Université de Montréal

Assessing early white matter predictors of syntactic abilities in post-stroke aphasia using HARDI-based tractography

Par

Mariem Boukadi

Département de Psychologie

Faculté des Arts et des Sciences

Thèse présentée en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)

en Psychologie, option Neuroscience Cognitive

Juin, 2020

Université de Montréal

Département de Psychologie, Faculté des Arts et des Sciences

---

cette thèse intitulée

Assessing early white matter predictors of syntactic abilities in post-stroke aphasia using HARDI-based tractography

Présentée par

Mariem Boukadi

A été évaluée par un jury composé des personnes suivantes

Sarah Lippé
Présidente-rapporteur

Simona Maria Brambati
Directrice de recherche

Karine Marcotte
Codirectrice de recherche

Gregory West
Membre du jury

Marc F. Joanisse
Examinateur externe

Résumé

La recherche de prédicteurs d'habilités langagières en aphasie post-accident vasculaire cérébral (AVC) basés sur la matière blanche a récemment vu un élan. Cela a été motivé par l'émergence du modèle à double-voie où des faisceaux de matière blanche dorsaux et ventraux jouent un rôle important dans le langage, ainsi que par l'avènement de la tractographie basée sur l'imagerie par résonance magnétique (IRM) de diffusion permettant l'étude *in-vivo* des faisceaux de matière blanche et de leurs propriétés structurelles. Les caractéristiques structurelles et la charge lésionnelle des faisceaux de matière blanche ont permis de prédire les troubles langagiers dans la phase chronique dans quelques études. Cependant, les prédicteurs aigus de matière blanche des habilités syntaxiques en aphasie post-AVC chronique sont méconnus.

L'exploitation de la tractographie dans l'étude des faisceaux langagiers de matière blanche a été limitée par plusieurs défis méthodologiques, dont la difficulté de reconstruire des faisceaux ayant une architecture complexe. Des progrès méthodologiques ont été récemment introduits afin d'adresser ces limites, dont le plus important est la tractographie basée sur l'imagerie à haute résolution angulaire (« HARDI »). Cependant, la fiabilité test-retest de la reconstruction et des propriétés structurelles d'une approche de tractographie HARDI de pointe n'a pas encore été évaluée.

Le premier article de cette thèse visait à évaluer la fiabilité test-retest de la reconstruction et des propriétés structurelles (anisotropie fractionnelle, FA; diffusivité moyenne, axiale et radiale, MD, AD, RD; nombre d'orientations de fibres, NuFO; volume du faisceau; longueur moyenne des « streamlines ») de faisceaux langagiers majeurs (arqué, inférieur fronto-occipital, inférieur longitudinal, unciné, AF, IFOF, ILF, UF) obtenus avec une approche de tractographie HARDI de

pointe. La majorité des mesures de propriétés structurelles ont montré une bonne ou excellente fiabilité. Ces résultats ont des implications importantes pour l'utilisation d'une telle approche pour l'étude des faisceaux langagiers de matière blanche, car ils renforcent la confiance dans la stabilité des reconstructions et les propriétés structurelles obtenus avec la tractographie HARDI.

Le second article de cette thèse visait à déterminer si et quelles propriétés structurelles (FA, AD, volume du faisceau), et la charge lésionnelle, de l'AF et l'UF gauches dans la phase aigüe (≤ 3 jours), obtenus avec l'approche de tractographie HARDI utilisée dans le premier article, prédisent les habilités syntaxiques dans le discours spontané en aphasie post-AVC chronique (≥ 6 mois). Des régressions multiples ascendantes ont révélé que le volume de l'AF prédit la production des verbes, la complexité des phrases et la complexité de la structure argumentale du verbe. Le volume de l'UF a amélioré la prédiction de cette dernière. Ces résultats indiquent que le volume semble être un bon prédicteur précoce des habilités syntaxiques dans le discours spontané en aphasie post-AVC chronique.

Mis ensemble, les résultats de cette thèse soulignent l'utilité d'une approche de tractographie HARDI de pointe et son potentiel pour le développement futur de biomarqueurs précoces pouvant améliorer le pronostic de patients ayant une aphasie post-AVC chronique. Cela pourrait promouvoir l'optimisation des soins et le développement de thérapies pour le bienfait des patients et leurs familles.

Mots-clés: IRM de diffusion, tractographie, HARDI, matière blanche, aphasie post-AVC, AVC, syntaxe, discours spontané.

Abstract

The search for white matter predictors of language abilities in post-stroke aphasia has gained momentum in recent years. This growing interest has been driven by the emergence of the dual-stream framework where dorsal and ventral white matter bundles play an important functional role in language, as well as the advent of diffusion magnetic resonance imaging (MRI)-based tractography which allows the *in-vivo* investigation of white matter bundles and their structural properties. Structural characteristics, as well as the lesion load, of white matter bundles have been previously found to predict language impairments in the chronic phase. However, little is known about acute white matter predictors of syntactic abilities in chronic post-stroke aphasia.

Leveraging tractography to study white matter language bundles has been limited by several methodological challenges, such as the difficulty of reconstructing white matter bundles with a complex fiber architecture. A number of methodological advances have been introduced fairly recently to address these limitations, the most important of which is the advent of tractography based on High Angular Resolution Imaging (HARDI). However, the test-retest reliability of the reconstruction and structural properties of a state-of-the-art HARDI-based tractography pipeline has not been previously assessed.

The first article of the present thesis aimed to assess the test-retest reliability of the reconstruction and structural properties (fractional anisotropy, FA; mean, axial, radial diffusivity, MD, AD, RD; number of fiber orientations, NuFO; bundle volume; mean length of streamlines) of major white matter language bundles (arcuate, inferior fronto-occipital, inferior longitudinal, and uncinate fasciculi, AF, IFOF, ILF, UF) obtained using a state-of-the-art HARDI-based tractography pipeline. Most measures of structural properties showed good to excellent test-retest

reliability. These findings have important implications for the use of such a pipeline for the study of white matter language bundles, as they increase our confidence that the reconstructions and structural properties obtained from the tractography pipeline are stable and not due to random variations in measurement.

The second article of the thesis aimed to determine whether and which structural properties (FA, AD, bundle volume), as well as the lesion load, of the left AF and UF in the acute phase post-stroke (≤ 3 days), obtained with the same state-of-the-art HARDI-based tractography pipeline used in the first article, predict syntactic abilities in connected speech in chronic post-stroke aphasia (≥ 6 months). Forward multiple regressions revealed that the left AF's volume predicted the percentage of verbs produced, the structural complexity of sentences, as well as verb-argument structure complexity. The left UF's volume improved the prediction of verbs with a complex argument structure. These findings indicate that the bundle volume may be a good early predictor of syntactic ability in connected speech in chronic post-stroke aphasia.

Overall, the findings of this thesis highlight the usefulness of a state-of-the-art HARDI-based tractography approach and its potential for the future development of early biomarkers that could improve the prognosis and personalized care of patients with chronic post-stroke aphasia. This would promote the optimization of patient care and the development of therapies for the benefit of patients and their families.

Keywords: diffusion MRI, tractography, HARDI, white matter, post-stroke aphasia, stroke, syntax, connected speech.

Table of contents

List of tables

List of figures

List of acronyms and abbreviations

ACT	Anatomically constrained tracking
AD	Axial diffusivity
ADNI	Alzheimer's Disease Neuroimaging Initiative
AF	Arcuate fasciculus
BNT	Boston Naming Test
CHARMED	Composite Hindred and restricted ModEl of Diffusion
CSD	Constrained spherical deconvolution
CST	Corticospinal tract
DKI	Diffusion kurtosis imaging
DTI	Diffusion tensor imaging
DWI	Diffusion weighted imaging
EmC	Extreme capsule
FA	Fractional anisotropy
FAT	Frontal aslant tract
fODF	Fiber orientation distribution function
FSL	FMRIB software library
HARDI	High Angular Resolution Diffusion Imaging
IFG	Inferior frontal gyrus
IFOF	Inferior fronto-occipital fasciculus
ILF	Inferior longitudinal fasciculus
MCA	Middle cerebral artery

MD	Mean diffusivity
MldF	Middle longitudinal fasciculus
MLS	Mean length of streamlines
MNI	Montreal Neurological Institute
MRI	Magnetic resonance imaging
NAVS	Northwestern Assessment of Verbs and Sentences
NP	Noun phrase
NuFO	Number of fiber orientations
PLORAS	Predicting Language Outcome and Recovery After Stroke
PPA	Primary progressive aphasia
RD	Radial diffusivity
ROI	Region of interest
SLF	Superior longitudinal fasciculus
SPM	Statistical Parametric Mapping
STG	Superior temporal gyrus
TBSS	Tract-based spatial statistics
UF	Uncinate fasciculus
VLSM	Voxel-based lesion-symptom mapping
WM	White matter
WMQL	White matter language query

To my Mother, my blessing

« Everything connects to everything else. »

Leonardo da Vinci

Acknowledgements

I would like to thank the people and organizations who have played a part, big or small, in this Ph.D. story.

I would like to thank my Ph.D. supervisors Dr. Simona M. Brambati and Dr. Karine Marcotte for trusting me with this work and providing me with their time, resources, and constructive feedback. I am very grateful that they gave me the opportunity to work on such exciting and challenging projects.

I was very honored and fortunate to have received the *Canadian Institutes of Health Research* (CIHR) Frederick Banting and Charles Best Doctoral scholarship (2017-2020), as well as the doctoral award of the *Fonds de recherche en santé – Québec* (FRSQ) in partnership with the Heart and Stroke Foundation (HSF Québec) (2016-2017). I am very grateful to the CIHR, FRQS, and HSF for awarding me these doctoral awards to support me in my Ph.D. studies.

Special thanks to all my family, friends, mentors, and colleagues who have been incredibly supportive throughout the years.

Chapter I – Theoretical Context

1. General introduction

Aphasia (i.e., an impairment of language abilities) is one of the most common consequences of stroke. The search for white matter predictors of language impairment and recovery in post-stroke aphasia has seen growing interest in recent years. The emergence of the dual-stream network where white matter pathways play an important functional role has motivated research into the contribution of white matter language bundles to language impairment and recovery in post-stroke aphasia. A few studies have reported that the structural characteristics of white matter bundles could predict language abilities such as naming, repetition, and comprehension (Hillis et al., 2018; Ivanova et al., 2016; Xing et al., 2017). However, not much is known about white matter predictors of syntactic abilities in post-stroke aphasia.

Research into the role of white matter bundles in language was made possible by the advent of tractography, a neuroimaging technique based on diffusion weighted imaging (DWI), which allows the *in-vivo* reconstruction of white matter bundles and the assessment of their structural properties. However, tractography suffers from several challenges (such as a difficulty to reconstruct white matter bundles with a complex fiber architecture). Recent methodological advances have introduced new approaches to tractography that remedy, to a certain extent, some of the method's chief limitations.

In what follows, an introduction to the two articles of the present thesis will be presented. First, a definition of the cornerstones of this thesis will be presented, namely post-stroke aphasia, syntax in connected speech production, and tractography based on diffusion magnetic resonance imaging (MRI). We will then present the methodological challenges and advances of tractography and its reliability. This will be followed by an overview of the dorsal and ventral white matter language pathways in the dual-stream framework. A review of the studies that investigated the

white matter predictors of language impairments in post-stroke aphasia, then of those that looked into structural determinants of syntax production, will then be provided. Finally, the objectives and hypotheses of this thesis will be presented.

2. Post-stroke aphasia

2.1. Definition and clinical portrait

Ischemic stroke is a major cerebrovascular disease (Grefkes & Ward, 2014). It is caused by an occlusion of a cerebral artery which results in cell death and dysfunction (Yang et al., 2013). One of the most common cognitive sequelae of an ischemic stroke is an impairment of language functions known as aphasia. About a third of people who survive a stroke will present with post-stroke aphasia in the first few days (Heart and Stroke Foundation of Canada, 2017), and some will continue to live with it months or even years after the event (Johnson et al., 2019). Most often, post-stroke aphasia occurs following an occlusion of the left middle cerebral artery (MCA) which is one of the major arteries that supply blood to the brain and irrigates the left perisylvian language areas (Kemmerer, 2015).

Thus, post-stroke aphasia is an acquired language disorder that manifests in a variety of symptoms at the level of one or more linguistic processes (phonology, morphology, syntax, or semantics) affecting language production and comprehension (Zumbansen & Thiel, 2014). People with post-stroke aphasia may be classified into two broad categories: fluent (i.e., their language impairments are mostly receptive, which means that they have difficulties with comprehension) or non-fluent (i.e., their language impairments are mostly expressive, which means that they show difficulties with production of speech sounds, words, or sentences). Some patients may present with both expressive and receptive language impairments (known as mixed aphasia). Additionally, language impairments may vary in severity across individuals, as well as over the course of

recovery. Language impairments in post-stroke aphasia include anomia (i.e., word finding difficulties, including noun and verb production in confrontation naming or connected speech), word, pseudoword and sentence repetition difficulties, word and sentence comprehension difficulties, sentence production difficulties, and speech production difficulties (e.g., slower speech rate, phonemic paraphasia, hesitations, and connected speech impairments).

Post-stroke aphasia may sometimes, though not systematically, co-occur with verbal and non-verbal memory deficits (particularly short-term and working memory). For example, comprehension deficits in post-stroke aphasia have been found to be predicted by auditory short-term memory (Leff et al., 2009). Aphasia severity has also been found to be related to memory (working, short-term, and spatial) impairments in some patients with post-stroke aphasia (Laures-Gore et al., 2011; Potagas et al., 2011), which suggests that some common mechanism might underly language and memory impairments in some cases of post-stroke aphasia. Additionally, impairments in non-verbal tests of semantic memory are often observed in people with post-stroke aphasia (Fonseca et al. 2019), though it has been argued that such a deficit might be more one of semantic control than of storage (Jefferies & Lambon Ralph, 2006). However, not all post-stroke aphasia patients show such memory deficits and no cause-effect relationship between these two types of impairments has been identified.

Stroke and the aphasia it induces are dynamic conditions whose evolution broadly progresses over three main phases: acute, subacute, and chronic (Saur et al., 2006). The acute phase is identified as being sometime between the first 24 hours and the first week. The subacute phase may span a period between two weeks to three months post-stroke. The chronic phase is usually considered to be around three (according to some studies) to six months post-stroke (chronic aphasia is more often studied around this later time-window) and later.

2.2.Syntax in post-stroke aphasia

People with fluent and non-fluent post-stroke aphasia may also present with syntactic impairments that manifest in difficulties during the production or comprehension of complex sentences or verbs, particularly those with a complex verb-argument structure (Thompson et al., 2012). People with post-stroke aphasia may also show a reduction of grammatically well-formed sentences in connected speech (Edwards, 2005; Thompson et al., 2013).

Syntax can be defined as a hierarchical incremental structure building process of which the first building block is the verb. Sentence structure building starts with the selection of the verb lemma which encodes syntactic information, such as the verb-argument structure (Levelt et al., 1999). Verb-argument structure refers to the number and types of thematic roles or arguments a verb requires. For example, the transitive verb *put* requires three arguments, an agent, a theme, and a location, as in (a), and therefore has a complex argument structure, while the intransitive verb *sleep* requires one argument, an experiencer, as in (b), which means that it has a simple argument structure.

(a) John$_{\text{AGENT}}$ put the book$_{\text{THEME}}$ on the shelf$_{\text{LOCATION}}$.

(b) John$_{\text{EXPERIENCER}}$ sleeps.

The integration of a verb with its arguments requires the specification of the thematic roles it assigns and their relationship to the verb, as well as their syntactic realization (also known as subcategorization frames). For example, the theme role for the verb *see* can have two syntactic expressions, either a noun phrase (NP), as in (c) or a clause, as in (d):

(c) John saw [a movie]$_{\text{NP}}$

(d) John saw [that Mary was happy]$_{\text{CLAUSE}}$

The retrieval of the verb lemma projects its syntactic tree (which includes the argument slots), then each argument (e.g., a noun lemma) is selected and projects its realization, a NP, which is then integrated in its appropriate slot in the argument structure (e.g., in the subject or object position) (Ferreira, 2000; Thompson et al., 2015). This incremental hierarchical structure building process is characterized by the speaker's planning ahead of the sentence structure, resulting in the retrieval of multiple lemmas guided by their hierarchical structural relationships (Thompson et al., 2015). This incremental hierarchical process applies to simple structures (e.g. phrases) and extends to more complex ones as well, including complex sentences constituted of two or more clauses. Clauses are syntactic structures comprised of a subject and a main verb. Sentences constituted of two or more clauses are considered complex.

Syntactic impairments have received little attention in research, in comparison to other impairments in post-stroke aphasia. Syntactic impairments in production are also understudied in comparison to syntactic comprehension impairments. Additionally, adequate standardized clinical tests for the assessment of syntactic abilities in production are lacking. In English, the Northwestern Assessment of Verbs and Sentences (NAVS, Cho-Reyes & Thompson, 2012) has been developed and assesses verb and sentence production in terms of syntactic and verb-argument complexity. However, such a standardized clinical battery for syntax in language production does not exist in French. In this context, connected speech analysis can be a potent tool to assess the syntactic abilities of persons with aphasia.

Connected speech broadly refers to language production in a discourse context. It has the advantage of being more ecologically valid than standardized tests. As previously mentioned, standardized tests allow the controlled assessment of specific aspects of syntax (e.g. sentences with specific syntactic structures that are not naturally elicited by people with aphasia, such as object-

relatives and passives) in language production at the isolated verb and sentence levels. On the other hand, connected speech production allows for the study of these aspects in a discourse context that is more reflective of functional communication (Dipper et al., 2018). This has particularly important clinical implications. Indeed, the priority for patients with post-stroke aphasia and their relatives and the outcome they hope to achieve is better functional communication in their daily life (Boles, 1998; Edwards, 1998). In that context, measures of syntactic ability obtained from standardized test batteries like the NAVS, even if they provide valuable information about specific syntactic processes, might not reflect functional communication. Therefore, it is crucial to also assess syntactic ability in connected speech which provides better insight into functional communication than standardized tests, albeit an imperfect one. Additionally, connected speech may be sensitive to impairments that are not apparent in standardized tests, or may reveal compensatory strategies (Stark, 2019).

The most commonly used structured method for the elicitation of connected speech in clinical contexts is complex picture description (e.g., the Picnic Scene from the Western Aphasia Battery) (Bryant et al., 2016). Picture description is a task in which participants are asked to describe in depth a complex scene depicted in a single picture. It provides structured speech samples and has the advantage of restricting the content of the discourse produced to the contents of the picture (Bryant et al., 2016). This reduces sources of uncontrolled variability. Additionally, picture description produces samples of a manageable length for research and clinical purposes. As a result, this task is usually the most preferred one for the elicitation of connected speech. While expositional discourse (i.e., picture description) may not be an exact reflection of functional communication given the limited length of samples and the type of discourse it elicits, it remains a type of assessment that may afford better insights into everyday communication abilities than standardized clinical tests. Connected speech has been used to assess different aspects of the

syntactic abilities of fluent and non-fluent aphasic speakers (Bastiaanse, 2011; Hsu & Thompson, 2018; Mirman et al., 2019). The most commonly used measures obtained from connected speech to assess syntactic abilities are the number of free or embedded clauses (as a measure of syntactic complexity of sentences), the percentage of verbs in the sample, and the number of well-formed sentences. The verb-argument structure has also been investigated using connected speech (Bastiaanse & Jonkers, 1998; Hsu & Thompson, 2018; Thompson et al., 1995).

Studies on syntactic impairments in post-stroke aphasia using connected speech have revealed that both fluent and non-fluent aphasic speakers may experience syntactic impairments. For example, patients with post-stroke aphasia have been found to show impaired production of free or embedded clauses (Edwards, 1995; Edwards & Bastiaanse, 1998; Edwards & Knott, 1994; Hsu & Thompson, 2018; Llinàs-Grau & Martínez-Ferreiro, 2014; Stark, 2019; Webster et al., 2007), difficulties in the production of verbs in connected speech (Edwards, 1998; Edwards & Bastiaanse, 1998; Saffran et al., 1989; Thompson et al., 1995), fewer well-formed sentences (Edwards, 2005; Thompson et al., 2013), and difficulties with the production of verbs with a complex argument structure (Kim & Thompson, 2004; Thompson et al., 1995; Webster et al., 2007).

3. Diffusion MRI-tractography

The specification of white matter connectivity in the language network, as well as the investigation of its contribution to post-stroke aphasia has been made possible by the advent of tractography. This technique leverages diffusion MRI data to virtually reconstruct white matter bundles *in-vivo* (Dell'Acqua & Catani, 2012).

Diffusion MRI or diffusion weighted imaging (DWI) probes the diffusion of water molecules in biological tissues, thereby allowing us to recover information about the structural

properties of white matter (Descoteaux, 2015). In most parts of the brain, such as the cerebrospinal fluid or grey matter, the diffusion of water molecules is unrestricted (i.e., the molecules move randomly in different directions). In white matter, however, water molecules move along myelinated axons, making their movement restricted to one main direction perpendicular to axons (Catani & Forkel, 2019). This movement is called anisotropic (see Figure 1). Tractography emerged about 20 years ago and, as previously mentioned, has been crucial in advancing our understanding of the role of white matter pathways in the language network. Even more recently, tractometry was introduced. This technique is closely related to tractography in that it allows a detailed study of each reconstructed white matter bundle's structural characteristics. Several metrics and measures that reflect the micro- and macro-structural properties of white matter can be extracted from the bundles reconstructed by means of tractography. The most common microstructural measures include fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD) (Yeatman et al., 2012). Macrostructural measures such as bundle volume or the mean length of streamlines (MLS; i.e., the length of the virtually reconstructed fibers) may also be obtained and provide information about the anatomy of the bundle (Catani & Forkel, 2019; Girard et al., 2014).

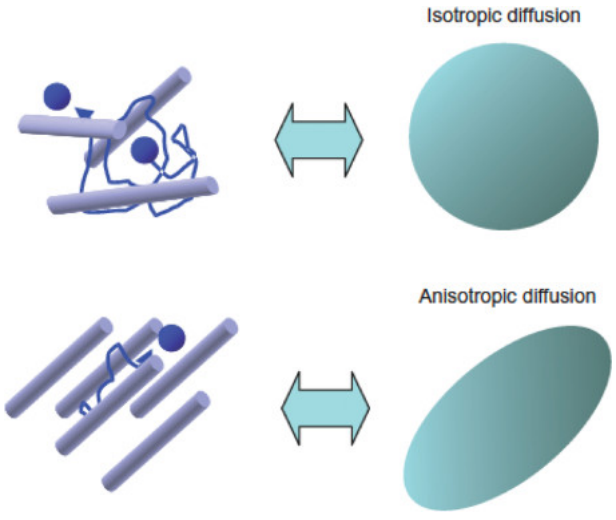


Figure 1. Illustration of isotropic and anisotropic diffusion.
On the left, a representation of isotropic (top of the illustration) vs. anisotropic (bottom of the illustration) diffusion of water molecules is shown. On the right, the shape of the diffusion tensor given isotropic vs. anisotropic diffusion is illustrated. Reprinted from *Introduction to Diffusion Tensor Imaging and Higher Order Models*, 2nd Edition, Mori, S. and Tournier, J.-D., Principle of Diffusion Tensor Imaging, Page 29, Copyright (2014), with permission from Elsevier.

Measures of micro- and macro-structural properties provide different types of information. While microstructural measures provide information about the degree of diffusivity and anisotropy in white matter, macrostructural measures can be used to inform us about the morphology and anatomy of the white matter bundles. FA, MD, RD, and AD have been found to be particularly sensitive to microstructural abnormalities in white matter following a stroke (Alexander et al., 2007; Tournier, Masterton, & Seitz, 2012). FA is a widely used metric that reflects the degree of tissue anisotropy (Beaulieu, 2014). It ranges from zero (i.e. perfectly isotropic diffusion) to 1 (i.e., highly anisotropic diffusion). A decrease in the mean FA along a white matter bundle reflects lower anisotropy (i.e., less restricted movement of water molecules) and is believed to be sensitive to axonal damage. In other words, damage at the level of white matter could cause axons to degrade, lose myelination, or become disorganized inside a bundle of fibers. This would in turn lead to freer, less restricted movement of water molecules. A drop in FA values has been reported in patients with white matter damage, such as demyelination or Wallerian degeneration (Beaulieu, 2002; Pierpaoli et al., 2001; Tournier, Masterton, & Seitz, 2012). MD is another metric that assesses white matter microstructure. It reflects the average movement of water molecules in a voxel (Salat, 2014). An increase in MD values is thought to reflect decreased restriction of the diffusion of water molecules (Madhyastha et al., 2014). AD and RD indicate the diffusion of water molecules parallel and perpendicular, respectively, to the principal fiber direction. Studies have found that these different metrics (FA, MD, AD, RD) may be affected differently by different types of white matter

injury (Pierpaoli et al., 2001; Song et al., 2003, 2002; Wheeler-Kingshott & Cercignani, 2009). For example, AD would be affected by acute axonal damage, while RD would be affected by demyelination. This entails that these different metrics may be differentially affected in different neuropathologies or at different stages of a neurological condition such as stroke. At the macrostructural level, a reduction of the volume of bundles has been reported in the chronic phase post-stroke (e.g., Jang & Lee, 2014).

Finally, tractography has also been leveraged in post-stroke aphasia studies to investigate the degree of overlap between a white matter bundle and the lesion (Geva et al., 2015; Hillis et al., 2018; Marchina et al., 2011). This approach combines lesion information with the *in-vivo* reconstruction of white matter bundles to obtain a measure known as the lesion load. Thus, the lesion load combines information about lesion size and location specific to white matter and is considered an indirect measure of damage to specific white matter bundles.

4. Methodological challenges and improvements of tractography

Since its inception, tractography has faced several challenges that have limited its potential for the study of white matter bundles and their characteristics. The biggest challenge for tractography is that of complex fiber architectures (e.g., crossing, fanning, bending, or kissing fibers) which the diffusion tensor imaging (DTI) model fails to properly represent. Other limitations and challenges include false negatives (streamlines that stop in white matter and do not reach the grey matter), intra-rater variability in ROI approaches to tractography, and false positives (i.e., streamlines that are not anatomically plausible) (Rheault, Poulin, at al., 2020). Solutions have been developed to try and remedy these limitations and biases. In what follows, each limitation or challenge of tractography and the corresponding solution that has been developed by the field will be introduced.

4.1. Challenge 1: Complex fiber architectures

DTI was the first approach to model the diffusion of water molecules in white matter. It has been tremendously influential (Reuter & Fischl, 2011), to the point that it is often used to refer to the diffusion MRI imaging modality as a whole. In this approach, the diffusion process is modeled using an object called the tensor with three eigenvectors and three eigenvalues describing its diffusion properties (Descoteaux, 2015). The FA, MD, RD, and AD metrics are derived from these eigenvalues. However, the tensor can only represent one fiber population per voxel, while up to 90% of voxels in the white matter contain a complex fiber configuration, namely kissing, crossing, fanning, bending, or twisting fibers (see Figure 2) (Descoteaux & Deriche, 2008; Jeurissen et al., 2014). Thus, DTI-based tractography is limited in white matter areas with a complex architecture which happen to represent a majority of white matter voxels. Because it assumes a single main fiber orientation per voxel (when in fact there are many), DTI often yields invalid streamlines and fails to reconstruct anatomically valid streamlines or entire portions of white matter pathways (Farquharson & Tournier, 2016; Farquharson et al., 2013; Jeurissen et al., 2013; Pierpaoli et al., 2001). This has come to be known as *the crossing fiber problem* which refers to any type of complex fiber configuration (Descoteaux, 2015).

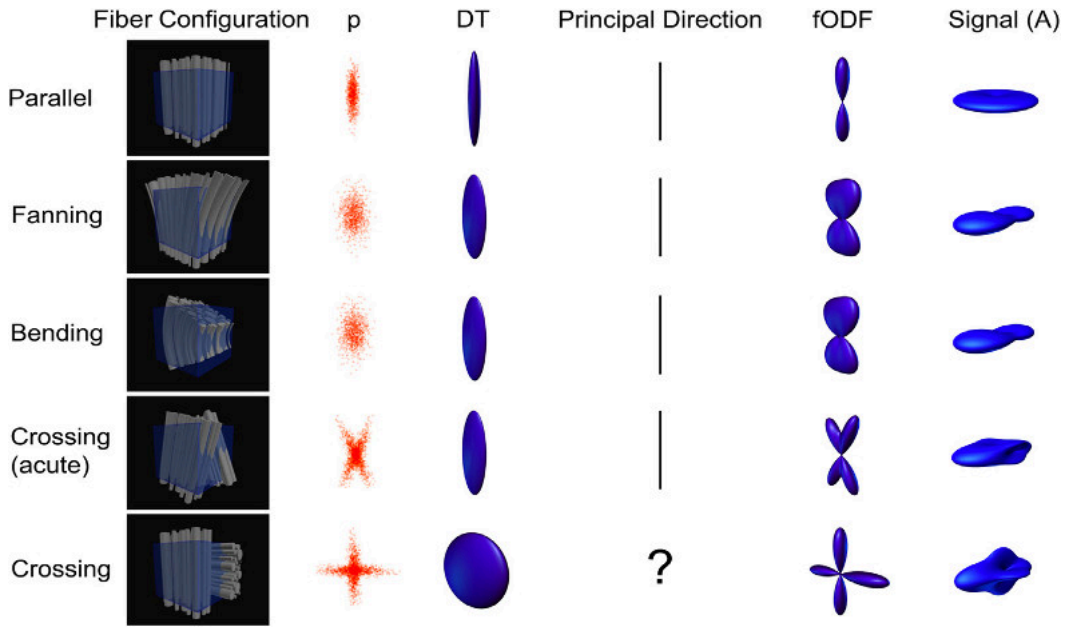


Figure 2. Illustration of the diffusion tensor's (DT) and the fODF's representations of different fiber configurations.

The left column illustrates different fiber configurations in the white matter. The third column illustrates the DT's representation of these configurations. As can be seen, the DT only models the principal direction and is not able to define it in a case of complete fiber crossing (last row). By contrast, the fODF is better able to model these complex configurations (Seunarine & Alexander, 2014). Reprinted from *Diffusion MRI: From Quantitative Measurement to In-vivo Neuroanatomy*, 2nd Edition, Seunarine, K. K. and Alexander, D. C., Multiple fibers: Beyond the diffusion tensor, Page 106, Copyright (2014), with permission from Elsevier.

4.2.Solution 1: HARDI-based tractography

High angular resolution diffusion imaging (HARDI) was introduced to remedy the crossing fiber problem. There exist a variety of HARDI-based approaches, known as higher-order models, that model the diffusion process (Mori & Tournier, 2014). These include q-ball imaging, Composite Hindred and restricted ModEl of Diffusion (CHARMED), and diffusion kurtosis imaging (DKI) (Descoteaux, 2015). However, such approaches require multi-shell DWI data (i.e., several b-values in the MRI acquisition), which involve a very time-consuming acquisition and are

therefore unsuitable for clinical contexts. The most commonly used HARDI-based approach is constrained spherical deconvolution (CSD) which is based on a single-shell acquisition and models the diffusion process by means of an object known as the fiber orientation density function (fODF). The fODF represents an estimation of the distribution of fiber orientations in a voxel (Tournier et al., 2004). Thus, CSD-tractography is able to better account for complex fiber configurations in white matter than DTI-tractography (see Figure 2), though it does not solve the crossing fiber problem completely (Côté et al., 2013; Descoteaux, 2015).

CSD-tracking has been shown to yield more anatomically valid and plausible reconstructions of white matter bundles than DTI and to successfully reconstruct many pathways with complex fiber configurations (Behrens et al., 2007; Catani & Forkel, 2019; Dell'Acqua & Tournier, 2019; Farquharson & Tournier, 2016; Farquharson et al., 2013; Jeurissen et al., 2017; Jeurissen et al., 2011; Tournier, Calamante, & Connelly, 2012). For example, probabilistic CSD-tractography has been found to successfully reconstruct the anatomically valid fan-shaped configuration of the sensorimotor white matter pathways, while DTI-tractography (both deterministic and probabilistic) failed to reconstruct the lateral projections of the same pathways (Farquharson et al., 2013). Others have shown that DTI fails to correctly reconstruct the lateral projections of the corpus callosum which cross more dense and dominant streamlines from the corticospinal tract, while CSD-tractography successfully reconstructed these projections (Dell'Acqua & Tournier, 2019).

CSD-tractography can be leveraged to extract measures about the underlying fiber architecture, such as the number of fiber orientations (NuFO) which is a HARDI-specific measure based on the number of local maxima of the fiber orientation distribution (Dell'Acqua et al., 2013). The macrostructural measures of bundle volume and MLS are not specific to HARDI or DTI but

may be potentially affected by the approach used for tractography (for example, white matter bundles reconstructed by CSD-tractography have been shown to have a larger volume than when the reconstruction was done using DTI-tractography, Kristo et al., 2013).

4.3. Challenge 2: False negatives

One of the main challenges of tractography are false negatives, namely streamlines that are not reconstructed or stop abruptly in the white matter. False negatives are due to a variety of reasons which are sources of uncertainty, such as a complex fiber architecture (Behrens et al., 2007; Jeurissen et al., 2011), choice of tracking algorithm (i.e. probabilistic or deterministic) (Mori & Tournier, 2014; Tournier, Calamante, & Connelly, 2012), or even noise, which is inherent to diffusion imaging (Mori & Tournier, 2014). These sources of uncertainty can compound the difficulty of tracking bundles with a curved shape, such as the uncinate fasciculus (UF) (Girard et al., 2014).

4.4. Solution 2: Probabilistic tractography

Because HARDI-based tractography performs better than DTI tractography in areas with a complex fiber architecture, it helps to reduce the false negatives problem. However, as mentioned above, the choice of tracking algorithm (deterministic or probabilistic) also plays a role in this regard. Deterministic tractography uses the peak orientation of the ODF to carry out tracking, which does not take into account the uncertainty in fiber orientations, thereby leading to false negatives (Smith et al., 2012). Probabilistic tractography was introduced to reduce the rate of false negatives, as this approach accounts for the uncertainty in diffusion data by sampling different fiber orientations from the fODF (Mori & Tournier, 2014). CSD-based probabilistic tractography has been found to outperform CSD-based deterministic tracking (Descoteaux et al., 2009).

4.5. Challenge 3: False positives

The combination of CSD-based tractography with probabilistic tracking has several advantages, as previously mentioned, but comes at the price of an increased number of false positives. These are invalid, not anatomically plausible connections. False positives are an inherent issue of tractography, but their rate increases with the use of HARDI-based models and probabilistic tracking (as compared to DTI-based models and deterministic tracking).

4.6. Solution 3: Anatomically constrained tracking and tract-filtering

A number of methodological advances have been made to help address the false positives problem. These include anatomically constrained tracking (ACT) (Smith et al., 2012), tract-filtering, and the use of a-priori anatomical knowledge to guide the reconstruction process. ACT leverages the high resolution segmented T1-weighted image to guide the tractography. It does so by using anatomical priors, namely the classification of different tissue types (cortical, subcortical grey matter, white matter, and cerebrospinal fluid) to set criteria for the termination of streamlines and their inclusion or exclusion based on biological plausibility (Smith et al., 2012). These criteria include terminating and accepting a streamline that enters cortical grey matter or rejecting a streamline that terminates in white matter or the cerebrospinal fluid (Rheault, Poulin, et al., 2020). Tract-filtering algorithms remove streamlines that are considered to be outliers (i.e. implausible connections resulting from tractography biases) from the tractogram (i.e., the set of reconstructed white matter streamlines in the whole brain). These algorithms use different criteria, such as biological plausibility based on anatomical priors (as in ACT) or length of streamlines (Girard et al., 2014; Smith et al., 2013). Tract-filtering is carried out after the tractogram is constructed (i.e., after ACT has been applied) and therefore adds another level of tractography-bias reduction. While these approaches may not completely eliminate false positives, they significantly reduce them.

4.7. Challenge 4: Inter- and intra-rater variability of ROI approaches to tractography

To reconstruct or extract bundles from the tractogram, manual approaches have for a long time been the method of choice. These approaches mainly consist in manually delineating one or two regions of interest (ROIs) to seed the bundles of interest. However, this approach is highly dependent on the expert neuroanatomical knowledge of the experimenter or rater and is inevitably biased by inter- and intra-rater variability (Rheault, Poulin, et al., 2020). Thus, the reconstruction of a given bundle may vary not only from one subject to another, but also from one testing timepoint to another, as well as between raters (so-called virtual dissections are often carried out by more than one single rater given their time-consuming nature) (Rheault, De Benedictis, et al., 2020).

4.8. Solution 4: Semi-automatic approaches

Automatic or semi-automatic approaches to bundle segmentation have been introduced in recent years. One of the most intuitive and user-friendly is the White Matter Query Language (WMQL) which can be implemented in the software package TractQuerier (Wassermann et al., 2016). This approach eschews the inter- and intra-rater reliability issue of the manual ROI placement approach. Indeed, WMQL allows the extraction of bundles from the tractogram by using commands describing the bundle of interest in anatomical definitions based on existing atlases of white matter. This method is dependent on the quality of the grey and white matter parcellation carried out on the T1-weighted image, since it relies on it for the anatomical definitions. It also requires some fine-tuning of the definitions but once they are written up, the same queries will be used for all subjects. In other words, the by-default commands developed by Wasserman et al. (2016) may be modified by the user by following an Atlas of white matter pathways such as Catani & Thiebaut de Schotten's (2012) to ensure that the reconstruction is based on valid prior anatomical knowledge of the bundles. Using one subject as a reference, one may then add inclusion and

exclusion cortical and subcortical regions to the command or query, then checking the results on all subjects, and iterating until reaching a satisfactory result (i.e. that the bundle is reconstructed in all subjects and that it is biologically valid based on prior anatomical knowledge of the bundle). The final command can then be applied to all subjects across all timepoints to reconstruct their bundles in a fully automatic way.

To summarise, using a HARDI-based higher-order model such as CSD for tractography helps to address (at least to some extent) the crossing-fiber problem and its ensuing issues (such as false negatives). Using probabilistic tractography helps with the false negatives issue as well. Using ACT and tract-filtering remedies to some extent the challenge of false positives. And finally, using automatic or semi-automatic reconstruction approaches such as WMQL reduces the biases entailed by the conventional manual ROI-based reconstruction approach. Thus, a combination of the abovementioned methodological advances in a tractography pipeline allows to limit several shortcomings and biases of conventional tractography. However, the reliability of the reconstructions and the diffusion measures yielded by such a pipeline has not been assessed.

5. Reliability of tractography

As mentioned above, diffusion MRI is an inherently noisy process and tractography comes with several biases. Before using tractography and tractometry to study white matter macro- and microstructure in healthy and pathological populations, it is crucial to assess the stability of these techniques. This is important to ensure that any variations observed in white matter anatomy and microstructure are due to biological phenomena rather than random variations in measurement or reconstructions (Cousineau et al., 2017). Test-retest reliability assesses the stability of measurement instruments to ensure that any variation is due to true differences rather than chance or random factors (Multon, 2012). In other words, test-retest reliability is essential to ensure that

any variations we would observe in reconstructions and structural characteristics between patients are due to true inter-individual differences rather than random variability caused by instability at one or several levels of the tractography pipeline.

Good to excellent test-retest reliability of diffusion metrics has been previously reported using a variety of approaches such as tract-based spatial statistics (TBSS), atlas ROI-based approaches and voxel-based approaches (Cole et al., 2014; Duan et al., 2015; Jovicich et al., 2014; Madhyastha et al., 2014; Papinutoo et al., 2013). A few studies have assessed the test-retest reliability of diffusion measures extracted from white matter bundles reconstructed with DTI-based tractography among healthy subjects (Buchanan et al., 2014; Ciccarelli et al., 2003; Danielian et al., 2010; Heiervang et al., 2006; Vollmar et al., 2010; Wang et al., 2012). These studies showed good to excellent test-retest reliability of most diffusion measures, though some showed inconsistent results with poor reliability for some metrics or white matter bundles. For example, Danielian et al. (2010) found poor reliability of AD in all studied white matter bundles, while Wang et al. (2012) found that FA and MD had good reliability in some bundles but poor reliability in others. The test-retest reliability of CSD-tractography has been previously assessed in three studies (Besseling et al., 2012; Cousineau et al., 2017; Kristo et al., 2013). These studies also showed inconsistent results. For example, Besseling et al. (2012) found that microstructural measures of the arcuate fasciculus (AF) were stable while its volume showed poor reliability. Cousineau et al. (2017) found that the reconstruction of some white matter bundles showed good reliability while others did not. CSD addresses only one of tractography's biases. Other sources of uncertainty or variability need to be addressed for tractography and tractometry to be stable by leveraging the aforementioned methodological advances (ACT, tract-filtering, probabilistic tracking, etc.). However, the test-retest reliability of the reconstructions and diffusion measures yielded by such a

tractography pipeline has not yet been assessed. Additionally, the only language bundle included in previous CSD-tractography reliability studies was the AF and the test-retest reliability of its reconstruction and characteristics was inconsistent.

6. The dual-stream language network

6.1. A brief history

The Wernicke-Lichtheim-Geschwind classic model (illustrated in Figure 3), the first model of language organization in the brain, dates back to the late 19th century and emerged from the pioneering work of surgeon Paul Broca (1861), who identified the posterior two thirds of the inferior frontal gyrus (IFG) as the neural correlate of language production, and neurologist Carl Wernicke (1874), who identified the superior temporal gyrus (STG) as the region supporting language comprehension (Catani & Budisavljevic, 2014). The model was first described by Wernicke then later refined and illustrated by Lichtheim in the late 19th century, and finally revised in the late 20th century by the neurologist Geschwind with the inclusion of the AF which had been described and studied earlier by neuroanatomists such as Constantin von Monakow and Déjèrine (Dick & Tremblay, 2012; Tremblay & Dick, 2016). The classic model consisted of two main cortical areas, at the time referred to as Broca's (IFG) and Wernicke's (posterior STG) areas connected by a large bundle of fibers known as the AF (Tremblay & Dick, 2016). In this framework, the only role ascribed to the AF was that of verbal repetition. This came mainly from lesion studies with stroke patients who presented with conduction aphasia (i.e., repetition impairments) following a lesion to their AF (Dick & Tremblay, 2012).

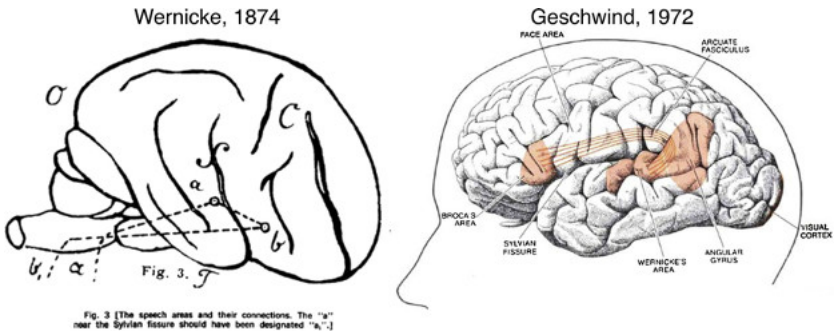


Figure 3. Illustration of the classic model of language organization in the brain.
On the left, the original model by Wernicke in 1874 and on the left, Geschwind's later version of the classic model with Broca and Wernicke's areas and the AF connecting them (Tremblay & Dick, 2016). Reprinted from *Brain & Language*, Vol. 162, Tremblay and Dick, Broca and Wernicke are dead, or moving past the classic model of language neurobiology, Page 62, Copyright (2016), with permission from Elsevier.

Advances in neuroimaging have led to a departure from the classic model and the rise of the dual-stream framework for language (Dick et al., 2014; Hickok & Poeppel, 2007). Drawing on an analogy with the dual-stream model of visual processing (Ungerleider & Haxby, 1994), the dual stream model by Hickok and Poeppel (2000, 2004, 2007) identified a functionally-specified language network of speech perception. In this network, a dorsal stream maps speech sounds to articulation and a ventral stream maps sound to meaning. This model was mainly a proposal for a neurofunctional basis of speech perception (and, to a limited extent, speech production). It also did not include a specification of the structural connectivity supporting this network. In other words, there was no description of white matter connectivity in this model (since it was largely supported by evidence from the functional MRI literature).

Shortly thereafter, a series of studies and theoretical proposals sought to determine the neuroanatomical basis of Hickok and Poeppel's (2007) dual-stream model by identifying the association white matter pathways connecting the previously identified dorsal and ventral cortical areas. These proposals extended the dual-stream model of speech perception to other language

domains and functions, such as syntactic processing (Friederici et al., 2006), and object naming (Duffau et al., 2014). Most of this work converges on the following dual-stream architecture (see Figure 4 for an illustration): the AF is the direct white matter pathway anchoring the dorsal stream. It connects the left IFG (*pars triangularis* and *pars opercularis*) and the left posterior STG, as well as the middle and inferior temporal gyrus (Dick et al., 2014). The ventral stream is anchored by direct and indirect pathways (Duffau et al., 2014). The direct pathway is constituted by the inferior fronto-occipital fasciculus (IFOF), putatively connecting the occipital, parietal, temporal, and ventrolateral frontal cortex. The indirect pathway is constituted by the inferior longitudinal fasciculus (ILF) connecting the occipital and temporal lobes, and the uncinate fasciculus (UF) connecting the anterior temporal lobe to the orbitofrontal cortex. According to the dual-stream framework of language, the ventral stream is bilateral, while the dorsal stream is left-dominant (Hickok & Poeppel, 2007), though a role for the right AF in prosody has also been suggested (Glasser & Rilling, 2008).

A few other white matter bundles have also been suggested to play a role in language. These include the three components of the superior longitudinal fasciculus (SLF I, II, III), the middle longitudinal fasciculus (MdlF) and the extreme capsule (EmC), as well as, more recently, the frontal aslant tract (FAT) (Catani et al., 2013). However, these white matter bundles are anatomically ill-defined and their existence in the human brain remains controversial. For example, the EmC is a large structure of white matter fibers which has been identified in non-human primates. In humans, the EmC has been found to correspond to the IFOF (which exists in humans but not in macaques), and to include fibers from the UF (Thiebaut de Schotten et al., 2012). Similarly, the MdlF has been identified in the macaque but its existence in the human brain is controversial (Dick et al., 2014). As for the FAT, it has mainly been identified in a few tractography

studies and its anatomical terminations and trajectory remain highly underspecified. Additionally, the functional role of these secondary bundles in language remains underspecified and controversial. Thus, to date, the major and most studied white matter bundles of language are the AF, IFOF, ILF, and UF.

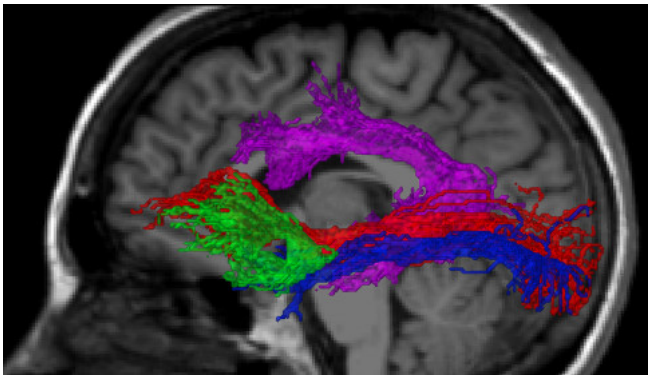


Figure 4. An illustration of the major dorsal and ventral white matter bundles of the dual-stream framework of language organization in the brain.
Purple: the AF; Red: the IFOF; Blue: the ILF; Green: the UF.

Different roles have been ascribed to the major dorsal and ventral white matter language fiber bundles. The AF has been suggested to be involved in phonological processing (Duffau, 2016), as well as in syntactic processing, more specifically in the production and comprehension of complex sentences (Catani & Bambini, 2014; Friederici & Gierhan, 2013). The IFOF is considered to be the main direct ventral pathway essential for semantic processing, as well as language comprehension (Duffau et al., 2014; Friederici & Gierhan, 2013). The ILF and the UF have also been suggested to play a role in semantic and lexical processing (Catani & Bambini, 2014; Duffau et al., 2014). The ILF has also been suggested to play a role in visual-orthographic processing (Dick et al., 2014). In addition to its proposed role in lexical and semantic processing, the UF has also been suggested to play a role in syntactic processing (Friederici, 2011; Friederici et al., 2006).

6.2. The dual-stream framework of syntax

The dual-stream framework of syntax has identified the left AF as the dorsal pathway which would play a role in global syntactic processes, such as processing complex sentences (Friederici et al., 2006; Zaccarella & Friederici, 2015). In the ventral stream, the UF would be involved in local processes, such as basic structure building (i.e., the hierarchical ordering of words and phrases to build basic syntactic structures like phrases or simple sentences, such as the integration of a verb with its arguments), and combinatory processes (such as assigning thematic roles that the verb requires to each argument) (Friederici et al., 2006; Zaccarella & Friederici, 2015).

7. The contribution of white matter to outcome in post-stroke aphasia

The emergence of the dual-stream framework of language organization in the brain has led to a reconsideration of the role of the AF in language impairments, namely that its damage may contribute to a variety of language functions and not just to repetition impairments as previously believed in the context of the classic language model. Additionally, it has motivated the study of the contribution of other major white matter language bundles in language impairments post-stroke. Thus, the role of white matter language bundles' damage and their structural characteristics to language outcomes in post-stroke aphasia has seen increasing interest in recent years. Previous studies have found that the lesion load of the AF (an indirect measure of damage that represents the degree of overlap between a lesion and a white matter bundle) in the chronic phase predicts chronic language abilities, such as naming, speech fluency, comprehension and repetition (Geva et al., 2015; Marchina et al., 2011). Others have found that diffusion measures (FA, AD, MD, RD, or bundle volume) of different dorsal and ventral language bundles in the chronic phase correlate with aphasia outcome or different language functions, such as word and sentence comprehension, naming, and sentence construction (Ivanova et al., 2016; Xing et al., 2017; Jang & Tak, 2014).

However, most of these studies focused on white matter damage or structural characteristics in the chronic phase.

A few studies have investigated whether white matter characteristics of the left AF a few days after the stroke predict language outcome in the chronic phase. One study has reported that the AF's lesion load in the acute phase predicted naming outcome in the chronic phase (Hillis et al., 2018). Another has shown that no diffusion measure in the left AF predicted general language outcome in the chronic phase (Forkel et al., 2014). However, none of these studies have looked into early predictors of syntactic abilities in connected speech.

8. The dorsal and ventral structural basis of syntax during language production in aphasia

Only a handful of studies in post-stroke aphasia have investigated the structural determinants of syntax in language production. These studies have shed some light on the dorsal and ventral structural determinants of syntax in language production. Using voxel-based lesion symptom mapping (VLSM), Henseler et al. (2014) investigated the structural correlates of several dimensions of connected speech (assessed using a semi-standardized interview), including syntactic structure in a group of chronic post-stroke aphasia patients. This study found that syntactic structure, assessed using a single general measure that encompassed correctness of grammar and syntactic complexity, was correlated with lesions to the posterior and middle STG. In another VLSM study, Faroqi-Shah et al. (2014) found that production of sentences of different levels of complexity (assessed by pictures that elicit different types of syntactic structures, such as passives, actives, and object-relatives which include embedded clauses) was related to dorsal cortical lesions primarily in the IFG, as well as in the STG and supramarginal gyrus.

A few VLSM studies investigating verb naming (the first building block of syntactic structure) in chronic post-stroke aphasia patients, found that verb naming (assessed using an experimental picture naming task where participants were asked to name actions or objects) was related to lesions in dorsal and ventral frontal cortical regions, such as the left IFG (Akinina et al., 2019; Alyahya et al., 2018; Piras & Marangolo, 2007), orbitofrontal cortex (Alyahya et al., 2018; Piras & Marangolo, 2007), the precentral gyrus (Akinina et al., 2019; Alyahya et al., 2018), and other frontal regions. Verb naming was also related to dorsal and ventral temporal cortical regions, including the left posterior middle and inferior temporal gyri, the anterior middle temporal gyrus, the temporal fusiform cortex, planum polare, the temporal pole (Alyahya et al., 2018), as well as the left anterior temporal lobe and the STG (Marangolo & Piras, 2010). Akinina et al. (2019) found that their VLSM map for verb naming (assessed using an experimental picture naming task) in a group of patients with chronic post-stroke aphasia intersected with an out-of-sample probabilistic map of several dorsal and ventral association white matter bundles (left AF, UF, IFOF, SLF, FAT, frontal orbito-polar, frontal inferior longitudinal tracts, and fronto-insular tract 4).

Little is known about the white matter correlates of syntax in language production in post-stroke aphasia. Den Ouden et al. (2019) recently investigated the chronic structural (grey and white matter) predictors of syntactic impairments in language production in chronic post-stroke aphasia. This study used VLSM to assess grey matter damage and a connectome approach which assessed white matter connectivity between pairs of cortical regions of interest. The whole-brain connectome was reconstructed using DTI-based probabilistic tractography. Syntactic impairments were assessed using the NAVS battery (Cho-Reyes & Thompson, 2012). This study found that patients with low NAVS scores were characterized by damage to the medial posterior STG and by reduced connectivity in the dorsal in ventral streams. The verb-argument structure production test

scores were associated with damage to a middle to posterior region of the STG and to the angular gyrus (a grey matter region which lies at the intersection between the dorsal and ventral stream). Additionally, the production of verbs with a complex argument structure (i.e., verbs that require three arguments) was related to dorsoventral white matter connectivity between the middle temporal gyrus and the insula. The results of this study have indicated that syntax in language production, and particularly verb-argument structure, in chronic post-stroke aphasia relies on both dorsal and ventral grey and white matter regions. In a DTI-tractography study carried out in the chronic phase post-stroke, Ivanova et al. (2016) investigated the relationship between the microstructural characteristics (FA, MD, AD, RD) of several white matter bundles (AF, UF, ILF, IFOF, corticospinal tract, and corpus callosum) and language functions such as picture naming, word and sentence comprehension, and sentence complexity. The latter was assessed using a sentence construction test where participants had to produce sentences of increasing complexity elicited by pictures of actions. Only the FA of the left AF was related to scores on the sentence construction test. These findings highlighted the role of the left AF in structural sentence complexity.

Some evidence regarding the structural correlates of syntax in language production also comes from a handful of studies in the primary progressive aphasia (PPA) literature. The production of syntactically complex sentences (i.e., sentences constituted of two or more clauses) in PPA patients has been found to be related to atrophy in the left IFG and prefrontal cortex, located in the dorsal stream (Gunawardena et al., 2010; Wilson et al., 2010). One study has also found that the left AF's FA is related to impaired syntax in language production (rated on a 7-point scale by a speech-language pathologist) in a group of PPA patients (Wilson et al., 2011). However, the specific role of white matter language bundles in syntax remains, in large part, a pending question.

The handful of studies carried out on the subject have mainly focused on sentence complexity or a general assessment of syntax in production. Measures that reflect local structure building (e.g., verb-argument structure) have not been investigated in these studies (except for den Ouden et al., 2019 who did not use bundle-specific analyses). As a result, the role of the left UF in local structure building proposed by Friederici et al. (2006) remains to be fully elucidated.

9. Objectives and hypotheses

9.1. Article 1

9.1.1. Objective

The aim of the first article is to assess the test-retest reliability of the reconstruction as well as the micro- and macrostructural characteristics (FA, MD, RD, AD which are tensor-based measures, NuFO which is a HARDI-based measure, as well as bundle volume and MLS) of the left and right AF, UF, IFOF, ILF reconstructed using a state-of-the-art probabilistic CSD-tractography pipeline in a sample of older healthy individuals.

9.1.2. Hypotheses

We expect the reconstruction of the left and right AF, UF, IFOF, and ILF to show good morphological overlap between the first and second scanning occasions (one week apart). We also expect micro- and macro-structural diffusion measures (i.e., FA, MD, RD, AD, NuFO, volume, and MLS) extracted from the four major white matter bundles (in the left and right hemispheres) to show good to excellent test-retest reliability.

9.2. Article 2

9.2.1. Objective

The aim of the second article is to determine whether and which of the structural characteristics (FA, a widely used measure of anisotropy, Beaulieu, 2002; AD, the diffusivity measure which has been shown to be particularly sensitive to axonal damage in the acute phase post-stroke, Moulton et al., 2019; and bundle volume) of the left AF and UF reconstructed in the acute phase using our state-of-the-art probabilistic CSD-based tractography pipeline predict the production of verbs, sentences, as well as sentence and verb-argument structure complexity in the connected speech of chronic post-stroke aphasia patients. Additionally, given the fact that the lesion load of white matter bundles has been previously identified as a predictor of outcome in chronic post-stroke aphasia (Geva et al., 2015; Hillis et al., 2018; Marchina et al., 2011; Xing et al., 2017), we aimed to determine whether it would predict syntactic abilities in chronic post-stroke aphasia.

9.2.2. Hypotheses

We expect the structural characteristics of the left AF and the UF to predict the production of verbs and verb-argument structure complexity, since previous studies support the involvement of dorsal and ventral structures in verb production (e.g., Akinina et al., 2019; den Ouden et al., 2019) and that a role for the UF in local syntactic processing, such as verb-argument structure, has been postulated (Friederici et al., 2006). We also expect the production of grammatically well-formed sentences to be predicted by the structural characteristics of both the left AF and UF, since the production of sentences involves both global and local syntactic processes which have been postulated to be supported by the dorsal and ventral pathways, respectively (Friederici et al., 2006). Finally, we expect sentence complexity to be predicted by the left AF, since previous studies

indicate that the production of complex sentences is supported by the dorsal stream (Gunawardena et al., 2010; Wilson et al., 2010).

Chapter II – Methods and Results

Article 1: Test-retest reliability of diffusion measures extracted along white matter language fiber bundles using HARDI-based tractography

Mariem Boukadi[1,2], Karine Marcotte[3,4], Christophe Bedetti[1], Jean-Christophe Houde[6], Alex Desautels[3,5], Samuel Deslauriers-Gauthier[7], Marianne Chapleau[1,2], Arnaud Boré[1], Maxime Descoteaux[6], & Simona M. Brambati[1,2*]

[1]Centre de recherche de l'Institut universitaire de gériatrie de Montréal, Montréal, QC, Canada

[2]Département de psychologie, Université de Montréal, Montréal, QC, Canada

[3]Centre de recherche du CIUSSS du Nord-de-l'île-de-Montréal, Hôpital du Sacré-Cœur de Montréal, Montréal, QC, Canada

[4]École d'orthophonie et d'audiologie, Faculté de médecine, Université de Montréal, Montréal, QC, Canada

[5]CIUSSS du Nord-de-l'île-de-Montréal, Hôpital du Sacré-Cœur de Montréal, Montréal, QC, Canada

[6]Sherbrooke Connectivity Imaging Lab, Département d'informatique, Université de Sherbrooke, Montréal, QC, Canada

[7]Université Côte d'Azur, Inria, France

Published in *Frontiers in Neuroscience: Brain Imaging Methods* (2019)

doi: https://doi.org/10.3389/fnins.2018.01055

Abstract

High angular resolution diffusion imaging (HARDI)-based tractography has been increasingly used in longitudinal studies on white matter macro- and micro-structural changes in the language network during language acquisition and in language impairments. However, test-retest reliability measurements are essential to ascertain that the longitudinal variations observed are not related to data processing. The aims of this study were to determine the reproducibility of the reconstruction of major white matter fiber bundles of the language network using anatomically-constrained probabilistic tractography with constrained spherical deconvolution based on HARDI data, as well as to assess the test-retest reliability of diffusion measures extracted along them. Eighteen right-handed participants were scanned twice, one week apart. The arcuate, inferior longitudinal, inferior fronto-occipital, and uncinate fasciculi were reconstructed in the left and right hemispheres and the following diffusion measures were extracted along each tract: fractional anisotropy, mean, axial, and radial diffusivity, number of fiber orientations, mean length of streamlines, and volume. All fiber bundles showed good morphological overlap between the two scanning timepoints and the test-retest reliability of all diffusion measures in most fiber bundles was good to excellent. We thus propose a fairly simple, but robust, HARDI-based tractography pipeline reliable for the longitudinal study of white matter language fiber bundles, which increases its potential applicability to research on the neurobiological mechanisms supporting language.

1. Introduction

The characterization of the brain and language network and its development, disruption, and changes over time represents one of the central themes of cognitive neuroscience. Diffusion magnetic resonance imaging (dMRI)-based tractography has been proven to be a valuable tool for the in vivo identification of white matter (WM) fiber bundles involved in language and the extraction of measures of their micro- and macro-structural characteristics. However, the ability of this tool to reproduce the same language fiber bundles' morphology and micro- and macro-structural characteristic measurements when dMRI data is acquired twice from the same participant under the same conditions (i.e. test-retest reliability), has yet to be clearly demonstrated. In fact, while test-retest reliability has already been reported for other neuroimaging techniques that are usually employed in evaluating longitudinal changes in the language brain network (such as resting state and task-based functional MRI, voxel-based morphometry, and cortical thickness; e.g. Birn et al., 2013; Braun et al., 2012; Jovicich et al., 2009; Lin et al., 2015; Madan & Kensinger, 2017; Powers et al., 2013; Seiger et al., 2015; Wang, Jin, Zhang, & Wang, 2016; Zhang et al., 2011; Zhang, Chen, Zhang, & Shen, 2017), test-retest reliability of dMRI-based tractography has received comparatively less attention. This represents a first necessary step to validate the use of this approach in longitudinal studies on language.

It is increasingly accepted that WM associative fiber bundles play a crucial role in mediating the transfer of information among specialized language brain areas, distributed along two main processing streams, namely the dorsal and ventral streams (Dick et al., 2014; Hickok & Poeppel, 2000, 2007; Poeppel, Emmorey, Hickok, & Pylkkänen, 2012; Saur et al., 2008). The central and most widely studied WM fiber bundle of the dorsal stream is the arcuate fasciculus (AF), putatively connecting Broca's and Wernicke's territories (Catani & Thiebaut de Schotten, 2012). The AF has

been suggested to play a central role in the processing of phonological information and complex syntax in both language production and comprehension (Brauer, Anwander, & Friederici, 2011; Brauer, Anwander, Perani, & Friederici, 2013; Duffau et al., 2002; Duffau, Gatignol, Denvil, Lopes, & Capelle, 2003; Friederici et al., 2006; Wilson et al., 2011). The major fiber bundles of the ventral stream are the inferior longitudinal fasciculus (ILF), the inferior fronto-occipital fasciculus (IFOF), and the uncinate fasciculus (UF). Their specific contribution to language processing is still a matter of debate. Both the ILF and IFOF are bundles of long association fibers originating in the occipital lobe (Dick et al., 2014). The ILF connects occipital and temporal lobes, while the IFOF connects the occipital and frontal lobes (Catani & Thiebaut de Schotten, 2008; Dick et al., 2014; Thiebaut de Schotten, Dell'Acqua, Valabregue, & Catani, 2012). Both of these bundles have been suggested to play a key role in semantic processing, more specifically in reading and naming (Duffau, 2008; Duffau et al., 2005, 2014; Gil-Robles et al., 2013; Han et al., 2013; Turken & Dronkers, 2011). The UF is a long-range association fiber bundle connecting the anterior temporal lobe with the orbital and polar frontal cortex (Thiebaut de Schotten et al., 2012). While the role of this bundle in language is still controversial, it has been suggested to support semantic retrieval (Catani & Mesulam, 2008; Grossman et al., 2004; Lu et al., 2002) and simple syntactic operations (e.g. processing of phrases) (Friederici et al., 2006).

The use of advanced probabilistic fiber tracking based on high angular resolution diffusion imaging (HARDI) has proven to be particularly suitable for the reconstruction of fiber bundles with complex configurations (i.e., crossing, kissing, or fanning fibers), such as language-related fiber bundles (Alexander, Barker, & Arridge, 2002; Descoteaux, 2015; Farquharson & Tournier, 2016; Jeurissen, Descoteaux, Mori, & Leemans, 2017; Maier-Hein et al., 2017; Tournier, Masterton, & Seitz, 2012; Tuch et al., 2002). Up until recently, diffusion tensor imaging (DTI) tractography

based on dMRI data has been considered a standard tool for the in vivo reconstruction of fiber bundles. However, it has been demonstrated that DTI fails to adequately represent the complex architecture of WM fibers in the brain (Alexander, Lee, Lazar, & Field, 2007; Behrens, Berg, Jbabdi, Rushworth, & Woolrich, 2007; Descoteaux, 2015; Descoteaux, Angelino, Fitzgibbons, & Deriche, 2007; Descoteaux, Deriche, Knösche, & Anwander, 2009; Frank, 2001; Jeurissen, Leemans, Tournier, Jones, & Sijbers, 2010; Prckovska, Descoteaux, Poupon, ter Haar Romeny, & Vilanova, 2012). Standard DTI analysis can represent only one fiber population per voxel, whereas about 66% to 90% of voxels contain a complex fiber configuration (Descoteaux & Deriche, 2008; Jeurissen et al., 2010, 2014). HARDI has been introduced to mitigate some of DTI's limitations in WM areas with a complex geometry (Alexander et al., 2002; Descoteaux, 2015; Farquharson & Tournier, 2016; Tournier et al., 2012a; Tuch et al., 2002). HARDI measures the diffusion signal along 60 or more gradient directions taken on the sphere in q-space (Descoteaux, 2015). HARDI-based reconstruction techniques such as constrained spherical deconvolution (CSD) aim to estimate the distribution of different fiber orientations within a voxel using a mathematical object known as the fiber orientation distribution function (fODF) (Seunarine & Alexander, 2014b). As opposed to the tensor, the fODF allows the estimation of more than one fiber population per voxel, which allows better characterization of WM in regions with a complex architecture (Côté et al., 2013; Descoteaux, 2015).

The combination of micro- and macro-structural measures allows a more comprehensive analysis of WM fiber bundle characteristics. Microstructural properties of bundles reconstructed with tractography are usually inferred from the extraction of different scalar metrics, such as fractional anisotropy (FA), mean diffusivity (MD), radial diffusivity (RD), and axial diffusivity (AD). These measures are sensitive to different fiber properties such as axonal ordering,

myelinization, and density (Jones, Knösche, & Turner, 2013). Although the specific interpretation of these measures is still a matter of debate, they are routinely used in both fundamental and clinical neuroscience studies to provide insights into WM fiber bundles' profile (Tournier et al., 2012b). The development of CSD based on HARDI data allows the estimation of the number of fiber orientations (NuFO) using the number of local maxima of the fiber orientation distribution (FOD) (Dell'Acqua, Simmons, Williams, & Catani, 2013). NuFO indicates the number of distinct fiber orientations in each voxel, thus providing valuable information on WM complexity. Interestingly, NuFO maps are highly consistent across individuals, which could represent a sensitive marker of age-related changes in WM complexity among healthy populations or changes observed in clinical populations (Dell'Acqua et al., 2013). Macrostructural measures provide complementary information regarding the morphology of the bundles, which includes the volume of the fiber bundles and the mean length of streamlines (MLS) (Girard, Whittingstall, Deriche, & Descoteaux, 2014).

While there is growing interest in the use of tractography and tractometry in longitudinal studies to investigate language-related fiber bundles' changes over time (e.g., Forkel et al. 2014; Lam et al. 2014; Mandelli et al. 2016; Takeuchi et al. 2016; Asaridou et al. 2017; Chow and Chang 2017), the test-retest reliability of HARDI-based tractography and tractometry for language-related fiber bundles has yet to be demonstrated. Test-retest reliability refers to the reproducibility of a measure repeated twice for the same participant (Berchtold, 2016). In order for an instrument to be used to detect a change, it has to be able to distinguish between a real change in individuals and a random variation due to the measurement instrument itself (Guyatt, Walter, & Norman, 1987). This entails that one of the most crucial aspects to look at when assessing the reliability of a method for longitudinal designs is the test-retest reliability of measurement instruments (Berchtold, 2016). To

date, most studies have not integrated reproducibility assessment of their diffusion measures and fiber bundles. This is a critical issue because different factors may affect intra-subject reproducibility such as imaging acquisition parameters (e.g. Bisdas, Bohning, Besenski, Nicholas, & Rumboldt, 2008; Gao, Zhu, & Lin, 2009; Jones, 2004), tractography pipelines (Cousineau et al., 2017; Kristo et al., 2013; Wang, Abdi, Bakhadirov, Diaz-arrastia, & Devous, 2012), and subject physiological noise (e.g. Farrell et al., 2007; Pfefferbaum, Adalsteinsson, & Sullivan, 2003). Previous studies have provided evidence of good to excellent test-retest reliability for other methods of analysis of dMRI data, such as tract-based spatial statistics (TBSS), region-of-interest (ROI)-based approaches, and DTI-tractography (Ciccarelli et al., 2003; Cole et al., 2014; Danielian, Iwata, Thomasson, & Kay, 2010; Heiervang, Behrens, Mackay, Robson, & Johansen-Berg, 2006; Magnotta et al., 2012; Vollmar et al., 2010; Wang et al., 2012). However, the test-retest reliability of HARDI-tractography and tractometry has received less attention. Promising evidence of test-retest reliability of this approach comes from the work of Cousineau et al. (2017), Besseling et al. (2012), and Kristo et al. (2013). These studies have demonstrated the overlap of WM fiber bundles reconstructed by means of HARDI-based tractography and the reproducibility of their micro- and macro-structural measures, based on dMRI data obtained from healthy subjects in separate MRI acquisition sessions. Even though these studies were crucial in determining the potential of this approach in longitudinal studies, the only language-related bundle included in all of them is the AF which yielded conflicting results. Thus, the test-retest reliability of HARDI-tractography and tractometry in the main language-related fiber bundles remains to be validated.

In order to fill this gap, the aim of the present study is to assess test-retest reliability of the reconstruction, as well as the micro- and macro-structural characteristics of the major WM fiber bundles associated with language processing reconstructed using probabilistic HARDI-

tractography. To this aim, we have collected dMRI data from a sample of healthy individuals at two time-points, one week apart, and reconstructed major WM fiber bundles supporting language functions within the left and right hemispheres (AF, ILF, IFOF, and UF). We expect that no measurable changes in the micro- and macro-structural characteristics of the tracts under study would be observed in that short time period. The test-retest reliability of the fiber bundles' morphology was obtained by calculating, for each subject, the spatial overlap between each tract's reconstruction at the two time-points as proposed in Cousineau et al. (2017). Additionally, macro-structural characteristics such as volume and MLS, as well as mean microstructural measures such as the tensor-based metrics FA, MD, RD, AD, and the FOD-based measure NuFO (Dell'Acqua et al., 2013) were extracted for each bundle and their reproducibility was assessed.

2. Methods

2.1. Participants

Eighteen right-handed cognitively-unimpaired participants (age: M = 64.61 y.o. ± 7.99; education: M = 16.16, ± 3.42 years; 9 women, 9 men) with no history of psychiatric or neurological conditions were scanned at two time-points, one week apart. The study was approved by the research ethics committee of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Ile-de Montréal (Project #MP-32-2018-1478) and written informed consent was obtained from all participants.

2.2. Image acquisition

The diffusion MRI protocol was acquired using a Skyra 3T MRI scanner (Siemens Healthcare, USA) at the radiology department of Hôpital du Sacré-Coeur of Montreal. At each of the two scanning occasions participants underwent the same acquisition sequence. One high resolution 3D T1-weighted (T1w) image (TR = 2200 ms, TE = 2.96 ms, TI = 900 ms, FOV = 250

mm, voxel size = 1x1x1 mm3, matrix = 256x256, 192 slices, flip-angle = 8°) was acquired using a Magnetization Prepared Rapid Gradient Echo (MP-RAGE) sequence. A diffusion weighted imaging (DWI) sequence was also acquired (TR = 8051 ms, TE = 86 ms, FOV= 230 mm, voxel size = 2x2x2 mm3, flip angle = 90°, bandwidth = 1698; EPI factor = 67; 68 slices in transverse orientation) with one image (b = 0 s/mm²) and 64 images with non-collinear diffusion gradients (b=1,000 s/mm²) in a posterior-anterior (PA) acquisition, as well as two additional images (b = 0 s/mm²): one in a PA acquisition, namely in the same direction as the diffusion gradients, and the other in an anterior-posterior (AP) acquisition, namely in the opposite direction of the diffusion gradients.

2.3.dMRI data analysis

All analysis steps were conducted using the Toolkit for Analysis in Diffusion MRI (TOAD) pipeline (http://www.unf-montreal.ca/toad).

2.3.1. Pre-processing

Pre-processing steps included denoising, motion/eddy/distortion corrections, upsampling, registration, segmentation and parcellation, and masking. First, DWI was noise-corrected using overcomplete local principal component analysis (PCA) using the Matlab toolbox DWI Denoising Software (Manjo, Concha, Buades, & Collins, 2013). The FMRIB Diffusion toolbox EDDY of FSL 5.0.11 (publicly available neuroimaging software: http://www.fmrib.ox.ac.uk/fsl/) (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) was used to correct all images for subject movement, eddy-currents, and susceptibility-induced distortions using AP-PA images. Gradient directions were corrected corresponding to motion correction parameters (motion for each subject at each timepoint is reported in the supplementary materials). T1w images were processed with Freesurfer's pipeline 6.0.0 (Dale, Fischl, & Sereno, 1999; Desikan et al., 2006) for segmentation

and parcellation of grey and WM into anatomical regions. DWI was upsampled to 1mm isotropic resolution using a trilinear interpolation (Dyrby et al., 2014; Girard & Descoteaux, 2012; Raffelt et al., 2012; Smith, Tournier, Calamante, & Connelly, 2012; Tournier, Calamante, & Connelly, 2012) and the segmented and parcellated T1w was registered to the DWI using FMRIB's linear registration tool (FLIRT) from FSL. This step allowed us to carry out anatomically-constrained tracking (ACT) (see next section for further details). Finally, a mask image was obtained from the segmented T1w image and served to seed streamlines on the grey matter-white matter interface (Tournier et al., 2012a).

2.3.2. Tractography

Fiber orientation distribution functions (fODFs) were estimated using CSD. A whole-brain tractogram was computed using MRtrix3's probabilistic tractography algorithm with ACT (https://github.com/jdtournier/) (Tournier et al., 2012a). ACT uses the T1w (i.e. the segmented anatomical image obtained from Freesurfer) to limit potential false-negatives (i.e. no-connections) and improve WM coverage in general (Girard & Descoteaux, 2012; Girard et al., 2014; Guevara et al., 2011; Mori & Tournier, 2014; Smith et al., 2012). The AF, ILF, IFOF, and UF were reconstructed from the tractogram using the White Matter Query Language (WMQL) (Wassermann, Makris, Rathi, Shenton, et al., 2016). WMQL is a user-friendly method to carry out WM bundle extraction from tractography in a nearly automatic way. It allows us to consistently define bundles across subjects without manually specifying regions of interest. It consists in writing queries with the WMQ language describing the WM bundles to be reconstructed using anatomical definitions from Freesurfer's Desikan/Killiany atlas. In order to be able to extract the fiber bundles using the written queries, Freesurfer's grey and WM parcellation was overlaid on the tractogram. The queries were then automatically interpreted by tractography tools. The queries used to

reconstruct the fiber bundles are presented in the supplementary materials. Outlier streamlines were then removed from each tract using a tract-filtering algorithm (Côté, Garyfallidis, Larochelle, & Descoteaux, 2015). The following diffusion and bundle measures were extracted along each fiber bundle for each participant: FA, MD, AD, RD, NuFO (Dell'Acqua et al., 2013), volume, and MLS. All tractography steps were performed in native space since non-linear normalization with diffusion MRI data requires local reorientations and warping which affects the gradient table at every voxel (bval/bvec) (Vollmar et al., 2010). Bringing the T1-w image into native diffusion space with linear affine registration and using the Freesurfer parcellation in this space is more robust (Girard, Fick, Descoteaux, Deriche, & Wassermann, 2015; Girard et al., 2014).

2.4. Statistical analysis

Test-retest analyses were carried out in two steps. First, we used the weighted dice similarity coefficient (wDSC) to determine the degree of overlap between the reconstructed fiber bundles at Times 1 and 2 as in Cousineau et al. (2017). DSC is a statistical metric that ranges between 0 and 1 and is used to assess the degree of overlap between two volumes (Dice, 1945). The wDSC is a variation of this metric and gives more weight to voxels with more streamlines. This is important to take into account considering the fact that WM bundles have more streamlines in their middle than in the extreme portions (Cousineau et al., 2017). In the two previous studies which used this measure to assess the test-retest reliability of CSD-based reconstruction of WM tracts (Cousineau et al., 2017; Besseling et al., 2012), the minimum value of Dice was .70. Therefore, this value was used as the acceptable threshold for a good wDSC in our study. The wDSC was computed using the following formula from Cousineau et al. (2017):

$$D(W_i, W_j) = \frac{\Sigma_{v'} W_{i,v'} + \Sigma_{v'} W_{j,v'}}{\Sigma_v W_{i,v'} + \Sigma_v W_{j,v}}$$

where *Wi* and *Wj* respectively represent the bundles at Time 1 and Time 2 and *v*' represents the voxels from the two reconstructions of the bundles (*Wi* and *Wj*) that overlap.

To do so, T1-weighted images in diffusion space taken at Time 1 were registered linearly to anatomical images taken at Time 2 (i.e., seven days later) for each subject with Advanced Normalization Tools (ANTs), version>=2.1 (Tustison et al., 2014) (http://stnava.github.io/ANTs/). Transformation matrices were applied to all Time 1 bundles using TractQuerier's tract_math tool (Wassermann, Makris, Rathi, Shenton, et al., 2016). Once the two fiber bundles of each subject were in the same space, wDSCs were computed with the tractometry pipeline from the Sherbrooke Connectivity Imaging Lab (SCIL) http://scil.dinf.usherbrooke.ca/?lang=fr. The right UF bundle could not be reconstructed in one participant. Analyses were therefore conducted with a sample of 17 participants for that bundle.

In a second step, we combined two complementary analyses, the intra-class correlation coefficient (ICC) and the Bland-Altman plots to assess the test-retest reliability of each of the measures extracted in each reconstructed fiber bundle. The intra-class correlation coefficient (ICC) (McGraw & Wong, 1996; Shrout & Fleiss, 1979) is a widely used statistical approach to assess agreement in test-retest reliability studies in different fields, including neuroimaging (e.g., Birn et al., 2013; Braun et al., 2012; Duan, Zhao, He, & Shu, 2015; Duda, Cook, & Gee, 2014; Zhang et al., 2011). The ICC is calculated from an analysis of variance and can be broadly defined as the ratio of between-subject variance to the total variance (including within-subject variance and residue) (Berchtold, 2016). ICC values range from 0 to 1 and can be categorized into four levels of test-retest reliability: excellent (ICC > .75), good (ICC = .60 to .74), fair (ICC = .40 to .59), and poor (ICC < .40) (Fleiss, 2003). ICC estimates and their 95% confidence intervals were calculated

using SPSS version 25 based on a single measurement, absolute-agreement, two-way mixed-effects model. The formula used for computing this ICC (McGraw & Wong, 1996) is as follows:

$$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where MS_R is the mean square for rows, MS_C is the mean square of columns, MS_E is the mean square for error, *k* is the number of measurements, and *n* is the number of subjects.

We also created Bland & Altman plots which provide a visual assessment of the agreement of the two time-points (test and retest) of each measure in all four fiber bundles bilaterally (Bland & Altman, 1999). The created graphs are scatter plots with the Y axis representing the difference between the measurements at the two timepoints and the X axis representing the mean of these measures. Good agreement between measurements at two time-points exists if 95% of the data falls within ±2 standard-deviations of the mean of differences.

3. Results

The degree of overlap was good for all four reconstructed fiber bundles (AF, ILF, IFOF, and UF, bilaterally) between Time 1 and Time 2, with wDSC values ranging between .71 and .87 (values for each fiber bundle are reported in Table 1). Figure 1 illustrates the bundle overlap for a representative subject. One must note that the figure reflects the raw bundle overlap rather than the weighted overlap represented by the wDSC which gives more weight to voxels with more streamlines.

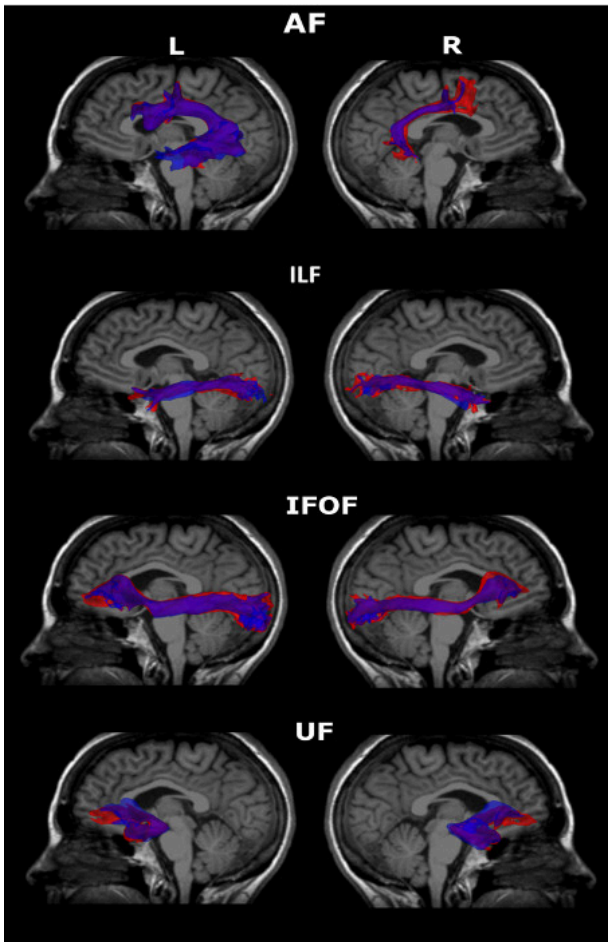


Figure 1. Overlapped 3D volume representations of the reconstructed fiber bundles at the two scanning time-points in a representative subject.
Please note that these do not reflect the wDSC values. Blue = time 1, red = time 2, purple indicates the overlap. AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus, IFOF = inferior fronto-occipital fasciculus; uncinate fasciculus; L = Left; R = Right.

Table 1 shows the ICC estimates, their 95% confidence intervals, and *p* values of the diffusion measures of interest, namely FA, MD, RD, AD, NuFO, volume, and MLS. FA, AD, MD, RD, and MLS measures showed consistently good to excellent test-retest reliability (ICCs = .62 - .95) across all four WM fiber bundles, bilaterally. Volume showed fair reliability in the right IFOF and UF (ICC = .41 - .58), and good to excellent reliability in all other bundles. NuFO showed the lowest reliability; test-retest reliability was fair in the ILF bilaterally and good in all other bundles.

Table 1. wDSC values, ICC estimates and their 95% confidence intervals for all measures and fiber bundles

		AF		*ILF*		*IFOF*		*UF*	
		Left	Right	Left	Right	Left	Right	Left	Right
	wDSC	.86	.83	.79	.71	.84	.87	.78	.83
FA	ICC	.89**	.85***	.78**	.86***	.87**	.85***	.62**	.74***
	CI	.61-.90*	.65-.92*	.52-.86*	.67-.90*	.70-.92*	.65-.93*	.22-.87*	.42-.90*
AD	ICC	.91**	.71***	.95**	.88***	.89**	.92***	.85**	.82***
	CI	.78-.96*	.38-.84*	.86-.97*	.70-.93*	.72-.94*	.79-.95*	.65-.93*	.58-.93*
MD	ICC	.91**	.86***	.95**	.90***	.84**	.86***	.83**	.80***
	CI	.79-.96*	.67-.92*	.86-.97*	.75-.94*	.62-.92*	.66-.93*	.60-.92*	.52-.92*
RD	ICC	.92**	.89***	.88**	.89***	.83**	.84***	.76**	.77***
	CI	.81-.96*	.73-.93*	.70-.97*	.73-.94*	.60-.92*	.62-.93*	.45-.91*	.48-.91*
NuFO	ICC	.62**	.68**	.50*	.56**	.62**	.63**	.61**	.69**
	CI	.22-.87*	.33-.87*	.05-.84*	.14-.87*	.22-.86*	.25-.87*	.21-.87*	.32-.87*
Volume	ICC	.83**	.79***	.79**	.75***	.70**	.58**	.71**	.41*
	CI	.61-.94*	.52-.90*	.53-.94*	.46-.89*	.14-.90*	.12-.81*	.38-.89*	-.08-.67
MLS	ICC	.71**	.87***	.84**	.89***	.70**	.69**	.68**	.82***
	CI	.37-.90*	.68-.95*	.62-.95*	.65-.95*	.35-.90*	.33-.89*	.33-.90*	.56-.93*

Note: AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; FA = fractional anisotropy; AD = axial diffusivity; MD = medial diffusivity; RD = radial diffusivity; NuFO = number of fiber orientations; MLS = mean length of streamlines; wDSC = weighted dice similarity coefficient; ICC = Intra-class correlation coefficient estimates; CI = 95% confidence intervals of the ICC.
*$p < .05$; **$p < .01$; ***$p < .001$.

In Figure 2, we only present the Bland & Altman plots created for the FA measure for the sake of brevity and clarity. Bland–Altman analysis showed high reproducibility (95% CI = .019 , -.01 for the left AF; .025, -.04 for the right AF; .03 , -.04 for the left ILF; .03 , -.03 for the right

ILF; .02, -.02 for the left IFOF; .03 , -.02 for the right IFOF; .04 , -.04 for the left UF; and .2 , -.2 for the right UF) with little difference (mean difference = .005 for the left AF, -.006 for the right AF, -.004 for the left ILF, -.003 for the right ILF, -.0008 for the left IFOF, .002 for the right IFOF, .001 for the left UF, and .03 for the right UF). A total of 88% (right AF, left and right UF) to 94% (Left AF, left and right ILF, left and right IFOF) of data points were within these limits. The 6 other plots are reported in the supplementary materials. All plots were consistent with the ICC analyses.

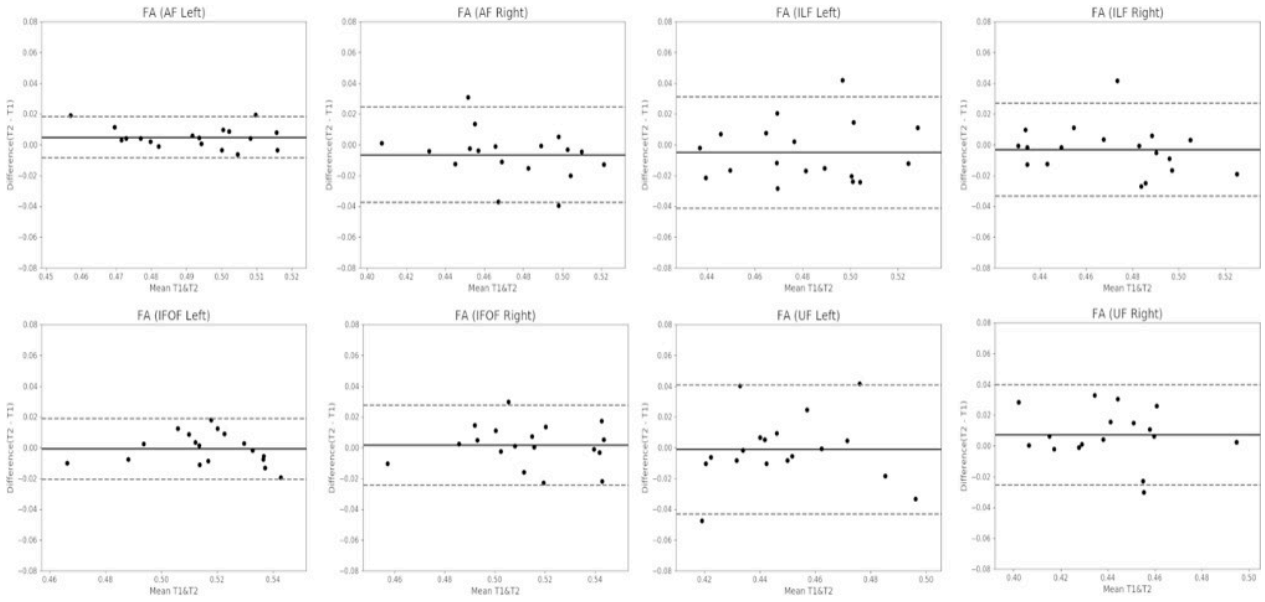


Figure 2. Bland-Altman Plots for the FA metric in all four fiber bundles, bilaterally.
The Y axis represents the mean difference between the measurements at the two timepoints and the X axis represents the mean of these measures. The upper and lower dashed lines represent the two limits of agreements at ±2 standard-deviations of the mean of differences (i.e. the 95% confidence interval). The solid line represents the mean of the differences between the two timepoints. The dots represent the individual subjects. FA = fractional anisotropy; AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; T1 = time 1; T2 = time 2.

4. Discussion

The aim of this study was to demonstrate the test-retest reliability of the reconstruction and micro- and macro-structural characteristics of major WM language fiber bundles using probabilistic CSD-tractography based on HARDI data. The dMRI data were obtained on a group of healthy subjects at two timepoints, spanning one week. First, the results demonstrated that all the reconstructed fiber bundles have a good overlap between the two timepoints. Secondly, tract-specific measures usually used in studying microstructural WM characteristics, such as FA, MD, RD, and AD, as well as the macrostructural measure MLS showed good to excellent test-retest reliability in the AF, ILF, IFOF and UF, bilaterally. Volume, another macrostructural property, showed good to excellent reproducibility for some fiber bundles (AF, ILF, as well as the IFOF and UF in the left hemisphere) but only fair reproducibility for others (IFOF and UF in the right hemisphere). NuFO showed good test-retest reliability for all fiber bundles, except the ILF which showed only fair test-retest reliability. Our results agree with and, at the same time, critically expand on previous studies that investigated the test-retest reliability of probabilistic CSD-tractography (Cousineau et al., 2017; Besseling et al., 2012). These results represent a first necessary validation protocol for longitudinal studies in research in the cognitive neuroscience of language. Assessing test-retest reliability of the reconstruction of fiber bundles and of their micro- and macro-structural measures is of paramount importance for the use of this approach in longitudinal studies, as it allows to ascertain that the observed variations truly reflect the changes that may take place in WM over time and are not due to the variability inherent to dMRI data processing, instead (Cousineau et al., 2017).

Diffusion MRI tractography is presently the only method that allows the reconstruction of WM fiber bundles in-vivo. For this reason, in the last decades it has gained tremendous popularity

in the field of neuroscience and its potential to map the human connectome is widely recognized. In the last years, an increasing number of big data initiatives has been developed in order to collect longitudinal dMRI data in healthy individuals with the ultimate goal to describe the changes of dMRI over the lifespan and to link these changes to cognitive performance (Howell et al., 2017). In order to fully benefit from the potential of longitudinal dMRI data, it is necessary to demonstrate the test-retest reliability of dMRI-based tractography. The present work provides critical information to investigate two important questions. When we obtain dMRI data in two separate acquisition sessions one week apart in the same subjects, using probabilistic CSD-tractography with ACT and tractometry based on HARDI data, can we 1) reconstruct overlapping WM language fiber bundles? and 2) extract similar micro-and macro-structural measure values?

Regarding the first question, our data seem to provide an affirmative response. The obtained wDSC values determining the degree of overlap of the bundles reconstructed at the two timepoints using probabilistic CSD-tractography ranged between 0.71 and 0.87. Based on the minimum value (.70) of Dice found in the two studies which used this metric to assess the test-retest reliability of CSD-based reconstruction of WM tracts (Cousineau et al., 2017; Besseling et al., 2012), our wDSC values indicate that all the fiber bundles investigated in the present study have good test-retest reliability. In addition, our wDSC values are consistent with the values obtained in previous studies aimed at validating test-retest reliability of probabilistic CSD-tractography in other fiber bundles, such as the cingulum, optic radiation, and the corpus callosum (Cousineau et al., 2017; Besseling et al., 2012). We also report excellent overlap for the AF and IFOF, which is consistent with Cousineau et al. (2017) who used a similar tractography pipeline. The overlap obtained in our study is greater than what has been reported by Besseling et al. (2012) in which, in order to reconstruct the AF, they only used seed and target ROIs. Considering the complex anatomy of the AF, a

tractography method allowing the use of more specific anatomical priors, as we did in the present study, might improve the reconstruction of this complex tract and thus allow for a better reproducibility of its morphology.

Our study also confirms the reproducibility of tensor metrics and MLS. Test-retest reliability of tensor metrics (FA, MD, RD, and AD) has been previously studied using DTI-based tractography (Buchanan, Pernet, Gorgolewski, Storkey, & Bastin, 2014; Ciccarelli et al., 2003; Danielian et al., 2010; Heiervang et al., 2006; Vollmar et al., 2010; Wang et al., 2012). However, studies using this approach have not always reported satisfactory results. For example, in one study, poor test-retest reliability was observed with AD across all studied fiber bundles (Danielian et al.,2010), whereas others reported tract-specific variability of the reproducibility of FA and MD (Wang et al., 2012). There are several sources of variability in diffusion MRI that can affect the test-retest reliability of tractography or the measures of WM structural characteristics (Danielian et al., 2010). These include, but are not limited to, partial volume effects introduced by the DTI model, bad anatomical priors, as well as potential inter- and intra-rater reliability of ROI placement in seed-based approaches for tractography (Wakana et al., 2007; Danielian et al., 2010; Cousineau et al., 2017). In the present study, we used an approach that attempts to reduce variability from these sources by using HARDI-based state-of-the-art tracking algorithms based on ACT and probabilistic tracking algorithms which have the potential to yield fuller, longer bundles that better reach the cortex (Mori and Tournier 2014; Maier-Hein et al., 2017), novel approaches to extract the bundles from the tractogram (i.e. WMQL), as well as good anatomical priors (Catani, Howard, Pajevic, & Jones, 2002; Conturo et al., 1999; Hagmann et al., 2003; Huang, Zhang, Van Zijl, & Mori, 2004; Wakana et al., 2007). Using this approach, we were able to demonstrate good to excellent reliability of all tensor-based metrics which are the microstructural measures most

commonly used in dMRI studies on language, and MLS, a macrostructural measure, in all language fiber bundles. This represents an important step towards the validation of this approach in the longitudinal study of language fiber bundles. On the other hand, the test-retest reliability of NuFO was less than good in some tracts. To the best of our knowledge, no previous study has investigated the test-retest reliability of this measure. Our results seem to encourage further longitudinal validation of this measure before adopting it in longitudinal studies. Additionally, our results for the volume, another macrostructural measure, were consistent with previous studies that reported inconsistent test-retest reliability for this measure across fiber bundles, using probabilistic CSD-tractography (Besseling et al., 2012) or DTI-based tractography (Wang et al., 2012).

Even though the present results are very promising, particularly for the tensor metrics and MLS, future studies should be designed in order to confirm our findings. First, these results should be reproduced in larger groups. Secondly, the use of CSD allows to resolve multiple fiber orientations at reasonable angles with a properly data-driven response function at lower b-values and 64 directions as in the present study (Descoteaux et al., 2009; Raffelt et al., 2012; Tournier, Calamante, & Connelly, 2007). Nevertheless, utilizing multi-b-value sequences, such as b = 1000 s/mm², b = 2000 s/mm², b = 3000 s/mm², or b = 1000 s/mm² and b = 3000 s/mm², could help to interpret the differences obtained in the present study by considering other available measures, such as intracellular, extracellular, and isotropic volume (Raffelt et al., 2012).

In conclusion, in an era where initiatives to collect dMRI longitudinal data are multiplying and fiber tracking is considered one of the most popular tools to follow changes in the language network over time, the question of test-retest reliability of dMRI tractography is of paramount importance. Our study provides critical evidence indicating the test-retest reliability of probabilistic CSD-tractography. As in previous studies which demonstrated test-retest reliability of TBSS or

DTI-tractography (e.g., Forkel et al., 2014; Kitamura, Kiuchi, & Taoka, 2013; Poudel et al., 2015), the present results support the use of probabilistic CSD-tractography to study language fiber bundles in longitudinal studies in healthy and clinical populations interested in language related fiber bundles.

Conflict of interest

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

Author contributions

MB drafted the manuscript, contributed to the design of the study, reviewed the literature, and collected, analyzed and interpreted the data. SB and KM designed the study, contributed to data collection, supervised data analysis and interpretation, and contributed to the drafting of the manuscript. MD contributed to the design of the study and to the development of the tractography pipeline. AD contributed to the design of the study and MRI data acquisition. CB and MC helped with the data analysis. CB, MD, AB, JH, SD developed the tractography pipeline. All authors revised the final version of the manuscript.

Funding

This work was supported by doctoral awards from the Canadian Institutes of Health Research -Frederick Banting and Charles Best Canada Graduate Scholarships and the Fonds de Recherche du Québec – Santé (FRQS) in partnership with Heart and Stroke Foundation – Quebec to M.B. (grant numbers: N.A.). S.M.B. and K.M. hold a Career Award from FRQS (grant number: N.A.). This work was also supported by grants from Réseau Québécois de la Recherche sur le Vieillissement (RQRV) to S.M.B. (grant number: N.A.); Heart and Stroke Foundation of Canada

to S.M.B. and K.M. (grant number: G-16-00014039); the National Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant program (grant number: RGPIN-2015-05297) and the institutional Université de Sherbrooke Research Chair in Neuroinformatics to M.D. (grant number: N.A.).

Acknowledgments

We would like to thank the participants who accepted to take part of this study. We also would like to thank Pamela Ross, Audrey Borgus and Amélie Brisebois for their valuable assistance with the recruitment and testing of participants. We also thank the radiology team at Hôpital du Sacré-Coeur de Montréal (HSCM) for their support with the scanning of participants.

References

Alexander, A. L., Lee, J. E., Lazar, M., & Field, A. S. (2007). Diffusion tensor imaging of the brain. *Neurotherapeutics, 4*(3), 316–329.

Alexander, D. C., Barker, G. J., & Arridge, S. R. (2002). Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine, 48*, 331–340. https://doi.org/10.1002/mrm.10209

Asaridou, S. S., Demir-Lira, O. E., Goldin-Meadow, S., & Small, S. L. (2017). The pace of vocabulary growth during preschool predicts cortical structure at school age. *Neuropsychologia, 98*, 13–23. https://doi.org/10.1016/j.neuropsychologia.2016.05.018

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage, 34*(1), 144–155. https://doi.org/10.1016/j.neuroimage.2006.09.018

Berchtold, A. (2016). Test-retest: Agreement or reliability? *Methodological Innovations, 9*, 1–7. https://doi.org/10.1177/2059799116672875

Besseling, M. H., Jansen, J. F. A., Overvliet, G. M., Vaessen, M. J., Braakman, H. M. H., Hofman, P. A. M., … Backes, W. H. (2012). Tract Specific Reproducibility of Tractography Based Morphology and Diffusion Metrics. *Plos One, 7*(4). https://doi.org/10.1371/journal.pone.0034125

Birn, R., Molloy, E., Patriat, R., Parker, T., Meier, T., Kirk, G., … Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage, 83*, 550–558. https://doi.org/10.1111/j.1743-6109.2008.01122.x.Endothelial

Bisdas, S., Bohning, D. E., Besenski, N., Nicholas, J. S., & Rumboldt, Z. (2008). Reproducibility, interrater agreement, and age-related changes of fractional anisotropy measures at 3T in healthy subjects: Effect of the applied b-value. *AJNR, 29,* 1128–1133. https://doi.org/10.3174/ajnr.A1044

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Brauer, J., Anwander, A., & Friederici, A. D. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cerebral Cortex, 21*(2), 459–466. https://doi.org/10.1093/cercor/bhq108

Brauer, J., Anwander, A., Perani, D., & Friederici, A. D. (2013). Dorsal and ventral pathways in language development. *Brain and Language, 127*(2), 289–295. https://doi.org/10.1016/j.bandl.2013.03.001

Braun, U., Plichta, M. M., Esslinger, C., Sauer, C., Haddad, L., Grimm, O., … Meyer-Lindenberg, A. (2012). Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures. *NeuroImage, 59*(2), 1404–1412. https://doi.org/10.1016/j.neuroimage.2011.08.044

Buchanan, C. R., Pernet, C. R., Gorgolewski, K. J., Storkey, A. J., & Bastin, M. E. (2014). NeuroImage Test – retest reliability of structural brain networks from diffusion MRI. *NeuroImage, 86*, 231–243. https://doi.org/10.1016/j.neuroimage.2013.09.054

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*, 77–94. https://doi.org/10.1006/nimg.2002.1136

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex, 44*(8), 953–961. https://doi.org/10.1037/a0030561

Catani, M., & Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual in vivo dissections. *Cortex, 44*(8), 1105–1132. https://doi.org/10.1016/j.cortex.2008.05.004

Catani, M., & Thiebaut de Schotten, M. (2012). *Atlas of Human Brain Connections.* New York, NY: Oxford University Press.

Chow, H. M., & Chang, S.-E. (2017). White matter developmental trajectories associated with persistence and recovery of childhood stuttering. *Human Brain Mapping.* https://doi.org/10.1002/hbm.23590

Ciccarelli, O., Parker, G. J. M., Toosy, A. T., Wheeler-Kingshott, C. A. M., Barker, G. J., Boulby, P. A., … Thompson, A. J. (2003). From diffusion tractography to quantitative white matter tract measures: A reproducibility study. *NeuroImage, 18*(2), 348–359. https://doi.org/10.1016/S1053-8119(02)00042-3

Cole, J. H., Farmer, R. E., Rees, E. M., Johnson, H. J., Frost, C., Scahill, R. I., & Hobbs, N. Z. (2014). Test-Retest Reliability of Diffusion Tensor Imaging in Huntington's Disease. *PLoS Currents, 6*, 1–20. https://doi.org/10.1371/currents.hd.f19ef63fff962f5cd9c0e88f4844f43b

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., … Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences*, *96*(18), 10422–10427. https://doi.org/10.1073/pnas.96.18.10422

Côté, M. A., Garyfallidis, E., Larochelle, H., & Descoteaux, M. (2015). Cleaning up the mess: tractography outlier removal using hierarchical QuickBundles clustering. In *Proceedings of ISMRM*, vol. 23 (p. 2844).

Côté, M. A., Girard, G., Boré, A., Garyfallidis, E., Houde, J. C., & Descoteaux, M. (2013). Tractometer: Towards validation of tractography pipelines. *Medical Image Analysis, 17*(7), 844–857. https://doi.org/10.1016/j.media.2013.03.009

Cousineau, M., Jodoin, P., Garyfallidis, E., Côté, M., Morency, F. C., Rozanski, V., … Descoteaux, M. (2017). A test-retest study on Parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical, 16*(July), 222–233. https://doi.org/10.1016/j.nicl.2017.07.020

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage, 194*, 179–194.

Danielian, L. E., Iwata, N. K., Thomasson, D. M., & Kay, M. (2010). Reliability of fiber tracking measurements in diffusion tensor imaging for longitudinal study. *NeuroImage, 49*(2), 1572–1580. https://doi.org/10.1016/j.neuroimage.2009.08.062

Dell'Acqua, F. D., Simmons, A., Williams, S. C. R., & Catani, M. (2013). Can spherical deconvolution provide more information than fiber orientations ? Hindrance modulated

orientational anisotropy, a true-tract specific index to characterize white matter diffusion. *Human Brain Mapping, 34*, 2464–2483. https://doi.org/10.1002/hbm.22080

Descoteaux, M. (2015). High angular resolution diffusion imaging (HARDI). In *Wiley Encyclopedia of Electrical and Electronics Engineering.* https://doi.org/10.1002/047134608X.W8258

Descoteaux, M., Angelino, E., Fitzgibbons, S., & Deriche, R. (2007). Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine, 58*(3), 497–510. https://doi.org/10.1002/mrm.21277

Descoteaux, M., & Deriche, R. (2008). Mapping neuronal fiber crossings in the human brain. *SPIE Newsroom*, 2–4. https://doi.org/10.1117/2.1200807.1205

Descoteaux, M., Deriche, R., Knösche, T. R., & Anwander, A. (2009). Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE Transactions on Medical Imaging, 28*(2), 269–286. https://doi.org/10.1109/TMI.2008.2004424

Desikan, R. S., Se, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage, 31*, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Dice, L. R. . (1945). Measures of the amount of ecologic association between species. *Ecology, 26*(3), 297–302. https://doi.org/10.2307/1932409

Dick, A. S., Bernal, B., & Tremblay, P. (2014). The Language Connectome: New Pathways, New Concepts. *The Neuroscientist, 20*(5), 453–467. https://doi.org/10.1177/1073858413513502

Duan, F., Zhao, T., He, Y., & Shu, N. (2015). Test-retest reliability of diffusion measures in cerebral white matter: A multiband diffusion MRI study. *Journal of Magnetic Resonance Imaging, 42*(4), 1106–1116. https://doi.org/10.1002/jmri.24859

Duda, J. T., Cook, P. A., & Gee, J. C. (2014). Reproducibility of graph metrics of human brain structural networks. *Frontiers in Neuroinformatics, 8*(46), 1–11. https://doi.org/10.3389/fninf.2014.00046

Duffau, H. (2008). The anatomo-functional connectivity of language revisited. New insights provided by electrostimulation and tractography. *Neuropsychologia, 46*(4), 927–934. https://doi.org/10.1016/j.neuropsychologia.2007.10.025

Duffau, H., Capelle, L., Sichez, N., Denvil, D., Lopes, M., Sichez, J.-P., … Fohanno, D. (2002). Intraoperative mapping of the subcortical language pathways using direct stimulations. An anatomo-functional study. *Brain, 125*(Pt 1), 199–214.

Duffau, H., Gatignol, P., Denvil, D., Lopes, M., & Capelle, L. (2003). The articulatory loop: study of the subcortical connectivity by electrostimulation. *Neuroreport, 14*(15), 2005–2008. https://doi.org/10.1097/01.wnr.0000094103.16607.9f

Duffau, H., Gatignol, P., Mandonnet, E., Peruzzi, P., Tzourio-Mazoyer, N., & Capelle, L. (2005). New insights into the anatomo-functional connectivity of the semantic system: A study using cortico-subcortical electrostimulations. *Brain, 128*(4), 797–810. https://doi.org/10.1093/brain/awh423

Duffau, H., Moritz-Gasser, S., & Mandonnet, E. (2014). A re-examination of neural basis of language processing: Proposal of a dynamic hodotopical model from data provided by brain stimulation mapping during picture naming. *Brain and Language, 131*, 1–10. https://doi.org/10.1016/j.bandl.2013.05.011

Dyrby, T. B., Lundell, H., Burke, M. W., Reislev, N. L., Paulson, O. B., Ptito, M., & Siebner, H. R. (2014). Interpolation of diffusion weighted imaging datasets. *NeuroImage, 103*, 202–213. https://doi.org/10.1016/j.neuroimage.2014.09.005

Farquharson, S., & Tournier, J.-D. (2016). High Angular Resolution Diffusion Imaging. In W. Van Hecke, L. Emsell, & S. Sunaert (Eds.), *Diffusion Tensor Imaging* (pp. 383–406). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3118-7

Farrell, J. A. D., Landman, B. A., Jones, C. K., Smith, S. A., Prince, J. L., Van Zijl, P. C. M., & Mori, S. (2007). Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging-derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *Journal of Magnetic Resonance Imaging, 26*(3), 756–767. https://doi.org/10.1002/jmri.21053

Fleiss, J. L. (2003). The measurement of interrater agreement. In J. L. Fleiss, B. Levin, & M. C. Paik (Eds.), *Statistical Methods for Rates and Proportions* (pp. 598–626). New Jersey, NY: John Wiley & Sons Inc.

Forkel, S. J., De Schotten, M. T., Dell'Acqua, F., Kalra, L., Murphy, D. G. M., Williams, S. C. R., & Catani, M. (2014). Anatomical predictors of aphasia recovery: A tractography study of bilateral perisylvian language networks. *Brain, 137*(7), 2027–2039. https://doi.org/10.1093/brain/awu113

Frank, L. R. (2001). Anisotropy in high angular resolution diffusion-weighted MRI. *Magn Reson Med, 45*(6), 935–939. https://doi.org/10.1002/mrm.1125 [pii]

Friederici, A. D., Bahlmann, J. r., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*(7), 2458–2463. https://doi.org/10.1073/pnas.0509389103

Gao, W., Zhu, H., & Lin, W. (2009). An unified optimization approach for diffusion tensor imaging technique. *NeuroImage, 44*(3), 729–741. https://doi.org/10.2217/nnm.12.167.Gene

Gil-Robles, S., Carvallo, A., Jimenez, M. D. M., Gomez Caicoya, A., Martinez, R., Ruiz-Ocaña, C., & Duffau, H. (2013). Double dissociation between visual recognition and picture naming: A study of the visual language connectivity using tractography and brain stimulation. *Neurosurgery, 72*(4), 678–686. https://doi.org/10.1227/NEU.0b013e318282a361

Girard, G., & Descoteaux, M. (2012). Anatomical tissue probability priors for tractography. In *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'12) - Computational Diffusion MRI Workshop* (pp. 174–185). Nice.

Girard, G., Fick, R., Descoteaux, M., Deriche, R., & Wassermann, D. (2015). AxTract: Microstructure-Driven Tractography Based on the Ensemble Average Propagator. *Inf. Proc. Med. Imaging, 24*, 675–686. doi: 10.1007/978-3-319-19992-4_53

Girard, G., Whittingstall, K., Deriche, R., & Descoteaux, M. (2014). Towards quantitative connectivity analysis: Reducing tractography biases. *NeuroImage, 98*, 266–278. https://doi.org/10.1016/j.neuroimage.2014.04.074

Grossman, M., McMillan, C., Moore, P., Ding, L., Glosser, G., Work, M., & Gee, J. (2004). What's in a name: Voxel-based morphometric analyses of MRI and naming difficulty in Alzheimer's disease, frontotemporal dementia and corticobasal degeneration. *Brain, 127*(3), 628–649. https://doi.org/10.1093/brain/awh075

Guevara, P., Duclap, D., Rivière, D., Cointepas, Y., Poupon, C., & Mangin, J. (2011). Accurate tractography propagation mask using T1-weighted data rather than FA. In *Proceedings of the 19th Scientific Meeting International Society for Magnetic Resonance in Medicine*, Montreal.

Guyatt, G., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Disorders, 40*(2), 171–178. https://doi.org/10.1016/0021-9681(87)90069-5

Hagmann, P., Thiran, J. P., Jonasson, L., Vandergheynst, P., Clarke, S., Maeder, P., & Meuli, R. (2003). DTI mapping of human brain connectivity: Statistical fibre tracking and virtual dissection. *NeuroImage*, *19*(3), 545–554. https://doi.org/10.1016/S1053-8119(03)00142-3

Han, Z., Ma, Y., Gong, G., He, Y., Caramazza, A., & Bi, Y. (2013). White matter structural connectivity underlying semantic processing: Evidence from brain damaged patients. *Brain, 136*(10), 2952–2965. https://doi.org/10.1093/brain/awt205

Heiervang, E., Behrens, T. E. J., Mackay, C. E., Robson, M. D., & Johansen-Berg, H. (2006). Between session reproducibility and between subject variability of diffusion MR and tractography measures. *NeuroImage, 33*(3), 867–877. https://doi.org/10.1016/j.neuroimage.2006.07.037

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4*(4), 131–138. https://doi.org/10740277

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience, 8*(May), 393–402. https://doi.org/10.1038/nrn2113

Howell, B. R., Styner, M. A., Gao, W., Yap, P. T., Wang, L., Baluyot, K., … Elison, J. T. (2017). The UNC/UMN Baby Connectome Project (BCP): An overview of the study design and protocol development. *NeuroImage, 185*, 891-905. https://doi.org/10.1016/j.neuroimage.2018.03.049

Huang, H., Zhang, J., Van Zijl, P. C. M., & Mori, S. (2004). Analysis of noise effects on DTI-based tractography using the brute-force and multi-ROI approach. *Magnetic Resonance in Medicine*, *52*(3), 559–565. https://doi.org/10.1002/mrm.20147

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jeurissen, B., Descoteaux, M., Mori, S., & Leemans, A. (2017). Diffusion MRI fiber tractography of the brain. *NMR in Biomedicine*. https://doi.org/10.1002/nbm.3785

Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., & Sijbers, J. (2010). Estimating the number of fiber orientations in diffusion MRI voxels: a constrained spherical deconvolution study. In *Proceedings of the International Society for Magnetic Resonance in Medicine.*

Jeurissen, B., Tournier, J. D., Dhollander, T., Connelly, A., & Sijbers, J. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage, 103*, 411–426. https://doi.org/10.1016/j.neuroimage.2014.07.061

Jones, D. K. (2004). The Effect of Gradient Sampling Schemes on Measures Derived from Diffusion Tensor MRI: A Monte Carlo Study. *Magnetic Resonance in Medicine, 51*(4), 807–815. https://doi.org/10.1002/mrm.20033

Jones, D. K., Knösche, T. R., & Turner, R. (2013). White matter integrity, fiber count, and other fallacies: The do's and don'ts of diffusion MRI. *NeuroImage, 73*, 239–254. https://doi.org/10.1016/j.neuroimage.2012.06.081

Jovicich, J., Czanner, S., Han, X., Salat, D., Kouwe, A. Van Der, Quinn, B., … Dale, A. (2009). MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *NeuroImage, 46*(1), 177–192. https://doi.org/10.1016/j.neuroimage.2009.02.010.MRI-derived

Kitamura, S., Kiuchi, K., & Taoka, T. (2013). Longitudinal white matter changes in Alzheimer's disease: A tractography-based analysis study. *Brain Research, 1515*, 12–18. https://doi.org/10.1016/j.brainres.2013.03.052

Kristo, G., Leemans, A., Raemaekers, M., Rutten, G. J., De Gelder, B., & Ramsey, N. F. (2013). Reliability of two clinically relevant fiber pathways reconstructed with constrained spherical deconvolution. *Magnetic Resonance in Medicine, 70*(6), 1544–1556. https://doi.org/10.1002/mrm.24602

Lam, B. Y. K., Halliday, G. M., Irish, M., Hodges, J. R., & Piguet, O. (2014). Longitudinal white matter changes in frontotemporal dementia subtypes. *Human Brain Mapping, 35*(7), 3547–3557. https://doi.org/10.1002/hbm.22420

Lin, Q., Dai, Z., Xia, M., Han, Z., Huang, R., Gong, G., … He, Y. (2015). A connectivity-based test-retest dataset of multi-modal magnetic resonance imaging in young healthy adults. *Scientific Data, 2*. https://doi.org/10.1038/sdata.2015.56

Lu, L. H., Crosson, B., Nadeau, S. E., Heilman, K. M., Gonzalez-Rothi, L. J., Raymer, A., … Roper, S. N. (2002). Category-specific naming deficits for objects and actions: Semantic attribute and grammatical role hypotheses. *Neuropsychologia, 40*(9), 1608–1621. https://doi.org/10.1016/S0028-3932(02)00014-3

Madan, C. R., & Kensinger, E. A. (2017). Test–retest reliability of brain morphology estimates. *Brain Informatics, 4*(2), 107–121. https://doi.org/10.1007/s40708-016-0060-4

Magnotta, V. A., Matsui, J. T., Liu, D., Johnson, H. J., Long, J. D., Bolster, B. D., … Paulsen, J. S. (2012). MultiCenter Reliability of Diffusion Tensor Imaging. *Brain Connectivity, 2*(6), 345–355. https://doi.org/10.1089/brain.2012.0112

Maier-Hein, K. H., Neher, P. F., Houde, J. C., Côté, M. A., Garyfallidis, E., Zhong, J., … Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications, 8*(1). https://doi.org/10.1038/s41467-017-01285-x

Mandelli, M. L., Vilaplana, E., Brown, J. A., Hubbard, H. I., Binney, R. J., Attygalle, S., … Gorno-Tempini, M. L. (2016). Healthy brain connectivity predicts atrophy progression in non-fluent variant of primary progressive aphasia. *Brain, 139*(Pt 10), 2778–2791. https://doi.org/10.1093/brain/aww195

Manjo, V., Concha, L., Buades, A., & Collins, D. L. (2013). Diffusion Weighted Image Denoising Using Overcomplete Local PCA. *PLOS One, 8*(9), 1–12. https://doi.org/10.1371/journal.pone.0073021

McGraw, K., & Wong, S. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46.

Mori, S., & Tournier, J.-D. (2014). *Introduction to Diffusion Tensor Imaging and Higher-Order Models.* (S. Mori & J.-D. Tournier, Eds.). Oxford, UK: Academic Press.

Pfefferbaum, A., Adalsteinsson, E., & Sullivan, E. V. (2003). Replicability of Diffusion Tensor Imaging Measurements of Fractional Anisotropy and Trace in Brain. *Journal of Magnetic Resonance Imaging, 433*, 427–433. https://doi.org/10.1002/jmri.10377

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *The Journal of Neuroscience, 32*(41), 14125–14131. https://doi.org/10.1523/JNEUROSCI.3244-12.2012

Poudel, G. R., Stout, J. C., Domínguez D., J. F., Churchyard, A., Chua, P., Egan, G. F., & Georgiou-Karistianis, N. (2015). Longitudinal change in white matter microstructure in Huntington's disease: The IMAGE-HD study. *Neurobiology of Disease, 74*(1), 406–412. https://doi.org/10.1016/j.nbd.2014.12.009

Powers, J. P., McMillan, C. T., Brun, C. C., Yushkevich, P. A., Zhang, H., Gee, J. C., & Grossman, M. (2013). White matter disease correlates with lexical retrieval deficits in primary progressive aphasia. *Frontiers in Neurology, 4*(212), 1–9. https://doi.org/10.3389/fneur.2013.00212

Prckovska, V., Descoteaux, M., Poupon, C., ter Haar Romeny, B. M., & Vilanova, A. (2012). Classification study of DTI and HARDI anisotropy measures for HARDI data simplification. In *New Developments in the Visualization and Processing of Tensor Fields* (pp. 1–23). Berlin: Springer.

Raffelt, D., Tournier, J., Rose, S., Ridgway, G. R., Henderson, R., Crozier, S., … Connelly, A. (2012). Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted

magnetic resonance images. *NeuroImage, 59*(4), 3976–3994. https://doi.org/10.1016/j.neuroimage.2011.10.045

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., … Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America, 105*(46), 18035–18040. https://doi.org/10.1073/pnas.0805234105

Seiger, R., Hahn, A., Hummer, A., Kranz, G. S., Ganger, S., Küblböck, M., … Lanzenberger, R. (2015). Voxel-based morphometry at ultra-high fields. A comparison of 7T and 3T MRI data. *NeuroImage, 113*, 207–216. https://doi.org/10.1016/j.neuroimage.2015.03.019

Seunarine, K. K., & Alexander, D. C. (2009). Multiple fibers: Beyond the diffusion tensor. In H. Johansen-Berg & T. E. J. Behrens (Eds.), *Diffusion MRI: From Quantitative Measurements to in Vivo Neuroanatomy* (2nd edition, pp. 105–123). London, UK: Academic Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability.1. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979, 86:420–8. *Psychological Bulletin, 86*(2), 420–428. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18839484

Smith, R. E., Tournier, J., Calamante, F., & Connelly, A. (2012). Anatomically-constrained tractography : Improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage, 62*(3), 1924–1938. https://doi.org/10.1016/j.neuroimage.2012.06.005

Takeuchi, H., Taki, Y., Hashizume, H., Asano, K., Asano, M., Sassa, Y., … Kawashima, R. (2016). Impact of reading habit on white matter structure: Cross-sectional and longitudinal analyses. *NeuroImage, 133*, 378–389. https://doi.org/10.1016/j.neuroimage.2016.03.037

Thiebaut de Schotten, M., Dell'Acqua, F., Valabregue, R., & Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex, 48*(1), 82–96. https://doi.org/10.1016/j.cortex.2011.10.001

Tournier, J., Calamante, F., & Connelly, A. (2012a). MRtrix : Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology, 22*, 53–66. https://doi.org/10.1002/ima.22005

Tournier, J. D., Calamante, F., & Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage, 35*(4), 1459–1472. https://doi.org/10.1016/j.neuroimage.2007.02.016

Tournier, J. D., Masterton, R. A. J., & Seitz, R. J. (2012b). Imaging techniques provide new insights. In L. M. Carey (Ed.), *Stroke Rehabilitation: Insights from Neuroscience and Imaging* (pp. 35–53). New York, NY: Oxford University Press.

Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., & Wedeen, V. J. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine, 48*, 577–582. https://doi.org/10.1002/mrm.10268

Turken, A. U., & Dronkers, N. F. (2011). The Neural Architecture of the Language Comprehension Network: Converging Evidence from Lesion and Connectivity Analyses. *Frontiers in System Neuroscience, 5*(1), 1–20. https://doi.org/10.3389/fnsys.2011.00001

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., … Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage, 99*, 166–179. https://doi.org/10.1016/j.neuroimage.2014.05.044

Vollmar, C., Muircheartaigh, J. O., Barker, G. J., Symms, M. R., Thompson, P., Kumari, V., … Koepp, M. J. (2010). Identical, but not the same : Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3 . 0 T scanners. *NeuroImage, 51*(4), 1384–1394. https://doi.org/10.1016/j.neuroimage.2010.03.046

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., … Mori, S. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *NeuroImage*, *36*(3), 630–644. https://doi.org/10.1016/j.neuroimage.2007.02.049

Wang, H., Jin, X., Zhang, Y., & Wang, J. (2016). Single-subject morphological brain networks: Connectivity mapping, topological characterization and test-retest reliability. *Brain and Behavior, 6*(4), 1–21. https://doi.org/10.1002/brb3.448

Wang, J. Y., Abdi, H., Bakhadirov, K., Diaz-arrastia, R., & Devous, M. D. (2012). A comprehensive reliability assessment of quantitative diffusion tensor tractography. *NeuroImage, 60*(2), 1127–1138. https://doi.org/10.1016/j.neuroimage.2011.12.062

Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., & Westin, C.-F. (2016). The white matter query language: a novel approach for describing human white matter anatomy. *Brain Structure and Function, 221,* 4705–4721. https://doi.org/10.1007/s00429-015-1179-4

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., … Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron, 72*(2), 397–403. https://doi.org/10.1016/j.neuron.2011.09.014

Zhang, H., Chen, X., Zhang, Y., & Shen, D. (2017). Test-retest reliability of "high-order" functional connectivity in young healthy adults. *Frontiers in Neuroscience, 11*. https://doi.org/10.3389/fnins.2017.00439

Zhang, H., Zhang, Y.-J., Duan, L., Ma, S.-Y., Lu, C.-M., & Zhu, C.-Z. (2011). Is resting-state functional connectivity revealed by functional near-infrared spectroscopy test-retest reliable? *Journal of Biomedical Optics, 16*, 67008–67009. https://doi.org/10.1117/1.3591020

Article 2: Early white matter predictors of syntactic ability in connected speech in chronic post-stroke aphasia

Mariem Boukadi[1,2], Karine Marcotte[3,4], Amélie Brisebois[3,4], Christophe Bedetti[1,2], Bérengère Houzé[1,2], Johémie Boucher[1,2], Alex Desautels[4,5], & Simona M. Brambati[1,2]

[1]Department of Psychology, Université de Montréal

[2]Research Center, Institut universitaire de gériatrie de Montréal (CRIUGM)

[3]École d'orthophonie et d'audiologie, Faculté de médecine, Université de Montréal, Montréal, QC, Canada

[4]CIUSSS Nord-de-l'île-de-Montréal (Hôpital du Sacré-Cœur de Montréal)

[5]Department of Neuroscience, Faculty of Medicine, Université de Montréal

Abstract

Syntactic ability in connected speech production is central to communication but its white matter predictors have seldom been studied. The aim of this study was to determine whether and which structural characteristics of the left AF and UF as well as the lesion load of the left AF in the acute phase post-stroke predicted syntactic ability in the connected speech of people with chronic post-stroke aphasia. Sixteen participants with post-stroke aphasia took part in this study. A magnetic resonance imaging scan was acquired within the first three days post-stroke and connected speech was assessed by means of the Picture Description test of the Western Aphasia Battery at least six months following the stroke. Four syntactic measures were extracted from the connected speech samples: the percentage of verbs produced, the number of grammatically well-formed sentences, the number of clauses per utterance (a measure reflecting sentences' syntactic complexity), and the percentage of verbs with a complex argument structure. The left AF and UF were reconstructed using a state-of-the-art constrained spherical convolution (CSD)-tractography pipeline and their volume, axial diffusivity (AD), and fractional anisotropy (FA) were extracted. The lesion load of the AF was also computed. Forward multiple linear regressions revealed that the volume of the left AF in the acute phase predicted sentence complexity, verb production ability, as well as the production of verbs with a complex argument structure. The latter's prediction was improved by the volume of the UF in the acute phase. Our study is the first to demonstrate that the volume of the left AF in the acute phase post-stroke is the sole predictor of the production of verbs and complex sentences and that the production of verbs with a complex argument structure relies on both the AF and UF.

Keywords: syntax, post-stroke aphasia, tractography, white matter, stroke, connected speech

1. Introduction

Post-stroke aphasia is an acquired language disorder most often caused by an ischemic stroke to the left middle cerebral artery. It is a heterogenous condition where one or several linguistic domains (phonology, morphology, syntax, and semantics) can be impaired, affecting language production and comprehension (Zumbansen & Thiel, 2014). One of the key challenges in the study of aphasia is the search for early neuroimaging predictors of outcome (i.e., language performance in the chronic phase post-stroke). Such an endeavour is critical for a better understanding of how brain structural alterations in the early phases define chronic language impairment following a stroke. The emergence of the dual-stream framework whereby language functions are supported by large-scale networks of grey matter regions connected via white matter pathways organized in dorsal and ventral streams (Hickok & Poeppel, 2004, 2007), as well as advances in neuroimaging over the past 20 years have placed white matter pathways at the forefront of research on the neural underpinnings of language and its breakdown in post-stroke aphasia and other neurological diseases (Dick et al., 2014; Poeppel et al., 2012; Wilson et al., 2012). It is therefore important to better understand how damage and the structural characteristics of key white matter fiber bundles in the early phase post-stroke contribute to language abilities in chronic aphasia.

Verb and sentence production are two aspects of discourse that are central to communication in everyday life. Individuals with post-stroke aphasia (fluent and non-fluent) have been found to experience verb production deficits (Berndt et al., 1997; Cho-Reyes & Thompson, 2012), as well as a reduced number of well-formed grammatical sentences in connected speech (Edwards, 2005; Thompson et al., 2013). Additionally, studies have shown that people with post-stroke aphasia experience greater difficulty producing verbs with a complex argument structure

(i.e., transitive verbs with two or three arguments) than intransitive verbs (i.e. those with one argument) (Thompson et al., 2012), and produce less complex sentences in terms of the number of clauses produced per sentence in connected speech (Hsu & Thompson, 2018; Llinàs-Grau & Martínez-Ferreiro, 2014; Stark, 2019). Thus, the syntactic structural complexity at the level of the verb-argument structure and the sentence seems to be important for verb and sentence production abilities in post-stroke aphasia.

Verb and sentence production in chronic post-stroke aphasia may be assessed by means of standardized clinical tests or connected speech analysis. While standardized clinical tests allow to investigate specific aspects of syntactic abilities, they lack ecological validity. Additionally, such tests do not exist for French-speaking populations. Connected speech production, which may be used as a surrogate assessment of functional communication, allows the simultaneous assessment of a range of measures and has been used to gain insight into syntactic performance during language production in post-stroke aphasia (e.g., Bastiaanse, 2011; Hsu & Thompson, 2018; Mirman et al., 2019). Verb production ability may be investigated by analyzing the proportion of verbs produced, or the number of arguments produced with each verb. At the sentence level, connected speech analysis allows the extraction of variables such as the number of clauses which indicates the level of sentence complexity and the number of well-formed sentences (Edwards, 2005; MacWhinney & Fromm, 2016).

Thus far, only a few studies have investigated syntactic abilities in language production and its structural brain substrates in aphasia (induced by a stroke or a neurodegenerative disease), with a particular focus on grey matter. For example, production of syntactically complex sentences in connected speech in primary progressive aphasia (PPA) has been found to be related to atrophy in left frontal and prefrontal regions (left posterior inferior frontal gyrus, IFG, superior frontal sulcus

and adjacent prefrontal areas, and the supplementary motor area) (Wilson et al., 2010). Studies on the structural correlates of verb retrieval deficits in post-stroke aphasia have also uncovered a relationship with damage to left frontal, temporal and parietal grey and white matter regions in the ventral as well as in the dorsal streams of the language network (Aggujaro et al., 2006; Alyahya et al., 2018; Piras & Marangolo, 2007). Impaired production of verbs with a complex argument structure in chronic post-stroke aphasia has been found to be predicted by damage to posterior superior temporal and angular gyrus (a structure at the intersection between the dorsal and ventral streams) and to dorsoventral white matter connections between the temporal and frontal lobes in the chronic phase post-stroke (den Ouden et al., 2019). The grey matter regions found in these studies are structurally connected via the arcuate fasciculus (AF) and the uncinate fasciculus (UF), respectively. Additionally, the dual-stream framework of the neural organization of syntax have proposed the AF and UF as the white matter pathways of the dorsal and ventral networks supporting syntactic processing (Friederici, 2012; Friederici & Gierhan, 2013; Zaccarella & Friederici, 2015). In this framework, the dorsal stream would support global syntactic operations such as complex sentence processing (Friederici & Gierhan, 2013; Zaccarella & Friederici, 2015). The ventral stream, on the other hand, is postulated to be involved in local syntactic operations, such as phrase structure building (e.g., verb-argument integration) and syntactic-semantic integration (Friederici et al., 2006; Friederici, 2018; Zaccarella & Friederici, 2015). The contribution of white matter fiber bundles' structural characteristics, particularly in the early phase post-stroke, to long-term syntactic abilities remains a pending question.

Growing evidence from studies on neuroimaging predictors of aphasia outcome has highlighted the left AF and UF as predictors of language performance in the chronic phase post-stroke. The AF's lesion load (i.e., the degree of overlap between a white matter fiber bundle and

the stroke-induced brain lesion) was found to predict speech fluency (Marchina et al., 2011), object naming (Geva et al., 2015; Marchina et al., 2011), repetition (Geva et al., 2015), and sentence comprehension (Geva et al., 2015; Xing et al., 2017) in chronic post-stroke aphasia. Microstructural properties of the AF have also been found to be related to chronic naming and sentence production (fractional anisotropy, FA in the chronic phase; Ivanova et al., 2016) and chronic global language outcome (axial diffusivity, AD in the acute phase; Moulton et al., 2019), while bundle volume (i.e., its macrostructure) has been related to overall aphasia severity in the chronic phase (Tak & Jang, 2014). The microstructural characteristics of the UF have been found to be related to word comprehension (Xing et al., 2017). However, none of these studies have specifically determined whether the AF and UF's structural characteristics in the early phase of post-stroke aphasia predict syntactic abilities in the chronic phase.

The objective of this study is to determine whether and which structural characteristics of the left AF and UF, as indexed by the bundle volume, the FA and AD in the acute phase of post-stroke aphasia (within three days post-stroke) predict the production of verbs and grammatically well-formed sentences and the syntactic structural complexity at the sentence and verb-argument structure levels in connected speech in the chronic phase (≥ six months post-stroke). We expect the number of clauses per utterance (a measure of syntactic complexity at the sentence level) in the chronic phase to be predicted by structural characteristics of the left AF in the acute phase (Wilson et al., 2010), and the percentage of verbs and verb-argument structure complexity, as well as the number of well-formed sentences to be predicted by structural characteristics of the left AF and UF in the acute phase (Akinina et al., 2019; den Ouden et al., 2019; Friederici et al., 2006). Additionally, since the lesion load is a variable that has also been found by some studies to predict language functions (Geva et al., 2015; Hillis et al., 2018; Marchina et al., 2011), we aimed to

determine whether it would predict verb and sentence production abilities in chronic post-stroke aphasia.

2. Methods

2.1. Participants

Sixteen participants were included in this study. They presented with aphasia due to a first single ischemic stroke in the left middle cerebral artery. No criteria concerning aphasia severity or lesion size were adopted. All participants were diagnosed by a neurologist at the Stroke Unit at Hôpital du Sacré-Coeur de Montréal and screened for eligibility. The chronic language assessments took place at least six months post-onset (chronic phase). Participants underwent a magnetic resonance imaging (MRI) scan within the first three days after stroke onset (acute phase). Exclusion criteria included presenting bi-hemispheric infarcts, previous head injury or intracranial surgery, history of major psychiatric illness, alcohol or drug abuse, learning difficulties before the accident, an uncorrected hearing or visual disturbance, contraindication to magnetic resonance (i.e. claustrophobia, metallic implant, etc.), and not being a French speaker. Additionally, twenty cognitively unimpaired controls matched with the patients on age and education served as a reference group for the behavioral analyses.

Demographic data of the participants is presented in Table 1. The study was approved by the research ethics committee of the Centre intégré universitaire de santé et de services sociaux du Nord-de-l'Ile-de Montréal (Project #MP-32-2018-1478) and written informed consent was obtained from all participants.

Table 1. Demographic, language, and neuropsychological data for the post-stroke aphasia (PSA) and control (CTRLs) groups

	PSAs (n = 16)		CTRLs (n = 20)		Intergroup differences (independent t-test)
	Mean	SD	Mean	SD	
Demographics					
Sex (% females)	50%		55%		
Age	72.0	13.08	67.55	9.01	$t(34) = -1.21$, $p = .236$
Education	13.0	4.05	14.55	2.28	$t(22.45)^a = 1.37$, $p = .185$
Time post-stroke (acute; days)	2.38	1.15	NA	NA	NA
Time post-stroke (chronic; months)	8.6	2.13	NA	NA	NA
Language and neuropsychological assessment					
Naming[b]	-3.79	7.07	.39	.60	$t(15.17)^a = 2.36$, $p < .05$
Repetition					
Words	26.25	5.66	29.05	1.32	$t(16.30)^a = 1.94$, $p = .07$
Sentences	1.88	1.09	2.6	.50	$t(20.10)^a = 2.46$, $p < .05$
Auditory comprehension					
Word and sentence comprehension	40.19	7.46	46.5	1.19	$t(15.61)^a = 3.35$, $p < .01$
Token Test	26.53	9.88	34.06	2.05	$t(15.06)^a = 2.90$, $p < .05$
Verbal fluency					
Free	28.0	24.80	68.05	16.94	$t(34) = 5.75$, $p < .001$
Orthographic	11.13	12.16	28.85	7.30	$t(34) = 5.42$, $p < .001$
Semantic	13.13	7.40	29.5	6.12	$t(29.06)^a = 7.11$, $p < .001$
PPTT	41.81	9.30	50.4	1.23	$t(15.42)^a = 3.67$, $p < .01$
Bells Test	29.43	9.89	NA	NA	NA

[a]Homogeneity of variance assumption not met according to Levene's test. Welch's t-test with adjusted degrees of freedom is reported.

[b]Since two patients were evaluated with the TDQ-60, while the rest were evaluated with the DO-80, we computed a z-score for Naming by using each test's normative data.

2.2. Behavioral data

2.2.1. Language assessment

A language battery in French was administered ≥ six months post-stroke to the participants to assess the following language functions: confrontational naming (DO-80 and TDQ-60, Deloche & Hannequin, 1997; Macoir et al., 2018), repetition (the word and pseudo-word repetition subtests of the Protocole Montreal-Toulouse d'examen linguistique de l'aphasie, MT-86, Nespoulous et al., 1992), and auditory comprehension (using the word and sentence comprehension subtests of the MT-86; Nespoulous et al., 1992, and the Token Test, short version; De Renzi & Faglioni, 1978). Connected speech was assessed using The Picnic Scene from the Western Aphasia Battery (Kertesz, 2006). Additionally, executive functions were assessed using the Verbal Fluency (free, phonemic, semantic) test from the Montréal Évaluation de la Communication (MEC) battery (Joanette et al., 2004), semantic memory was assessed using the Pyramids and Palm Trees test (Howard & Patterson, 1992), and visual neglect was assessed using the Bells Test (Gauthier et al., 1989). Table 1 presents the test results of the post-stroke aphasia and control groups. Aphasia severity was determined using the Boston Denomination Aphasia Examination (BDAE) severity scale (Goodglass et al., 2001). Twelve patients had mild or mild to moderate aphasia, three were moderate, and one was moderate to severe.

2.2.2. Connected speech analysis

Speech samples were transcribed in the CHAT format by a speech-language pathologist (A.B.). Inter-rater reliability was carried out in a previously published study by our group (Brisebois et al., 2020) and was found to be high (intra-class correlation coefficient >.80). The MOR program in the CLAN software was run to tag parts-of-speech (Forbes et al., 2012; MacWhinney et al., 2011). CLAN's EVAL program was then used to extract the following measures: the percentage of verbs

produced (over the total number of words) as a measure of verb production ability; the number of verbs per utterance which corresponds to the number of clauses per utterance (Macwhinney, 2019) as a measure of syntactic complexity at the sentence level (Brisebois et al., 2020; Thorne & Faroqi-Shah, 2016); and the number of grammatically well-formed sentences. Additionally, the percentage of verbs with a complex argument structure (i.e., verbs with two or three arguments) produced by each participant was extracted by a team member (M.B.). As previously mentioned, this variable reflects verb-argument structure complexity.

2.3.Neuroimaging data

2.3.1. Image acquisition

The diffusion MRI protocol was acquired using a Skyra 3T MRI scanner (Siemens Healthcare, USA) at the radiology department of Hôpital du Sacré-Coeur of Montreal. One high resolution 3D T1-weighted (T1w) image (TR = 2200 ms, TE = 2.96 ms, TI = 900 ms, FOV = 250 mm, voxel size = 1x1x1 mm3, matrix = 256x256, 192 slices, flip-angle = 8°) was acquired using a Magnetization Prepared Rapid Gradient Echo (MP-RAGE) sequence. A diffusion weighted imaging (DWI) sequence was also acquired (TR = 8051 ms, TE = 86 ms, FOV= 230 mm, voxel size = 2x2x2 mm3, flip angle = 90°, bandwidth = 1698; EPI factor = 67; 68 slices in transverse orientation) with one T2-weighted image (b = 0 s/mm²) and 64 images with non-collinear diffusion gradients (b=1,000 s/mm²) in a posterior-anterior (PA) acquisition, as well as two other T2-weighted images (b = 0 s/mm²): one in a PA acquisition and the other in an anterior-posterior (AP) acquisition.

2.3.2. dMRI data pre-processing and tractography

Pre-processing and tractography were carried out following the same procedure as in Boukadi et al. (2019). Noise-correction of all DWI images was carried out using overcomplete local principal component analysis (PCA) (Manjo et al., 2013) and AP-PA images (the two b0 images) were used to correct susceptibility distortion. DWI images were also corrected for subject movement, geometric distortions, and eddy-currents with the FMRIB Diffusion toolbox EDDY of FSL 5.0.11 (http://www.fmrib.ox.ac.uk/fsl/) (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012) and gradient directions were aligned with motion correction parameters. DWI images were upsampled to 1mm isotropic resolution using a trilinear interpolation (Dyrby et al., 2014; Girard & Descoteaux, 2012; Raffelt et al., 2012; Smith et al., 2012; Tournier et al., 2012). T1 weighted (T1w) images were segmented and parcelled into grey and white matter anatomical regions using Freesurfer's pipeline 6.0.0 (Dale et al., 1999; Desikan et al., 2006). Then, T1w-to-DWI registration was carried out with FMRIB's linear registration tool (FLIRT) from FSL.

Using constrained spherical deconvolution, fiber orientation distribution functions (fODFs) were estimated and a whole-brain tractogram was computed using MRtrix3's probabilistic anatomically constrained tractography algorithm (https://github.com/jdtournier/) (Tournier et al., 2012). The T1w parcellation was overlaid on the tractogram and the left AF and UF were then reconstructed with the White Matter Query Language (WMQL) (Wassermann et al., 2016). WMQL is a method that consists in writing anatomic definitions of the bundles of interest in the form of queries, using the anatomic regions from Freesurfer's Desikan/Killiany atlas. Queries used to reconstruct the left AF and UF have been presented in Boukadi et al. (2019). Finally, we ran a tract-filtering algorithm (Côté et al., 2015) to remove outlier streamlines (i.e., false positives). The volume, FA, and AD of the AF and UF were then extracted.

2.3.3. Lesion delineation

The stroke-induced lesion was delineated for each patient using a semi-automatic method and verified with a fully manual method. First, the Clusterize toolbox (De Haan et al., 2015) was used in SPM12 to semi-automatically delineate the lesion on the mean diffusivity (MD) map extracted from the DWI image acquired in the acute phase (an example of the lesion drawn on the MD image, as well as the MD and T1-w images for one participant can be viewed in the Appendix). Clusterize has been previously shown to have a good reliability for acute lesions' delineation in stroke patients (De Haan et al., 2015), and the mean diffusivity map was suggested to be a good tool for lesion visualization in the acute phase (Schaefer et al., 2006). Clusterize automatically computed hypo-intense clusters of voxels on mean diffusivity maps (default parameters were used). Clusters of interest corresponding to the lesion in each slice were then manually selected and adjusted to accurately fit the lesion by a neuroscientist (B.H.) and the entire lesion was extracted for each patient. Secondly, each lesion image was counter-verified and adjusted (as needed) using the MI-brain software (Imeka Solutions Inc.) by a neuroscientist experienced in lesion delineation (S.M.B.). The mean diffusivity and b0 maps were used to support this visual verification. Both delineators were blind to behavioral scores. Finally, the lesion volume (in ml) in the DWI native space was computed for each patient using Clusterize. T1w images and lesions were then normalized to Rorden et al.'s (2012) age-matched template from 30 healthy controls (mean age: 61.3 years; 17 men) using the enantiomorphic normalization method of the Clinical Toolbox (Rorden et al., 2012) in SPM12. The lesion overlap map of all patients is displayed in Figure 1 on the age-matched template in MNI space.

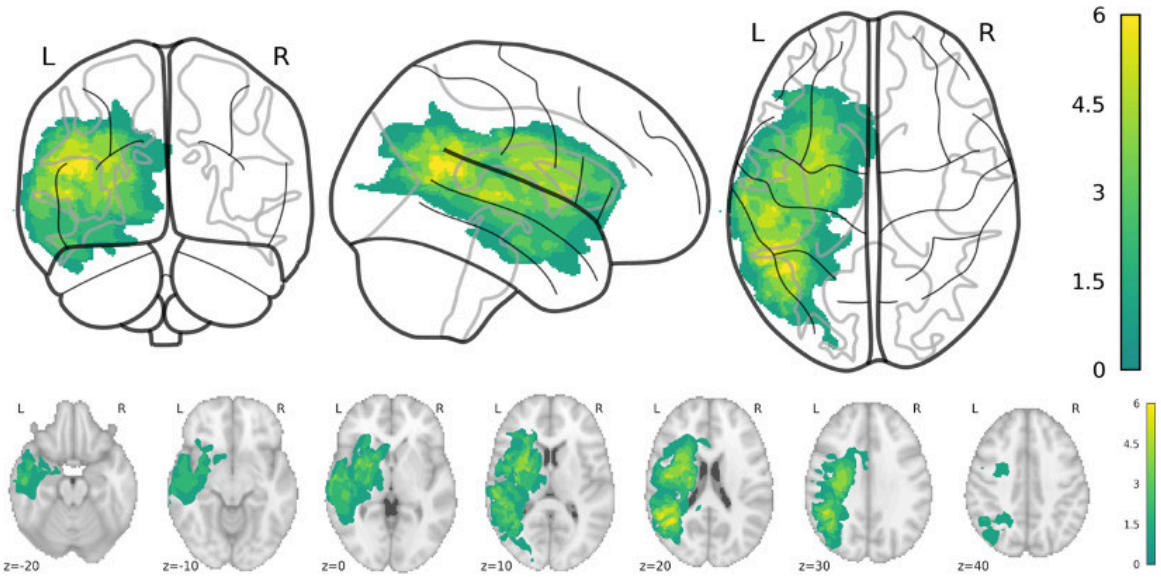


Figure 1. Lesion overlap map. Lighter green on the color scale indicates the greatest number of patients with a lesion in the same location.

2.3.4. Lesion load calculation

Lesion loads of the AF and UF were obtained following the same methodology as in Marchina et al. (2011). The fiber bundles of eighteen cognitively unimpaired controls reconstructed in a previous study (Boukadi et al., 2019) were transformed into binary maps and normalized to the MNI152 space using SPM12. A study-specific atlas or map for each bundle was then created by calculating the probability that a voxel was part of the fiber bundle across all control subjects. The lesions were also normalized to the MNI152 space using SPM12. The lesion load, namely the overlap between the fiber bundle and the lesion, was calculated as the sum of the intensities of all shared voxels between the AF or UF's fiber bundle map and the patient's lesion. Figure 2 shows the lesion-fiber bundle overlap in one aphasic participant.

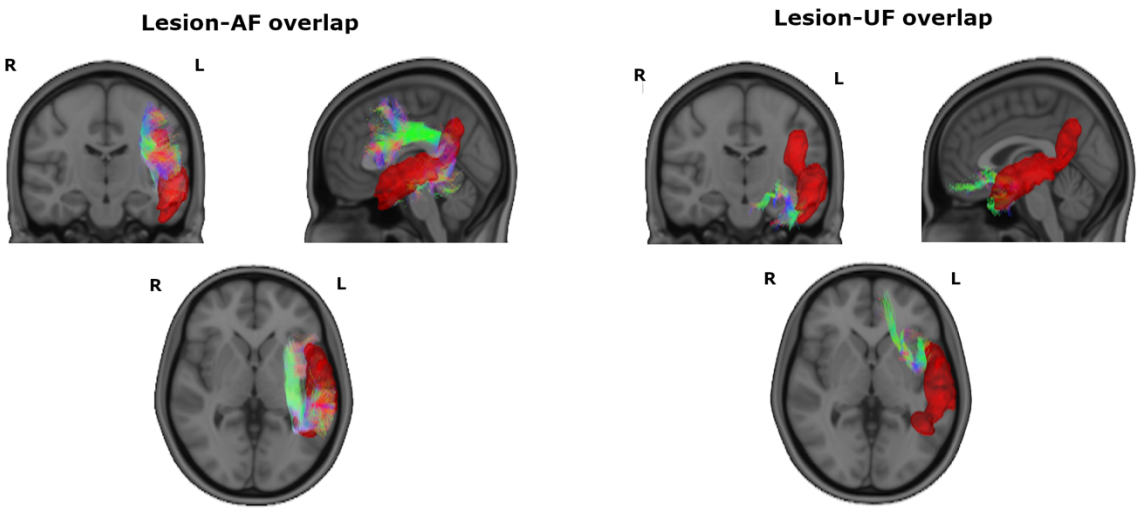


Figure 2. Overlap between the white matter bundles and the lesion in one participant.

3. Statistical analyses

All statistical analyses were run on SPSS25. T-tests were carried out with the group (participants with post-stroke aphasia and controls) entered as an independent variable and each syntactic measure (percentage of verbs, number of grammatically well-formed sentences, number of clauses per utterance, and percentage of verbs with a complex argument structure) entered as a dependent variable. The homogeneity of variance assumption was assessed with Levene's test and Welch's *t*-test with adjusted degrees of freedom was reported whenever equality of variances was not met. In order to determine which of the structural characteristics of the left AF and UF (bundle volume, FA, AD) would best predict our syntactic variables, we ran forward multiple linear regressions with our group of patients with post-stroke aphasia. Only the lesion load of the AF (and not that of the UF) was included in the analyses, since all patients but one showed an overlap between their AF and the lesion while only four patients showed an overlap between their UF and the lesion. The percentage of verbs produced, the percentage of verbs with a complex argument

structure), the number of clauses per utterance, as well as the number of grammatically well-formed sentences were each separately entered as a dependent variable.

4. Results

The mean and standard deviation of the syntactic variables extracted from the connected speech analysis for the chronic post-stroke aphasia and control groups, as well as the results of the *t*-tests are presented in Table 2. No differences at the group level were identified for any of the syntactic measures.

Table 2. Mean and standard deviation (SD) for syntactic measures for PSAs and CTRLs

	PSAs		CTRLs		
	Mean	SD	Mean	SD	Intergroup differences (independent *t*-test)
Number of clauses per utterance	.37	.32	.427	.231	$t(34) = .93$, $p = .357$
Number of well-formed sentences	19.81	12.28	20.75	8.87	$t(34) = .27$, $p = .792$
% verbs	4.00	2.32	4.059	1.45	$t(34) = .17$, $p = .863$
% verbs with a complex argument structure	68.87	37.97	77.62	16.17	$t(19.35)^a = .86$, $p = .399$

[a]Homogeneity of variance assumption not met according to Levene's test. Welch's *t*-test with adjusted degrees of freedom is reported.

The forward multiple regression analysis we ran on our group of patients with post-stroke aphasia revealed that the volume of the left AF in the acute phase was selected as the only predictor of the number of clauses ($R^2 = .448$, $F(1,13) = 10.566$, $p < .01$; $\beta = .67$, $p < .01$) and percentage of verbs produced ($R^2 = .579$, $F(1,13) = 17.866$; $\beta = .761$, $p < .01$). Additionally, the volume of both the left AF and UF in the acute phase emerged as significant predictors of the percentage of verbs with a complex argument structure. The AF was selected as the most important predictor and entered in the first step of the model, accounting for 53% ($F(1, 13) = 14.561$, $p < .01$) of the variance of verbs with a complex argument structure. The volume of the UF was added as a predictor in the second step of the model, significantly improving the prediction by 14% ($F(1,12) = 4.98$, $p < .05$).

Thus, the regression model with the volumes of the left AF and UF accounted for 67% of the variance of verbs with a complex argument structure ($F(2, 12)= 11.999$, $p < .01$; $\beta_{AFvolume} = .65$, $p < .01$, $\beta_{UFvolume} = .38$, $p < .05$). There was no multicollinearity among the two predictors included in the last model (variance inflation factor = 1.05, which indicates that there is no collinearity issue; Hair et al., 2014). No significant predictors were found for the number of well-formed sentences.

5. Discussion

In this study, we investigated whether the structural characteristics of the left AF and UF, as well as the lesion load of the left AF in the acute phase post-stroke predicted the production of verbs and grammatically well-formed sentences, as well as sentence and argument structure complexity, in the connected speech of people with chronic post-stroke aphasia. Our results revealed that the number of clauses per utterance (an index of syntactic complexity at the sentence level), the percentage of verbs produced, as well as the percentage of verbs with a complex argument structure produced in the chronic phase are predicted by the volume of the left AF in the acute phase. Additionally, the volume of the left UF in the acute phase significantly improved the prediction of the percentage of verbs with a complex argument structure. The number of grammatically well-formed sentences produced was not predicted by any diffusion measure. Neither the lesion load of the AF nor the FA or AD of the AF and UF emerged as predictors of the syntactic measures under study.

Our group of participants showed impaired performance on all language functions (except word repetition) assessed using standardized clinical tests. However, comparisons with controls indicate that this group of patients was unimpaired on the syntactic measures extracted from connected speech (i.e. the Picnic Scene from the Western Aphasia Battery). This seemingly reflects normal syntactic ability in a discourse context in the chronic phase. It could also be due to the fact

that all participants presented with mild or moderate aphasia (with a majority of patients showing a mild language impairment). Previous studies also reported unimpaired performance, as compared to controls, on different connected speech measures in mild/moderate post-stroke aphasia. For example, Yorkston and Beukelman (1980) found no differences between mild/moderate aphasic speakers and healthy individuals on the number of content units (a measure of informativeness) produced in connected speech. In a sample of mostly mild aphasic speakers, Stark (2019) found no impairments (as compared to controls) on connected speech measures such as noun-verb, open-closed class, and type-token ratios. Thus, it is possible that some connected speech measures may be less sensitive than others to impairments in mild/moderate aphasic speakers. Future work would have to assess this hypothesis by comparing mild/moderate and severe aphasic speakers on a wide range of connected speech measures. That being said, participants with mild/moderate post-stroke aphasia in the present study do appear to be unimpaired on the syntactic measures that we extracted from picture description. Additionally, the absence of a difference between controls and aphasic speakers does not discount the presence of inter-individual variability in syntactic ability among our participants with post-stroke aphasia.

Our findings highlight the central role of the left AF in syntactic processing, as the volume of this bundle emerged as the sole acute predictor of chronic verb production and complex sentences, and as the most important predictor of chronic verb-argument structure complexity. This is in line with structural studies in chronic post-stroke aphasia (Ivanova et al., 2016) and primary progressive aphasia (Wilson et al., 2011) that found that damage to the left AF was related to syntactic production. While these studies reported correlations between syntactic impairments in production and microstructural variations (i.e., the FA) in the left AF, ours revealed that

macrostructural variation (i.e., the bundle volume) in this fiber bundle is a determinant of relatively preserved syntactic ability in connected speech.

In addition to the volume of the left AF, we found that the volume of the left UF in the acute phase also predicted the production of verbs with a complex argument structure in the chronic phase. Friederici and Gierhan's (2013) theoretical framework proposed a differential role for the left AF and UF in syntactic processing (i.e., that the AF would be involved in hierarchical structure building of complex sentences and the UF would be involved in local and simple syntactic operations). Instead, our finding highlights a potential synergistic relationship between the AF and the UF in the production of verbs with a complex argument structure. This is in line with previous work on language comprehension suggesting a synergy between the dorsal and ventral streams in syntactic comprehension in the chronic phase (Griffiths et al., 2013; Rolheiser et al., 2011). It is possible that the AF would be involved in building the hierarchical syntactic structure necessary for the realization of the argument structure. The more arguments a verb requires, the more complex the structure. This hypothesis would be in line with the role generally attributed to the AF in structural syntactic complexity (Catani & Bambini, 2014; Friederici & Gierhan, 2013; Zaccarella & Friederici, 2015). The UF on the other hand, would be involved in verb-argument integration (i.e. assigning thematic roles that the verb requires to each argument), which would be more in line with its proposed role in combinatory processes, as well as local structure (i.e. phrase structure) building (Friederici et al., 2006; Zaccarella & Friederici, 2015).

Verb production in general (i.e., the percentage of verbs produced over the total number of words in the speech sample) in the chronic phase was predicted by the left AF's volume and not the UF's. This is somewhat inconsistent with voxel-based lesion symptom mapping studies in chronic post-stroke aphasia which found that grey and white matter structures in both the dorsal

and ventral streams correlated with verb production in confrontation naming tasks (Akinina et al., 2019; Alyahya et al., 2018; Piras & Marangolo, 2007). However, these studies were carried out in the chronic phase where damage is usually much more extensive in grey and white matter than in the acute phase and focused on correlating structural damage with verb production impairments, while our study reveals that inter-patient variability in the volume of the left AF in the acute phase predicts relatively preserved verb production in the chronic phase post-stroke.

Additionally, in our study neither the AF's nor the UF's volume predicted the number of grammatically well-formed sentences produced by our participants. It is possible that since the production of well-formed sentences involves different processes (syntactic but also semantic and phonological), it would engage a widespread network of interacting grey and white matter structures rather than rely primarily on a specific white matter bundle or the limited set of cortical regions it connects. Further work using connectome-based analyses would be needed to assess this hypothesis. Another potential explanation could be that in such relatively short speech samples as the ones elicited by picture description, the raw number of sentences produced could be biased by the overall inter-subject variability in the length of output. In this context, the number of well-formed sentences might reflect syntactic ability less specifically.

Our study is, to the best of our knowledge, the first to reveal that bundle volume in the acute phase is a good predictor of connected speech measures in the chronic phase. There are only a handful of studies that investigated acute white matter predictors of language abilities in chronic post-stroke aphasia and only two which investigated the predictive value of the bundle volume (Forkel & Catani, 2018; Forkel et al., 2014) which did not emerge as a predictor of language ability in the chronic phase. However, in that study, a general measure of language impairment severity in the chronic phase was used (i.e. the Aphasia Quotient composite score), which may rely on a

network of bundles rather than on one specific bundle, as the authors themselves suggested (Forkel & Catani, 2018). By contrast, in the present study we measured specific linguistic features in connected speech that might be better related to the structural properties of specific white matter bundles. In other words, structural characteristics of white matter bundles are more likely to predict specific language abilities for which they are known to play a role, rather than a general language measure (i.e., the Aphasia Quotient composite score). Future research using bundle volume as an acute-phase predictor would have to be carried out with other specific language measures in the chronic phase in order to assess this hypothesis.

Neither the lesion load, the AD, nor the FA were selected as predictors of any of our syntactic measures in the forward regression analyses. The lesion load of the AF (a surrogate measure of damage) in the acute phase has been found to predict the degree of language impairment in the subacute or chronic phases in previous studies (Hillis et al., 2018; Moulton et al., 2019; Osa García et al., 2020). In other words, a bigger lesion overlap to the AF in the acute phase predicts more severe aphasia in the chronic phase. By contrast, in our study, patients were unimpaired on the syntactic measures obtained from connected speech. Thus, the lesion load seems to serve more as a predictor of aphasia severity. As for the AD and FA, these measures have been previously found to undergo rapid dynamic changes in the early phase post-stroke and only stabilize in later stages (Fung et al., 2011; Green et al., 2002; Sorensen et al., 1999; Yang et al., 1999). These changes may vary rapidly in the first few hours and days post-stroke (Green et al., 2002; Yang et al., 1999). Since our measures were taken at different timepoints across patients within the three-day window of the acute phase, this variability could have limited the predictive value of these tensor metrics.

The present study has a few limitations. First, our sample size is rather small and further research with a bigger sample size is required to confirm and generalize our findings. Additionally, a bigger sample size could potentially allow the investigation and identification of a larger number of predictors. Secondly, our language assessment did not include standardized clinical tests that specifically assess syntactic abilities, such as the Northwestern Assessment of Verbs and Sentences (Cho-Reyes & Thompson, 2012). This would have allowed us to confirm whether our group of participants had no syntactic impairments altogether or was simply unimpaired in a discourse context or on the specific measures extracted from the connected speech samples. Additionally, a connected speech task that elicits longer samples (e.g. the Cinderella story retelling) could help in the interpretation of these findings. Finally, the great majority of our participants had mild to moderate aphasia, which might have influenced our findings. It would be interesting for future studies to use a more varied sample with a wider range of mild to severe cases. Though this could represent an additional challenge at the level of patient recruitment.

In conclusion, our findings have implications for the theoretical literature on the neurocognitive mechanisms supporting syntactic processes, as well as for the clinical literature on the neuroanatomical predictors of syntax production in chronic post-stroke aphasia. With regards to theoretical implications, our findings that the volume of the left AF (as well as the volume of the left UF for verbs with a complex argument structure) predicts syntactic measures in connected speech in chronic post-stroke aphasia suggest that the left AF is a fiber bundle that plays a major role in syntactic production, either independently or in synergy with the left UF for local syntactic operations such as verb-argument integration. With regards to clinical implications, our study suggests that the volume of white matter fiber bundles in the acute phase could potentially serve as a prognostic marker that would help identify patients that might show preserved or least impaired

syntactic function in connected speech in the chronic phase. This could potentially have important applications in personalized speech rehabilitation in the long-term. Future studies would have to first replicate our findings with a bigger, more diverse aphasic sample.

References

Aggujaro, S., Crepaldi, D., Pistarini, C., Taricco, M., & Luzzatti, C. (2006). Neuro-anatomical correlates of impaired retrieval of verbs and nouns: Interaction of grammatical class, imageability and actionality. *Journal of Neurolinguistics, 19*(3), 175–194. https://doi.org/10.1016/j.jneuroling.2005.07.004

Akinina, Y., Dragoy, O., Ivanova, M. V., Iskra, E. V., Soloukhina, O. A., Petryshevsky, A. G., Fedina, O. N., Turken, A. U., Shklovsky, V. M., & Dronkers, N. F. (2019). Grey and white matter substrates of action naming. *Neuropsychologia, 131*, 249–265. https://doi.org/10.1016/j.neuropsychologia.2019.05.015

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018). Noun and verb processing in aphasia: Behavioural profiles and neural correlates. *NeuroImage: Clinical, 18*, 215–230. https://doi.org/10.1016/j.nicl.2018.01.023

Bastiaanse, R. (2011). The retrieval and inflection of verbs in the spontaneous speech of fluent aphasic speakers. *Journal of Neurolinguistics, 24*(2), 163–172. https://doi.org/10.1016/j.jneuroling.2010.02.006

Berndt, S. R., Mitchum, C. C., Haendiges, A. N., & Sandson, J. (1997). Verb retrieval in aphasia. *Brain and Language, 106*(56), 68–106.

Boukadi, M., Marcotte, K., Bedetti, C., Houde, J. C., Desautels, A., Deslauriers-Gauthier, S., Chapleau, M., Boré, A., Descoteaux, M., & Brambati, S. M. (2019). Test-retest reliability of diffusion measures extracted along white matter language fiber bundles using HARDI-based tractography. *Frontiers in Neuroscience: Brain Imaging Methods, 12*, 1–11. https://doi.org/10.3389/fnins.2018.01055

Brisebois, A., Brambati, S. M., Désilets-Barnabé, M., Boucher, J., Osa García, A., Rochon, E., … Marcotte, K. (2020). Narrative discourse recovery in acute post-stroke aphasia: the importance of thematic informativeness. *Aphasiology, 34*(4), 472–491. https://doi.org/10.1080/02687038.2019.1705661

Catani, M., & Bambini, V. (2014). A model for Social Communication and Language Evolution and Development (SCALED). *Current Opinion in Neurobiology, 28*, 165–171. https://doi.org/10.1016/j.conb.2014.07.018

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Côté, M. A., Garyfallidis, E., Larochelle, H., & Descoteaux, M. (2015). Cleaning up the mess: tractography outlier removal using hierarchical QuickBundles clustering. In *Proceedings of ISMRM*, vol. 23 (p. 2844).

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis. *NeuroImage, 194*, 179–194.

De Haan, B., Clas, P., Juenger, H., Wilke, M., & Karnath, H. O. (2015). Fast semi-automated lesion demarcation in stroke. *NeuroImage: Clinical, 9*, 69–74. https://doi.org/10.1016/j.nicl.2015.06.013

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex, 14*(1), 41–49.

Deloche, G., & Hannequin, D. (1997). *Test de dénomination orale d'images: DO-80*. Éditions du centre de psychologie appliquée.

den Ouden, D. B., Malyutina, S., Basilakos, A., Bonilha, L., Gleichgerrcht, E., Yourganov, G., … Fridriksson, J. (2019). Cortical and structural-connectivity damage correlated with impaired syntactic processing in aphasia. *Human Brain Mapping, 40*(7), 2153-2173. https://doi.org/10.1002/hbm.24514

Desikan, R. S., Se, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage, 31*, 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Dick, A. S., Bernal, B., & Tremblay, P. (2014). The language connectome: New pathways, new concepts. *The Neuroscientist, 20*(5), 453–467. https://doi.org/10.1177/1073858413513502

Dyrby, T. B., Lundell, H., Burke, M. W., Reislev, N. L., Paulson, O. B., Ptito, M., & Siebner, H. R. (2014). Interpolation of diffusion weighted imaging datasets. *NeuroImage, 103*, 202–213. https://doi.org/10.1016/j.neuroimage.2014.09.005

Edwards, S. (2005). Connected fluent aphasic speech. In *Fluent Aphasia* (pp. 92–121). Cambridge University Press.

Forbes, M. M., Fromm, D., & MacWhinney, B. (2012). AphasiaBank: A resource for clinicians. *Seminars in Speech and Language, 33*(3), 217–222. https://doi.org/10.1055/s-0032-1320041

Forkel, S. J., & Catani, M. (2018). Lesion mapping in acute stroke aphasia and its implications for recovery. *Neuropsychologia, 115*, 88–100. https://doi.org/10.1016/j.neuropsychologia.2018.03.036

Forkel, S. J., De Schotten, M. T., Dell'Acqua, F., Kalra, L., Murphy, D. G. M., Williams, S. C. R., & Catani, M. (2014). Anatomical predictors of aphasia recovery: A tractography study of bilateral perisylvian language networks. *Brain, 137*(7), 2027–2039. https://doi.org/10.1093/brain/awu113

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16*(5), 262–268. https://doi.org/10.1016/j.tics.2012.04.001

Friederici, A. D. (2018). The neural basis for human syntax: Broca's area and beyond. *Current Opinion in Behavioral Sciences, 21*, 88–92. https://doi.org/10.1016/j.cobeha.2018.03.004

Friederici, A. D., Bahlmann, J. R., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*(7), 2458–2463. https://doi.org/10.1073/pnas.0509389103

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology, 23*(2), 250–254. https://doi.org/10.1016/j.conb.2012.10.002

Fung, S. H., Roccatagliata, L., Gonzalez, R. G., & Schaefer, P. W. (2011). MR Diffusion Imaging in Ischemic Stroke. *Neuroimaging Clinics of North America, 21*(2), 345–377. https://doi.org/10.1016/j.nic.2011.03.001

Gauthier, L., Dehaut, F., & Joanette, Y. (1989). The Bells Test: A quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology, 11*(2), 49–54.

Geva, S., Correia, M. M., & Warburton, E. A. (2015). Contributions of bilateral white matter to chronic aphasia symptoms as assessed by diffusion tensor MRI. *Brain and Language, 150*, 117–128. https://doi.org/10.1016/j.bandl.2015.09.001

Girard, G., & Descoteaux, M. (2012). Anatomical tissue probability priors for tractography. In *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'12) - Computational Diffusion MRI Workshop* (pp. 174–185). Nice.

Goodglass, H., Kaplan, E., & Barresi, B. (2001). T*he Boston Diagnostic Aphasia Examination: BDAE-3 Long Form Kit.* Lippincott, Williams, and Wilkins.

Green, H. A. L., Peña, A., Price, C. J., Warburton, E. A., Pickard, J. D., Carpenter, T. A., & Gillard, J. H. (2002). Increased anisotropy in acute stroke: A possible explanation. *Stroke, 33*(6), 1517–1521. https://doi.org/10.1161/01.STR.0000016973.80180.7B

Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., & Tyler, L. K. (2013). Functional organization of the neural language system: Dorsal and ventral pathways are critical for syntax. *Cerebral Cortex, 23*(1), 139–147. https://doi.org/10.1093/cercor/bhr386

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate Data Analysis* (7th ed.). Peason Education Limited. https://doi.org/10.1002/9781118895238.ch8

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1–2), 67–99. https://doi.org/10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience, 8*, 393–402. https://doi.org/10.1038/nrn2113

Hillis, A. E., Beh, Y. Y., Sebastian, R., Breining, B., Tippett, D. C., Wright, A., Saxena, S., Rorden, C., Bonilha, L., Basilakos, A., Yourganov, G., & Fridriksson, J. (2018). Predicting recovery in acute poststroke aphasia. *Annals of Neurology, 83*(3), 612–622. https://doi.org/10.1002/ana.25184

Howard, D., & Patterson, K. (1992). *Pyramids and palm trees: A test of semantic access from pictures and words.* Thames Valley Test Company.

Hsu, C.-J., & Thompson, C. K. (2018). Manual versus automated narrative analysis of agrammatic production patterns: The Northwestern Narrative Language Analysis and computerized language analysis. *Journal of Speech Language and Hearing Research,* 1–13. https://doi.org/10.1044/2017

Ivanova, M. V., Isaev, D. Y., Dragoy, O. V., Akinina, Y. S., Petrushevskiy, A. G., Fedina, O. N., Shklovsky, V. M., & Dronkers, N. F. (2016). Diffusion-tensor imaging of major white matter tracts and their role in language processing in aphasia. *Cortex, 85*, 165–181. https://doi.org/10.1016/j.cortex.2016.04.019

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Joanette, Y., Ska, B., & Côté, H. (2004). Protocole Montreal d'Evaluation de la Communication MEC. Ortho Editions.

Kertesz, A. (2006). *Western Aphasia Battery.* Pearson.

Kristo, G., Leemans, A., Raemaekers, M., Rutten, G. J., De Gelder, B., & Ramsey, N. F. (2013). Reliability of two clinically relevant fiber pathways reconstructed with constrained spherical deconvolution. *Magnetic Resonance in Medicine, 70*(6), 1544–1556. https://doi.org/10.1002/mrm.24602

Llinàs-Grau, M., & Martínez-Ferreiro, S. (2014). On the presence and absence of that in aphasia. *Aphasiology, 28*(1), 62–81. https://doi.org/10.1080/02687038.2013.828345

Macoir, J., Beaudoin, C., Bluteau, J., Potvin, O., & Wilson, M. A. (2018). TDQ-60 - a color picture-naming test for adults and elderly people: validation and normalization data. *Aging,*

Neuropsychology and Cognition, 25(5), 753–766. https://doi.org/10.1080/13825585.2017.1372355

Macwhinney, B. (2019). *Tools for Analyzing Talk Part 2: The CLAN Program*. https://doi.org/10.21415/T5G10R

MacWhinney, B, & Fromm, D. (2016). AphasiaBank as BigData. *Seminars in Speech and Language, 37*(1), 10–22. https://doi.org/10.1002/cncr.27633.Percutaneous

MacWhinney, Brian, Fromm, D., Forbes, M., & Holland, A. (2011). Aphasiabank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893

Manjo, V., Concha, L., Buades, A., & Collins, D. L. (2013). Diffusion weighted image denoising using overcomplete local PCA. *PLOS One, 8*(9), 1–12. https://doi.org/10.1371/journal.pone.0073021

Marchina, S., Zhu, L. L., Norton, A., Zipse, L., Wan, C. Y., & Schlaug, G. (2011). Impairment of speech producton predicted by lesion load of the left arcuate fasciculus. *Stroke, 42*(8), 2251–2256. https://doi.org/10.1161/STROKEAHA.110.606103.Impairment

Mirman, D., Kraft, A. E., Harvey, D. Y., Brecher, A. R., & Schwartz, M. F. (2019). Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. *Cognitive, Affective, & Behavioral Neuroscience, 19,* 1286-1298. https://doi.org/10.3758/s13415-019-00729-9

Moulton, E., Magno, S., Valabregue, R., Amor-Sahli, M., Pires, C., Lehéricy, S., Leger, A., Samson, Y., & Rosso, C. (2019). Acute diffusivity biomarkers for prediction of motor and language outcome in mild-to-severe stroke patients. *Stroke, 50*(8), 2050–2056. https://doi.org/10.1161/STROKEAHA.119.024946

Nespoulous, J. L., Lecours, A. R., Lafond, D., Lemay, M. A., Puel, M., Joanette, Y., Cot, F., & Rascol, A. (1992). *Protocole Montréal-Toulouse d'examen linguistique de l'aphasie: MT-86 module standard initial, M1b* (2e édition révisée par Renée Béland et Francine Giroux). Ortho Éditions.

Osa García, A., Brambati, S. M., Brisebois, A., Désilets-Barnabé, M., Bedetti, C., Rochon, E., Leonard, C., Desautels, A., & Marcotte, K. (2020). Predicting early post-stroke aphasia outcome from initial aphasia severity. *Frontiers in Neurology, 11*. https://doi.org/10.3389/FNEUR.2020.00120

Piras, F., & Marangolo, P. (2007). Noun-verb naming in aphasia: A voxel-based lesion-symptom mapping study. *NeuroReport, 18*(14), 1455–1458. https://doi.org/10.1097/WNR.0b013e3282ef6fc9

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience, 32*(41), 14125–14131. https://doi.org/10.1523/JNEUROSCI.3244-12.2012

Raffelt, D., Tournier, J., Rose, S., Ridgway, G. R., Henderson, R., Crozier, S., Salvado, O., & Connelly, A. (2012). Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. *NeuroImage, 59*(4), 3976–3994. https://doi.org/10.1016/j.neuroimage.2011.10.045

Rolheiser, T., Stamatakis, E. A., & Tyler, L. K. (2011). Dynamic processing in the human language system: Synergy between the arcuate fascicle and extreme capsule. *Journal of Neuroscience, 31*(47), 16949–16957. https://doi.org/10.1523/JNEUROSCI.2725-11.2011

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization Christopher. *NeuroImage, 61*(4), 957–965. https://doi.org/10.1038/mp.2011.182.doi

Schaefer, P. W., Copen, W. A., Lev, M. H., & Gonzalez, R. G. (2006). Diffusion-weighted imaging in acute stroke. *Magnetic Resonance Imaging Clinics of North America, 14*(2), 141–168. https://doi.org/10.1016/j.mric.2006.06.005

Smith, R. E., Tournier, J., Calamante, F., & Connelly, A. (2012). Anatomically-constrained tractography: Improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage, 62*(3), 1924–1938. https://doi.org/10.1016/j.neuroimage.2012.06.005

Sorensen, A. G., Wu, O., Copen, W. A., Davis, T. L., Gonzalez, R. G., Koroshetz, W. J., Reese, T. G., Rosen, B. R., Wedeen, V. J., & Weisskoff, R. M. (1999). Human acute cerebral ischemia: Detection of changes in water diffusion anisotropy by using MR imaging. *Radiology, 212*(3), 785–792. https://doi.org/10.1148/radiology.212.3.r99se24785

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology, 28*(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Tak, H. J., & Jang, S. H. (2014). Relation between aphasia and arcuate fasciculus in chronic stroke patients. *BMC Neurology, 14*(1), 1–5. https://doi.org/10.1186/1471-2377-14-46

Thompson, C. K., Lukic, S., King, M. C., Mesulam, M. M., & Weintraub, S. (2012). Verb and noun deficits in stroke-induced and primary progressive aphasia: The Northwestern Naming Battery. *Aphasiology, 26*(5), 632–655. https://doi.org/10.1080/02687038.2012.676852

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

Thorne, J., & Faroqi-Shah, Y. (2016). Verb Production in Aphasia: Testing the Division of Labor between Syntax and Semantics. *Seminars in Speech and Language, 37*(1), 23–33. https://doi.org/10.1055/s-0036-1571356

Tournier, J., Calamante, F., & Connelly, A. (2012). MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology, 22*, 53–66. https://doi.org/10.1002/ima.22005

Wassermann, D., Makris, N., Rathi, Y., Shenton, M., Kikinis, R., Kubicki, M., & Westin, C.-F. (2016). The white matter query language: a novel approach for describing human white matter anatomy. *Brain Structure and Function, 221,* 4705-4721. https://doi.org/10.1007/s00429-015-1179-4

Wilson, S. M., Galantucci, S., Tartaglia, M. C., & Gorno-Tempini, M. L. (2012). The neural basis of syntactic deficits in primary progressive aphasia. *Brain and Language, 122*(3), 190–198. https://doi.org/10.1016/j.bandl.2012.04.005

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., Ogar, J. M., DeLeon, J., Miller, B. L., & Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron, 72*(2), 397–403. https://doi.org/10.1016/j.neuron.2011.09.014

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

Xing, S., Lacey, E. H., Skipper-Kallal, L. M., Zeng, J., & Turkeltaub, P. E. (2017). White matter correlates of auditory comprehension outcomes in chronic post-stroke aphasia. *Frontiers in Neurology, 8*, 1–9. https://doi.org/10.3389/fneur.2017.00054

Yang, Q., Tress, B. M., Barber, P. A., Desmond, P. M., Darby, D. G., Gerraty, R. P., Li, T., & Davis, S. M. (1999). Serial study of apparent diffusion coefficient and anisotropy in patients with acute stroke. *Stroke, 30*(11), 2382–2390. https://doi.org/10.1161/01.STR.30.11.2382

Yorkston, K. M., & Beukelman, D. R. (1980). An analysis of connected speech samples of aphasic and normal speakers. *Journal of Speech and Hearing Disorders*, 27–36.

Zaccarella, E., & Friederici, A. D. (2015). Syntax in the Brain. In *Brain Mapping: An Encyclopedic Reference* (pp. 461–468). Elsevier.

Zumbansen, A., & Thiel, A. (2014). Recent advances in the treatment of post-stroke aphasia. *Neural Regeneration Research, 9*(7), 2012–2015.

Chapter III – General Discussion

The objective of the present thesis was two-fold. The aim of the first article was to assess the test-retest reliability of the reconstructions and structural characteristics of major white matter language bundles using a state-of-the-art probabilistic CSD-tractography pipeline. The aim of the second article was to determine whether and which acute structural characteristics of the left AF and UF, reconstructed using the same tractography pipeline as in the first article, as well as the lesion load, predicted syntactic abilities in connected speech in chronic post-stroke aphasia. The main findings are 1) that the tractography pipeline we used yields stable reconstructions and structural characteristics of the bundles of interest, and 2) that the bundle volume of the left AF and UF assessed in the acute phase predict syntactic abilities (i.e., verb production and verb-argument structure and sentence complexity) in connected speech in the chronic phase of post-stroke aphasia.

In this final chapter, a summary of the main findings of the two articles will first be presented, followed by a discussion of the methodological, clinical, and theoretical contributions and implications of this thesis. The limitations of the two articles and suggestions for future research directions will then be discussed. Finally, general conclusions for this thesis will be presented.

1. Summary of the two articles

1.1. Article 1

In the first article of this thesis, we aimed to assess the test-retest reliability of the reconstruction and the diffusion measures extracted from major white matter language bundles using a state-of-the-art tractography pipeline. The tractography pipeline used included the main recent methodological advances in the field. Thus, we used probabilistic CSD-tractography with ACT, and a semi-automatic approach for the reconstruction of the white matter bundles (i.e. WMQL). To this aim, we scanned a group of older cognitively unimpaired participants at two

timepoints, one week apart. We reconstructed the four major white matter language bundles (the AF, IFOF, ILF, and UF) in both hemispheres. We then extracted the following average diffusion measures of the white matter structural properties from each bundle: FA, MD, AD, RD, NuFO, bundle volume, and MLS.

First, we expected the reconstruction of the white matter bundles to show good morphological overlap between the two scanning sessions. This hypothesis was confirmed, since the degree of morphological overlap for the AF, IFOF, ILF, and UF in both hemispheres was good. Secondly, we expected the diffusion measures extracted from the white matter bundles to show good to excellent test-retest reliability. This hypothesis was partially confirmed, since the FA, MD, RD, AD, and MLS showed good to excellent test-retest reliability in the AF, IFOF, ILF, and UF, bilaterally. However, bundle volume showed good to excellent reliability in the AF and ILF bilaterally and the left IFOF and UF but only fair reliability in the right IFOF and UF. The NuFO measure also showed good reliability for all bundles but only fair reliability for the ILF, bilaterally. Overall, this study showed that using our state-of-the-art CSD-tractography pipeline, we can obtain stable reconstructions of major white matter language bundles and, for the most part, stable bundle-specific diffusion measures. Thus, we ascertained the stability of the reconstruction and diffusion measures obtained with the tractography pipeline we used.

1.2. Article 2

The aim of the second article of this thesis was to determine whether and which structural characteristics of the left AF and UF, as well as the lesion load, assessed in the acute phase post-stroke predict syntactic ability (as reflected by verb and sentence production and sentence and verb-argument structure complexity) in the connected speech of people with chronic post-stroke aphasia. We used the tractography pipeline from the first article to reconstruct the left AF and UF and extract

their average FA and AD, as well as the bundle's volume in our sample of patients with post-stroke aphasia. To assess syntactic ability in the connected speech of our participants with chronic post-stroke aphasia, we extracted the following measures from speech samples elicited by the Western Aphasia Bank's Picnic Scene picture description task: the percentage of verbs produced, the number of grammatically well-formed sentences produced, the percentage of verbs with a complex argument structure, and the number of clauses per utterance (an index of syntactic complexity at the sentence level).

Since this is the first study of its kind, we had no specific hypotheses as to which measures of structural characteristics (FA, AD, or bundle volume) or lesion load (if any) would predict our syntactic measures. We expected verb and sentence production, as well as verb-argument structure complexity to be predicted by the structural characteristics of the left AF and UF, and sentence complexity to be predicted by the structural characteristics of the left AF. First, our findings revealed that, out of the three diffusion measures of structural characteristics included (i.e., the FA, AD, and bundle volume), only the bundle volume emerged as a predictor of syntactic ability in connected speech. The lesion load did not emerge as a predictor either. Secondly, our findings revealed that, unlike what was hypothesized, only verb-argument structure complexity was predicted by the volume of both the left AF and UF. The production of verbs was predicted by the volume of the left AF (but not the UF). Additionally, as expected, sentence complexity was predicted by the volume of the left AF. Finally, the production of grammatically well-formed sentences was not predicted by any of our measures. Taken together, these findings showed that the bundle volume in the acute phase is a good predictor of preserved syntactic abilities in connected speech in chronic post-stroke aphasia. They also indicate that some syntactic abilities, such as the production of verbs with a complex verb-argument structure, may call upon the synergy

of the left AF in the dorsal stream and the left UF in the ventral stream, as suggested by some studies on syntax comprehension (Griffiths et al., 2013; Rolheiser et al., 2011), while other syntactic abilities may only rely on the left AF.

2. Contributions and implications of the two articles

2.1. The reliability of the reconstructions and diffusion measures of major white matter language bundles using a state-of-the-art tractography pipeline

The first article on the test-retest reliability allowed to show for the first time that a state-of-the-art probabilistic CSD-based tractography pipeline that includes most of the recent methodological advances in the field (i.e., a higher-order HARDI-based probabilistic tracking algorithm, ACT, tract-filtering, and a semi-automatic approach for bundle segmentation) can achieve a stable reconstruction as well as stable diffusion measures of white matter language bundles. Good test-retest reliability of our tractography approach demonstrates that variability in the reconstructions and the diffusion measures extracted from them that could be observed between participants is closer to reflecting true inter-individual variability rather than measurement error. The first article of the present thesis also highlights the fact that for one given diffusion measure, reliability may vary from one bundle to the other, including bundles that represent similar anatomical structures but are located in different hemispheres (e.g. the bundle volume showed good reliability in the left IFOF and UF but only fair reliability in these bundles' right hemisphere homologues). It is remarkable that these bundles showed good (in the left hemisphere) and fair (in the right) but not poor reliability, because they are bundles that are notoriously difficult to track due to their complex architecture (Hau et al., 2017; Rheault, Poulin, et al., 2020). This demonstrates

that the diffusion measures obtained using our tractography pipeline are stable, even for hard-to-track bundles.

The fact that no measure showed poor reliability indicates the stability of the pipeline used in this study for our bundles of interest compared to other test-retest reliability studies of tractography. For example, the reliability of average diffusion measures of one major language bundle, the AF, reconstructed using probabilistic CSD-tractography has been previously assessed in one study but yielded conflicting results (Besseling et al., 2012). In that study, the AF showed good reproducibility of microstructural measures, but the bundle's volume showed poor reproducibility. However, the CSD-tractography pipeline used in that study suffered from most of the shortcomings mentioned in Chapter I (Section 4). No tract-filtering or ACT were used, and a manual ROI-based approach was employed for bundle segmentation, which, as we previously mentioned, results in bias and less accurate reconstruction of white matter bundles. Previous DTI tractography reliability studies also showed similar inconsistencies. Additionally, the reliability of diffusion measures derived from CSD-tractography has been previously found to be comparable (and not better) to that yielded by DTI-tractography (Kristo et al., 2013). This would seem to indicate that using CSD-tractography alone might not confer better reliability than DTI-tractography. Our study is the first to show reproducibility of major white matter language bundles using a state-of-the-art probabilistic CSD-tractography pipeline leveraging the latest methodological advances in tractography.

Cousineau et al. (2017) used a pipeline similar to ours, yet not all the bundles they reconstructed had good morphological overlap. Indeed, less established, harder-to-track white matter bundles in their study showed poor reproducibility (e.g. a bundle of streamlines connecting the sensorimotor cortex to the caudate which had been identified by Sharman et al., 2013 using

functional MRI-guided tractography). Larger and well-established bundles included in Cousineau et al.'s (2017) study, such as the corticospinal tract (CST) showed good reproducibility (by established bundles, we mean bundles that correspond to anatomical structures that have been identified in post-mortem dissection atlases rather than white matter connections identified by *in-vivo* tractography alone). In our study, we included only established major white matter language bundles. Even though some of them are considered difficult to track (the UF and IFOF), their reconstruction showed good reproducibility. This highlights the importance of using good prior anatomical knowledge to define the bundles of interest.

Prior anatomical knowledge about established white matter language bundles is derived from white matter atlases based on post-mortem dissection studies that have identified these major white matter bundles in the human brain and were then further confirmed and refined by tractography (Martino et al., 2011). This knowledge can be leveraged to define the reconstruction of the bundles, either manually or with semi-automatic approaches such as the one we used in the present thesis (i.e., WMQL). The more extensive the existing knowledge on the anatomical definition of a white matter bundle is, the better its reconstruction can be. Such anatomical knowledge includes the specific cortical terminations of a white matter bundle, as well as the path of its fibers through and near other structures and bundles (e.g., the IFOF has terminations in the orbitofrontal cortex and the occipital cortex and crosses through the insula and the temporal stem). This is crucial to achieve as stable and anatomically accurate as possible reconstructions of the white matter bundles (Catani & Forkel, 2019; Maier-Hein et al., 2017; Rheault, Poulin, et al., 2020). This is particularly crucial for harder-to-track bundles because of their complex architecture (such as the UF which is particularly tricky to reconstruct because of its sharply curved shape and its terminations in the orbitofrontal cortex that are difficult to track), and for bundles that converge

to bottlenecks. The latter are regions where the fibers of two bundles run parallel to each other before they diverge and pose challenges for the reconstruction that cannot be addressed without sound knowledge of the anatomy of the fiber bundles (Maier-Hein et al., 2017; Rheault, Poulin, et al., 2020). Accordingly, in the present thesis, we leveraged previously established anatomical knowledge to track major white matter language bundles, which we believe contributed to the good reproducibility we obtained for these bundles.

In sum, the findings of the first article of the present thesis increase our confidence that the reconstructions and structural properties obtained from our tractography pipeline are stable and not due to random variations in measurement.

2.2. Tractography-based predictors of syntactic ability in chronic post-stroke aphasia

The second article of this thesis is the first study to ever show that the volume of white matter bundles in the acute phase predicts syntactic ability in the chronic phase of post-stroke aphasia. There are several levels of novelty to this study. First, very few studies have investigated acute structural predictors of chronic post-stroke aphasia, perhaps owing to the difficulty of data collection in the acute phase and the challenge of follow-up in the chronic phase. Thus, our study provides new data to this important line of research. Secondly, bundle volume has not been investigated as a potential predictor in previous studies as much as microstructural measures such as the FA or AD. This is possibly due to the fact that such an assessment was only made possible recently with the advent of tractography and, even more recently, tractometry which allows to obtain bundle-specific measures (by comparison, assessment of tensor metrics was possible by other means, such as TBSS, before tractography became more common in patient studies). The finding of bundle volume as a predictor in our study paves the way to further investigate this

measure as an acute predictor of other language functions and processes in the chronic phase. Third, no study has previously assessed acute predictors of syntactic ability. Finally, the second article of the present thesis is the first to combine several diffusion measures of micro- and macrostructural white matter characteristics to predict outcome in chronic post-stroke aphasia.

The findings of the second article demonstrate that the bundle volume may be a good acute predictor of chronic syntactic ability in connected speech production. One previous study where the volume of the left AF was assessed in the acute phase failed to predict chronic general language outcome using this measure (Forkel et al., 2014). This could potentially be accounted for by the fact that the dependent variable in that study was a general measure of aphasia severity which would be related to a network of bundles rather than one specific structure (Forkel & Catani, 2018). By contrast, in our study we predict specific syntactic measures that might have a more direct relationship to measures of the structural properties of specific white matter bundles (i.e., the left AF and UF).

With regards to other predictors that have been previously investigated, the second article of this thesis provides complementary information to the existing literature. Indeed, our study is the first to predict specific language measures extracted from connected speech in the chronic phase of recovery using a diffusion measure obtained in the acute phase. Previous studies have focused on a general measure of aphasia severity (Forkel et al., 2014; Moulton et al., 2019; Osa García et al., 2020) or naming impairment, the hallmark deficit of post-stroke aphasia, assessed by means of a standardized clinical test (Hillis et al., 2018). These studies have reported measures of white matter damage, lesion load or AD, as acute predictors of long-term language outcome (either in the chronic or subacute phases). By contrast, in our study, neither the lesion load nor AD (or FA) seem

to be determinants of measures of syntactic ability, such as verb or sentence production obtained from a test that approximates functional communication such as connected speech.

First, based on previous studies, the lesion load in the acute phase seems to be more of a predictor of degree of severity of subacute or chronic general language outcome (Hillis et al., 2018; Moulton et al., 2019; Osa-Garcia et al., 2020). Secondly, anisotropy and diffusivity in white matter undergo dynamic and rapid changes throughout the different phases post-stroke. The acute phase is characterized by an increase in anisotropy and decrease in diffusivity (most noticeable along the long axis of white matter fibers, namely the AD), reflecting restricted diffusion. This is followed by a progressive decrease in anisotropy and decreased diffusivity in the subacute phase. Finally, a drop in anisotropy and increase in diffusivity is observed in the chronic phase, as diffusion of water molecules becomes less restricted (Fung et al., 2011; Yang et al., 1999). However, heterogeneity in the timeline of the post-stroke evolution of white matter anisotropy and diffusivity across stroke patients has been observed (Yang et al., 1999), as these dynamic changes depend on a number of factors that vary across patients, such as time and rate of reperfusion and heterogeneity in lesion size, as well as in imaging time and stroke onset (Fung et al., 2011). Since our patients were assessed at different timepoints within the first three days post-stroke, it is possible that this variability could have been reflected in different patterns of increases and decreases in anisotropy and diffusivity across patients. By contrast, the only previous study that identified the AD in the left AF as an acute predictor of language outcome in chronic phase has scanned patients within a 24-hour window, which potentially could have reduced variability in anisotropy and diffusivity patterns in their sample (Moulton et al., 2019).

In sum, different white matter predictors could have different predictive value depending on a number of factors, such as the time post-stroke, the types of language measures used, and

potentially degree of impairment or recovery. This has important clinical implications. Given the dynamic nature of ischemic stroke and of the aphasia it induces, it is likely that there is no one single white matter predictor of language abilities in chronic post-stroke aphasia. A more optimal approach could be to combine different tractography-based predictors to capture a more comprehensive picture of the white matter's state post-stroke.

2.3. Implications for the neurocognitive basis of syntax in language production

The findings of the second article of this thesis also have theoretical implications for the neurocognitive basis of syntax in production. Indeed, this study is the first to shed light on the potential role of specific white matter language bundles in specific syntactic abilities in connected speech production in post-stroke aphasia. Other existing studies have either focused on syntactic abilities in comprehension (Griffiths et al., 2013), one aspect of syntactic function in production (e.g., sentence complexity, Ivanova et al., 2016), or by using standardized clinical tests (den Ouden et al., 2019). While informative, studies that report on the predictors of only one dimension of syntax in production are not sufficient to capture the full spectrum of syntactic abilities and the structural brain network that underpins them. Additionally, studying different dimensions of syntax during language production by using standardized tests (e.g., den Ouden et al., 2019) does not provide information about different dimensions of syntax and their white matter correlates in a discourse context. Indeed, syntactic abilities in a discourse context are bound to be different from those reflected by standardized clinical tests. Discourse is much closer to reflect the hierarchical structure building process that characterizes syntax at different levels (even though current approaches to discourse elicitation, such as picture description, limit our ability to fully gauge syntactic abilities of people with post-stroke aphasia in discourse). If the ultimate goal of this line

of research is to improve our ability to predict long-term functional language ability after stroke, we must assess these abilities using approaches that are closer to reflect functional communication (such as connected speech tasks).

The fact that the volume of left AF emerged as the sole predictor (or main one in the case of verbs with a complex argument structure) of syntactic measures suggests that this white matter bundle plays an important role in different aspects of syntax production in a discourse context. A crucial role for the left AF in syntax comprehension has been previously demonstrated in the chronic phase of post-stroke aphasia (Griffiths et al., 2013). Additionally, the fact that the volume of the left UF emerged as an independent predictor of verb-argument structure complexity, along with the left AF, suggests that both of these white matter bundles play a role in local structure building in connected speech production. In sum, our findings suggest that different aspects of syntax might rely differently on the dorsal and ventral white matter pathways, with some relying solely on the dorsal pathway (left AF), and at least one aspect of syntax (i.e., verb-argument structure complexity) relying on both the dorsal and ventral pathways (i.e., left AF and UF). Previous studies on the structural determinants of syntactic ability in language production in chronic post-stroke aphasia have highlighted a dominant role for the dorsal stream (Faroqi-Shah et al., 2014; Henseler et al., 2014; Ivanova et al., 2016) while others found the involvement of both dorsal and ventral streams (Akinina et al., 2019; den Ouden et al., 2019). However, studies uncovering a role for the ventral stream in syntax in production (verb production or complex verb-argument structure) have not revealed the involvement of one specific white matter bundle. By contrast, in our study we highlight the involvement of the left UF in one aspect of syntax in language production in post-stroke aphasia.

According to the dual-stream framework of syntax (Friederici, 2011; Friederici et al., 2006), the left AF and UF play differential roles in syntax structure building, with the left AF being involved in syntactic structural complexity at the sentence level (e.g. sentences with two or more clauses) and the UF being involved in basic local structure building (e.g. verb-argument structure integration) . However, the finding that verbs with a complex argument structure rely not only on the left UF but also the left AF is more in line with a view where the dorsal and ventral pathways may have a synergistic relationship when it comes to aspects of syntax in language production that require both complex hierarchical structure building and basic local structure building relationships. Such a synergistic relationship between dorsal and ventral white matter pathways in syntax has been previously suggested in language comprehension studies (Griffiths et al., 2013; Rolheiser et al., 2011).

This reflects the hierarchical incremental nature of syntactic structure building which has been described in contemporary psycholinguistic frameworks of syntax in language production (Thompson et al., 2015). Indeed, complex structure building may happen at the local level, with the verb projecting argument structures of increasing complexity, or at a more global level with sentences of increasing complexity constituted of two or more clauses organized in a hierarchical relationship to one another. This incremental hierarchical structure building of increasing complexity may be where the left AF plays a functional role, which would explain its involvement in both syntactic complexity at the sentence and verb-argument structure levels. The left UF, on the other hand, may be potentially more involved in combinatory processes necessary for the assignment of thematic roles by the verb and their integration.

3. Limitations

The main limitation of the first article of this thesis is that the reported test-retest reliability of the reconstructions and diffusion measures is specific to the tractography pipeline used. Changing one level of the pipeline could potentially change the degree of reliability. In other words, in order to be sure to achieve a similar degree of stability obtained in our study, the tractography pipeline used should follow similar steps (probabilistic algorithm for tracking, ACT, tract-filtering, etc.). For example, using a deterministic algorithm instead of a probabilistic one or using another semi-automatic approach instead of WMQL could potentially result in less reproducible reconstructions or structural characteristics. Additionally, a different acquisition scheme than the one we used (i.e. *b*-value = 1000, 64 diffusion directions) could also result in differences at the level of reproducibility. However, as we demonstrate in the second article of the present thesis, the classical HARDI acquisition scheme we used is feasible in terms of acquisition time, even with acute stroke patients, and can be easily implemented in clinical as well as experimental settings. The acquisition scheme we used takes about 30 minutes and can be carried out easily in a clinical context since our acute stroke patients were able to endure it (except for those with severe comprehension deficits). Additionally, all of the tools and software we used in our tractography pipeline are open source and accessible to users of tractography. As we report in the introduction of this thesis, the tractography pipeline we used addresses most of the major challenges of tractography and we therefore recommend using a similar pipeline in future studies.

The second article of this thesis has a few limitations. First, the fact that no standardized clinical tests of syntax were used somewhat limits our interpretation of the behavioral findings. Indeed, it is difficult to determine with certainty whether the preserved syntactic abilities in connected speech observed in our sample reflect complete absence of syntactic impairments or is

due to lack of sensitivity of connected speech to milder deficits. However, while a few standardized clinical tests for the assessment of syntax in language production exist in English, none have been developed for French-speaking populations, which prevented us from including such an assessment. At any rate, our findings reflect normal syntactic ability in connected speech elicited by a picture description task. Secondly, the nature and size of our sample somewhat limits the potential of generalization of our findings. First, since many patients presented mild/moderate aphasia in the chronic phase (though some had severe aphasia in the acute phase), it is possible that other measures (such as the lesion load) would predict severe impairments better than the bundle volume. Additionally, it is possible that other, secondary, independent predictors could be studied and identified with a bigger sample size. For example, it is possible that the volume of the left UF could have emerged as a potential secondary predictor of the percentage of verbs produced (over the total number of words) with a bigger sample size.

4. Future research directions

The work presented herein opens new perspectives for research in the field of tractography, the assessment of syntactic ability in language production, and the identification of acute white matter predictors of chronic language abilities in post-stroke aphasia.

First, since we demonstrated the stability of the reconstructions and structural characteristics of four major white matter language bundles bilaterally using a state-of-the-art probabilistic CSD-tractography pipeline, future research should investigate whether these findings extend to other established white mater bundles. For example, the components of the SLF (i.e., SLF I, II, and III) are white matter bundles that have been suggested to contribute to different cognitive functions, including speech, emotion, attention, visuospatial processing, and motor function (Makris et al., 2005; Mesulam, 1998; Petrides & Pandya, 2002). Assessing the test-retest

reliability of the reconstruction and diffusion measures of the three major components of the SLF using probabilistic CSD-tractography would be useful to leverage this approach for the study of these white matter bundles.

Another important avenue of research is the assessment of syntactic abilities in post-stroke aphasia. First, as previously mentioned, there is a need for the development of standardized clinical tests for the assessment of syntax production in French (Quebec French in our case). One approach could be to adapt the existing English tests (e.g. NAVS, Cho-Reyes & Thompson, 2012) to French. However, to the best of our knowledge, the sensitivity of these tests has not been assessed and there is therefore no guarantee that they would detect mild impairments in post-stroke aphasia. The development of standardized clinical tests sensitive to mild impairments in aphasia is a challenge in and of itself and is an important area of research on the assessment of language impairments (Ross & Wertz, 2004). Additionally, such tests have their own limitations. The sentence production priming subtest of the NAVS has important comprehension demands which can confound the results, since a patient may have comprehension but no production impairments. In this test, two pictures are presented to the patient; the examiner produces the sentence for the first picture, which is supposed to prime the production of a similar structure for the second picture by the patient. Therefore, new tests need to be developed. Another limitation of such tests is that they lack ecological validity. Thus, a more optimal approach to the assessment of syntax in language production could be to combine data from both standardized clinical tests and connected speech, elicited by different tasks, such as picture description and story retelling to be able to better reflect functional communication. Such an approach would provide a more complete clinical portrait of syntax deficits in language production.

The identification of early neuroimaging biomarkers of long-term post-stroke aphasia outcome has been deemed a priority (Boyd et al., 2017) but is a research avenue that is still in its infancy. Data collection in the acute phase is a particularly considerable challenge for this line of research. Initiatives to create an open-access database for stroke that include imaging and language data are therefore critical to the advancement of our knowledge on early prognostic markers of post-stroke aphasia. Such databases exist for other neurological conditions (e.g. the Alzheimer's Disease Neuroimaging Initiative, ADNI) but pose a greater challenge in stroke. A global initiative focused on Predicting Language Outcome and Recovery After Stroke (PLORAS) is ongoing. However, this project does not collect diffusion MRI data and is open only to the research teams that are currently a part of it. Thus, future initiatives ought to collect diffusion MRI data and make them available to the global research community. Another post-stroke aphasia global database is the AphasiaBank which is an open-access database of connected speech samples elicited by different discourse types (expositional, narrative, procedural, interviews) in eight different languages. However, this database does not include or collect neuroimaging data. Additionally, more studies carried out in the acute phase are needed to identify early biomarkers of language abilities in post-stroke aphasia. In the second article of the present thesis, the data collection was carried out in a hospital setting, thereby demonstrating that this is feasible in the acute phase.

The identification of biomarkers that allow for tailored rehabilitation and prognosis is the new frontier in stroke research (Simpkins et al., 2019). Both stroke and the aphasia it induces are highly complex and dynamic phenomena. Therefore, the identification of appropriate biomarkers is no small challenge. This research field still has some way to go before tractography-based predictors of post-stroke aphasia, such as the one identified in our study and others previously reported can be used as biomarkers in clinical practice. Future methodological studies need to focus

on ways to make tractography and tractometry easily implemented in clinical protocols. At this time, these analyses are time-consuming and are not accessible to clinicians.

5. Conclusion

The findings presented in this thesis pave the way for the development and use of early tractography-based biomarkers of language ability in the chronic phase post-stroke. Indeed, tractography has the potential to provide powerful early biomarkers that could improve prognosis of post-stroke aphasia outcome and better stratification of patients for rehabilitation. Leveraging the latest methodological advances in tractography, such as higher-order models based on HARDI data, ACT, probabilistic tracking, and semi-automatic tracking approaches, is crucial to attain this goal because they allow us to capture white matter architecture more accurately. The present thesis offers important methodological, clinical, and theoretical contributions that constitute a stepping stone towards a better understanding of the neural structural basis of language abilities post-stroke and the identification of sensitive and good predictors of long-term aphasia outcome.

References

Akinina, Y., Dragoy, O., Ivanova, M. V., Iskra, E. V., Soloukhina, O. A., Petryshevsky, A. G., … Dronkers, N. F. (2019). Grey and white matter substrates of action naming. *Neuropsychologia*, *131*(May), 249–265. https://doi.org/10.1016/j.neuropsychologia.2019.05.015

Akinina, Y., Dragoy, O., Ivanova, M. V., Iskra, E. V., Soloukhina, O. A., Petryshevsky, A. G., … Dronkers, N. F. (2019). Grey and white matter substrates of action naming. *Neuropsychologia*, *131*, 249–265. https://doi.org/10.1016/j.neuropsychologia.2019.05.015.Grey

Alexander, A. L., Lee, J. E., Lazar, M., & Field, A. S. (2007). Diffusion tensor imaging of the brain. *Neurotherapeutics*, *4*(3), 316–329.

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018). Noun and verb processing in aphasia: Behavioural profiles and neural correlates. *NeuroImage: Clinical*, *18*(December 2017), 215–230. https://doi.org/10.1016/j.nicl.2018.01.023

Bastiaanse, R. (2011). The retrieval and inflection of verbs in the spontaneous speech of fluent aphasic speakers. *Journal of Neurolinguistics*, *24*(2), 163–172. https://doi.org/10.1016/j.jneuroling.2010.02.006

Bastiaanse, R., & Jonkers, R. (1998). Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia Verb retrieval in action naming and spontaneous speech in agrammatic and anomic aphasia. *Aphasiology*, *12*(11), 951–969. https://doi.org/10.1080/02687039808249463

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR in Biomedicine*, *15*(7–8), 435–455. https://doi.org/10.1002/nbm.782

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, *34*(1), 144–155. https://doi.org/10.1016/j.neuroimage.2006.09.018

Besseling, M. H., Jansen, J. F. A., Overvliet, G. M., Vaessen, M. J., Braakman, H. M. H., Hofman,

P. A. M., … Backes, W. H. (2012). Tract Specific Reproducibility of Tractography Based Morphology and Diffusion Metrics. *Plos One*, *7*(4). https://doi.org/10.1371/journal.pone.0034125

Boles, L. (1998). Conversational discourse analysis as a method for evaluating progress in aphasia: A case report. *Journal of Communication Disorders*, *31*(3), 261–274. https://doi.org/10.1016/S0021-9924(98)00005-7

Boyd, L. A., Hayward, K. S., Ward, N. S., Stinear, C. M., Rosso, C., Fisher, R. J., … Cramer, S. C. (2017). Biomarkers of stroke recovery: Consensus-based core recommendations from the Stroke Recovery and Rehabilitation Roundtable. *International Journal of Stroke*, *12*(5), 480–493. https://doi.org/10.1177/1747493017714176

Bryant, L., Ferguson, A., & Spencer, E. (2016). Linguistic analysis of discourse in aphasia: A review of the literature. *Clinical Linguistics and Phonetics*, *30*(7), 489–518. https://doi.org/10.3109/02699206.2016.1145740

Buchanan, C. R., Pernet, C. R., Gorgolewski, K. J., Storkey, A. J., & Bastin, M. E. (2014). Test-retest reliability of structural brain networks from diffusion MRI. *NeuroImage*, *86*, 231–243. https://doi.org/10.1016/j.neuroimage.2013.09.054

Catani, M., & Budisavljevic, S. (2014). Contribution of diffusion tractography to the anatomy of language. In H. Johansen-Berg & T. E. J. Behrens (Eds.), *Diffusion MRI: From Quantitative Measurement to In-vivo Neuroanatomy* (2nd ed., pp. 511–529). Oxford, UK: Academic Press.

Catani, M., & Thiebaut de Schotten, M. (2012). *Atlas of Human Brain Connections*. New York, NY: Oxford University Press.

Catani, Marco, & Bambini, V. (2014). A model for Social Communication And Language Evolution and Development (SCALED). *Current Opinion in Neurobiology*, *28*, 165–171. https://doi.org/10.1016/j.conb.2014.07.018

Catani, Marco, & Forkel, S. J. (2019). Diffusion Imaging Methods in Language Sciences. In G. I. de Zubicaray & N. O. Schiller (Eds.), *The Oxford Handook of Neurolinguistics*. New York, NY: Oxford University Press.

Catani, Marco, Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., …

Rogalski, E. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain*, *136*(8), 2619–2628. https://doi.org/10.1093/brain/awt163

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology*, *26*(10), 1250–1277. https://doi.org/10.1016/j.physbeh.2017.03.040

Ciccarelli, O., Parker, G. J. M., Toosy, A. T., Wheeler-Kingshott, C. A. M., Barker, G. J., Boulby, P. A., … Thompson, A. J. (2003). From diffusion tractography to quantitative white matter tract measures: A reproducibility study. *NeuroImage*, *18*(2), 348–359. https://doi.org/10.1016/S1053-8119(02)00042-3

Cole, J. H., Farmer, R. E., Rees, E. M., Johnson, H. J., Frost, C., Scahill, R. I., & Hobbs, N. Z. (2014). Test-Retest Reliability of Diffusion Tensor Imaging in Huntington's Disease. *PLoS Currents*, 1–20. https://doi.org/10.1371/currents.hd.f19ef63fff962f5cd9c0e88f4844f43b

Côté, M. A., Girard, G., Boré, A., Garyfallidis, E., Houde, J. C., & Descoteaux, M. (2013). Tractometer: Towards validation of tractography pipelines. *Medical Image Analysis*, *17*(7), 844–857. https://doi.org/10.1016/j.media.2013.03.009

Cousineau, M., Jodoin, P., Garyfallidis, E., Côté, M., Morency, F. C., Rozanski, V., … Descoteaux, M. (2017). A test-retest study on Parkinson ' s PPMI dataset yields statistically signi fi cant white matter fascicles. *NeuroImage: Clinical*, *16*(July), 222–233. https://doi.org/10.1016/j.nicl.2017.07.020

Danielian, L. E., Iwata, N. K., Thomasson, D. M., & Kay, M. (2010). Reliability of fiber tracking measurements in diffusion tensor imaging for longitudinal study. *NeuroImage*, *49*(2), 1572–1580. https://doi.org/10.1016/j.neuroimage.2009.08.062

Dell'Acqua, F., & Catani, M. (2012). Structural human brain networks : hot topics in diffusion tractography. *Current Opinion in Neurology*, *25*(4), 375–383. https://doi.org/10.1097/WCO.0b013e328355d544

Dell'Acqua, F. D., Simmons, A., Williams, S. C. R., & Catani, M. (2013). Can spherical deconvolution provide more information than fiber orientations ? Hindrance modulated orientational anisotropy, a true-tract specific index to characterize white matter diffusion.

Human Brain Mapping, *34*, 2464–2483. https://doi.org/10.1002/hbm.22080

Dell'Acqua, F., & Tournier, J. D. (2019). Modelling white matter with spherical deconvolution: How and why? *NMR in Biomedicine*, *32*(4), 1–18. https://doi.org/10.1002/nbm.3945

den Ouden, D. B., Malyutina, S., Basilakos, A., Bonilha, L., Gleichgerrcht, E., Yourganov, G., … Fridriksson, J. (2019). Cortical and structural-connectivity damage correlated with impaired syntactic processing in aphasia. *Human Brain Mapping*, (August 2018), 1–21. https://doi.org/10.1002/hbm.24514

Descoteaux, M. (2015). High angular resolution diffusion imaging (HARDI). In *Wiley Encyclopedia of Electrical and Electronics Engineering*. https://doi.org/10.1002/047134608X.W8258

Descoteaux, M., & Deriche, R. (2008). Mapping neuronal fiber crossings in the human brain. *SPIE Newsroom*, 2–4. https://doi.org/10.1117/2.1200807.1205

Descoteaux, M., Deriche, R., Knösche, T. R., & Anwander, A. (2009). Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE Transactions on Medical Imaging*, *28*(2), 269–286. https://doi.org/10.1109/TMI.2008.2004424

Dick, A. S., Bernal, B., & Tremblay, P. (2014). The Language Connectome: New Pathways, New Concepts. *The Neuroscientist*, *20*(5), 453–467. https://doi.org/10.1177/1073858413513502

Dick, A. S., & Tremblay, P. (2012). Beyond the arcuate fasciculus : consensus and controversy in the connectional anatomy of language. https://doi.org/10.1093/brain/aws222

Dipper, L., Pritchard, M., Walkden, E., & Cruice, M. (2018). How do speakers with and without aphasia use syntax and semantics across two discourse genres? *Aphasiology*, *32*(6), 720–738. https://doi.org/10.1080/02687038.2018.1447642

Duan, F., Zhao, T., He, Y., & Shu, N. (2015). Test-retest reliability of diffusion measures in cerebral white matter: A multiband diffusion MRI study. *Journal of Magnetic Resonance Imaging*, *42*(4), 1106–1116. https://doi.org/10.1002/jmri.24859

Duffau, H. (2016). White Matter Pathways in the Human. *Neurobiology of Language*, 129–137.

https://doi.org/10.1016/B978-0-12-407794-2.00011-0

Duffau, H., Moritz-Gasser, S., & Mandonnet, E. (2014). A re-examination of neural basis of language processing: Proposal of a dynamic hodotopical model from data provided by brain stimulation mapping during picture naming. *Brain and Language*, *131*, 1–10. https://doi.org/10.1016/j.bandl.2013.05.011

Edwards, S. (1995). Profiling fluent aphasic spontaneous speech: a comparison of two methodologies. *International Journal of Language & Communication Disorders*, *30*(3), 333–345. https://doi.org/10.3109/13682829509021446

Edwards, S. (1998). Single Words are not Enough: Verbs, Grammar and Fluent Aphasia. *International Journal of Language & Communication Disorders*, *33*(S1), 190–195. https://doi.org/10.3109/13682829809179421

Edwards, S. (2005). Connected fluent aphasic speech. In *Fluent Aphasia* (pp. 92–121). New York, NY: Cambridge University Press.

Edwards, S., & Bastiaanse, R. (1998). Diversity in the lexical and syntactic abilities of fluent aphasic speakers. *Aphasiology*, *12*(2), 99–117. https://doi.org/10.1080/02687039808250466

Edwards, S., & Knott, R. (1994). Assessing spontaneous language abilities of aphasic speakers. *Language Testing*, *11*(1), 49–64. https://doi.org/10.1177/026553229401100105

Faroqi-Shah, Y., Kling, T., Solomon, J., Liu, S., Park, G., & Braun, A. (2014). Lesion analysis of language production deficits in aphasia. *Aphasiology*, *28*(3), 258–277. https://doi.org/10.1080/02687038.2013.853023

Farquharson, S., & Tournier, J.-D. (2016). High Angular Resolution Diffusion Imaging. In W. Van Hecke, L. Emsell, & S. Sunaert (Eds.), *Diffusion Tensor Imaging* (pp. 383–406). New York, NY: Springer. https://doi.org/10.1007/978-1-4939-3118-7

Farquharson, S., Tournier, J. D., Calamante, F., Fabinyi, G., Schneider-Kolsky, M., Jackson, G. D., & Connelly, A. (2013). White matter fiber tractography: Why we need to move beyond DTI. *Journal of Neurosurgery*, *118*(6), 1367–1377. https://doi.org/10.3171/2013.2.JNS121294

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of Language Production* (pp. 291–330). Philadelphia, PA: Psychology Press. https://doi.org/10.4324/9781315804453

Fonseca, J., Raposo, A., & Martins, I. P. (2019). Cognitive functioning in chronic post-stroke aphasia. *Applied Neuropsychology:Adult*, *26*(4), 355–364. https://doi.org/10.1080/23279095.2018.1429442

Forkel, S. J., & Catani, M. (2018). Lesion mapping in acute stroke aphasia and its implications for recovery. *Neuropsychologia*, *115*(March), 88–100. https://doi.org/10.1016/j.neuropsychologia.2018.03.036

Forkel, S. J., Thiebaut de Schotten, M., Dell'Acqua, F., Kalra, L., Murphy, D. G. M., Williams, S. C. R., & Catani, M. (2014). Anatomical predictors of aphasia recovery : a tractography study of bilateral perisylvian language networks. *Brain*, *137*, 2027–2039. https://doi.org/10.1093/brain/awu113

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A. D., Bahlmann, J. r., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, *103*(7), 2458–2463. https://doi.org/10.1073/pnas.0509389103

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254. https://doi.org/10.1016/j.conb.2012.10.002

Fung, S. H., Roccatagliata, L., Gonzalez, R. G., & Schaefer, P. W. (2011). MR Diffusion Imaging in Ischemic Stroke. *Neuroimaging Clinics of North America*, *21*(2), 345–377. https://doi.org/10.1016/j.nic.2011.03.001

Geva, S., Correia, M. M., & Warburton, E. A. (2015). Contributions of bilateral white matter to chronic aphasia symptoms as assessed by diffusion tensor MRI. *Brain and Language*, *150*, 117–128. https://doi.org/10.1016/j.bandl.2015.09.001

Girard, G., Whittingstall, K., Deriche, R., & Descoteaux, M. (2014). Towards quantitative

connectivity analysis: Reducing tractography biases. *NeuroImage*, *98*, 266–278. https://doi.org/10.1016/j.neuroimage.2014.04.074

Glasser, M. F., & Rilling, J. K. (2008). DTI tractography of the human brain's language pathways. *Cerebral Cortex*, *18*(11), 2471–2482. https://doi.org/10.1093/cercor/bhn011

Grefkes, C., & Ward, N. S. (2014). Cortical Reorganization After Stroke: How Much and How Functional? *The Neuroscientist*, *20*(1), 56–70. https://doi.org/10.1177/1073858413491147

Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., & Tyler, L. K. (2013). Functional organization of the neural language system: Dorsal and ventral pathways are critical for syntax. *Cerebral Cortex*, *23*(1), 139–147. https://doi.org/10.1093/cercor/bhr386

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

Hau, J., Sarubbo, S., Houde, J. C., Corsini, F., Girard, G., Deledalle, C., … Petit, L. (2017). Revisiting the human uncinate fasciculus, its subcomponents and asymmetries with stem-based tractography and microdissection validation. *Brain Structure & Function*, *222*(4), 1645–1662. https://doi.org/10.1007/s00429-016-1298-6

Heart and Stroke Foundation of Canada. (2017). *Different strokes: Recovery triumphs and challenges at any age*.

Heiervang, E., Behrens, T. E. J., Mackay, C. E., Robson, M. D., & Johansen-Berg, H. (2006). Between session reproducibility and between subject variability of diffusion MR and tractography measures. *NeuroImage*, *33*(3), 867–877. https://doi.org/10.1016/j.neuroimage.2006.07.037

Henseler, I., Regenbrecht, F., & Obrig, H. (2014). Lesion correlates of patholinguistic profiles in chronic aphasia: Comparisons of syndrome-, modality-and symptom-level assessment. *Brain*, *137*(3), 918–930. https://doi.org/10.1093/brain/awt374

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131–138. https://doi.org/10740277

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99. https://doi.org/10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews: Neuroscience*, *8*(May), 393–402. https://doi.org/10.1038/nrn2113

Hillis, A. E., Beh, Y. Y., Sebastian, R., Breining, B., Tippett, D. C., Wright, A., … Fridriksson, J. (2018). Predicting recovery in acute poststroke aphasia. *Annals of Neurology*, *83*(3), 612–622. https://doi.org/10.1002/ana.25184

Hsu, C.-J., & Thompson, C. K. (2018). Manual versus automated narrative analysis of agrammatic production patterns: The Northwestern Narrative Language Analysis and computerized language analysis. *Journal of Speech Language and Hearing Research*, 1–13. https://doi.org/10.1044/2017

Ivanova, M. V., Isaev, D. Y., Dragoy, O. V., Akinina, Y. S., Petrushevskiy, A. G., Fedina, O. N., … Dronkers, N. F. (2016). Diffusion-tensor imaging of major white matter tracts and their role in language processing in aphasia. *Cortex*, *85*, 165–181. https://doi.org/10.1016/j.cortex.2016.04.019

Ivanova, M. V., Isaev, D. Y., Dragoy, O. V., Akinina, Y. S., Petrushevskiy, A. G., Fedina, O. N., … Dronkers, N. F. (2016). Diffusion-tensor imaging of major white matter tracts and their role in language processing in aphasia. *Cortex*, *85*, 165–181. https://doi.org/10.1016/j.cortex.2016.04.019

Jang, S. H., & Lee, H. Do. (2014). Recovery of injured arcuate fasciculus in the dominant hemisphere in a patient with an intracerebral hemorrhage. *American Journal of Physical Medicine & Rehabilitation*, *93*(12), e15–e18. https://doi.org/10.1097/PHM.0000000000000202

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: A case-series comparison. *Brain*, *129*(8), 2132–2147. https://doi.org/10.1093/brain/awl153

Jeurissen, B., Descoteaux, M., Mori, S., & Leemans, A. (2017). Diffusion MRI fiber tractography

of the brain. *NMR in Biomedicine*, (July), e3785. https://doi.org/10.1002/nbm.3785

Jeurissen, B., Leemans, A., Jones, D. K., Tournier, J. D., & Sijbers, J. (2011). Probabilistic fiber tracking using the residual bootstrap with constrained spherical deconvolution. *Human Brain Mapping*, *32*(3), 461–479. https://doi.org/10.1002/hbm.21032

Jeurissen, B., Leemans, A., Tournier, J. D., Jones, D. K., & Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Human Brain Mapping*, *34*(11), 2747–2766. https://doi.org/10.1002/hbm.22099

Jeurissen, B., Tournier, J. D., Dhollander, T., Connelly, A., & Sijbers, J. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage*, *103*, 411–426. https://doi.org/10.1016/j.neuroimage.2014.07.061

Johnson, L., Basilakos, A., Yourganov, G., Cai, B., Leonardo, B., Rorden, C., & Fridriksson, J. (2019). Progression of aphasia severity in the chronic stages of stroke. *AJSLP*, *28*(May), 639–649.

Jovicich, J., Marizzoni, M., Bosch, B., Bartrés-faz, D., Arnold, J., Benninghoff, J., … Frisoni, G. B. (2014). Multisite longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging of healthy elderly subjects. *NeuroImage*, *101*, 390–403. https://doi.org/10.1016/j.neuroimage.2014.06.075

Kemmerer, D. (2015). *Cognitive Neuroscience of Language*. New York, NY: Psychology Press.

Kim, M., & Thompson, C. K. (2004). Verb deficits in Alzheimer's disease and agrammatism: Implications for lexical organization. *Brain and Language*, *88*(1), 1–20. https://doi.org/10.1016/S0093-934X(03)00147-0

Kristo, G., Leemans, A., Raemaekers, M., Rutten, G. J., De Gelder, B., & Ramsey, N. F. (2013). Reliability of two clinically relevant fiber pathways reconstructed with constrained spherical deconvolution. *Magnetic Resonance in Medicine*, *70*(6), 1544–1556. https://doi.org/10.1002/mrm.24602

Laures-Gore, J. S., Marshall, R. S., & Verner, E. (2011). Performance of individuals with left hemisphere stroke and aphasia and individuals with right brain damage on forward and

backward digit span tasks. *Aphasiology*, *25*(1), 43–56. https://doi.org/10.1080/02687031003714426

Leff, A. P., Schofield, T. M., Crinion, J. T., Seghier, M. L., Grogan, A., Green, D. W., & Price, C. J. (2009). The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: Evidence from 210 patients with stroke. *Brain*, *132*(12), 3401–3410. https://doi.org/10.1093/brain/awp273

Levelt, W. J., Roelofs, A., & Meyer, a S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11301520

Llinàs-Grau, M., & Martínez-Ferreiro, S. (2014). On the presence and absence of that in aphasia. *Aphasiology*, *28*(1), 62–81. https://doi.org/10.1080/02687038.2013.828345

Madhyastha, T., Mérillat, S., Hirsiger, S., Bezzola, L., Liem, F., Grabowski, T., & Jancke, L. (2014). Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. *Human Brain Mapping*, *35*, 4544–4555. https://doi.org/10.1002/hbm.22493

Maier-Hein, K. H., Neher, P. F., Houde, J. C., Côté, M. A., Garyfallidis, E., Zhong, J., … Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01285-x

Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., & Pandya, D. N. (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: A quantitative, in vivo, DT-MRI study. *Cerebral Cortex*, *15*(6), 854–869. https://doi.org/10.1093/cercor/bhh186

Marangolo, P., & Piras, F. (2010). Language and its interacting components: the right hemisphere hypothesis in derivational morphology. *Brain Research*, *1320*, 114–122. https://doi.org/10.1016/j.brainres.2010.01.019

Marchina, S., Zhu, L. L., Norton, A., Zipse, L., Wan, C. Y., & Schlaug, G. (2011). Impairment of speech producton predicted by lesion load of the left arcuate fasciculus. *Stroke*, *42*(8), 2251–2256. https://doi.org/10.1161/STROKEAHA.110.606103.Impairment

Martino, J., de Witt Hamer, P. C., Vergani, F., Brogna, C., de Lucas, E. M., Vázquez-Barquero,

A., … Duffau, H. (2011). Cortex-sparing fiber dissection: An improved method for the study of white matter anatomy in the human brain. *Journal of Anatomy*, *219*(4), 531–541. https://doi.org/10.1111/j.1469-7580.2011.01414.x

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(6), 1013–1052. https://doi.org/10.1093/brain/121.6.1013

Mirman, D., Kraft, A. E., Harvey, D. Y., Brecher, A. R., & Schwartz, M. F. (2019). Mapping articulatory and grammatical subcomponents of fluency deficits in post-stroke aphasia. *Cognitive, Affective and Behavioral Neuroscience*. https://doi.org/10.3758/s13415-019-00729-9

Mori, S., & Tournier, J.-D. (2014). *Introduction to Diffusion Tensor Imaging and Higher-Order Models*. (Susumu Mori & J.-D. Tournier, Eds.). Oxford, UK: Academic Press.

Mori, Susumu, & Tournier, J.-D. (2014). Fiber-tracking: 3-Dimensional Tract Reconstruction. In *Introduction to Diffusion Tensor Imaging: And Higher Order Models* (2nd ed., pp. 79–96). Oxford, UK: Amercian Press.

Moulton, E., Magno, S., Valabregue, R., Amor-Sahli, M., Pires, C., Lehéricy, S., … Rosso, C. (2019). Acute Diffusivity Biomarkers for Prediction of Motor and Language Outcome in Mild-to-Severe Stroke Patients. *Stroke*, *50*(8), 2050–2056. https://doi.org/10.1161/STROKEAHA.119.024946

Multon, K. D. (2012). Test-retest reliability. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 1496–1498). Thousand Oaks: SAGE Publications, Inc.

Osa García, A., Brambati, S. M., Brisebois, A., Désilets-Barnabé, M., Bedetti, C., Rochon, E., … Marcotte, K. (2020). Predicting early post-stroke aphasia outcome from initial aphasia severity. *Frontiers in Neurology*, *11*(February), 120. https://doi.org/10.3389/FNEUR.2020.00120

Papinutoo, D. N., Maule, F., & Jovicich, J. (2013). Reproducibility and biases in high field brain diffusion MRI : An evaluation of acquisition and analysis variables. *Magnetic Resonance Imaging*, *31*(6), 827–839. https://doi.org/10.1016/j.mri.2013.03.004

Petrides, M., & Pandya, D. N. (2002). Comparative cytoarchitectonic analysis of the human and

the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *European Journal of Neuroscience*, *16*(2), 291–310. https://doi.org/10.1046/j.1460-9568.2001.02090.x

Pierpaoli, C., Barnett, A., Pajevic, S., Chen, R., Penix, L. R., Virta, A., & Basser, P. (2001). Water diffusion changes in Wallerian degeneration and their dependence on white matter architecture. *NeuroImage*, *13*(6 Pt 1), 1174–1185. https://doi.org/10.1006/nimg.2001.0765

Piras, F., & Marangolo, P. (2007). Noun-verb naming in aphasia: A voxel-based lesion-symptom mapping study. *NeuroReport*, *18*(14), 1455–1458. https://doi.org/10.1097/WNR.0b013e3282ef6fc9

Potagas, C., Kasselimis, D., & Evdokimidis, I. (2011). Short-term and working memory impairments in aphasia. *Neuropsychologia*, *49*(10), 2874–2878. https://doi.org/10.1016/j.neuropsychologia.2011.06.013

Reuter, M., & Fischl, B. (2011). Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage*, *57*(1), 19–21. https://doi.org/10.1016/j.bandl.2011.02.005.Mechanisms

Rheault, F., De Benedictis, A., Daducci, A., Maffei, C., Tax, C. M. W., Romascano, D., … Descoteaux, M. (2020). Tractostorm: The what, why, and how of tractography dissection reproducibility. *Human Brain Mapping*, (August 2019), 1859–1874. https://doi.org/10.1002/hbm.24917

Rheault, F., Poulin, P., Valcourt Caron, A., St-Onge, E., & Descoteaux, M. (2020). Common misconceptions, hidden biases and modern challenges of dMRI tractography. *Journal of Neural Engineering*, *17*(1). https://doi.org/10.1088/1741-2552/ab6aad

Rolheiser, T., Stamatakis, E. A., & Tyler, L. K. (2011). Dynamic processing in the human language system: Synergy between the arcuate fascicle and extreme capsule. *Journal of Neuroscience*, *31*(47), 16949–16957. https://doi.org/10.1523/JNEUROSCI.2725-11.2011

Ross, K. B., & Wertz, R. T. (2004). Accuracy of formal tests for diagnosing mild aphasia: An application of evidence-based medicine. *Aphasiology*, *18*(4), 337–355. https://doi.org/10.1080/02687030444000002

Saffran, E. M., Berndt, R. S., & Scwartz, M. F. (1989). The quantitative analysis of agrammatic production; Procedure and data. *Brain and Language*, *37*, 440–479.

Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., & Weiller, C. (2006). Dynamics of language reorganization after stroke. *Brain*, *129*(6), 1371–1384. https://doi.org/10.1093/brain/awl090

Seunarine, K. K., & Alexander, D. C. (2014). Multiple fibers: Beyond the diffusion tensor. In H. Johansen-Berg & T. E. J. Behrens (Eds.), *Diffusion MRI: From Quantitative Measurement to In-vivo Neuroanatomy* (2nd ed., pp. 105–121). Oxford, UK: Academic Press.

Sharman, M., Valabregue, R., Perlbarg, V., Marrakchi-Kacem, L., Vidailhet, M., Benali, H., … Lehéricy, S. (2013). Parkinson's disease patients show reduced cortical-subcortical sensorimotor connectivity. *Movement Disorders*, *28*(4), 447–454. https://doi.org/10.1002/mds.25255

Simpkins, A. N., Janowski, M., Oz, H. S., Roberts, J., Bix, G., Doré, S., & Stowe, A. M. (2019). Biomarker Application for Precision Medicine in Stroke. *Translational Stroke Research*. https://doi.org/10.1007/s12975-019-00762-3

Smith, R. E., Tournier, J., Calamante, F., & Connelly, A. (2012). Anatomically-constrained tractography : Improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage*, *62*(3), 1924–1938. https://doi.org/10.1016/j.neuroimage.2012.06.005

Smith, R. E., Tournier, J. D., Calamante, F., & Connelly, A. (2013). SIFT: Spherical-deconvolution informed filtering of tractograms. *NeuroImage*, *67*, 298–312. https://doi.org/10.1016/j.neuroimage.2012.11.049

Song, S.-K., Sun, S.-W., Ju, W.-K., Lin, S.-J., Cross, A. H., & Neufeld, A. H. (2003). Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *NeuroImage*, *20*(3), 1714–1722.

Song, S.-K., Sun, S.-W., Ramsbottom, M. J., Chang, C., Russell, J., & Cross, A. H. (2002). Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *NeuroImage*, *17*(3), 1429–1436.

Stark, B. C. (2019). A comparison of three discourse elicitation methods in aphasia and age-matched adults: Implications for language assessment and outcome. *American Journal of Speech-Language Pathology*, *28*(3), 1067–1083. https://doi.org/10.1044/2019_AJSLP-18-0265

Thiebaut de Schotten, M., Dell'Acqua, F., Valabregue, R., & Catani, M. (2012). Monkey to human comparative anatomy of the frontal lobe association tracts. *Cortex*, *48*(1), 82–96. https://doi.org/10.1016/j.cortex.2011.10.001

Thompson, C. K., Faroqi-Shah, Y., & Lee, J. (2015). Models of Sentence Production. In A. E. Hillis (Ed.), *The Handbook of Adult Language Disorders* (pp. 332–353). New York, NY: Psychology Press. https://doi.org/10.1007/978-1-4612-4568-1_2

Thompson, C. K., Lukic, S., King, M. C., Mesulam, M. M., & Weintraub, S. (2012). Verb and noun deficits in stroke-induced and primary progressive aphasia: The Northwestern Naming Battery. *Aphasiology*, *26*(5), 632–655. https://doi.org/10.1080/02687038.2012.676852

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

Thompson, C. K., Shapiro, E. D., Li, L., & Schendel, L. (1995). Analysis of verb and verb-argument structure: A method for quantification of Aphasic language production. *Clinical Aphasiology*, *23*, 121–140.

Tournier, J-D, Masterton, R. A. J., & Seitz, R. J. (2012). Imaging techniques provide new insights. In L. M. Carey (Ed.), *Stroke Rehabilitation: Insights from Neuroscience and Imaging* (pp. 35–53). New York, NY: Oxford University Press.

Tournier, J.-D., Calamante, F., & Connelly, A. (2012). MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology*. https://doi.org/10.1002/ima.22005

Tournier, J., Calamante, F., & Connelly, A. (2012). MRtrix : Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology*, *22*, 53–66.

https://doi.org/10.1002/ima.22005

Tournier, J. D., Calamante, F., Gadian, D. G., & Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage*, *23*(3), 1176–1185. https://doi.org/10.1016/j.neuroimage.2004.07.037

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, *162*, 60–71. https://doi.org/10.1016/j.bandl.2016.08.004

Ungerleider, L. G., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165. https://doi.org/10.1039/c0dt00096e

Vollmar, C., Muircheartaigh, J. O., Barker, G. J., Symms, M. R., Thompson, P., Kumari, V., … Koepp, M. J. (2010). Identical , but not the same : Intra-site and inter-site reproducibility of fractional anisotropy measures on two 3 . 0 T scanners. *NeuroImage*, *51*(4), 1384–1394. https://doi.org/10.1016/j.neuroimage.2010.03.046

Wang, J. Y., Abdi, H., Bakhadirov, K., Diaz-arrastia, R., & Devous, M. D. (2012). A comprehensive reliability assessment of quantitative diffusion tensor tractography. *NeuroImage*, *60*(2), 1127–1138. https://doi.org/10.1016/j.neuroimage.2011.12.062

Wassermann, D., Makris, N., Rathi, Y., & Shenton, M. (2016). The white matter query language : a novel approach for describing human white matter anatomy. *Brain Structure and Function*, *221*(9), 4705–4721. https://doi.org/10.1007/s00429-015-1179-4

Webster, J., Franklin, S., & Howard, D. (2007). An analysis of thematic and phrasal structure in people with aphasia: What more can we learn from the story of Cinderella? *Journal of Neurolinguistics*, *20*(5), 363–394. https://doi.org/10.1016/j.jneuroling.2007.02.002

Wheeler-Kingshott, C. A. M., & Cercignani, M. (2009). About "axial" and "radial" diffusivities. *Magnetic Resonance in Medicine*, *61*(5), 1255–1260. https://doi.org/10.1002/mrm.21965

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., … Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron*, *72*(2), 397–403. https://doi.org/10.1016/j.neuron.2011.09.014

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

Xing, S., Lacey, E. H., Skipper-Kallal, L. M., Zeng, J., & Turkeltaub, P. E. (2017). White matter correlates of auditory comprehension outcomes in chronic post-stroke aphasia. *Frontiers in Neurology*, *8*(FEB), 1–9. https://doi.org/10.3389/fneur.2017.00054

Yang, Q., Tress, B. M., Barber, P. A., Desmond, P. M., Darby, D. G., Gerraty, R. P., … Davis, S. M. (1999). Serial study of apparent diffusion coefficient and anisotropy in patients with acute stroke. *Stroke*, *30*(11), 2382–2390. https://doi.org/10.1161/01.STR.30.11.2382

Yang, Y., Hou, L., Li, Y., Ni, J., & Liu, L. (2013). Neuronal necrosis and spreading death in a Drosophila genetic model. *Cell Death & Disease*, *4*, e723. https://doi.org/10.1038/cddis.2013.232

Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A., & Feldman, H. M. (2012). Tract Profiles of White Matter Properties : Automating Fiber-Tract Quantification. *PLoS ONE*, *7*(11), 1–15. https://doi.org/10.1371/journal.pone.0049790

Zaccarella, E., & Friederici, A. D. (2015). Syntax in the Brain. In *Brain Mapping: An Encyclopedic Reference* (pp. 461–468). Elsevier Inc. https://doi.org/10.1016/B978-0-12-397025-1.00268-2

Zumbansen, A., & Thiel, A. (2014). Recent advances in the treatment of post-stroke aphasia. *Neural Regeneration Research*, *9*(7), 2012–2015.

Appendix 1

Supplementary figures - Article 1

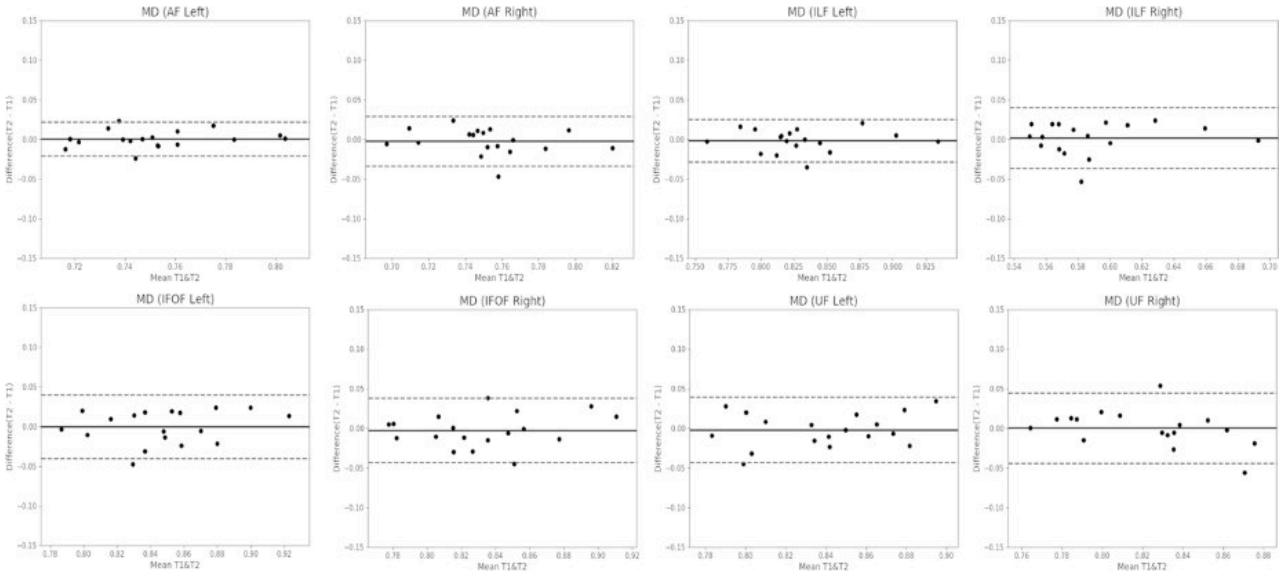


Figure 1. Bland-Altman Plots for the MD metric in all four fiber bundles, bilaterally.
The Y axis represents the mean difference between the measurements at the two timepoints and the X axis represents the mean of these measures. The upper and lower dashed lines represent the two limits of agreements at ±2 standard-deviations of the mean of differences (i.e. the 95% confidence interval). The solid line represents the mean of the differences between the two timepoints. The dots represent the individual subjects. MD = mean diffusivity; AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; T1 = time 1; T2 = time 2.

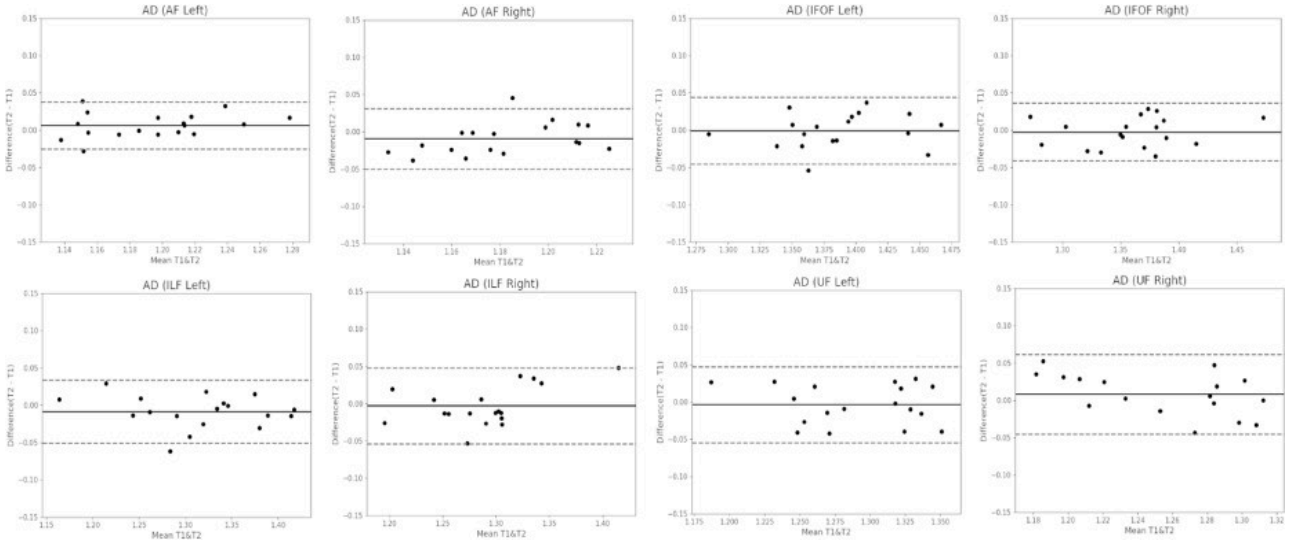


Figure 2. Bland-Altman Plots for the AD metric in all four fiber bundles, bilaterally.

The Y axis represents the mean difference between the measurements at the two timepoints and the X axis represents the mean of these measures. The upper and lower dashed lines represent the two limits of agreements at ±2 standard-deviations of the mean of differences (i.e. the 95% confidence interval). The solid line represents the mean of the differences between the two timepoints. The dots represent the individual subjects. AD = axial diffusivity; AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; T1 = time 1; T2 = time 2.

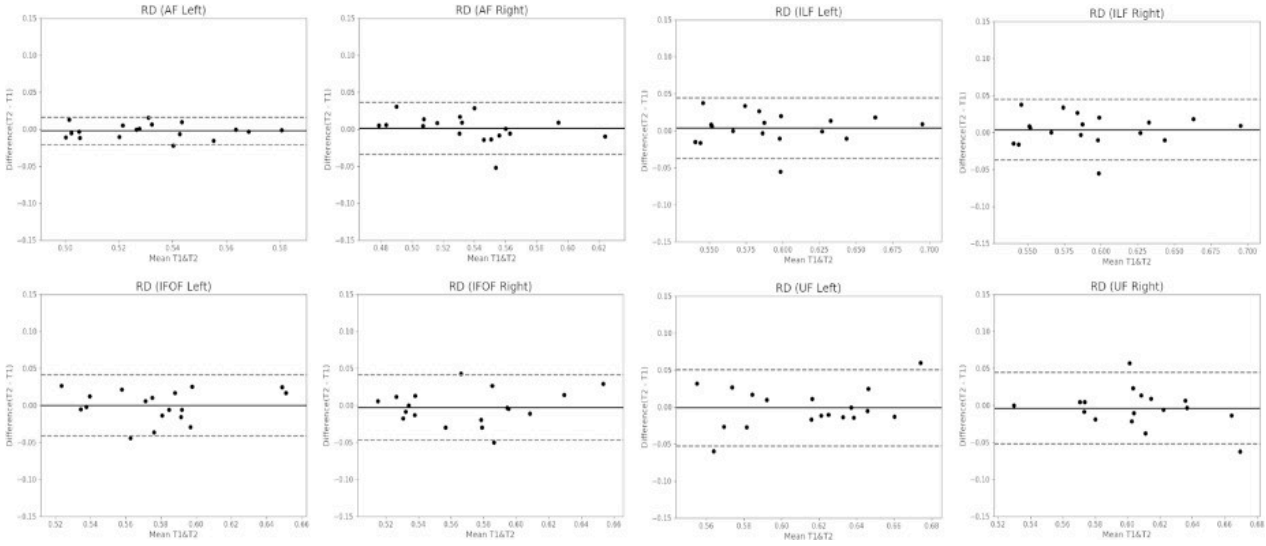


Figure 3. Bland-Altman Plots for the RD metric in all four fiber bundles, bilaterally.

The Y axis represents the mean difference between the measurements at the two timepoints and the X axis represents the mean of these measures. The upper and lower dashed lines represent the two limits of agreements at ±2 standard-deviations of the mean of differences (i.e. the 95% confidence interval). The solid line represents the mean of the differences between the two timepoints. The dots represent the individual subjects. RD = radial diffusivity; AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; T1 = time 1; T2 = time 2.

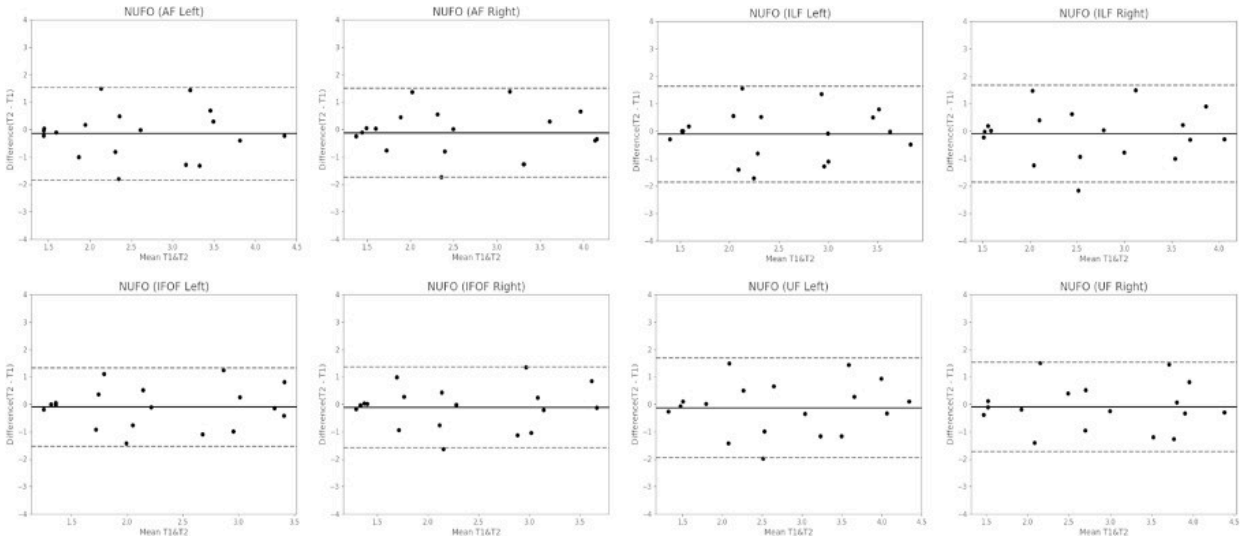


Figure 4. Bland-Altman Plots for the NuFO measure in all four fiber bundles, bilaterally.

The Y axis represents the mean difference between the measurements at the two timepoints and the X axis represents the mean of these measures. The upper and lower dashed lines represent the two limits of agreements at ±2 standard-deviations of the mean of differences (i.e. the 95% confidence interval). The solid line represents the mean of the differences between the two timepoints. The dots represent the individual subjects. NuFO = Number of fiber orientations; AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; T1 = time 1; T2 = time 2.

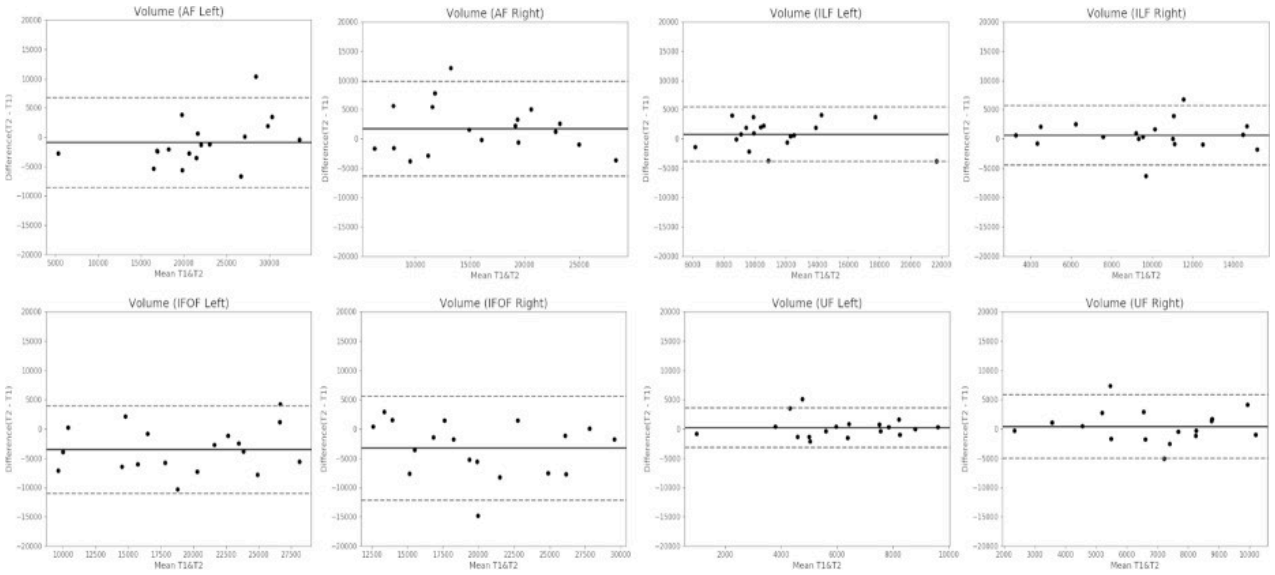


Figure 5. Bland-Altman Plots for the Volume measure in all four fiber bundles, bilaterally. The Y axis represents the mean difference between the measurements at the two timepoints and the X axis represents the mean of these measures. The upper and lower dashed lines represent the two limits of agreements at ±2 standard-deviations of the mean of differences (i.e. the 95% confidence interval). The solid line represents the mean of the differences between the two timepoints. The dots represent the individual subjects. AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; T1 = time 1; T2 = time 2.

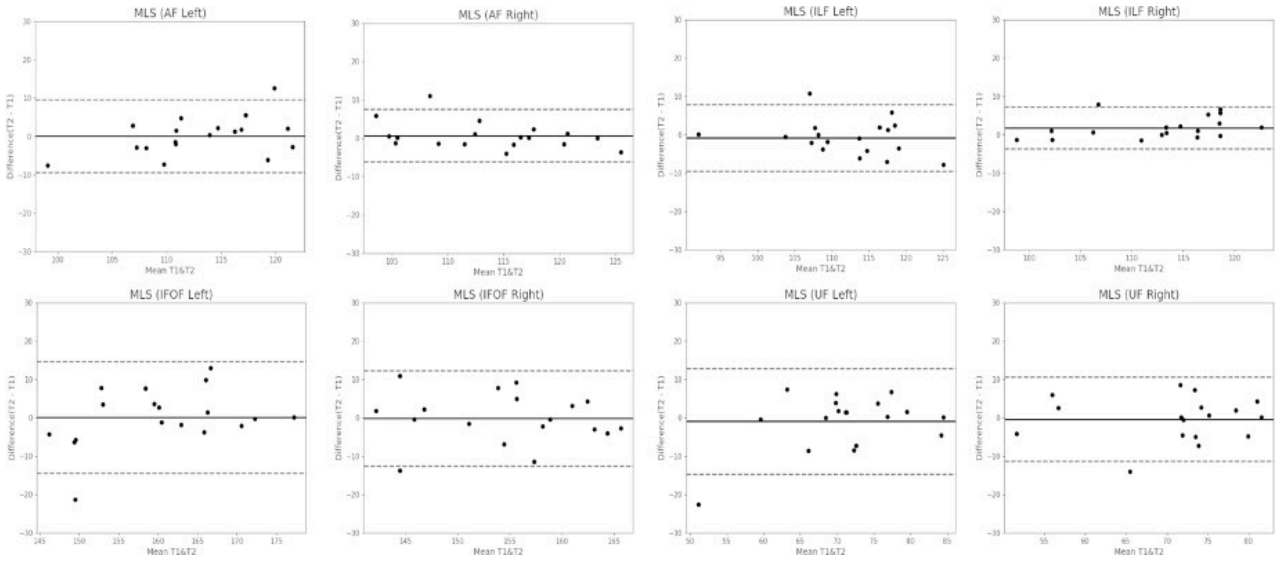


Figure 6. Bland-Altman Plots for the MLS measure in all four fiber bundles, bilaterally.

The Y axis represents the mean difference between the measurements at the two timepoints and the X axis represents the mean of these measures. The upper and lower dashed lines represent the two limits of agreements at ±2 standard-deviations of the mean of differences (i.e. the 95% confidence interval). The solid line represents the mean of the differences between the two timepoints. The dots represent the individual subjects. MLS = mean length of streamlines; AF = arcuate fasciculus; ILF = inferior longitudinal fasciculus; IFOF = inferior fronto-occipital fasciculus; UF = uncinate fasciculus; T1 = time 1; T2 = time 2.

Appendix 2

Supplementary material – Article 1

Modified White Matter Query Language Queries from Article 1

#Arcuate Fasciculus (AF)

AF.side = (inferior_frontal_gyrus.side or middle_frontal_gyrus.side or precentral.side) and (superiortemporal.side or middletemporal.side) not in hemisphere.opposite not in medial_of(supramarginal.side) not in ILF_final_1.side not in IFOF_final_1.side not in temporalpole.side not in frontalpole.side not in subcortical.side not in rostralmiddlefrontal.side not in lateralorbitofrontal.side not in ec.side not in superiorfrontal.side not in ctx_superiortemporal.side not in ctx_insula.side

#Inferior Longitudinal Faciculus (ILF)

ILF.side= only(temporal.side and occipital.side) and anterior_of(hippocampus.side) not in parahippocampal.side not in ctx_lingual.side not in ctx_lateraloccipital.side not in ctx_cuneus.side not in ctx_pericalcarine.side

#Inferior Fronto-Occipital Fasciculus (IFOF)

IFOF.side= endpoints_in(orbitofrontalgyrus.side or inferior_frontal_gyrus.side) and endpoints_in(occipital.side) and temporal.side and insula.side not in inferiorparietal.side not in hemisphere.opposite

#Uncinate Fasciculus (UF)

UF.side= insula.side and endpoints_in(orbitofrontalgyrus.side) and endpoints_in(temporal_anterior_section.side) not in posterior_of(putamen.side) not in centrum_semiovale.side not in superiorfrontal.side not in postcentral.side not in precentral.side

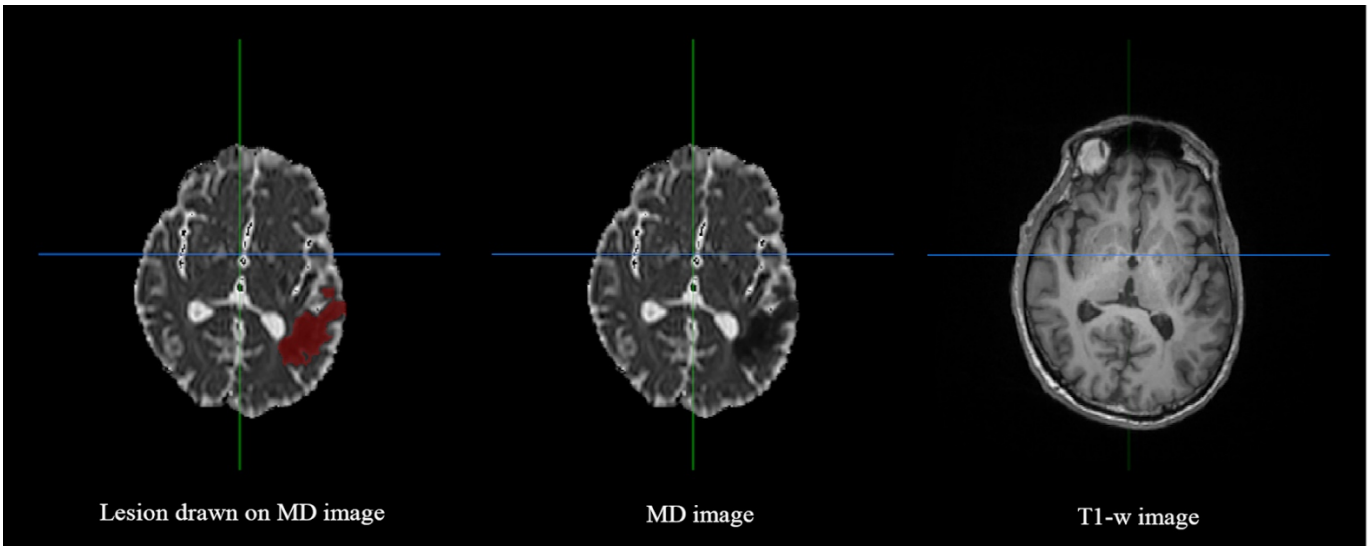
Appendix 3

Supplementary material – Article 1

Motion of each participant at each timepoint

	Time 1		Time 2	
Participant	Mean	SD	Mean	SD
1	0.06177309	0.04426779	0.0569417	0.03484703
2	0.05257956	0.03107867	0.05408551	0.02999035
3	0.0740012	0.04511743	0.07651165	0.04921323
4	0.08372465	0.06744107	0.07903643	0.04782817
5	0.07697474	0.03897778	0.04636438	0.02777494
6	0.04968429	0.03638177	0.04530571	0.03038225
7	0.06372794	0.03709305	0.05533408	0.02970052
8	0.07719768	0.06353363	0.05609605	0.02513792
9	0.06572525	0.03452804	0.06984521	0.04321858
10	0.04871767	0.02983834	0.0477965	0.029187
11	0.08445199	0.05201282	0.07557234	0.0489541
12	0.04386039	0.02260433	0.04060202	0.02549169
13	0.06665881	0.037032	0.07046448	0.04341005
14	0.07411467	0.04105924	0.08721011	0.05380671
15	0.0562165	0.02560989	0.06681639	0.03486903
16	0.10179909	0.05675531	0.08396169	0.05246243
17	0.06013741	0.03774356	0.05252436	0.03768069
18	0.06618107	0.02970252	0.09723072	0.04166339

Appendix 4



Lesion overlayed over the mean diffusivity (MD) image used to delineate it, the MD image, and the T1-weighted (T1-w) image (for comparison purposes) of one stroke participant.