. E., Dippel, D. W. J., Koudstaal, P. J., & Visch-Brink, E. G. (2013). Recovery of aphasia after stroke: A 1-year follow-up study. *Journal of neurology, 260*(1), 166-171. doi:10.1007/s00415-012-6607-2.

Eurocode (2006).Prevalence of dementia in Europe. Retrieved from https://ec.europa.eu/health/archive/ph_information/dissemination/diseases/docs/eurocode.pdf

European Medicines Agency, (2016). Draft guideline on the clinical investigation of medicines for the treatment of Alzheimer´s Disease and other dementias. Retrieved from www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2016/02/WC500200830.pdf

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*(3), 189-198.

Forlenza, O. V., Radanovic, M., Talib, L. L., Aprahamian, I., Diniz, B. S., Zetterberg, H., & Gattaz, W. F. (2015). Cerebrospinal fluid biomarkers in Alzheimer's disease: Diagnostic accuracy and prediction of dementia. *Alzheimer's & Dementia, 1*(4), 455-463. doi:10.1016/j.dadm.2015.09.003.

Foxe, D. G., Irish, M., Hodges, J. R., & Piguet, O. (2013). Verbal and visuospatial span in logopenic progressive aphasia and Alzheimer's disease. *Journal of the International Neuropsychological Society, 19*(3), 247-253. doi:10.1017/S1355617712001269.

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. doi:10.1016/j.cortex.2012.12.006.

Gianaros, P. J., Muth, E. R., Mordkoff, J. T., Levine, M. E., & Stern, R. M. (2001). A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space and Environmental Medicine, 72*(2), 115-119.

Giannini, L. A. A., Irwin, D. J., McMillan, C. T., Ash, S., Rascovsky, K., Wolk, D. A., . . . Grossman, M. (2017). Clinical marker for Alzheimer disease pathology in logopenic primary progressive aphasia. *Neurology, 88*(24), 2276-2284. doi:10.1212/WNL.0000000000004034.

Goldman, J. S., Farmer, J. M., Wood, E. M., Johnson, J. K., Boxer, A., Neuhaus, J., . . . Miller, B. L. (2005). Comparison of family histories in FTLD subtypes and related tauopathies. *Neurology, 65*(11), 1817-1819. doi:10.1212/01.wnl.0000187068.92184.63.

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

Gorno-Tempini, M. L., Dronkers, N. F., Rankin, K. P., Ogar, J. M., Phengrasamy, L., Rosen, H. J., . . . Miller, B. L. (2004). Cognition and anatomy in three variants of primary progressive aphasia. *Annals of Neurology, 55*(3), 335-346. doi:10.1002/ana.10825.

Gorno-Tempini, M. L., Hillis, A. E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S. F., . . . Grossman, M. (2011). Classification of primary progressive aphasia and its variants. *Neurology, 76*(11), 1006-1014. doi:10.1212/WNL.0b013e31821103e6.

Gorno-Tempini, M. L., Ogar, J. M., Brambati, S. M., Wang, P., Jeong, J. H., Rankin, K. P., . . . Miller, B. L. (2006). Anatomical correlates of early mutism in progressive nonfluent aphasia. *Neurology, 67*(10), 1849-1851. doi:10.1212/01.wnl.0000237038.55627.5b.

Grossman, M. (2012). The non-fluent/agrammatic variant of primary progressive aphasia. *The Lancet Neurology, 11*(6), 545-555. doi:10.1016/s1474-4422(12)70099-6.

Guariglia, C. C., & Nitrini, R. (2009). Topographical disorientation in Alzheimer´s disease. *Arquivos de Neuro-psiquiatria, 67*(4), 967-972. doi:10.1590/S0004-282X2009000600001.

Hall, J. R., Vo, H. T., Johnson, L. A., Wiechmann, A., & O'Bryant, S. E. (2012). Boston Naming Test: Gender differences in older adults with and without Alzheimer's Dementia. *Psychology, 3*(6), 485-488. doi:10.4236/psych.2012.36068.

Harris, J. M., Gall, C., Thompson, J. C., Richardson, A. M., Neary, D., du Plessis, D., . . . Jones, M. (2013). Classification and pathology of primary progressive aphasia. *Neurology, 81*(21), 1832-1839. doi:10.1212/01.wnl.0000436070.28137.7b.

Harris, J. M., Jones, M., Gall, C., Richardson, A. M., Neary, D., du Plessis, D., . . . Thompson, J. C. (2016). Co-occurrence of language and behavioural change in frontotemporal lobar degeneration. *Dementia and Geriatric Cognitive Disorders extra, 6*(2), 205-213. doi:10.1159/000444848.

Heister, J., Würzner, K. -M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB-eine lexikalische Datenbank für die psychologische und linguistiche Forschung. Psychologische Rundschau, 62, 10-20. Hogrefe Verlag. doi:10.1026/0033-3042/a000029.

Hensel, A., Angermeyer, M. C., & Riedel-Heller, S. G. (2007). Measuring cognitive change in older adults: Reliable change indices for the Mini-Mental State Examination. *Journal of Neurology, Neurosurgery, and Psychiatry, 78*(12), 1298-1303. doi:10.1136/jnnp.2006.109074.

Hodges, J. R., Martinos, M., Woollams, A. M., Patterson, K., & Adlam, A. L. (2008). Repeat and Point: Differentiating semantic dementia from progressive non-fluent aphasia. *Cortex, 44*(9), 1265-1270. doi:10.1016/j.cortex.2007.08.018.

Hodges, J. R., Mitchell, J., Dawson, K., Spillantini, M. G., Xuereb, J. H., McMonagle, P., . . . Patterson, K. (2010). Semantic dementia: Demography, familial factors and survival in a consecutive series of 100 cases. *Brain, 133*(Pt 1), 300-306. doi:10.1093/brain/awp248.

Hodges, J. R., Patterson, K., Oxbury, S. , & Funnell, E. (1992). Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. *Brain, 115*(Pt 6), 1783-1806. doi:10.1093/brain/115.6.1783.

Holmes, C., Boche, D., Wilkinson, D., Yadegarfar, G., Hopkins, V., Bayer, A., . . . Nicoll, J. A. (2008). Long-term effects of Abeta42 immunisation in Alzheimer's disease: Follow-up of a randomised, placebo-controlled phase I trial. *Lancet, 372*(9634), 216-223. doi:10.1016/s0140-6736(08)61075-2.

Hope, T., Keene, J., McShane, R. H., Fairburn, C. G., Gedling, K., & Jacoby, R. (2001). Wandering in Dementia: A Longitudinal Study. *International Psychogeriatrics, 13*(2), 137-147. doi:10.1017/S1041610201007542.

Howard, D., & Patterson, K. E. (1992). The pyramid and palm trees test : a test of semantic access from words and pictures [Assessment instrument]. Bury St Edmunds: Thames Valley Test Company.

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). Aachener Aphasie Test [Assessment instrument]. Göttingen: Hogrefe

Huff, F. J., Corkin, S., & Growdown, J. H. (1986). Semantic impairment and anomia in Alzheimer's disease. *Brain and language, 28*(2), 235-249. doi: 10.1016/0093-934X(86)90103-3.

Hutchings, R., Hodges, J. R., Piguet, O., Kumfor, F., & Boutoleau-Bretonnière, C. (2015). Why should I Care? Dimensions of socio-emotional cognition in younger-onset dementia. *Journal of Alzheimer's disease, 48*(1), 135-147. doi: 10.3233/JAD-150245.

Hutchinson, A. D., & Mathias, J. L. (2007). Neuropsychological deficits in frontotemporal dementia and Alzheimer's disease: A meta-analytic review. *Journal of Neurology, Neurosurgery, and Psychiatry, 78*(9), 917-928. doi: 10.1136/jnnp.2006.100669.

Irish, M., Kumfor, F., Hodges, J. R., & Piguet, O. (2013). A tale of two hemispheres: Contrasting socioemotional dysfunction in right-versus left-lateralised semantic dementia. *Dementia & Neuropsychologia, 7*(1), 88-95. doi:10.1590/S1980-57642013DN70100014.

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L.. M., Aisen, P. S., Weiner, M. W., . . . Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology, 9*(1), 119-128. doi:10.1016/s1474-4422(09)70299-6.

Jaeger, J. J., Lockwood, A. H., Kemmerer, D. L., Van Valin, R. D., Murphy, B. W., & Khalak, H. G. (1996). A positron emission tomographic study of regular and irregular verb morphology in English. *Language, 72*(3), 451-497. doi:10.2307/416276.

Jahn, D. R., Mauer, C. B., Menon, C. V., Edwards, M. L., Dressel, J. A., & O'Bryant, S. E. (2013). A brief Spanish-English equivalent version of the Boston Naming Test: A project FRONTIER study. *Journal of Clinical and Experimental Neuropsychology, 35*(8), 835-845. doi:10.1080/13803395.2013.825234.

Jessen, F., Amariglio, R. E., van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., . . . Subjective Cognitive Decline Initiative Working, Group (SCD-I). (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia, 10*(6), 844-852. doi:10.1016/j.jalz.2014.01.001.

Jonker, C., Geerlings, M. I., & Schmand, B. (2000). Are memory complaints predictive for dementia? A review of clinical and population based studies. *International Journal of Geriatric Psychiatry, 15(11),* 983-991. doi:10.1177/2050312115574796.

Josephs, K. A., Duffy, J. R., Strand, E. A., Whitwell, J. L., Layton, K. F., Parisi, J. E., . . . Petersen, R. C. (2006). Clinicopathological and imaging correlates of progressive aphasia and apraxia of speech. *Brain, 129*(Pt 6), 1385-1398. doi:10.1093/brain/awl078.

Kamminga, J., Kumfor, F., Burrell, J. R., Piguet, O., Hodges, J. R., & Irish, M. (2015). Differentiating between right-lateralised semantic dementia and behavioural-variant frontotemporal dementia: An examination of clinical characteristics and emotion processing. *Journal of Neurology, Neurosurgery, and Psychiatry, 86*(10), 1082-1088. doi:10.1136/jnnp-2014-309120.

Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test* [Assessment instrument] Philadelphia: Lea & Febiger.

Karin, A., Hannesdottir, K., Jaeger, J., Annas, P., Segerdahl, M., Karlsson, P., . . . Miller, F. (2014). Psychometric evaluation of ADAS-Cog and NTB for measuring drug response. *Acta Neurologica Scandinavica, 129*(2), 114-122. doi:10.1111/ane.12153.

Katsumata, Y., Mathews, M., Abner, E. L., Jicha, G. A., Caban-Holt, A., Smith, C. D., . . . Fardo, D. W. (2015). Assessing the discriminant ability, reliability, and comparability of multiple short forms of the Boston Naming Test in an Alzheimer's disease center cohort. *Dementia and Geriatric Cognitive Disorders, 39*(3-4), 215-227. doi:10.1159/000370108.

Kertesz, A. (1982). The Western aphasia battery. New York :Grune & Stratton.

Knibb, J. A., Woollams, A. M., Hodges, J. R., & Patterson, K. (2009). Making sense of progressive non-fluent aphasia: An analysis of conversational speech. *Brain, 132*(Pt 10), 2734-2746. doi:10.1093/brain/awp207.

Knibb, J. A., Xuereb, J. H., Patterson, K., & Hodges, J. R. (2006). Clinical and pathological characterization of progressive aphasia. *Annals of Neurology, 59*(1), 156-165. doi:10.1002/ana.20700.

Lansing, A. E., Ivnik, R. J., Munro Cullum, C., & Randolph, C. (1999). An empirically derived short form of the Boston Naming Test. *Archives of Clinical Neuropsychology, 14*(6), 481-487. doi:10.1016/S0887-6177(98)00022-5.

Lebedeva, E., Huang, M., & Koski, L. (2016). Comparison of alternate and original items on the Montreal Cognitive Assessment. *Canadian Geriatrics Journal, 19*(1), 15-18. doi:10.5770/cgj.19.216.

Lee, H., Lee, Y., Choi, H., & Pyun, S. B. (2015). Community integration and quality of life in aphasia after stroke. *Yonsei Medical Journal, 56*(6), 1694-1702. doi:10.3349/ymj.2015.56.6.1694.

Leyton, C. E., Villemagne, V. L., Savage, S., Pike, K. E., Ballard, K. J., Piguet, O., . . . Hodges, J. R. (2011). Subtypes of progressive aphasia: Application of the international consensus criteria and validation using beta-amyloid imaging. *Brain, 134*(Pt 10), 3030-3043. doi:10.1093/brain/awr216.

Liberman, J., Stewart, W., Seines, O. & Gordon, B. (1994). Rater agreement for the Rey-Osterrieth Complex Figure Test. *Journal of Clinical Psychology, 50*(4), 615-624. doi:10.1002/1097-4679(199407)50:43.0.CO;2-R.

Lin, C. Y., Chen, T. B., Lin, K. N., Yeh, Y. C., Chen, W. T., Wang, K. S., & Wang, P. N. (2014). Confrontation naming errors in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 37*(1-2), 86-94. doi:10.1159/000354359.

Lowe, D. A., Balsis, S., Benge, J. F., & Doody, R. S. (2015). Adding delayed recall to the ADAS-cog improves measurement precision in mild Alzheimer's disease: Implications for predicting instrumental activities of daily living. *Psychological Assessment, 27*(4), 1234-1240. doi:10.1037/pas0000133.

Magen, H. S. (2014). A behavioral study of regularity, irregularity and rules in the English past tense. *Journal of Psycholinguistic Research, 43*(6), 791-814. doi: 10.1007/s10936-013-9276-0.

Magis, D., & Facon, B. (2013). Item purification does not always improve DIF detection: A counterexample with Angoff's Delta Plot. *Educational and Psychological Measurement, 73*(2), 293-311. doi:10.1177/0013164412451903.

Magis, D., & Facon, B. (2014). deltaPlotR: An R package for differential item functioning analysis with Angoff´s delta plot. *Journal of Statistical Software, 59*(1), 1-19.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D., & O'Keefe, J. (1998). Knowing where and getting there: A human navigation network. *Science, 28*(5365), 921-924. doi:10.1126/science.280.5365.921.

Mair, P., Hatzinger, R., & Maier, M. J. (2015).eRm : Extended Rasch Modeling. 0.15-6. http://erm.r-forge.r-project.org/

Manton, K. C., Gu, X. L., & Ukraintseva, S. V. (2005). Declining prevalence of dementia in the U.S. elderly population,. *Advances in Gerontology, 16*, 30-37.

Matthews, F. E., Arthur, A., Barnes, L. E., Bond, J., Jagger, C., Robinson, L., & Brayne, C. (2013). A two-decade comparison of prevalence of dementia in individuals aged 65 years and older from three geographical areas of England: Results of the Cognitive Function and Ageing Study I and II. *Lancet, 382*(9902), 1405-1412. doi:10.1016/s0140-6736(13)61570-6.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Jr., Kawas, C. H., . . . Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia, 7*(3), 263-269. doi:10.1016/j.jalz.2011.03.005.

Merten, T. (2004). Development of a German short version of the Boston Naming Test. *Neurologie und Rehabilitation, 10*(6), 305-311.

Mesulam, M. M. (1982). Slowly progressive aphasia without generalized dementia. *Annals of Neurology, 11*(6), 592-598. doi:10.1002/ana.410110607.

Mesulam, M. M., Wicklund, A., Johnson, N., Rogalski, E., Léger, G. C., Rademaker, A., . . . Bigio, E. H. (2008). Alzheimer and frontotemporal pathology in subsets of primary progressive aphasia. *Annals of Neurology, 63*(6), 709-719. doi:10.1002/ana.21388.

Mesulam, M. M., Wieneke, C., Rogalski, E., Cobia, D., Thompson, C., & Weintraub, S. (2009). Quantitative template for subtyping primary progressive aphasia. *Archives of Neurology, 66*(12), 1545-1551. doi:10.1001/archneurol.2009.288.

Mesulam, M. M., Wieneke, C., Thompson, C., Rogalski, E., & Weintraub, S. (2012). Quantitative classification of primary progressive aphasia at early and mild impairment stages. *Brain, 135*(Pt 5), 1537-1553. doi:10.1093/brain/aws080.

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. doi:10.1016/j.bandl.2009.03.007.

Meteyard, L., Quain, E., & Patterson, K. (2014). Ever decreasing circles: Speech production in semantic dementia. *Cortex, 55*(1), 17-29. doi: 10.1016/j.cortex.2013.02.013.

Miller, N., Willmes, K., & De Bleser, R. (2010). The psychometric properties of the English language version of the Aachen Aphasia Test (EAAT). *Aphasiology, 14*(7), 683-722. doi:10.1080/026870300410946.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke´s Cognitive Examination Revised (ACE-R) : A brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry, 21*(11), 1076-1085. doi:10.1002/gps.1610.

Mitruschina, M., & Satz, P. (1991). Reliability and validity of the Mini-Mental State Exam in neurologically intact elderly. *Journal of Clinical Psychology, 47*(4), 537-543. doi:10.1002/1097-4679(199107)47:4<537::AID-JCLP2270470411>3.0.CO;2-9.

Mohs, R. C., Knopman, D. S., Petersen, R. C., Ferris, S. H., Ernesto, C., Grundman, M., . . . & Thal, L. J. (1997). Development of cognitive instruments ofr use in clinical trials of antidementia drugs : Additions to the Alzheimer´s Disease Assessment Scale that broaden its scope. The Alzheimer´s disease cooperative study. *Alzheimer Disease and Associated Disorders, 11*(S2), 13-21. doi:10.1097/00002093-199700112-00003.

Moms, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., . . . Clark, C. (1989). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assesment of Alzheimer's disease. *Neurology, 39*(9), 1159-1159. doi:10.1212/wnl.39.9.1159.

Morgan, D., Diamond, D. M., Gottschall, P. E., Ugen, K. E., Dickey, C., Hardy, J., . . . Arendash, G. W. (2000). A beta peptide vaccination prevents memory loss in an animal model of Alzheimer´s disease. *Nature, 408*(6815), 982-985. doi:10.1038/35050116.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699. doi:10.1111/j.1532-5415.2005.53221.x.

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., . . .Benson, J. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology, 51*(6), 1546-1554. doi: https://doi.org/10.1212/WNL.51.6.1546.

Osterrieth, P. A. (1944). Le test de copie d'une figure complexe: Contribution à l'étude de la perception et de la mémoire [The test of copying a complex figure: A contribution to the study of perception and memory]. *Archives de Psychologie, 30*, 286-356.

Pai, M. C., & Jacobs, W. J. (2004). Topographical disorientation in community-residing patients with Alzheimer's disease. *International Journal of Geriatric Psychiatry, 19*(3), 250-255. doi:10.1002/gps.1081.

Panda 3D (Game engine). (2010). Retrieved from https://www.panda3d.org/

Patterson, K., & Hodges, J. R. (1992). Deterioration of word meaning: Implications for reading. *Neuropsychologia, 30*(12), 1025-1040. doi:10.1016/0028-3932(92)90096-5.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia, 39*(7), 709-724. doi:10.1016/S0028-3932(01)00008-2.

Peirce, J. W. (2007). PsychoPy--Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. doi: 10.1016/j.jneumeth.2006.11.017.

Pengas, G., Patterson, K., Arnold, R. J., Bird, C. M., Burgess, N., & Nestor, P. J. (2010). Lost and found: Bespoke memory testing for Alzheimer's disease and semantic dementia. *Journal of Alzheimer´s Disease, 21*(4). 1347-1365. doi:10.3233/jad-2010-100654.

Pengas, G., Williams, G. B., Acosta-Cabronero, J., Ash, T. W., Hong, Y. T., Izquierdo-Garcia, D., . . . Nestor, P. J. (2012). The relationship of topographical memory performance to regional neurodegeneration in Alzheimer's disease. *Frontiers in Aging Neuroscience, 4*(17), 1-10. doi:10.3389/fnagi.2012.00017.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment : Clinical characterization and outcome. *Archives of Neurology, 56*(3), 303-308. doi:10.1001/archneurol.2009.266.

Pinker, S. (1991). Rules of language. *Science, 253*(2019), 530-535. doi:10.1126/science.1857983.

Posner, H. B., Cano, S., Carrillo, M. C., Selnes, O., Stern, Y., Thomas, R. G., . . . Alzheimer's Disease Neuroimaging, Initiative. (2013). Establishing the psychometric underpinning of cognition measures for clinical trials of Alzheimer's disease and its precursors: A new approach. *Alzheimer's & Dementia, 9*(1 Suppl), 56-60. doi:10.1016/j.jalz.2012.10.008.

Pracharitpukdee, N., Phanthumchinda, K., Huber, W., & Willmes, K. (1998). The Thai version of the German Aachen Aphasia Test (AAT) : Description of the test and performance in normal subjects. *Journal of the Medical association of Thailand, 81*(6), 402-412.

Price, J. L., Ko, A. I., Wade, M. J., Tsou, S. K., McKeel, D. W., & Morris, J. C. (2001). Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. *Archives of Neurology, 58*(9), 1395-1402. doi:10.1001/archneur.58.9.1395.

Price, J. L., & Morris, J. C. (1999). Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Annals of Neurology, 45*(3), 358-368. doi:10.1002/1531-8249(199903)45:3<358::AID-ANA12>3.0.CO;2-X.

Puente, A.E., & Puente, A.N. (2009). The challenges of measuring abilities and competencies in Hispanics Latinos. In E. L. Grigorenko (Ed.), *Multicultural psychoeducational assessment* (pp. 417-441). New York: Springer Publishing Company.

R Core Team (2013). R: A language and environment for statistical computing [Computer Software]. Vienna, Austria : R Foundation for Statistical Computing. URL http://www.R-project.org/.

Rabinovici, G. D., Jagust, W. J., Furst, A. J., Ogar, J. M., Racine, C. A., Mormino, E. C., . . . Gorno-Tempini, M. L. (2008). Abeta amyloid and glucose metabolism in three variants of primary progressive aphasia. *Annals of Neurology, 64*(4), 388-401. doi:10.1002/ana.21451.

Rafii, M. S., & Aisen, P. S. (2015). Advances in Alzheimer's disease drug development. *BMC Medicine, 13*(62), 1-7. doi: 10.1186/s12916-015-0297-4.

Rentz, D. M., Mormino, E. C., Papp, K. V., Betensky, R. A., Sperling, R. A., & Johnson, K. A. (2017). Cognitive resilience in clinical and preclinical Alzheimer's disease: The association of Amyloid and Tau burden on cognitive performance. *Brain Imaging and Behavior, 11*(2), 383-390. doi:10.1007/s11682-016-9640-4.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie, 28*, 215-285.

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and Rationale. *Clinical Aphasiology, 24*, 121-133.

Rockwood, K., Fay, S., & Gorman, M. (2010). The ADAS-cog and clinically meaningful change in the VISTA clinical trial of galantamine for Alzheimer's disease. *International Journal of Geriatric Psychiatry, 25*(2), 191-201. doi:10.1002/gps.2319.

Rockwood, K., Fay, S., Gorman, M., Carver, D., & Graham, J. E. (2007). The clinical meaningfulness of ADAS-Cog changes in Alzheimer's disease patients treated with donepezil in an open-label trial. *BMC neurology, 7*(26), 1-8. doi:10.1186/1471-2377-7-26.

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. *The Journal of Neuroscience, 31*(9), 3344-50. doi:10.1523/JNEUROSCI.5544-10.2011.

Rohrer, J. D., Rossor, M. N., & Warren, J. D. (2009). Neologistic jargon aphasia and agraphia in primary progressive aphasia. *Journal of the Neurological Sciences, 277*(1-2), 155-159. doi:10.1016/j.jns.2008.10.014.

Rohrer, J. D., Sauter, D., Scott, S., Rossor, M. N., & Warren, J. D. (2012). Receptive prosody in nonfluent primary progressive aphasias. Cortex, 48, 308–316. doi:10.1016/j.cortex.2010.09.004.

Rosen, H. J., Allison, S. C., Ogar, J. M., Amici, S., Rose, K., Dronkers, N., . . . Gorno-Tempini, M. L. (2006). Behavioral features in semantic dementia vs other forms of progressive aphasias. *Neurology, 67*(10), 1752-1756. doi:10.1212/01.wnl.0000247630.29222.34.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry, 141*(11), 1356-1364. doi:10.1176/ajp.141.11.1356.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), Mechanisms of language aquisition (pp. 195-248). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Ryu, S. Y., Lee, S. B., Kim, T. W., & Lee, T. J. (2016). Memory complaints in subjective cognitive impairment, amnestic mild cognitive impairment and mild Alzheimer's disease. *Acta Neurologica Belgica, 116*(4), 535-541. doi:10.1007/s13760-016-0604-7.

Sajjadi, S. A., Patterson, K., Arnold, R. J., Watson, P. C, & Nestor, P. J. (2012). Primary progressive aphasia: A tale of two syndromes and the rest. *Neurology, 78*(21),1670-1677. doi:10.1212/WNL.0b013e3182574f79

Sajjadi, S. A., Patterson, K., & Nestor, P. J. (2014). Logopenic, mixed, or Alzheimer-related aphasia? *Neurology, 82*(13), 1127-1131. doi:10.1212/WNL.0000000000000271.

Sajjadi, S. A., Patterson, K., Tomek, M., & Nestor, P. J. (2012a). Abnormalities of connected speech in semantic dementia vs Alzheimer's disease. *Aphasiology, 26*(6), 847-866. doi: 10.1080/02687038.2012.654933.

Sajjadi, S. A., Patterson, K., Tomek, M., & Nestor, P. J. (2012b). Abnormalities of connected speech in the non-semantic variants of primary progressive aphasia. *Aphasiology, 26*(10), 1219-1237. doi:10.1080/02687038.2012.710318.

Salmon, D. P., & Bondi, M. W. (2009). Neuropsychological assessment of dementia. *Annual Review of Psychology, 60*, 257-282. doi: 10.1146/annurev.psych.57.102904.190024.

Salthouse, T. A., & Tucker-Drob, E. M. (2008). Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology, 22*(6), 800-811. doi:10.1037/a0013091.

Seelaar, H., Kamphorst, W., Rosso, S. M., Azmani, A., Masdjedi, R., de Koning, I., . . . van Swieten, J. C. (2008). Distinct genetic forms of frontotemporal dementia. *Neurology, 71*(16), 1220-1226. doi:10.1212/01.wnl.0000319702.37497.72.

Seidenberg, M. S., & Hoeffner, J. H. (1998). Evaluating behavioral and neuroimaging data on past tense processing. *Language, 74*(1), 104-122. doi:10.2307/417567.

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist, 5*(1-2), 165-173. doi:10.1300/J018v05n01_09.

Silagi, M. L., Feirreira Bertolucci, P. H., & Ortiz, K. Z. (2015). Naming ability in patients with mild to moderate Alzheimer's disease: What changes occur with the evolution of the disease? *Clinics, 70*(6), 423-428. doi: 10.6061/clinics/2015(06)07.

Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioural neurology, 2*(3), 167-182.

Snowden, J. S., Neary, D., Mann, M. A., Goulding, P. J., & Testa, H. J. (1992). Progressive language disorder due to lobar atrophy. *Annals of Neurology, 31*(2), 174-183. doi:10.1002/ana.410310208.

Spencer, A. (1994). Morphological theory and English. *Links & Letters, 1*, 71-84.

Spinelli, E. G., Mandelli, M. L., Miller, Z. A., Santos-Santos, M. A., Wilson, S. M., Agosta, F., . . . Gorno-Tempini, M. L. (2017). Typical and atypical pathology in primary progressive aphasia variants. *Annals of Neurology, 81*(3), 430-443. doi:10.1002/ana.24885.

Statistiches Bundesamt (2015). Bevölkerung Deutschlands bis 2060: 13. koordinierte Bevölkerungsvorausberechnung. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/VorausberechnungBevoelkerung/BevoelkerungDeutschland2060Presse5124204159004.pdf?__blob=publicationFile

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology, 11*(11), 1006-1012. doi:10.1016/s1474-4422(12)70191-6.

Stomrud, E., Hansson, O., Blennow, K., Minthon, L., & Londos, E. (2007). Cerebrospinal fluid biomarkers predict decline in subjective cognitive function over 3 years in healthy elderly. *Dementia and Geriatric Cognitive Disorders, 24*(2), 118-124. doi:10.1159/000105017.

Thompson, C. K., Ballard, K. J., Tait, M. E., Weintraub, S., & Mesulam, M. M. (1997). Patterns of language decline in non-fluent primary progressive aphasia. Aphasiology, 11, 297–321. doi:10.1080/02687039708248473.

Thompson, C. K., Cho, S., Hsu, C. J., Wieneke, C., Rademaker, A., Weitner, B. B., . . . Weintraub, S. (2012). Dissociations between fluency and agrammatism in primary progressive aphasia. *Aphasiology, 26*(1), 20-43. doi:10.1080/02687038.2011.584691.

Thompson, C. K., Meltzer-Asscher, A., Cho, S., Lee, J., Wieneke, C., Weintraub, S., & Mesulam, M. M. (2013). Syntactic and morphosyntactic processing in stroke-induced and primary progressive aphasia. *Behavioural Neurology, 26*(1-2), 35-54. doi:10.3233/BEN-2012-110220.

Tuokko, H., Garrett, D. D., McDowell, I., Silverberg, N., & Kristjansson, B. (2003). Cognitive decline in high-functioning older adults: Reserve or ascertainment bias? *Aging & Mental Health, 7*(4), 259-270. doi:10.1080/1360786031000120750.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive neuroscience, 9*(2), 266-276. doi:10.1162/jocn.1997.9.2.266.

U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), & Center for Biologics Evaluation and Research (CBER). (2018). Early Alzheimer's Disease: Developing Drugs for Treatment Guidance for Industry. (Draft guidance). Retrieved from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596728.pdf

Vandenberghe, R. (2016). Classification of the primary progressive aphasias: Principles and review of progress since 2011. *Alzheimer's Research & Therapy, 8*(1), 16; 1-9. doi:10.1186/s13195-016-0185-y.

Villarejo-Galende, A., Llamas-Velasco, S., Gomez-Grande, A., Puertas-Martin, V., Contador, I., Sarandeses, P., . . . Herrero-San MartÍn, A. (2017). Amyloid pet in primary progressive aphasia: Case series and systematic review of the literature. *Journal of neurology, 264*(1), 121-130. doi:10.1007/s00415-016-8324-8.

Walenski, M., Sosta, K., Cappa, S., & Ullman, M. T. (2009). Deficits on irregular verbal morphology in Italian-speaking Alzheimer's disease patients. *Neuropsychologia, 47*(5), 1245-1255. doi:10.1016/j.neuropsychologia.2008.12.038.

Walter, H., Poeck, K., Weniger, D., & Willmes. K.(1983) Aachener Aphasie Test [Assessment instrument]. Göttingen: Hogrefe.

McKenna, P., & Warrington, E.K. (1983). Graded naming test: test manual. Windsor: NFER-Nelson.

Warrington, E.K., & James, M. (1991). The visual object and space perception battery. [Assement instrument]. Bury St Edmunds: UK: Thames Valley Test Company.

Wechsler, D. (1997a). *WAIS-III: Wechsler Adult Intelligence Scale – (3rd ed) administration and scoring manual (WAIS-III)*. [Assement instrument]. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997b). *Wechsler Memory Scale- (3rd ed).* San Antonio: TX : The Psychological Corporation.

Weintraub, S., Mesulam, M. M., Wieneke, C., Rademaker, A., Rogalski, E., & Thompson, C. K. (2009). The northwestern anagram test: Measuring sentence production in primary progressive aphasia. *American Journal of Alzheimer´s Disease and Other Dementias, 24*(5), 408-416. doi:10.1177/1533317509343104.

Weintraub, S., Wicklund, A. H., & Salmon, D. P. (2012). The neuropsychological profile of Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine, 2*(4), a006171. doi: 10.1101/cshperspect.a006171.

Welch, L. W., Doineau, D., Johnson, S., & King, D. (1996). Educational and gender normative data for the Boston Naming Test in a group of older adults. *Brain and Language, 53*(2), 260-266. doi:10.1006/brln.1996.0047.

Welsh, K.A., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1991). Detection of abnormal memory decline in mild cases of Alzheimer´s disease using CERAD neuropsychological measures. *Archives of Neurology, 48*(3), 278-281. doi:10.1001/archneur.1991.00530150046016.

Welsh, K.A., Butters, N., Hughes, J.P., Mohs, R.C., & Heyman, A. (1992). Detection and staging of dementia in Alzheimer´s Disease. Use of the neuropsychological measures developed for the Consortium to Establish a Registry for Alzheimer´s Disease. *Archives of Neurology, 49*(5), 448-452. doi:10.1001/archneur.1992.00530290030000.

Wicklund, M. R., Duffy, J. R., Strand, E. A., Machulda, M. M., Whitwell, J. L., & Josephs, K. A. (2014). Quantitative application of the primary progressive aphasia consensus criteria. *Neurology, 82*(13), 1119-1126. doi:10.1212/WNL.0000000000000261.

Wilson, S. M., Brambati, S. M., Henry, R. G., Handwerker, D. A., Agosta, F., Miller, B. L., . . . Gorno-Tempini, M. L. (2009). The neural basis of surface dyslexia in semantic dementia. *Brain, 132*(Pt 1), 71-86. doi: 10.1093/brain/awn300.

Wilson, S. M., Brandt, T. H., Henry, M. L., Babiak, M., Ogar, J. M., Salli, C., . . . Gorno-Tempini, M. L. (2014). Inflectional morphology in primary progressive aphasia: An elicited production study. *Brain and Language, 136*, 58-68. doi:10.1016/j.bandl.2014.07.001.

Wilson, S. M., Henry, M. L., Besbris, M., Ogar, J. M., Dronkers, N. F., Jarrold, W., . . . Gorno-Tempini, M. L. (2010). Connected speech production in three variants of primary progressive aphasia. *Brain, 133*(Pt 7), 2069-2088. doi: 10.1093/brain/awq129.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14*(1), 1-33. doi:10.1017/S0142716400010122.

Woollams, A. M., Ralph, M. A., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review, 114*(2), 316-339. doi:10.1037/0033-295X.114.2.316.

Yau, W.-Y. W., Tudorascu, D. L., McDade, E. M., Ikonomovic, S., James, J. A., Minhas, D., . . . Klunk, W. E. (2015). Longitudinal assessment of neuroimaging and clinical markers in autosomal dominant Alzheimer's disease: A prospective cohort study. *The Lancet Neurology, 14*(8), 804-813. doi: 10.1016/s1474-4422(15)00135-0.

Yin, J. X., Maalouf, M., Han, P., Zhao, M., Gao, M., Dharshaun, T., . . . Shi, J. (2016). Ketones block amyloid entry and improve cognition in an Alzheimer's model. *Neurobiology of Aging, 39*, 25-37. doi:10.1016/j.neurobiolaging.2015.11.018.

Yourganov, G., Smith, K. G., Fridriksson, J., & Rorden, C. (2015). Predicting aphasia type from brain damage measured with structural MRI. *Cortex, 73*, 203-215. doi: 10.1016/j.cortex.2015.09.005.

Ziegler, J. C., Perry, C., & Coltheart, M. (2000). The DRC model of visual word recognition and reading aloud: An extension to German. *European Journal of Cognitive Psychology, 12*(3), 413-430. doi:10.1080/09541440050114570.

Appendix

I. The graded Object-Naming Task (GONT)

Retained items sets with the percentage of namability (percentage of people who could name the item in the respective groups) in HC and patients.

Item in German	***Translation***	***Namability in HC***	***Namability in patients***
Easy item set			
Rüssel	Trunk	100	100
Rose	Rose	100	100
Hahnkamm	Comb	100	97
Docht	Wick	100	94
Storch	Stork	100	91
Zebra	Zebra	100	91
Geweih	Antlers	99	97
Ellbogen	Elbow	99	95
Thermometer	Thermometer	99	91
Gerüst	Scaffolding	99	88
Anker	Anchor	99	87
Sieb	Sieve	99	83
Schaukelstuhl	Rocking-Chair	98	85
Feuerlöscher	Fire-extinguisher	98	85
Globus	Globe	96	85
Gitarre	Guitar	96	87
Lupe	Magnifying glass	96	79
Hammer	Hammer	95	90
Amboss	Anvil	95	81
Schuppen	Scales	95	77
Kompass	Compass	94	66
Grübchen	Dimple	93	68
Stethoskop	Stethoscope	91	59
Boje	Buoy	90	71
Pelikan	Pelican	90	59

Item in German	*Translation*	*Namability in HC*	*Namability in patients*
Intermediate item set			
Stimmgabel	Tuning Fork	88	70
Sonnenuhr	Sun-dial	88	65
Ventilator	Fan	87	58
Wäscheklammer	Peg	85	94
Sanduhr	Hourglass	85	61
Wetterhahn	Weather vane	84	49
Bullauge	Porthole	79	61
Kehlkopf	Adam´s apple	79	50
Vorhängeschloss	Padlock	77	73
Schädel	Skull	76	21
Schießscharte	Loophole	74	54
Bommel	Pom-pom	73	60
Klarinette	Clarinet	72	49
Lama	Lama	72	44
Graffiti	Graffiti	71	28
Wirbel	Vertebra	66	42
Qualle	Jellyfish	66	33
Golfball	Golfball	66	14
Gorilla	Gorilla	65	25
Zauberwürfel	Rubik´s Cube	62	19
Mandoline	Mandolin	61	31
Blasebalg	Bellows	61	30
Ellipse	Ellipse	59	34
Eckzahn	Canine	59	32

Item in German	*Translation*	*Namability in HC*	*Namability in patients*
Hard item set			
Fünfeck	Pentagon	57	44
Sehne	Tendon	56	41
Krabbe	Crab	55	32
Iris	Iris	52	27
Skorpion	Scorpion	48	27
Objektiv	Lens	46	14
Querflöte	Flute	43	15
Schnauze	Snout	38	52
Spindel	Spindle	38	20
Sextant	Sextant	34	22
Stößel	Pestle	33	9
Wimpernzange	Eyelash Curler	33	6
Abakus	Abacus	28	8
Chamäleon	Chameleon	27	3
Metronom	Metronome	24	13
Oboe	Oboe	20	8
Fagott	Bassoon	17	7
Barteln	Barbel	13	17
Seeigel	Sea Urchin	13	5
Kapuzineraffe	Capuchin	10	9
Chinchilla	Chinchilla	9	9
Joch	Yoke	9	8
Petschaft	Seal	8	5
Wetzstahl	Sharpening steel	7	6
Ammonit	Ammonite	7	3

II. Sentence Comprehension Test (SECT) [English]

Practice items

Thomas is heavier than Brian. *Who is lighter?*

Alan was forgiven by Tom. *Who was guilty?*

The market, the store is behind, is busy. *What is busy?*

Passive sentences

9 words sentences

Bill was arrested by Kevin. *Who was the suspect?*

The lion was eaten by the tiger. *Who survived?*

John was hit by Adam yesterday. *Who got hurt?*

James was sacked by Claire. *Who was the boss?*

12 words sentences

The boy was taught by the girl yesterday. *Who was the teacher?*

Joe was treated by Mary in the hospital. *Who was the doctor?*

Jack was visited by John at his home. *Who was the host?*

The girl was defeated by the boy in the match. *Who won?*

Embedded sentences

9 words sentences

The boy, Charlotte likes, is clever. *Who is clever?*

The runner, Jack overtook, was tall. *Who was tall?*

The girl, Tom chases, is thin. *Who is thin?*

The man, Simon visited, was sick. *Who was sick?*

12 words sentences

The cycle, the red car is behind, is old. *What is old?*

The bowl, the big fish is in, is red. *What is red?*

The village, the town is far from, is crowded. *What is crowded?*

The carpet, the big table is on, is brown. *What is brown?*

Comparative sentences

9 words sentences

Robert is younger than his brother. *Who is older?*

Kate runs much faster than Alex. *Who is quicker?*

Ann is richer than Jane. *Who has more money?*

Wendy dances better than Katy. *Who wins the prize?*

12 words sentences

The lounge is bigger than the room. *Which one has more space?*

The car is cheaper than the new bike. *Which one costs more?*

The tiger is much stronger than the lion. *Which one is weaker?*

Andrew is taller than Peter. *Who is the shorter of the two?*

III. Make A Sentence Test (MAST)

Presented Words	**Example of a correct answer**
Active sentences	
Practice : Ken lose key house	*Ken lost his key in the house.*
Emma bake pie party	*Emma baked a pie for the party.*
Thomas hit ball bat	*Thomas hit the ball with his bat.*
Fred put coat chair	*Fred put his coat on the chair.*
Passive sentences	
Practice: Dinner cook George yesterday	*The dinner was cooked by George yesterday.*
Key leave car yesterday	*The key was left in the car yesterday.*
Cake eat girl yesterday	*The cake was eaten by the girl yesterday.*
File keep cabinet office	*The files are kept in the cabinet in the office.*

Interrogative sentences	
Practice: Who teacher ask question?	*Who did the teacher ask the question from?*
Jake come party tonight?	*Is Jack coming (Does…come) to the party tonight?*
Sally go dinner yesterday?	*Did Sally go to the dinner yesterday?*
Ask Joe mend fence?	*Who asked (did…ask) Joe to mend the fence?*
Relative sentences	
Practice: Simon choose trousers blue	*Simon chose the trousers that were blue.*
Doctor treat boy sick	*The doctor treated the boy who was sick.*
Plumber change pipe leak	*The plumber changed the pipe that was leaking.*
Adam find cat lose	Adam found the cat that was lost.

Embedded sentences	
Practice: Dress Helen buy pretty	*The dress that Helen bought was pretty.*
Food Tom bring delicious	*The food that Tom brought was delicious.*
Prize student win special	*The prize that the student won was special.*
Car John buy fast	*The car that John bought is/was fast.*

IV. Verb inflection

Frequency was computed thanks to the dlexDB (Heister et al., 2011). The number of occurrences of an annotated type in corpus is case sensitive.

Verb	***Translation***	***dlex annotated type: absolute frequency***
Regular		
High frequency		
Lernen	To learn	4944
Folgen	To follow	4863
Öffnen	To open	2283
Sagen	To say	34383
Fühlen	To feel	2603
Fragen	To ask	5842
Low frequency		
Konstruieren	To construct	341
Glühen	To glow	98
Reimen	To rhyme	57
Ölen	To oil	44
Lächeln	To smile	365
Gähnen	To yawn	56
Irregular		
High frequency		
Bringen	To bring	18346
Beginnen	To begin	3967
Zwingen	To force	2094
Nehmen	To take	16955
Treten	To step	3914
Bleiben	To stay	17036
Low frequency		
Bestechen	To bribe	140
Blasen	To blow	342
Schlingen	To gorge	63
Reiten	To ride	207

V. SEntence Comprehension Test Visual (SECT V) [German]

Practice items

Steffi hat einen Kuchen gebacken. Wer hat einen Kuchen gebacken?

Jana wird von Paul ins Restaurant eingeladen. Wer hat bezahlt?

Der Mann, den die Frau gestern getroffen hat, ist Lehrer. Wer ist Lehrer?

Tobias ist kleiner als Felix. Wer ist größer?

Active sentences

Der Mann ist heute krank geschrieben. Wer ist heute krank geschrieben?

Der Fahrer hat das Radio angemacht. Wer hat das Radio angemacht?

Der Reisende hat seinen Ausweis vergessen. Wer hat seinen Ausweis vergessen?

Das Ehepaar hat eine neue Couch gekauft. Wer hat eine neue Couch gekauft?

Die Frau hat sich heute beim Ratgeber beschwert. Wer hat sich beim Ratgeber beschwert?

Die Touristen haben ein Hotel am Meer gebucht. Wer hat ein Hotel am Meer gebucht?

Passive sentences

Franziska wird von Sandra unterrichtet. Wer ist Lehrerin?

Claudia wird von Victoria betrogen. Wer ist sauer?

Katrin wird von Markus festgenommen. Wer ist schuldig?

Fabian wird von Thomas im Restaurant bedient. Wer ist Kellner?

Der Frau wurde von der Nachbarin geholfen. Wer brauchte Hilfe?

Der Chef wurde von seinen Mitarbeitern verklagt. Wer ist schuldig?

Embedded sentences

Der Kater, der dem Hund folgt, ist braun. Wer ist braun?

Der Arzt, der meinen Cousin untersucht hat, ist jung. Wer ist jung?

Der Arzt, der den Patienten geheilt hat, ist krank. Wer ist krank?

Die Krankenschwester, die den Patienten betreut hat, hat blonde Haare. Wer hat blonde Haare?

Die Küche, die in dem neuen Haus ist, ist sehr groß. Was ist sehr groß?

Der Zeuge, der gegen den Angeklagten ausgesagt hat, ist letzten Monat gestorben. Wer ist letzten Monat gestorben?

Comparative sentences

Die Küche ist dunkler als das Wohnzimmer. Was ist heller?

Meine Tante ist älter als meine Mutter. Wer ist jünger?

Der Fisch ist weniger klug als die Katze. Wer ist dümmer?

Das Hotel ist teurer als das Gasthaus. Was ist preiswerter?

Im Deutschen ist die Rechtschreibung leichter als die Grammatik. Was ist schwerer?

Das Wetter ist weniger schlecht im Süden als im Norden. Wo ist das Wetter besser?

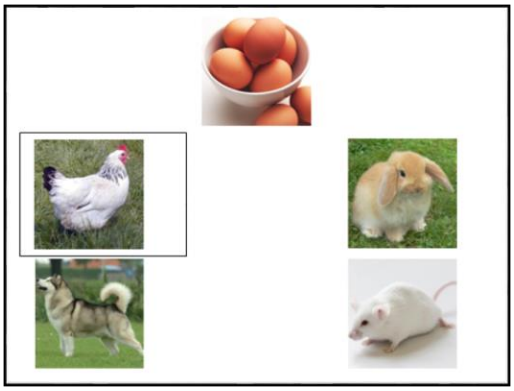
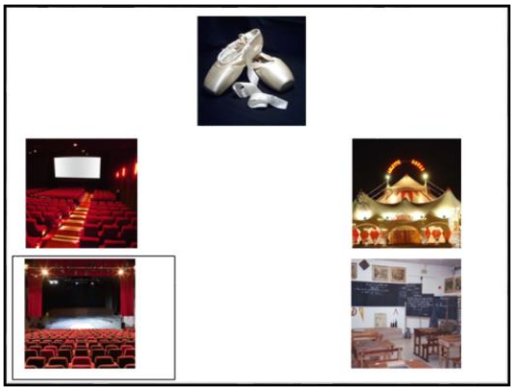
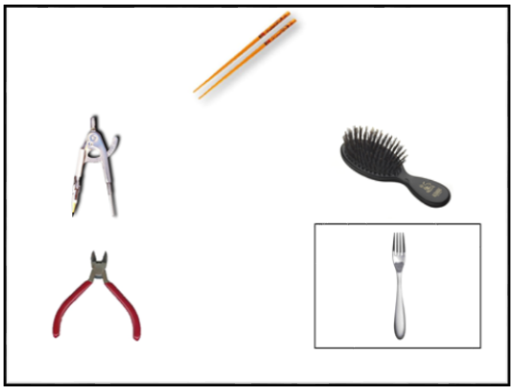
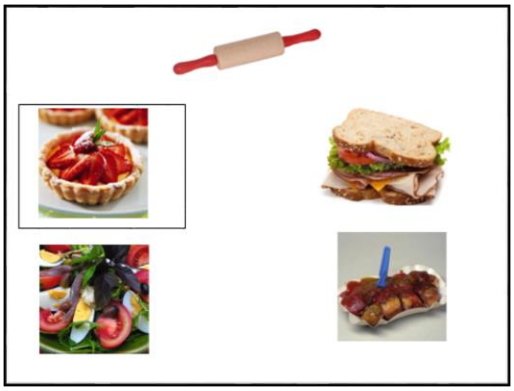
VI. Kaffee Kuchen Test (KKT)

The framed picture is the correct answer.

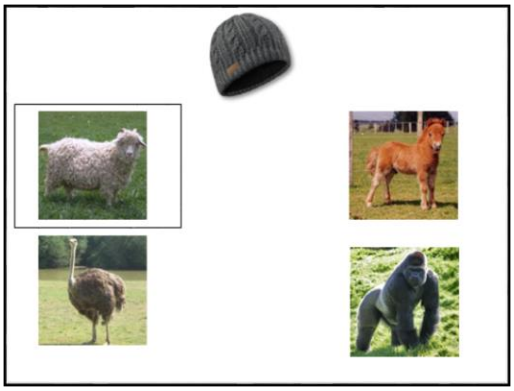
Practice item

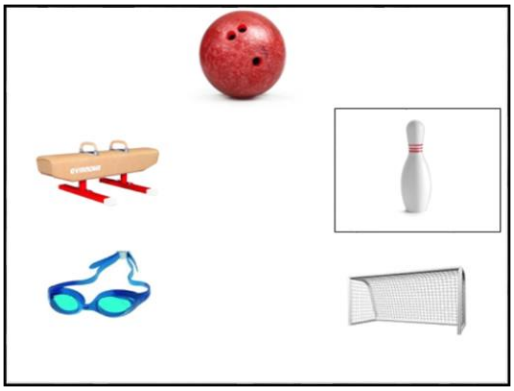
"You are going to see 5 pictures. Please tell me which of the four bottom pictures matches best with the one on the top of the screen"

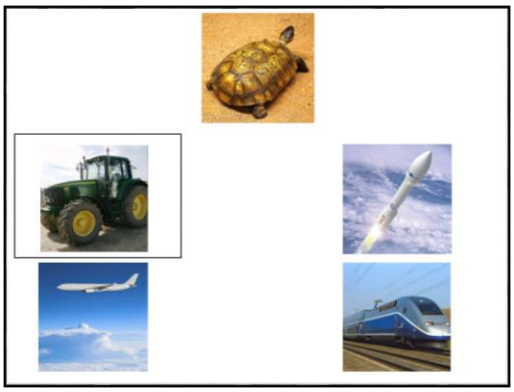
Test items

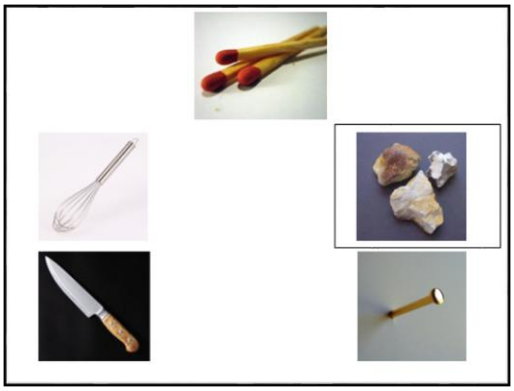





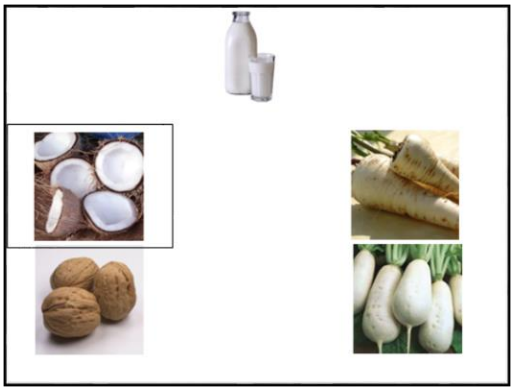


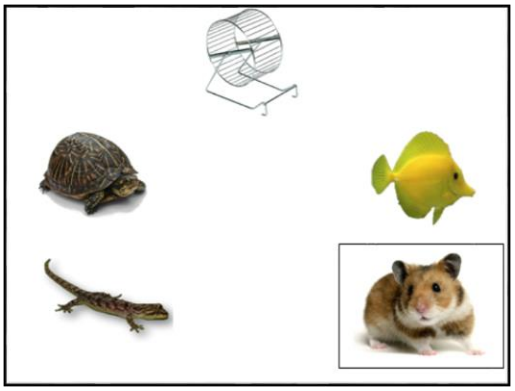


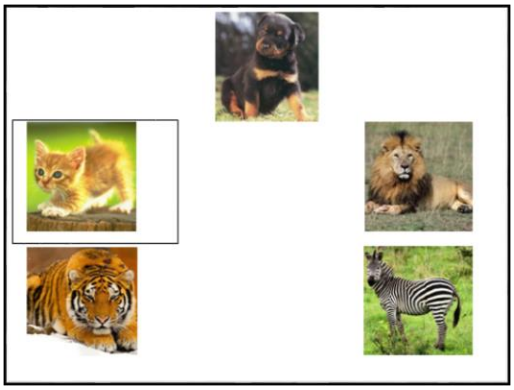


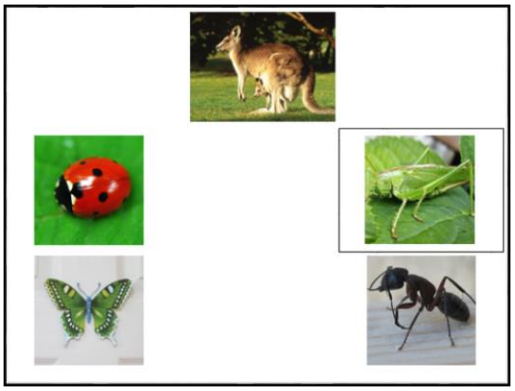












Lebenslauf

PERSÖNLICHE INFORMATION

ORNELLA BILLETTE
E-Mail: ORNELLA.BILLETTE@GMAIL.COM
Familienstand: verheiratet, 1 Sohn
Staatsangehörigkeit: französisch
Alter: 29 ~ geboren am 03.05.1989
Geburtsort: Reims, Frankreich

AKADEMISCHER UND SCHULISCHER WERDEGANG

Sep. 10 - Jul. 12	Master of Science in Neuropsychologie und kognitive Psychologie an der Universität Paris Descartes (Université Sorbonne Paris Cité), Frankreich (mit Auszeichnung)
Sep. 09 - Aug. 10	Auslandsjahr, Bachelor of Science in Psychologie, University of Malta, Malta
Sep. 07 - Jul. 10	Bachelor of Science in Psychologie an der Universität Reims Champagne-Ardenne, Frankreich (mit Auszeichnung)
Sep. 04 - Jun. 07	Abitur am Gymnasium „Marc Chagall“ in Reims, Frankreich (mit Auszeichnung)

BERUFSERFAHRUNG, PRAKTIKA, PROJEKTE

Mai 17- Jun. 18	Elternzeit
Jul. 15 - Mär. 17	Doktorandenvertreterin des DZNE Magdeburg
seit Feb. 13	Wissenschaftliche Mitarbeiterin/ Doktorandin des DZNE, AG Kognitive Neurologie und Neurodegeneration, Magdeburg Betreuer: Prof. Peter Nestor
Sep. 11- Jul. 12	Praktikum in klinischer Neuropsychologie auf der Neurologiestation, Krankenhaus Maison Blanche, Reims, Frankreich Masterarbeit II: Suche nach Gedächtnisunterschieden für die Differenzialdiagnostik von kortikaler und subkortikaler neurodegenerativer Pathologie (mit Auszeichnung) Betreuer: Dr. Nathalie Ehrlé und Dr. Julien Barra
Jul. 11	Praktikum an der deutsch-französischen Kita „Au clair de la lune“, Magdeburg, Beobachtung der bilingualen Sprachentwicklung
Mär. 11 - Jun. 11	Praktikum in klinischer Neuropsychologie auf der Geriatrie Station, Émile Roux Krankenhaus, Limeil-Brévannes, Frankreich
Sep. 10 - Jun. 11	Masterarbeit I: phonologische und semantische Entwicklung an zweijährigen einsprachigen und bilingualen Babys mit Hilfe der Eye-Tracking Methode (mit Auszeichnung), Laboratorium der Wahrnehmungspsychologie, CNRS, Paris Betreuer: Dr. Pia Rämä
Sep. 08 - Aug. 09	freiwillige Arbeit im GENEPI, Bildung für Insassen

Sep. 07 - Aug. 08 freiwillige Arbeit im AFEV, Schulhilfe für benachteiligte Schüler

PUBLIKATIONEN

Mai. 15 **Billette, O.V,** Sajjadi, S.A., Patterson, K. & Nestor, P. (2015). SECT and MAST: new tests to assess grammatical abilities in primary progressive aphasia.

Erklärung

Hiermit erkläre ich, dass ich die von mir eingereichte Dissertation zu dem Thema:

"Improving cognitive testing in the context of dementia: Tools for an earlier and better-characterized diagnosis of Alzheimer´s dementia and Primary Progressive Aphasia"

selbständig verfasst, nicht schon als Dissertation verwendet habe und die benutzten Hilfsmittel und Quellen vollständig angegeben wurden.

Weiterhin erkläre ich, dass ich weder diese noch eine andere Arbeit zur Erlangung des akademischen Grades doctor rerum naturalium (Dr. rer. nat.) an anderen Einrichtungen eingereicht habe.

Magdeburg, Juni 2018

Ornella Billette