Universiteit Utrecht

Master thesis

Diagnostic Instrument for Mild Aphasia: adaptation and standardization in French

A preliminary study

Supervisors: Djaina Satoer (Erasmus MC, Rotterdam), Bert Le Bruyn (Utrecht University)

Research Master in Linguistics, UU

Héloïse Pierret, 6230660

August 2019

Erasmus MC
Universitair Medisch Centrum Rotterdam

Abstract

The Diagnostic Instrument for Mild Aphasia (DIMA) (Satoer et al., 2019) is a valid tool developed in Dutch to evaluate the language of patients with suspected mild aphasia at production and perception levels in the most important linguistic domains: phonology, semantics and (morpho-) syntax. It was developed for pre- and postoperative use on patients with low-grade gliomas. Due to the brain tumor's slow growth, neural networks undergo reorganization and mild gradual degradation can be observed in higher functions (Duffau, 2014). However, there is a lack of short and sensitive diagnostic instruments (Satoer et al., 2013). No such battery exists in French either (Le Rhun et al., 2009). Therefore, the purpose of this research was to adapt the Dutch battery in terms of linguistic and psycholinguistic variables into French. Two healthy French-speaking populations were compared (Belgians and French participants, total n=67) and other potential effects of demographic variables such as age, gender, level of education (and early and late bilingualism) were explored. Ceiling results in accuracy rates were registered for all tasks. Although no effect of nationality was found, there was an effect of age for most of the tasks, and education and gender also influenced performance in some tasks. Bilingualism did not affect performance. We can thus consider that the final version of the French DIMA as presented here is a valid instrument and that the standardization and clinical validation can be pursued. The application can also be extended to patients with other neurological diseases (such as strokes or traumatic brain injury) with suspected aphasia. In addition to enhancing the reliability of the protocol, creating an internationally valid battery enables more patients to be included in anatomo-clinical studies.

Keywords: DIMA, Low-Grade Glioma, mild aphasia, French, standardized language test adaptation

Acknowledgments

First and foremost, I would like to thank all the people who contributed to this Master's thesis directly or indirectly.

Above all, thank you to my supervisors Djaina Satoer and Bert Le Bruyn for their guidance, their feedback and their availability. Thank you Bert for your recommendations and for being so quick at answering my questions. Thank you Djaina for your advice and the interesting discussions, for your enthusiasm, and for inviting me to the block.

I would also like to thank Isabelle Poisson and Marion Barberis, my internship supervisors at the Lariboisière Hospital in Paris, for their time and their constructive criticism when I showed them the first version of the adaptation of the DIMA, as well as for all the valuable knowledge I acquired during my stay in Paris.

Thank you to Piet van Tuil for the advice session on the statistics.

Rachel, thank you for the language proof reading and for being so fast at it. That was perfect!

Many thanks to my family and friends who supported me and motivated me all along, and thanks to whom I could take nice breaks when I came back to Belgium or went to Paris. Special thanks to Sarah and Daphné who gave me invaluable advice on how to *start*; that was a turning point. Last but not least, a huge thank you to my best housemate Anouk for actually everything: from your smart comments to putting up with my weird hours during intense work. I think that being in the same *galère* (here's a variant for you!) was also part of the fun!

Finally, thank you to all the participants in this study, who often gathered together so that I could test several people in one testing session, and a special thanks to the people who contributed to recruiting participants. Thank you to my sister Mélusine and my Bonne-Maman for spreading the word. It was a huge help!

Table of content

Il faudra à Nacht encore trois autres semaines pour pouvoir reparler presque à son habituel niveau de compétence, dans les trois langues. Deux mois après l'opération, aucune altération de ses capacités verbales ne restera, de l'extérieur, encore détectable. Lui sait pourtant fort bien que, depuis cette chirurgie, sa fatigue physique et neurologique est plus rapidement atteinte qu'auparavant. Et surtout que, fatigué, les mots peuvent lui venir plus lentement, voire aller se cacher quelques instants avant de resurgir plus tard… Mais pour prolonger de quelques années sa vie pleine et consciente, ce prix est dérisoire.

– Patrick Declerck, *Crâne*

1. Introduction

Human language consists of sets of rules mapping arbitrary symbols onto concepts. They are organized into intertwined systems, combining sounds and rhythms (phonology), pieces of words (morphology), words and meanings (lexicon and semantics), words into sentences (syntax) and in turn into longer utterances (discourse), and language use and the context (pragmatics). Those rules are learnt by newborns no matter what their mother tongue is. It is the basis of human communication. Language is everywhere, under different modalities: written and verbal, relying both on visual and auditory abilities. Language needs to be perceived but also produced, and speech therefore requires muscles and movements (neural motor commands, articulators). For a message to be meaningful, efficient and relevant, speakers also rely upon working memory, attention, inhibition, etc. (executive functions). All those functions are located as networks in the brain and make language a so-called 'higher cognitive function' (Luria, 2012).

Any of those functions and modalities can be affected to different degrees due to a brain disease, such as a stroke for instance. Examples that can result are anomia (i.e. impairment in naming due to difficulties of access to the lexicon), apraxia of speech (i.e. a speech impairment linked with neural motor commands) or Wernicke's aphasia (i.e. impairment in the comprehension of speech). We use the term 'aphasia' to name disturbances in language production or perception that can result from different brain injuries. It seems thus evident that a severe or even moderate aphasia can have a huge impact on communication. Nonetheless, patients' quality of life can be highly impacted even by mild aphasia (Cruice, Worral & Flickson., 2006). It can be difficult for instance to maintain dynamic communication, social relationships and a professional activity, and has consequently been reported to correlate with depression (Le Rhun, Delgeuck, Devos, Pasquier & Dubois, 2009).

Specifically, mild aphasia can be caused by brain tumors such as low-grade gliomas. Due to their slow growth, neural plasticity is facilitated and thus allows functional reorganization (Duffau, 2014). Brain tumors are very different from strokes as they are gradual and require resection as part of the treatment. The Dutch Intraoperative Language Protocol (DuLIP) (De Witte et al., 2015) is a test battery that was developed to target intraoperative language mapping during awake surgery with maximal tumor resection. Although this procedure is designed to avoid language impairments and preserve the quality of life of patients, language recovery following the intervention needs to be controlled (Satoer et al., 2018). Nevertheless, the DuLIP is very time-consuming, and items are easy enough for patients with mild aphasia to perform in the awake surgery setting. As such, they are not sensitive for mild impairments in the preoperative phase.

Based on these considerations, Satoer and colleagues (2019) developed the Diagnostic Instrument for Mild Aphasia (DIMA) (Satoer et al., 2019), a valid tool created in Dutch to evaluate the language performance of patients with a suspicion of mild aphasia. The DIMA is based on the DuLIP from which the most complex tasks have been selected and revised, since impairments are often too subtle for regular aphasia diagnostic instruments. The authors show that the protocol can be used preoperatively as a baseline for task selection for the DuLIP intraoperatively, and in the postoperative phase to assess recovery. The DIMA is built to detect subtle impairments in production and/or comprehension, targeting the most important linguistic domains (i.e. phonology, semantics and syntax) and includes complex tasks (with reaction time). It was standardized on 211 participants from Flanders and The Netherlands and validated on patients with brain tumors (low-grade glioma and meningioma) (Satoer et al., in progress). The authors also suggest that it can be applied to patients with other neurological etiologies.

Lack of an instrument to diagnose mild aphasia, especially linked with low-grade gliomas, has been reported in other languages as well, namely in French (Le Rhun et al., 2009). We have thus adapted the DIMA into this language and a preliminary standardization in 67 healthy participants from France and Belgium has been carried out. This thesis describes the adaptation of the tasks according to the relevant (psycho-)linguistic variables and the results from the healthy population. It will be structured as follows. The next section discusses how low-grade gliomas are implicated in language deficits, introducing the concept of neuroplasticity and the DuLIP battery. Section 3 describes the DIMA in further details. Section 4 describes the current study. The implications linked with standardizing a French battery in France and in Belgium are examined. Section 5 states the research questions and hypotheses. Section 6 discusses the methods, followed by the results in section 7. Section 8 is a discussion including clinical application, limitations and further directions. Section 9 is the conclusion.

2. Diffuse Low-Grade Glioma and language

Diffuse low-grade gliomas (LGG) are primary brain tumors arising from glial cells (i.e. non-neuronal cells in the brain) (Duffau, 2014). Due to their preferential location in direct proximity of so-called 'eloquent' areas, i.e. essential for language, sensory or motor functions (Duffau & Capelle, 2004), their growth and their treatment may impair language functions. The following sections focus on why this type of tumor can cause mild aphasia, what is the current 'gold standard' treatment and how to target the language deficits.

2.1. Neuroplasticity

As primary brain tumors, LGG are characterized by a slow growth during several years (Mandonnet et al., 2003). Despite their slow progression, they usually end up mutating to high-grade gliomas. Crucially, LGG display almost no behavioral symptoms until their size and nature has reached such a point that they cause sudden seizures (DeAngelis, 2001). DeAngelis mentions that, although it is possible to diagnose such tumors with magnetic resonance imaging, it is often not seen until a seizure has occurred due to the absence of perceptible neural deficits as the slow growth facilitates neural plasticity. Gradual degradation can then be observed in higher functions, working memory, attention, executive functions, learning and in some emotion processing functions (Duffau, 2014), which are symptoms of the limits of the plasticity of the system facing neural destruction (Desmurget, Bonnetblanc & Duffau, 2007). This type of brain lesion occurs however quite seldom compared to strokes (Desmurget et al., 2007) and their analysis was therefore largely lacking from research during the last century.

Most of our knowledge about brain functions was indeed based on findings such as Broca's and Wernicke's on stroke patients at the end of the 19^{th} century. They are the fathers of the localizationist approach on neurology, which claims that damage in the brain in functional regions responsible for language comprehension and production causes irreversible language impairments. However, strokes are very different from infiltrating tumors in terms of onset, time course, growth and localization (Andersson, 2017). The more recent connectionist philosophy argues that the regions at stake are not solely responsible for a function. Instead, functions are carried out by networks with cortical epicenters at the surface of the brain and white fiber bundles located deeper in the white matter, connecting the epicenters together (Catani, Howard, Pajevic & Jones, 2002; Catani et al., 2012). This claim is supported by evidence for recovery from lesions such as diffuse LGG, which specifically infiltrate white matter bundles (Duffau, 2014). Epicenters are thus involved in several functions, enabling reorganization whenever necessary and possible. The factor time is especially crucial for the reorganization potential and its role came to light with the findings concerning slowly growing tumors (Desmurget et al., 2007).

Starting in the 1930s, those discoveries lead to the development of neurosurgery in an awake setting with Direct Electrical Stimulation (DES) (Ojemann, 1983; Penfield & Roberts, 2014). Initially designed to treat untraceable epilepsy, DES is now also used for preventive tumor resection. This method consists of directly mapping language functions onto the cortical and subcortical areas of the brain responsible for reorganized brain networks in the areas of language, visuospatial cognition, calculation or emotion (Papagno, 2017). Although relatively time-consuming for both patients and medical staff, DES still appears as the 'gold standard' for intraoperative use as it increases chances of survival in addition to enabling a better quality of life (De Witt Hamer et al., 2012). Besides limiting the growth of the LGG, awake surgery indeed allows the core of the tumor to be removed without damaging the cognitive functions. Out of the 103 patients in Duffau et al. (2003) who had been diagnosed with LGG, 93% have been able to resume work within a year after surgery without showing perceptible deficits, and the remaining ones only showed mild cognitive impairments.

With regards to language, there is no real generalization yet as to when and how such subtle deficits are present. A thorough assessment is however invariably recommended to detect differences between normal and near normal language (Antonsson et al., 2018; Satoer et al, 2012). Antonsson et al. reported that, before surgery, only a limited number of patients reported subjective deficits (2018). Their findings from language assessments correlated as they found very light, yet significant, impairment for lexical retrieval (naming and word fluency), a finding which is in line with Satoer et al. (2012) and Santini et al. (2012). Antonsson (2017) also reports a slight deficit in writing pre-operatively, also found by Santini and colleagues (2012).

Satoer et al. (2013) investigated spontaneous speech. The analyses do reveal abnormalities before surgery, namely that patients produce more incomplete sentences than controls, and that this was due to incomplete sentences with the omission of content words. This confirms findings on a lexical retrieval deficit. They also found a link between impairments and tumor location. In addition, cognitive impairments add to the linguistic ones, namely concerning divided attention, processing speed and episodic verbal and non-verbal memory (Le Rhun et al., 2009; Santini et al., 2012; Talacchi, Santini, Savazzi & Gerosa, 2010). The authors note that those deficits have a significant link with the quality of life and can help diagnose the tumor or its regrowth.

In the acute postoperative period, impairment is generally found, specifically in word retrieval with picture naming (Santini et al., 2012; Talacchi et al., 2010). At three months postoperatively, Satoer et al. (2012) found that patients had a decline in language functions on the Boston Naming Test (Kaplan, Goodglass & Weintraub, 2001) and fluency tasks and executive functions (Stroop, TMTA&B). As to the written language skills (reading and writing) and non-words repetition, van Ierschot et al. found partial recovery at the individual patient level compared to before surgery (2016). Antonsson (2017) found slower writing rate and more pauses than in the preoperative phase. Between three and six months postoperatively, Santini et al., found a trend of improvement in language functions (including writing skills) although preoperative performance was not reached (2012). In this time-span, Papagno et al. (2012), also show that patients operated on the uncinated fasciculus had only partially recovered for famous faces naming and objects naming.

With regard to spontaneous speech, Satoer et al. (2013) found that, in addition to incomplete sentences, patients' utterance length was also below normal mean scores. At twelve months post-operatively, there was still evidence of deviant language: category fluency was deteriorating although naming and letter fluency showed improvement compared to three months after surgery, and improvements in memory were found (Satoer et al., 2014). Spontaneous speech also remained impaired on the long term, thus being more sensitive than comprehensive tests (Satoer et al., 2018). The authors suggest that assessment should therefore encompass all linguistic levels and not only naming. Similarly, it was argued that naming does not reflect the complexity of 'real' speech (Prins & Bastiaanse, 2004). Nevertheless, the results presented show variation between studies, especially due to patient sample sizes and variable treatments.

The next section discusses the DuLIP, a protocol developed for the assessment of language intraoperatively.

2.2. DuLIP

To target the needs for a standardized linguistic protocol to be used in the pre- intra- and postoperative settings (De Witte & Mariën, 2013; Rofes et al., 2017), the Dutch Linguistic Intraoperative Protocol was developed by De Witte et al. (2015). It facilitates intraoperative mapping of eloquent language areas and is a valid test battery, standardized in a population of 250 native Dutch participants. The battery provides language specific tasks as suggested by Talacchi (2013) through linguistic tests that assess the main areas of productive and perceptive language for adequate communication, including phonology-, semantics-, syntax- (and articulation-) oriented tasks, and the total duration is about 1.5 hours. The DuLIP was designed for awake surgery with DES, hence it is built on a 'location-function-task' model. This means that each task targets (a) language function(s) that is/are associated to a preferred brain area. The tasks included allow thus to identify (un)impaired functions and map functional brain structures in order to evaluate the potential for maximal tumor resection.

The anatomo-clinical correlations mentioned in the DuLIP are made for the language dominant hemisphere. Figure 1 illustrates the network of cortical sites and subcortical pathways that are the basis for the 'location-function-tasks' paradigms for cortical mapping. The tasks are presented in Table 1 below.

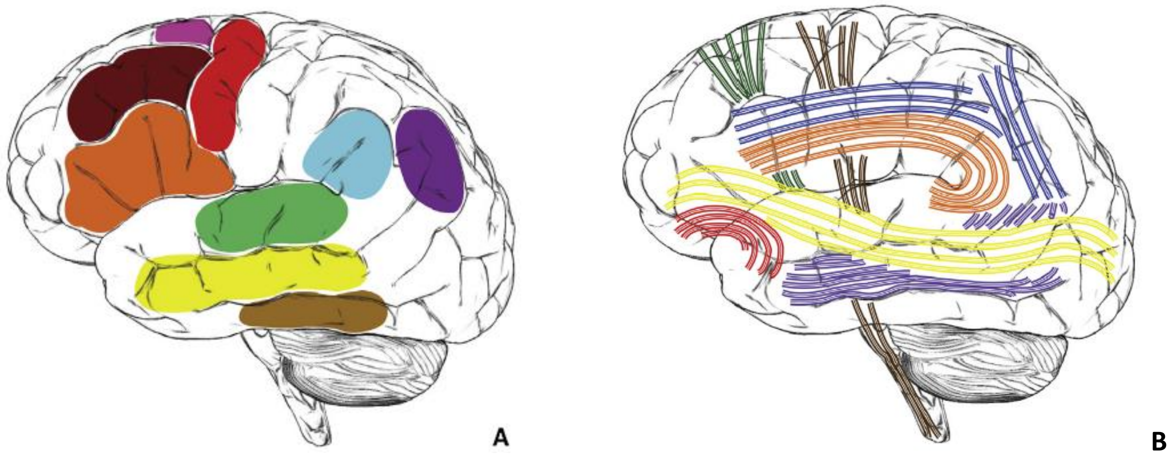

Figure 1: (A) Cortical sites and (B) subcortical pathways of the functions for which tasks are developed in the DuLIP (Adapted from De Witte et al., 2015).

In conclusion, the DuLIP is a valid test battery for pre-, intra- and postoperative use in patients with a brain tumor in eloquent areas. Adaptations are currently being carried out in several languages, among which French (Quinchon et al., 2019; Quinchon & Tahiri, 2019). Nevertheless, it is an extended protocol which is time-consuming and in which tasks are not all as sensitive. Since they are focused on the intra-operative procedure, they are designed for patients with mild language deficits to be able to perform them in this setting, while pre- and postoperative assessments necessitate more sensitive tasks. (De Witte, Satoer, Visch-Brink & Mariën, 2016). There is thus still a need for a diagnostic tool for mild language impairments that can be administered quickly for this purpose. The next section describes the Diagnostic Instrument for Mild Aphasia (DIMA) and how it is developed based on the DuLIP.

Table 1: Intraoperative linguistic tasks from the DuLIP and those adopted in the DIMA (Adapted from De Witte et al., 2015).

Linguistic level	**Tasks**	**DIMA**	**Stimuli (examples)**
Phonology (with DES)	- repetition of words:	v	
	• repetition of 3-syllabic words with alternating word accents		- discussie (discussion)
	• repetition of 2-syllabic words		- wortel (carrot)
	• repetition of words with phonemic similarities		- individu (individual)
	• repetition of syllabic words with consonant clusters		- programma (programme)
	- reading with phonological odd word out[a]		- wijn, pijn, lijn, kat (wine, pain, line, cat) → answer kat (cat)
Semantics (with DES)	- reading with semantic odd word out[a]		- been, arm, raam, voet (leg, arm, window, foot) → answer raam (window)
	- naming with semantic odd word out[a]	v	- pictures of 'borstel, hond, kat' (brush, dog, cat) → answer borstel (brush)
	- semantic association task[a]		- auto, fiets … (car, bike) → answer e.g. bus
	- sentence completion[a] (semantically induced sentences)	v	- Hij snijdt met een … (He cuts with a …) → answer e.g. mes (knife)
Syntax (with DES)	- verb generation		- bal (ball) → goien/werpen (to throw)
	- action naming (3rd person singular, transitive verbs) 60 actions[a] (RUG – Vrije Universiteit Brussel)		- de man … (the man) → answer loops (runs) → picture of a man who is running
Articulation – Praxis	- verbal diadochokinesis test[a]		- repeat /papapa/, /pataka/, /papopu/, /pafpafpaf/, /dafnaflaf/, /pafpofpuf/, /pafpaspag/ x5
Naming (with DES)	- objects naming, 100 objects[a]		- black and white drawings of objects Dit is … (This is …) → e.g. answer hond (dog) → picture of a dog
Phonology	- phonological sentence judgment	v	- De hokkel eet een gersie. (The hokkel eats a gersie.) → 'hokkel' and 'gersie' non-existing words
Semantics	- semantic sentence judgment	v	- De gieter smeert een boterham. (The watering-can prepares a sandwich. → wrong sentence
	- sentence completion (less semantically induced sentences) [a]	v	- Om 5 uur … (At 5 O'clock …) → possible answer: ga ik naar huis (I go home)
Syntax	- syntactic sentence judgment	v	- Er was over niets meer van het scheerapparaat. (There was over nothing left of the razer.) → wrong word order

[a] Presented with powerpoint slides.

3. DIMA

The DIMA was developed to counter the lack of a standardized test to detect mild aphasia in a short time span (Satoer et al., 2019). This instrument was developed in Dutch and targets the main linguistic domains, namely phonology, semantics and grammar/morphosyntax, including some cognitively more complex tasks as suggested in the study of Satoer et al. (2018). It is standardized in 211 native Dutch-speaking healthy participants from Flanders and The Netherlands. Clinical validation was carried out in patients with low-grade glioma and meningioma and is ongoing for patients with mild language impairments from other neurological etiologies. Potential for inclusion in standard language assessments for mild language disorders is suggested by the authors.

For its primary use is language assessment in pre- and postoperative phases of LGG, the DIMA is based on the DuLIP. Similarly to the DuLIP, the DIMA is built according to theoretical and clinical models linking language functions to cortico-subcortical anatomy (Papagno et al., 2012; De Witte et al., 2015). Hence, the tasks included in the DIMA mainly overlap with those in the DuLIP; they have been selected and adapted from this protocol by taking samples from the more complex ones and adapting others to make them cognitively more demanding (e.g. including reaction time) (Satoer et al., in progress).

Tasks included in the DIMA are indicated in Table 1 above. The battery contains in total: 1. Phonology – a repetition task with words, compound words, pseudo-words and sentences, and a phonological sentence judgment task; 2. Semantics – a naming task with semantic odd picture out targeting nouns and verbs, a sentence completion task with elicitation from narrow contexts (first part), and a semantic sentence judgment task; 3. Syntax – the second part of the sentence completion task with elicitation from broad contexts and a syntactic sentence judgment task.

Complexification of the tasks is achieved through the following adaptations. The repetition of pseudo-words and compound-words was added as there is evidence that non-existing words and compounds are processed differently to simple existing words (Semenza & Mondini, 2006). More specifically, it has been reported that aphasia patients were more impaired in compounds compared to long matched single words, and that depending on the type of impairment, they could have difficulties retrieving the verbal component (in V-N compounds) or made more constituent errors in transparent compounds compared to opaque ones (Lorenz, Heide & Burchert, 2014). Findings about impaired morphological processing in compounds rather than monomorphemic words have been reported for patients with fluent types of aphasia as well (Eiesland & Lind, 2012). Longer and phonologically more complex sentences have been added to the repetition task. In the naming task with semantic odd picture out, semantic categories were made closer than in the DuLIP and the 4-seconds paradigm from DES was kept. The sentence judgement task combines all three linguistic judgment types in visual modality and hand coordination (to press on the accurate key). Measurements of reaction time are included in the latter two tasks as it has been shown that although patients with mild aphasia could obtain high to ceiling scores in accuracy, they were hindered by their processing speed (Hendriks et al., 2018; Hickin et al., 2015; Le Rhun et al., 2009; Moritz-Gasser, Herbert, Maldonado & Duffau, 2012).

Results from Satoer et al. (2019) report no effect of gender nor handedness in all tasks for the 2 versions of the DIMA. As in the DuLIP, there was however an effect of age (the older the subject, the lower the scores) and of education (the more well educated, the higher the scores). The accuracy rate was above 80% for all tasks apart from the semantic odd picture out with actions. There is evidence that verb naming requires more cognitive resources (Bastiaanse & van Zonneveld, 2004) and the authors suggest that the time-paradigm might be adapted for this task. The clinical application of the test was performed on a LGG patient who showed severe impairment in the repetition tasks and mild impairment in the sentence completion task, although she performed within normal range at the Boston Naming Test, Token Test and verbal fluency. Another patient with meningioma one year after surgery performed within normal limits at the object naming task, Token Test and letter fluency but not category fluency and was severely impaired in the DIMA for two of the repetition tasks (Satoer et al., in progress).

The DIMA seems to be a sensitive tool for the diagnosis of mild aphasia. By coupling behavioral results from the DIMA and neuroimaging techniques, more insight can be reached in anatomo-clinical correlations and the battery can be further improved (Satoer et al., in progress).

4. Current study

The general lack of diagnosis instruments for subtle language disorders also exists in French. For instance, the DO80 was reported to lack sensibility (Rousset & Gatignol, 2014), and recent batteries now include reaction times in their scoring systems for finer assessments (e.g. the Batterie d'Evaluation des Troubles Lexicaux (BETL) by Mai, Grob-Nicolas & Muchembled, 2012). With regards to mild aphasia and LGG, Le Rhun et al. (2009) specifically mention the need for tests to evaluate processing speed, divided attention and memory.

As follows, the present study concerns the adaptation of the DIMA into French and its preliminary standardization in healthy controls. Currently the DIMA is being adapted to different languages in collaboration with several universities/hospitals. The value of the test does indeed increase in several ways. Besides broadening the scope of patients able to benefit from it, the adaptation adds scientific reliability to the protocol: the tasks, their feasibility and their outcome can be compared across different language populations. In addition, the creation of standardized protocols allows more patients and controls to be included in anatomo-clinical studies internationally and hence strengthens the results.

Tasks in the French version of the DIMA overlapping with the DuLIP have been selected from the French version of the DuLIP (Quinchon & Tahiri, 2019) and the others were created based on the original DIMA (see section 6.1 on the description of the tests including the adaptation of linguistic and psycholinguistic variables). All the tasks of the original DIMA have been adopted in French with the exception of the semantic odd picture out with actions due to low success rate in healthy participants.

4.1. Belgium versus France

The aim of this preliminary standardization study is to create a French version of the DIMA that may be used with valid norms both in France and in Belgium. Although it is not always specified, some test batteries mention being normalized in Francophonie, such as the batteries GRÉMOTS (Bézy, Renard & Pariente, 2016), Lexique (de Partz de Courtray, Bilocq, De Wilde, Seron & Pillon, 2001), and GRECOVASC (Roussel & Godefroy, 2016), although they all provide a single norm rather than different norms per country. It is however well-known that there are some regional usages and lexical variations generally accepted in the French speaking part of Belgium, Wallonia, known as *belgicisms*. Although 'standard' French, i.e. the language described in grammar books and dictionaries, is widely spread through literature and the media, some regionalisms remain very frequent and are the norm in Belgium (for inventories and dictionaries, see Bal, Doppagne & Gausse, 1994; Delcourt, 1998-1999; Francard, Geron, Wilmet & Wirth, 2010; Goosse, 2011; Massion, 1987). It may thus be the case that items are, or elicit responses that are, less/more frequent than the ones expected.

Given the potential for variation, the selected stimuli were thus controlled for not containing any regional language form and is based on 'standard' French such as described in reference works. Nevertheless, as mentioned previously, frequencies may vary for a single item, and non-standard responses might be received. We will therefore look at whether Belgian participants and French participants react similarly to the stimuli and whether they can be considered as one group or should rather have their own norms, possibly for specific tasks only.

5. Research questions and hypotheses

The research questions aim thus to evaluate test performance of the French population, how it is affected by different socio- and psycholinguistic variables, and whether it differs from the performance of the Dutch speaking population in the original standardization study. In addition, we evaluate the validity of the French version. More specifically, the research questions are listed below:

1. How does the French-speaking healthy population perform in the French version of the DIMA at task and item levels?
    a. What is the internal validity of the French adaptation like? Which items should be included in the definitive French version?
    b. Is there a difference in test performance affected by origin (i.e. between the group of French participants and the group of Belgian participants)? If so, could it be explained by other demographic and/or linguistic variables?

c. Is there a difference in test performance affected by gender, age and education level?

2. Is there a difference between the French and Dutch standardization studies? If so, is it affected by demographic and/or linguistic variables?

As the participants in this standardization study are healthy controls, the hypothesis is that results similar to the ones obtained in the DIMA from Dutch speaking participants will be replicated. We thus expect to find ceiling results for all subtests with mean scores above 80% in the definitive version.

In the literature, age and education level have been reported to affect performance on neurological assessments, with gender sometimes adding to the factors, and so across languages (Brucki & Rocha, 2004; Merck et al., 2011; Snitz et al., 2009). We therefore expect to find an effect of age on most subtests scores and reaction times (RTs), with lower scores and longer RTs for older participants. An additional effect of education is also expected to impact scores and RTs, with higher education being correlated with higher scores and faster RTs in all timed tasks.

Besides, a significant effect of dialect (Dutch vs. Flemish) was found in the original DIMA for some tasks, which authors suggest might be affected by potential methodological differences and by a more prominent multilingualism in the Flemish region where testing was carried out (Satoer et al, in progress). In the current study, as all participants were tested by the same experimenter, we do not expect methodological differences. However, given that some participants were recruited in a non-French environment (The Netherlands), an effect of linguistic interference might be expected and will therefore be taken into account.

6. Methods

6.1. Description of tests

6.1.1. Adaptation of psycholinguistic variables

The items were partly selected from the French version of the DuLIP (Quinchon et al., 2019; Quinchon & Tahiri, 2019) when there was overlap with existing tasks (i.e. repetition of words, repetition of sentences, sentence completion and sentence judgment tasks), and created when there was no DuLIP equivalent item (i.e. repetition of compound words, repetition of non-words, repetition of sentences, naming with semantic odd word out). The French database LEXIQUE (New, Pallier, Ferrand & Matos, 2001) and the study by Bonin et al. (2003) were used to control for the main linguistic and psycholinguistic variables: frequency, imageability, word length and word form. Images are retrieved from the database by Snodgrass and Vanderwart (1980). Any potential distressing emotional value was avoided in an item (e.g. death, brain…). In addition, cultural, psychological and social appropriateness also play a role in the cross-linguistic adaptation and were thus given special attention (Fyndanis et al., 2017). Within a task, items were presented in order of increasing complexity as errors are more frequent following a previously failed item. The variables per task are described below.

6.1.2. Tasks

As mentioned earlier, the French version of the DIMA features the same tasks as in the original version except for the semantic odd picture out with actions. Consequently, in order to reach ten items for naming in the final version, additional items were created for the naming task with semantic odd picture out – nouns and animals. Additional items were also created for each new task in order to select the ten most successful ones for the final version. Table 2 provides an overview of the tasks per linguistic domain included in the French DIMA with example items and the relevant linguistic variables. Subsequently, each task is described and information is given about the design of the items and the relevance of the linguistic variables for the assessment of each linguistic level.

Table 2: List of tasks included in the French DIMA

Linguistic level	Tasks	Examples French Stimuli	Main linguistic variables
Phonology	- Repetition of words:		
	• repetition of 3-syllabic words with alternating word accents, with/out phonemic similarities and consonant clusters	- aventure (adventure) - domino (domino) - cicatrice (scar)	- baseline (3-syllables) - phonological similarity - consonant cluster
	• repetition of 3- to 6-syllabic compound words with/out phonemic similarities and consonant clusters[b]	- hors-la-loi (out-law) - presse-citron (lemon squeezer) - auto-stoppeur (hitchhiker)	- 3-syll - consonant cluster - phonological similarity (4-syll)
	• repetition of pseudo-words with/out phonemic similarities and consonant clusters[b]	- titaubau - cuicaupréteur	- phonological similarity (3-syll) - consonant cluster (4-syll)
	• repetition of sentences with/out phonemic similarities and consonant clusters	- Il regarde les éléphants (He is looking at the elephants.) - Maman crie que les crêpes sont prêtes. (Mum shouts that the pancakes are ready.)	- phonological similarity (4 words) - consonant clusters and alliterations (6 words) - sentence structure
+ Reaction time[b]	- Phonological sentence judgement [a]	- Le ramion va a Paro (The ramion goes to Paro)	- presence of non-words
Semantics	- Naming with semantic odd picture out [a]	- pictures of 'serpent, chien, chat (snake, dog, cat) → answer serpent (snake)	- all three pictures belong to the same semantic category
	- Sentence completion[a] (semantically induced sentences)	- Je me lave les mains avec… (I wash my hands with…) → answer e.g. du savon (soap)	- single word induction
+ Reaction time[b]	- Semantic sentence judgment [a]	- Le passager but une carte. (The passenger drank a card)	- violation of semantic restriction
Syntax	- Sentence completion (less semantically induced sentences)	- A cinq heures … (At 5 O'clock …) → possible answer: je bois du thé (I drink tea.)	- verbal phrase induction
+ Reaction time[b]	- Syntactic sentence judgment [a]	- Nous avons déjà avoir du café. (We have already have coffee.) - Ce n'est pas le vôtre, c'est le nous. (That's not yours, it's we.) - Il aide sa petite sœur de manger. (He is helping his little sister of eat)	- wrong verb form - wrong pronoun - wrong preposition

[a] Presented visually (Powerpoint slides for the naming task with odd picture out, Praat for sentence judgment tasks).

[b] Tasks not present in the DuLIP

Repetition tasks: phonology

The repetition task assesses the phonological input- and output routes (De Witte et al., 2015) as well as articulation and verbal working memory. Participants are instructed to repeat each item of a list of 10 single existing nouns, compound-nouns, pseudo-words (i.e. words that do not exist in the language but that follow the phonotactic rules of that language) and sentences.

For each item, word length (three-syllable nouns, three- to six-syllable compound-nouns, two- and three-syllable pseudo-words), the absence or presence of consonant clusters, and phonological similarity (i.e. similar phonemes within a word) are used to increase the articulatory and phonological complexity (Gierut et al., 2007). The latter two were shown to increase the risks of errors (Nespoulous & Moureau, 1998; Shattuck-Hufnagel & Klatt, 1980). Words containing mutual phonological or semantic similarities were not presented consecutively to avoid the risk of perseveration. In Dutch, a variable syllable-stress pattern also contributes to increasing phonological complexity. This variable could not be adopted in French as stress is not lexically specified (Dell & Vergnaud, 1984). Imageability values are given for the simple words.

Pseudo-words and compounds are an addition to the DuLIP as it has been shown that they are processed differently to existing monomorphemic words and therefore could be problematic for patients depending on their brain lesion. Romance languages, although they differ from Germanic languages in their way of compounding, showed similar results (Mondini, Luzzatti, Zonca, Pistarini & Semenza, 2004; Semenza, Luzzeratti & Carabelli, 1997).

In the sentence repetition, some items have been adopted from the French version of the DuLIP but the majority was created to present participants with stimuli increasing in complexity. Sentences contain phonological similarities (e.g. *Il regarde les éléphants.* – He is looking at the elephants.), alliterations and consonant clusters (e.g. *Maman crie que les crêpes sont prêtes.* – Mum shouts that the pancakes are ready.), in addition to having an increasing amount of words (from four to ten). Verbal tense was also controlled.

Naming with Semantic odd picture out

Naming is a common task in language assessments. By combining the naming task with the semantic odd picture out, this test targets verbal semantic judgment and processing to assess language in a complex task (Satoer et al., 2015). Involved are: semantic knowledge and processing, divided attention, lexical access and inhibition. Participants are presented with ten combinations of three pictures belonging to closely related semantic categories, two of which belong to the same category (e.g. snake, dog, cat; all three of them are animals but the snake is not a pet). The semantic categories were made closer than in the DuLIP. The frequency of the items and the random order of the target item are checked. All images are black and white pictures (Snodgrass & Vanderwart, 1980). Reaction time is limited to 4 seconds (automatically timed slides on PowerPoint) and accuracy is registered.

Sentence completion in closed and broad contexts: semantics and syntax

As in the DuLIP, the sentence completion task measures ‘spontaneous speech in context’, which includes phonological processing, language dynamics, and production of semantically and syntactically appropriate speech (De Witte et al., 2015). It does so through the “production of parts of speech elicited by a specific context” (closed context) and “speech production within a sentence frame with obligatory and less obligatory parts of speech” (broad context) (De Witte et al., 2015, p.38-39). The items are here presented verbally by the instructor rather than visually.

The ten sentences are taken from the French version of the DuLIP and include three semantically induced instances (i.e. where a word has to be added) and seven less semantically induced instances (i.e. where a finite, and thus grammatically correct, phrase has to be added). Participants are instructed to complete the sentences spontaneously in a meaningful way. Sentence structure, transitivity and verbal tense are checked.

Sentence judgement tasks: phonology, semantics and syntax

This task assesses the phonological, semantic and morphosyntactic decoding, awareness and judgment occurring via the visual input route (unlike in the DuLIP where the stimuli is presented verbally) and in a complex attentional setting (participants have to use the laptop keyboard to indicate their answers). Participants are asked to discriminate between correct and incorrect sentences on a total of thirty sentences. Erroneous sentences include five instances of each category (phonological, semantic and syntactic): phonological errors – incorrect sentences contain non-words (e.g. *Le ramion va a Paro* – The ramion goes to Paro); semantic errors – incorrect sentences contain violations of semantic selection restrictions (e.g. *Le passager but une carte.* – The passenger drank a card); syntactic errors – incorrect sentences include verb form errors (e.g. *Il lui a demandé si elle a été froid.* – He asked her whether she had cold.), incorrect pronouns (e.g. *Ce n’est pas le vôtre, c’est le nous.* – That’s not yours, it’s we.), and incorrect prepositions (e.g. *Il aide sa petite soeur de manger.* – He is helping his little sister of eat). All sentences are active sentences in the declarative modality. The variables tense, sentence length and transitivity were checked in all tasks in addition to phonological complexity (phonological judgment) and error type (syntactic judgment). Reaction time and accuracy rates are measured.

6.2. Procedure

The procedure was identical for all participants. The battery is administered by the experimenter (i.e. the author of this thesis) in a quiet environment (private setting). The tasks are presented in an identical order and practice trials precede each task to ensure participants have understood the instructions. No repetition (for auditory stimuli), long hesitation or self-correction is allowed. All answers are transcribed and if necessary audio-recorded to be transcribed in orthographic script after the session. One point is awarded for each correct answer. Visual stimuli for the naming task with odd picture out is presented via PowerPoint slides on a laptop. Participants had to answer within 4 seconds to the naming task with odd picture put and the first answer is registered. The visual stimuli for the sentence judgement task is coded in Praat (Boersma & Weenink, 2019) and presented on a laptop. Participants have to judge the sentences by clicking on the keys F (*faux* – incorrect) or J (*juste* – correct) and their reaction times are recorded automatically. Those labels were selected because they are central and on the same row on the keyboard. The total duration of the session is approximately fifteen minutes. Uncertainties about the correctness of certain answers were discussed between experimenter and supervisor afterwards and qualitative terms were agreed. All data were qualitatively and quantitatively analyzed.

6.3. Subjects

A total of 67 native French-speaking adult participants took part in this preliminary standardization, recruited from diverse socio-cultural backgrounds and across a broad age range to cover most of the population, in Wallonia (Belgium), in Paris (France) and from the French speaking expats community in Utrecht (The Netherlands). They were divided by nationality (24 French people and 43 Belgians), handedness (right, left, ambidextrous), gender (male, female), age groups (18-54 and above 55 years old) and education groups (≤12 and >12 years). The education groups were divided as so because pursuing education after 12 years implies a higher education cycle (and having obtained the BAC for participants who were educated in the French educational system), which is a cut-off often used in normative studies (Merck et al., 2011; Miatton, Wolters, Lannoo & Vingerhoets, 2004). This cut-off is the same in the original Dutch DIMA, and so is the age cut-off (age at which cognitive decline may begin (Rönnlund, Nyberg, Backman, & Nilsson, 2005)). Participants were volunteers; they were informed of the goals of the study and gave a written informed consent. This research has been approved by the *Ethische Toetsingscommissie Linguïstiek* (Ethical Committee for Linguistics) of the Utrecht Institute of Linguistics OTS, University Utrecht.

All subjects taking part in the standardization fulfilled inclusion criteria which consisted of the following: (1) native speaker of French, (2) no (history of) cardiovascular, neurological, psychiatric developmental language and/or speech disorders, (3) normal (or corrected) hearing and (4) vision, (5) no toxic substance abuse (drugs/alcohol), (6) no excessive use of sleep medication and (7) no use of psychofarmaca (6-7 are medication known to affect cognitive functioning). Table 3 summarizes the demographic data of both nationality groups.

Table 3: Demographic characteristics of the 67 healthy participants.

BELGIAN PARTICIPANTS (n = 43)				
Demographics		***Mean (median)***	***SD***	***Range***
Age		48 (49)	20,81	19-80
Education in years		15 (15)	2,80	8-21
Groups		***Mean***	***Number of subjects***	***Percentage (%)***
Age	18-54 y	32,8	25	58,14
	>55 y	69,2	18	41,86
Education level	≤ 12 y	11,69	13	30,23
	>12 y	16,47	30	69,77
Gender	F		29	67,44
	M		14	32,56
Handedness	R		38	88,37
	A		3	6,98
	L		2	4,65
FRENCH PARTICIPANTS (n = 24)				
Demographics		***Mean (median)***	***SD***	***Range***
Age		35,4 (35,5)	12,99	21-69
Education in years		16,9 (17)	2,64	11-21
Groups		***Mean***	***Number of subjects***	***Percentage (%)***
Age	18-54 y	32,4	22	91,67
	>55 y	68,5	2	8,33
Education level	≤ 12 y	11,67	3	12,50
	>12 y	17,62	21	87,50
Gender	F		19	79,17
	M		5	20,83
Handedness	R		22	91,67
	A		2	8,33
	L		-	-

Legend:

SD = Standard Deviation

y = years, F = female, M = Male, R = right-handed, A = ambidextrous (Laterality Quotient ≥ -40 and ≤ +40), L = left-handed

6.4. Statistics

The data obtained from the 67 participants is analyzed in R (R Core Team, 2017). First, an analysis of internal reliability was carried out. Assumptions of normality were checked using the Shapiro-Wilk's normality test. Items selected for the final version are those with the highest accuracy rates above 80%. For tied scores, selection was based on the coefficient of ordinal reliability alpha since the response scale was ordinal with only two response options (right, wrong) and showed skewness (Gadermann, Ghun & Zumbo, 2012). The ordinal alphas for each task were computed using the R package *psych* (Revelle, 2018).

Second, a linear regression analysis of the relationship between task scores and demographic variables was performed using the R package *lme4* (Bates, Maechler, Bolker & Walker, 2015). The dependent variable is the number of correct answers per task per participant, or the reaction time, both treated as linear variables. For the analyses, the same binary group divisions were applied to the data as in the original DIMA: two nationality groups were compared (Belgians, French people), two gender groups (Female, Male), two education groups (≤ 12 years of education, >12 years of education), and two age groups (18-54 y, >55 y). Those cut-offs allow comparison with the original study. In addition, we analyzed the effects of early bilingualism (i.e. learned more than one language before the age of 3 years old or not) and late bilingualism (i.e. daily use of a foreign language learnt after 3 years old or not) (Sullivan, Poarch & Bialystok, 2018). This addition was made to test for potential effects of linguistic interference because some participants were recruited from a multilingual environment. Factorial ANOVAs were used to examine the main effects of the six previously mentioned groupings on the test scores, as well as the two-way interaction effects. P-values were obtained by likelihood ratio tests of the full model with the effect in question against the model without the effect in question, and the best fitting model was selected.

Finally, descriptive statistics are computed per subgroup defined on the above mentioned background variables (age and education) and for the whole group. Per linguistic task, mean, SD, median and range are calculated. This prepares for the calculation of cut-off values to be used in the clinical practice.

7. Results

7.1. Items analysis

At least ten items in each task (and five per sentence judgment category) from the pilot version had accuracy rates above the 80% threshold, and most of them had accuracy rates above 90%. The tasks which had not yet been standardized in the DuLIP contained fifteen items, namely the repetition of compounds, pseudo-words, and sentences, and the semantic odd picture out (which contained sixteen items). Besides, the syntactic judgment task contained six items: four of them were used in the DuLIP already so two were added to choose from. In the words and compounds repetition tasks, the sentence completion task, and the semantic and phonological sentence judgment tasks, all items were above 80%, in the pseudo-words repetition and sentence repetition tasks, twelve were above 80%, in the semantic odd picture out and syntactic judgment tasks, all items but one were above the threshold. The items included in the final version were those registering the highest accuracy rates above the 80% threshold. A list of all items tested in the pilot version is presented in Table 7 in Appendix A.

Internal consistency for all tasks in the final version is relatively high as all items included registered above 80% accuracy. Except for one task (repetition of sentences: ordinal alpha = .62), the ordinal alphas for all tasks are above .70 and most of the tasks have alphas above .80 as illustrated in Table 4. Coefficients above .70 are accepted in research and above .80 are recommended for applied purposes (Nunnally, 1978). The repetition of compounds has no variance as all items reached the 100% accuracy rate except one, which reached 98,5%.

Table 4: Ordinal alpha for each task

Task	Ordinal alpha
Repetition of words (/10)	.86
Repetition of compound-words (/10)	no variance
Repetition of pseudo-words (/10)	.84
Repetition of sentences (/10)	.62
Naming with semantic odd word out (/10)	.87
Sentence completion task (/10)	.77
Sentence judgment: correct and incorrect items (/30)	.94
Syntactic sentence judgment (/5)	.74
Semantic sentence judgment (/5)	.89
Phonological sentence judgment (/5)	.87

7.2. Tasks analysis

With regards to the tasks, ceiling effects were observed for all subtests, with median values for the whole sample all being perfect scores (10/10 or 5/5). Overall, accuracy rate ranged from 99,85% (repetition of compounds) to 91,34% (semantic sentence judgment). This shows that the healthy population performed well, with the semantic sentence judgment being the most difficult task (quickly followed by the syntactic sentence judgment with 92,24%), although still above 90%. The distribution is therefore not normal. For the whole sample, means, medians and accuracy rates are presented in Table 5.

Table 5: Mean scores, standard deviations and accuracy rates per linguistic task for the whole group (n=67).

Task	Mean score	SD	Accuracy rate (%)
Phonology			
Repetition of words (/10)	9,93	± 0.32	99,25
Repetition of compounds (/10)	9,99	± 0.12	99,85
Repetition of pseudo-words (/10)	9,58	± 0.74	95,82
Repetition of sentences (/10)	9,64	± 0.60	96,36
Total	*39,14*	*± 1.20*	*97,84*
Semantics			
Naming with odd picture out (/10)	9,49	± 0.96	94,93
Syntax			
Sentence completion (/10)	9,46	± 0.68	94,63
Mixed sentence judgment task			
Phonology : accuracy (/5)	4,96	± 0.72	99,10
Phonology: mean Reaction Time	1,34623E+16		
Semantics : accuracy (/5)	4,57	± 0.96	91,34
Semantics: mean Reaction Time	1,39753E+16		
Syntax : accuracy (/5)	4,61	± 0.21	92,24
Syntax: mean Reaction Time	1,50209E+16		
Total: accuracy (/15)	*14,13*	± 1.31	*94,23*
Total: Reaction Time	*4,24584E+16*		

On a qualitative note, regarding the first subtests of the repetition task (words and compounds), the few errors were due to almost homophones that were mistaken for the target items (e.g. *dominant* instead of *domino*). Most failed answers in the semantic odd picture out were either due to incorrect lexical item retrieved, or were delayed answers. Some additional errors were due to self-correction since only the first answer was taken into account. Regarding the sentence completion, most incorrect answers occurred in the broad context, due to too long hesitations. With regards to the sentence judgment task, the subset semantics obtained the lowest accuracy rate. Some participants reported being confused about whether only syntactic violations were incorrect or whether semantic oddities were to be taken into account as well (besides phonological ones more easily targetable), and that the ‘poetic’ aspect of some semantically incorrect sentences induced them to respond to the sentence as correct.

7.3. Demographic analysis

Effects of the demographic variables were obtained from the linear regression analysis. We constructed linear models of score as a function of respectively nationality, age, education, gender, early bilingualism and late bilingualism, and looked at the interactions, which are shown in Table 6. It appeared that age has a significant main effect on most of the tasks scores (repetition of words: $p<.05$); pseudo-words, and sentences, and sentence completion: $p<.01$; phonology total and semantic odd picture out: $p<.001$) but not on the repetition of compounds, the sentence judgment task (as well as each subset independently), and on reaction times. Older participants registered lower scores than the younger ones. Education also impacted significantly the scores in the total of the phonology tasks ($p<.05$) and in the syntactic sentence judgment task ($p<.01$), with lower scores for the lower educated group. It significantly improved the model in the pseudo-word repetition task although it did not have significance on its own. A main effect of gender ($p<.05$) was observed in the non-word repetition task, where men registered lower scores. In addition, an interaction of age and education was observed in the repetition of pseudo-words tasks ($p<.05$), leading to better performance for older and higher educated participants. No effect of nationality nor early bilingualism was found to be significant in any of the subtasks. The effect of laterality was not considered in the models because of the too low number of participants in the left-handed and ambidextrous groups.

A linear mixed model was performed on the whole test, including all subtasks, with score and the demographic variables as fixed effects. As random effects, we had intercepts for subjects and tasks. In all tasks combined, no effect of any of the factors was observed.

Table 6: Results of the factorial ANOVA. For each task and tasks-group, F-values and degrees of freedom of the best fitting model are presented and significance levels are indicated for the relevant factors in main effects and interactions.

			Main effects			Interactions
Test	**df**	**F-value**	**Age**	**Education**	**Gender**	**Age * Education**
Word repetition	1,65	4.714	*			
Compound repetition	66					
Pseudo-word repetition	4,62	9.595	**	-	*	*
Sentence repetition	1,64[1]	8.547	**			
Phonology: total	2,264	13.96	***	*		
Semantic odd picture out	1,65	15.62	***			
Sentence completion	1,65	9.06	**			
SJ Syntax: Accuracy	1,65	8.171		**		
SJ Syntax: RT	66					
SJ Semantics: Accuracy	66					
SJ Semantics: RT	66					
SJ Phonology: Accuracy	66					
SJ Phonology: RT	66					
SJ Total: Accuracy	66					

Legend: df = degrees of freedom; SJ = sentence judgment task; RT = reaction time;

- ➔ part of the model although not significant, * ➔ p<.05, ** p <.01; *** ➔ p <.001

Descriptive statistics are computed for the whole group and for age and education groups, and are provided in Appendix C.

8. Discussion

The DIMA was adapted into French based on the relevant (psycho-)linguistic variables. The normative data was standardized on 67 native French-speaking participants (24 French people and 43 Belgians), and the results of this pilot study are discussed below. They show that the French version of the DIMA can be standardized on a wider healthy population as well as on patients. Expectations with regards to the clinical application are discussed in the following sections and limitations and future directions are mentioned.

8.1. Normative data

8.1.1. Items

The French DIMA globally has a high internal validity since all items registered accuracy rates above the threshold of 80% accuracy. As there was no test-retest validation possible, ordinal alpha scores for the tasks were computed. They confirm a good internal validity as most tasks have alpha

[1] One participant did not take part in the sentence repetition task.

coefficients above .80 which is recommended for applied purposes, and two of them have alphas above .70 (the sentence completion task and syntactic sentence judgment), which is the threshold for research purposes (Nunnally, 1978). Only one task has an alpha between .60 and .70 (sentence repetition task). The coefficient depends on the number of items and on co-variance between them. In this battery, not all tasks are built the same way: some have several items of the same level of difficulty for a variable (e.g. repetition of words, compounds and pseudo-words, and naming with semantic odd picture out), while other tasks include items gradually increasing in difficulty (e.g. repetition of sentences, and to a lesser extent sentence completion). That the sentence repetition task had a lower alpha could be due to the fact that it contains sentences ranging from five to nine words, with different structural and articulatory levels of difficulty, creating less co-variance. Similarly, the sentence completion task contains closed and open contexts and, in the latter, different levels of automaticity. A suggestion could be to add items to create more co-variance between them as it would make the gradual difficulty smoother. With regards to the sentence judgment task, it could be hypothesized that the lower coefficient value for the syntactic judgment could be due to the more varied type of syntactic violations (prepositions, pronouns, verbs) compared to the phonological and semantic ones.

From a qualitative point of view, the items that were kept in the final version correspond in terms of linguistic and psycholinguistic variables with the original Dutch study. This means that the proportion of successful items per level of difficulty is similar between both versions and that the variables under consideration are preserved. One adaptation may however be suggested with regards to the repetition of sentences. The level of difficulty for the last successful item in the list does not totally match with the Dutch version. The last item in Dutch is *'De Griek ontdekte vier nietjes in de band van zijn fiets'*[2]. It is an eleven-word sentence with a main verb in the past tense and its structure is subject-verb-object-complement where the complement contains an embedded prepositional group. Two matching French items were created, namely a) *'Le grec découvrit que quatre grosses agrafes perçaient son pneu.'*[3] (ten-words, past tense, embedded phrase) and *b) 'L'été dernier, mon ami est allé à pied en Italie.'*[4] (eleven-words, past tense, several complements), and had accuracy rates of 53% and 84% respectively. Although above the 80% threshold, the latter item was not selected because ten items had higher accuracy rates. The last item in the French list is therefore *'Stéphanie va chez le coiffeur trois fois par an.'*[5] (nine-words, present tense, subject-verb-object-complement, and 100% accuracy

[2] Translation: The Greek discovered four staples in the tire of his bike.
[3] Translation: The Greek discovered that four staples were piercing his tire.
[4] Translation: Last summer, my friend went on foot to Italy.
[5] Translation: Stéphanie goes to the hairdresser three times a year.

rate). It could thus be relevant to replace this item by b) *'L'été dernier […]'* to have a more accurate increase in difficulty in the French list of items for this task.

8.1.2. Tasks and demographic factors

In terms of mean accuracy rates, ceiling effects (i.e. above 90%) for all tasks indicate that the healthy population performed well. In general, responses counted as errors were due to the strict correction criteria, such as hesitations exceeding 4 seconds (mainly in the semantic odd picture out and sentence completion tasks). The semantic and syntactic judgment tasks were also apparently more cognitively demanding for healthy participants. This was especially caused by the confusion arising from mixing several types of accuracy violations within the same task, which is reflected in the lower accuracy rates.

With regards to the demographics, the results of the ANOVAs did not show any effect of nationality on the results. We can thus conclude that the items included in the French version of the DIMA are neutral in terms of cultural and linguistic norms. This means that both French people and Belgians can be included in the next steps of the standardization of the French DIMA and that no separate norms are necessary in the future when testing Belgian or French patients with this instrument.

As expected, the best predictor of the results was age, which significantly affected accuracy performance in all tasks except for the compounds repetition and the sentence judgment task (any of the subtests). This effect was also found in the original DIMA (Satoer et al., 2019) and in the DuLIP (in Dutch: De Witte et al, 2015; in French: Quinchon et al., 2019; Quinchon & Tahiri, 2019), as well as in other studies (Lezak, Howieson, Bigler & Tranel, 2012). So, older participants had lower scores on average than younger participants. Education level, however, was less of a recurrent predictor. Education had a significant effect in interaction with age on the repetition of pseudo-words, and significantly affected the total of the phonology tasks as well as the accuracy of the syntactic sentence judgment task. Participants with a higher education level scored higher on the syntactic judgment than participants with a lower education level. However, this is the only factor affecting any of the sentence judgment subtests, including reaction times. It can be hypothesized that participants with a lower education background were less confident in assuming that a sentence was syntactically incorrect compared to those with a higher education background. Additionally, gender affected accuracy in the pseudo-word repetition, with lower scores registered for men than for women. However, both gender and education effects had a significance level of $p<.05$, which could be due to imbalance in the groups (total, n = 67; males, n=19; education ≤12 y, n = 16).

Comparing the results with those of the original DIMA described in Section 3, it was unexpected that no effect of age was found on the reaction times, as longer RTs were found for older participants in the original DIMA. It could be that this result is partly due to participants being distracted by the content of the sentences and it should therefore be stressed that this task is timed. Nonetheless, the results should be replicated with more participants as they could also be due to the unequal participants groups. As concerns education levels, more effects were found in the Dutch study as well. For this factor, the number of participants with lower educational background was lower in the French study.

8.2. Clinical application

Given the successful accuracy rates of the healthy population, the French DIMA is applicable on patients for normalization. On the basis of the descriptive analysis of the normative data, in which means and standard deviations are provided, z-scores can be calculated. The cut-off scores commonly used in clinical practice are, for clinically and pathologically impaired, 1.5 and 2 standard deviations respectively (Palmer, Boone, Lesser & Wohl, 1998; Satoer et al., in progress), and can be computed for direct application. In addition, patients' results can be compared with their performance on 'classical' language tests commonly administered. The original DIMA, when administered to patients, appeared to be more sensitive than the most classical tests including the shortened version of the Token Test, the Boston Naming Test, verbal fluencies (category and letter) and the Trailmaking Test A&B. Only the verbal fluency tasks were impaired which is explained by the multidimensional background of the task, and the authors suggest therefore 'to administer a cognitive screening […] in order to disentangle non-verbal cognitive disturbances from language impairments' (Satoer et al., in progress, p. 13). For the French version, a comparison with similar tasks from classical tests is also recommended, such as with the DO80, the repetition of sentences of the ELEA (Batterie d'Évaluation du Langage Élaboré de l'Adulte), the BECS (Batterie d'Évaluation des Connaissances Sémantiques) and fluencies, as were used for the comparison of the French DuLIP (Quinchon & Tahiri, 2019).

Originally intended for patients of brain tumors, such as LGGs, the DIMA responds to the demand for short test batteries able to detect mild aphasias. Mild aphasia had indeed received limited consideration in the literature of the last century compared to severe or even moderate aphasia, resulting in patients often being underdiagnosed (Hickin, Mehta & Dipper, 2015). In addition, usual aphasia diagnostic tests, generally designed for aphasia following strokes, are not sensitive enough to detect the more subtle deficits, such as processing speed and attention or grammatically complex language (Le Rhun et al., 2009; Satoer, Vincent, Smits, Dirven & Visch-Brink, 2013). This lack of adequate instrument is also expressed in the literature for the French clinical application (Le Rhun et al., 2009). Its application is recommended by Satoer and colleagues (in progress) as an instrument to assess pre- and postoperative language in patients who have to undergo (awake) surgery. This means that on the basis of the results of the DIMA preoperatively, relevant tasks from the DuLIP can be selected to be administered intraoperatively. The authors also describe the administration of the DIMA to a meningioma patient one year after surgery, as cognitive impairments are also known to occur with this type of brain tumor. They also suggest a broader application for the DIMA to patients with other neurological diseases. Such cases could include Traumatic Brain Injury or strokes (e.g. Armstrong, Fox & Wilkinson, 2013; Hickin, Mehta & Dipper, 2015; Hunting-Pompon, Kendall & Bacon More, 2011), but also incipient dementia such as Primary Progressive Aphasia in its early phase when symptoms are mild. It has been reported, for instance, that polysyllabic items, non-words and longer sentences were necessary for PPA early diagnoses, although lacking from the Western Aphasia Battery (Pressman & Gorno-Tempini, 2016).

8.3. Limitations

There are some limitations to this adaptation. Considering this work was carried out as a Master thesis, a constraint is that all the answers were rated by a single experimenter. It is however advised in the literature to perform reliability tests such as inter-raters reliability. Therefore, hesitations were discussed with the first supervisor and indications were applied consistently on the similar cases. Moreover, all answers were recorded and could be heard several times.

Next, some groups within the healthy population sample are underrepresented, such as the lower education group, but also male participants. Although special attention was paid to recruit participants from varied socio-economic statuses and demographic backgrounds, some groups remained unbalanced and the results must therefore be interpreted carefully.

Besides, a limitation was already mentioned regarding the sentence repetition task. This task registered a low ordinal alpha (ordinal alpha <.70). Additionally, we suggested an adaptation in order to keep the level of difficulty closer to the original Dutch battery. However, all items met the 80% accuracy threshold and the global level of difficulty in each task is preserved.

Finally, the reliability of the timing in the sentence judgment task could possibly be increased. Since test administration was carried out in the home setting, it was possible that some distractions occurred. In addition, some participants liked to comment on the task while taking the test, possibly delaying some answers. Although participants were informed about the timed component of the task, extra attention should be paid to clearly informing participants before the start that they should try and answer the questions without being distracted.

8.4. Future directions

Because of the urgent need for a quick diagnostic instrument for mild language impairments, standardization should therefore be extended on a wider French-speaking healthy population. Since no effect of nationality was observed on performance between Belgium and France, we conclude that two specific norms are not needed. It would therefore be relevant to compare populations from other French speaking countries/regions that are culturally similar, such as Switzerland and Quebec. This way, more patients can be included in studies about anatomo-clinical correlations, thus improving our understanding of the brain. Validation on patients of LGG and different neurological diseases can also be started.

The DIMA covers verbal language production and perception and requires intact reading. However, depending on lesion location, different impairments can be present. Hence we present several possible additions. First of all, regarding the semantic odd picture out, a more detailed grading system could be implemented to obtain a more detailed profile of the patients, as can be found in the recent literature (Wilson, Eriksson, Schneck & Lucanie, 2018). It could indeed be possible to split the score into two halves since two main components are tested: the picture selection and the picture naming tasks.

Then, verbs and nouns are stored in different places and can therefore be impaired differently (e.g. Broca's aphasia). Specifically, verbs appear to be more difficult since arguments and theta roles need to be visualized (Bastiaanse & van Zonneveld, 2004). A version of the semantic odd picture out task with actions was proposed in the original DIMA, although the accuracy rate of the task was below 80% hence it should be revised (Satoer et al., in progress). Once this task is adapted, it can also be included in the French version and be standardized.

Next, it would be a relevant addition to test participants on their writing and reading skills with sensitive tasks for mild impairments (e.g. text comprehension, summarizing). Writing impairments in patients with LGG before and after surgery have indeed been registered (Antonsson et al., 2018; van Ierschot et al., 2016). In addition, it also appears that mild aphasia impacts spontaneous speech (Satoer et al., 2013; Satoer et al., 2018), so an evaluation of spontaneous speech should be added to the protocol when administered to patients.

Finally, since the DIMA can be used to assess language on a long term basis, it is very likely that patients have to be tested several times with the same items. To prevent an effect of test-retest, it would be useful to develop two parallel versions of the French DIMA. As there are some remaining items that met the 80% accuracy rate threshold, they could be used as a starting point for making a second version.

9. Conclusion

In this study, a French adaptation of the DIMA (Satoer et al., 2019) was developed. Initially, the DIMA was developed in Dutch and is derived from the DuLIP (De Witte et al., 2015). It captures language deficits in production and perception in the most important linguistic domains (i.e. phonology, semantics and (morpho-)syntax). Its use is illustrated as a pre- and postoperative assessment of the language performance of patients with suspected subtle language deficits due to brain tumor. As a shorter version of the DuLIP, the DIMA targets patients of LGGs and mainly serves as a selection instrument for intraoperative tasks from the DuLIP in the awake surgery conditions for tumor resection. At the same time, the battery answers the demand in the literature for a short diagnostic instrument for mild aphasia as increasing attention is being brought to the quality of life of patients with mild language deficits. It is suggested that the DIMA can also be applied to other types of tumors causing mild language impairments such as meningioma, but also to other types of neurological etiologies that can cause mild aphasia and, so, possibly become part of a standard language assessment.

The purpose of this research was to adapt the Dutch battery in terms of linguistic and psycholinguistic variables into French. In addition, the goal was to compare two healthy French-speaking populations, namely Belgians and French people, as well as to explore potential effects of demographic variables such as age, gender, level of education (and early and late bilingualism). The analysis was carried out on 67 participants. The results showed that the proposed final version of the French DIMA is a valid instrument that can be normalized on patients and standardized on a larger scale. Ceiling results in accuracy rates were registered for all tasks. Although no effect of nationality was found, there was an effect of age for most of the tasks, and education and gender also influenced performance in some tasks. Those results are in line with other findings in the literature. Finally, further directions are suggested regarding the standardization and possible additions to the DIMA. To conclude, we can consider that the final version of the French DIMA as presented here is a valid instrument and that standardization and clinical validation can be pursued.

References

Antonsson, M. (2017). *Language ability in patients with low-grade glioma-detecting signs of subtle dysfunction* (Doctoral dissertation). Retrieved from: https://gupea.ub.gu.se/handle/2077/53612

Antonsson, M., Johansson, C., Hartelius, L., Henriksson, I., Longoni, F., & Wengelin, Å. (2018). Writing fluency in patients with low-grade glioma before and after surgery. *International journal of language & communication disorders*, *53*(3), 592-604.

Antonsson, M., Longoni, F., Jakola, A., Tisell, M., Thordstein, M., & Hartelius, L. (2018). Pre-operative language ability in patients with presumed low-grade glioma. *Journal of neuro-oncology*, *137*(1), 93-102.

Armstrong, E., Fox, S., & Wilkinson, R. (2013). Mild aphasia: is this the place for an argument?. *American Journal of Speech-Language Pathology*.

Bal, W., Doppagne, A., & Goosse, A. (1994). *Belgicismes. Inventaire des particularités lexicales du français en Belgique*. Louvain-la-Neuve: Duculot.

Bastiaanse, R., & Van Zonneveld, R. (2004). Broca's aphasia, verbs and the mental lexicon. *Brain and Language*, *90*(1-3), 198-202.

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

Bézy, C., Renard, A., & Pariente, J. (2016). *GRÉMOTS: évaluation du langage dans les pathologies neurodégénératives*. De Boeck Superieur.

Boersma, P., & Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.53, retrieved 26 May 2019 from http://www.praat.org/

Bonin, P., Méot, A., Aubert, L. F., Malardier, N., Niedenthal, P., & Capelle-Toczek, M. C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *L'année Psychologique*, *103*(4), 655-694.

Brucki, S. M. D., & Rocha, M. S. G. (2004). Category fluency test: effects of age, gender and education on total scores, clustering and switching in Brazilian Portuguese-speaking subjects. *Brazilian journal of medical and biological research*, *37*(12), 1771-1777.

Catani, M., Dell'Acqua, F., Bizzi, A., Forkel, S. J., Williams, S. C., Simmons, A., ... & de Schotten, M. T. (2012). Beyond cortical localization in clinico-anatomical correlation. *Cortex*, *48*(10), 1262-1287.

Catani, M., Howard, R. J., Pajevic, S., & Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage*, *17*(1), 77-94.

Cruice, M., Worral, L. & Flickson L. (2006). Perspectives of quality of life by people with aphasia and their family: Suggestions for successful living. *Topics in Stroke Rehabilitation 13*, 14-24.

DeAngelis, L. M. (2001). Brain tumors. *New England Journal of Medicine*, *344*(2), 114-123.

Delcourt, C. (1998-1999). *Dictionnaire du français de Belgique* (Vols. 1-2). Brussels: Le Cri Édition.

Dell, F., & Vergnaud, J.-R. (1984). Les développements récents en phonologie : quelques idées centrales. In F. Dell, D. Hirst, & J.- R. Vergnaud (Eds.), *Forme sonore du langage* (pp. 1–42). Paris: Hermann

Desmurget, M., Bonnetblanc, F., & Duffau, H. (2007). Contrasting acute and slow-growing lesions: a new door to brain plasticity. *Brain*, *130*(4), 898-914.

De Witt Hamer, P., C., Robles, S. G., Zwinderman, A. H., Duffau, H., & Berger, M. S. (2012). Impact of intraoperative stimulation brain mapping on glioma surgery outcome: a meta-analysis. *J Clin Oncol*, *30*(20), 2559-2565.

De Witte, E., & Mariën, P. (2013). The neurolinguistic approach to awake surgery reviewed. *Clinical neurology and neurosurgery*, *115*(2), 127-145.

De Witte, E., Satoer, D., Robert, E., Colle, H., Verheyen, S., Visch-Brink, E., & Mariën, P. (2015). The Dutch Linguistic Intraoperative Protocol: A valid linguistic approach to awake brain surgery. *Brain and Language*, *140*, 35-48.

De Witte, E., Satoer, D., Visch-Brink, E., & Mariën, P. (2016). Taaltests voor, tijdens en na wakkere hersenchirurgie: de rol van DuLIP. *Neuropraxis*, *20*(3), 83-90.

Duffau, H., & Capelle, L. (2004). Preferential brain locations of low-grade gliomas: Comparison with glioblastomas and review of hypothesis. *Cancer*, *100*(12), 2622-2626.

Duffau, H., Capelle, L., Denvil, D., Sichez, N., Gatignol, P., Taillandier, L., ... & Bitar, A. (2003). Usefulness of intraoperative electrical subcortical mapping during surgery for low-grade gliomas located within eloquent brain regions: functional results in a consecutive series of 103 patients. *Journal of neurosurgery*, *98*(4), 764-778.

Duffau, H. (2014). Diffuse low-grade gliomas and neuroplasticity. *Diagnostic and interventional imaging*, *95*(10), 945-955.

Eiesland, E. A., & Lind, M. (2012). Compound nouns in spoken language production by speakers with aphasia compared to neurologically healthy speakers: An exploratory study. *Clinical linguistics & phonetics*, *26*(3), 232-254.

Francard, M., Geron, G., Wilmet, R. & Wirth, A. (2010). *Dictionnaire des belgicismes.* Brussels: Éditions De Boeck Duculot.

Fyndanis, V., Lind, M., Varlokosta, S., Kambanaros, M., Soroli, E., Ceder, K., ... & Gavarró, A. (2017). Cross-linguistic adaptations of the Comprehensive Aphasia test: challenges and solutions. *Clinical linguistics & phonetics*, *31*(7-9), 697-710.

Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: a conceptual, empirical, and practical guide. Practical Assessment, Research & Evaluation, 17(3).

Gierut, J. A. (2007). Phonological complexity and language learnability. *American Journal of Speech-Language Pathology*, *16*(1), 6-17.

Goosse, A. (2011). *Façons belges de parler.* Brussels: Académie royale de langue et de littérature françaises and Le Cri.

Hendriks, E. J., Habets, E. J., Taphoorn, M. J., Douw, L., Zwinderman, A. H., Vandertop, W. P., ... & De Witt Hamer, P. C. (2018). Linking late cognitive outcome with glioma surgery location using resection cavity maps. *Human brain mapping*, *39*(5), 2064-2074.

Hickin, J., Mehta, B., & Dipper, L. (2015). To the sentence and beyond: A single case therapy report for mild aphasia. *Aphasiology*, *29*(9), 1038-1061.

Hunting-Pompon, R., Kendall, D., & Bacon Moore, A. (2011). Examining attention and cognitive processing in participants with self-reported mild anomia. *Aphasiology*, *25*(6-7), 800-812.

Ierschot, F. van, Miozzo, A., Santini, B., Spena, G., Talacchi, A., & Miceli, G. (2016, September). *Language preservation in brain tumor patients undergoing awake surgery: Does monitoring object naming suffice to spare other language skills?.* Conference paper: 17th International Science of Aphasia Conference.

Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test, ed 2.* Philadelphia: Lippincott, Williams & Wilkins.

Le Rhun, E., Delbeuck, X., Devos, P., Pasquier, F., & Dubois, F. (2009). Troubles cognitifs dans les gliomes de grade II et III de l'adulte: à propos d'une série de 15 patients. *Neurochirurgie*, *55*(3), 303-308.

Lezak, M., Howieson, D., Bigler, D., & Traner, D. (2012). *Neuropsychological assessment*. Oxford University Press, USA.

Lorenz, A., Heide, J., & Burchert, F. (2014). Compound naming in aphasia: Effects of complexity, part of speech, and semantic transparency. *Language, Cognition and Neuroscience*, *29*(1), 88-106.

Luria, A. R. (2012). *Higher cortical functions in man*. Springer Science & Business Media.

Mai, T. T., Grob-Nicolas, I., & Muchembled, A. (2012). *La Batterie d'Évaluation des Troubles Lexicaux: Normalisation de la version écrite de la BETL.* (Unpublished master thesis). Retrieved from : http://bu.univ-lille2.fr/fileadmin/user_upload/memoires_ortho/2012/LIL2_SMOR_2012_070.pdf

Mandonnet, E., Delattre, J. Y., Tanguy, M. L., Swanson, K. R., Carpentier, A. F., Duffau, H., ... & Capelle, L. (2003). Continuous growth of mean tumor diameter in a subset of grade II gliomas. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *53*(4), 524-528.

Massion, F. (1987). *Dictionnaire de* belgicismes (Vols. 1-2). Frankfurt am Main: Lang.

Merck, C., Charnallet, A., Auriacombe, S., Belliard, S., Hahn-Barma, V., Kremin, H., ... & Roussel, M. (2011). La batterie d'évaluation des connaissances sémantiques du GRECO (BECS-GRECO): validation et données normatives. *Revue de neuropsychologie*, *3*(4), 235-255.

Mondini, S., Luzzatti, C., Zonca, G., Pistarini, C., & Semenza, C. (2004). The mental representation of Verb–Noun compounds in Italian: Evidence from a multiple single-case study in aphasia. *Brain and Language*, *90*(1-3), 470-477.

Moritz-Gasser, S., Herbet, G., Maldonado, I. L., & Duffau, H. (2012). Lexical access speed is significantly correlated with the return to professional activities after awake surgery for low-grade gliomas. *Journal of neuro-oncology*, *107*(3), 633-641.

Nespoulous, J. L., & Moreau, N. (1998). Repair strategies and the production of segmental errors in aphasia: Epentheses vs. syncopes in consonantal clusters. *Linguistic levels in aphasiology*, 133-145.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™//A lexical database for contemporary french: LEXIQUE™. *L'année Psychologique*, *101*(3), 447-462.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Ojemann, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *Behavioral and Brain Sciences*, *6*(2), 189-206.

Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1998). Base rates of “impaired” neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology*, *13*(6), 503-511.

Papagno, C., Casarotti, A., Comi, A., Gallucci, M., Riva, M., & Bello, L. (2012). Measuring clinical outcomes in neuro-oncology. A battery to evaluate low-grade gliomas (LGG). *Journal of neuro-oncology*, *108*(2), 269-275.

Papagno, C. (2017). Studying cognitive functions by means of direct electrical stimulation: a review. *Neurological Sciences*, *38*(12), 2079-2087.

Partz de Courtray, M. P. de, Bilocq, V., De Wilde, V., Seron, X., & Pillon, A. (2001). *Lexis. Tests pour le diagnostic des troubles lexicaux chez le patient aphasique.* De Boek Superieur.

Penfield, W., & Roberts, L. (2014). *Speech and brain mechanisms* (Vol. 62). Princeton University Press.

Pressman, P., & Gorno-Tempini, M.L. (2016) The logopenic variant of primary progressive aphasia. In Miller, B. L., & Boeve, B. F. (Eds.). (2016). *The behavioral neurology of dementia*. Cambridge University Press.

Prins, R., & Bastiaanse, R. (2004). Analyzing the spontaneous speech of aphasic speakers. *Aphasiology*.

Quinchon, C., & Tahiri., E. (2019). *Évaluation du langage oral en chirurgie éveillée : adaptation d’une batterie et étude préliminaire.* (Unpublished master’s thesis). Retrieved from: Sciences cognitives ffdumas-02168421f

Quinchon, C., Tahiri, E., Pierret, H., Barberis, M., Mandonnet, E., Satoer, D., & Poisson, I. (2019, June). *Oral language assessment in awake surgery: battery adaptation and preliminary study.* Unpublished abstract presented at the 14th European Low Grade Glioma Network Conference, London.

Revelle, W. (2018). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA. https://CRAN.R-project.org/package=psych Version = 1.8.12.

Rofes, A., Mandonnet, E., Godden, J., Baron, M. H., Colle, H., Darlix, A., ... & Lubrano, V. (2017). Survey on current cognitive practices within the European Low-Grade Glioma Network: towards a European assessment protocol. *Acta neurochirurgica*, *159*(7), 1167-1178.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L. G. (2005). Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. *Psychology and aging*, *20*(1), 3.

Roussel, M., & Godefroy, O. (2016). *La batterie GRECOGVASC: Evaluation et diagnostic des troubles neurocognitifs vasculaires avec ou sans contexte d'accident vasculaire cérébral*. De Boeck Superieur.

Rousset, J., & Gatignol, P. (2014). Intérêt d'un nouvel étalonnage de tests: réflexion et mise en pratique autour de la batterie de dénomination orale d'images DO80. *Revue Neurologique*, *170*, A210.

Santini B, Talacchi A, Squintani G, Casagrande F, Capasso R, Miceli G. (2012). Cognitive outcome after awake surgery for tumors in language areas. *J Neurooncol,* 108(2), 319-326.

Satoer, D., De Witte, E., Bastiaanse, R., Vincent, A., Mariën, P. & Visch-Brink, E. (2019). *Diagnostic Instrument for Mild Aphasia (DIMA): standardization and clinical application.* Conference Abstract: Academy of Aphasia 55th Annual Meeting. Front. Hum. Neurosci. doi: 10.3389/conf.fnhum.2017.223.00103

Satoer, D., Vincent, A., Smits, M., Dirven, C., & Visch-Brink, E. (2013). Spontaneous speech of patients with gliomas in eloquent areas before and early after surgery. *Acta neurochirurgica*, *155*(4), 685-692.

Satoer, D., Vincent, A., Ruhaak, L., Smits, M., Dirven, C., & Visch-Brink, E. (2018). Spontaneous speech in patients with gliomas in eloquent areas: Evaluation until 1 year after surgery. *Clinical neurology and neurosurgery*, *167*, 112-116.

Satoer, D., Vork, J., Visch-Brink, E., Smits, M., Dirven, C., & Vincent, A. (2012). Cognitive functioning early after surgery of gliomas in eloquent areas. *Journal of neurosurgery*, *117*(5), 831-838.

Satoer, D., Visch-Brink, E., Smits, M., Kloet, A., Looman, C., Dirven, C., & Vincent, A. (2014). Long-term evaluation of cognition after glioma surgery in eloquent areas. *Journal of neuro-oncology*, *116*(1), 153-160.

Semenza, C., Luzzatti, C., & Carabelli, S. (1997). Morphological representation of compound nouns: A study on Italian aphasic patients. *Journal of Neurolinguistics*, *10*(1), 33-43.

Semenza, C., & Mondini, S. (2006). Neuropsychology of compound words. *The representation and processing of compound words*, 71-95.

Shattuck-Hufnagel, S., & Klatt, D. H. (1980). How single phoneme error data rule out two models of error generation. *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*, 35-46.

Snitz, B. E., Unverzagt, F. W., Chang, C. C. H., Vander Bilt, J., Gao, S., Saxton, J., ... & Ganguli, M. (2009). Effects of age, gender, education and race on two tests of language ability in community-based older adults. *International Psychogeriatrics*, *21*(6), 1051-1062.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, *6*(2), 174.

Sullivan, M. D., Poarch, G. J., & Bialystok, E. (2018). Why is lexical retrieval slower for bilinguals? Evidence from picture naming. *Bilingualism: Language and Cognition*, *21*(3), 479-488.

Talacchi, A., Santini, B., Casartelli, M., Monti, A., Capasso, R., & Miceli, G. (2013). Awake surgery between art and science. Part II: language and cognitive mapping. *Functional neurology*, *28*(3), 223.

Talacchi, A., Santini, B., Savazzi, S., & Gerosa, M. (2010). Cognitive effects of tumour and surgical treatment in glioma patients. *Journal of neuro-oncology*, *103*(3), 541-549.

Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PloS one*, *13*(2), e0192773.

Appendix

A. Items and accuracy rates

Table 7: Items and their accuracy rate. Grey items are not part of the final version.

Item (item number)	Accuracy in %
Repetition of words (/10)	
différence	98,5%
hérisson	100,0%
domino	94,0%
caméra	100,0%
instruction	100,0%
abricot	100,0%
cicatrice	100,0%
procédé	100,0%
criminel	100,0%
photographe	100,0%
Repetition of compound-words (/10)	
timbre-poste	100,0%
sous-marin	100,0%
auto-stoppeur	100,0%
porte-documents	100,0%
avant-propos	100,0%
rez-de-chaussée	100,0%
moissonneuse-batteuse	100,0%
micro-organisme	100,0%
arrière-grands-parents	100,0%
anti-inflammatoire	98,5%
presse-citron	98,5%
année-lumière	98,5%
sous-directrice	98,5%
gastro-entérite	98,5%
contre-proposition	97,0%
Repetition of pseudo-words (/10)	
odonan	97%
esponel	99%
mirérie	97%
ruchuto	94%
éjournette	100%
ipsoda	100%
notonsonnar	94%
citrupauphile	90%
lodipopel	91%
sagéléfique	97%
adamube	61,0%
torbignamban	52,0%
chèclamutine	87,0%
grinvrizanin	48,0%
cuicaupréteur	87,0%
Repetition of sentences (/10)	

Henri roule dans la rue	100%
Sa sœur a attendu une heure.	100%
Cette pluie arrose le jardin.	100%
Dans quelle lettre l'a-t-il lu?	100%
Nous sommes allés en Bretagne.	100%
Où avez-vous rangé les serviettes?	100%
Maman crie que les crêpes sont prêtes.	86%
Le mercredi j'ai cours d'escrime.	100%
Mon vieux frère veut une veste coupe-vent.	91%
Stéphanie va chez le coiffeur trois fois par an.	100%
Le chien chatouilleux a léché son chiot.	84%
Une énième sirène réveilla Ségolène en sursaut.	67%
La brise rendait l'air frais à l'ombre.	79%
Le grec découvrit que quatre grosses agrafes perçaient son pneu.	53%
L'été dernier mon ami est allé à pied en Italie	84%
Naming with semantic odd word out (/10)	
serpent	97%
chaussette	99%
vélo	91%
nounours	96%
poisson	96%
marteau	94%
moufle	94%
bateau	100%
trompette	93%
pantalon	91%
carotte	90%
oiseau	88%
pied	85%
pinceau	82%
bureau	73%
lit	87%
Sentence completion task (/10)	
Elle s'asseoit sur…	100%
La voiture est dans…	100%
J'écoute de…	96%
La fille a commencé…	97%
Le voisin pense que	94%
La femme sur la plage	99%
Chaque jour	90%
Il n'est pas venu parce que	93%
Quand l'arbre est tombé	81%
Il s'assure que	99%
Syntactic sentence judgment (/5)	
Il s'inquiète à quelque chose.	94%
Ce n'est pas le vôtre, c'est le nous.	96%
Il lui a demandé si elle a été froid.	96%
Léa chante une chanson hier.	93%
Il aide sa petite sœur de manger.	84%

Le boulanger vendaient de délicieuses tartelettes.	30%
Semantic sentence judgment (/5)	
L'usine riait de la voiture.	*94%*
Le chat a acheté un canapé.	*85%*
Le plombier répare l'arc-en-ciel.	*90%*
L'escargot pédalait vers les buissons.	*90%*
Les fleurs ont regardé leurs yeux.	*99%*
Phonological sentence judgment (/5)	
Le steil fait fondre la ningle.	99%
Le racancier achète une blace.	99%
Il marque vers Loudre.	100%
Le fril est trop four.	99%
La neau coule du tansan.	100%

B. R outputs of the factorial ANOVAs

Word repetition

```
> summary(model8)

Call:
lm(formula = score ~ age_group + late_bil, data = sub_repmots)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.60371 -0.05346 -0.05346  0.17141  0.18967 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.05346    0.05357 187.673  < 2e-16 ***
age_group2  -0.24313    0.08397  -2.896  0.00517 ** 
late_bil1   -0.20662    0.08668  -2.384  0.02012 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2978 on 64 degrees of freedom
Multiple R-squared:  0.1436,	Adjusted R-squared:  0.1169 
F-statistic: 5.368 on 2 and 64 DF,  p-value: 0.006996
```

Compound repetition

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_repmotsC)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.98507  0.01493  0.01493  0.01493  0.01493 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.98507    0.01493     669   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1222 on 66 degrees of freedom
```

Pseudo-word repetition

```
> summary(model5)

Call:
lm(formula = score ~ age_group * yoe_group + gender, data = sub_repNmots)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4555 -0.1137  0.1147  0.2281  1.2281 

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)             9.7383     0.2470  39.420  < 2e-16 ***
age_group2             -0.9664     0.3129  -3.088  0.00301 ** 
yoe_group2              0.1470     0.2632   0.559  0.57839    
genderM                -0.4298     0.1655  -2.596  0.01175 *  
age_group2:yoe_group2   0.9530     0.3763   2.533  0.01386 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6013 on 62 degrees of freedom
Multiple R-squared:  0.3823,	Adjusted R-squared:  0.3425 
F-statistic: 9.595 on 4 and 62 DF,  p-value: 4.188e-06
```

Sentence repetition

```
> summary(model2)

Call:
lm(formula = score ~ age_group, data = sub_repph)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.7660 -0.3158  0.2340  0.2340  0.6842 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.76596    0.08262 118.209  < 2e-16 ***
age_group2  -0.45017    0.15398  -2.924  0.00478 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5664 on 64 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.1178,	Adjusted R-squared:  0.104 
F-statistic: 8.547 on 1 and 64 DF,  p-value: 0.00478
```

Phonology: total

```
> summary(model3)

Call:
lm(formula = score ~ age_group + yoe_group, data = sub_reptot)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.90209  0.09791  0.09791  0.28928  0.53993 
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.71072    0.07877 123.273  < 2e-16 ***
age_group2  -0.25066    0.07412  -3.382  0.00083 ***
yoe_group2   0.19137    0.07968   2.402  0.01701 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5082 on 264 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.09563,	Adjusted R-squared:  0.08878 
F-statistic: 13.96 on 2 and 264 DF,  p-value: 1.728e-06
```

Semantic odd picture out

```
> summary(model2)

Call:
lm(formula = score ~ age_group, data = sub_opo)

Residuals:
   Min     1Q Median     3Q    Max 
-2.850  0.150  0.234  0.234  1.150 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   9.7660     0.1266  77.131  < 2e-16 ***
age_group2   -0.9160     0.2317  -3.952 0.000193 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.868 on 65 degrees of freedom
Multiple R-squared:  0.1938,	Adjusted R-squared:  0.1814 
F-statistic: 15.62 on 1 and 65 DF,  p-value: 0.0001933
```

Sentence completion

```
> summary(model2)

Call:
lm(formula = score ~ age_group, data = sub_complph)

Residuals:
   Min     1Q Median     3Q    Max 
-1.617 -0.617  0.383  0.383  0.900 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  9.61702    0.09385  102.47  < 2e-16 ***
age_group2  -0.51702    0.17177   -3.01  0.00372 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6434 on 65 degrees of freedom
Multiple R-squared:  0.1223,	Adjusted R-squared:  0.1088 
F-statistic:  9.06 on 1 and 65 DF,  p-value: 0.003716
```

Sentence judgment

Syntax: accuracy

```
> summary(model3)

Call:
lm(formula = score ~ yoe_group, data = sub_jugph_synt)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.7451 -0.1875  0.2549  0.2549  0.8125 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.1875     0.1702  24.605  < 2e-16 ***
yoe_group2    0.5576     0.1951   2.858  0.00571 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6808 on 65 degrees of freedom
Multiple R-squared:  0.1117,	Adjusted R-squared:  0.098 
F-statistic: 8.171 on 1 and 65 DF,  p-value: 0.005715
```

Syntax: RT

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_time_synt)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.073e+16 -3.156e+15  1.115e+14  3.397e+15  1.381e+16 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.502e+16  6.066e+14   24.76   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.966e+15 on 66 degrees of freedom
```

Semantics: accuracy

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_jugph_sem)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5672  0.4328  0.4328  0.4328  0.4328 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.5672     0.1169   39.06   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.957 on 66 degrees of freedom
```

Semantics: RT

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_time_sem)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.090e+16 -2.457e+15  3.073e+13  2.310e+15  8.424e+15 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.398e+16  5.016e+14   27.86   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.106e+15 on 66 degrees of freedom
```

Phonology: accuracy

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_jugph_phono)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.95522  0.04478  0.04478  0.04478  0.04478 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.95522    0.02546   194.7   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2084 on 66 degrees of freedom
```

Phonology: RT

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_time_phono)

Residuals:
       Min         1Q     Median         3Q        Max 
-8.534e+15 -3.118e+15 -2.599e+14  2.612e+15  1.152e+16 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.346e+16  5.256e+14   25.61   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.302e+15 on 66 degrees of freedom
```

Total: accuracy

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_jug_tot)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.1343 -0.1343  0.8657  0.8657  0.8657 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  14.1343     0.1604   88.11   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.313 on 66 degrees of freedom
```

Total: RT

```
> summary(model1)

Call:
lm(formula = score ~ 1, data = sub_time_tot)

Residuals:
       Min         1Q     Median         3Q        Max 
-1.949e+16 -6.622e+15  1.124e+15  5.683e+15  1.565e+16 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4.246e+16  1.038e+15   40.91   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.495e+15 on 66 degrees of freedom
```

C. Descriptive statistics tables

Table 8: Descriptive statistics for the repetition tasks, semantic odd picture out and sentence completion for the whole sample and for age and education groups.

Test	Statistics	Gr1 <55 ≤12y (n=6)	Gr1 <55 >12y (n=41)	Gr1 >55 ≤12y (n=10)	Gr1 <55 >12y (n=10)	Total (n=67)
Word repetition	Mean (±sd) Median Range	10 ± 0.00 10 10-10	9.98 ± 0.16 10 9-10	9.8 ± 0.42 10 9-10	9.8 ± 0.63 10 8-10	9,93 ± 0.32 10 8-10
Compound repetition	Mean (±sd) Median Range	10 ± 0.00 10 10-10	10 ± 0.00 10 10-10	9.9 ± 0.32 10 9-10	10 ± 0.00 10 10-10	9,99 ± 0.12 10 9-10
Pseudo-word repetition	Mean (±sd) Median Range	9.67 ± 0.52 10 9-10	9.78 ± 0.57 10 7-10	8.6 ± 0.97 8.5 7-10	9.7 ± 0.48 10 9-10	9,58 ± 0.74 10 7-10

Sentence repetition	Mean (±sd) Median Range	9.67 ± 0.52 10 9-10	9.78 ± 0.47 10 8-10	9.22 ± 0.67 9 8-10	9.4 ± 0.84 10 8-10	9,64 ± 0.60 10 8-10
Phonology: total	Mean (±sd) Median Range	39.33 ± 1.03 40 38-40	39.54 ± 0.81 40 36-40	36.6 ± 2.99 38 36-39	38.9 ± 1.20 39 37-40	39,14 ± 1.20 40 36-40
Semantic odd picture out	Mean (±sd) Median Range	9.33 ± 1.21 10 7-10	9.83 ± 0.44 10 8-10	8.8 ± 1.48 9.5 6-10	8.9 ± 1.20 9 7-10	9,49 ± 0.96 10 6-10
Sentence completion	Mean (±sd) Median Range	9.5 ± 0.85 10 8-10	9.63 ± 0.49 10 9-10	8.9 ± 0.88 9 8-10	9.3 ± 0.82 9.5 8-10	9,46 ± 0.68 10 8-10

Table 9: Descriptive statistics for the Sentence judgment task.

Test	**Statistics**	**Gr1 <55 ≤12y (n=6)**	**Gr1 <55 >12y (n=41)**	**Gr1 >55 ≤12y (n=10)**	**Gr1 <55 >12y (n=10)**	**Total (n=67)**
Syntax: Accuracy	Mean (±sd) Median Range	4 ± 1.1 4 2-5	4.73 ± 0.63 5 2-5	4.3 ± 0.82 4.5 3-5	4.8 ± 0.42 5 4-5	4,61 ± 0.72 5 2-5
Syntax: RT	Mean (±sd) Median Range					15021 ± 4965 15132 4290-28833
Semantics: Accuracy	Mean (±sd) Median Range					4,57 ± 0.96 5 1-5
Semantics: RT	Mean (±sd) Median Range					13975 ± 4106 14006 3014-22400
Phonology: Accuracy	Mean (±sd) Median Range					4,96 ± 0.21 5 4-5
Phonology: RT	Mean (±sd) Median Range					13462 ± 4302 13202 4928-24980
Total: Accuracy	Mean (±sd) Median Range					14,13 ± 1.31 15 8-15
Total: RT	Mean (±sd) Median Range					42458 ± 8495 43582 22967-58105