Natural Language Processing Technologies in Compensatory Supports for People with Aphasia and Acquired Reading Impairments

Giorgia Cistola

TESI DOCTORAL UPF / 2022

Dra. Mireia Farrús and Dra. Ineke van der Meulen

Dra. Mireia Farrús, Language and Computation Centre, Universitat de Barcelona

Dra. Ineke van der Meulen, Erasmus MC, University Medical Center Rotterdam, Department of Rehabilitation Medicine, The Netherlands; Rijndam Rehabilitation, Rotterdam, The Netherlands

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES

upf. Universitat Pompeu Fabra Barcelona

To all the people with aphasia. Especially the patients that I had the privilege to encounter in my work experience who inspired me to keep going and ask questions.

Acknowledgements

Embarking on a PhD is a big leap, and these four years felt like a journey looking for a treasure: at times searching for clues and direction, others running excitedly on a clear path towards the final destination. For sure, it has been an intense and compelling adventure all the way to the end. Looking back on these years, I can only feel grateful for the opportunity and how incredibly lucky I was for the guides and people I have met on the path.

First of all, I would like to thank Mireia, without whom this PhD would have never happened. As a supervisor, you saw potential in my "odd" background, and you believed in my capabilities and guided me on the right path. Your trust in my idea, support, and unceasing understanding has been invaluable throughout the journey, and I will always cherish you as an example to follow for kind leadership.

Then, I want to deeply thank my co-supervisor Ineke, for coming on board with this project and seeing interest and validity in it. Your clinical research experience has added crucial value to the project, and I have learned very much from it. I also want to thank you for welcoming me to a different country: your caring for my stay in the Netherlands and making me feel part of the group was very important to me, and I will always be grateful to you for this.

I could not have wished for a better team, both in terms of scientific expertise and human relationships. You both set a wonderful example of guidance to aspire to.

I also want to give a special thank you to Horacio for his advice on text simplification and to Kim for helping me with the technical setup for one of my experiments: without them the second part of the project would not have been possible.

Next, I want to thank all the members of the TALN group that I had the privilege to encounter during these four years. Your hard work, talent, positive attitude, and enthusiasm for research were inspiring! You are an incredible team, and you made my stay in Barcelona feel like home. A round of special thanks goes to the current TALNers Laura and Roberto, the former members Ahmed, Alex, Janine and Carla, and other UPF family members Victor, Veronica and Fabio. Thank you for your friendship and for being there in the ups and downs of this journey, you made every day at UPF special.

I also want to thank the members of the rehabilitation medicine department at the Erasmus MC in Rotterdam, especially the PhD fellows, for your warm welcome and help in improving my experiments with your insights and suggestions. It was a real pleasure working with you, and I am sorry we could spend only a limited time together. Special thanks go to Majanka for the kind help in the analysis stage.

Next, this project could not have happened without the wonderful participants from *Hospital de la Santa Creu i Sant Pau*. My thanks to doctor Carmen Garcia Sanchez for her collaboration and help in recruiting the participants and allowing me to run my experiment in the hospital; also, thanks to Nicholas Grunder and Anna Gassiot for helping in recruiting and assessing the participants.
A special thank you also goes to the medical and research centre *Vithas* in Valencia for being so available and responsive in helping me find patients for my experiments despite the challenging times of the first months of the covid-19 pandemic. Thank you especially to doctor Loles Navarro Pérez and doctor Enrique Noe for offering their time to help me despite the troubled time.

Thank you to two people I accidentally met in Barcelona and whom I have now the privilege of calling friends! Yan Bai and Eriselda Danaj, you are marvellous researchers, your friendship has been crucial to me in these four years, and I am looking forward to keeping each other motivated and curious for the future to come.

A big thank to my family for supporting me with every means and believing in me and my path. And last but not least, many thanks to the person I share my life with, Pietro, who is always ready to cheer me up and celebrate my accomplishments.

Abstract

Aphasia is often associated with reading impairments. Nonetheless, supports designed for people with aphasia and acquired reading impairments allowing autonomous access to written material are lacking. These individuals frequently use tools intended for other populations, thus encountering difficulties. This thesis includes two experiments exploring whether two natural language processing techniques - speech synthesis and automatic text simplification – combined facilitate access to written content. The first experiment assessed which speech synthesis technology was aphasia-friendly considering comprehension rate, reaction times and preferences; it showed that the concatenative speech synthesis was the one causing longer reaction times in the participants. Therefore, we selected the deep learning-based output voice for the second experiment. Then, in the second experiment, we investigated whether automatic text simplification improves comprehension of sentences when combined with a text-to-speech system. The simplification of passive sentences did not increase comprehension, differently from the simplification of complex relative clauses. However, this test was conducted online, and the findings need to be considered carefully. Overall, this thesis's contribution could benefit clinicians and researchers both in the aphasia and natural language processing fields.

Resum

L’afàsia s’associa sovint amb problemes de lectura. No obstant, falten suports pensats per a persones amb afàsia i discapacitat lectora adquirida que permetin l’accés autònom al material escrit. Aquests individus utilitzen freqüentment eines destinades a altres poblacions, amb les dificultats que això comporta. Aquesta tesi inclou dos experiments que exploren si dues tècniques de processament del llenguatge natural, la síntesi de la parla i la simplificació automàtica de textos, faciliten l'accés al contingut escrit. En el primer experiment es va avaluar quina tecnologia de síntesi de parla anava més bé per a l’afàsia, tenint en compte la velocitat de comprensió, els temps de reacció i les preferències; i es va veure que la síntesi de la parla concatenativa era la que provocava temps de reacció més llargs en els participants. Per tant, vam seleccionar la veu de sortida basada en l’aprenentatge profund per al segon experiment. Després, en el segon experiment, vam investigar si la simplificació automàtica del text millorava la comprensió de les frases quan es combinava amb un sistema de conversió de text a parla. La simplificació de frases passives no augmentava la comprensió, a diferència de la simplificació de les oracions de relatiu complexes. Tanmateix, aquest experiment es va realitzar en línia i les conclusions s’han d’analitzar amb precaució. A grans trets, la contribució d’aquesta tesi podria beneficiar metges i investigadors tant en l'àmbit de l’afàsia com del processament del llenguatge natural.

Preface

This thesis contribution is valuable in the context of compensative devices for people with aphasia and acquired reading impairments. Our research considers patients who do not benefit from the visual support of a text and have to rely preponderantly on the auditory channel; this population does not have a tool of reference when it comes to reading materials they were used to read before the aphasia acquisition.

This field of research is still scarcely investigated; however, the problem with PWA not being able to access texts autonomously is very much present in clinical practice and it is an issue often reported by the patients. Indeed, PWA with acquired reading impairments often experience low self-worth, frustration and depression, inevitably leading to a lower quality of life. Although this thesis moves only the first steps in outlining the main features necessary for a compensative tool for this population, we believe that the contribution is valuable, and it shows possible future lines of development.

To achieve the goal of the thesis and reach the full contribution of this thesis, we had to interact with patients. Indeed, this project requires patients' participation, the support of the neuropsychologist responsible and in-presence testing at the hospital. Unfortunately, the schedule for the second part of the thesis happened to be in the middle of the outbreak of the covid-19 pandemic. We put a great deal of effort into operating under the circumstances of emergency, but inevitably the planned development of the project was affected. Nonetheless, the covid-19 situation also leads us to extract some observations that will be useful for further investigations.

For the same reasons, participation in conferences has been limited. During the core period of the PhD, the conferences where we planned to submit the manuscripts were suspended.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

In this section, we report the motivations that started this thesis project, then the main goals of this work, along with the approach and the steps that we took to accomplish them. Next, we describe the organization and structure of the thesis, and, to conclude, there is the list of publications and presentations given during the PhD period.

1.1 Motivation

Patients with aphasia and acquired reading impairments do not have a compensative tool of reference when it comes to reading materials they were used to read before the aphasia acquisition. The devices some patients use are not designed considering their difficulties, hence the challenges they face when resorting to them are various. However, existing natural language technologies could work in synergy with the aphasia rehabilitation field and possibly help these patients access reading material autonomously. Hence, the observations just mentioned motivated the current thesis.

Reading is fundamental to every aspect of life: socialising, family, leisure and work (Kjellén et al., 2017; Knollman-Porter et al., 2015; Lynch et al., 2013). Arguably, its importance has only increased due to the growing prevalence of online textual information (Dietz et al., 2011). Reading impairments are often associated with aphasia, and they vary in degree and severity. Difficulties in decoding and comprehending written material can occur at every level: single words, sentences, paragraphs and the text (DeDe, 2012, 2013; Knollman-Porter et al., 2015; Webster et al., 2018). From qualitative studies, it has emerged that people with aphasia with acquired reading impairments desire and seek out effective methods and supports for decoding written text—not only in rehabilitation settings but also, and especially, in everyday situations (e.g., work, leisure and personal healthcare) (Kjellén et al., 2017; Parr, 1995). According to Knollman-Porter et al. (2015), these patients desire to read even if they were not avid readers before their stroke. They want to gather information about the community, connect with friends and family through social media, and practise communication skills (Knollman-Porter et al., 2015). Reading comprehension deficits often impose restrictions on daily life tasks and engagement in pre-aphasia activities and roles, causing lower self-esteem, reduced cognitive stimulation and frustration (Brown et al., 2019; Knollman-Porter et al., 2015). Ultimately, such restrictions might reduce the individual's quality of life (Knollman-Porter et al., 2019; Parr, 1995; Rose et al., 2011) and cause depression and anxiety (Døli et al., 2017).

Although researchers and clinicians in the field of aphasia are aware of these difficulties affecting patients' lives and the necessity to compensate for these challenges, research on compensative tools for this specific population and their efficacy is limited. That is, tools and devices specifically designed for this population are still missing. The great diversity in clinical pictures of aphasia seems to call for software with a high level of flexibility and personalisation.

Even though natural language processing and language rehabilitation pertain to two different fields, they can be successfully combined to provide helpful compensative methods. Indeed, new natural language processing technologies could potentially be adapted and integrated with aphasia rehabilitation and play a critical role in overcoming issues that have hardly been covered as yet, such as compensative tools for people with aphasia with acquired reading impairments. Technological advancements in natural language processing are evolving towards more customisable software that can conveniently be merged with research in aphasia. Indeed, the possibility of personalised text adaptations would represent a relevant improvement for individuals with aphasia, as the difficulties experienced when reading or comprehending auditory speech can vary substantially across individuals. Therefore, flexible tools such as those already developed for other populations, such as dyslexic individuals, could also be highly desirable and beneficial for people with aphasia with acquired reading impairments.

1.2 Goal and approach

The main goal of this PhD thesis was to investigate whether combining two natural language processing technologies – speech synthesis and automatic text simplification – produces a noticeable improvement in sentence comprehension in people with aphasia and reading impairments. Alongside this main objective, we had other three minor goals:

- to identify if there is a synthetic speech technology more aphasia-friendly than others, and if so, which one,

- to investigate why the previously found speech synthesis method is more suitable for people with aphasia,

- to examine if automatically simplified passive and relative clauses enhance comprehension for people with aphasia when spoken with a text-to-speech system.

To achieve these objectives, we proceeded with the following steps:

i. We reviewed the literature on research on compensative reading tools for people with aphasia and acquired reading impairments and other technologies that they use to compensate for their reading difficulties.

ii. We interviewed six volunteers from a hospital in Barcelona to explore the interest of these participants in having a device that could help them read text that they were used to read before the stroke onset.

iii. We selected three text-to-speech systems using different synthesis technologies that synthesise in the Spanish language to employ in our first experiment.

iv. We designed and conducted a pilot experiment on twenty-two volunteers from the hospital. The pilot was for the first experiment of the thesis and had the objective to identify the most aphasia-friendly text-to-speech system among the three, using a digitalised human voice as a baseline.

v. We modified the experiment investigating the more aphasia-friendly text-to-speech system following the observations gathered from the pilot study. Subsequently, we conducted the experiment with patients and chose the text-to-speech to use for the second experiment of the thesis.

vi. We investigated the available automatic text simplification systems for the Spanish language and selected the one to use for the second experiment.

vii. We designed the second experiment at first as presential and subsequently online, asking for collaboration to a Spanish research and medical centre. The second experiment aimed at investigating the comprehension level improvement of complex sentences when spoken with the synthetic voice selected from the first experiment paired with the text simplification system.

1.3 Structure of the thesis

The thesis is structured as follows:

In **Chapter 2**, we present the definition and the description of aphasia and acquired dyslexia. We report the different aphasia types and acquired dyslexia syndromes, the neurocognitive models of language perception, production, and reading. This chapter also highlights the cognitive skills involved in reading and reading comprehension. Lastly, we tackle the issue related to the text accessibility options for people with aphasia and acquired dyslexia and discuss what kind of tools are missing to obviate these individuals' reading challenges.

In **Chapter 3**, we present a description of the natural language processing tools considered in the thesis. In the first part of the chapter, we describe the different types of synthetic synthesis currently available and their advantages and disadvantages. Subsequently, we also describe the systems that we used in the experiments. This chapter also discusses the past and current research on text-to-speech systems and aphasia. In the second part, we examine all the steps of text simplification systems and describe the one we choose to conduct our study. Lastly, we report the literature on text simplification in aphasia.

Chapter 4 is divided into three main parts: the pilot study, the first experiment, and the study of the voice parameters that could have influenced the first experiment results. The first part of the chapter reports the aim, methodology, and lessons learnt from the pilot. We observed the shortcomings and recalibrated those parts to reach the final design. In the second part, there is a description of the methodology we used for the study, the materials, and the first experiment results. The third part of the chapter uses the significant results of the experiment and investigates the reason behind the significant results concerning the reaction times. In order to do so, we run intelligibility and naturalness analysis and voice parameters analysis (jitter and shimmer). Through

these analyses, we wanted to investigate if a specific voice parameter caused the delayed reaction time in one voice (as resulted from the first experiment).
Chapter 5 reports the methodologies and the findings from the second experiment involving the text simplification system. This experiment has been transformed from presential into online due to the outburst of the Covid-19. The results that we obtained after conducting the study online are findings and observations that we gathered following the difficult task of testing people with aphasia and acquired reading impairments online and unsupervised. We discussed the findings and reported considerations about conducting such a challenging experiment.
Lastly, **Chapter 6** summarises the main conclusions and describes potential future work.

1.4 Publications and presentations

The main contributions of this thesis project were presented in the following publications and presentations:

Cistola, G., Farrús, M., & van der Meulen, I. (2021). Aphasia and acquired reading impairments: What are the high-tech alternatives to compensate for reading deficits?. International Journal of Language & Communication Disorders, 56(1), 161-173.

Cistola, G., Peiró-Lilja, A., Cámbara, G., van der Meulen, I., & Farrús, M. (2021). Influence of TTS Systems Performance on Reaction Times in People with Aphasia. Applied Sciences, 11(23), 11320.

Participation and live presentation: "Identifying an adaptable aphasia-friendly Text-To-Speech system for Spanish patients with acquired reading impairments", in the online British Aphasiology Conference between the 8th and the 10th of September 2021.

Chapter 2

APHASIA AND ACQUIRED DYSLEXIA

Aphasia and acquired dyslexia are two complex acquired neurological conditions that present a great variety of combinations between them. Indeed, it is uncommon to witness two individuals with the same set of auditory and reading challenges. This chapter presents some concepts related to aphasia and acquired reading impairments. By describing them, we want to convey the complexity of potential clinical profiles characterising people with aphasia and acquired dyslexia.

This chapter is structured as follows: in the first part, there is a description of aphasia and some historical background, the incidence and prevalence, the different syndromes and their characteristics. In the second part, we report the definition, historical background, incidence and the main characteristics of acquired dyslexia syndromes. In addition, we also report the cognitive functions fundamental to reading. In the third part, we discuss the compensative devices available to individuals who acquired aphasia and dyslexia and the missing tools in this field.

2.1 Aphasia

Aphasia is a term for a group of acquired receptive and expressive language deficits that occur as a result of brain damage. The more common etiology is a stroke of the middle cerebral artery, as well as tumours and traumatic brain injury (Stefaniak et al., 2020). It is important to note that aphasia is not an articulatory deficit (*dysarthria*) or a voice impairment (*aphonia or dysphonia*)(Code, 2021). The term aphasia is usually reserved for language impairments caused by damage to the left brain hemisphere that result in deficits in one or more of the core domains of language: semantics (word to text meaning), syntax (grammatical structure), morphology (word structure) and phonology (structure and characteristics of sound) (Code, 2021). This definition of aphasia excludes some complex language impairments, such as those resulting from damage to the right hemisphere, which includes difficulties with using verbal jokes, understanding inferences and metaphors, and using pragmatic aspects of language that deal with the behavioural context of language use.

The first studies on aphasia and the brain lesions associated with it date back to the 19th century when Broca (1861, 1865), Wernicke (1874, 1881) and others unveiled a new understanding of the neural organisation of language. The finding that still stands nowadays is that language impairments typically result from damage to the left hemisphere. Broca (1865) observed that the more anterior part of the brain, the posterior left frontal gyrus (also known as Broca's area), was damaged in those patients who spoke poorly or limitedly. Similarly, in 1881, Wernicke observed that patients with a damage in more posterior regions in the left temporal lobe (if in the absence of damage to the frontal region) showed well-articulated and fluent but meaningless speech. Figure 2.1.1 shows the main left brain areas involved in language processing. These early observations provided the bases for what is known as Classic Model often referred to as the ‘‘Broca–Wernicke–Lichtheim–Geschwind model” (e.g. Geranmayeh et al., 2014; Hickok & Poeppel, 2004), the ‘‘Wernicke–Lichtheim– Geschwind model” (e.g., Hagoort, 2013, 2014, 2016), or simply the ‘‘Wernicke-Lichtheim model” (e.g., Graves, 1997). The Classic Model is essentially a collection of models, and there are significant changes in each of the model's historical variants, and some appear to be better supported by recent data than others (Weiller et al., 2011).

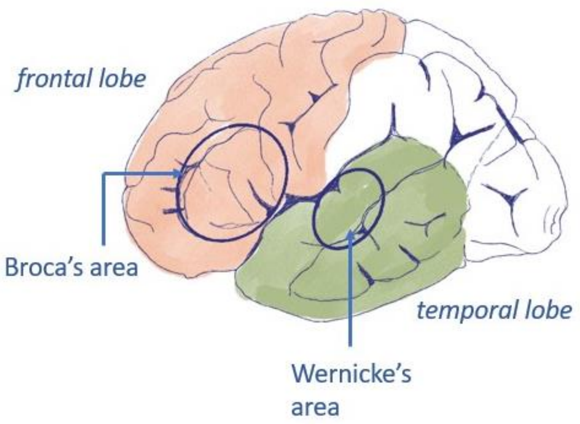


Figure 2.1.1 Schematic depiction of the main language-relevant brain areas: Broca's and Wernicke's areas in the left brain hemisphere.

2.1.1 Prevalence and Incidence of Aphasia

The exact prevalence of aphasia is still unknown; due to the different etiology and degree of severity, the incidence and prevalence of aphasia are difficult to determine.

The National Aphasia Association in the United States estimates that between 25 to 40 per cent of stroke survivors will acquire aphasia and that every year there are roughly 100.000 new cases (Engelter et al., 2006; National Aphasia Association, n.d.). Since aphasia can also occur as a result of a brain injury (McDonald et al., 2013) or dementia (Klimova & Kuca, 2016), determining the incidence and prevalence in such cases is even more difficult than in the population after stroke (Code, 2021). Similarly, globally, more than ten million new stroke cases are reported yearly (Benjamin et al., 2017), and at least one-third of the affected individuals will have symptoms of aphasia (Engelter et al., 2006). These are the numbers only for patients who suffered a stroke and visited a hospital that kept and contributed to the survey about incidence or prevalence. If it were to include all the etiology, such as people who suffer from traumatic brain injury, progressive aphasia and dementia, then both incidence and prevalence numbers would increase (Code, 2021). Patients' degrees of recovery vary, and the rate of improvement reduces over time (Maas et al., 2012; Yagata et al., 2017). Clinical knowledge holds that after 6–9 months after a stroke, language function reaches a plateau (Pedersen et al., 1995); nonetheless, minor changes (both positive and negative) might occur at later stages, and it might also cause the subtype and severity of aphasia to alter over time as people recover (Elkana et al., 2013; Hope et al., 2017).

2.1.2 Types of aphasic syndromes

The types of aphasia described in this sub-chapter refer to the classic model of language (Fig. 2.1.2). The aphasia syndromes classification (developed in the 1960s and 1970s and still used in clinical practice today) is based on the classic language model. Figure 2.1.2 depicts the classic model based on the latest version of Wernicke–Lichtheim–Geschwind, and it shows how the information flows among the different language areas and what lesion causes each of the eight aphasia syndromes.

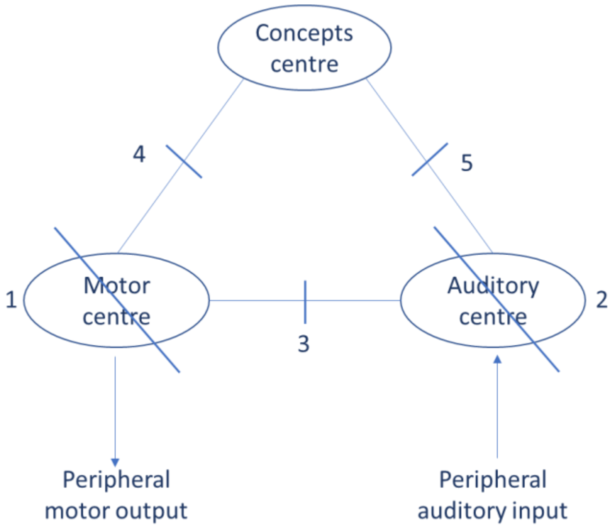


Figure 2.1.2 Representation of the Classical language model in the Wernicke-Lichtheim-Geschwind version. (*Features extracted from* Mendez, 2012 and Small & Hickok, 2016).

The eight syndromes are: Broca's, Wernicke's, conduction, global, anomic, transcortical motor, transcortical sensory, and mixed transcortical. Broca's aphasia is caused by damage to the motor centre (1), which is responsible for spoken language output, while Wernicke's aphasia follows an injury to the auditory centre (2), which is responsible for storing the sound representation of words. The interruption of the pathway that connects the auditory centre to the motor centre (3) causes conduction aphasia. Transcortical motor aphasia is caused by an interruption of the pathway that connects the concept centre to the motor centre (4), and transcortical sensory aphasia is caused by an interruption of the pathway that connects the auditory centre to the concept centre (5); mixed transcortical aphasia derives from the disconnection between both pathways connecting the auditory and motor centres with the concept centre; lastly, global aphasia results from the entire or the majority of the system (Kemmerer, 2014, pp.74-75). The aphasic syndrome that does not find its place in Figure 2.1.2 is the anomic syndrome. Unlike the others, anomia is not strictly associated with a specific brain lesion or the disruption of a pathway between two language areas (Kemmerer, 2014, p. 88). What can be observed is that a specific difficulty in retrieval is often associated with damages to a specific brain area (i.e., difficulties in retrieving nouns are often observed after lesions to the left temporal pole and the middle and inferior temporal gyri) (Damasio et al., 2004).

This classification system categorises the eight classic syndromes that could be diagnosed using the Boston Diagnostic Aphasia Examination (BDAE), developed in 1983 by Goodglass et al. but still widely used as an assessment method. Many language scientists agree that the classic model is outdated (Poeppel, 2012), and several shortcomings proved the model wrong by indisputable pieces of evidence (Tremblay & Dick, 2016). The main drawbacks are i) in the syndrome-based approach, one specific syndrome is diagnosed if one specific symptom is either present or absent; however, clinically, more than all-or-nothing phenomena, symptoms in every aspect of language come in a degree of severity; ii) it is centred on only two main language regions (Tremblay & Dick, 2016); iii) it concentrates on cortical structures and omits subcortical structures and related connections for the most part (Crosson, 2013); iv) the syndrome-approach considers a set of linguistic symptoms and damage to a particular brain region or network of regions, but these deficit-lesion correlations are not straightforward as this classification might lead to think. For instance, individuals with Broca's aphasia, in most cases, have lesions to the Broca region; however, there are cases where an individual with the same syndrome presented damages extended beyond this area and in some cases, even when the Broca's area is actually spared (Kemmerer, 2014, p. 75). Nonetheless, such classification continues to be used today. It represents a useful reference point for clinicians to communicate effectively among themselves and, in some cases, to identify the appropriate treatment plan (Kemmerer, 2014). In Table 2.1.1, we report a summary of the language difficulties in production and comprehension associated with each aphasia syndrome (Kemmerer, 2014, pp.76-90).

Table 2.1.1: Summary of language production and comprehension characterising the eight aphasia syndromes.

Type of aphasia	**Production**	**Comprehension**
Broca's	- Non-fluent; - strong reliance on memorised formulaic expression; - severe word retrieval deficits, worse retrieval of verbs than nouns; - marked impairment of closed-class elements (e.g. determiners, auxiliary verbs, propositions, suffixes); - reduced syntactic complexity.	- Relatively preserved understanding of colloquial conversation; - poor comprehension of some types of syntactically complex grammatical constructions.

	- articulatory difficulties	
Wernicke's	- Fluent; sometimes excessively - frequent phonemic and semantic paraphasias which could be so severe to result in neologistic jargon; - morphological and syntactic substitution errors; - patients are often unaware of their deficits.	- Impaired comprehension of sentences, phrases and in many cases, also words.
Conduction	- Fluency between Broca's and Wernicke's aphasia; - frequent phonemic paraphasias; - recurrent attempts to correctly produce the sound structure of the desired expression. - unable to repeat words	- Relatively preserved; - difficulties may arise when confronted with long, complex sentences with high short-term memory demands.
Global	Severely impaired	Severely impaired
Anomic	- Fluent but with hesitations; - marked word-finding attempts (sometimes worse for certain categories than others); - frequent "tip-of-the-tongue" states.	Relatively preserved
Transcortical Motor aphasia	- Poor planning and initiation of the speech; - speech severely impaired; mainly “empty phrases”; - they can benefit from nonlinguistic motor prompts (e.g., nodding their head) to initiate speech. - better repetition than spontaneous speech	Relatively preserved
Transcortical sensory aphasia	- Phonemic paraphasias, neologism and semantic substitutions; - extensive use of generic filler terms such as "thing", "one", "does", which make the speech rather empty.	Impaired
Mixed transcortical	A combination of the other two transcortical syndromes.	Impaired

Much of our discussion so far has been framed around aphasic syndromes, which are helpful concepts for drawing generalisations and smoothing out the idiosyncrasies of individual cases; however, we need to keep in mind that one of the most intriguing elements of aphasia is that the breakdown of language is not unitary. With advancements in neuroimaging methods over the last 20 years, we know that patient lesions are seldom localised and often include cortical and subcortical systems. No two people with aphasia have the same lesion profile, and variances exist not only in the magnitude of the lesion but also in the degree of damage to a specific location (Blumstein, 2016).

One of the more recent frameworks of the functional neuroanatomy of language postulates two streams of language processing (Bornkessel-Schlesewsky & Schlesewsky, 2013; Hickok & Poeppel, 2000, 2004, 2007; Wilson et al., 2011). The dual routes model – depicted in Figure 2.1.3 – holds that a ventral stream, which involves structures in the superior and middle portions of the temporal lobe, is involved in processing speech signals for comprehension. A dorsal stream, which involves structures in the posterior planum temporale region (at the parietal-temporal junction) and the posterior frontal lobe, translates acoustic-based representations of speech signals into articulatory representations, essential for speech production (Hickok & Poeppel, 2007). Such findings imply that impairments may represent brain networks rather than being caused only by local disease (Hickok & Poeppel, 2007; Metter et al., 1990). This model introduced an important novelty in the view of how language works, such as the idea that the right hemisphere ventral stream has a linguistic capacity or that the metalinguistic perceptual tasks depend on the dorsal stream. However, there is also some considerable continuity with the classic model (i.e., the ventral stream essentially corresponds to the mapping between the auditory centre and the centre of the concept in the Wernicke-Lichtheim-Geschwind model, Figure 2.1.3, while the dorsal stream corresponds to the link between the auditory centre and the motor centre).

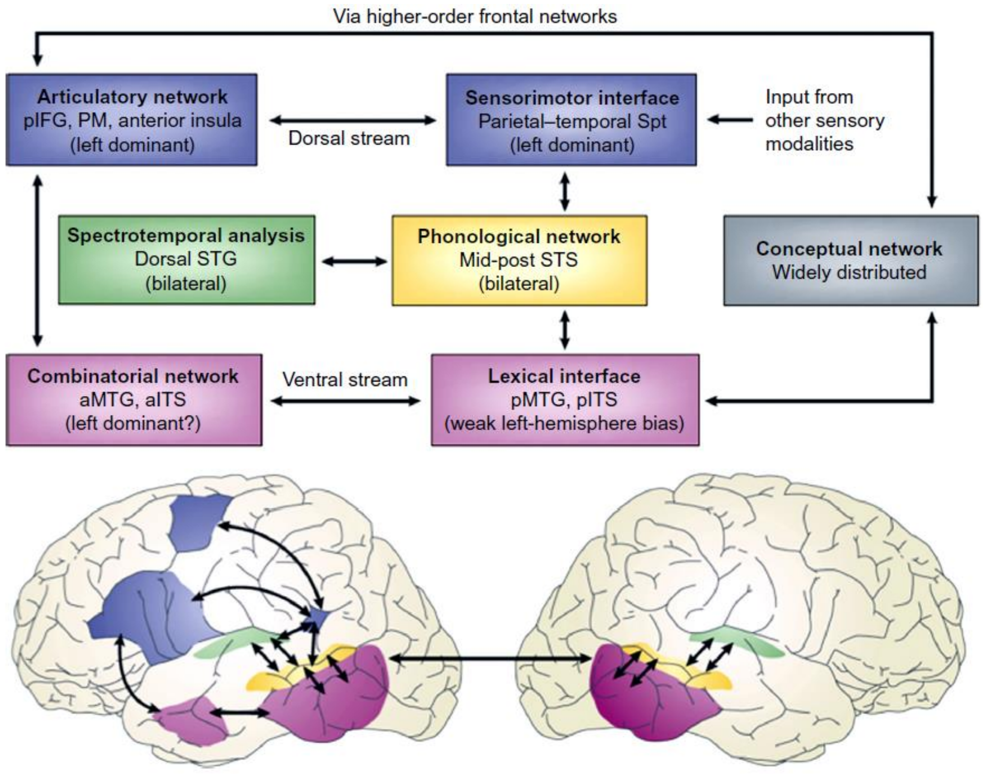


Figure 2.1.3.: Dual route model of speech processing described by Hickok & Poeppel (2000, 2004 and 2007). *(IFG, inferior frontal gyrus; ITS, inferior temporal sulcus; MTG, middle temporal gyrus; PM, premotor; Spt, Sylvian parietal-temporal; STG, superior temporal gyrus; STS, superior temporal sulcus; "a", anterior; "p" posterior)*. Figure extracted from Hickok and Poeppel (2007).

2.2 Acquired dyslexia

Acquired dyslexia is a disorder of the reading abilities occurring in previously literate individuals due to brain damage (Woollams, 2014). This disorder is rarely observed in isolation; for instance, it could be associated with aphasia and/or writing/spelling abilities (agraphia) (Beeson & Hillis, 2001; Hillis, 2019) or deficits in phonology (Cipolotti & Warrington, 1996; Woollams, 2014) or other aspects of visual processing such as face recognition (Behrmann & Plaut, 2014). An individual suffers from acquired dyslexia when the reading difficulties are the predominant impairment, even in association with other language impairments, or the language impairments could not explain the significant difficulties in reading (Ellis, 2016). It is well known that acquired dyslexia is very often associated with aphasia. Even though incidence data are still uncertain, various studies agreed on identifying acquired dyslexia in the majority of individuals with aphasia. For instance, a study conducted by Brookshire et al. (2014) identified that 68 % of their sample population who suffered from aphasia (99 participants in total) met their criteria of acquired dyslexia. Similar results about acquired dyslexia in individuals with aphasia have been shown in other studies such as Beeson & Hillis (2001), Cherney (2004) and Webb & Love (1983).

The first reference to acquired dyslexia dates back to the late nineteenth century (Dejerine, 1892), to be again investigated only starting from the 1970s by researchers such as Marshall & Newcombe (1973), Warrington & Shallice (1980), Ellis et al. (1988) and Coltheart (1987). It was in the 1980s that authors such as Coltheart (1987), Ellis (1982), Ellis et al. (1988), and Patterson et al. (1985) designed and developed the "dual-route models" for written words (see Figure 1.2.1).

In clinical practice, the term acquired dyslexia refers to preponderant difficulties at the word reading level compared to the auditory comprehension level. However, many patients in clinical practice have difficulties in sentence or text level reading and in these cases, usually, clinicians use the term acquired reading impairments. In our studies, we have a diverse spectrum of reading problems that range from PWA that have acquired dyslexia -hence single word reading or grapheme level difficulty- and those who have difficulties at a higher level (i.e., at sentence and text level). For this reason, we will use the term acquired reading impairments throughout the thesis, except in this chapter, where we describe the acquired dyslexia model based on the single-word reading.

2.2.1 Model of single-word reading and types of acquired dyslexia

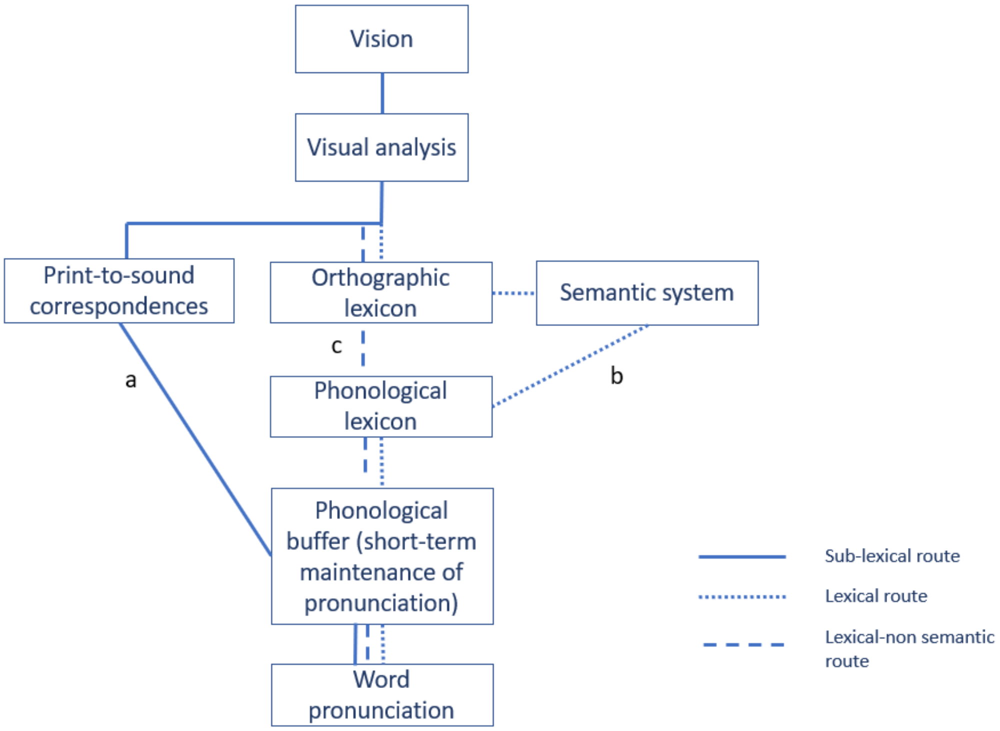


Figure 2.2.1: Simplified model for single word reading processes. The features are extracted from Ellis et al. (1988) and Beeson et al. (2002). This model is also known as the dual-route model.

The visual analysis system is the first step of the single-word reading model (depicted in Figure 2.2.1). This first aspect is essential to i) recognise the signs on the page as letters and ii) to pay attention to the position of the letter in a word. For instance, "TOP" and "POT" present the same letters but different positioning of such letters (Ellis, 2016). A disruption at this level usually causes one of the peripheral dyslexia syndromes as the impairments are at the very beginning of the word recognition process (Warrington & Shallice, 1980).

The *orthographic lexicon system,* which stores the written representation of familiar words (this step of the process is also referred to as "word recognition units"), is where we process if the word we see is a word or a nonword (Coltheart et al., 2001). For instance, the orthographic lexicon system stores the representation for the word "TOP" but not for "TEP". This stage is only an antecedent to the word meaning and pronunciation which are stored in the semantic system (Ellis, 2016). An impairment at the orthographic lexicon system-level usually causes surface dyslexia. Individuals presenting surface dyslexia show an exclusive reliance on the sublexical route; that is, they convert letters into sound using the print-to-sound correspondences system, hence treating all words, also the ones that used to be familiar, as unfamiliar (Marshall & Newcombe, 1973). The main feature of surface dyslexia is the inability to read words with irregular or exceptional print-to-sound correspondence, making often regularisation errors, and a reasonably well-preserved ability to read regular words (Karalyn Patterson et al., 2017). For instance, the word ISLAND would become "izland" or SUGAR "sudger" ("z" and "dg" being part of the international phonetic alphabet) (Ellis, 2016). A similar error pattern could also be shown when the disruption occurs at the phonological lexicon system level. There is also a difference: An individual with impairments in the orthographic lexicon system will not know the meaning of a familiar word that is pronounced incorrectly. In contrast, an individual with a lesion at the phonological lexicon system could still know the meaning of a familiar mispronounced word as the process passing by the semantic system is still intact. In fact, in the case of the impairments of the orthographic lexicon system, the semantically mediated mechanism is corrupted (Ellis et al., 1988).

If the word in front of the reader is unfamiliar, the reading process will proceed through the print-to-sound correspondences process. The *print-to-sound correspondences* system is a catalogue of correspondences between letters or sequences of letters and speech sounds. This process does not entail any stored information about words and is used to process unfamiliar words or nonwords as this process represents the sublexical route of reading, and the disruption to this system causes phonological dyslexia (Coslett & Turkeltaub, 2016). Phonological dyslexia represents the mirror counterpart of surface dyslexia. Individuals with phonological dyslexia rely exclusively on the lexical route of reading. Therefore, the main characteristic of phonological dyslexia is the inability to read unfamiliar words or nonwords, while they are able to read familiar words. In this syndrome, individuals tend to make lexicalisation errors, transforming nonwords into words familiar to them. For instance, the word "COBE" is read as "COMB", or "PLOON" as "SPOON" (Ellis, 2016).

Impairments to the lexical-semantic system alongside the disruption of the print-to-sound correspondences system usually cause deep dyslexia. Among the various characteristics observable in the deep dyslexia syndrome (see Table 2.2.1 for a detailed description), there is the impossibility to read unfamiliar or nonwords, semantic errors (for instance, reading "FOREST" as "TREE"), visual errors ("SIGNAL" read as "SINGLE") or the combination of the two ("SYMPATHY" as "ORCHESTRA"), and the imageability effect, meaning the difficulty to read words with abstract meaning (Ellis, 2016). According to Marshall & Newcombe (1973), the semantic errors and errors with abstract nouns suggest that the individuals attempt to read through the semantic system and such systems present impairments; moreover, the complete inability to read nonwords suggests that the sublexical print-to-sound correspondences system is disrupted. Since the clinical profile of deep dyslexia is broad and involves different deficits, many researchers have proposed different explanations throughout the years. Some researchers suggest that deep dyslexia is the consequence of multiple deficits (e.g., Morton & Patterson, 1980), others suggest that it is a continuum of phonological dyslexia (e.g., Crisp et al., 2011), while others argue that it derives from deficits of the phonological lexicon after that the correct semantic information has been gathered (e.g., Caramazza & Hillis, 1990).

The model described is based on error patterns observed in patients. For example, patients who cannot read nonwords but do read existing words are classified based on their difficulty with the nonwords; the other way round, patients who can read nonwords but have difficulties in existing words will be classified based on their challenge in reading existing words.

The dual-route reading model is composed of the sublexical and lexical routes. The sublexical route processes words at individual letters and sound levels, and the meaning extraction is achieved via an explicit phonological analysis. In Figure 2.2.1, the sublexical route is the straight line indicated with the letter "a"; it starts from the visual analysis system passing through the print-to-sound correspondences system to arrive at the phonological buffer and then the word pronunciation. This system is responsible for reading nonwords, low-frequency words and novel words (Beeson et al., 2008; Howard & Gatehouse, 2006; Rapp & Caramazza, 1997). The lexical route, on the contrary, processes words in their entirety and not in their single units (in Figure 2.2.1, this is represented by the dotted line indicated by the letter "b"). The lexical route is the primary reading route; this path is responsible for successfully reading i) familiar irregularly spelt words, namely the ones that do not follow the regular grapheme-phoneme conversion (e.g., height or yacht), and ii) high-frequency words both with regular and irregular spelling as they are memorised and stored in the semantic system so that the words are retrieved fast and efficiently. A standard reading process involves both routes (Beeson et al., 2002; Tainturier & Rapp, 2001). Lastly, some authors argue that there is a third route, the lexical-nonsemantic route, which in Figure 2.2.1 is indicated with a dashed line and the letter "c". In this route, the reader recognises and speaks the target word, but the meaning is lacking (Schwartz et al., 1980).

One of the main ways in which acquired dyslexias are distinguished was proposed by Warrington & Shallice (1980), who divided the types of acquired dyslexia into peripheral and central. According to the authors, peripheral dyslexia was characterised by a deficit in processing visual aspects of the stimulus (letters) in words, hence the visual analysis system is damaged. On the other hand, central dyslexias are caused by impairments in the procedure that connects visual words to meaning or speech production mechanism; therefore, after surpassing the visual analysis system, and resulting in impairments in comprehending or reading words (Warrington & Shallice, 1980). Table 2.2.1 summarises the different types of acquired dyslexia – peripheral and central – based on the Warrington & Shallice (1980) classification.

Table 2.2.1. Summary of the different types of dyslexia and their respective main characteristics.

Types of acquired dyslexia	**Characteristics**
Peripheral dyslexias	
Pure Alexia (or alexia without agraphia)	- Impairment in reading fast in the sense of automatic word recognition (lexical route) and error-prone; - compensatory strategy: usually, these individuals read letter by letter converting the letters into their name (e.g., H: "aitch" or V: "vee"); - longer words take more time to read than short ones; - performance is not influenced by factors like the word is abstract or concrete or whether it is orthographically correct; - they can read irregular words as well as regular words.[1]
Neglect dyslexia	A deficit in which patients commit consistently lateralised letter omission, addition, and/or substitution errors when reading single words. Patients can present left neglect dyslexia and read "IMPARTIAL" as "PARTIAL", or can have right neglect dyslexia and read "IMPARTIAL" as "IMPART" .[2]

Attention dyslexia	- Relative preservation of single-word reading, but disruption of reading when there are several words in text or sentence format; - they might show difficulties in reading letters within words even if they can read the word as a whole. Individuals with attentional dyslexia often make "migration" errors; for instance, presenting the words "GLOVE" and "SPADE", they could read "GLADE" .[3]
Central dyslexias	
Phonological dyslexia	- Selective deficit in translating print into sound (sublexical route); - good performance with real and familiar words as these individuals have resort to the lexical route of reading; - impairments in nonword reading or unfamiliar letter strings as the sublexical route is disrupted.
Surface dyslexia	- Disruption of semantically mediated mechanism (they treat once familiar words as unfamiliar); - inability to read words with irregular or exceptional print-to-sound correspondences as these individuals read irregular words as regular (regularisation errors)("YACHT" as "YAKT"); - ability to read regular words and nonwords preserved;
Deep dyslexia	Production of semantic errors: - target word substituted by a word semantically related in meaning ("FOREST" read as "TREE"); - frequently present "visual" errors ("SKATE" read as "SCALE"); - morphological errors in which a suffix or a prefix is added or deleted or substituted; - often associated with severe phonological processing impairments; - nonwords widely impaired: frequent lexicalisation errors of nonwords ("FLIG" read as "FLAG") - high imageability words (e.g., table, chair) read more successfully than low imageability words (e.g., fate, destiny). This also involves functors (e.g., which, that, under) that are elaborated as abstract words; - better reading of nouns compared to modifiers (e.g., adjectives and adverbs) or even verbs; - vocabulary limited to words they know.

[1] (Ellis, 2016; Karalyn Patterson & Kay, 1982) [2] (Caramazza & Hillis, 1990a; Ellis et al., 1987; Ellis & Young, 2013; Jackson & Coltheart, 2013; Moore & Demeyere, 2019; Vallar et al., 2010; Moore & Demeyere, 2017). [3] (Cipolotti & Warrington, 1996; Ellis, 2016; Price & Humphreys, 1995; Saffran, 1996)

2.2.2 Reading combinations of words

The attention to the single-word reading system and the disruption of one or more parts of such model usually gather more interest than the mechanisms behind sentence, paragraph or text reading, although in the majority of the cases, and in everyday life, reading occurs in combinations of words (sentences, paragraphs or texts). However, according to Perfetti (1985, 1992) and his verbal-efficiency account of reading theory, fast word identification processes are the foundation for text comprehension. Rapid and automatic word identification reduces the attentional load in the individuals and allows working memory to be accounted exclusively for integrating text propositions and meaning construction. On the other hand, less skilled readers have inefficient word identification abilities, have lower attentional resources employed in word identification, and consequently, a reduced working memory capacity necessary for comprehension (Jenkins et al., 2003). In support of the verbal-efficiency theory, various studies associate text comprehension with speed word reading, being the word in isolation (McCormick & Samuels, 1979; Perfetti & Hogaboam, 1975) or in context (Deno et al., 1982; Fuchs et al., 1988, 2001; Jenkins & Jewell, 1993).

In 2019, Smith & Clark studied comprehension and reading errors at paragraph level made by individuals with aphasia compared to healthy individuals. This study showed that the frequency and the types of errors were consistent with the language and speech production profiles of the individuals with aphasia. For instance, participants with anomic aphasia made semantic and morphological errors during reading, while individuals with Broca's aphasia, except for one participant, made morphological errors and unrelated function words, consistently with the grammatical characteristics of non-fluent aphasias (Maher, 2017). The semantic errors (typical of the deep dyslexia subtype) were consistent with the findings that deep dyslexia is observed more often in the Broca's or global aphasia rather than the fluent types of aphasia (Coslett & Turkeltaub, 2016). The results about the reading errors mirroring the aphasia severity or the overall language processing severity are in line with other studies such as Webb & Love (1983), or more recent studies as Cherney (2010a, 2010b) suggesting that language processes engaged in reading are also involved with the more general language processes abilities. In fact, according to Kay et al. (1996), alongside processes exclusive to reading (e.g., visual analysis or grapheme-to-phoneme conversion), others are shared with spoken language, such as lexical-semantic process or morphosyntactic processing. Regarding the reading comprehension at paragraph level, the results of Smith & Clark (2019) study, except for one participant, were in line with Perfetti's (1985,1992) verbal efficiency account of reading ability theory. In fact, a relationship was observed between the number of errors' frequency and reading comprehension scores and aphasia severity: higher number of errors' frequency (in both oral and silent reading) was associated with lower reading comprehension and more severe aphasia, and lower number in errors were resulted in better reading comprehension and less severe aphasia (Smith & Clark, 2019), which is in line with other works such as the study conducted by Brookshire et al. (2014).

However, not all individuals who acquired aphasia have impairments in single word reading; in fact, the majority who acquired reading impairments have difficulties decoding longer texts (paragraph and text level) without single word reading errors (Breznitz et al., 2013; Webb & Love, 1983). Text comprehension is a complex process that entails more than typical reading processes for single words (Meteyard et al., 2015). Neuroimaging studies investigating the brain areas involved in text reading showed that reading engages not only the typical language regions located on the left hemisphere (i.e., the inferior frontal gyrus, middle and superior temporal gyri, anterior temporal lobe) but also the right hemisphere counterparts of the same regions.

In 2000, Perfetti outlined a framework for understanding text-level reading. The author recognised three broad stages. The first stage is the visual processes; deficits or damages at this level usually result in peripheral dyslexias. The second stage involves decoding and identification of orthography, lexical representations, semantics, morphology and syntax (Perfetti, 2000). Impairments at this level cause one of the central dyslexias (phonological, surface and deep dyslexia). At this stage, the information about the text (i.e., access to vocabulary, morphology and syntax) is crucial to reading comprehension in adult readers (Guo et al., 2011; Long & Chong, 2001). These linguistic skills are usually impaired in people with acquired aphasia (Kertesz, 1982), and for some of these individuals, the linguistic deficits could cause or exacerbate text comprehension impairments (Meteyard et al., 2015). The third stage entails the cognitive skills that cooperate with linguistic skills to allow the reader to understand the meaning of the text (Perfetti, 2000). This model is still less specific than the model for reading at the single-word level. The research on the linguistic processes underlying text comprehension and the way in which these processes can be disrupted in PWA is still very limited.

2.2.3 Reading and related cognitive skills

Literature on the cognitive functions taking part in the reading process recognises the following main processes: reading speed, language skills, representation of the text base, inferencing, working memory and meta-cognitive skills (Meteyard et al., 2015). While reading speed and language skills (i.e., single word and sentence reading, or linguistic knowledge such as the use of the pronouns) are self-explanatory, the other cognitive skills need further discussion. The textbased representation is part of the discourse comprehension model and refers to the meaning of the text encoded by words and phrases in the text and the linguistic relationships between them (Kintsch & Van Dijk, 1978; Van Dijk & Kintsch, 1983). When the text-based representation meets the reader's knowledge about the world, this combination generates situation models that are rich in meaning and provoke inferences (Kintsch & Van Dijk, 1978; Perfetti, 2000). See Figure 2.2.2 for a schematic representation of an example of textbase and situation models comprehension levels.

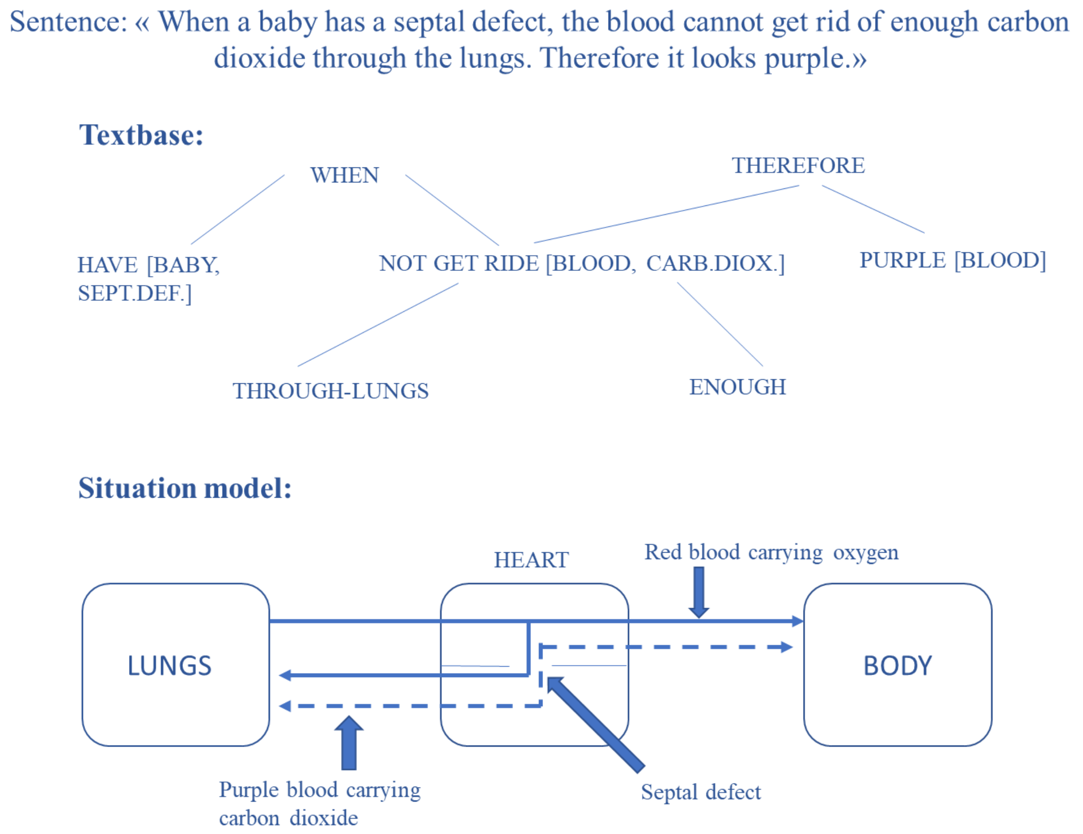


Figure 2.2.2.: Schematic representation of an example of textbase and situation models comprehension levels. (the features are extracted from Kintsch (1994). The textbase

representation shows the linguistic relationships between the elements of the sentence, whereas the situational model refers to the individuals' knowledge about the blood circulatory system.

Inferences are another key element in the cognitive skills necessary for text reading and comprehension. Inferences could be local, allowing text comprehension through linking together the elements (e.g., words, sentences or premises), or they could be global, completing the meaning of a text through background knowledge (Long & Chong, 2001). An example of inference is: "Marc played games, had cake and snacks, then watched as his friend opened his gifts."; from this sentence, it is possible to infer that Marc was at a birthday party. Text comprehension difficulties might arise, for example, from deficits in identifying implied meaning in a sentence (Sohlberg et al., 2014). In the case of our example sentence, a patient could not derive that Marc, the subject, was at a birthday party from the elements used in the sentence.

Working memory represents a critical cognitive skill when reading and comprehending a text since it is a key mediator in the reading process (e.g., Carretti et al., 2009; McVay & Kane, 2012). Working memory interacts with text comprehension as it is the capacity to briefly store information and allocate attention and control to relevant information(McVay & Kane, 2012). Impairments in working memory are connected with text comprehension difficulties in aphasia (Mayer & Murray, 2012). Deficits in working memory could negatively impact individuals who already present linguistic deficits (Caplan et al., 2013) as, for instance, the patient might already struggle in lexical retrieving, which will be further exacerbated by the difficulty to use the retrieved information to build the meaning of the text (Meteyard et al., 2015). Lastly, the meta-cognitive skills represent an important ability when, for example, it comes to self-monitoring the comprehension level (i.e., the individuals are aware and self-conscious that they understand the sentence or the text they are reading) or detecting possible errors when reading (e.g., Ehrlich et al., 1999; Mokhtari & Reichard, 2002; Schreiber, 2006).

2.3 Text accessibility for people with aphasia and acquired reading impairments

A literature review on compensative devices designed for PWA with acquired reading impairments revealed that there is no scientific literature investigating the use and the efficacy of tools designed for this target population (Cistola et al., 2021). The results of the literature review highlight the lack of devices, software or applications designed for PWA with acquired reading impairments. The review highlighted two ways in which PWA compensate for their reading problems: (1) using technology designed for a different target population (for instance, people with dyslexia); and (2) adapting the technology embedded in mainstream devices (e.g., iPad accessibility features and e-readers) to their own needs (Cistola et al., 2021).

The use of one of these technologies does not exclude the use of the other. This kind of overlapping was found in some investigations on the topic. For instance, in one of the first investigations on devices for PWA with acquired reading impairments, the study of Caute et al. (2016), the authors observed the use of both types of reading technologies in their sample population. Indeed, one patient used already a tool designed for another population (ClaroRead™) and then trained in using Amazon Kindle. More specifically, Caute et al. (2016) investigated whether PWA with acquired reading impairments could learn to use the Amazon Kindle, an e-book reader produced by Amazon, and if the use of this device might improve reading comprehension, participation and enjoyment of reading for these patients. The study involved four participants; at the time of the study, one was already using another reading compensative tool, ClaroRead™. ClaroRead™ (developed by Claro Software Ltd) is a software application created for people with dyslexia. It is compatible with Windows, Google Chrome, Mac, iOS, Android and PDF files. It provides text-to-speech functionality with the option to adjust the speech rate and choose a male or female voice. The patient who was already proficient in using ClaroRead™ reported no benefits in using the Kindle: he found it difficult to navigate and strongly disliked the text-to-speech voice, which he described as too mechanical. Moreover, at the time of the study, the Kindle e-reader did not allow the reading speed to be altered – a feature that the participant considered a fundamental help. This patient was able to use ClaroRead™ successfully with the text-to-speech function to listen to online news, recipes and short stories. However, he only used this device once or twice a week for 10 minutes, due to the fatigue he experienced when reading – even when using ClaroRead™. This patient had mild dyslexia and good comprehension of single words and sentences, but impairments in reading paragraphs and texts, and mild working memory impairments. The other three participants had a positive experience with Amazon Kindle and reported having regained confidence and enjoyment in reading.

In 2019, Caute et al. conducted a quasirandomised wait-list controlled design study that extended the findings just described in Caute et al. (2016). In this study, the authors investigated the improvements following a six-week treatment that consisted in training the participant in using a high-tech support. The 21 participants were randomised to an immediate or a delayed therapy group. The delayed group started the technology-enhanced therapy with six weeks delay. The authors investigated improvements in a) reading comprehension (especially when assisted by the trained technology), b) reading confidence and enjoyment and c) functional communication, mood and quality of life. The reading supports considered were ClaroRead™ and Amazon's Fire 7 tablet. The choice between the two technologies depended on the severity of the participant's reading impairment: less impaired patients were trained in the Fire 7, whereas those with more pronounced difficulties were trained in ClaroRead™. Another critical factor besides the technology training was the highly individualised goal setting. Every participant was trained with their specific reading goal in mind (e.g. reading a novel or sharing a book with a grandchild). The results of this study were quite positive and similar in both groups. The patients consistently used the two high-tech devices as a compensative tool for at least six weeks follow-up, and they also self-reported enjoyment and improved confidence in reading after the training treatment. Nonetheless, there was no wider change in mood, functional communication or quality of life. The results of this study are quite different from the results of Caute et al. (2016), which the authors speculated might be due to the extensive therapy training (14 hours twice a week) or the personalised reading goals associated with the technology training. Regardless of the differences between these two similar studies, the positive outcomes of Caute et al. (2019) encouraged studies more oriented towards investigating individually tailored compensative supports.

In 2016, Caute & Woolf carried out a single case study on a patient with severe dysgraphia and acquired dyslexia, whose spoken language comprehension was moderately impaired. The authors wanted to investigate whether training the patient in using voice recognition software (VRS) and a reading support tool with a text-to-speech system could help him improve his participation in everyday activities. Specifically, the participant desired to go back to work or partially recover his role at work, which was not possible with his current writing and reading difficulties. In the study, Dragon NaturallySpeakingRTM (a voice recognition software application) was utilised to assist with writing, and Read&Write GoldRTM (reading support for people with developmental dyslexia) to assist with reading. Initially, the patient was trained to use Dragon, and Read&Write GoldRTM was only introduced in a second stage. The reading support was introduced as the text-to-speech system incorporated into Dragon had a low-quality text-to-speech system that impeded the participant's already-compromised auditory comprehension. Moreover, the text-to-speech system embedded in Dragon could only read what was written in Dragon software. The patient found Read&Write GoldRTM extremely useful; he slowed the inter-word pause and was able to follow the text through the highlight. With Read&Write GoldRTM he was able to read the emails he received and to check possible errors in what he had written. In the months following therapy, the patient continued to use both devices to interact with colleagues and friends successfully. He used them for the majority of his communications, and reported a better quality of life. There was an increase in the number of people and activities he engaged with. He continued using Read&Write GoldRTM to read emails and online news and purchase items online. He also added an e-book reader to his routine, to read books even if it was not part of the training.

Other studies have investigated the experiences that PWA with acquired reading impairments have with mainstream devices such as e-readers (especially Amazon's Kindle) and audiobooks, and whether these tools are valuable compensative strategies for text comprehension (Kelly et al., 2016; Kjellén et al., 2017; Knollman-Porter et al., 2015). These studies are qualitative and do not provide data on possible comprehension improvements related to the use of audiobooks or e-readers. Nonetheless, the qualitative information from the patients is encouraging, and also important to consider when thinking about supports for specific populations, since the end user must be engaged in using the tool in the first place. Although these tools are not designed for PWA with acquired reading impairments, these patients do find benefits in some of their features. For instance, in 2017, Kjellén et al. interviewed 12 patients with aphasia and literacy difficulties. The results of this study highlighted that each patient has their own preferences, and takes advantage of different features to support their writing or reading difficulties. Some participants in this study have used, or are still using, supports such as text-to-speech software, audiobooks or electronic dictionaries when reading. Although most of the participants interviewed reported positive experiences in using audiobooks or text-to-speech software, other participants reported difficulties in using high-tech reading or writing supports, such as one participant who found the audiobook's speech rate too fast (Kjellén et al., 2017). In other studies, some participants reported a rediscovered pleasure in reading novels when using the Amazon Kindle with text-to-speech (Knollman-Porter et al., 2015). Others are pleased that they can listen to long novels that they would not otherwise be able to read (Kelly et al., 2016; Kjellén et al., 2017). Other studies reported how patients favour the use of screen-reader applications (e.g. the iPad's Speak Selection) or computer-generated text-to-speech technology to independently access text they are interested in when using the device (Kelly et al., 2016; Knollman-Porter et al., 2015; Knollman-Porter & Julian, 2019). For instance, a participant in the study carried out by Knollman-Porter et al. (2015) reported using the iPad screen reader to make purchases online. It has become very hard for this patient to navigate the Web or do any "reading" without a text-to-speech system. In 2014, Szabo and Dittelman investigated the use of embedded iPad features such as Reader (in Safari), Speak Selection and Siri, everyday mobile applications and websites (i.e., n2y, news-2-you; www.n2y.com) used by PWA with acquired reading difficulties. The study focused on training patients on mobile apps and accessibility features in iPads, and investigating the satisfaction level and the real usefulness experienced by participants. The authors also examined the use and satisfaction level in websites, they observed that participants particularly enjoyed their subscriptions to n2y; the website offers worksheets along with articles of different difficulty levels, and the facility to reproduce the content of the article out loud while the words are highlighted. A new article is uploaded each week. In the case of the n2y website, the only information available is the degree of enjoyment of the patients; the authors provide no information on comprehension improvements. The results from this study highlighted that the accessibility features embedded in the iPad were not easily accessible for people with aphasia (Szabo & Dittelman, 2014). For instance, the iPad provides Speaker Selection: an accessibility option that reads the text presented on the device's screen aloud. Although participants with reading difficulties reported a positive experience overall, they still encountered challenges in using this feature because it requires fine motor skills, which are often compromised in patients with aphasia (Szabo & Dittelman, 2014).

From these studies investigating the different experiences of PWA and reading impairments when using different kinds of devices, it emerged that alongside positive experiences, researchers also observed difficulties associated with the use of high-tech mainstream tools. This variability in usefulness is mostly because these tools were not designed with PWA (Szabo & Dittelman, 2014).

2.4 What tools are missing for people with aphasia and acquired reading impairments

The great diversity in clinical pictures of aphasia seems to call for software with a high level of flexibility and personalisation. For instance, some patients can present a low level of reading impairment, where the combined modality (written and auditory input) or the font adjustment might be sufficient (as in the case of ClaroReadTM or tablets' and e-readers' accessibility features). Other patients, meanwhile, can present severe reading difficulties where none of these supports would help. Moreover, PWA often have difficulties in processing long and complex sentences or low-frequency words via the auditory channel. Indeed, in these cases, it would not be sufficient to convert the text on the screen into speech, or to provide the option to follow the text in combined modality, if the text on the screen is long and articulated. Alongside strictly linguistic impairments, PWA often also experience cognitive impairments such as working memory or executive-function deficits (e.g. task monitoring and planning and organising actions), as well as motor impairments. These impairments hinder not only reading ability itself, but also the movements and actions required to activate and use a reading-support programme. Such kind of device, which can reach a high level of personalisation is still missing so that patients have to rely on a variety of devices not designed for their specific difficulties or strengths.

In the last few years, researchers conducted various studies aiming at identifying the aphasia-friendly features of the devices used from PWA and acquired reading impairments. All these characteristics together, with the possibility to calibrate them based on each patient necessity, could represent an initial step toward a successful compensative reading tool for these patients. The accessibility features common to these devices that are most helpful for PWA when they try to access written material are a) the presence of a text-to-speech system or a recorded voice, b) the option to see the text highlighted as it is read out loud by the text-to-speech system, c) the option to modify the text font and add blank space between lines and words to improve readability, and d) the possibility to calibrate the speech rate. Combined modality – namely, the combination of speech and text – is a feature that reportedly helps PWA with acquired reading difficulties (Brown et al., 2019; Wallace et al., 2019). It seems to decrease the demands on working memory, thus promoting an improved performance in text comprehension as patients can focus their cognitive resources on understanding the content (McNeil et al., 1991; Murray, 1999). The second recurring feature – that is, highlighting text as it is read aloud – is usually associated with text-to-speech systems, and is welcomed by patients with reading impairments as it allows them to follow the text easily. Despite the possibility that this feature can enhance comprehension, there is still limited evidence for this as yet (Brown et al., 2021). This feature is one of the overt aphasia-friendly principles, together with simple sentence structure and vocabulary (Brennan et al., 2005; Caute et al., 2016; Rose et al., 2003; Worrall et al., 2005). The last two aphasia-friendly features - text adaptation of adjusting font size and line spacing and speech rate personalisation - are integrated into most devices as an essential accessibility tool.

Alongside these well-investigated aphasia-friendly aspects, the same cannot be said for the possibility to convert the text into a simpler version, which represents one of the obstacles for PWA and acquired reading impairments (Caute et al., 2016; Knollman-Porter et al., 2015). Usually, a caregiver or the speech therapist executes this operation manually, giving many patients a sense of dependence and frustration (Dalemans et al., 2010; Knollman-Porter et al., 2015). Research on PWA with acquired reading difficulties has yet to investigate the possibility of patients selecting and simplifying a text passage autonomously. However, the capability to modify complex and lengthy texts into simpler, shorter versions has already been developed in the field of natural language processing (NLP). In this area, research on automatic text simplification for populations with special needs has been ongoing for two decades and continues to make remarkable improvements. Examples of NLP tools such as Dyswebxia (Rello & Baeza-Yates, 2014), Lexi (Bingel et al., 2018) or Simplext (Saggion et al., 2015) suggest that although the fields of NLP and language rehabilitation are separate, they can be successfully combined to provide helpful compensative methods. Indeed, new NLP technologies could potentially be adapted and integrated with aphasia rehabilitation and play a critical support role in overcoming issues that have hardly been covered as yet, such as compensative tools for PWA with acquired reading impairments. Technological advancements in NLP are evolving towards more customisable software that can conveniently be merged with research in aphasia. Indeed, the possibility of personalised text adaptations would represent a relevant improvement for PWA, as the difficulties experienced when reading or comprehending auditory speech can vary substantially across individuals. Therefore, flexible tools such as those already developed for other populations, such as dyslexic individuals, could also be highly desirable and beneficial for PWA with acquired reading impairments.

Chapter 3

TECHNICAL TOOLS: TEXT-TO-SPEECH SYSTEMS AND AUTOMATIC TEXT SIMPLIFICATION SYSTEMS

With the rapid development of computer science, artificial intelligence, and robot control technology, the field of human-computer interaction has witnessed exponential interest and development. One of the main aspects characterising human-computer interaction is speech communication, which comprises many subcategories such as voiceprint recognition, speech recognition and speech synthesis (Mu et al., 2021). For this research project, we focus on speech synthesis and text simplification. To synthesise an artificial voice as close as possible to the human voice – with also all the various styles that characterise human speech – the systems require an extensive amount of data to train and highly complex models involving diverse disciplines such as acoustics, linguistics and computer science (Mu et al., 2021). Another technology part of the human-computer interaction field is text simplification. The benefits of automatic text simplification (ATS) as a tool for reading assistance, which makes written material more accessible to a larger audience, have driven research in the last few decades (Saggion, 2017). We believe that a calibrated combination of these two technologies could help individuals with aphasia and acquired reading impairments to access texts they can no longer read.

This chapter is structured as follows: first, we overview each system in their structural parts and how they work; then, we describe the text-to-speech systems and the text simplification system we decided to use in the studies on persons with aphasia and acquired reading impairments (see Chapter 2, sections 2.1 and 2.2); lastly, we report the literature already present on the topics text-to-speech systems and aphasia as well as text simplification and aphasia.

3.1 Text-To-Speech systems

When speech synthesis has a text as input, it is also called *text-to-speech* (TTS). TTS is currently a frontier technology in the field of information processing, and its application extends to various aspects of people's daily life, from intelligent assistants or intelligent customer service to newer applications such as article reading, language education, video dubbing, and rehabilitation therapy (Mu et al., 2021). The speech synthesis process from text to voice has two phases: text analysis and digital signal processing. In the text analysis phase, the text is analysed in its linguistic units as sentences, phrases and words and then transcribed into a phonetic representation. The steps in this phase are: pre-processing, morphological analysis, contextual analysis, syntactic analysis, phonetisation and prosody generation (Yin, 2018) (see Figure 3.1.1 for the steps sequence and Table 3.1.1 for a summary of the characteristics of each step of the process).

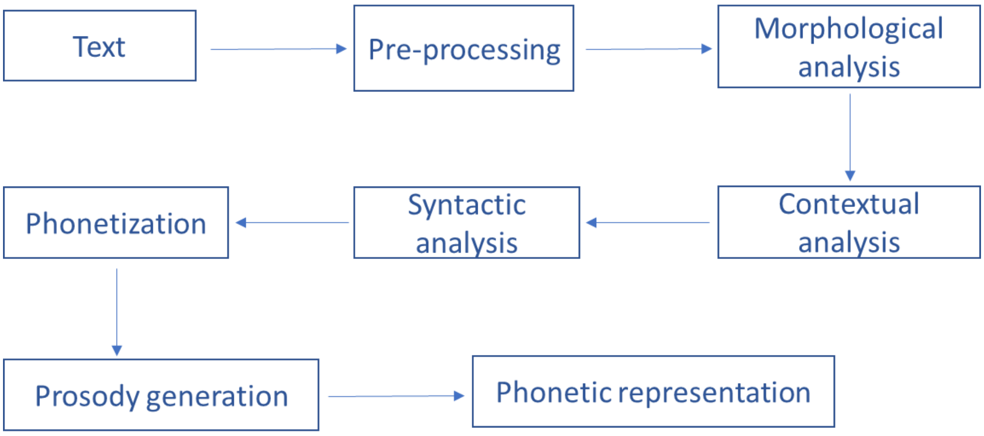


Figure 3.1.1: Modules of the text analysis phase *(features extracted from Yin, 2018).*

Table 3.1.1: Summary of the characteristics of each text analysis step.

<table>
<tr><th>Process phase</th><th>Principal characteristics</th></tr>
<tr><td>Pre-processing</td><td>The step that corrects the input text errors and irregularities (i.e., numbers, abbreviation and acronyms) into a full text and then it breaks the text into sentences (usually marked by punctuation).</td></tr>
<tr><td>Morphological analysis</td><td rowspan="2">Usually, these two parts of the analysis work together. The sentences are divided into words that are then categorised into parts of the speech (POS: the category assigned to a word in accordance with its syntactic functions) based on their spelling. This step is normally also called word segmentation or word tokenisation. While for some languages, like English, the boundaries between words are clear</td></tr>
<tr><td>Contextual analysis</td></tr>
</table>

	(i.e., white space), there are no clear separators between different words in other languages, such as Chinese, which is character-based. In these cases, contextual analysis is the only step available to assign words to their possible part of the speech. The contextual analysis could also accelerate the process to assign part of the speech to the words by considering the neighbouring words of each word.
Syntactic analysis	This step parses the syntactic structure using the information of word segmentation and word classes done in the previous step. Generally, this step starts the prediction of the input sentence prosody.
Phonetisation	In this step, after creating a pronunciation lexicon, the sequence of phonetic symbols for each word is created, as the phonetic symbols for each word are produced based on the language's lexicon.
Prosody generation	This last step generates all the prosodic features, such as duration, pause information, pitch curve, stress, and rhythm. Despite being the last step of the process, it has a crucial role in the naturalness level of the TTS and, to some extent, in its intelligibility. Prosody, for instance, could help the listener in understanding sentence chunks.

The final result of the text analysis is an input text converted into its corresponding phonetic representation.

The second part of the speech synthesis process, the conversion of the phonetic representation into speech sounds, has two types of speech synthesis methods: rule-driven and data-driven. The rule-driven method comprises the articulatory and formant synthesis; it synthesises speech based on acoustic or articulatory rules and is less performant than the data-driven. Rule-driven synthesis methods are old systems that are not used anymore and were never used for commercial purposes (Khan & Chitode, 2016).

The data-driven method is recognised as superior to the rule-based because it uses real speech data and usually performs better in the two indicators of voice quality: naturalness and intelligibility (Yin, 2018). Naturalness is defined as the correct use of emphasis, intonation, pitch, intensity, and pauses according to the message and intention (Sanders et al., 1981), while intelligibility of synthetic speech is considered as the listener's ability to recognise/identify phonemes and words when they are presented in isolation (Giannouli & Banou, 2020). Data-driven methods comprise i) concatenative synthesis, ii) statistical parametric synthesis, and iii) deep neural networks synthesis (DNN, which is based on a statistical approach). Figure 3.1.2 shows a simplified diagram of the different speech synthesis methods, while table 3.1.2 rates the naturalness and the intelligibility of the systems compared among them and summarises the main limitations of these methodologies.

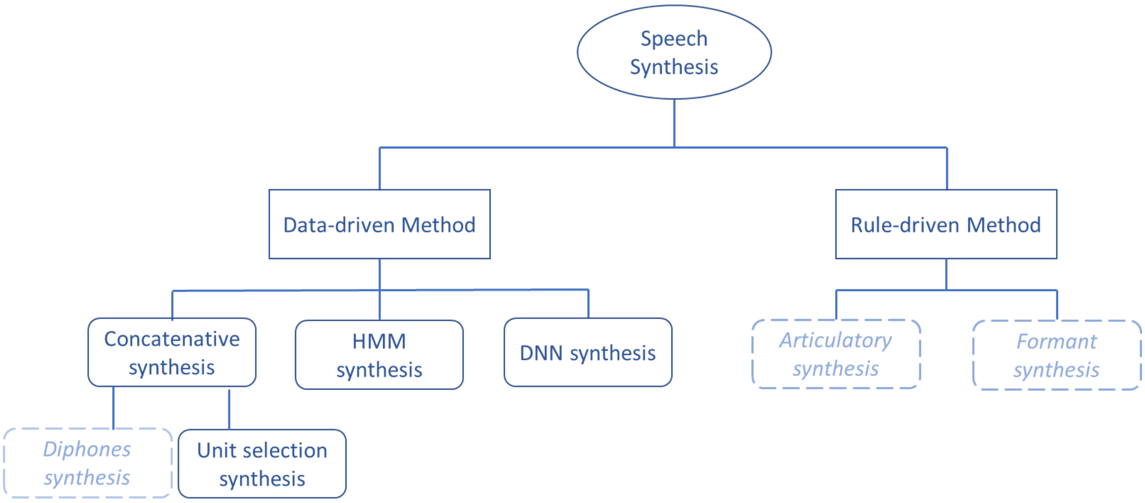


Figure 3.1.2: Modules of the different speech synthesis methods *(features extracted from Yin (2018).* Diphones, articulatory and formant synthesis are in lighted colour and dashed box as they are methodologies not used anymore. Yin (2018) did not report diphones synthesis, but we added this methodology out of completeness.

Table 3.1.2 Summary of the different data-driven speech synthesis methods' characteristics.

Method	Naturalness	Intelligibility	Limitations
Unit selection based concatenative system	Good	Poor	Presence of speech artefacts affecting intelligibility, and requires a large speech database

Statistical parametric synthesis (HMM)	Poor	Good	Poor naturalness of the synthesised speech
Deep neural network (DNN)	Very good (similar human voice)	Good	Requires larger computational resources

The concatenative speech synthesis is generated by selecting and combining small voice units (e.g., words, syllables and phones) previously recorded. The speech quality is affected by the length of the units: e.g., longer units (e.g., words or sentences) increase naturalness as the number of concatenation points decreases, but the size of the speech corpus to train and the memory needed to operate is considerable, and the flexibility of the TTS reduced. On the other hand, shorter units, such as phones, cost less in memory and speech corpus needed, but the sounds quality will be negatively affected as the concatenation points increase (Khan & Chitode, 2016; Thomas et al., 2006). Initially, concatenative synthesis was represented by diphones based synthesis, that later developed into unit selection synthesis and that nowadays is not used anymore. In the diphone-based approach, there is only a single instance of all speech units available in the speech inventory (Thomas et al., 2006). Diphones are two linked half-phones that begin in the middle of the first phones and finish in the middle of the second phones (Khan & Chitode, 2016). Since there is no more than one instance of the speech units, various signal processing methods were necessary to have the correct prosody (e.g., PSOLA) (Khan & Chitode, 2016). More recently, the concatenative speech synthesis is mainly represented by unit selection synthesis. In this method, multiple examples of speech units with different prosodic features are stored. This method usually selects the longer unit segment according to the text that needs to be synthesised. An optimal unit selection process considers the i) target cost, which evaluates how similar the prosodic attributes of a database speech unit are to those of the target speech unit (candidate unit), and ii) the concatenation cost, which measures how well two speech units join and match each other, hence concatenate (Khan & Chitode, 2016). These observations about the concatenative synthesis suggest that having a good speech corpus is crucial for sound quality. The main limitation of concatenative speech synthesis is poor intelligibility. The availability of speech units with acceptable prosody in the database, as well as the continuity of acoustic characteristics at the concatenation point, such as fundamental frequency, amplitude, and speaking rate, are the key elements impacting the intelligibility of synthesised speech (Khan & Chitode, 2016; Thomas et al., 2006).

Statistical parametric speech synthesis (SPSS) infers acoustic parameters from speech data. The hidden Markov model (HMM) is the most used statistical parametric synthesis technique. This approach has the advantage of i) allowing flexibility in changing voice characteristics, ii) requiring a smaller speech corpus compared to the concatenative method, and iii) the training time is shorter than theconcatenativ method. In this approach, frequency spectrum, excitation, and speech duration are simultaneously modelled by context-dependent statistical models (HMMs), and speech waveforms are generated from the HMMs themselves (Yin, 2018; Zen et al., 2009). The main limitation of this approach is the naturalness of the synthesised speech that could be degraded by vocoding, accuracy of acoustic models and over-smoothing. During the vocoding process, the speech signal is converted into a more compact representation to be transmitted; this process causes what is referred to as a "buzzy" quality of the synthetic speech (King, 2011). The acoustic models generate the speech parameters of the HMM voice, and therefore their accuracy affects the quality of the synthesised speech. For this reason, a more precise statistical model produces better voices. Lastly, over-smoothing refers to the process that removes some detailed characteristics of the speech parameters when averaging many frames of speech with different spectral properties; this deletion cannot be recovered in the synthesis part, and it gives a muffled quality to the output voice (King, 2011; Zen et al., 2009).

The more recent technology in speech synthesis is deep neural networks (DNNs). The fundamental goal of DNNs is to mimic how the brain works. To do so, DNN models acquire the more relevant items from speech by constructing a machine learning model with multiple hidden layers and big data training, hence reducing the issues associated with manually selecting features. In this case, the voice quality is similar to natural speech. The main disadvantage of this approach is that it requires more computation than other methods and the prediction of pitch is not as good as the HMM-based approach (Zen et al., 2013). The TTS methods that could produce speech similar to natural voice are HMM and the deep neural network-based, and they represent the best balance between the amount of data and memory and the quality of the voice (Yin, 2018).

The DNNs are a subfield of neural networks, a subset of artificial intelligence (Alom et al., 2019). DNNs present two main architectures: convolutional neural network (CNN) (Lecun et al., 1998) and recurrent neural network (RNN) (Elman, 1990). See figure 3.1.3 for a schematic representation of these fields and how they interconnect. This figure does not want to be exhaustive and detailed in every aspect of artificial intelligence technology for synthetic speech synthesis. It aims to give an idea of how the technologies described in this chapter are interconnected.

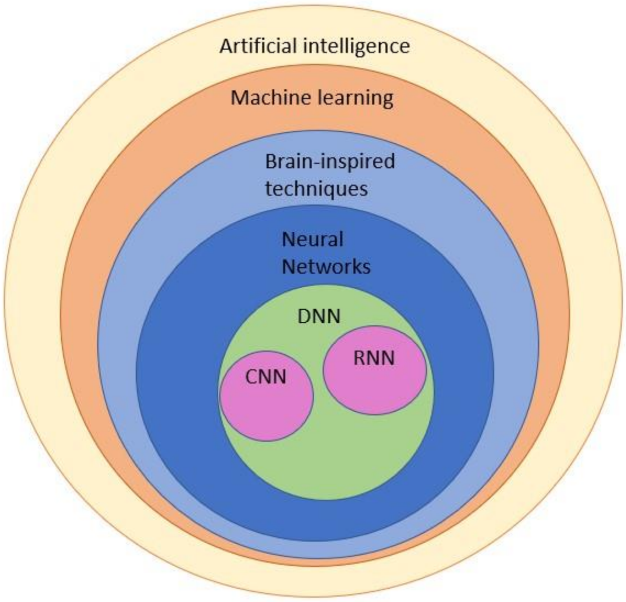


Figure 3.1.3 Diagram representing the relationship between the different technologies described in this chapter. DNN: deep neural networks, CNN: convolution neural networks, RNN: recurrent neural networks. *(features extracted from Alom et al. (2019))*

CNN architecture is hierarchical, while RNNs are sequential. CNNs exploit the fact that many natural signals are compositional hierarchies, in which higher-level elements are created by assembling lower-level ones. For instance, local combinations of edges produce motifs in images, motifs assemble into parts, and parts assemble into objects. In the same way, hierarchies exist in speech and text, from sounds to phones, phonemes, syllables, words, and sentences (LeCun et al., 2015). For this reason, CNNS have brought breakthroughs in processing images, video, speech and audio. RNNs process an input sequence one element at a time, storing a 'state vector' in their hidden units that contains information on the history of all the previous elements. In this way, RNNs use all of the available input information to predict the output features of each frame

(LeCun et al., 2015; Mu et al., 2021). For this reason, RNNs are frequently used on sequential inputs such as text and speech (LeCun et al., 2015), and researchers use them to capture the long-term dependence of speech frames to increase the quality of synthesised speech (Mu et al., 2021).

3.1.1 Text-to-speech systems in the experiments

In the two experiments included in this thesis, we used different synthetic voices to identify if there was one more suitable for PWA and acquired reading impairments among the TTS currently available for the Spanish language. In what follows, we report the characteristics of the variuos TTS systems that we considered in our experiments. We divided this section in two parts: in the first part we describe the three TTS used in the first experiment and in the second part the TTS used in the second experiment.

Experiment 1: Systems description

In the first experiment (Chapter4, subsection 4.2), we used three different TTS systems corresponding to the three major types of technology: i) a computer-generated voice by IBM, which at the time of the experimental set up was called Bluemix and now renamed Watson, ii) Ogmios synthetic speech from the Universitat Politècnica de Catalunya, and iii) FestivalTTS system from the University of Edinburgh. Bluemix was selected as an example of a proprietary high-quality close-to-human voice TTS system as, at the time of the experimental setup, to the extent of our knowledge, there was not an open-source deep-learning-based TTS for the Spanish language. Lastly, Ogmios and FestivalTTS were selected to represent open-source TTS systems built with two different speech synthesis technologies.

Ogmios represented the concatenative synthesis system, specifically the unit selection approach (Bonafonte et al., 2007). Ogmios is a multilingual speech synthesis designed primarily for Spanish and Catalan languages and subsequently also for English. Ogmios contains the following modules: i) the symbolic analysis, where the input text is classified in tokens as punctuation, acronyms, abbreviation, data expressions, numbers etc.; the text is labelled with part-of-speech (POS) tags using a statistical tagger and word pronunciation is taken from a dictionary and pronunciation of unknown words is predicted; ii) the prosody generator to produce a natural sounding voice and it entails different tasks of prosodic modelling: phrasing (breaking long sentences into smaller prosodic phrases where boundaries are defined by a pause, tonal change or length of the last syllable), duration, intensity and intonation; and iii) waveform generation synthesis that derives from the concatenation of recorded segments extrapolated from a large database (Bonafonte et al., 2007). The online demo of Ogmios[1] allows writing a text of 250 characters and selecting between three languages and fifteen voices. Each voice has different hours of speech data used to build the voice; for instance, Spanish voices have both male and female voices for up to ten hours. In the experiment described in the thesis, we used the Spanish male voice (Antonio).

FestivalTTS was chosen to represent the statistical parametric (HMM-based) synthesis. It is a multilingual system, and it synthesises for English (British and American), Italian, Czech and Spanish, although the most advanced voices are the English ones. Other languages are available as prototypes. FestivalTTS has three synthesis methods: the statistical parametric approach, the unit selection concatenative approach and the single instance diphone concatenation. The single instance diphone is the first TTS generation technology, and it was used from the mid-1980s to the mid-1990s. The online demo[2] presents a wide variety of voices, especially for the HMM-based method, and the majority of them are for the English language. In the experiment, we used UVIGO voice, the only Spanish voice available on the online demo.

IBM Bluemix TTS represents the deep learning-based synthesis (now Watson TTS[3]) and it is among one of the TTS using the most recent technology for building artificial voices. Specifically, Watson IBM TTS is built using recurrent neural networks (RNNs) trained on natural human speech to predict acoustic features. The RNN method used in Watson TTS has the advantage of allowing for quick and simple training and independent control of each component. The system first analyses the input text and establishes the content. Specifically, the model considers each phone in a sequence of phones in the context of the preceding and following two phones. From this evaluation, the model produces a set of acoustic units that are then evaluated based on the

[1] http://ttsdemo.talp.cat/, accessed 04/03/2022
[2] https://www.cstr.ed.ac.uk/projects/festival/morevoices.html, accessed 04/03/2022
[3] https://www.ibm.com/demos/live/tts-demo/self-service/home, accessed 04/03/2022

appropriateness level. This phase minimises the complexity by limiting it to only those units that meet certain contextual criteria and eliminating all others (Kons et al., 2019). IBM Watson uses three DNNs: prosody prediction, acoustic feature prediction and neural vocoder that predicts the pitch and the phoneme duration, the spectral structure and the waveform of the speech. Each of these three blocks has its model, and they are trained independently for each voice (Kons et al., 2019). Watson TTS had the advantage of creating good quality voices using a smaller set of audio data (5-20 minutes) and without an expensive GPU support that is usually required in neural network multi-speaker models, as it is usually a very slow and heavy synthesis process (Kons et al., 2019). Watson TTS presents thirteen different output languages; we used "Enrique", the Castillan Spanish male voice option. In the case of Spanish, Watson TTS allows choosing also between Latin American or North American Spanish.

Experiment 2: System description

In the second experiment (see Chapter 5), the TTS system used for the tasks was *DCTTS* (Deep Convolution text-to-speech). DCTTS is fully based on convolutional neural networks (CNNs), one of the latest state-of-the-art methods in the deep-learning approach (Tachibana et al., 2018). This approach takes as input a sequence of characters or phonemes and outputs a raw waveform or a spectrogram (Park & Mulc, 2019; Tachibana et al., 2018). One of the main novelties of DCTTS is the use of a public database – LJ Speech Dataset (https://keithito.com/LJ-Speech-Dataset/). Recent studies have shown that CNN-based synthesis is much faster and requires less sophisticated machines than recurrent neural networks (RNN), and the voice quality is still acceptable (Goodfellow et al., 2016). DCTTS allows speech synthesis for ten different languages; in the case of our experiment, we used the Spanish male voice.

3.1.2 Text-to-speech and aphasia

In recent years, work exploring the impact of TTS systems on people who acquired aphasia and reading impairments has increased. Even though the number of studies is still limited, investigations on compensatory strategies are gaining interest. Researchers have been exploring the subject from different points of view, depicting a progressively more complete picture of the role and impact of TTS systems on people with aphasia and acquired reading impairments. One of the main points that many researchers try to uncover is to investigate why people with aphasia and acquired reading impairments have limited use of text-to-speech systems, even though it could improve post-aphasia text comprehension (Hux et al., 2021; Knollman-Porter et al., 2019; Wallace et al., 2021). This phenomenon regarding PWA contrasts with the growing number of individuals – with and without reading challenges – that are more and more resorting to TTS technology (Wallace et al., 2021).

The researchers found that for PWA to adopt TTS systems in daily life, they need the opportunity to explore, manipulate, and comment on such technology (Wallace et al., 2021). Wallace and colleagues gave practice sessions illustrating different TTS systems' use and modifiable features. The results from this study confirmed that PWA do try to read despite their persistent language challenges and that the strategies and preferences in reading are highly individualised, as also reported in other studies (Kjellén et al., 2017; Knollman-Porter et al., 2015; Wallace et al., 2021; Webster et al., 2021). The participants endorsed the belief that TTS technology would enhance their comprehension and diversity of reading materials (e.g., from only short to longer texts), but it might not be necessary or useful for short material or materials with supplementary images or material not directly linked to technology (e.g. calendars). In general, if PWA are exposed and guided by therapists or caregivers to this technology and become acquainted with its functionalities, they are more likely to develop an interest and willingness to incorporate it into their daily lives (Hux et al., 2021).

Alongside the basic necessity of training, other aspects have to be considered when trying to understand the underuse of TTS in the aphasia population, such as personalisation. The need for personalisation expresses itself as calibration of i) the preferred speech output voice, ii) gender of the output voice, iii) speech output rate, iv) pausing, and v) highlighting. Characteristics such as voice and highlight selection and setting speech output rate to the preferred speed are important to people with aphasia and need to be considered when selecting a device for people with aphasia (Hux et al., 2021). One of the first quantitative studies investigating the voice characteristics aspect is the study of Hux et al. (2017). In this study, the researchers investigated the impact on comprehension rate of three different voices, two artificial (David from the Window platform and Alex from the Macintosh platform) and one human digitalised. The patients were also asked to perform a preference task. This study revealed that the comprehension accuracy was significantly better in the digitalised human voice than in the other two artificial voices. Similarly, the preference tasks showed that the digitalised human voice and David TTS were preferred over Alex TTS. Hux et al. (2017) study was the first to investigate TTS voices' impact on PWA with acquired reading impairments, and it set an important structure and direction for the subsequent studies in the field. However, TTS technology has significantly improved in the last few years. The differences between synthetic voices and digitalised human voices are so small that a TTS built with state-of-the-art technologies usually does not pose significant comprehension difficulties or marked preferences.

More recent research like the one conducted by Knollman-porter et al. (2021) tends towards identifying the optimal balance between the personalisation of the TTS features and the comprehension accuracy of a text. Knollman-Porter et al. (2021) investigated PWA's comprehension and processing time when reading supported by TTS systems with preferred voice settings, speech rate, highlighting type and highlighting colour versus unsupported reading. Secondarily, the researchers aimed at examining the i) initial support and feature preference selections, ii) changes in preferences following the training to TTS and iii) the anticipated reading activity when using TTS systems. Results showed that participants tend to select the voice matching their gender and that the preferred rating was 160 wpm, comparable to the results of other studies investigating speech rate in PWA (Hux et al., 2021). Regarding the highlight, the participants, in general, preferred sentence or word highlighting over no highlighting (also in this case, the result is comparable to the one conducted in a similar study by Brown et al. (2021); lastly, the colour selected to highlight the text varied across participant matching personal preferences. However, the results and trends observed did not show significant results in comprehension between TTS and unsupported conditions. On the other hand, there was a significant difference in processing time comparing TTS supported and unsupported tasks; namely, the participants showed shorter processing time in a TTS supported modality than in the unsupported condition. These results suggested that PWA are more efficient when using customised TTS than when they read without the TTS support.

The second part of the experiment investigated the changes in preferences following TTS exposure and the projected use of TTS in functional reading. The results in this second part of the experiment were diverse, showing how difficult it is to depict a univocal picture regarding perception and preferences. However, one relevant observation from this investigation is that the practitioners should provide PWA with multiple opportunities to explore the features and the different options of TTS systems.

Other evidence of the impact of speech rate on preferences and comprehension is the one conducted by Hux et al. (2021), who studied the impact on comprehension of three different text-to-speech rates in adults who acquired aphasia when accessing newspaper articles in combined modality. The researchers also collected data about the time spent reviewing the text after the synthetic speech ceased, the rate preference, preference consistency, and participant justification for their preferences. The results showed that although there was no significant difference between the three speech rates conditions, there was a trend for better comprehension with the slower presentation rate. Similarly, the reviewing time of the passages was significantly longer for the faster speech rate. The participants in this study indicated a preferred speech rate of 150 words per minute (wpm) – the medium presentation rate. Hux et al. (2021) suggested that the clinicians use this speech rate when introducing TTS technology to PWA. Adjustments in faster or slower could be made subsequently based on individual differences.

In light of these recent studies on TTS systems and PWA, findings suggest that the first step to consider is to engage with this population and guide them in the different settings and functionalities of TTS systems and show them how it could improve their reading or, more in general, the understanding of written content. This goal could be achieved through different calibration steps. As the studies highlighted, personalisation for each individual is the key to the successful use of TTS.

3.2 Automatic Text simplification

Automatic text simplification has the goal of transforming complex sentences and lexicon in a way that it allows better understanding and accessibility to a wider audience without changing its original meaning (Saggion, 2017; Štajner et al., 2019; Štajner & Saggion, 2018).

Several applications of automatic text simplification (ATS) have been proven useful in different scenarios, such as reading aids for children (Siddharthan, 2002; Watanabe et al., 2009), second language learning (Siddharthan, 2002), tools for people with neurodivergent profile – i.e., people with autism (Barbu et al., 2015; Orasan et al., 2013), dyslexia (Matausch & Peböck, 2010; Rello et al., 2013), or aphasia (Carroll et al., 1999). Alongside these applications for specific user populations, ATS could also be useful as a first processing step to improve results in many natural language processing (NLP) tasks such as parsing (Chandrasekar et al., 1996) information extraction (Evans, 2011; Jonnalagadda & Gonzalez, 2010), question generation (Bernhard et al., 2012), text summarisation (Siddharthan et al., 2004), and machine translation (Štajner & Popović, 2016). However, this last kind of application to other NLP tasks will not be tackled in this thesis as we are interested solely in ATS applied to reading tools for individuals with special needs, specifically in PWA and acquired reading impairments.

Automatic text simplification usually presents two main components: lexical and syntactic simplification. Lexical simplification refers to simplifying difficult and complex words into their more common and simpler equivalent. Even though it might appear like an easy task, the lexical simplification process requires extensive lexical and world knowledge. This task presents two main steps: finding a set of synonyms and replacing the complex word with a simpler one. A successful substitution replaces difficult words, still maintaining the core meaning and the grammaticality of the word in the sentence (Saggion, 2017). The parameters to define a word as difficult vary, but the main factors reduce to i) the word frequency (Carroll et al., 1998) and the word length (Bautista et al., 2011), or the combination of the two (Bott et al.,2012). Simplification systems usually first identify the complex words followed by the substitution generation, the substitution selection, and substitution ranking (Paetzold & Specia, 2016). For the English language, the state-of-the-art lexical simplification systems are data-driven. They range from supervised approaches using parallel corpora (e.g. Simple English Wikipedia in combination with the ordinary English Wikipedia) (Kauchak, 2013) to unsupervised techniques that rely on word embedding trained on larger corpora (Glavas & Stajner, 2015; Paetzold & Specia, 2016) until more recent techniques such as supervised lexical simplification systems with neural architectures (Nisioi et al., 2017).

In the case of the Spanish language, the alternatives available are also data-driven, either unsupervised (Bott et al., 2012), fully supervised (Štajner, Calixto, et al., 2015) or a hybrid of the two (Ferrés et al., 2018). The data-driven supervised lexical simplification usually presents the main problem of having small size parallel corpora, especially in languages other than English (Štajner et al., 2019). On the other hand, the main limitation of the unsupervised approach is that it performs word-to-word substitution so that lexical phrases or word reorderings in lexical phrases (e.g. "alternative solvents" -> "other solvents") do not give a successful result (Štajner, Saggion, et al., 2019).

Another technique that is becoming more common is the phrase-based statistical machine translation (PBSMT) approach, particularly in English (e.g., Štajner, Béchara, et al., 2015), Spanish (Štajner, Calixto, et al., 2015) and Brazilian Portuguese (Specia, 2010). A successful lexical simplification for Spanish was proposed by Štajner et al. (2019), who built a new simplification-specific dataset of synonyms and paraphrases that performed well in i) number of transformations, ii) grammaticality, iii) simplicity and iv) meaning preservation compared to the other lexical simplification systems for Spanish (Štajner et al., 2019). It outperformed both the more recent Spanish lexical simplification systems: LexSiS (Bott et al., 2012), which executes one-to-one substitution, and CASSA (Baeza-Yates et al., 2015), which only offers a list of unordered simpler substitution candidates.

Regarding syntactic simplification, ATS entails the transformation of syntactic structures that might impede comprehension in a simpler paraphrase, thus facilitating the readability and comprehension of texts (Saggion, 2017). An example of syntactic simplification is the following:

> Complex: "*Hu Jintao, who is the current Paramount Leader of the People's Republic of China, was visiting Spain.*"
> Simplified: *"Hu Jintao was visiting Spain. Hu Jintao is the current Paramount Leader of the People's Republic of China."*

Typical complex syntactic structures are subordinate sentences, coordinates, relative clauses and sentences that do not follow the canonical word order (e.g., subject-verb-object). The types of syntactic simplification vary from a simplification that maintains the whole sentence material (Chandrasekar et al., 1996; Ferrés et al., 2015; Siddharthan, 2011) to a simplification approach that entails the reduction of the sentence eliminating the superfluous details in order to keep only the essential information (Barlacchi & Tonelli, 2013; Glavaš & Štajner, 2013). The more common operations of syntactic simplification are: i) splitting, ii) reordering, iii) dropping, and iv) substituting (Sheang, 2020). Usually, these rules are handcrafted, making them time-consuming and hard to define since they require linguistic expertise and corpus analysis (Saggion, 2017). Generally, an expert linguist is needed to computationally model the complex phenomena in the text in grammatical terms and to be able to recognise them in new texts. However, when the rules are properly made, they guarantee a broad coverage of sentences and a precise system. Another issue with the handcrafted approach is that the simplification rules are language-dependent, meaning that to use the same system with another language is necessary to re-write new rules (Saggion, 2017; Sheang, 2020).

The newest approaches to ATS come from neural machine translation (Nisioi et al., 2017; Wang et al., 2016); syntax-based (Xu et al., 2016), phrase-based (Coster & Kauchak, 2011; Specia, 2010; Štajner, Calixto, et al., 2015; Wubben et al., 2012), and tree-based machine translation (Woodsend & Lapata, 2011; Zhu et al., 2010). Thanks to their good performance and the non-requirement of handcrafted rules, these approaches became fairly popular. Zhao et al. (2018) have recently introduced two new neural network-based sentence simplification approaches that provided results that outperformed all the previous state-of-the-art models in sentence simplification.

3.2.1 The Tuner simplifier

For the second experiment (see Chapter 5) of this research project, we used Tuner simplifier to transform lexically and syntactically complex sentences extracted from Spanish newspapers into easier-to-understand clauses for PWA and acquired reading impairments. Tuner project had the objectives of i) providing language resources and tools to process textual data in English, Spanish, Catalan, Basque and Galician; ii) acquiring and integrating knowledge through the integration of a broad-coverage knowledge base into the Multilingual Central Repository and adapted to specific domain; iii) developing inference and semantic reasoning engines using the knowledge integrated into the multilingual central repository; and iv) providing effective approaches for constructing robust and advanced semantic processing systems adaptable to different areas and languages (Agerri et al., 2018). Tuner presents both lexical and syntactic simplification processes for the four major Ibero-Romance Languages: Catalan, Galician, Portuguese, and Spanish, which share high lexical similarities (above 85 %) and have the advantages of having processes and lexical resources that can be easily adapted semi-automatically (Ferrés et al., 2018). Tuner's lexical simplifier was constructed utilising corpus-based methodologies mixed with a hybrid morphological generator that combines both a freely accessible lexicon and a decision tree-based algorithm and an easy to modify rule-based context re-writing module. The availability of a powerful multilingual generator is critical in the rich morphological languages such as the ones addressed by Tuner (Ferrés et al., 2018).

Alongside the lexical simplification, Tuner was built to tackle also the syntactic complexity such as passive sentences, appositive construction, relative clauses, coordinated constructions, correlated correlatives, subordinate clauses and adverbial clauses. However, at the time of our experiment, the passive clauses simplification was not available; therefore, it was impossible to simplify them for the Spanish language. In order to analyse what the results would be if this functionality was available, we simulated the automatic simplification by manually transforming the passive constructions into active ones.

In Tuner, the syntactic simplification is organised as a two-phase process: document analysis to identify which syntactic structure needs to be simplified and the sentence generation that produces the simplified clause (Ferrés et al., 2016). In the sentence generation phase, the information provided by the analysis phase is used to generate a simpler structure by applying a set of specific rules for each complex structure. These rules perform the common simplification steps: sentence splitting, reordering of words or phrases, word substitution, verbal tense adaptation, personal pronouns transformation, and capitalisation and de-capitalisation of words (Ferrés et al., 2016).

3.2.2 Text simplification and aphasia

Looking at text simplification for PWA, the research on TS dates back between 1998 and 2006, when Carroll et al. (1998), Devlin (1998), Canning et al. (2000), and Devlin and Unthank (2006) developed and published their pipeline of a full simplification system – Practical Simplification of English Texts (PSET) – for PWA, that was part of a UK project aiming at adapting texts for PWA (Saggion, 2017). The steps to arrive at the final system were various. For instance, Devlin (1998) investigated what type of text simplification is most suitable for PWA. Their proposition of sentences characteristics was in line with the characteristics clinically recognised: i) subject-verb-object (SVO) structure, ii) verbs in active voice, iii) short sentences, no coordinate structures, iv) one adjective per noun, v) the order of the sentences should mirror the chronological order of the events, vi) sentences semantically irreversible. The authors part of the UK project also examined the newspaper discourse and identified features that are challenging to process for PWA. For instance, long sentences, overuse of compound nouns and adjectives (e.g., mother-of-two) and the adverbials in the sentences (Saggion, 2017). In 1998, Carroll et al., observing these difficulties encountered by PWA when processing newspaper texts, designed an automatic simplification system for English newspapers available on the Internet. At that time, the system combined the state-of-the-art NLP tools.

Three main parts compose the PSET system: i) syntactic simplifier, ii) anaphora resolver and substitution component, and iii) lexical simplifier. The syntactic simplifier used in PSET was SYSTAR - SYntactic Simplification of Text for Aphasic Readers - (Briscoe & Carroll, 1995). SYSTAR presents three modules: anaphora resolution, syntactic simplification and anaphora replacement (Canning & Tait, 1999). For the lexical simplification, a more common synonym for each difficult word was created using WordNet lexical database (Miller et al., 1990). The synonyms were analysed using the Oxford Psycholinguistic Database (Kučera-Francis frequency list) (Quinlan, 1992)) to find the most frequent synonym and replace the original word.
Devlin & Unthank (2006) continued and described this work when they published a paper describing the HAPPI (Helping Aphasic People Process Information) project. This project aimed to develop a web-based system to help individuals with aphasia access online information such as news. The main feature is the possibility to select

easier and more common synonyms of words that the PWA find difficult (Devlin and Unthank, 2006). Another aspect considered by these researchers was the possibility of pairing difficult words with images of the concept or allowing the word to be spoken. Both these additional measures were thought to adapt to all the different levels of aphasia. Despite this extensive study about an online text simplification web-based system, we were unable to find any further information about the testing of such tool on PWA participants, nor any information on whether this system is in use by this target population. In fact, to the best of our knowledge, this system has not yet been applied to an interface destined for PWA with acquired reading impairments.

After describing the characteristics of aphasia and acquired reading impairments, and the main characteristics of the technical tools used in the thesis project, we describe the first experiment of the thesis in the next chapter, which is aimed at exploring which TTS system is particularly suitable for PWA and acquired reading impairments.

Chapter 4

SPEECH SYNTHESIS FOR PEOPLE WITH APHASIA AND ACQUIRED READING IMPAIRMENTS

Speech synthesis technology is present in devices on the market. In the last decades, improvements in the quality of artificial voices are such that nowadays, the text-to-speech (TTS) systems associated with mainstream devices (e.g., smartphones and tablets) are increasingly similar to human speech. While the TTS systems are well understandable to the healthy population, and possible lower voice quality does not represent an obstacle to the listener, investigations on populations with special needs, such as people with aphasia (PWA), are limited and not up to date. The auditory comprehension in PWA is frequently impaired, and when considering a TTS to use for this population, it is crucial to consider the quality of the voices and characteristics that might amplify auditory difficulties for PWA. This chapter investigates the impact of different TTS on auditory comprehension in PWA measuring comprehension accuracy and reaction times. Studies showed that reaction times are indicators of cognitive load: the longer the reaction time, the longer the time required to process and understand the auditory message. Indeed, a few psycholinguistic studies (Friedmann et al., 2008; McAllister et al., 2009; Shapiro et al., 1989; Shapiro & Levine, 1990) show that in PWA, longer reaction times are associated with higher cognitive cost. By collecting reaction times, it would be possible to investigate if there is a voice amongst the others that hampers auditory comprehension compared to the other voices. Compared with the state-of-the-art research, what is missing is a study investigating the single auditory modality using more recent TTS systems. Not all PWA have residual reading abilities; therefore, they might be able to use only audio input. A deep understanding of what technological architecture and what speech characteristics are best suited for PWA is crucial in these cases.

The main contributions of this chapter are:

- The comparison of the impact on text comprehension rate and reaction times of three different TTS technologies,

- The investigation of the preferences of the participants in relation to the type of TTS,

- The analysis of the results in reaction times is further explored through intelligibility, naturalness, and voice quality parameters.

The chapter is structured as follows: section 4.1 illustrates the pilot study on the target population and control group population and the considerations about what we learned from running it; section 4.2 describes the formal experiment designed to identify the TTS suitable for the PWA; section 4.3 describes the analysis of the parameters conducted on the human and artificial voices used in the experiment on patients; in 4.4 we tackled the main limitations of the studies, and lastly, in section 4.5, we discuss the conclusions concerning the experiments depicted in this chapter.

4.1 Pilot study

The pilot study was conducted on a target population of PWA and a control group. We run the pilot study to identify possible weaknesses and points of improvement for the formal experiment. The results of the pilot are not reliable due to the limitations of the experimental design and because we did not use all the sentences and sets of images that we selected for the experiment. So the pilot results are on just sample sentences. For these reasons, we do not report the results – hence the analysis – in the thesis, but we will discuss the lessons we learned from the pilot study and how we improved the final experiments based on them. During the first meeting with the patients, we illustrated our study and the motivation of the experiment. We conducted a short questionnaire on six people who decided to volunteer for the pilot in this session. In the questionnaire, we asked what they thought was their difficulty. Based on the answer, we asked if they were willing to either be helped in speech production or read by a machine (see appendix A for a complete example of the questionnaire). The question about the speech production and the machine support was because we were still refining the research question of the experiments, and we wanted to be sure that text comprehension was, indeed, felt like a problem by the volunteers, and they were open to the possibility to use a technological tool.

Four participants declared that they would be interested in such a tool. The remaining two participants either did not have a specific opinion on the matter or declared that what they wanted was to regain the ability to read so that they would rather put more effort into rehabilitation.

4.1.1 Methodology

Participants

The participants in the pilot study were twenty-two patients recruited at the "*Hospital de la Santa Creu i Sant Pau"* in Barcelona, and ten healthy participants for the control group were recruited through the university department. The PWA participating in the pilot volunteered for the study. They were not previously selected for aphasia severity, auditory comprehension, or cognitive difficulties, or having specific difficulties in reading; therefore, the composition of the target group was heterogeneous in aphasia severity, linguistic and cognitive difficulties. PWA age ranged from 45 to 75, while the control group age was between 32 and 72.

Experimental design

The pilot was structured in three parts: a) calibration phase, b) auditory comprehension task, and c) a MOS test. The auditory task was designed as a picture-point test. The reason for this choice was that PWA has difficulties executing other kinds of tests, such as paper and pencil tests or tasks without images. Even though this is not always the case, the participants in the pilot were heterogeneous in severity, so a picture-point task was appropriate. In this task, we measured the reaction times and the comprehension accuracy rate (dependent variables) in relation to the three different artificial voices and the digitalised human voice (independent variables). The pilot had a repeated measurement design.

The MOS test instead was designed as a Likert scale from one to five. The MOS test was used to investigate how the participants perceived the clarity, naturalness and ease of understanding of the voices used in the test. We decided to include this task in the pilot to gather information on the likability of the voices as it represents a crucial factor when considering supports for a specific population.

Materials

Sentences

The formulation of the sentences to use in the pilot was conducted using the following principles: a) short sentences of max nine words, b) high-frequency vocabulary, c) the use of only active verbs, d) no coordinates or subordinates. For instance, these sentences were used in the pilot: "*El niño tiene un perro en sus brazos*" [*"The boy holds a dog in his arms"*] or "*El padre toma una foto de los niños*" [*"The father takes a picture of the* *kids"*]. The sentences were formulated based on the images we found in the Norman Rockwell drawing (www.Saturdayeveningpost.com).; specifically, I made a set of 4 images for the comprehension task, and subsequently, I formed the sentences for the pilot.

Images

The images used are 224 images part of the Norman Rockwell[4] drawing. The images of Norman Rockwell are high-context images that help in formulating sentences with different agents and contexts.

Voices

See Chapter 3 subchapter 3.1 for detailed information about text-to-speech systems in general and section 3.1.1 for information about the systems used in this experiment.

Procedure

The session took place one-on-one with each participant in a separate room of the day hospital ward. The experiment was conducted in the Spanish language. Each session lasted approximately 20 minutes and involved a calibration phase; then, the experiment itself was divided into two parts: a) an auditory comprehension task and b) a MOS test.

Calibration phase

In the calibration phase, the participants performed a short mock test to ensure that the instruction and the goals of the experiment were understood and that the pc screen and the volume of the audio were comfortable for the execution of the test.

Auditory comprehension task

The task was structured as a picture point task to investigate the effect on comprehension of the four different voices selected for the experiment: three computer-generated voices and one digitalised human voice. We recorded both accuracy and reaction times for each item. The auditory comprehension task consisted of 56 sentences. The presentation order of the sentences was randomised for each patient. During the experiment, the participants looked at a four-image set on the computer screen in a 2 x 2 grid on the laptop monitor. Every time a set of pictures appeared, an audio stimulus sentence was played describing only one of the images. Thus, the four

[4] www.Saturdayeveningpost.com

voices considered in the experiment spoke 14 sentences each (always the same 14 sentences for every patient). For each set of images, there was one correct image and three foils. The experiment was self-paced; once the participant decided which of the images matched the stimulus sentence, they had to point at the image on the screen, and the researcher promptly pressed the corresponding image number on the laptop keyboard.

MOS Test

We built a MOS test to investigate the voice quality perceived by the participants. We asked to rate from 1 to 5 the following voice characteristics: ease of understanding, clarity and naturalness. Before each parameter, the patients listened to a sample sentence that was the same for each participant. After listening to the sentence, the researcher stated the main question that, in the meantime, was present on the pc screen (e.g., "Lo natural que sonaba esta voz?" ["*How natural sounds this voice?*"]). The questions – with their five levels boxes – were asked one at a time. We described each of the three parameters in an easy way to make it more understandable to the participant. For instance, the ease of understanding was rephrased "how easy is it for you to understand the voice?"; for clarity, we asked, "how clean do you perceive the voice? Do you hear any rasp?" lastly, for the naturalness, we asked, "do you hear this voice as human or similar to a robot and fake?". After every question, the patients had to point at one of the five boxes, indicating the level at which they perceived the clarity, the ease of understanding and the naturalness. The rating went from the lower level on the left side box (e.g., “muy poco clara” [“*not clear at all”*]) to the highest level on the right side box (e.g. “muy clara” [*“really clear”*]). After every question and explanation of the question, the researcher also pointed at each box, reading aloud the five possible responses. In figure 4.1.1, there is an example of the MOS test questions that we used during the pilot.

¿Cuán clara es esta voz?

1	2	3	4	5
Muy poco clara	Poco clara	no es clara o poco clara	Clara	Muy clara

¿Qué tan fácil es entender esta voz?

1	2	3	4	5
Muy difícil	Difícil	no es fácil o difícil	Fácil	Muy Fácil

Figure 4.1.1: Example of the MOS test proposed to the participants in the pilot study. Each question appeared alone on the screen. Here are together in one figure for simplification purposes.
"Lo natural que sonaba esta voz?" [*How natural sounds this voice?]*; "Cuán clara es esta voz?" [*How clear is this voice?*]; "Qué tan fácil es entender esta voz?" [*How simple is to understand this voice?*].

4.1.2 Lessons learnt from the pilot

By conducting the pilot experiment, we could identify the issues and the difficulties encountered by the patients and the weaker parts of the experimental design. The following issues have been identified: a) the speech rate is too fast for some PWA, b) the sentences are not rationally divided into specific grammar structures, and c) the MOS test was too difficult to understand for PWA.

Speech rate

At the end of the comprehension task, the patients were asked if they found the speech rate too fast. Half of the participants found the speech rate was too fast. In fact, during the test, the researchers observed some difficulties (i.e., many participants asked to repeat the sentence). For this reason, we decided to modify the speech rate. This modification was indeed necessary. In the formal experiment, we introduced the possibility to execute the experiment 25 and 50% slower than the standard speech rate, and all the PWA selected one of the two reduced speeds.

Sentences

We reduced the length of the experiment from 56 sentences to 48. We observed that the task was too tiring for some PWA, especially those with a more severe condition and in combination with the second task (i.e., MOS test). Moreover, we realised that the organisation of the comprehension task lacked structure. We decided to divide the 48 sentences into four different grammatical categories in order to control the results for grammar structure.

MOS Test

The execution of the MOS test was problematic for most participants. It was difficult to explain concepts like the ease of understanding, clarity and naturalness to PWA. We realised that these concepts were too complicated for PWA to understand, and even though they tried to complete the task, the answer was hardly reliable for most of them. Therefore, although we should consider the results despite what is expected (e.g., the human voice being more natural than one of the three voices) as this test is subjective, we decided to redesign it as a preference task.

4.2 Which text-to-speech system facilitates comprehension

This subsection will illustrate the experiment conducted on PWA and acquired reading impairments to investigate their comprehension rate and reaction times in relation to the type of synthetic voice. The study was structured to measure the impact of the synthetic voice on the two dependent variables, leaving out other factors that could cause comprehension and cognitive load difficulties such as speech rate, grammatical structure and vocabulary.

4.2.1 Methodology

Participants

Participants with aphasia

The study participants are seven adults (three females and four males) with chronic aphasia with reading impairments secondary to a cerebrovascular event in the language-dominant hemisphere (see details in Table 4.2.1). They were recruited at the "*Hospital de la Santa Creu i Sant Pau*" in Barcelona. The participants were native Spanish speakers or bilingual Spanish and Catalan. Participants' age ranged from 45 to 72. The inclusion criteria for the experiment were:

(a) chronic aphasia caused by a vascular event,

(b) patients reporting either difficulties or the inability to read in their everyday life,

(c) aphasia severity ranging from moderate to mild (WAB AQ 47.5-85)

(d) being a Spanish native speaker,

(e) no premorbid dyslexia,

(f) no hearing impairment,

(g) sight corrected to normal, guaranteeing the perception of images on the computer screen.

Table 4.2.1: Participants' details. (Age and years for patient eight are missing.)

ID	Age	Gender	Years post-stroke
1	47	male	6
2	57	female	11
3	63	female	20
4	51	female	5
5	72	male	10
6	58	male	12
7	45	male	9
8		female	
9	64	male	10

All participating PWA were assessed by the neuropsychologist and her assistants. They used the following tests (see Table 4.2.2 to check the assessment of each participant):

- Western Aphasia Battery (WAB) (Kertesz, 1982) to determine the severity of the aphasia.
- Corsi block-tapping test (onwards and backwords) (Kessels et al., 2000) examines patients' working memory. A score between 5 and 7 on the onward Corsi block-tapping test means no working memory impairments.
- The Boston Diagnostic Aphasia Examination (BDAE) reading subtest (Goodglass et al., 2001).
- All participants signed the informed consent (see appendix B for the check template we used).

The study was approved by the Institutional Committee for Ethical Review of Projects (CIREP) at Universitat Pompeu Fabra, approval Nr 0118.

Table 4.2.2: Assessment results for each participant. AQ: Aphasia Quotient, WR: word reading, SR: sentences reading, SC: sentences comprehension, PR: paragraph reading.

ID	WAB		Corsi block-tapping test		Boston Diagnostic Aphasia Examination - reading subtest			
	AQ (max 100)	Aphasia type	Onward (max.16)	Backward (max.14)	WR (max.30)	SR (max.10)	SC (max.5)	PR (max.10)
1	82.5	anomic	6	6	30	10	4	10
2	68.5	anomic	6	4	0	0	0	0
3	47.6	wernicke	5	2	15	2	0	0
4	85	anomic	4	6	30	10	2	7
5	45.8	broca	8	2	0	0	0	0
6	74.8	anomic	5	7	26	10	5	7
7	61.3	broca	5	7	0	0	0	0

Based on the inclusion and exclusion criteria, the initial 9 participants who volunteered in the experiment became 7. The last two participants enlisted in Table 4.2.1 were excluded because both a) had a WAB AQ above 85 and b) did not report any difficulties in reading.

Among the seven patients considered in the experiment, one patient did not present any measurable reading difficulty (patient 1); however, in the preliminary short interview, the participant reported difficulties reading material that he used to read before the stroke. Specifically, he complained about not being able to read the sports newspaper. The lack of a corresponding result of these complaints in the BDAE might be because the reading subtests of the BDAE are not sensible enough to capture mild difficulties in an everyday setting. Still, the goal of the study is oriented toward PWA who report complaints in the everyday reading activity, reason why we decided to include patient 1 in the study.

Control group

The control group consisted of 15 persons. The participants were recruited from the university department, and they volunteered in participating in the experiment. Also, in this case, the participants were anonymised. The only personal information asked was the age, which ranged from 30 to 58 years old.

Materials

Sentences

The 48 sentences of the experiment were divided into four different types of sentences, 12 for each of the following grammatical structures: subject + transitive verb + object (e.g. "*Los niños miran los peces"* [the boys look at the fishes]); subject + transitive verb + object + adjunct (e.g. *"El hombre tiende ropa en el patio"* [the man hangs the laundry in the courtyard]; subject + intransitive verb + adjunct (e.g. *"La pareja camina en el bosque"* [the couple walk in the woods]); complex subject + be (e.g. *"La madre y el niño están bajo el paraguas"* [the mother and the boy are under the umbrella]). The syntactic structures are simple structures that usually do not cause auditory comprehension impairments in PWA with residual comprehension skills. The grammatical structures present in the study are the ones considered to be easily understandable for PWA (DeDe, 2013).

Among the easy grammatical structures for PWA, there are some that are easier than others. For instance, complex subject + be verb is a structure that is less easy than the other three in the experiment. The challenge lies in the plurality of the subject. For instance, there are items where the images present only one of the subjects in the environment and act as indicated by the verb (see Figure 4.2.1 as an example). In this case, the patients have to pay attention to the plurality of the subject in addition to the action and the environment in which it takes place. The simpler structure, on the other hand, is the transitive one. Since the type of transitive sentence is not reversible – except for one sentence – these sentences do not represent an obstacle for the patients. Lastly, the transitive + adjunct and the intransitive + adjunct are structures that do not pose considerable difficulties as the transitive verbs are, for the majority, non-reversible, and the intransitive verbs do not allow ambiguity in the processing between subject and object. In Table 4.2.3, there is the list of the sentences we used in the experiment divided by grammar structure. The choice was motivated by wanting to focus the experiment as much as possible on the comprehensibility of the voices and not on the syntactic structure or the vocabulary. For this reason, there was preferred the use of short sentences with easy syntactic structures and high-frequency vocabulary (DeDe, 2012, 2013; Levy et al., 2012). The sentences' speech rate was manipulated using the 2.4.2 version of Audacity® recording and editing software [5].

Figure 4.2.1: Set of images extracted from the experiment as an example of a sentence with a complex subject followed by the verb "be". In this case the target sentence is : "*La madre y el niño están bajo el paraguas*" [The mum and the child are under the umbrella]. The target image is the number four. The other images in the set either have a different subject or a different location.

Table 4.2.3: Sentences used in the experiment divided by grammar structure

	Transitive	**Transitive + adjunct**	**Intransitive + adjunct**	**Complex subject + be verb**
1	El niño toca el violín	La madre busca algo debajo del sofá	Los niños juegan con la cometa en el prado	Dos niños están en el bosque
—	*The boy plays the violin*	*The man looks for something under the couch*	*Children play with kite in the meadow*	*Two children are in the forest*
2	El señor mira las	El señor toma el	El chico duerme	Los escolares y la

[5] Audacity team, 2020. https://www.audacityteam.org/

	postales	sole en el jardín.	en su habitación	maestra están en el museo
	The man looks at the postcards	*The man sunbathes in the garden*	*The boy sleeps in his room*	*The schoolchildren and the teacher are in the museum*
3	El niño cruza la calle	La chica prueba el vestido en su habitación.	El señor descansa debajo de un arbol	El padre y los niños están en el baño
	The boy crosses the street	*The girl tries the dress in her room*	*The man rests under a tree*	*The father and the children are in the bathroom*
4	El señor corta el seto	El hombre en el techo repara la antena	La mujer descansa sentada en el sótano	Los señores están acampando
	The man cut the hedge	*The man on the roof repairs the antenna*	*The woman rests sitting in the basement*	*The men are camping*
5	La niña abraza a su madre	Los niños bajan las escalera deprisa	La mujer corre bajo la lluvia	La madre y el niño están bajo el paraguas
	The girl hugs the mother	*The children come down the stairs quickly*	*The woman runs in the rain*	*The mother and child are under the umbrella*
6	El jugador firma el autógrafo.	El padre tiene al niño en sus brazos.	La mujer habla por teléfono	Los coches están parados en el tráfico
	The player signs the autograph	*The father holds the child in his arms*	*The woman speaks on the phone*	*Cars are stopped in traffic*
7	Los niños secan los platos	El niño ve un caballo fuera de la ventana	La pareja camina en el bosque	Los pescadores están en el barco de pesca
	The children dry the dishes	*The boy sees a horse outside the window*	*The couple walks in the woods*	*The fishermen are on the fishing boat*
8	Los niños miran los peces	El niño tiene un perro en sus brazos	Las chicas juegan a las cartas en la playa	Dos niños están en la pastelería
	The children look at the fishes	*The boy has a dog in his arms*	*The girls play cards on the beach*	*Two children are in the pastry shop*

9	La mujer persigue el autobús escolar	El hombre tiende ropa en el patio	La mujer en la tienda ríe	Un señor y una mujer están en una cafetería
___	*The woman chases the school bus*	*The man tends clothes in the yard*	*The woman in the store laughs*	*A man and a woman are in a cafeteria*
10	La anciana lee una carta	Un hombre lleva madera en la nieve	El bebé sentado en el suelo esta llorando	Una niña y su madre están en la playa
___	*The old lady reads a letter*	*A man carries wood in the snow*	*The baby sitting on the floor is crying*	*A girl and her mother are on the beach*
11	El zapatero arregla un zapato.	El hombre recoge el correo en su buzón	La señora y la chica conversan en el salón	Las niñas están en la escuela de danza
___	*The shoemaker fixes a shoe*	*Man collects mail in his mailbox*	*The lady and the girl talk in the living room*	*The girls are in the dance school*
12	Dos hombres miran la televisión	Los niños hacen ruido en el autobús	Los perros corren en la nieve	Los padres y el niño están en el médico
___	*Two men watch the television*	*Children make noise on the bus*	*Dogs run in the snow*	*The parents and the child are at the doctor*

Images

Like in the pilot study, the images used in the experiments were taken from Norman Rockwell's drawing. In this experiment, we used 192 images. The licence to use the images was granted from the Saturday evening post.

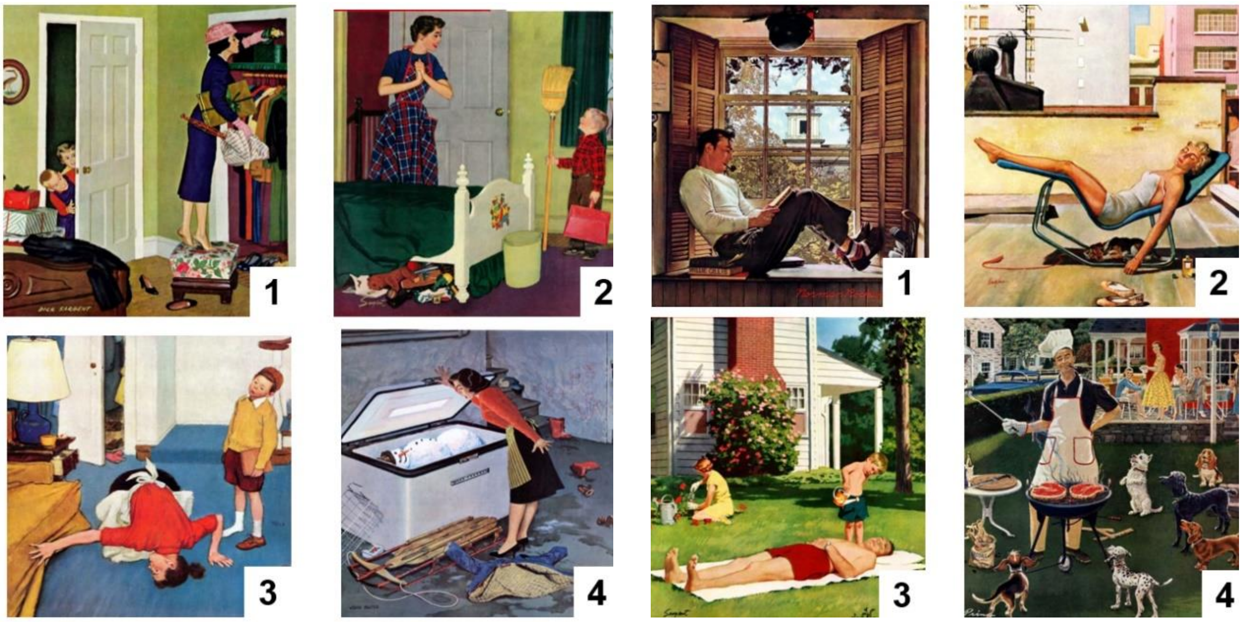


Figure 4.2.2: Examples of two items belonging to the four-set images set up. The target sentence for the set on right is "*El señor toma el sole en el jardín*" and the target sentence for the set on the left is "*La madre busca algo debajo del sofá*"

Visual support tools

The auditory comprehension task was carried out using a laptop where the DMDX software (Forster and Forster, 2003) was run. DMDX is a Windows-based program designed primarily for language processing experiments. It is designed to precisely record the time of presentation of text, audio, graphical and video material and record reaction times (RTs) (Forster & Forster, 2003).

Voices

See Chapter 3 subchapter 3.1 for detailed information about text-to-speech systems in general and section 3.1.1 for information about the systems used in this experiment.

Procedure

The session took place one-on-one with each participant in a separate room of the day hospital ward. The experiment was conducted in the Spanish language. Each session lasted approximately 20 minutes and involved a calibration phase, and then the experiment itself divided into two parts: a) an auditory comprehension task and b) a preference task.

Calibration phase

In the calibration phase, the participants needed to select the audio volume and the preferred speech rate at which to perform the experiment. Subsequently, they performed a short mock test to ensure that the instruction and the goals of the experiment were understood. The normal speaking rate in the Spanish language has an average of 210.86 words per minute (SD 32.58) (Trauzettel-Klosinski; Dietz; the IReST Study Group, 2012). From this reference, the patients could select a speech rate either a) 25% slower (168.69 words/minute (SD 26.07)) or b) 50% slower (140.57 words/minute (SD 21.72)). In order to complete this part, we used three recordings of the same sentence modified in the three different rates. We started from the standard speech rate and progressively slowed it down. After hearing all three rates, the patients selected the one they felt more comfortable with.

Auditory comprehension task

The first test was structured as a picture point task to investigate the effect on comprehension accuracy and reaction times of four different voices: three computer-generated voices and one digitalised human voice. We recorded both accuracy and reaction times for each item. The auditory comprehension task consisted of 48 sentences. The presentation order of the sentences was randomised for each patient. During the experiment, the participants looked at a four-image set on the computer screen in a 2 x 2 grid on the laptop monitor. Every time a set of pictures appeared, an audio stimulus sentence was played describing only one of the images. The four voices considered in the experiment spoke 12 sentences each (always the same 12 sentences for every patient). For each set of images, there was one correct image and three foils. The experiment was self-paced; once the participant decided which of the images matched the stimulus sentence, they had to point at the image on the screen, and the researcher promptly pressed the corresponding image number on the laptop keyboard.

Preference task

The preference task compared each voice to the other three in dyads. The participants had to indicate which of the two voices they preferred. If possible, they tried to explain the reason for their choice. In this task, they had to listen to six couples of voices speaking the same sentence. The experiment was carried out on PowerPoint. The participants looked at the computer screen where they could see four different boxes with four different answers: "voice 1", "voice 2", "both the voices equally", and "none of them". Every time one of the six couples of voices was played, the researcher read aloud the four answers while indicating the corresponding box displayed on the laptop screen to reiterate the possible answer and their location on the screen. The possibility to answer "both the voices equally" and "none of them" was because part of the focus of the experiment was to select the voice that the patients liked better, and forcing an answer between two voices could partially hinder the final goal of the task. There were no images to support this experiment. The control group also performed the preference task, and in this case, the task was shared via email through Google Form. The participants, opening the email, were invited to click on the link and execute the short preference task online. In the control group's case, the participants had to click on the links playing the sentences and then answer between the same four options through a multiple-choice grid. In both cases, patients and the control group, the items were randomised.

Data analysis

The data collected from the experiments were analysed using cross-tabulation and chi-square analysis and a generalised estimating equation. The cross-tabulation was used to analyse the comprehension accuracy and the preference task, while the generalised estimating equation for the reaction times because the data were not normally distributed, and they represented repeated measurements.

All the statistical analysis was performed using IBM SPSS Statistics for Windows version 23.

4.2.2 Results

This section will illustrate the analysis of the auditory comprehension experiment and the preference task. The analysis of the results was carried out using a cross-tabulation and chi-square analysis in the comprehension accuracy test and the preference task, while the generalised estimating equations (GEE) were used to analyse the reaction times. The GEE was used as the repeated measurements of the reaction times are not normally distributed.

Comprehension accuracy

In table 4.2.4, we can observe the number of correct answers and differences in means of correctness across conditions only in PWA, while in table 4.2.5 there is the multiple comparison analysis of significance of the correctness accuracy between conditions. The difference in means of accuracy between the conditions was not significant: X^2 (3, N = 336) = 1.36, p = 0.71. The comprehension accuracy test was carried out also in the control population using a Google Form – that does not collect reaction times. The accuracy rate had a ceiling effect for this group.

Table 4.2.4: Number of correct answers and differences in means of correctness for the four conditions.

	DHV	Bluemix	FestivalTTS	Ogmios
Number of correct answers out of 84	75	75	71	72
Mean of correctness (SD)	.89 (.311)	.89 (.311)	.85 (.364)	.86 (.352)

Table 4.2.5: Multiple comparisons for the analysis of significance of the correctness accuracy between conditions.

Condition (a)	condition (b)	Mean difference *(a-b)*	*p*	95% CI
DHV	Bluemix	.000	1.000	(-.10, .10)
	FestivalTTS	.048	.358	(-.05, .15)
	Ogmios	.036	.491	(-.07, .14)
Bluemix	DHV	.000	1.000	(-.10, .10)
	FestivalTTS	.048	.358	(-.05, .15)
	Ogmios	.036	.491	(-.07, .14)
FestivalTTS	Bluemix	-.048	.358	(-.15, .05)
	DHV	-.048	.358	(-.15, .05)
	Ogmios	-.012	.818	(-.11, -.09)
Ogmios	Bluemix	-.036	.491	(-.14, .07)
	FestivalTTS	.012	.818	(-.09, .11)
	DHV	-.036	.491	(-.14, .07)

Reaction times

Table 4.2.6 presents the results of the reaction times of PWA for each type of TTS. There was a significant difference between Ogmios and Festival TTS and the DHV. In fact, the RTs associated with sentences spoken by Ogmios were significantly longer. There was a similar trend for the difference between Ogmios and Bluemix, but the difference did not reach significance here.

Table 4.2.6: Pairwise comparisons of the reaction times in the four conditions.

Condition (a)	*condition (b)*	Mean difference *(a-b)*	*p*	95% CI
DHV	*Bluemix*	-2032.6982	.088	(-4367.1247, 301.7283)
	FestivalTTS	-1460.5433	.301	(-4230.2161, 1309.1294)
	Ogmios	-6871.6389 [a]	**.001**	(-10985.5738, -2757.7040)
Bluemix	*DHV*	2032.6982	.088	(-301.7283, 4367.1247)
	FestivalTTS	572.1549	.767	(-3216.6500, 4360.9598)
	Ogmios	-4838.9407	.057	(-9817.0946, 139.2132)
FestivalTTS	*Bluemix*	-572.1549	.767	(-4360.9598, 3216.6500)
	DHV	1460.5433	.301	(-1309.1294, 4230.2161)
	Ogmios	-5411.0956 [a]	**.000**	(-7712.8527, -3109.3384)
Ogmios	*Bluemix*	4838.9407	.057	(-139.2132, 9817.0946)
	FestivalTTS	5411.0956 [a]	**.000**	(3109.3384, 7712.8527)
	DHV	6871.6389 [a]	**.001**	(2757.7040, 10985.5738)

Voice preference

The second part of the experiment aimed at investigating whether the participants have a clear preference for one voice over the others. The data analysis did not show any clear preference for one voice. The most preferred voice was the DHV (12 preferences), followed by Bluemix (10), Ogmios (8) and Festival (4). Some participants also selected "none" (3) and "both equally" (5) as an answer.

The results of the control group are comparable with the ones given by the target population. Also in this case, the DHV appears to be the preferred voice by far compared to any voice. The scale of preference was the DHV, with 42 preferences, followed by Bluemix (20), Ogmios (10) and then Festival TTS (7). In the task conducted by the control group, in some cases, participants selected "both equally" (5) and "none of them" (6) as an answer.

4.2.3 Discussion

The study's main objective was to investigate which TTS system is more suitable for PWA and reading difficulties. For this purpose, comprehension rate and comprehension delay (reaction times) were measured. In addition, it was also investigated whether the patients had a preference among the four different voices or not. Individual preferences are crucial to consider when assessing tools intended for specific populations.

We found no significant difference in comprehension accuracy between the four different voices. The three TTS performed similarly among them, also in comparison with the digitalised human voice. A possible explanation for this is two-fold: i) the quality of the artificial voices has now reached a quality that does not hinder the auditory comprehension in populations with special needs as are PWA; ii) the possibility for the patients to select the preferred speech rate in which to conduct the experiment had a major impact on the comprehension accuracy.

In 2017, Hux et al. conducted an experiment that, similarly, investigated whether PWA comprehended synthetic speech with sufficient accuracy to justify the pursuit of using TTS systems as a strategy to compensate for persistent reading difficulties. The results showed a significant difference in comprehension accuracy between the digitalised human voice and the two TTS systems selected for the study: Alex TTS, available through the Macintosh platform and David TTS, available for Windows that, at the time, were considered two representatives of high-quality synthetic speech. However, differently from our experiment, the speech rate was the same for every participant. In Hux et al. (2017) study, the speaking rate was 205.56 words per minute (SD 27.23), while in our current study, participants could select the preferred speech rate, and none of them selected the default one (210.864 words per minute (SD 32.58)) - which was similar to the speech rate in Hux et al. (2017) - but they mostly selected the 25% slower

(168.69 words per minutes (SD 26.07)) followed by the 50% slower (140.57 words/minute (SD 21.72)).
A more recent study by the same authors, Hux, Brown and Wallace (2020), investigated the impact on comprehension and revision time of three different TTS speech rates in PWA. This study concluded that the most preferred rate was the medium rate of 154 words per minute (wpm) and that the faster rate caused extra revision time. The comprehension accuracy did not significantly change across different rates (slow rate-113 wpm and fast rate 200 wpm). However, this study was conducted in a combined modality – audio with text support - which might have played a substantial role as the visual support, when possible to use it, facilitates comprehension, which means that the visual support might have caused the lack of difference in comprehension rate across speech rate. The two Hux studies (2017 and 2020) confirm to some extent the impact of speech rate on comprehension in PWA with reading impairments. In fact, if the first Hux experiment (2017) was not in line with our results, the second experiment, on the other hand, confirmed that, in general, PWA and reading impairments prefer and have lower revision time in a speech rate around 150 wpm. Our experiment poses itself between these two studies as we investigated different speech rates in a single modality comparing different TTS systems.
We observed that one artificial voice – Ogmios - caused longer reaction times compared to the other three voices that, on the other hand, performed similarly. Ogmios has been built using the unit selection synthesis technology, which, compared with the most modern technology (i.e. IBM Bluemix developed using deep neural networks), produces a voice easily attributable to a computer-generated speech. We conducted a study aiming at analysing the possible reasons that caused this significant result. (This study is extensively described in section 4.3 following the discussion.)

The data analysis of the voice preference did not reveal any specific predilection. Contrary to expectations, PWA did not significantly prefer one voice over the others. However, the pattern of the answers is similar to the one observed for the control group. In fact, in general, the digitalised human voice is the preferred voice, followed by Bluemix, Ogmios and Festival TTS. The voice quality of the digitalised human voice and Bluemix is considerably higher than the other two, which would lead to thinking that the answer would be unanimous towards the higher quality voices compared to one of the other two lower-quality voices. In general, the variety of the answers in both groups was somewhat unexpected, leading to the belief that the preference for synthetic voices is not as straightforward for people outside the field of study as it might be thought. Also, individual preferences play a major role, which is essential to consider, especially if the main goal is a long-term acceptance and use of the compensative technology (Scherer, 2005).

4.3 Influence of text-to-speech systems performance on reaction times

The reaction times results were subsequently investigated by assessing the samples voices in their intelligibility, naturalness and voice quality parameters (i.e., jitter and shimmer). The following experiment was based on the experiment and the results reported in the previous section (4.2); therefore, the methodological part concerning procedure, the comprehension task materials and the information about participants are the same as described in subchapter 4.2.

4.3.1 Synthetic speech assessment

First of all, for each patient, we observed their test outcomes and presented an overview of the results. Secondly, we automatically evaluated the test samples to score the two main ways of assessing synthesised speech: intelligibility and naturalness. The former defines how well the words spelt in the speech message are understood by the listener, while the latter reflects how natural it is compared to a human voice.

Intelligibility

In the current work, we assessed the intelligibility of a synthetic utterance by means of an automatic speech recognition system (ASR). As shown by the authors in Baby et al. (2020), there is a correlation between the accuracy of an ASR system and the intelligibility of a speech utterance. All elements that may cause speech to be difficult to understand directly increase the error rate in an ASR system, such as mispronunciations, audio distortions, or inconsistent variations of speech loudness and rate. Therefore, we transcribed all of the synthetic audio samples with an ASR system, consequently measuring the word error rate (WER) plus the token error rate (TER) per sample. Note that since our ASR system outputs graphemes as tokens, our TER score is just the grapheme error rate. Lower error rates denote higher speech intelligibility, whereas higher error rates may reflect less intelligible speech. WER and TER measurements are computed by comparing the transcription from the ASR system with the ground truth human transcription. Both scores are based on the Levenshtein distance (Levenshtein, 1966), which is the minimum number of edits that need to be done in a text sequence to match another. For example, the WER score is computed by counting the number of word substitutions (S), deletions (D), and insertions (I) needed to convert the ASR transcription into the ground truth one. Then, these counts are summed up, dividing the result by the number of words in the ground truth transcription. The computation of TER is done the same way, but by counting token edits instead of whole word edits. Regarding the ASR system specifics, we used a pre-trained transformer based model from wav2letter (Pratap et al., 2019). The model was trained with the Spanish Multilingual LibriSpeech dataset (Pratap et al., 2020), and its recipe can be found in wav2letter's repository. To obtain the transcriptions, we decoded the acoustic model outputs with the Viterbi algorithm without any additional language model (LM). Given that we wanted to assess pure acoustic intelligibility, we avoided using an LM since it might correct some pronunciation mistakes that we wanted to detect. The WER and TER analyses were automatically computed. An input file with the .wav files—audio sentences—and the ground truth transcription was run through the handler programme, an executive file, and the ASR model (a previously trained deep learning model). The output file of this analysis reports (i) the hypothesis prediction text, namely, the ASR output for each audio sentence, (ii) the reference text (the ground truth), and (iii) the WER and TER for each sentence.

Naturalness

Naturalness is defined as when the speaker's voice has the correct use of emphasis, intonation, pitch, intensity, and pauses according to the message and intention (Sanders et al., 1968). We used Non-Intrusive Quality Assessment (NISQA) to predict the naturalness of the voices used in the experiment. NISQA is a model able to automatically evaluate a super-wideband speech quality without the need of a clean reference signal (Soni & Patil, 2016). This same model was used to predict the mean opinion score (MOS) in terms of naturalness of the speech samples (Mittag & Möller, 2020). Initially, the MOS test was used to evaluate global voice quality, and it was initially used for phone speech, and it required human evaluators for subjective measurements. Nowadays, the MOS test is frequently used to score naturalness, intelligibility and the quality of TTS systems. Since the initial ways to conduct MOS test (i.e. with human evaluators) was expensive and time-consuming, the recent MOS tests used in synthetic speech are models – such as NISQA – that predict the answers of human evaluator, without resorting to them. However, these models are developed and trained using human MOS ratings. Its architecture consists of a stack of CNN layers that predict spectral frame-level quality scores in an intermediate step, followed by a RNN LSTM module to predict the final MOS score taking the output of the CNN together with the Mel-Frequency Cepstral Coefficients (MFCC). Manually annotated MOS data from Voice Conversion Challenge (VCC) 2018 and Blizzard were used to train NISQA. The model is publicly available for research purposes.

Jitter and Shimmer parameters

Jitter and shimmer are two acoustic measurements widely used as metrics for pathological voice detection and acoustic analysis more in general (Farrús et al., 2007). Jitter and shimmer are measurements of the fundamental frequency disturbance, and they have been proven to be useful in describing vocal characteristics (Teixeira et al., 2013). Jitter is defined as the parameter of the fundamental frequency variation of successive periods, while shimmer refers to the amplitude variation of the sound wave (Zwetsch et al., 2006). Jitter is affected mostly by the reduced or lack of control of vibration of the vocal cords. Shimmer, instead, is correlated with breathiness and the presence of noise in vocal emission; it usually changes with the reduction of the glottal resistance and lesions on the vocal cords (Teixeira et al., 2013).

Once the onset time of the glottal pulse is recognised, jitter can be determined using different measures. The current study considered jitter local and jitter relative average perturbation (RAP). The former represents the average absolute difference between two consecutive periods, divided by the average period; its threshold limit to identify pathologies is 1.04%, and it is expressed as a percentage (Farrús et al., 2007):

$$\text{Jitter (local)} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i}$$

The latter is the relative average perturbation, meaning the average absolute difference between a period and the average of it and its two neighbours, divided by the average period; its threshold value for detecting pathologies is 0.68% (Teixeira & Gonçalves, 2014), and it is expressed as:

$$\text{Jitter RAP} = \frac{(1/N-2)\ \sum_{i=2}^{N-1} |T_i - ((T_i + T_{i-1} + T_{i+1})/3)|}{\frac{1}{N}\sum_{i=1}^{N} T_i}$$

Shimmer refers to irregular period-to-period variation in the peak amplitude of the signal. In the current study, shimmer local was taken into consideration. Shimmer local is defined as the average absolute difference between the amplitudes of two consecutive periods, divided by the average amplitude. The limit for detecting human voice pathologies is 3.81% (Zwetsch et al., 2006), and it is expressed as a percentage:

$$\text{Shimmer (local)} = \frac{\frac{1}{N-1}\ \sum_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i}$$

The parameters described here are commonly measured for long sustained vowels. However, in our case, the samples are short sentences. For this reason, the interpretation of the results will be made keeping in mind this aspect. Another aspect to consider was that, according to (Slyh et al., 1999), significant differences could occur in jitter and shimmer measurements between different speaking styles, especially concerning the shimmer measure. Praat software was used to extract these measurements from all our test samples (Boersma & Weenink, 2020).

4.3.2 Results

The analysis of the intelligibility and naturalness was executed both at the system-level (i.e., analyse the performance based on TTS system) and at the sentence-level (i.e., analyse the performance based on the type of sentences). We decided to investigate also at the sentence-level because there was a sentence that required to be re-listened as some participants could not understand it the first time. Both samples with 50% and 25% slower speech rate reduction were considered. Table 4.3.1 shows the RTs results from the experiment described in subchapter 4.2. The longest RTs for most of the patients occurred with the Ogmios synthesis samples. In general, Bluemix and DHV achieved closer to average RTs, although Festival TTS took the lead in a couple of patients. Surprisingly, Patient 6 had longer RTs for the DHV and Bluemix.

Table 4.3.1: Patients' average reaction times per TTS system. (Bold values are the largest RT of each patient.)

Average Reaction Times (s)							
	P1	P3	P4	P5	P6	P7	P8
DHV	6.5	7.58	9.72	14.43	23.57	6.9	19.64
Ogmios	**11.27**	**15.55**	**19.52**	**23.83**	20.85	**9.85**	**35.87**
Festival	8.02	11.23	9.06	15.85	18.54	8.05	27.9
Bluemix	5.65	8.47	10.1	14.45	**31.32**	6.97	25.71

System-level analysis

The plots in Figure 4.3.1 show DHV and TTS systems’ performance in terms of WER, TER and MOS. We assessed all sentences with 25% and 50 % reduction in speech rate. As expected, the DHV achieved the best scores in terms of naturalness and intelligibility compared to the TTS systems. However, the WER increased when the speech velocity was reduced to 50%. This may happen because the ASR system is not robust enough for such a reduction. Bluemix obtained very similar MOS scores to the human ones, while WER and TER were higher than Festival, the TTS system that achieved the lowest MOS score. Ogmios obtained the highest WER and TER values, meaning it had lower intelligibility than the other voices, although it presented a better score compared to Festival TTS in terms of naturalness. It must be considered that, differently from the unit selection (Ogmios), the statistical parametric synthesis technology (Festival TTS) tends to be better in terms of intelligibility. However, at the same time, the naturalness of the speech easily degrades, especially in long sentences (King, 2011). In the context of this experiment, the sentences were short enough not to allow a relevant worsening of the naturalness and were still better in terms
of intelligibility compared to Ogmios. This might explain the difference in intelligibility and naturalness between Ogmios and Festival TTS.

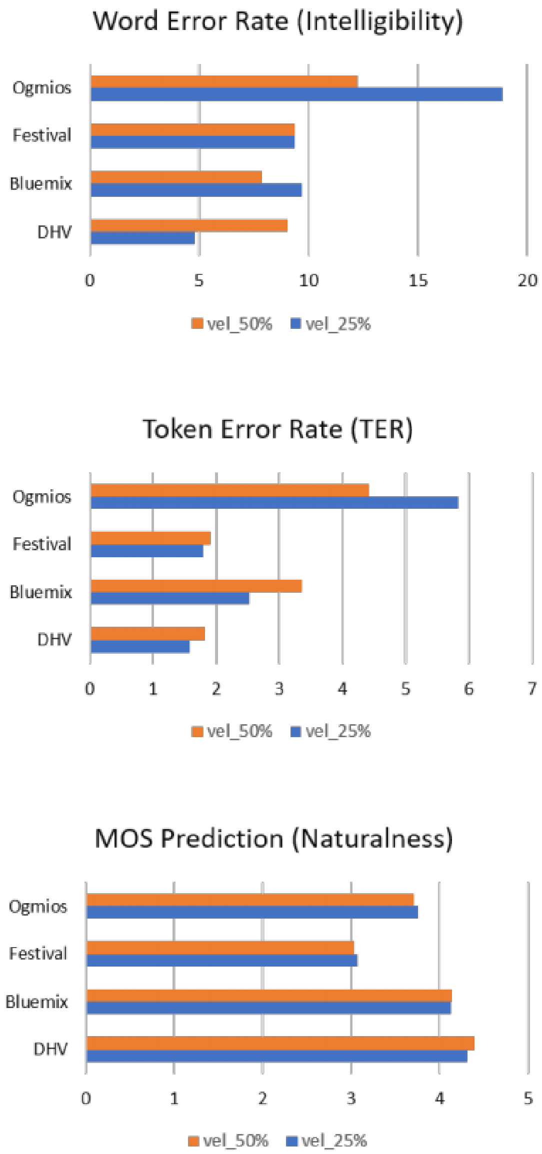


Figure 4.3.1: System-level analysis of the WER, TER and MOS prediction test of the three synthetic voices and the DHV at their 25% and 50% slower speech rate.

Jitter and shimmer analysis

Table 4.3.2 shows the resulting average jitter local, Jitter RAP and Shimmer local of each synthesis system and DHV at the original speech rate, Table 4.3.3 at 25% and Table 4.3.4 at 50% slower rate.

Table 4.3.2: Average jitter, jitter RAP and shimmer in (%) from samples at original speech rate.

	DHV	Bluemix	Ogmios	Festival
Jitter local (%)	2.67	2.56	3.70	1.83
Jitter RAP (%)	1.18	0.92	1.53	0.51
Shimmer local (%)	9.56	9.85	10.39	8.96

Table 4.3.3: Average jitter, jitter RAP and shimmer in (%) from samples at 25% reduced speed.

	DHV	Bluemix	Ogmios	Festival
Jitter local (%)	1.298	1.612	2.183	1.292
Jitter RAP (%)	0.425	0.416	0.782	0.287
Shimmer local (%)	7.757	7.723	8.115	7.824

Table 4.3.4: Average jitter, jitter RAP and shimmer in (%) from samples at 50% reduced speed

	DHV	Bluemix	Ogmios	Festival
Jitter local (%)	1.16	1.39	1.91	1.05
Jitter RAP (%)	0.37	0.35	0.67	0.23
Shimmer local (%)	6.5	6.49	6.8	6.5

The relative values for jitter were found to be higher in the Ogmios synthesis, which denotes that the perturbation in f0 cycles caused by concatenative synthesis artefacts was higher than the other synthetic voices. In fact, the Ogmios system, once again, showed the highest values in all metrics. Considering the mentioned thresholds for the pathological voices, all systems, including DHV, surpassed the jitter and shimmer local, while the jitter RAP threshold was only surpassed by Ogmios. This could have happened as we measured over the whole signal, so pitch and intensity present more variation than measuring a single vowel, for instance. Thus, even though these results do not show that Ogmios is the only voice with value beyond the pathological threshold

(except for the jitter RAP), it is the voice that presented the highest values compared to the others, which gives an orientation of TTS systems' performance. The results, particularly for jitter and shimmer, should
be regarded as indicators of speech deterioration when compared against pathological voice threshold—specifically, that Ogmios is more prone to pathological features than the other voices.

Sentence-level analysis

Observing the comprehension test results, 4 out of the 7 patients had difficulties understanding one of the Ogmios samples the first time they heard it, so their RTs increased substantially. Specifically, the sentence that needed a second listen was sentence n 36 ("la mujer en la tienda ríe." [*the lady in the shop laughs*]). This phenomenon was further analysed to investigate whether it was the reason why Ogmios presented such a significant difference in RTs compared to the other voices. Figure 4.3.2 depicts the RT distribution for every item per person aggregated by the different voices to investigate if the seven items (one for each participant) corresponding to sentence 36 were outliers, indicating that this sentence exclusively caused the significant RT difference between them Ogmios and the other voices. However, after investigating the possible outliers related to sentence 36, the results do not indicate the items corresponding to this sentence (items 177, 189, 201, 213, 225, 237, and 249) are outliers (Figure 4.3.2). In fact, from figure 4.3.2, it is possible to observe that the highest value in RT (an extreme outlier) was achieved by item 222, which was an item corresponding to sentence 21 in participant 6. In the case of item 222, the considerably higher RT could explain why, in the clinical experiment, the significance between Ogmios and Bluemix in terms of RTs is not as clear as in the case of Festival TTS and the DHV.

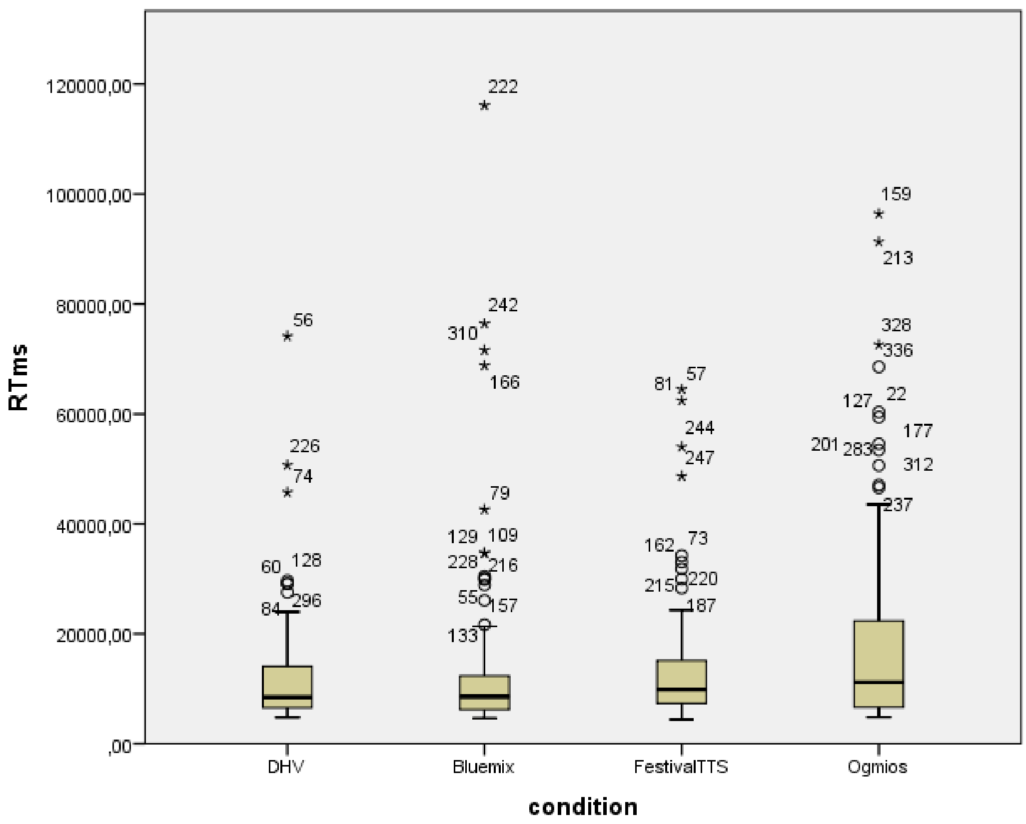


Figure 4.3.2: Box plot illustrating the RTs distribution of the single items by the four voices.

The correlation between RTs and MOS, WER and TER measurements was also investigated; however, the results of the Pearson's correlation analysis (ranging between -0.02 and 0.23) did not show correlation between the parameters considered.

4.3.3 Discussion

This study aimed to answer the following research questions: (i) Are the significant results in reaction times in Ogmios TTS caused by one or more acoustic voice characteristics? (ii) If so, which one? After conducting a system-level automatic assessment of the intelligibility and naturalness, the sample voices were analysed using jitter and shimmer parameters. Ogmios, in general, performed worse than the other voices in these measurements; therefore, there are indeed acoustic voice characteristics that contributed to the longer RTs in Ogmios: intelligibility, jitter, and shimmer, while the naturalness was better in Ogmios compared with Festival TTS – the statistical parametric synthesis – and the other voices. The results concerning intelligibility and naturalness are in line with the characteristics of the different speech synthesis methods (see chapter 3, subchapter 3.1).

4.4 Limitations

Despite the careful planning of the experiments described in this chapter, there are still aspects of the experimental design that could be improved. In the procedure, the reaction times were recorded in relation to when the researcher pressed the keyboard number corresponding to the image that the patient indicated as the answer. A better solution for this problem would have been to use a tablet; however, the only means available was a personal computer. Moreover, the programme we used to execute the experiment (DMDX) did not present a touch screen input modality if not for a specific version of the software (6.1.2.0 DMDX) and to interface with eye trackers. For this reason, we resort to the pc keyboard to record the answers and the reaction times. However, the extra time for the researcher to press the number on the keyboard was added to all the answers, therefore, the inaccuracy of the recording was distributed across all the items.

The last part of the experiment that could represent a limitation is that for the control group, there is no information about the reaction times. During the pilot experiment, we tried to send the experiment built with DMDX via email, but the process of downloading the compressed file generated by DMDX was too problematic for some participants, so we excluded this modality from the final experiment. The reason why the experiment with the control group was not done presentially is because it would have reached a lower number of people and mostly people from the research group acquainted with text-to-speech, so the preference task would have been highly biased.

4.5 Conclusions

In this chapter, the main goal was to have a better insight into synthetic speech in relation to PWA and reading impairments. We investigated the impact of four different TTS on comprehension and cognitive load; next, we further investigated what are the voice parameters and quality performance that impact on the cognitive load of these patients. The results of these experiments show that the current TTS systems do not affect comprehension accuracy when the experiment is carried out with the speech rate preferred by the participant. On the other hand, the results on reaction times suggest that indeed, there are differences between the different types of TTS when it comes to cognitive cost. These results were further explored by analysing the intelligibility, naturalness and pathological voice parameters to identify what might have caused this measurable difficulty in reaction times.

To conclude, when considering a TTS for PWA and acquired reading impairments, it should be kept in mind that there is no one-fits-all solution. From our experiment, it appeared that for the Spanish language, the concatenative speech synthesis available is not suitable for the target population considered. These results are not generalisable to all concatenative speech systems for every language, and we realise it is TTS-specific. However, our results would suggest that an initial analysis of jitter, shimmer, and WER and TER intelligibility measures could give initial information about whether a synthetic voice could be used with PWA with acquired reading impairments. Nonetheless, repeating the experiment with another TTS for another language with a larger sample size could complete the results gathered in this analysis. For further investigations, the results obtained from the acoustic analysis and automatic investigation of intelligibility and naturalness could be paired and further investigated based on Cohn and Zellou (2020) results. Cohn and Zellou reported that in a noisy environment, neural network TTS systems (such as Bluemix in the current study) have lower intelligibility in a healthy population than concatenative systems. This is due to the fact that neural networks have more coarticulated and human-like speech than concate native systems, which have a more hyper articulated voice. As a result, the concatenative system is more understandable than neural networks in a noisy, competing environment. A further extension of the current study could investigate which acoustic characteristics are not detrimental for PWA and acquired reading impairments in a functional, realistic situation such as a noisy environment and do not cause longer RTs.

Chapter 5

THE RELEVANCE OF SYNTACTIC SIMPLIFICATION IN AUDITORY CONTENT COMPREHENSION

There is evidence that people who acquired aphasia and present mild, moderate or severe reading difficulties benefit from simplified lexicon, syntax or paragraphs (Brennan et al., 2005; Rose et al., 2003, 2011). In clinical practice, speech and language therapists usually manually simplify sentences and texts or look for different levels of complexity in written material to train and improve the reading abilities of their patients. At the same time, when the patients do not follow a rehabilitation program anymore, their chances to have simplified material drop dramatically. In fact, the possibility to have simplified material is generally absent, and reading or simplifying texts could become a task that weights on the caregivers (Hux et al., 2017). Even after a rehabilitation cycle, the patients could still have difficulties with long and complex sentences, paragraphs or low-frequency words.

Text simplification (TS) is still an ongoing discipline that entails the possibility to transform lexically and syntactically complex texts or sentences into a simpler version that could be easily understood by a specific population, as in our case, people with aphasia (PWA) and acquired reading impairments (Štajner, Calixto, et al., 2015). This chapter presents the experiment we designed to investigate whether automatic text simplification improves patients' comprehension of written material. If so, TS, associated with text-to-speech systems, could help PWA with acquired reading difficulties to access written material more independently. In this experiment, we decided to start with smaller units than text (i.e., sentences) to proceed in steps. Our target population are PWA who acquired difficulties in reading, specifically individuals who do not benefit from a combined modality of presentation (audio and written input). For this reason, the experiment resorts exclusively to auditory comprehension ability. For an extensive description of text simplification systems and their application, refer to chapter 3, section 3.2. The main contributions of this chapter are:

- The identification of critical points in the current ATS technologies with specific regards to their use in tools for PWA and acquired reading impairments;
- What lessons we learned from the transposition of the study from in presence to online;

This chapter will be structured as follows: section 5.1 illustrates the methods that we planned to use for the experiment and the design and procedure that we actually used as a consequence of the outburst of the coronavirus crisis; section 5.2 reports the findings of the study; section 5.3 discusses the results and reports the conclusions of the Chapter with also comments concerning the lessons we learned from the study.

5.1 Methodology

The initial plan of the methodology involved in-person testing of the same participants of the first experiment (see Chapter 4, subsection 4.2.1). In this study, the main goal was to test if automatically simplified sentences are easier to understand for the patients compared to complex relative and passive clauses, for the most part, extracted from newspaper articles. In the in-person experiment plan, the patients hear a target sentence followed by a test sentence that either corresponds to the target sentence or not. On the screen, there were two options: a) yes, they have the same meaning, b) no, they do not have the same meaning. The participants had to point to one of these two options immediately after listening to the target and the test sentences. The target sentences were either complex clauses extracted from newspapers or a simplified version of these sentences created by Tuner, an automatic text simplification system that simplifies for the Spanish language (see Chapter 3, subsection 3.2.1). We planned to allow the patients to choose the speech rate they preferred and supervise the experiment to ensure they understood the instructions and were ready to start the task. Also, we planned to use the same software – DMDX (see Chapter 4, section 4.2.1) – as the first experiment to record the reaction times alongside the accuracy rate. At the end of the experiment, we planned to collect comments and observations from the patients regarding the simplified clauses (e.g., if the sentences sounded forced and weird or if they felt comfortable with the simplification) and the voice paired with them. Even though the TTS was already assessed in the first experiment (see Chapter 4), in this second study, they did not have any visual support; therefore, the ease of understanding and its likability might slightly change.

However, unfortunately, we had to recalibrate and modify the investigation due to the COVID-19 pandemic outburst just a month before this second experiment was scheduled. Therefore, the methodology described in this Chapter is a readaptation of the original one. The new design was thought and built quickly to meet the necessity to continue developing the research project even during the sanitary emergency; thus, the experiment shifted from in presence to online modality. The online testing was possible thanks to the contact with a Spanish medical centre and research centre, Vithas, which has various branches throughout Spain. Specifically, our main contact was a researcher in the Vithas Valencia facility.

5.1.1 Recruitment

Initially, we sent to our contact in Vithas Valencia a mock format of our online test to give the sense of our investigation to the researcher and the speech and language therapists (SLT) who volunteered to help us actualise the test. The mock format was considerably shorter than the final one, but it gave a general idea. Before sending the official links of the online experiment, alongside the mock test, we also sent an extensive description of i) the target population we were interested in (specifying the assessments we needed), ii) the detailed description of the goal of the experiment, iii) the detailed description of how the experiment procedure.

The final official links to the online experiment were sent to the contact in Vithas, who subsequently distributed the links to the SLT who decided to help in the investigation. The links were sent through an Excel file containing a list of 25 different links. Each link was unique to allow us to associate the anonymous ID of the patients that the SLT provided us with the results. The SLT noted what link was given to which patient, and afterwards, the SLT compiled the initial Excel file with anonymous IDs connecting patients' clinical assessments to the results.

The participants had to listen to two recordings and indicate if two recordings had the same meaning. The first – the main sentence - of the two recordings was either a complex sentence or the same sentence in content but automatically simplified; the second recording was a test sentence, a short clause to measure if the participants understood the first sentence. This experiment has complex and simplified sentences as independent variables and the comprehension rate as dependent variables.

5.1.2 Materials

This section illustrates the materials we used for the online experiment. We aimed to build a study that was as much similar as possible to everyday materials to have results as close as possible to the use of automatic text simplification in reading in real. For this reason, the majority of the sentences were extracted from articles we found in some of the major Spanish and Catalan newspapers: "La Vanguardia", "El Mundo", and "El País". We looked for coordinates relative and passive sentences. The sentences in the study not extracted from the newspaper were created manually; specifically, we found it challenging to find passive structures in coordinate sentences. Hence, we created some of the coordinate passive sentences manually.
In this Chapter, we use the term *cluster* to indicate a group of clauses composed of i) a complex sentence, ii) one or two simplified sentences (depending on if the complex sentence has a passive verb), and iii) a test clause. We use the term *query* to indicate the complex and the simplified sentences part of a cluster. For example, in figure 5.1.1, Q58, Q60 and Q62 are the complex passive sentences, the automatically simplified and manually simplified sentences.

Sentences

The experiment presented 90 sentences divided by the grammatical structure and level of complexity. Among these 90 sentences, 36 were complex sentences, namely relative clauses and passive sentences. Among other complex syntactic structures, usually these ones cause challenges in PWA both at auditory and reading comprehension levels (Caplan, 2006; Caplan et al., 2007; DeDe & Kelleher, 2021), therefore, for this experiment, we used complex relative (e.g., "*Algunos países constituyen una excepción, entre los cuales España, que fija la fecha de reapertura para el primero de julio.*" [Eng: *Some countries constitute an exception, among which Spain, which sets the reopening date for the first of July])* and passive clauses (e.g., "*Richard Osman fue influido quizás por la figura de su abuelo, que fue policía en Brighton." [*Eng: *Richard Osman was perhaps influenced by the figure of his grandfather, who was a policeman in Brighton]).* Both types of sentences presented coordinate syntactic structures. The coordinate clauses are made when connecting two or more clauses (called conjuncts) through coordinating conjunction (e.g., "and", "or", "but").

The remaining 54 sentences were automatically simplified or a mix of automatically and manually simplified sentences. The passive sentences have two simplification options because Tuner (the automatic simplifier used in the study and described in Chapter 3, subsection 3.2.1) does not simplify passive structures. Therefore, in the specific case of the passive clauses, firstly, we used Tuner to simplify the complex syntactic construct present alongside the passive verbs, and then, from Tuner's simplification, we manually added the simplification of the passive verbs. In Figure 5.1.1, there is an example of the simplification process we used in the experiment. The first sentence (Q58) is the sentence we extracted from the newspaper; the second sentence (Q60) is the sentence resulting from Tuner simplification. Specifically, in Q60, Tuner simplified the relative clause by forming two separate sentences removing the relative structure. The third sentence (Q62) is where we manually simplified the passive verb leaving the rest of the clause unvaried. Hence Q62 is the result of the automatic simplification applied in Q60 plus our manual simplification.

Q58: Richard Osman fue influido quizás por la figura de su abuelo, que fue policía en Brighton.

[Richard Osman was perhaps influenced by the figure of his grandfather, who was a policeman in Brighton.]

Test sentence: El abuelo de Richard Osman podría haberlo influido.
[Richard Osman grandfather might have influenced him]

Q60: Tuner simplified sentence: Richard Osman fue influido quizás por la figura de su abuelo. Este su abuelo fue policía en Brighton.

[Richard_Osman was perhaps influenced by the figure of his grandfather. This his grandfather was a policeman in Brighton.]

Q62: Manually simplified (passive structure simplification): El abuelo de Richard Osman podría haberlo influido. Este su abuelo fue policía en Brighton.

[Richard Osman's grandfather could have influenced him. This his grandfather was a policeman in Brighton]

Figure 5.1.1 Steps of the simplification process from Q58 being the sentence extracted from the newspaper to Q62, the sentence presenting both the automatic simplification and the manual simplification. All these sentences together represent a cluster.

Table 5.1.1 summarises the type of clauses we used. The first column includes the kind of sentence and type of simplification we used --if we used any-- (e.g. a relative complex clause: "*Richard Osman, who in another life would have liked to be a detective, but his journey has been in different directions before meeting the world of crime again in the form of extraordinary literary success*"). In the second column, we specify if the test sentence matched the meaning of the main clause (e.g., the test sentence matched the original sentence: *"Richard Osman had a literary success"*). The test sentence is a short sentence reporting only part of the content of the main sentence. The test sentence could have i) the same meaning, ii) the opposite meaning, iii) a different action verb, iv) a different object from the main sentence. We used it to measure the comprehension rate of the complex clauses versus the simplified clauses. The last column shows the number of sentences per type.

Table 5.1.1: Summary of the different types of sentences and the kind of simplification applied. The 9 test sentences with different meanings for any type of main sentence were divided as follows: 3 with reversed meaning, 3 with a different action verb, and 3 with a different object.

Main sentence types	**Test sentence**	**number**
Relative clause complex	Same meaning	9
Relative clause simplified by Tuner	Same meaning	9
Relative clause complex	Different meaning	9
Relative clause simplified by Tuner	Different meaning	9
Passives	Same meaning	9
Passive simplified by Tuner (simplification of the relative and coordinate structures)	Same meaning	9
Passive manually simplified (Tuner simplification + passive structure manually simplified)	Same meaning	9
Passives	Different meaning	9
Passive simplified by Tuner (simplification of the relative and coordinate structures)	Different meaning	9

Passive manually simplified (Tuner simplification + passive structure manually simplified)	Different meaning	9
Total number of sentences		**90**

The following images include the script of a cluster of sentences used in the experiment. Every query with the test sentence appeared as a separate item during the experiment. Specifically, Q2 and the test clause appeared separately from Q5 and the test sentence; however, both are part of the same cluster. See the following example:

Q2: En su opera prima, Mortensen encarna a un hombre, John Peterson, que choca de manera aparatosa con su padre Willis cuando éste acude a visitarle a su casa de California.

[In his debut, Mortensen plays a man, John Peterson, who clashes in a spectacular way with his father Willis when he visits his home in California.]

Test Sentence: John Peterson pelea con su padre.
[John Peterson fights with his father]

Figure 5.1.2: Query 2: complex sentence extracted from a newspaper. The test sentence is to test the participants' auditory comprehension.

Q5: simplified sentence: En su opera prima, Mortensen encarna a un hombre, John Peterson. Este un hombre, John Peterson choca de manera aparatosa con su padre Willis cuando éste acude a visitarle a su casa de California.

[In his debut, Mortensen plays a man, John Peterson. This one man, John Peterson clashes dramatically with his father Willis when he comes to visit him at his home in California.]

Test Sentence: John Peterson pelea con su padre.
[John Peterson fights with his father]

Figure 5.1.3: Query 5: sentence simplified by Tuner. The test sentence is the same as the Q2 as they belong to the same cluster.

Voice

For this experiment, we resort to DCTTS, a neural network TTS system. For more detailed information about this TTS, see Chapter 3, section 3.1.1. We decided to use this system as the results from the experiment described in Chapter 4 (subsection 4.2) suggested that deep learning-based or statistical speech synthesis are a good fit for PWA. In light of these results, we decided to use the TTS built with the more cutting-edge technology. The speech rate in this experiment was set at 168.69 words/minute (SD 26.07), which was one of the two speech rates preferred by the participants in the first experiment (for reference, check Chapter 4, section 4.2).

Online testing tool

The experiment was carried out using Qualtrics software[6]. Qualtrics allows for complex layering structure, for instance, giving feedback after an answer; moreover, it easily allows audio clips as test material differently from other free tools for online surveys. Furthermore, Qualtrics collects the response time and enables inserting an auto-advance option into the experiment (an option that automatically moves respondents to the next page after they answer). This tool is usually implemented in psychology or economics as an online survey tool. To our knowledge, it was the first time it was used in the context of online testing for people with aphasia. However, due to the sanitary emergency, this was the only viable option that responded to the necessity to transport the experiment online.

5.1.3 Participants

We provided the inclusion criteria for the participants in the experiment to our contact in Vithas medical centre. The inclusion criteria were the following:

(a) individuals with aphasia and acquired reading difficulties after a stroke (event not prior to six months);

(b) being a Spanish native speaker;

(c) aphasia severity ranging from moderate to mild (WAB AQ 50-82) (Kertesz, 1982);

(d) no working memory impairments, operationalised with an onward Corsi block tapping test (Kessels et al., 2000) score higher than 5;

[6] (https://www.qualtrics.com)

(e) patients reporting either difficulties or the inability to read in their everyday life;

(f) reading comprehension is worse than auditory comprehension.

For this study, five participants were recruited. Table 5.1.2 shows the information of the patients with the assessment sent by the SLT.

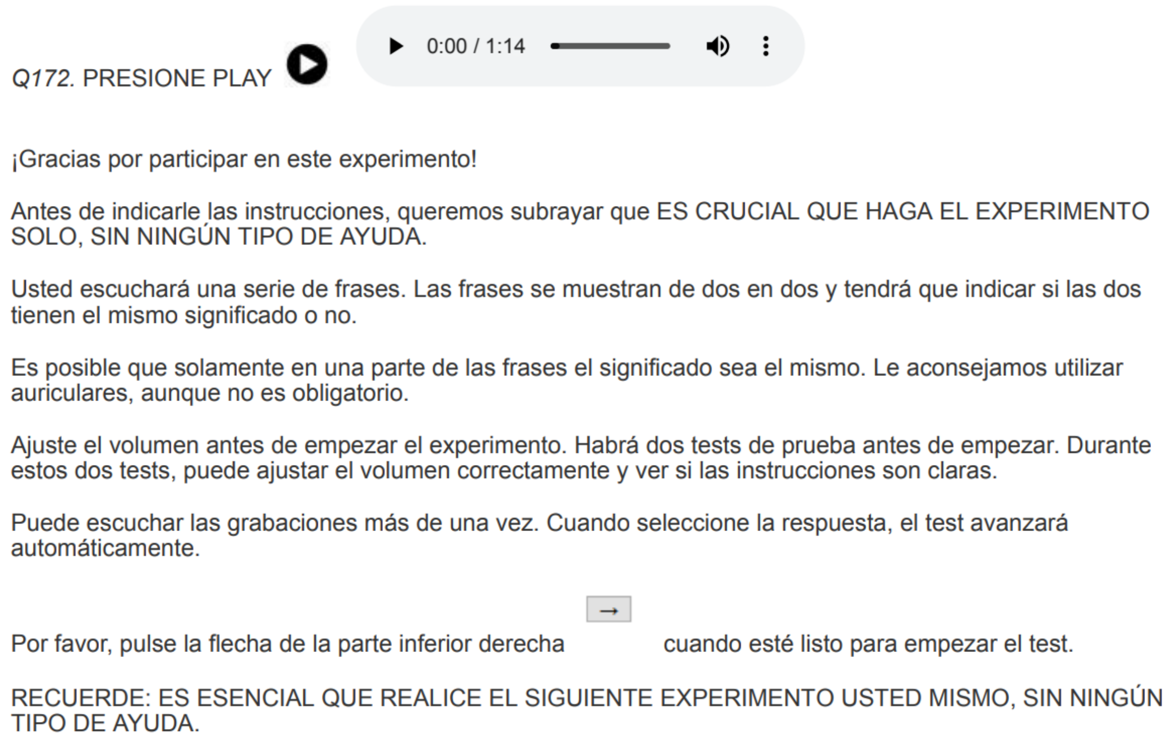
Table 5.1.2: Evaluation carried by the speech and language therapists in Vithas Hospitals. * YoE: Years of education; *MAST C: Mississippi Aphasia Screening Test.

ID	Age	YoE*	MASTsp-R *	Corsi block-tapping test		Boston Diagnostic Aphasia Examination - reading subtests	
				Onward (max. 16)	Backward (max. 14)	Word reading (max. 30)	Sentences reading (max. 10)
1/M	60	8	48/50	7	5	30	10
2/F	30	17	32/50	9	7	0	0
3/M	58	11	48/50	7	7	30	10
4/F	65	12	49/50	8	6	30	10
5/M	53	10	48/50	6	4	27	10

The SLT from the medical centre assessed the participants using the Spanish version of the Mississippi Aphasia Screening Test (MASTsp). The MAST test is a screening tool for detecting potential alterations in different language areas (Nakase-Thompson et al., 2002; Nakase-Thompson et al., 2005) and allows a rapid determination of which key language aspects require a more thorough analysis. According to Nakase-Thompson et al. (2005), this test accuracy has been demonstrated in patients with language disorders secondary to a different array of cerebral events. This test presents nine subtests, and its score range from 0 (severe aphasia) to 100 (healthy individual). The MAST for Spanish also offers four receptive subtests (MASTsp-R) and five expressive subtests (MASTsp-E). In both subsections, the scores range from 0 to 50. In the case of the current study, the SLT tested the participants only in the receptive subsection, giving us information on their comprehension level in listening and reading. The Corsi block-tapping test and the Boston Diagnostic Aphasia Examination are the same tests used in the first experiment; for more information about these two tests and what they assess, refer to chapter 4.

5.1.4 Procedure

Assuming that our instructions were met, the online experiment was structured as follows: i) the patients received and opened the link (or were helped in opening it), ii) they saw the introduction to the experiment. The introduction was written and recorded to allow the participant to understand the instructions even if they could not read. Figure 5.1.4 shows a screenshot of what the participants saw once they opened the link.



Q172. PRESIONE PLAY

0:00 / 1:14

¡Gracias por participar en este experimento!

Antes de indicarle las instrucciones, queremos subrayar que ES CRUCIAL QUE HAGA EL EXPERIMENTO SOLO, SIN NINGÚN TIPO DE AYUDA.

Usted escuchará una serie de frases. Las frases se muestran de dos en dos y tendrá que indicar si las dos tienen el mismo significado o no.

Es posible que solamente en una parte de las frases el significado sea el mismo. Le aconsejamos utilizar auriculares, aunque no es obligatorio.

Ajuste el volumen antes de empezar el experimento. Habrá dos tests de prueba antes de empezar. Durante estos dos tests, puede ajustar el volumen correctamente y ver si las instrucciones son claras.

Puede escuchar las grabaciones más de una vez. Cuando seleccione la respuesta, el test avanzará automáticamente.

Por favor, pulse la flecha de la parte inferior derecha → cuando esté listo para empezar el test.

RECUERDE: ES ESENCIAL QUE REALICE EL SIGUIENTE EXPERIMENTO USTED MISMO, SIN NINGÚN TIPO DE AYUDA.

Figure 5.1.4. Screenshot of the introduction to the experiment. [ENG: "Thank you for taking part in this test! Before illustrating to you the instructions, we want to underline that IT'S CRUCIAL THAT YOU PERFORM THE EXPERIMENT ALONE WITHOUT ANY HELP. You will have to listen to some sentences. The sentences are displayed in pairs, and you will need to indicate if the sentences have the same meaning. Only a part of the sentences may be the same. We advise using a headphone, however, it is not compulsory. Adjust the volume before starting the test. You can do it right now. There will be two trials before the start of the test. During these trials, you can set up the volume properly and see if the instructions are clear. You can listen to the recordings more than once. When you select the answer, the test will advance automatically. Please, press the arrow at the bottom right when you are ready to proceed with the test.REMEMBER: IT IS ESSENTIAL THAT YOU EXECUTE THE FOLLOWING EXPERIMENT ON YOUR OWN, WITHOUT ANY KIND OF HELP."]

After the initial introduction, two trials were provided to the patient to understand the instructions better. For each example, we also provided feedback to either confirm they understood the instructions or correct them. Figure 5.1.5 depicts one of the trial tests, while Figures 5.1.6 and 5.1.7 show screenshots of the feedback answers. The recordings of the general introductions, feedback, and the instructions in each item are recordings of a human voice. The TTS system synthesised only the two sentences part of the comprehension task.

Figure 5.1.5. Trial test number one. [ENG: "This is the first trial item. Please adjust the volume making sure you can hear well and see the screen well. Also, make sure you understand the instructions correctly. Indicate if A and B recordings have the same meaning. Press play for listening to the recorded sentences. Remember, press the arrow "next "in the bottom right corner to proceed to the next item."]

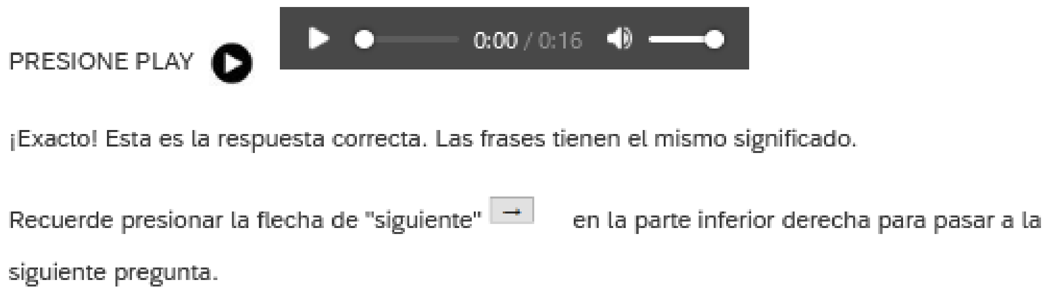


Figure 5.1.6. Screenshot of what appeared on the screen in case of the participant answered correctly during the trial test. [ENG: "Exactly! This is the right answer. The sentences have the same meaning. Remember, press the arrow "next "in the bottom right corner to proceed to the next item."]

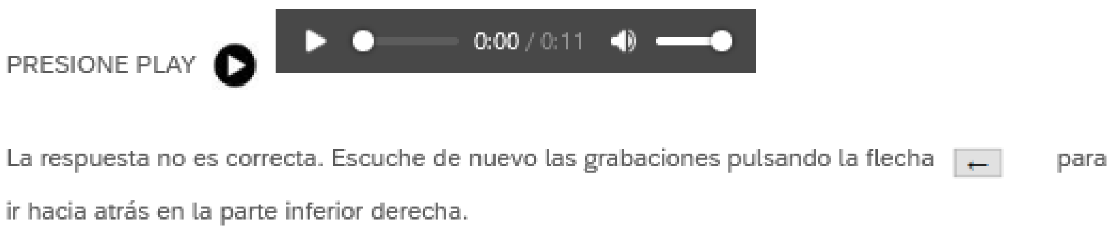


Figure 5.1.7. Screenshot of what appeared on the screen in case of the participant answered wrongly during the trial test. [ENG: "The answer is not correct. Listen again to the recording pressing the arrow at the bottom right to go back."]

Following these introductory and illustrative parts, the actual tasks started. Figure 5.1.8 depicts an example of the first item of the experiment. At the top of the page, the instructions are repeated both in written and audio modalities. The recording named A is the complex or simplified clause, while B is the test sentence.

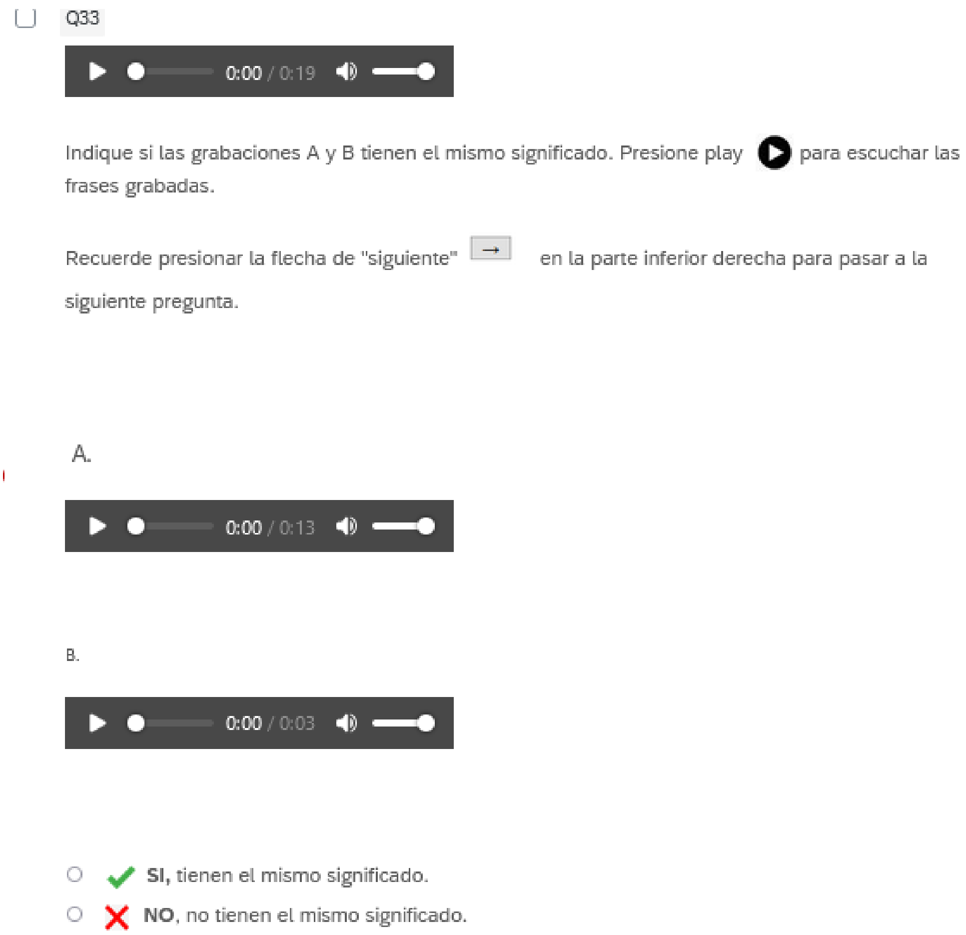


Figure 5.1.8. Example of one item of the experiment. [ENG: "Indicate if the recordings A and B have the same meaning. Press play to listen to the recorded sentences. Remember, press the "next" arrow in the bottom right part of the screen to proceed to the next item" The answers are: "Yes, they have the same meaning", "No, they don't have the same meaning".]

5.2 Findings

Table 5.2.1: Summary of the percentage of correctness per type of clause and cluster of sentences.

	% of correctness
Relative clauses	
Automatically simplified sentences	69%
Complex sentences	61%
Passive clauses	
Automatically simplified sentences	68,8%
Manually simplified sentences	67,7%
Complex sentences	61%

Table 5.2.2: Percentage of correctness when the test sentences have the same content or a different content compared to the main clause.

	% of correctness
Passive clauses – test sentence	
Same meaning	36,6%
Different meaning	40%
Passive clauses – test sentence	
Same meaning	28%
Different meaning	28,8%

Visual inspection of Table 5.2.1 shows that automatic simplification of the relative clauses did cause a slight improvement in comprehension. On the contrary, transforming the passive verb into an active form does not show the improvement in comprehension that we expected. In fact, in the three-step simplification process of the passive structures, the more relevant improvement was not simplifying the passive structure but the simplification of the complex structure from a complex long sentence (61 %) to a shorter and without relative grammar structure sentence (68,8 %). This part of the simplification was made by Tuner. The last step of simplification – from the automatically simplified to the manually simplified sentence - did not cause an important comprehension increase, but actually a slight decrease in comprehension (from 68,8 % to 67,7%). However, we are cautious in interpreting the findings, will be reported and commented not taking into consideration the results in relation to the single participants, nor we will give statistics about the significance as we cannot consider the results sufficiently reliable for the following reasons: i) all the participants received the same link; ii) the inclusion criteria were not met for all participants; iii) we are not sure that the participants completed the test without guidance as requested. Firstly, the links to the online experiment were not distributed as we intended to: all participants received the same links, making it impossible for us to track down and pair the link with the participants' assessment and the correspondent results. For this reason, the results could not be analysed based on the participants' profiles. Secondly, from the evaluation tests provided by the SLT, among the five participants, only one seems to meet all the inclusion criteria of the study. Lastly, we cannot be sure that participants completed the test unaided. We did not receive further information about the execution of the task; therefore, we cannot be sure that the participants were not guided to any extent during the experiment. Alongside the online test results, the SLT who agreed to help in the distribution of the experiment sent us a few comments regarding the difficulties the participants encountered. The main problem was the TTS voice that presented lower voice quality (especially in terms of clarity) on longer sentences. Another point brought up by the SLT was that the DCTTS did not pronounce the English names accurately; in fact, the TTS read the sentences using the Spanish phonetics, resulting in unclear terms. Lastly, they believed that the experiment was too difficult for the target population required, and many withdrew from the study before completing it.

5.3 Discussion

This experiment aimed to investigate whether pairing a TTS with aphasia-friendly features (i.e., slower speech rate) with an automatic text simplifier would help PWA understand complex sentences that might be found in everyday reading material such as newspapers. We decided to use two syntactic structures that usually cause comprehension difficulties in PWA and reading impairments: relative clauses and passive clauses. The sentences selected were either extracted from a Spanish newspaper or were created manually.

Looking at the findings of this experiment, we need to be aware that the results might not belong to the participants listed in the participants' assessment table (Table 5.1.1) due to the issue related to the distribution of the links. Therefore, the consideration hereby presented are limited observations that may not refer to the target population we had in mind when designing the initial experiment.

The results (shown in table 5.2.1 and table 5.2.2) display that the simplified sentences (both for the relative and passive clauses) have a slightly higher percentage of correctness than the complex sentences. In the case of the passive sentences, the simplification was done following two steps (refer to Figure 5.1.1 for clarification). The first step was to automatically simplify a complex sentence using Tuner. Using Tuner, we simplified only coordinates syntactic structure and relative structure. The second step was to simplify the passive verb manually. The manual simplification was essential as Tuner could not transform passive verbs into their active form. We intended to simplify the passive structure to verify if the simplification of the passive verbs was essential in the understanding process from a complex to a simpler sentence in PWA and acquired reading impairment. Unlike what we expected, the sentences only automatically simplified (hence with passive verbs) presented a slightly higher percentage of comprehension rate (68,8%) than the sentences with the manual simplification of the passive verbs (67,7%). This result might raise questions about whether the simplification of passive verbs is a fundamental feature when considering a simplification tool for PWA and acquired reading impairments. However, the difference between the two types of sentences was minimal, suggesting that we should rerun the experiment in-person on a wider and properly selected sample population to have a more reliable view of the topic. Nonetheless, since we found a low number of passive sentences in the newspapers (forcing us to create the missing sentences manually), if these results are confirmed even after repeating the experiment in-person, we could consider the automatic simplification of the passive sentences not crucial. However, further investigation is needed to present such conclusions. Another observation raised by the limited presence of passive sentences in newspapers is about the clinical validity of having aphasia assessment tests with a large part of the item investigating the ability of the patient to understand passive forms. Do the language assessments that present an extensive investigation of the passive verbs comprehension mirror the real challenges that PWA will face operating in the real world? From the clinical and diagnostic point of view, it is interesting to interrogate the ecological validity of thoroughly assessing passive structures when investigating the level of language deficits and, subsequently, the real necessity for a device to simplify passive verb structures. Is it necessary to thoroughly assess passive grammar structure comprehension and have an automatic text simplification tool for PWA if passive grammar structures are very limited and fairly difficult to encounter?

Another consideration concerns the TTS system. When a TTS system reads international news, a wide portion of the words used in the text will be names and nouns stemming from a variety of different languages; in fact, foreign names, especially English, commonly occur in different types of text content such as in newspapers or books (Sitaram & Black, 2016). Phonetisation of foreign names based on letter-to-sound rules from the TTS target language will typically be faulty (Mendelson et al., 2017). However, the majority of TTS systems work on the assumption that the input text is fully in a single language – i.e., in the TTS systems target language (Sitaram et al., 2019), and DCTTS is no exception. Indeed, DCTTS for the Spanish language cannot read – therefore speak - English names with the right English pronunciation, as the utterance follows the Spanish phonatization. This issue was both observed during the experimental design and reported by the SLT. To the extent of this experiment, we could not find another open-source TTS for the Spanish language that recognises foreign names or words and reads them aloud accordingly to the right phonetisation. While this problem might represent a manageable inconvenience for individuals whose first language is the TTS system's target language, we cannot overlook it in the case of PWA, as this difficulty adds to the already present reading and auditory challenges. This issue concerning TTS systems and code-switching is well known, and current approaches in handling this aspect fall into three broad categories: i) phone mapping, ii) multilingual, and iii) polyglot synthesis (Sitaram et al., 2019). In phone mapping, the foreign language phones are replaced with the primary language closest sounding phones, resulting in a speech with a strong accent. The multilingual approach synthesises the different text parts in a different language and then synthesises them by the corresponding monolingual TTS system. In this last approach, the risk is that each TTS will have different voices. Lastly, the polyglot solution consists of a single system trained using data from a multilingual speaker.

5.3.1 Lessons learned

The sanitary emergency urged us in finding another viable way to perform the experiment and collect the data. We were lucky enough to have promptly met the interest of a medical and research centre in Valencia. Nevertheless, the transposition of the experiment from in presence to online required more effort and time than initially expected. In the online experimental set-up and during the findings observation, we could identify a few aspects to consider carefully.

Online platform for the experiment:

The first issue with the online platform for the study was, indeed, to find a suitable tool that met our requirement to make online testing as much aphasia-friendly as possible. This first step took quite a long time as the main free platforms did not allow some key features, such as adding audio clips that could be opened and listened quickly and intuitively. The online tools explored were: SurveyMonkey, Typeform, SoGoSurvey, Google Form and Survey Gizmo. Despite these tools being excellent for online surveys, they lack some structural features that were either completely absent or restricted to subscription fees. Specifically, using these tools for free did not allow us to easily add audio clips, for instance, with Google Form, it is possible to add a link to an audio file saved in a Google Drive folder. These steps are way too many that could hinder the execution of the experiment. Another recurring feature that was missing in the free versions - or in general in these platforms - was the possibility to customise and add feedback to all the different answers. To this extent, Qualtrics met our expectations, and we were lucky enough that the economics department of Pompeu Fabra University has the licence of this software, and they agreed to share with us the access. Even though having access to Qualtrics as it was exactly the kind of platform we needed and we had the luck to have access to it, setting up the experiment was not without technical challenges. The major technical problem was that the auto-advance option did not work for everyone who received the link. We tried to solve this problem by contacting Qualtrics' client service; however, the problem was not solved, forcing us to add the instruction of pressing the "next" arrow, which we wanted to avoid to make the process easier for the participants. Regarding this first point, we learned that it could be useful to consider an online alternative to the experiments and, more in general, be aware of the software that might serve the purpose of the investigation in order to save time and effort in the case of need.

Third-party involved:

The main problem encountered for the final realisation of the experiment was the multi-level communication we went through. This part refers mostly to the issue concerning the distribution of the links. If, on the one hand, having a medical centre interested in our study and willing to help in providing us with participants for the online test was invaluable and crucial, on the other hand, the level of communication between the patients and us multiplied, causing miscommunications and, eventually, the disruption of the crafted organisation of the study. We learned about this circumstance that an online modality mediated by a third party is not recommended. We would generally discourage performing such experiments online, especially with people who acquired aphasia and reading difficulties. However, most certainly, if it is not possible to avoid an online modality, we believe it is crucial to directly connect with the patients, or at least with the SLT involved.

5.4 Conclusion

To conclude, we tried to control the experimental design and procedure as much as possible, especially in consideration of the sanitary emergency that outburst just before the scheduled start of the experiment. Even though the experiment presented various critical issues from the set-up to the findings, we still believe that the study was needed and that there are valuable observations that we could draw from it.

In general, the lessons we learnt point towards the need to execute the experiment in person, if possible, but if not, we suggest: i) knowing online platform suitable for the experiment beforehand; ii) keeping the level of communication between the researcher and the patients as lower as possible. Lastly, we observed from the current study that current technologies only partially meet PWA with acquired reading impairments necessities. The automatic text simplification tool and the TTS systems used in this experiment presented issues (i.e., not simplifying the passive verbs and not reading English names with the correct pronunciation) that possibly do not facilitate comprehension in this population, differently from what we aimed to do. In this regard, there is still the need for more accurate tools for compensatory supports for PWA and reading impairments.

Chapter 6

CONCLUSIONS AND FUTURE WORK

This thesis explored whether combining two natural language processing techniques could facilitate access to written content in individuals with aphasia and acquired reading difficulties. To this end, we used two different techniques: speech synthesis and automatic text simplification, which were tested individually and in combination. We were interested in identifying an aphasia-friendly text-to-speech system and subsequently pairing it with a text simplification system for the Spanish language. This investigation aimed to contribute to the expanding field of compensative instruments for PWA with acquired reading impairments. This research area still lacks an instrument able to allow autonomous access to written information to individuals who do not benefit from the combined modality of text presentation; hence, they could rely mostly on – or exclusively – the auditory channel. This area of research currently focuses on identifying which TTS systems' characteristics facilitate text comprehension and processing time in PWA when a text is presented with or without the support of a TTS system.

Next, we divide our conclusions as follows: section 6.1 describes the results from the literature review on reading compensative tools for PWA and acquired reading impairments; section 6.2 reports the main findings and conclusions about the first experiment about text-to-speech systems; section 6.2 shows the conclusions of the second experiment about automatic text simplification paired with the text-to-speech system selected from the first; lastly, in part 6.4 we report the future work connected with this thesis project.

6.1 Text accessibility

In Chapter 2, we gave an overview of the target population of the thesis. We described aphasia and acquired reading impairments in their common classification system, incidence and prevalence. At the end of the second chapter (see sections 2.3 and 2.4), we reported the literature review results that we published about text accessibility options for PWA and acquired reading impairments (Cistola et al., 2021). The literature review highlighted that, on a daily basis, PWA either do not use any reading aid or they use accessibility features and software that are not designed for them specifically. Hence, these tools' usefulness is limited and varies from one patient to another based on their language impairment severity or familiarity with technological instruments. The literature review suggested the need for a tool that considers all the variables in place when we talk about PWA with acquired reading impairments (e.g., cognitive difficulties possibly associated with aphasia or motor abilities difficulties). The individuals who acquired aphasia could present a great variability in the clinical picture. For this reason, a tool that could help PWA in accessing written content should present a considerable degree of customisation (both in speech output and automatic text simplification features) and ease of use (e.g., not too many steps to make the tool work).

6.2 Speech synthesis systems

In this thesis project, we wanted to identify a text-to-speech system that would not hinder PWA's auditory comprehension or reaction times. In recent years, there have been a growing number of studies aiming to find the gold standard aphasia-friendly TTS settings (see section 3.1.2).

From these studies, it appears that the recent advancements in speech synthesis do not pose any challenges to auditory comprehension in these patients. When the speech rate of the TTS is adjusted to the patients' preference, there is no difference in comprehension rate in aided and unaided conditions (e.g., Hux et al., 2021). Thus the presence of a TTS support, with a personalised speech rate, mostly impacts the processing time rather than improving comprehension of a text (e.g., Knollman-Porter et al., 2022).

In our thesis project, we were interested in looking at the TTS systems for PWA from another perspective. In fact, different technologies present different characteristics in voice parameters (see section 3.1), and we wanted to investigate if one technique - with its specific characteristics - was more accessible to PWA. In our findings, similarly to the results from other studies in the field, the comprehension rate was not impacted by the different speech synthesis systems. However, one specific TTS did cause significantly longer reaction times compared to the other voices. Ogmios, a system built using a concatenative unit selection-based technology, showed significantly longer reaction times in the participants in our experiment (see section 4.2.2). Concatenative technology usually has poor intelligibility compared to the other TTS systems (see section 3.1).

From further investigations trying to identify the reason behind this result, we examined the voice used in the first experiment in terms of intelligibility, naturalness, and voice quality parameters. Our results highlighted that Ogmios performed worst under every aspect except for naturalness compared to the statistical parametric synthesis. However, this last result is not surprising as the concatenative synthesis usually has better naturalness than the statistical speech synthesis (see section 4.3.2). This analysis has to be taken as pertaining only to Ogmios speech synthesis. The reason is that concatenative speech synthesis requires a large database to obtain a good voice, meaning that the voices could slightly differ from one speech synthesis to another, especially in languages with less vast resources than English.

The results of the TTS system identification process suggested two main conclusions:
i) when selecting a TTS system for PWA and reading difficulties, it is advisable to use either a statistical parametric speech synthesis or a deep neural network-based TTS and allow the participants to select their preferred speech rate;
ii) if possible, analyse beforehand the voice intended to use in terms of intelligibility, naturalness and voice quality parameters to exclude the possibility that the voice could pose any difficulties (e.g., cognitive load as in our experiment case) to the patients.

6.3 Automatic text simplification

The second part of the thesis project was to test patients with simplified sentences spoken by a deep learning-based TTS system that we selected based on the first experiment results. This second part aimed to investigate whether automatically simplifying complex sentences and reproducing them through a TTS would improve comprehension of complex sentences. Looking at the findings, we observed that comprehension improved when relative complex sentences were simplified, while there was no noticeable difference between the passive sentences and the simplified ones (i.e., passive sentences manually converted into active) (see tables 5.2.1 and 5.2.2). However, these findings were not sufficiently reliable to run statistical analysis on them or draw more than observations (see section 5.2); therefore, they need to be taken carefully. These results, paired with the difficulty that we encounter in finding complex passive structures in newspaper articles, raised the question of whether the simplification of passive sentences is, in fact, necessary.

In clinical practice, auditory and reading comprehension of passive sentences are usually extensively examined. However, in everyday life, the chances to encounter passive grammatical structures while reading a book or a newspaper, for instance, are few. From this consideration, more than drawing conclusions, we reflected on the real necessity of a text simplification device that simplifies passive sentences. If they are so rare to encounter, and their simplification does not facilitate sentence comprehension, should we reconsider the simplification of passive structures?

6.4 Future work

In all research, there are always some issues left for future research. Next, we describe the future work that could complete the research undertaken until now. We start discussing the remaining questions about text-to-speech systems for people with aphasia and acquired reading impairments; then, we report possible improvements and further work related to the application of automatic text simplification in sentence comprehension for the target population in question; lastly, there is the description of what I had in mind when starting pursuing the doctoral programme and the long way to go that it is ahead to accomplish it.

Text-to-Speech systems

For future work, it would be interesting to test the two TTS systems (IBM Bluemix and Festival TTS) that did not cause a delayed reaction time in a noisy environment. That is, to test if the speech synthesis characteristics hold in terms of comprehension and reaction times even when in a competing setting. This idea is inspired by Cohn & Zellou (2020) work, reporting that in a noisy environment, deep learning-based TTS has lower intelligibility than concatenative speech synthesis. Since we excluded the concatenative speech synthesis as a viable TTS for PWA, we would test the statistical parametric synthesis compared to the deep learning-based TTS system. Will a statistical parametric synthesis be easier to understand than a deep learning TTS in a noisy setting?

Automatic text simplification

Unfortunately, the second experiment of the thesis project involving automatic text simplification has been compromised in its design due to the transposition of the study from in presence to online. We believe that the research question involving automatic text simplification still stands, and it needs to be further explored by repurposing the experiment in presence in a controlled setting to answer this question. The question about the necessity to simplify passive structures needs to be further investigated as there might have been other factors affecting these findings that we are not aware of since we could not control for the patients that took the online test.

Proposal for a final tool

When I started the PhD programme, I worked towards the idea of outlining a highly customisable device that could help PWA with acquired reading difficulties in their everyday struggle in wanting to recover their reading habits. Ideally, the final tool would give the possibility – after the patient selected the text of interest online – to tailor the customisation options of i) the text-to-speech and ii) the automatic text simplification on their language difficulties level so that they could access the content of interest using this tool to maximise their residual language ability.

In this idea, there would be an initial training of the patients by the speech therapist or the caregiver, showing how to set the options, especially for the speech output, as there could be the need to change voice settings based on the location (e.g., environmental noise could require to slow down the speech rate or change the gender voice as, for instance, female voice are generally higher in pitch). After the initial training about settings, there would be a few sessions aiming to train the devices with the text simplifications features that are important for the individual. In this way, every time the patient uses the tool, it would directly simplify the text to make it immediately understandable to the patient. This part about the simplification process was inspired by the work done by Bingel et al. (2018) with their web extension, called Lexi, for individuals with developmental dyslexia.

I am aware that there is still a long way to go before accomplishing this goal, especially because we have to consider that this is a combination of two technologies that are exponentially improving year after year but still present some limitations. For instance, the pronunciation of names and words in a language different from the one selected for the text-to-speech (as we discussed in section 5.3) or some difficulties still present with text simplification and meaning preservation. There is still plenty of work to do to produce a fluent and simplified text and connect it with a TTS that adapts to the person who uses it. However, this is the main idea that underlined the research that went on during the PhD and even though the work to do is still plenty, I still believe in its validity.

Bibliography

Agerri, R., Bel, N., Rigau, G., & Saggion, H. (2018). TUNER: Multifaceted domain adaptation for advanced textual semantic processing. First results available. *Procesamiento de Lenguaje Natural*, *61*, 163–166. https://doi.org/10.26342/2018-61-23

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., & Asari, V. K. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics*, *8*(3), 292. https://doi.org/10.3390/electronics8030292

Baby, A., Vinnaitherthan, S., Adiga, N., Jawale, P., Badam, S., Adavanne, S., & Konjeti, S. (2020). *An ASR Guided Speech Intelligibility Measure for TTS Model Selection*. http://arxiv.org/abs/2006.01463

Baeza-Yates, R., Rello, L., & Dembowski, J. (2015). Cassa: A context-aware synonym simplification algorithm. *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1380–1385.

Barbu, E., Martín-Valdivia, M. T., Martínez-Cámara, E., & Ureña-López, L. A. (2015). Language technologies applied to document simplification for helping autistic people. *Expert Systems with Applications*, *42*(12), 5076–5086. https://doi.org/10.1016/j.eswa.2015.02.044

Barlacchi, G., & Tonelli, S. (2013). ERNESTA: A sentence simplification tool for children's stories in Italian. *International Conference on Intelligent Text Processing and Computational Linguistics*, 476–487.

Bautista, S., León, C., Hervás, R., & Gervás, P. (2011). Empirical identification of text simplification strategies for reading-impaired people. In *Everyday Technology for Independence and Care* (pp. 567–574). IOS Press.

Beeson, P M, & Hillis, A. E. (2001). *Comprehension and production of written words. In. R. Chapey (Ed.). Language intervention strategies in adult aphasia, (pp. 572-595). Baltimore, MD: Lippencott*. Williams & Wilkins.

Beeson, Pelagie M, Hirsch, F. M., & Rewega, M. A. (2002). Successful single-word writing treatment: Experimental analyses of four cases. *Aphasiology*, *16*(4–6), 473–491. https://doi.org/10.1080/02687030244000167

Beeson, Pélagie M, Rising, K., Kim, E. S., & Rapcsak, S. Z. (2008). A novel method for examining response to spelling treatment. *Aphasiology*, *22*(7–8), 707–717. https://doi.org/10.1080/02687030701800826

Behrmann, M., & Plaut, D. C. (2014). Bilateral hemispheric processing of words and faces: evidence from word impairments in prosopagnosia and face impairments in pure alexia. *Cerebral Cortex*, *24*(4), 1102–1118.

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., De Ferranti, S. D., Floyd, J., Fornage, M., & Gillespie, C. (2017). Heart disease and stroke statistics—2017 update: a report from the American Heart Association. *Circulation*, *135*(10), e146–e603.

Bernhard, D., De Viron, L., Moriceau, V., & Tannier, X. (2012). Question generation for French: collating parsers and paraphrasing questions. *Dialogue & Discourse*, *3*(2), 43–74.

Bingel, J., Paetzold, G. H., & Søgaard, A. (2018). Lexi: A tool for adaptive, personalized text simplification. *Proceedings Ofthe 27th International Conference on Computational Linguistics*, 245–258. https://www.readwithlexi.net.

Blumstein, S. E. (2016). Psycholinguistic approaches to the study of syndromes and

symptoms of aphasia. In *Neurobiology of Language* (pp. 923–933). Elsevier.
Boersma, P., & Weenink, D. (2020). *Praat: Doing phonetics by computer [Computer program](Version 6.1. 24)*.
Bonafonte, A., Adell, J., Agüero, P. D., Erro, D., Esquerra, I., Moreno, A., Pérez, J., & Polyakova, T. (2007). *The UPC TTS System Description for the 2007 Blizzard Challenge*.
Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal–ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76.
Bott, S., Rello, L., Drndarevic, B., & Saggion, H. (2012). Can spanish be simpler? LexSiS: Lexical simplification for spanish. *24th International Conference on Computational Linguistics - Proceedings of COLING 2012: Technical Papers*, *December*, 357–374.
Brennan, A. D., Worrall, L. E., & McKenna, K. T. (2005). The relationship between specific features of aphasia-friendly written material and comprehension of written material for people with aphasia: An exploratory study. *Aphasiology*, *19*(8), 693–711. https://doi.org/10.1080/02687030444000958
Breznitz, Z., Shaul, S., Horowitz-Kraus, T., Sela, I., Nevat, M., & Karni, A. (2013). Enhanced reading by training with imposed time constraint in typical and dyslexic adults. *Nature Communications*, *4*(1), 1–6.
Briscoe, T., & Carroll, J. (1995). Developing and evaluating a probabilistic LR parser of part-of-speech and punctuation labels. *ArXiv Preprint Cmp-Lg/9510005*.
Broca, P. (1861). Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech). *Bulletin de La Société Anatomique*, *6*, 330–357.
Broca, P. (1865). On the Site of the Faculty of Articulated Speech (English Translation). *Archives of Neurology*, *43*, 1065–7210.
Brookshire, C. E., Wilson, J. P., Nadeau, S. E., Gonzalez Rothi, L. J., & Kendall, D. L. (2014). Frequency, nature, and predictors of alexia in a convenience sample of individuals with chronic aphasia. *Aphasiology*, *28*(12), 1464–1480. https://doi.org/10.1080/02687038.2014.945389
Brown, J. A., Knollman-Porter, K., Hux, K., Wallace, S. E., & Deville, C. (2021). Effect of digital highlighting on reading comprehension given text-to-speech technology for people with aphasia. *Aphasiology*, *35*(2), 200–221. https://doi.org/10.1080/02687038.2020.1787728
Brown, J. A., Wallace, S. E., Knollman-Porter, K., & Hux, K. (2019). Comprehension of Single Versus Combined Modality Information by People With Aphasia. *American Journal of Speech-Language Pathology*, *28*(1S), 278–292. https://doi.org/10.1044/2018_AJSLP-17-0132
Canning, Y., & Tait, J. (1999). Syntactic simplification of newspaper text for aphasic readers. *ACM SIGIR'99 Workshop on Customised Information Delivery*, 6–11.
Canning, Y., Tait, J., Achibald, J., & Crawley, R. (2000). Cohesive Generation of Syntactically Simplified Newspaper Text. In *Text, Speech and Dialogue* (pp. 145–150). Springer Berlin Heidelberg.
Caplan, D. (2006). Aphasic deficits in syntactic processing. *Cortex*, *42*(6), 797–804.
Caplan, D., Michaud, J., & Hufford, R. (2013). Short-term memory, working memory, and syntactic comprehension in aphasia. *Cognitive Neuropsychology*, *30*(2), 77–109.
Caplan, D., Waters, G., DeDe, G., Michaud, J., & Reddy, A. (2007). A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and*

Language, *101*(2), 103–150.
Caramazza, A., & Hillis, A. E. (1990a). Spatial representation of words in the brain implied by studies of a unilateral neglect patient. *Nature*, *346*(6281), 267–269. https://doi.org/10.1038/346267a0
Caramazza, A., & Hillis, A. E. (1990b). Where do semantic errors come from? *Cortex*, *26*(1), 95–122.
Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246–251.
Carroll, J., Minnen, G., Canning, Y., Devlin, S., & Tait, J. (1998). Practical Simpliication of English Newspaper Text to Assist Aphasic Readers. *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, 7–10. http://alpha.mic.dundee.ac.uk/~slanger/workshop.html
Carroll, J., Minnen, G., Pearce, D., Canning, Y., Devlin, S., & Tait, J. (1999). Simplifying Text for Language-Impaired Readers. *European Capter of Association of Computational Linguistics*. http://osiris.sunderland.ac.uk/~pset/welcome.html
Caute, A., Cruice, M., Friede, A., Galliers, J., Dickinson, T., Green, R., & Woolf, C. (2016). Rekindling the love of books – a pilot project exploring whether e-readers help people to read again after a stroke. *Aphasiology*, *30*(2–3), 290–319. https://doi.org/10.1080/02687038.2015.1052729
Caute, A., & Woolf, C. (2016). Using voice recognition software to improve communicative writing and social participation in an individual with severe acquired dysgraphia: an experimental single-case therapy study. *Aphasiology*, *30*(2–3), 245–268. https://doi.org/10.1080/02687038.2015.1041095
Caute, A., Woolf, C., Wilson, S., Stokes, C., Monnelly, K., Cruice, M., Bacon, K., & Marshall, J. (2019). Technology-enhanced reading therapy for people with aphasia: Findings from a quasirandomized waitlist controlled study. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4382–4416. https://doi.org/10.1044/2019_JSLHR-L-18-0484
Chandrasekar, R., Doran, C., & Bangalore, S. (1996). Motivations and methods for text simplification. *COLING 1996 Volume 2: The 16th International Conference on Computational Linguistics*.
Cherney, L. R. (2010a). Oral Reading for Language in Aphasia (ORLA): Evaluating the Efficacy of Computer-Delivered Therapy in Chronic Nonfluent Aphasia. *Topics in Stroke Rehabilitation*, *17*(6), 423–431. https://doi.org/10.1310/tsr1706-423
Cherney, L. R. (2010b). Oral reading for language in aphasia: Impact of aphasia severity on cross-modal outcomes in chronic nonfluent aphasia. *Seminars in Speech and Language*, *31*(01), 42–51.
Cipolotti, L., & Warrington, E. K. (1996). Does recognizing orally spelled words depend on reading? An investigation into a case of better written than oral spelling. *Neuropsychologia*, *34*(5), 427–440.
Cistola, G., Farrús, M., & Meulen, I. (2021). Aphasia and acquired reading impairments: What are the high-tech alternatives to compensate for reading deficits? *International Journal of Language & Communication Disorders*, *56*(1), 161–173. https://doi.org/10.1111/1460-6984.12569
Code, C. (2021). Aphasia. In J. S. Damico, N. Müller, & M. J. Ball (Eds.), *The Handbook of Language and Speech Disorders* (pp. 286–309). Wiley. https://doi.org/10.1002/9781119606987.ch14
Cohn, M., & Zellou, G. (2020). Perception of concatenative vs. neural text-to-speech (TTS): Differences in intelligibility in noise and language attitudes. *Proceedings of*

the Annual Conference of the International Speech Communication Association, INTERSPEECH, *2020-Octob*(October), 1733–1737. https://doi.org/10.21437/Interspeech.2020-1336

Coltheart, M. (1987). Reading, phonological recoding, and deep dyslexia. *Deep Dyslexia*, 197–226.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Coslett, H. B., & Turkeltaub, P. (2016). Acquired dyslexia. In *Neurobiology of language* (pp. 791–803). Elsevier.

Coster, W., & Kauchak, D. (2011). Learning to Simplify Sentences Using Wikipedia. *Workshop on Monolingual Text-To-Text Generation, Pages 1–9, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, June*, 1–9.

Crisp, J., Howard, D., & Lambon Ralph, M. A. (2011). More evidence for a continuum between phonological and deep dyslexia: Novel data from three measures of direct orthography-to-phonology translation. *Aphasiology*, *25*(5), 615–641. https://doi.org/10.1080/02687038.2010.541470

Crosson, B. (2013). Thalamic mechanisms in language: A reconsideration based on recent findings and concepts. *Brain and Language*, *126*(1), 73–88. https://doi.org/https://doi.org/10.1016/j.bandl.2012.06.011

Dalemans, R. J. P., De Witte, L. P., Beurskens, A. J. H. M., Van Den Heuvel, W. J. A., & Wade, D. T. (2010). An investigation into the social participation of stroke survivors with aphasia. *Disability and Rehabilitation*, *32*(20), 1678–1685. https://doi.org/10.3109/09638281003649938

Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004). Neural systems behind word and concept retrieval. *Cognition*, *92*(1–2), 179–229. https://doi.org/10.1016/j.cognition.2002.07.001

DeDe, G. (2012). Effects of Word Frequency and Modality on Sentence Comprehension Impairments in People With Aphasia. *American Journal of Speech-Language Pathology*, *21*(2). https://doi.org/10.1044/1058-0360(2012/11-0082)

DeDe, G. (2013). Reading and Listening in People With Aphasia: Effects of Syntactic Complexity. *American Journal of Speech-Language Pathology*, *22*(4), 579–590. https://doi.org/10.1044/1058-0360(2013/12-0111)

DeDe, G., & Kelleher, D. (2021). Effects of animacy and sentence type on silent reading comprehension in aphasia: An eye-tracking study. *Journal of Neurolinguistics*, *57*, 100950. https://doi.org/10.1016/j.jneuroling.2020.100950

Dejerine, J. (1892). Contribution à l'étude anatomopathologique et clinique des différents variétés de cécité verbale. *Mémoires de La Société de Biologie*, *4*, 61–90.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, *49*(1), 36–47.

Devlin, S. (1998). The use of a psycholinguistic database in the simplification of text for aphasic readers. *Linguistic Databases*.

Devlin, S., & Unthank, G. (2006). Helping aphasic people process online information. *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility - Assets '06*, 225. https://doi.org/10.1145/1168987.1169027

Dietz, A., Ball, A., & Griffith, J. (2011). Reading and Writing with Aphasia in the 21st Century: Technological Applications of Supported Reading Comprehension and Written Expression. *Topics in Stroke Rehabilitation*, *18*(6), 758–769. https://doi.org/10.1310/tsr1806-758

Døli, H., Helland, T., & Andersen Helland, W. (2017). Self-reported symptoms of anxiety and depression in chronic stroke patients with and without aphasia. *Aphasiology*, *31*(12), 1392–1409. https://doi.org/10.1080/02687038.2017.1280595

Ehrlich, M.-F., Remond, M., & Tardieu, H. (1999). Processing of anaphoric devices in young skilled and less skilled comprehenders: Differences in metacognitive monitoring. *Reading and Writing*, *11*(1), 29–63.

Elkana, O., Frost, R., Kramer, U., Ben-Bashat, D., & Schweiger, A. (2013). Cerebral language reorganization in the chronic stage of recovery: a longitudinal fMRI study. *Cortex*, *49*(1), 71–81.

Ellis, A. W. (1982). Spelling and writing (and reading and speaking). *Normality and Pathology in Cognitive Functions*, 113–146.

Ellis, A. W. (2016). *Reading, Writing and Dyslexia (Classic Edition)*. Psychology Press. https://doi.org/10.4324/9781315669991

Ellis, A. W., Flude, B. M., & Young, A. W. (1987). "Neglect dyslexia" and the early visual processing of letters in words and nonwords. *Cognitive Neuropsychology*, *4*(4), 439–464. https://doi.org/10.1080/02643298708252047

Ellis, A. W., & Young, A. W. (2013). *Human cognitive neuropsychology: A textbook with readings*. Psychology Press.

Ellis, A. W., Young, A. W., & Anderson, C. (1988). Modes of word recognition in the left and right cerebral hemispheres. *Brain and Language*, *35*(2), 254–273. https://doi.org/10.1016/0093-934X(88)90111-3

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-E

Engelter, S. T., Gostynski, M., Papa, S., Frei, M., Born, C., Ajdacic-Gross, V., Gutzwiller, F., & Lyrer, P. A. (2006). Epidemiology of aphasia attributable to first ischemic stroke: Incidence, severity, fluency, etiology, and thrombolysis. *Stroke*, *37*(6), 1379–1384. https://doi.org/10.1161/01.STR.0000221815.64093.8c

Evans, R. J. (2011). Comparing methods for the syntactic simplification of sentences in information extraction. *Literary and Linguistic Computing*, *26*(4), 371–388.

Farrús, M., Hernando, J., & Ejarque, P. (2007). Jitter and shimmer measurements for speaker recognition. *Interspeech 2007*, 778–781. https://doi.org/10.21437/Interspeech.2007-147

Ferrés, D., Marimon, M., & Saggion, H. (2015). A web-based text simplification system for english. *Procesamiento Del Lenguaje Natural*, *55*, 191–194.

Ferrés, D., Marimon, M., Saggion, H., & AbuRa'ed, A. (2016). YATS: Yet Another Text Simplifier. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9612, Issue 9612, pp. 335–342). https://doi.org/10.1007/978-3-319-41754-7_32

Ferrés, D., Saggion, H., & Gómez Guinovart, X. (2018). *An Adaptable Lexical Simplification Architecture for Major Ibero-Romance Languages*. 40–47. https://doi.org/10.18653/v1/w17-5406

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 116–124. https://doi.org/10.3758/BF03195503

Friedmann, N., Taranto, G., Shapiro, L. P., & Swinney, D. (2008). The Leaf Fell (the Leaf): The Online Processing of Unaccusatives. *Linguistic Inquiry*, *39*(3), 355–377. https://doi.org/10.1162/ling.2008.39.3.355

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, *9*(2), 20–28.

Geranmayeh, F., Brownsett, S. L. E., & Wise, R. J. S. (2014). Task-induced brain activity in aphasic stroke patients: what is driving recovery? *Brain*, *137*(10), 2632–2648.

Giannouli, V., & Banou, M. (2020). The intelligibility and comprehension of synthetic versus natural speech in dyslexic students. *Disability and Rehabilitation: Assistive Technology*, *15*(8), 898–907. https://doi.org/10.1080/17483107.2019.1629111

Glavas, G., & Stajner, S. (2015). Simplifying lexical simplification: Do we need simplified corpora? *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference*, *2*, 63–68. https://doi.org/10.3115/v1/p15-2011

Glavaš, G., & Štajner, S. (2013). Event-centered simplification of news stories. *Proceedings of the Student Research Workshop Associated with RANLP 2013*, 71–78.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.

Goodglass, H., Barresi, B., Weintraub, S., & Segal, O. (2001). *BDAE : the Boston Diagnostic Aphasia Examination* (H. Goodglass & E. Kaplan (eds.); 3rd editio). Lippincott Williams & Wilkins.

Goodglass, H., Kaplan, E., & Weintraub, S. (1983). *Boston naming test*. Lea & Febiger Philadelphia, PA.

Graves, R. E. (1997). The Legacy of the Wernicke-Lichtheim Model. *Journal of the History of the Neurosciences*, *6*(1), 3–20.

Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: Is vocabulary knowledge a mediating variable? *Journal of Literacy Research*, *43*(2), 159–183.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology*, *4*, 416.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, *28*, 136–141.

Hagoort, P. (2016). MUC (Memory, Unification, Control): A model on the neurobiology of language beyond single word processing. In *Neurobiology of language* (pp. 339–347). Elsevier.

Hickok, G., & Poeppel, D. (2000). *Towards a functional neuroanatomy of speech perception*. *4*(4), 131–138.

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*(5), 393–402. https://doi.org/10.1038/nrn2113

Hillis, A. E. (2019). Alexia and agraphia in acute and chronic stroke. In O. Godefroy (Ed.), *The Behavioral and Cognitive Neurology of Stroke* (Vol. 90, Issue 5, pp. 86–100). Cambridge University Press. https://doi.org/10.1017/CBO9781139058988.010

Hope, T. M. H., Leff, A. P., Prejawa, S., Bruce, R., Haigh, Z., Lim, L., Ramsden, S., Oberhuber, M., Ludersdorfer, P., & Crinion, J. (2017). Right hemisphere structural adaptation and changing language skills years after left hemisphere stroke. *Brain*,

140(6), 1718–1728.
Howard, D., & Gatehouse, C. (2006). Distinguishing semantic and lexical word retrieval deficits in people with aphasia. *Aphasiology*, *20*(9), 921–950. https://doi.org/10.1080/02687030600782679
Hux, K., Knollman-Porter, K., Brown, J., & Wallace, S. E. (2017). Comprehension of synthetic speech and digitized natural speech by adults with aphasia. *Journal of Communication Disorders*, *69*, 15–26. https://doi.org/10.1016/j.jcomdis.2017.06.006
Hux, K., Wallace, S. E., Brown, J. A., & Knollman-Porter, K. (2021). Perceptions of people with aphasia about supporting reading with text-to-speech technology: A convergent mixed methods study. *Journal of Communication Disorders*, *91*(September 2020), 106098. https://doi.org/10.1016/j.jcomdis.2021.106098
Jackson, N. E., & Coltheart, M. (2013). *Routes to reading success and failure: Toward an integrated cognitive psychology of atypical reading*. Psychology Press.
Jane Moore, M., & Demeyere, N. (2017). Neglect Dyslexia in Relation to Unilateral Visuospatial Neglect: A Review. *AIMS Neuroscience*, *4*(4), 148–168. https://doi.org/10.3934/Neuroscience.2017.4.148
Jenkins, J. R., Fuchs, L. S., Van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of Individual Differences in Reading Comprehension and Reading Fluency. *Journal of Educational Psychology*, *95*(4), 719–729. https://doi.org/10.1037/0022-0663.95.4.719
Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children*, *59*(5), 421–432.
Jonnalagadda, S., & Gonzalez, G. (2010). BioSimplify: an open source sentence simplification engine to improve recall in automatic biomedical information extraction. *AMIA Annual Symposium Proceedings*, *2010*, 351.
Kauchak, D. (2013). Improving text simplification language modeling using unsimplified text data. *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, *1*, 1537–1546.
Kay, J., Lesser, R., & Coltheart, M. (1996). Psycholinguistic assessments of language processing in aphasia (PALPA): An introduction. *Aphasiology*, *10*(2), 159–180.
Kelly, H., Kennedy, F., Britton, H., McGuire, G., & Law, J. (2016). Narrowing the “digital divide”—facilitating access to computer technology to enhance the lives of those with aphasia: a feasibility study. *Aphasiology*. https://doi.org/10.1080/02687038.2015.1077926
Kemmerer, D. (2014). Cognitive neuroscience of language. In *Cognitive Neuroscience of Language*. https://doi.org/10.4324/9781315764061
Kertesz, A. (1982). *Western Aphasia Battery. Orlando, FL: Grune & Stratton*. Inc.
Kessels, R. P. C., Van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & De Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8
Khan, R. A., & Chitode, J. S. (2016). Concatenative Speech Synthesis: A Review. *International Journal of Computer Applications*, *136*(3), 1–6. https://doi.org/10.5120/ijca2016907992
King, S. (2011). An introduction to statistical parametric speech synthesis. *Sadhana - Academy Proceedings in Engineering Sciences*, *36*(5), 837–852. https://doi.org/10.1007/s12046-011-0048-y
Kintsch, W. (1994). Text comprehension, memory, and learning. *American*

Psychologist, *49*(4), 294–303. https://doi.org/10.1037/0003-066X.49.4.294
Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363.
Kjellén, E., Laakso, K., & Henriksson, I. (2017). Aphasia and literacy-the insider's perspective. *International Journal of Language & Communication Disorders*, *52*(5), 573–584. https://doi.org/10.1111/1460-6984.12302
Klimova, B., & Kuca, K. (2016). Speech and language impairments in dementia. *Journal of Applied Biomedicine*, *14*(2), 97–103. https://doi.org/10.1016/j.jab.2016.02.002
Knollman-Porter, K., Brown, J. A., Hux, K., Wallace, S. E., & Crittenden, A. (2022). Reading Comprehension and Processing Time When People With Aphasia Use Text-to-Speech Technology With Personalized Supports and Features. *American Journal of Speech-Language Pathology*, *31*(1), 342–358. https://doi.org/10.1044/2021_AJSLP-21-00182
Knollman-Porter, K., & Julian, S. K. (2019). Book Club Experiences, Engagement, and Reading Support Use by People With Aphasia. *American Journal of Speech-Language Pathology*, *28*(3), 1084–1098. https://doi.org/10.1044/2019_ajslp-18-0237
Knollman-Porter, K., Wallace, S. E., Brown, J. A., Hux, K., Hoagland, B. L., & Ruff, D. R. (2019). Effects of Written, Auditory, and Combined Modalities on Comprehension by People With Aphasia. *American Journal of Speech-Language Pathology*, *28*(3), 1206–1221. https://doi.org/10.1044/2019_AJSLP-19-0013
Knollman-Porter, K., Wallace, S. E., Hux, K., Brown, J., & Long, C. (2015). Reading experiences and use of supports by people with chronic aphasia. *Aphasiology*, *29*(12), 1448–1472. https://doi.org/10.1080/02687038.2015.1041093
Kons, Z., Shechtman, S., Sorin, A., Rabinovitz, C., & Hoory, R. (2019). High quality, lightweight and adaptable TTS using LPCNet. *Proc. Interspeech 2019*, 176–180. https://doi.org/10.21437/Interspeech.2019-1705
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539
Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791
Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, Issue 8, pp. 707–710). Soviet Union.
Levy, J., Hoover, E., Waters, G., Kiran, S., Caplan, D., Berardino, A., & Sandberg, C. (2012). Effects of Syntactic Complexity, Semantic Reversibility, and Explicitness on Discourse Comprehension in Persons with Aphasia and in Healthy Controls. *American Journal of Speech-Language Pathology*, *21*(2), S154-65. https://doi.org/10.1044/1058-0360(2012/11-0104)PWA
Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1424.
Lynch, K. E., Damico, J. S., Abendroth, K. J., & Nelson, R. L. (2013). Reading performance subsequent to aphasia: Strategies applied during authentic reading. *Aphasiology*, *27*(6), 723–739. https://doi.org/10.1080/02687038.2012.748182
Maas, M. B., Lev, M. H., Ay, H., Singhal, A. B., Greer, D. M., Smith, W. S., Harris, G. J., Halpern, E. F., Koroshetz, W. J., & Furie, K. L. (2012). The prognosis for aphasia in stroke. *Journal of Stroke and Cerebrovascular Diseases*, *21*(5), 350–357.

Maher, L. M. (2017). Broca's Aphasia and Grammatical Processing. *The Oxford Handbook of Aphasia and Language Disorders*, 145–160.
Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research*, *2*(3), 175–199.
Matausch, K., & Peböck, B. (2010). Easyweb–a study how people with specific learning difficulties can be supported on using the internet. *International Conference on Computers for Handicapped Persons*, 641–648.
Mayer, J. F., & Murray, L. L. (2012). Measuring working memory deficits in aphasia. *Journal of Communication Disorders*, *45*(5), 325–339. https://doi.org/10.1016/j.jcomdis.2012.06.002
McAllister, T., Bachrach, A., Waters, G., Michaud, J., & Caplan, D. (2009). Production and comprehension of unaccusatives in aphasia. In *Aphasiology* (Vol. 23, Issues 7–8). https://doi.org/10.1080/02687030802669518
McCormick, C., & Samuels, S. J. (1979). Word recognition by second graders: The unit of perception and interrelationships among accuracy, latency, and comprehension. *Journal of Reading Behavior*, *11*(2), 107–118.
McDonald, S., Togher, L., & Code, C. (2013). Social and communication disorders following traumatic brain injury. In *Social and Communication Disorders Following Traumatic Brain Injury*. https://doi.org/10.4324/9780203557198
McNeil, M. R., Odell, K., & Tseng, C.-H. (1991). Toward the integration of resource allocation into a general theory of aphasia. *Clinical Aphasiology*, *20*, 21–39.
McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*(2), 302.
Mendelson, J., Oplustil, P., Watts, O., & King, S. (2017). Nativization of Foreign Names in TTS for Automatic Reading of World News in Swahili. *Interspeech 2017*, *2017-Augus*, 2188–2192. https://doi.org/10.21437/Interspeech.2017-1398
Mendez, M. F. (2012). Language. In D. B. Arciniegas, C. A. Anderson, & C. M. Filley (Eds.), *Behavioral Neurology & Neuropsychiatry* (pp. 174–183). Cambridge University Press. https://doi.org/10.1017/CBO9781139016919.014
Meteyard, L., Bruce, C., Edmundson, A., & Oakhill, J. (2015). Profiling text comprehension impairments in aphasia. *Aphasiology*, *29*(1), 1–28. https://doi.org/10.1080/02687038.2014.955388
Metter, E. J., Hanson, W. R., Jackson, C. A., Kempler, D., Van Lancker, D., Mazziotta, J. C., & Phelps, M. E. (1990). Temporoparietal cortex in aphasia: Evidence from positron emission tomography. *Archives of Neurology*, *47*(11), 1235–1238.
Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, *3*(4), 235–244.
Mittag, G., & Möller, S. (2020). Deep Learning Based Assessment of Synthetic Speech Naturalness. *Interspeech 2020*, *2020-Octob*, 1748–1752. https://doi.org/10.21437/Interspeech.2020-2382
Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, *94*(2), 249–259. https://doi.org/10.1037/0022-0663.94.2.249
Moore, M. J., & Demeyere, N. (2017). Neglect dyslexia in relation to unilateral visuospatial neglect: A review. *AIMS Neuroscience*, 4(4)
Moore, M. J., & Demeyere, N. (2019). Neglect dyslexia as a word-centred impairment: A single case study. *Cortex*, *119*, 543–554. https://doi.org/10.1016/j.cortex.2018.10.024

Morton, J., & Patterson, K. (1980). A new attempt at an interpretation, or an attempt at a new interpretation. In J. Morton, K. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91–118). Routledge & Kegan Paul.

Mu, Z., Yang, X., & Dong, Y. (2021). *Review of end-to-end speech synthesis technology based on deep learning*. http://arxiv.org/abs/2104.09995

Murray, L. L. (1999). Review Attention and aphasia: theory, research and clinical implications. *Aphasiology*, *13*(2), 91–111. https://doi.org/10.1080/026870399402226

Nakase-Thompson, R, Manning, E., Sherer, M., Yablon, S. A., Vickery, C., Harris, C., & Dickson, S. (2002). Bedside screen of language disturbance among acute care admissions: Initial psychometrics of the Mississippi Aphasia Screening Test. *Archives of Clinical Neuropsychology*, *17*(8), 848.

Nakase-Thompson, Risa, Manning, E., Sherer, M., Yablon, S. A., Gontkovsky, S. L. T., & Vickery, C. (2005). Brief assessment of severe language impairments: initial validation of the Mississippi aphasia screening test. *Brain Injury*, *19*(9), 685–691.

Nisioi, S., Štajner, S., Ponzetto, S. P., & Dinu, L. P. (2017). Exploring neural text simplification models. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 85–91.

Orasan, C., Evans, R., & Dornescu, I. (2013). Towards multilingual europe 2020: A romanian perspective, chapter text simplification for people with autistic spectrum disorders. *Sals: Do They Exist*, 75–81.

Paetzold, G. H., & Specia, L. (2016). Benchmarking Lexical Simplification systems. *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 3074–3080.

Paetzold, G., & Specia, L. (2016). Unsupervised lexical simplification for non-native speakers. *Proceedings of the AAAI Conference on Artificial Intelligence*, *30*(1).

Park, K., & Mulc, T. (2019). CSS10: A collection of single speaker speech datasets for 10 languages. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2019-Septe*, 1566–1570. https://doi.org/10.21437/Interspeech.2019-1500

Parr, S. (1995). Everyday reading and writing in aphasia: Role change and the influence of pre-morbid literacy practice. *Aphasiology*, *9*(3), 223–238. https://doi.org/10.1080/02687039508248197

Patterson, K, Marshall, J. C., & Coltheart, M. (1985). *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading*. Erlbaum. https://books.google.it/books?id=3xtxmgEACAAJ

Patterson, Karalyn, & Kay, J. (1982). by-letter reading: Psychological descriptions of a neurologial syndrome. *The Quarterly Journal of Experimental Psychology Section A*, *34*(3), 411–441.

Patterson, Karalyn, Marshall, J. C., & Coltheart, M. (2017). *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (Vol. 8). Routledge.

Pedersen, P. M., Stig Jørgensen, H., Nakayama, H., Raaschou, H. O., & Olsen, T. S. (1995). Aphasia in acute stroke: incidence, determinants, and recovery. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *38*(4), 659–666.

Perfetti, C. A. (1985). *Reading ability.* oxford university Press.

Perfetti, C. A. (1992). *The representation problem in reading acquisition.*

Perfetti, C. A. (2000). Comprehending written language: a blueprint of the reader. In *The Neurocognition of Language* (Vol. 167, pp. 167–208). Oxford University

Press. https://doi.org/10.1093/acprof:oso/9780198507932.003.0006
Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*, *67*(4), 461.
Poeppel, D. (2012). The maps problem and the mapping problem: two challenges for a cognitive neuroscience of speech and language. *Cognitive Neuropsychology*, *29*(1–2), 34–55.
Pratap, V., Hannun, A., Xu, Q., Cai, J., Kahn, J., Synnaeve, G., Liptchinsky, V., & Collobert, R. (2019). Wav2Letter++: A Fast Open-source Speech Recognition System. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2019-May*, 6460–6464. https://doi.org/10.1109/ICASSP.2019.8683535
Pratap, V., Xu, Q., Sriram, A., Synnaeve, G., & Collobert, R. (2020). MLS: A large-scale multilingual dataset for speech research. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2020-Octob*, 2757–2761. https://doi.org/10.21437/Interspeech.2020-2826
Price, C. J., & Humphreys, G. W. (1995). Contrasting effects of letter-spacing in alexia: further evidence that different strategies generate word length effects in reading. *The Quarterly Journal of Experimental Psychology*, *48*(3), 573–597.
Quinlan, P. T. (1992). *The Oxford psycholinguistic database*. Oxford University Press.
Rapp, B., & Caramazza, A. (1997). From graphemes to abstract letter shapes: levels of representation in written spelling. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(4), 1130.
Reiff Cherney, L. (2004). Aphasia, Alexia, and Oral Reading. *Topics in Stroke Rehabilitation*, *11*(1), 22–36. https://doi.org/10.1310/VUPX-WDX7-J1EU-00TB
Rello, L., & Baeza-Yates, R. (2014). Evaluation of DysWebxia. *Proceedings of the 11th Web for All Conference on - W4A ’14*, 1–10. https://doi.org/10.1145/2596695.2596697
Rello, L., Baeza-Yates, R., Bott, S., & Saggion, H. (2013). Simplify or help? Text simplification strategies for people with dyslexia. *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, 1–10.
Rose, T. A., Worrall, L. E., Hickson, L. M., & Hoffmann, T. C. (2011). Aphasia friendly written health information: Content and design characteristics. *International Journal of Speech-Language Pathology*, *13*(4), 335–347. https://doi.org/10.3109/17549507.2011.560396
Rose, T. A., Worrall, L. E., & McKenna, K. T. (2003). The effectiveness of aphasia-friendly principles for printed health education materials for people with aphasia following stroke. *Aphasiology*, *17*(10), 947–963. https://doi.org/10.1080/02687030344000319
Saffran, E. M. (1996). “Attentional Dyslexia” in Alzheimer’s Disease: A Case Study. *Cognitive Neuropsychology*, *13*(2), 205–228. https://doi.org/10.1080/026432996382006
Saggion, H. (2017). Automatic Text Simplification. In Graeme Hirst (Ed.), *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers. http://www.morganclaypool.com/doi/10.2200/S00700ED1V01Y201602HLT032
Saggion, H., Štajner, S., Bott, S., Mille, S., Rello, L., & Drndarevic, B. (2015). Making it Simplext: Implementation and evaluation of a text simplification system for Spanish. *ACM Transactions on Accessible Computing*, *6*(4). https://doi.org/10.1145/2738046
Sanders, W. R., Gramlich, C., & Levine, A. (1981). Naturalness of Synthesized Speech.

Suppes-Corpus.Stanford.Edu, 487–502. http://suppes-corpus.stanford.edu/pdfs/CAI/II-8.pdf

Sanders, W. R., Gramlich, C., & Levinef, A. (1968). Naturalness of synthesized speech. *In P. Suppes (Ed.), University-Level*.

Scherer, M. J. (2005). Assessing the benefits of using assistive technologies and other supports for thinking, remembering and learning. *Disability and Rehabilitation*, *27*(13), 731–739. https://doi.org/10.1080/09638280400014816

Schreiber, F. J. (2006). Metacognition and self-regulation in literacy. In *Metacognition in literacy learning* (pp. 237–262). Routledge.

Schwartz, M., Saffran, E., Marin, O. S. M., Coltheart, M., Patterson, K. E., & Marshall, J. C. (1980). Fractionating the reading process in dementia: A case study. *Deep Dyslexia. London: Routledge and Kegan Paul*.

Shapiro, L P, Zurif, E. B., & Grimshaw, J. (1989). Verb processing during sentence comprehension: contextual impenetrability. *Journal of Psycholinguistic Research*, *18*(2), 223–243. https://doi.org/10.1007/BF01067783

Shapiro, Lewis P, & Levine, B. A. (1990). Verb processing during sentence comprehension in aphasia. *Brain and Language*, *38*(1), 21–47. https://doi.org/10.1016/0093-934X(90)90100-U

Sheang, K. C. (2020). Context-aware automatic text simplification. *CEUR Workshop Proceedings*, *2633*(Sepln), 56–62.

Siddharthan, A. (2002). An architecture for a text simplification system. *Language Engineering Conference, 2002. Proceedings*, 64–71.

Siddharthan, A. (2011). Text simplification using typed dependencies: A comparision of the robustness of different generation strategies. *Proceedings of the 13th European Workshop on Natural Language Generation*, 2–11.

Siddharthan, A., Nenkova, A., & McKeown, K. (2004). Syntactic simplification for improving content selection in multi-document summarization. *Proceedings of the 20th International Conference on Computational Linguistics - COLING '04*, 896-es. https://doi.org/10.3115/1220355.1220484

Sitaram, S., & Black, A. W. (2016). Speech synthesis of code-mixed text. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 3422–3428.

Sitaram, S., Chandu, K. R., Rallabandi, S. K., & Black, A. W. (2019). *A Survey of Code-switched Speech and Language Processing*. http://arxiv.org/abs/1904.00784

Slyh, R. E., Nelson, W. T., & Hansen, E. G. (1999). Analysis of mrate, shimmer, jitter, and F/sub 0/ contour features across stress and speaking style in the SUSAS database. *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, *4*, 2091–2094 vol.4. https://doi.org/10.1109/ICASSP.1999.758345

Small, S. L., & Hickok, G. (2016). Neurobiology of language. In S. S. Greg Hickok (Ed.), *Neurobiology of language*. Academic press.

Smith, K. G., & Clark, K. F. (2019). Error analysis of oral paragraph reading in individuals with aphasia. *Aphasiology*, *33*(2), 234–252. https://doi.org/10.1080/02687038.2018.1545992

Sohlberg, M., Griffiths, G., & Fickas, S. (2014). The effect of electronically delivered strategies on reading after mild-moderate acquired brain injury. *American Journal of Speech-Language Pathology*, *23*, 160–175.

Soni, M. H., & Patil, H. (2016). Non-intrusive Quality Assessment of Synthesized Speech using Spectral Features and Support Vector Regression. *SSW*.

Specia, L. (2010). *Translating from Complex to Simplified Sentences BT -*

Computational Processing of the Portuguese Language (T. A. S. Pardo, A. Branco, A. Klautau, R. Vieira, & V. L. S. de Lima (eds.); pp. 30–39). Springer Berlin Heidelberg.
Štajner, S., Béchara, H., & Saggion, H. (2015). A deeper exploration of the standard PB-SMT approach to text simplification and its evaluation. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 823–828.
Štajner, S., Calixto, I., & Saggion, H. (2015). Automatic text simplification for Spanish: Comparative evaluation of various simplification strategies. *International Conference Recent Advances in Natural Language Processing, RANLP*, *2015-Janua*, 618–626.
Štajner, S., Franco-Salvador, M., Rosso, P., & Ponzetto, S. P. (2019). CATS: A tool for customized alignment of text simplification corpora. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 3895–3903. https://aclanthology.org/L18-1615
Štajner, S., & Popović, M. (2016). Can text simplification help machine translation? *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, 230–242.
Štajner, S., & Saggion, H. (2018). Data-driven text simplification. *Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts*, 19–23.
Štajner, S., Saggion, H., & Ponzetto, S. P. (2019). Improving lexical coverage of text simplification systems for Spanish. *Expert Systems with Applications*, *118*, 80–91. https://doi.org/10.1016/j.eswa.2018.08.034
Stefaniak, J. D., Halai, A. D., & Lambon Ralph, M. A. (2020). The neural and neurocomputational bases of recovery from post-stroke aphasia. *Nature Reviews Neurology*, *16*(1), 43–55. https://doi.org/10.1038/s41582-019-0282-1
Szabo, G., & Dittelman, J. (2014). Using Mobile Technology with Individuals with Aphasia: Native iPad Features and Everyday Apps. *Seminars in Speech and Language*, *35*(1), 5–16. https://doi.org/10.1055/s-0033-1362993
Tachibana, H., Uenoyama, K., & Aihara, S. (2018). Efficiently trainable text-to-speech system based on deep convolutional networks with guided attention. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4784–4788.
Tainturier, M.-J., & Rapp, B. (2001). The spelling process. *The Handbook of Cognitive Neuropsychology: What Deficits Reveal about the Human Mind*, 263–289.
Teixeira, J. P., & Gonçalves, A. (2014). Accuracy of Jitter and Shimmer Measurements. *Procedia Technology*, *16*, 1190–1199. https://doi.org/10.1016/j.protcy.2014.10.134
Teixeira, J. P., Oliveira, C., & Lopes, C. (2013). Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. *Procedia Technology*, *9*, 1112–1122. https://doi.org/10.1016/j.protcy.2013.12.124
Thomas, S., Rao, M. N., Murthy, H. A., & Ramalingam, C. S. (2006). Natural sounding TTS based on syllable-like units. *European Signal Processing Conference*, *May 2012*.
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
Vallar, G., Burani, C., & Arduino, L. S. (2010). Neglect dyslexia: a review of the neuropsychological literature. *Experimental Brain Research*, *206*(2), 219–235.

https://doi.org/10.1007/s00221-010-2386-0
Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*.
Wallace, S. E., Hux, K., Knollman-Porter, K., Brown, J. A., Parisi, E., & Cain, R. (2021). Reading behaviors and text-to-speech technology perceptions of people with aphasia. *Assistive Technology*, *00*(00), 1–12. https://doi.org/10.1080/10400435.2021.1904306
Wallace, S. E., Knollman-Porter, K., Brown, J. A., & Hux, K. (2019). Narrative comprehension by people with aphasia given single versus combined modality presentation. *Aphasiology*, *33*(6), 731–754. https://doi.org/10.1080/02687038.2018.1506088
Wang, T., Chen, P., Amaral, K., & Qiang, J. (2016). An experimental study of LSTM encoder-decoder model for text simplification. *ArXiv Preprint ArXiv:1609.03663*.
Warrington, E. K., & Shallice, T. I. M. (1980). Word-form dyslexia. *Brain: A Journal of Neurology*, *103*(1), 99–112.
Watanabe, W. M., Junior, A. C., Uzêda, V. R., Fortes, R. P. de M., Pardo, T. A. S., & Aluísio, S. M. (2009). Facilita: reading assistance for low-literacy readers. *Proceedings of the 27th ACM International Conference on Design of Communication*, 29–36.
Webb, W. G., & Love, R. J. (1983). Reading problems in chronic aphasia. *Journal of Speech and Hearing Disorders*, *48*(2), 164–171.
Webster, J., Morris, J., Howard, D., & Garraffa, M. (2018). Reading for Meaning: What Influences Paragraph Understanding in Aphasia? *American Journal of Speech-Language Pathology*, *27*(1S), 423–437. https://doi.org/10.1044/2017_AJSLP-16-0213
Webster, J., Morris, J., Malone, J., & Howard, D. (2021). Reading comprehension difficulties in people with aphasia: investigating personal perception of reading ability, practice, and difficulties. *Aphasiology*, *35*(6), 805–823. https://doi.org/10.1080/02687038.2020.1737316
Weiller, C., Bormann, T., Saur, D., Musso, M., & Rijntjes, M. (2011). How the ventral pathway got lost–And what its recovery might mean. *Brain and Language*, *118*(1–2), 29–39.
Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Cohn.
Wernicke, C. (1881). *Lehrbuch der gehirnkrankheiten für aerzte und studirende* (Vol. 1). Fischer.
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.
Woodsend, K., & Lapata, M. (2011). Learning to simplify sentences with quasi-synchronous grammar and integer programming. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 409–420.
Woollams, A. M. (2014). Connectionist neuropsychology: Uncovering ultimate causes of acquired dyslexia. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634). https://doi.org/10.1098/rstb.2012.0398
Worrall, L., Rose, T., Howe, T., Brennan, A., Egan, J., Oxenham, D., & McKenna, K. (2005). Access to written information for people with aphasia. *Aphasiology*, *19*(10–11), 923–929. https://doi.org/10.1080/02687030544000137
Wubben, S., Van Den Bosch, A., & Krahmer, E. (2012). Sentence simplification by monolingual machine translation. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1015–1024.

Xu, W., Napoles, C., Pavlick, E., Chen, Q., & Callison-Burch, C. (2016). Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, *4*, 401–415.

Yagata, S. A., Yen, M., McCarron, A., Bautista, A., Lamair-Orosco, G., & Wilson, S. M. (2017). Rapid recovery from aphasia after infarction of Wernicke's area. *Aphasiology*, *31*(8), 951–980.

Yin, Z. (2018). An Overview of Speech Synthesis Technology. *2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)*, *7*, 522–526. https://doi.org/10.1109/IMCCC.2018.00116

Zen, H., Senior, A., & Schuster, M. (2013). Statistical parametric speech synthesis using deep neural networks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 7962–7966. https://doi.org/10.1109/ICASSP.2013.6639215

Zen, H., Tokuda, K., & Black, A. W. (2009). Statistical parametric speech synthesis. *Speech Communication*, *51*(11), 1039–1064. https://doi.org/10.1016/j.specom.2009.04.004

Zhao, S., Meng, R., He, D., Saptono, A., & Parmanto, B. (2018). Integrating Transformer and Paraphrase Rules for Sentence Simplification. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3164–3173. https://doi.org/10.18653/v1/D18-1355

Zhu, Z., Bernhard, D., & Gurevych, I. (2010). A monolingual tree-based translation model for sentence simplification. *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 1353–1361.

Zwetsch, I. C., Fagundes, R. D. R., Russomano, T., & Scolari, D. (2006). Digital signal processing in the differential diagnosis of benign larynx diseases [Abstract in English]. *Scientia Medica*, *16*(3), 109–114.

Appendix

Appendix A: Example of the questionnaire used during the first meeting with the patients volunteering for the pilot study. The questionnaire was done in the Spanish language. Here we report it translated into English.

1. What do you think is your difficulty?

- In speaking
- To understand
- Both

If the difficulty is in production, proceed with the following questions:

2. How do you feel about being helped to speak by a machine?

- I am favourable
- No, I don’t like it
- I am willing to try it

3. Do you ever have the impression that others do not understand if you are, for example, angry or sad or surprised when you say something? Yes/no

If the patient answered “yes” to question n 3, we asked:
“would you accept better a machine if it can help you convey your emotions?”

If the difficulty is in text comprehension, proceed with the following questions:

4. Do you also have difficulties in understanding texts? Yes/no

If the participant answered yes to this question, we asked the next two questions:

5. Your difficulties are because:

- The text is too long, and you lose attention
- The text is too difficult (for instance, you don’t know some words)

6. Would you like to be helped to comprehend a text by a machine? Yes/no

Appendix B: In the next pages we show the template of the informed consent approved by the CIREP and subsequently signed by the participants of the experiments.

El Consentimiento Informado

HOJA DE INFORMACIÓN

La investigación en la que está a punto de participar tiene el siguiente título: "Adaptación de sistemas de síntesis de voz para personas con afasia".

Supervisores del proyecto:
Dra. Mireia Farrús, correo electrónico: mireia.farrus@upf.edu, mfarrus@ub.edu
Dra. Ineke van der Meulen, correo electrónico: ivdmeulen@rijndam.nl

Estudiante de doctorado: Giorgia Cistola, correo electrónico: giorgia.cistola@upf.edu

Universitat Pompeu Fabra, grupo de Comunicación y Tecnologías de la Información, TALN (Tractament Automàtic del Llenguatge Natural).

El objetivo de este proyecto es investigar si es posible **mejorar** la forma en que las **personas con afasia acceden** al **material escrito.** Para alcanzar este objetivo, recurriremos a dos herramientas de alta tecnología. Uno convierte el texto en voz y el otro simplifica el texto para hacerlo más comprensible, reduciendo la longitud de las frases y transformando las palabras difíciles en más fáciles. De esta manera, pretendemos dar la posibilidad de **acceder de forma autónoma** al contenido de **los textos.**

El experimento tendrá lugar en una sala del Hospital de Sant Pau y tendrá una **duración** estimada de **20 minutos**. Usted ejecutará el estudio con un ordenador portátil proporcionado por el experimentador. El experimento se llevará a cabo en al menos dos sesiones diferentes. Sin embargo, podemos pedirle que participe en un máximo de dos sesiones más, dependiendo de los resultados de la primera ronda.

El experimento está **compuesto** de **dos partes**:

1) En la **primera** parte, su participación consistirá en escuchar 60 grabaciones. Inmediatamente después de finalizar la cada grabación, aparecerán cuatro imágenes en la pantalla del ordenador. En este punto, debe **indicar cuál de las cuatro imágenes** es la **descrita por la grabación que ha escuchado**. Solamente una de las imágenes es la correcta. No hay un límite de tiempo para indicar la respuesta correcta; **puede tomarse su tiempo** para responder.

2) En la **segunda** parte del estudio se escuchará una frase corta. Después, se le pedirá que califique la claridad de la voz, la

naturalidad y la facilidad de comprensión de la voz. Aquí hay un ejemplo de cómo es el gráfico que se le pedirá que complete:

Con la participación en este experimento, usted **no tendrá ningún beneficio ni daño directo**. No habrá ninguna compensación.
Usted puede decidir por sí mismo si participar o no en el estudio o si participar y renunciar en una fase posterior. **Ninguna** de estas condiciones **afectará el tratamiento** o **la atención** que esté recibiendo **en el hospital**.
Se informará tanto a los médicos como a los cuidadores sobre el estudio y su importancia. Asimismo, se les informará de que la participación o la falta de voluntad para participar no es una ventaja o una desventaja. Se informará a los cuidadores de que **no hay consecuencias de ningún tipo en el tratamiento** (positivas o negativas) de la participación en el estudio.

Para proteger su privacidad, **no lo identificaremos** por su **nombre**, **sino por un código** que **sólo** será **conocido** por los **investigadores** actuales del proyecto. La lista de coincidencias de códigos y nombres se mantendrá separada de los códigos que identifican sus respuestas. Sus datos serán almacenados en un lugar seguro, bajo llave o digitalmente, con mecanismos seguros de control de acceso para que solo los investigadores del proyecto actual puedan consultarlos. **En el caso de publicar los datos** referentes a la investigación, **se hará de forma anónima**.

Su **participación** es **voluntaria**:

- Su participación en este proyecto es **completamente gratuita**;
- Usted **puede retirarse u** optar por **no participar** en él en **cualquier momento sin** necesidad de **justificar** su decisión;
- Usted tiene el **derecho de omitir respuestas** a cualquier pregunta;
- No habrá **ningún** tipo **de penalización o respuesta** si decide **abandonar** el experimento.
- Abandonar o no participar en el experimento **no tendrá ningún tipo de impacto** en **su cuidado** en el hospital.

*Si tiene alguna pregunta sobre el proyecto de investigación, **puede ponerse en contacto** con:*

Giorgia Cistola: giorgia.cistola@upf.edu o 722 417 522

De acuerdo con lo dispuesto en el Reglamento General de Protección de Datos, Reglamento (UE) 2016/679, resumimos la información de protección de datos:

Responsable del tratamiento*: Universitat Pompeu Fabra. Pl. de la Mercé, 12. 08002 Barcelona. Tel. (+34) 935 422 000. Puede contactar con el Delegado de protección de datos de la UPF mediante la dirección dpd@upf.edu*
Finalidad*: Sus datos se conservarán durante el período de realización del proyecto y dos años más para su validación científica[7].*
Legitimación*: consentimiento de la persona interesada. Se puede retirar en cualquier momento.*
Destinatarios: *sus datos serán utilizados únicamente por la Universidad Pompeu Fabra, y no se cederán a terceros sin su consentimiento, excepto en los casos previstos por la ley.*
Derechos*: acceso a sus datos, rectificación, supresión, solicitud de portabilidad, oposición al tratamiento y solicitud de limitación mediante correo electrónico a gerencia@upf.edu. Si considera que sus derechos no han sido debidamente atendidos puede presentar reclamación ante la Autoridad Catalana de Protección de Datos.*

CONSENTIMIENTO INFORMADO

Título de la investigación: “Adaptación de síntesis de voz para personas con afasia”

Supervisor del proyecto de investigación: Dra. Mireia Farrús, correo electrónico: mireia.farrus@upf.edu

Estudiante de doctorado: Giorgia Cistola, correo electrónico: giorgia.cistola@upf.edu

Universitat Pompeu Fabra, grupo de Comunicación y Tecnologías de la Información, TALN (Tractament Automàtic del Llenguatge Natural).

CONFIRMO que:
- he leído la hoja de información del proyecto de investigación,
- he podido realizar preguntas sobre el proyecto,
- he recibido suficiente información sobre el proyecto.

ENTIENDO que mi participación es voluntaria y que puedo retirarme del estudio en cualquier momento sin tener que justificar mi decisión.

DOY MI CONSENTIMIENTO a participar en este proyecto.

Nombre y apellidos:
Firma:
Lugar y fecha