Usability of Partially Localised Websites in Switzerland: A Study with Screen Reader Users

Casalegno, Elisa

Elisa Casalegno

Usability of Partially Localised Websites in Switzerland: A Study with Screen Reader Users

Directrice : Silvia Rodríguez Vázquez
Jurée : Lucía Morado Vázquez

Mémoire présenté à la Faculté de traduction et d'interprétation (Département de Traitement Informatique Multilingue) pour l'obtention de la Maîtrise universitaire en traduction, mention Technologies de la traduction.

Université de Genève
Année Académique 2017-2018
Août 2018

Déclaration attestant le caractère original du travail effectué

J'affirme avoir pris connaissance des documents d'information et de prévention du plagiat émis par l'Université de Genève et la Faculté de traduction et d'interprétation (notamment la *Directive en matière de plagiat des étudiant-e-s,* le *Règlement d'études de la Faculté de traduction et d'interprétation* ainsi que l'*Aide-mémoire à l'intention des étudiants préparant un mémoire de Ma en traduction*).

J'atteste que ce travail est le fruit d'un travail personnel et a été rédigé de manière autonome.

Je déclare que toutes les sources d'information utilisées sont citées de manière complète et précise, y compris les sources sur Internet.

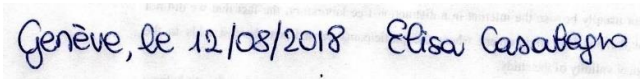
Je suis consciente que le fait de ne pas citer une source ou de ne pas la citer correctement est constitutif de plagiat et que le plagiat est considéré comme une faute grave au sein de l'Université, passible de sanctions.

Au vu de ce qui précède, je déclare sur l'honneur que le présent travail est original.

Nom et prénom :

CASALEGNO ELISA

Lieu / date / signature :



Acknowledgements

This master's thesis would not have been possible without the help and constant support of a number of people. First of all, I would like to thank Lucía Morado Vázquez for introducing me to the fascinating field of localisation, advising me during the first step of this research and accepting to be my second reader. A big thanks goes to Silvia Rodríguez Vázquez, who has guided me throughout the process, for her precious feedback and knowledge; without your kind encouragement and optimism, I would still be choosing my topic. Also, a special thanks to Dónal, for testing the accessibility of my recruitment form and to Tiffany and John for their invaluable help proofreading my English. I take full responsibility for any mistakes I may have introduced after their excellent work.

Merci beaucoup aux utilisateurs francophones pour leur participation et leur précieux commentaires, ainsi qu'à Maximilien Thilo et Corinne Doret Baertschi pour leur disponibilité. Je souhaite également remercier mes collègues de stage : le team de Production et les traducteurs du Secrétariat Général du Département fédéral de l'intérieur. Merci pour votre accueil et votre flexibilité, qui m'ont permis de concilier école et travail. Merci aux professeurs du Département de traitement informatique multilingue, spécialement Mme Bouillon et Mme Starlander, pour m'avoir montré un autre côté de la traduction, à mes collègues monitrices et à M Sossauer. Merci également à Emma, Jasmina et Caroline pour avoir relu mon français.

Vielen Dank an die deutschsprachigen Teilnehmer meiner Studie, sowie an Markus Riesch, Anton Bolfing und Alexander Wyssmann: Ihre Hilfe wurde sehr geschätzt. Ein herzlicher Dank geht an Alex und Philipp, die mich mit meinem Deutsch unterstützt haben, und Nadia, die meine Arbeit überprüft hat.

Grazie infinite ai partecipanti italofoni, difficili da scovare ma proprio per questo preziosissimi, a Denise, per avermi messo in contatto con uno di loro e a Tommaso, per aver accettato di testare la metodologia. Vorrei anche ringraziare Michelle e Marta per aver riletto i questionari in italiano e Ludovica, per il supporto logistico a Zurigo.

Un ringraziamento speciale va a tutte le persone che hanno contribuito, direttamente o indirettamente, a non farmi perdere la bussola in questi mesi. I miei amici italiani, lontani dagli occhi ma vicini al cuore, con le lunghe chiacchierate per aggiornarci a vicenda sugli ultimi avvenimenti. I miei amici ginevrini e quelli incontrati in Erasmus, che ormai sono sparsi per il mondo, ma che resteranno sempre i miei compagni d'avventura. I miei amici bernesi, coinquilini inclusi, che hanno reso la transizione più piacevole. Caroline, su cui posso sempre contare. Sabrina, per i consigli e il buon umore. Marta, per i finesettimana passati a scrivere, e perché se non ci fosse bisognerebbe inventarla. John, per la pazienza e l'incoraggiamento. E infine loro, che ci sono stati fin dal primo giorno. Senza l'affetto e il supporto della mia famiglia non sarei arrivata fino a qui.

Dedico questa tesi a nonna Lucia, nonno Tere, Daniela e Marco.

“Differences of habit and language are nothing at all if our aims are identical and our hearts are open.”
J.K. Rowling

Abstract

Equal opportunities for people with disabilities and, specifically, access to information technology and to education play a key role in the United Nations Convention on the Rights of Persons with Disabilities (CRPD). Given the increasing importance of the Internet in our everyday life, assistive technology (AT) is crucial for the inclusion of disabled people in society, as it allows them to access the Web.

A usability study involving ten visually impaired users was conducted on the websites of the Zurich University of Applied Sciences and the Bern University of Applied Sciences. Both websites were partially localised into French, English and, one of them, into Italian, but to different degrees: the first one had been reorganised, while the second one offered a mirror structure. The participants' native language was either German, French, or Italian and they were all experienced screen reader users. The tests were conducted in a real-life setting and were audio-taped by the researcher, who also took notes and interacted with the participants to guide them through the tasks.

The main goals were to observe the interaction of multilingual screen reader users with partially localised websites and to evaluate the impact of language selection by the user and the implementation of different localisation strategies on the overall usability of the test websites. The quantitative and qualitative data collected through the tests suggest that screen reader users generally prefer to browse in their native tongue and that they are used to find a way to change the language when needed. Users who browse the localised versions encounter more usability problems that are more often related to language than users who browse the original version. Finally, neither localisation degree proved satisfactory for screen reader users, even though the website that offered a mirror structure scored slightly better than the reorganised one.

In conclusion, we cannot recommend one of the studied localisation strategies over the other, but we have gathered evidence that show how localisation issues can affect the perceived usability of screen reader users.

Keywords: web usability, web accessibility, screen reader users, localisation levels, multilingual websites, university websites.

Contents

List of Tables

List of Figures

List of Abbreviations

ACM	Association for Computing's Machinery
ARIA	Accessible Rich Internet Applications
AT	Assistive Technology
ATAG	Authoring Tool Accessibility Guidelines
BFH	Bern University of Applied Sciences (Berner Fachhochschule)
BSI	British Standards Institute
CRPD	Convention on the Rights of Persons with Disabilitites
CSUQ	Computer System Usability Questionnaire
ETI	Enhanced Text User Interface
EWPL	Emotion Word Prompt List
FIGS	France Italy Germany and Spain
GILT	Globalisation Internationalisation Localisation Translation
GUI	Graphical Computer Interface
HCI	Human-Computer Interaction
HKB	Bern University of the Arts
ICT	Information and Communication Technology
IP	Internet Protocol
ISO	International Organisation for Standardisation
IT	Information Technology
LHAND	Federal Act on Equal Rights for People with Disabilities
LISA	Localization Industry Standards Association
SD	Standard Deviation
SIGCHI	Special Interest Group on Computer-Human Interaction
SLI	Swiss Literary Institute
SUS	System Usability Scale
TS	Translation Studies
UAAG	User Agent Accessibility Guidelines
UCD	User-Centred Design
URL	Uniform Resource Locator

USA/US United States of America

W3C World Wide Web Consortium

WAI Web Accessibility Initiaitve

WCAG Web Content Accessibility Guidelines

WHO World Health Organisation

ZHAW Zurich University of Applied Sciences (Zürcher Hochschule für Angewandte Wissenschaften)

1 Introduction

This master's thesis examines the impact of multilingualism and partial localisation strategies on the web navigation experience of screen reader users. More specifically, it investigates the navigation patterns of visually impaired users on multilingual websites, it examines whether the problems they encounter are related to language and it studies the impact of different translation strategies on the ease of use of two university websites.

1.1 Research Context

Information and communication technologies (ICT) play a key role in today's society (Access for all 2016, 3). During the last three decades, we have witnessed a continued growth of the Internet that is now present in most aspects of our lives. The Internet has given rise to the World Wide Web and both have revolutionised human communication and helped to interconnect the world (Folaron in Jiménez-Crespo 2013, 7). According to the Internet World Stats (2018), there were more than four billion Internet users, or 54% of the world population, as of December 2017[1]. To many people, the Web has become their principal source of information. As a result, not having access to the Web for any reason is increasingly becoming a serious disadvantage. There are many causes for exclusion, including economics, which is still one of the major factors, but there are also questions of physical access to such technologies. Indeed, by the time the Web was established, the graphical user interface (GUI) was standard and the normal means of access was through a keyboard, mouse and screen. In 1983, it was argued that successful human-computer interaction would require the user to have a perceptual system to receive sensory messages from the computer, a motor system to control the actions to provide input to the computer, and a cognitive system to connect the two systems and determine appropriate actions in response to the input received (Barreto 2008, 3). Since then, a variety of aid and adaptations have been developed enabling access for people who cannot use the conventional technologies because they may have difficulty typing on keyboards or pointing with a mouse due to physical impairments. Web accessibility issues also exist for people with visual disabilities because the Web has developed into a largely visual medium (Barreto 2008, 3; Edwards 2008, 142). Given the ubiquity of the GUI, "the demands placed on

[1] https://www.internetworldstats.com/stats.htm Last access:14 June 2018

the visual channel of the user's perceptual system have been raised beyond the capabilities of a significant proportion of potential Web users" (Barreto 2008, 3).

1.1.1 Visual impairment and assistive technology

Even though there are many kinds of barriers that prevent people from surfing the Web, this thesis analyses those that affect people with severe visual impairment and those who are completely blind. In 2015, a study estimated that 36 million people around the world were blind and 217 were moderately or severely visually impaired (Bourne et al. 2017). The International Statistical Classification of Diseases and Related Health Problems (2016, 10th Revision)[2] defines blindness as presenting distance visual acuity worse than 3/60 on the Snellen chart[3] in both eyes, and severe visual impairment as worse than 6/60.

The World Health Organisation (WHO) defines assistive technology (AT) as "any product, instrument, equipment or technology adapted or specially designed for improving functioning of a disabled person" (Cook and Polgar 2014, 2). In this sense, AT is used as an umbrella term for a broad range of devices, software and tools that help any person with a disability to complete a task. Generally speaking, this includes something as low tech as a crutch to help someone walk or a magnifying glass to assist with reading, or as high-tech as robotic prosthesis or image recognition software running on a tablet. It can include something as general as a browser zoom or as specific as a custom-designed game controller. It can be a separate physical device, like a braille display, or it can be fully implemented into software, like voice control. AT can also be built into an operating system, as with certain screen readers, or it can be a Web browser add-on, such as a Google Chrome extension[4].

For the purposes of this thesis we shall focus on assistive technology in the field of ICT, and specifically on those tools that help blind people surf the Web. For this reason, we will adopt the definition of the World Wide Web Consortium (W3C[5]), which is more specific and reads as follows: "hardware and/or software that acts as a user agent, or along with a mainstream user agent, to provide functionality to meet the requirements of users with disabilities that go beyond those offered by mainstream user agents." [6]

---

[2] http://apps.who.int/classifications/icd10/browse/2016/en#/H54.0 Last access: 10 June 2018

[3] https://www.britannica.com/science/Snellen-chart Last access: 9 June 2018

[4] https://de.udacity.com/course/web-accessibility--ud891 Last access: 10 June 2018

[5] The World Wide Web Consortium (W3C) is an international community that works for the development of Web standards. Led by Web inventor and Director Tim Berners-Lee and CEO Jeffrey Jaffe, W3C's mission is to lead the Web to its full potential https://www.w3.org/Consortium/ Last access: 11 July 2018

[6] https://www.w3.org/TR/WCAG21/#dfn-assistive-technologies Last access: 10 June 2018

Studies have shown that sighted users "do not read on the Web; instead they scan the pages trying to pick out a few sentences or even part of sentences to get the information they want" (Edwards 2008, 153). Blind users should have at their disposal a means to quickly scan a page to determine whether it interests them, so they can decide to read the page in detail or go elsewhere. One of the first types of assistive technology that was developed for this purpose was called the specialised browser and it was designed to render web pages in a non-visual form. A specialised browser was useful inasmuch as it could use the heading tags of a clearly structured web page to transform the method of access (ibid.). Nowadays, these programmes are no longer developed or supported. They have been replaced by standard browsers in conjunction with a screen reader, which is a piece of software that analyses the contents of the screen on a computer and turns the contents into a non-visual form, presenting it in synthetic speech or braille – or both (Edwards 2008, 154–55). Screen readers are specialised applications that read aloud text on a screen. Although most modern operating systems have built-in screen readers, commercial versions of this application are becoming increasingly popular (Cunningham 2012, 3). One of the reasons why screen readers have essentially replaced specialised browsers is that they offer, in combination with a standard browser, nearly all of the same features, with the advantage of flexibility because they make a whole range of software usable, not only browsers (Edwards 2008, 154).

1.1.2 Legal basis

Reviewing the accessibility-related legislation was also deemed relevant because it contextualises the present study, which deals with websites of public higher education institutions. This sub-section will begin with an overview of the international framework and then briefly illustrate how Europe and the United States have adapted their legislation. Finally, it will focus on Switzerland.

At an international level, the United Nations Convention on the Rights of Persons with Disabilities (CRPD) was adopted on 13 December 2006 in New York, and was opened for signature on 30 March 2007. The Convention aims at changing attitudes and approaches to persons with disabilities from "objects" of charity, medical treatment and social protection to "subjects" with rights, who are capable of claiming those rights and making decisions for their lives based on free and informed consent as well as being active members of society. It is intended as a human rights instrument with a social development dimension, adopts a broad categorisation of persons with disabilities and reaffirms that all persons with all types of disabilities must enjoy all human rights and fundamental freedoms. The Convention also

identifies (i) areas where adaptations have to be made for persons with disabilities to ensure they can effectively exercise their rights, (ii) areas where their rights have been violated and (iii) areas where protection of rights must be reinforced[7]. In terms of assistive technology, article 4 (*General obligations*) demands that States Parties

> […] undertake or promote research and development of, and […] promote the availability and use of new technologies, including information and communications technologies, mobility aids, devices and assistive technologies, suitable for persons with disabilities, giving priority to technologies at an affordable cost; […] provide accessible information to persons with disabilities about mobility aids, devices and assistive technologies, including new technologies, as well as other forms of assistance, support services and facilities; (CRPD, art. 4, sec. 1, pars. g-h)[8].

In terms of accessibility, article 9 affirms that States parties undertake to adopt immediate, effective and appropriate measures:

> To enable persons with disabilities to live independently and participate fully in all aspects of life, States Parties shall take appropriate measures to ensure to persons with disabilities access, on an equal basis with others, to the physical environment, to transportation, to information and communications, including information and communications technologies and systems, and to other facilities and services open or provided to the public, both in urban and in rural areas (CRPD, art. 9, sec. 1) [9].

The European Union ratified the CRPD in 2010, while the United States has signed but not ratified the convention[10]. In the USA, the first steps toward a Web accessibility policy were taken with the 1998 amendment to the Rehabilitation Act of 1973, or Section 508, which sets standards for accessibility of information technology (Barnum 2011, 109). These guidelines require all federal agencies to make their electronic and information technology accessible to people with disabilities and specify 16 standards that websites must meet (essentially a subset of the full Web Content Accessibility Guidelines 2.0, which will be presented in sub-section 2.4.1) (Albert and Tullis 2013, 231–32).

Switzerland signed the CRPD in 2014, committing itself not only to removing obstacles to ICT accessibility (see art. 9, sec. 1, par. b), but also to fostering measures guaranteeing accessibility at an early stage of development (see art. 9, sec. 2, par. h). Furthermore, the country should "[urge] private entities that provide services to the general public, including through the

[7] https://www.un.org/development/desa/disabilities/convention-on-the-rights-of-persons-with-disabilities.html Last access: 11 June 2018

[8] http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf Last access: 11 June 2018

[9] http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf Last access: 11 June 2018

[10] http://indicators.ohchr.org/ Last access: 11 June 2018

Internet, to provide information and services in accessible and usable formats for persons with disabilities;" (ibid.). The current Federal Constitution of the Swiss Confederation[11], in force since January 2000, states that "[n]o person may be discriminated against, in particular […] because of a physical, mental or psychological disability" (art. 8, par. 2) and that "[t]he law shall provide for the elimination of inequalities that affect persons with disabilities" (art. 8, par. 4). The ban on all discrimination, however, does not directly apply to private businesses that offer their services; these institutions are not required to make their offer accessible to people with disabilities.

Aimed at allowing disabled people to independently participate in daily life, the Federal Act on Equal Rights for People with Disabilities (LHand) was adopted in 2004[12]. While this represents a step forward in the legislation, only the State and state-owned companies must offer accessible services and adapt their websites. Federal institutions (e.g. municipalities, universities, schools, hospitals, libraries, associations, etc.) and private companies are only required to comply with the ban on discrimination. To encourage and support the fulfilment of the inclusion goals not just in the Internet offer of the governmental websites, but also of the cantons, municipalities and their institutions, an accessibility standard has been developed[13] by eCH, a Swiss association that coordinates and supports the expansion of electronic government services[14] (e-Government). The Accessibility Standard eCH-0059 is a recommendation based on the mandatory guidelines that apply to government websites (P028), based, in turn, on the World Wide Web W3C standards, and it applies to all public institutions (Access for all 2016, 51–52), including universities.

1.2 Motivation

As stated at the beginning of the present chapter, this thesis focuses on screen reader users and their navigation experience on multilingual websites. The researcher's interest in inclusion and her background in Translation Studies were the main factors behind this choice of topic.

Personal inclinations aside, there are multiple reasons to address access to the Web for people with visual impairment. First of all, morally speaking, it is the right thing to do, as the Web plays an important role and has significant benefits for people with disabilities (Theofanos and Redish 2003, 1–2). A study conducted by Theofanos and Redish (ibid.) shows that the

[11] https://www.admin.ch/opc/en/classified-compilation/19995395/index.html Last access: 11 June 2018
[12] https://www.admin.ch/opc/fr/classified-compilation/20002658/index.html Last access: 11 June 2018
[13] https://www.ech.ch/vechweb/page?p=dossier&documentNumber=eCH-0059 Last access: 21 July 2018
[14] https://www.egovernment.ch/en/ Last access: 22 July 2018

Internet has opened up a whole new world for blind users and has given them a sense of independence and freedom. Secondly, from a commercial perspective, disabled people represent a significant segment of potential customers for companies worldwide. We have seen in the previous section that the estimated number of visually impaired people is as high as 250 million[15]. According to Theofanos and Redish (ibid.), the number of people with disabilities is likely to increase, since the likelihood of having a disability increases with age and the overall population is aging. Thirdly, when improvements are made for people with disabilities, studies have shown that the user experience also improves for people without disabilities (Barnum 2011, 109). This is particularly true for the older population, people with low literacy or those who lack native language proficiency, people with low bandwidth connections or older technology and people with low Web literacy skills (ibid.). Users that are temporarily impaired can also benefit from such improvements. For example, a person who has broken her dominant arm will learn very quickly how difficult websites can be to navigate without a steady mouse; a person without headphones in a noisy environment will have trouble with websites that require sound and a person who has forgotten his glasses will be subjected to websites that do not deal with large text gracefully (Cunningham 2012, ix). Finally, equality of opportunities in education for people with disabilities and specifically access to higher education is one of the aspects advocated for by the United Nations in the above mentioned CRPD:

> States Parties shall ensure that persons with disabilities are able to access general tertiary education, vocational training, adult education and lifelong learning without discrimination and on an equal basis with others. To this end, States Parties shall ensure that reasonable accommodation is provided to persons with disabilities (CRPD, art. 24, sec. 5)[16].

This appeal has gained attention in Switzerland, as demonstrated by: the inclusion of university websites in the Access for All study (wich will be described in section 3.2), and the launch of a new Swiss project with a focus on barrier-free communication in higer education settings[17]. With this thesis we aim to make a contribution to this line of research.

In chapter *3. Related work*, numerous studies that focus on access to the web for people with visual impairments will be presented. Yet, to the best of our knowledge, it would appear that multilingualism is rarely considered by scholars. In this study, we offer our contribution to

[15] http://apps.who.int/classifications/icd10/browse/2016/en#/H54.0 Last access: 10 June 2018
[16] http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf Last access: 11 June 2018
[17] More on https://bfc.unige.ch/en/ Last access: 10 August 2016

bridging this knowledge gap by looking at the multilingual websites of a multilingual country, such as Switzerland.

1.3 Goals and research questions

In this section, we look closely at the aim of this thesis and the research questions that have guided our work. The broader objective of this research study is to explore various aspects of the relationship between usability, accessibility and localisation, whose definition can be found in the next chapter (*2. Research foundations*). To delimit the scope of our investigation, we focused on two more specific goals:

Goal 1: To observe the interaction of multilingual screen reader users with partially localised websites.

Goal 2: To evaluate the impact of language selection by the user and the implementation of different localisation strategies on the usability of multilingual websites.

We tried to achieve both goals by designing a multilingual usability study with visually impaired users in order to answer the following research questions, which will be reviewed in detail in chapter *4. Methodology*:

R1: What are the language navigation patterns of screen reader users browsing on partially localised websites?

R2: Are there specific usability issues related to partially localised websites that hamper an accessible browsing experience for screen reader users?

R3: Does the use of different localisation strategies have an impact on the overall usability of multilingual websites for screen reader users?

1.3.1 Hypotheses

In the framework of the second research question **R2**, it was deemed necessary to formulate two hypotheses that would require a deeper investigation within the usability issues encountered by the users. The first hypothesis highlights the type of issue, whereas the second one is focused on the quantity of issues.

H1: Usability issues vary depending on the navigation language chosen by the user.

H2: There are less usability issues in the original language version than in the localised version(s) of a multilingual website.

1.4 Methods

To investigate the above-mentioned research questions, multilingual screen reader users were recruited to carry out a usability test that involved completing a total of six task on two websites. The data collection methods included a think-aloud approach that encouraged the users to explain what they were trying to do and what were the issues that they encountered. The researcher audio-taped the test sessions and took observation notes, that were later transcribed. The experimental protocol also included two post-tasks questionnaires (containing an adapted version of the Computer System Usability Questionnaire (CSUQ) (Lewis 1995) and other questions about the navigation experience) and a post-test questionnaire (with demographic questions, as well as questions about computer and language skills). The researcher met the participants at a location of their choice and read out the instructions and the questions. Chapter 4 (*Methodology*) further illustrates the experimental design and the procedure.

The data was analised through a multi-method approach, in order to cope with the multiple data sources of the experimental setup. Section 5.2 of the *Findings and discussion* chapter provides a detailed account of how a descriptive statistical approach was adopted to analyse quantitative data, whereas a thematic analysis approach was adopted for quantitative data.

1.5 Structure of the thesis

To conclude this introductory chapter, the structure of the thesis will be briefly described. Chapter 2 sets the context of the thesis by addressing some key-concepts and comparing theories; Chapter 3 provides an overview the methods and findings of the work already carried out by other scholars on the topics covered in this dissertation; Chapter 4 examines the methodology used in this study and provides a detailed description ot the procedure of the test; Chapter 5 presents the results of the usability study and illustrates their contribution to the investigation; finally, Chapter 6 draws the conclusions of the work and outlines future research directions.

2 Research foundations

2.1 Overview

In this chapter, we will examine the most important concepts of this thesis by looking at some of the theoretical proposals from both academia and the industry. The first section (2.2) deals with the origins of the term 'localisation', as well as its definition and looks briefly at its history. Sub-section 2.2.1 examines the 'GILT process' and sub-section 2.2.2 deals with the different degrees of localisation. The subsequent section (2.3) attempts to define 'usability' and how to measure it. Sub-sections 2.3.1 and 2.3.2 deal with the usability-related concepts of 'user experience', 'user-centred design' and 'human-computer interaction'. Section 2.4 illustrates the concept of accessibility from different perspectives and presents the most popular guidelines on the subject (2.4.1), while sub-section 2.4.2 examines which guidelines are most relevant to screen reader users and how they should be implemented in the localisation process. Sub-section 2.4.3 reviews the main methods for assessing accessibility. Finally, section 2.5 highlights the relationship between usability and accessibility by presenting three different approaches (sub-sections 2.5.1, 2.5.2 and 2.5.3), including the one that was adopted in this thesis.

2.2 Localisation

The term 'localisation' derives from the notion of *locale*, which is the combination of linguistic and cultural options. It is usually a particular variety of a language, that includes local conventions regarding currency, date and hour settings, symbolic colour coding and so on (Pym 2004, 2). The ISO Standard 17100 defines locale as a "set of characteristics, information, or conventions specific to the linguistic, cultural, technical, and geographical conventions of a target audience" [18]. In a code or a text, the locale is expressed by the language code according to the ISO Standard 639 followed by the country code according to the ISO Standard 3166 (Jimenez-Crespo 2013, 13). For example, Portuguese spoken in Brazil (pt-BR) has a different language code than the one spoken in Portugal (pt-PT)[19].

[18] https://www.iso.org/obp/ui/#iso:std:iso:17100:ed-1:v1:en Last access: 14 June 2018
[19] http://www.lingoes.net/en/translator/langcode.htm Last access: 14 June 2018

As for localisation, the now defunct Localisation Industry Standard Association (LISA) published a popular definition in 2003: "[l]ocalization involves taking a product and making it linguistically and culturally appropriate to the target *locale* (country/region and language) where it will be used and sold". This definition, that incorporated elements from previous definitions, such as 'products' instead of 'texts' as the object of the process, a linguistic and a cultural component, and the previously mentioned notion of *locale* (Jimenez-Crespo 2013, 13–14), was later expanded by LISA itself (2007) in order to include websites and a wider range of services. The word 'services' was added to 'products' and the notion of 'market' instead of 'target locale' was introduced. It also worth noting that the term 'translation' was avoided as if to imply its distinct nature (ibid.).

The field of Translation Studies (TS) has tried to define localisation from a more academic perspective. For example, Dunne (2006, 4) defines it as

> [t]he processes by which digital content and products developed in one locale (defined in terms of geographical area, language and culture) are adapted for sale and use in another locale. Localisation involves: (a) translation of textual content into the language and textual conventions of the target language, (b) adaptation of non-textual content (from colors, icons and bitmaps, to packaging, form factors, etc.) as well as input, output and delivery mechanism to take into account the cultural, technical and regulatory requirements of that locale (ibid.).

Even though it comes from a TS scholar, this point of view follows a professional approach and focuses on industrial practices. Other authors argue that localisation falls under the umbrella term of 'translation-related phenomena' and that it is no more than a translation modality shaped by specific technological features (Jimenez-Crespo 2013, 17).

After reviewing the main definitions from the industry and from the translation studies perspective, the same author presents his own, that will be used as a reference throughout this thesis:

> a definition of web localisation departing from a pragmatic-cognitive, textual and technological process by which interactive digital texts are modified to be used in different linguistic and sociocultural contexts, guided by expectations of the target audience and the specifications and degree requested by initiators (Jimenez-Crespo 2013, 20).

While the LISA definition refers to localisation in general, this one specifically defines web localisation and highlights expectations because of the "interactive nature of digital texts and the distinct cognitive environment of reception" (ibid.), which is key in the context of this thesis. The terms 'e-localisation' (Schäler; Cronin), 'web-content localisation' (Esselink) and

'website localisation' (Williams and Chesterman) have also been used by experts to describe the same concept, as pointed out by Jimenez-Crespo (2013, 12).

In fact, some scholars have tried to define web localisation separately from localisation. For example, Gouadec (2007 in Jimenez-Crespo 2013, 19) defines it as an "[a]daptation of the contents and functionalities of a Web Site for a group of users who share a number of specific cultural and linguistic features different from those for whom the website was originally designed" (ibid.). Interestingly, Sandrini (2008, 9) includes two notions that we are going to analyse in the next sections (2.3 and 2.4), namely usability and accessibility. He defines web localisation as "a process of modifying an existing website to make it accessible, usable and culturally suitable to a target audience" (ibid.). While he probably refers here to accessibility in the sense of availability of the content in question for a different target audience and not conformance to guidelines, we welcome the introduction of the concept of usability and the idea that different users have different needs in terms of access.

Now that we have seen how the localisation industry and the TS scholars define localisation, we will take a step back and have a look at how it all began. The origins of localisation can be traced back to the late 1970s and early 1980s, when personal computers and software started to become popular among users who did not possess programming skills. Once they had succeeded in making their products widespread in the US, companies such as Sun Microsystems, Oracle and Microsoft turned their attention towards international markets like Japan and the so-called FIGS countries in Europe (France, Italy, Germany and Spain). The initial direction of the localisation flow was therefore from English into the languages of the aforementioned countries, and the first approach of developers was to hire linguists to translate the textual strings. By the 1980s and 1990s, the localisation industry had become the fastest-growing sector in translation and had expanded to a wide array of digital texts such as websites, videogames, smartphones and MP3 players (Jimenez-Crespo 2013, 7–9).

In the early 2000s, Web localisation surpassed the market share of software localisation, and the localisation process began to flow in the opposite direction, as websites were localised into English around the world. Also, the actors discovered that separating the development from the translation stages was impractical for a number of reasons and began to conceive localisation collaboratively from the start of the development cycle (ibid.). As a result, localisation became part of an interconnected process known as GILT, which is analysed in the next sub-section (2.2.1).

2.2.1 Globalisation, Internationalisation, Localisation and Translation

Globalisation, Internationalisation, Localisation and Translation (GILT) is a process involving the cooperation of various professionals, such as developers, managers, localisation engineers, localisers and translators (Jimenez-Crespo 2013, 9).

Globalisation represents the broader process in the cycle and focuses mainly on organisational issues. According to the LISA definition (2007 in Jimenez-Crespo, 24-25), globalisation refers to the ensemble of the business decisions required to make an organisation really international in scope and outlook and "[it] is the transformation of business and processes to support customers around the world, in whatever language, country, or culture they require" (ibid.).

Internationalisation is a set of processes to ensure that the general site has as few culture-specific features as possible, since those are the elements most likely to cause problems downstream. The internationalised site is therefore supposed to be neutral, functional, and constructed in such a way that later localisation teams are able to add elements (colours, images, references) that will make the site attractive to users in particular cultural locales (Pym 2010, 3).

Localisation is the next step in the GILT process and it refers to preparing, managing, engineering and quality testing the digital product. Translation, understood as the actual transfer of textual material, is often outsourced and performed by translator-localisers (Jimenez-Crespo 2013, 24–26). As previously mentioned, the localisation industry considers translation is just one of several services that form the localisation process, while some TS scholars consider localisation to be a specific translation modality with an emphasis on the technological component (ibid. 10). Within the framework of the present thesis, that examines university websites, we understand Web localisation as a specific translation modality that requires specific skills from translation professionals (ibid.), because we believe that the whole GILT model applies better to companies and corporate businesses, instead of local institutions.

2.2.2 Levels of web localisation

The degree of complexity of a website can vary greatly, and so can its localisation level. In fact, many scholars have attempted to classify websites according to the degree to which they are localised. Yunker (2003, 128–30), for example, adopted an industry approach and distinguished between (i) 'comprehensive localisation', which does not necessarily include everything in the source website, but offers an equivalent user experience (see sub-section

2.3.1); (ii) 'incremental localisation', that enables companies to start small and develop according to feedback and success; and (iii) 'customised localisation', that consists in redesigning the 'globally unfriendly' source website while localising. In 2001, Schewe (in Pym 2010, 4) proposed a straightforward distinction between monolingual, bilingual and multilingual websites. This categorisation stresses the number of languages instead of the degree of localisation.

Singh and Pereira (2005, 10-15) tried a different approach by applying cultural studies to web design and proposed five levels of localisation based on the role of cultural adaptation: (1) in 'standardised' websites, no effort is made in terms of translation, internationalisation or localisation and the same web content is offered to domestic and international users; (2) 'semi-localised' websites provide contact information about foreign subsidiaries; (3) 'localised' websites offer country-specific Web pages with translation where relevant; (4) 'highly localised' or 'extensively localised' websites offer country-specific URLs with translation where relevant and include high levels of localisation in terms of country specific information, time, date, zip code, etc.; (5) 'culturally customised' websites are designed in a way that reflects a complete 'immersion' in the culture of the target market. This 5-level categorisation relies on technical (URLs) as well as content (contact information, date…) and cultural elements (ibid.).

More recently, Jiménez-Crespo (2013, 35-36) based his localisation-level model on TS and distinguished between 'level 0', where the website offers translated PDF documents or machine translation links; 'level 1', where the website offers a paragraph or page in a different language (normally a brief description of the organisation and basic contact information); 'level 2', where several Web pages have been localised but all navigation menus are in the source language; 'level 3', where the website offers several localised pages with at least one navigation menu in the target language; and 'level 4', which is understood as a fully localised mirror website (ibid.). These levels of localisation are similar to those proposed by Sing and Pereira (2005), but they also introduce some new elements to consider, such as PDFs, machine translation and navigation menus.

In the *Methodology* chapter (4) these categories will be further analysed in order to understand how they have been taken into account for the purposes of this research. In what follows, two other important concepts for the context of this thesis will be introduced: usability (2.3) and accessibility (2.4).

2.3 Usability

According to the ISO definition 9241-11[20], usability is "[the] [e]xtent to which a product can be used by specific users to achieve specified goals of effectiveness, efficiency and satisfaction in a specified context of use." This definition highlights three important measures of usability: 'effectiveness' (task completion), 'efficiency' (time needed to complete the task), and 'satisfaction' (subjective sense of fulfilment). It also focuses on three critical aspects: usability applies to 'specific users', that is, the target users of a product (for example, prospective students are part of the target of a university website); it has 'specified goals', so the goal of the users corresponds to that of the product (for instance, an e-banking interface is used by someone who wants to check their financial situation); and it has a 'specific context of use', which means that the situation of usage plays an important role in the design of a usable product (the size of the text of a game that is designed to be played on smartphones needs to be large enough to be readable on a relatively small screen) (Barnum 2011, 11). Although the examples cited to illustrate the three critical elements of usability are all drawn from the field of interface usability (software or web applications), the concept of usability can be applied to any kind of product from tablets to home appliances and from medical devices to smart watches.

In the chapter titled *The Five Dimensions of Usability*, Quesenbery (2008, 82-88) illustrates the model she created in 2001 to expand on the three ISO 9241 characteristics of usability (efficient, effective, satisfying). The author lists five dimensions, or 'Es': effective, efficient, engaging, error tolerant, and easy to learn. Each of them describes an aspect of the user experience, a concept meant here as a synonym of usability, as we will discuss in the next sub-section (2.3.1), and they can be used to evaluate why an interface succeds or fails. She then proceeds to describe in more detail what these characteristics mean in the context of interface usability. 'Effective' refers to the "completeness and accuracy with which users achieve their goals" (ibid.). 'Efficient' indicates the speed (with accuracy) with which the user can complete the tasks. 'Satisfactory' becomes 'engaging' not merely for the sake of alliteration, but to create a sense of dynamic interaction and it refers to the degree to which a product is pleasant or satisfying to use. 'Error tolerant' means that the design prevents errors and helps users to recover from those that occur. Finally, 'easy to learn' refers to how well the interface supports both initial orientation and continuative use (ibid.).

[20] https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en Last access: 08 June 2016

A more informal definition can be found in Dumas and Redish (1999, 4): "Usability means that people who use the product can do so quickly and easily to accomplish their own tasks." Just as in the ISO definition mentioned in the introduction of the present section, users are the main focus of usability. According to the authors, they are busy people who use the product to accomplish tasks and they decide when a product is easy to use (ibid.).

After going through different definitions of the term 'usability', it is worth examining how usability is perceived or measured from a more practical perspective. According to Rubin, Chisnell, and Spool (2008, 4), what makes something usable is primarily the 'lack of frustration' when using it. A product or service is really usable when "the user can do what he or she wants to do the way he or she expects to be able to do it, without hindrance, hesitation or questions. […] To be usable, a product or service should be useful, efficient, effective, satisfying, learnable and accessible" (ibid.). Interestingly, accessibility is here listed as one of the fundamental characteristics of a usable product. We will further develop this idea in the section *Relationship between accessibility and usability* (2.5).

In fact, the same authors contend that 'true usability is invisible', arguing that we do not usually notice what is pleasant, e.g., the temperature in a room when it is comfortable. Product usability occurs along a continuum, so the right question is not "is this product usable?" but "how usable is this product?" (Rubin, Chisnell, and Spool 2008, 6). This perspective raises interesting questions about levels of usability and methods of measurement. The authors themselves wonder whether it is worth improving usability if users can already accomplish their goals. However, they are convinced that it is impossible to measure the usability of something. One can only measure how unusable a product is by determining what problems users encounter and their cause (ibid.).

The most common way to achieve this goal is to run a usability test by observing real users performing real and meaningful tasks (Barnum 2011, 11). Classic usability tests typically share five characteristics: (1) each test has a main goal, which is to improve the usability of a product, and other specific goals and concerns; (2) the participants are representative of the target users; (3) participants are asked to perform realistic tasks; (4) everything they do and say is observed and recorded; (5) at the end, the data is analysed in order to diagnose and recommend how to fix problems (Dumas and Redish 1999, 22). The *Methodology* chapter of the present work (4) illustrates how these aspects of usability testing have been addressed in this study (see section 4.2).

After an overview of the definition of usability and usability testing, it is also important to understand why anyone would want to measure and improve usability in the first place.

According to Dumas and Redish (1999, 14), everyone benefits from a usability focus: users will welcome the possibility to use more functionality with less effort and learn quickly how to use the product; companies will sell more of the product, enhance the company's reputation, and save money on support, training, update and maintenance costs (ibid.).

2.3.1 User experience and user-centred design

The field of usability partially shares its object of study with that of user experience. The two terms, sometimes used interchangeably, certainly have much in common, but they also differ from each other in some respects: while usability mainly focuses on the ability of the user to use the product to carry out a task successfully, user experience "takes a broader view, looking at the individual's entire interaction with the thing, as well as thoughts, feelings, and perceptions that result from that interaction" (Albert and Tullis 2013, 5).

Another field that revolves around users is that of user-centred design (UCD), an approach that has been known for decades under different names including but not limited to human factor engineering, ergonomics, and usability engineering (Rubin, Chisnell, and Spool 2008, 12). According to the ISO Standard 13407 (ibid.), UCD is "characterised by: the active involvement of users and a clear understanding of user and task requirements; an appropriate allocation of function between users and technology; the iteration of design solutions; and multidisciplinary design." Essentially, it combines the methods and procedures to design usable products, always placing the user at the centre of the process. The goal is to adapt the product or system to the way target users work, instead of forcing them to change how they use something (ibid.).

In this sense, UDC is the larger discipline that encompasses usability and accessibility (see sections 2.3 and 2.4) and is, in turn, part of an even larger, more holistic concept called experience design (ibid.).

2.3.2 Human-computer interaction

Formally founded in 1982, the year of the first conference on Human Factors in Computing Systems in Gaithersburg (Maryland, United States), that later turned into the annual Association for Computing Machinery's Special Interest Group on Computer-Human Interaction conference (ACM SIGCHI), the field of human-computer interaction (HCI) draws on expertise from other areas of research, such as computer science, sociology, communication, psychology, human factors engineering and many others. However, HCI work had been carried out long before then, when computers began to move out of laboratories into homes and offices.

(Lazar, Feng, and Hochheiser 2017, 1–2). According to Lazar, Feng and Hochheiser (ibid.) "[i]t was this move, away from large computers in secure rooms used only by highly trained technical people to personal computers on desktops and tin home dens used by nontechnical people in much greater numbers, that created the need for the field of HCI." In other words, the success or failure of these personal computers depended on their ease of use.

At the beginning, HCI research focused on the way people interacted with office automation programmes such as word processing, database or statistical software. During the early to mid 1990s there was a major shift in the field of HCI due to the wide acceptance of the Internet and the Web. The research community started to study new types of interfaces, such as websites, e-mail services, or chats. In the mid 2000s, the focus of research shifted again towards user-generated content like blogs, forums, wikis and later social networks. With time, the aspect of user variety gained more attention as the interaction of young users, older users, and users with disabilities with technology was finally taken into account by researchers (Lazar, Feng, and Hochheiser 2017, 3–4).

2.4 Accessibility

As we have seen in the previous section (2.3), usability is a broad concept that can be applied to both physical products and software, as well as services. Similarly, accessibility was originally meant as a lack of architectural barriers to a facility (as in the sentence "[t]he theater offers full wheelchair accessibility"[21]), and now applies to all sorts of access, including Web access. In this sense, "Web accessibility means that people with disabilities can perceive, understand, navigate, and interact with the Web, and that they can contribute to the Web" (WAI/W3C 2005 in Petrie, Savva and Power 2015, 5).

Petrie, Savva, and Power (2015) have recently tried to propose a unified definition of Web accessibility by analysing the existing ones and extracting the recurring core concepts to combine them together. The result is a very complete definition that reads:

> all people, particularly disabled and older people, can use websites in a range of contexts of use, including mainstream and assistive technologies; to achieve this, websites need to be designed and developed to support usability across these contexts (Petrie, Savva, and Power 2015, 1).

If we consider this definition and take into account the needs of the blind comunity, the challenge that emerges is rather clear: guaranteeing access to visually impaired people to a

[21] https://dictionary.cambridge.org/dictionary/english/accessibility Last accessed: 10 June 2018

computer system that generally utilises the user's visual channel by presenting patterns of light point sources on a screen (Barreto 2008, 3). As Cunningham puts it, "[t]he blind are particularly impacted by an inaccessible web. A page might be structured in a way that's nonsensical if a user is using a screen reader. They might miss out on vital information in a graph or image. They might have to sit through listening to the navigation with every page load" (Cunningham 2012, 1). However, the access problems of visually impaired people are different from those of people who cannot see at all. A 'one-size-fits-all' approach would not work, as there are many different causes and forms of visual impairment and corresponding variety of accommodations that can be provided (Edwards 2008, 150–51). For example, blind users might use a screen reader, while those who have some sight "generally prefer to make as much use of their vision as possible" (ibid.), so they may override the default styling of a web page in order to visualise colours that are of higher contrast or fonts that are legible (Cunningham 2012, 1).

To give developers an idea of what they should take into account to implement accessibility in their software applications or Web sites, the World Wide Web Consortium (W3C), a web standard organisation committed to an accessible Web, has published some guidelines. In the next sub-section (2.4.1), we shall briefly look at the latest update of these guidelines, which prescribe how websites can be authored so that the web content meets specific accessibility requirements (Barnum 2011, 109; Hanson, Richards, and Swart 2008, 216).

2.4.1 Web Content Accessibility Guidelines

Perhaps the most widely recognised web accessibility guidelines, the Web Content Accessibility Guidelines (WCAG) are part of the Web Accessibility Initiative, first launched in 1997 by the W3C. They are divided into four major principles (perceivable, operable, understandable and robust), 13 guidelines (general goals), and multiple success criteria divided into three levels of conformance (A, AA and AAA)[22]. Together with the Authoring Tool Accessibility Guidelines (ATAG), which address authoring tools for generating markup, and the User Agent Accessibility Guidelines (UAAG) for Web applications, they constitute the three-part approach to Web accessibility of the W3C (Barnum 2011, 109; Albert and Tullis 2013, 230–31; Hanson, Richards and Swart 2008, 216).

The latest version of the recommendation, WCAG 2.1, was published on the 5 June 2018 and contains, like the previous versions, success criteria written as testable statements that are not technology-specific. An example of best practice that is particularly relevant for blind

[22] https://www.w3.org/TR/WCAG21/#wcag-2-layers-of-guidance Last access 10 June 2018

people and refers to the first principle, 'perceivable', reads: "All non-text content that is presented to the user has a text alternative that serves the equivalent purpose, except for the situations listed below." It then lists a series of exceptions such as "[i]f non-text content is pure decoration, is used only for visual formatting, or is not presented to users, then it is implemented in a way that it can be ignored by assistive technology."[23]

Although the WCAG guidelines have been recently updated (version 2.1), the development of an increasingly dynamic Web required a different approach to keep up to speed with the complexity of websites. The Web Accessibility Initiative-Accessible Rich Internet Applications (WAI-ARIA[24]) was accepted into the W3C standards to fill this gap. It helps with dynamic content and advanced user interface controls developed with Ajax, JavaScript, and related technologies. Specifically, it allows programmers to declare the specific role of certain elements on the page so that disabled users can interact even with the most complex applications (Cunningham 2012, 23).

2.4.2 Key WCAG success criteria for Web Localisation

In an article titled *Localisation and web accessibility* (Gutiérrez y Restrepo and Martínez Normand 2010), the authors describe the success criteria that are most important for maintaining the level of accessibility of localised web content. The aim of this sub-section is to highlight the items that are particularly relevant for screen reader users. In the *Findings and discussion* chapter (section 5.4) we analyse the problems encountered by those who took part in our study and examine whether they could be prevented by implementing the following guidelines in the localisation process.

The key localisation issue for the 'perceivable' principle is to localise alternatives for non-text content, such as images, sound, video and interactive controls, because these elements are not perceivable for all users. This text alternative can be rendered in the modality that they can perceive, that is, auditory (text-to-speech), tactile (Braille) or visual (displayed on the screen) output. The success criteria related to alternatives are *1.1.1 Non-text content*, *1.2.1 Audio-only* and *Video-only*, *1.2.8 Media Alternative*, and *1.2.9 Audio-only*. An additional perception-related issue is the sequence of the content: a correct reading sequence should be available to the software that renders the content, especially if the sequence in which it is presented to the user affects its meaning. Given that the reading sequences are not the same in

[23] https://www.w3.org/TR/WCAG21/#text-alternatives Last access 10 June 2018
[24] https://www.w3.org/WAI/standards-guidelines/aria/#introduction Last access 10 June 2018

all languages, special attention should be paid to correctly localising this reading sequence. The related criterion is *1.3.2 Meaningful Sequence* (ibid.).

The key localisation issues for the 'operable' principle are related to bypassing blocks of content, page titles, focus order, link purpose and headings. This section is particularly relevant for blind users, as many criteria are related to the use of a screen reader. (a) The success criterion *2.4.1 Bypass Blocks*, for instance, refers to the need to provide a link (that is sometimes concealed from the users until it obtains keyboard focus) to skip navigation elements, and this link is, in some cases, concealed from the users until it obtains keyboard focus. Such links should be located and localised correctly in all versions of the website. (b) Another success criterion, *2.4.2 Page titled*, refers to the title that appears in the title bar of the browser and appears as the site name when storing bookmarks. The information provided by the title of the orignal version should be maintained across all localised versions. (c) The third relevant success criterion, *2.4.3 Focus order*, states that the order in which the components of a web page receive focus when pressing the tab key should preserve meaning and operability to ensure that keyboard users have a similar experience to users of pointing devices. From a localisation perspective, it is important to consider whether the destination language has a different reading order or direction than the original language. (d) The idea behind the purpose of links (success criteria *2.4.4 Link Purpose - In Context* and *2.4.9 Link Purpose - Link Only*) is that link content should indicate clearly where the user will land or what will happen when clicking on the link. Localisers should pay attention to providing adequate information about the link purpose. (e) Finally, headings, labels and sections should clearly represent their purpose to enable the users to easily navigate through the structure of the document (success criteria *2.4.6 Headings and Labels* and *2.4.10 Section Headings*). This is therefore another aspect that shoud be carefully considered when localising a website (ibid.).

The key localisation issues for the 'understandable' principle are related to language identification, unusual words, reading level, pronunciation, error management and help. To handle language changes and correctly present information to the user, identifying the human language used in web content is essential. (a) For example, the language needs to be correctly identified for text-to-speech technology to deliver the appropriate output and the search engines to find pages in the user-defined language. When localising web content, it is important to correctly indicate both the language of the page and the language of parts, using the codes described in the localisation section (2.2). The relevant success criteria are *3.1.1 Language of Page* and *3.1.2 Language of Parts*. (b) Also, according to the success criteria *3.1.3 Unusual Words* and *3.1.4 Abbreviations*, unusual words and abbreviations should be provided with

mechanisms for identifying their expanded form or meaning. Localisation has to maintain the definitions and expansions of the original content and could even add some more to enhance the understandability of the localised version. (c) Other relevant success criteria include *3.3.2 Labels or Instructions* and *3.3.5 Help* and apply mainly to intractive content, such as forms. Screen reader users benefit significantly from clear and adequate labels of form elements. Localisers should pay special attention to making the localised labels, instructions and help useful for the intended users (ibid.).

2.4.3 Accessibility evaluation

Just like for usability (see section 2.3), many scholars and field experts have attempted to find methods to measure accessibility. Generally speaking, it is possible to distinguish between three different approaches: (1) 'automated testing' is carried out by software tools without the need for human intervention; (2) 'manual testing' is performed by human evaluators; and (3) 'user testing', the approach adopted in this thesis, is also carried out by humans, but they need to be potential end-users of the product (Abou-Zhara 2008, 85–87).

According to Brajnik (2008, 113-14) the most popular method is the 'conformance review', also called 'expert review' or 'manual inspection', and it requires checking if a page satisfies a checklist of criteria. The checklist is usually based on the WCAG guidelines, but it can also reflect state-level guidelines or those of individual organisations. This method generally allows the evaluators to identify a large range of problems and is cost-effective, especially when coupled with automatic testing tools; however, it requires experienced evaluators and sometimes fails to distinguish between important and unimportant accessibility problems. Other methods for evaluating accessibility include 'screening techniques' (informal empirical techniques that somehow artificially reduce sensory, motor or cognitive capabilities of the user); 'subjective assessments', based on a panel of users instructed to explore and use a given website by themselves, and later provide feedback on what worked for them and what did not; 'barrier walkthrough', where an evaluator has to assess a number of predefined barriers which are interpretations and extensions of well-known accessibility principles; and finally the main method used in this thesis: user testing, based on informal empirical usability tests with disabled people (ibid.). Chapter 4 will expand on this aspect and further illustrate the methodology of this research work.

2.5 Relationship between usability and accessibility

In previous sections, usability and accessibility were described independently, but prior work has repeatedly studied them together, as they are closely related. When trying to explore their relationship, most authors use a problem-based approach to distinguish accessibility issues from usability issues and then study the connection or overlap between them.

2.5.1 Usability includes accessibility

One popular approach is to consider accessibility problems as a sub-set of usability problems and most accessibility definitions highlight the aspect of disability as the feature that distinguishes it from usability. For example, Albert and Tullis (2013, 228) believe that "accessibility is really just usability for a particular set of users." According to the definition of the British Standards Institute (BSI), Web accessibility is "[T]he extent to which a website can be used by users with specified disabilities to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (BSI 2010). This definition has a lot of elements in common with the previously mentioned ISO definition of usability (users, goals, effectiveness, efficiency and satisfaction, specified context of use) and the only different aspect is that of disabilities (see section 2.3). While it is not uncommon for standards to use similar wordings, it is striking to observe how similar they are. Visually, this point of view is represented in Figure 2.1:

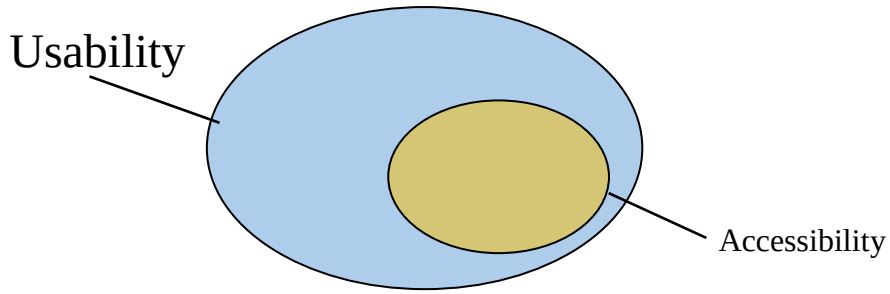


Figure 2.1 Relationship between usability and accessibility, where the latter is part of the former (1)

However, this theory fails to consider those problems that affect people with disabilities but have no impact on mainstream users: the fact that a title is not labelled as heading in the code but just written in a bigger font, for example, would probably be an issue only for screen reader users.

2.5.2 Accessibility includes usability

In their review of accessibility definitions, Petrie, Savva, and Power (2015) classify 50 definitions of web accessibility. Although most of them share the same key-features, some of them barely mention disability and do not mention usability or the verb 'use' at all, but focus instead on access to information regardless of the technology used. For instance, Martín, Mazalu and Cechich (2010 in Petrie, Savva and Power 2015, 11) argue that "Web Accessibility means universal access on the Web, regardless of the kind of hardware, software, network platform, language, culture, geographic location and users' capabilities." In this sense, it seems that accessibility is somehow the precondition to even consider usability issues, because it does not matter whether some software or Web page is easy to use if it cannot be accessed first[25]. Shneiderman (2000, in Petrie and Kheir 2007, 398), for example, argues that "access is not sufficient to ensure successful usage". In other words, not everything that is accessible is also usable. Figure 2.2 visually represent this concept:

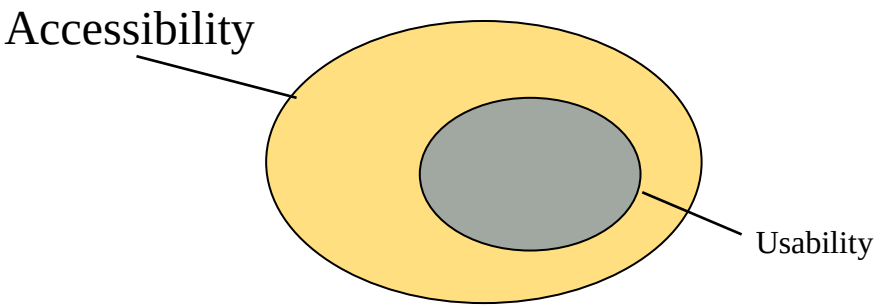


Figure 2.2 Relationship between usability and accessibility, where the former is part of the latter (2)

2.5.3 Overlap between usability and accessibility

The approach we have adopted in the present study is that of Petrie and Kheir (2007), which considers accessibility and usability as two interconnected yet independent areas of research. Petrie and Kheir (2007, 398) mention the distinction between 'usable accessibility', or usability for people with disability, and 'technical accessibility', that mainly relies on meeting technical criteria and guidelines in the underlying Web code (Petrie and Kheir 2007, 397). Figure 2.3 visually represents this distinction:

[25] This suggestion emerged from an informal conversation that took place in December 2017 between the researcher and Markus Riesch, Head of e-Accessibility of the Swiss Confederation (Geschäftsstelle E-Accessibility Bund) on the topic of accessibility research.

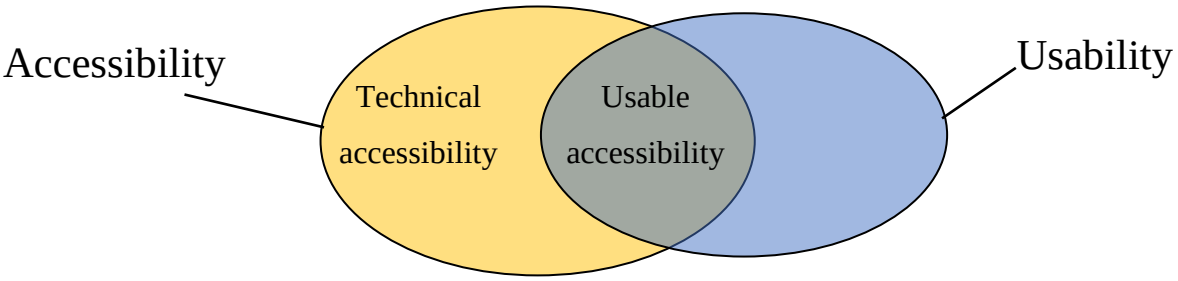


Figure 2.3 Relationship between usability and accessibility, where the two overlap (3)

The same authors explore the relationship between accessibility and usability by trying to classify problems into categories (Petrie and Kheir 2007, 398). This approach, that will be further described in the *Related work* chapter (sub-section *3.3.1 – Overlapping problems*), seems particularly suited to the present study, as identifying some of the challenges that screen reader users have to deal with when navigating on multilingual Web sites is crucial to answer our second research question (see section 5.4) . Yet, the dimension of multilingualism involves localisation and introduces a third element to consider.

Concretely, in terms of area of interest, the territory we are exploring is the intersection between accessibility, usability and localisation problems. Figure 2.2 visually represents the area of research (red dotted line) explored within the framework of this thesis:

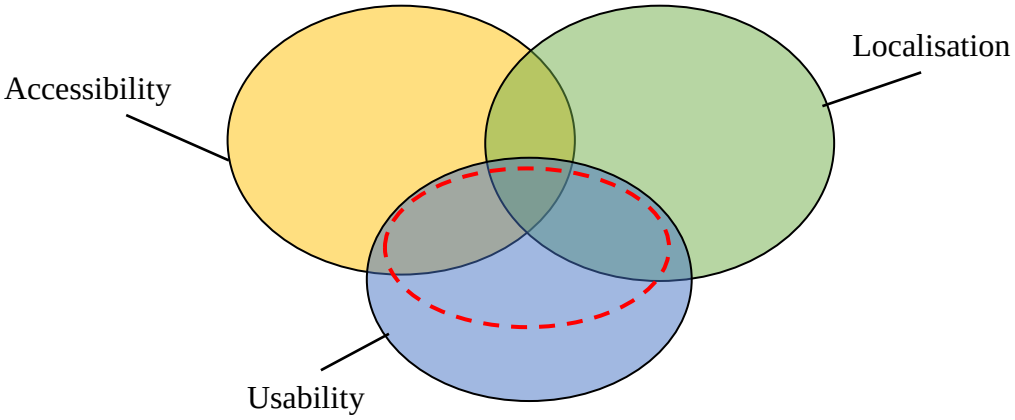


Figure 2.4 Intersection between accessibility, usability and localisation

Before moving to the presentation of the methodology followed in this study, it is worth reviewing the academic enquiry work carried out thus far on the subjects of localisation, usability and accessibility, with a particular focus on Switzerland and higher education institutions.

3 Related work

3.1 Overview

This chapter provides an overview of the research that has been carried out to date in academia and in the industry on the main topics covered by this thesis, with a focus on the studies that were relevant to us because of their findings or the methods used.

Section 3.2 presents a summary of Web accessibility studies with screen reader users aimed at evaluating accessibility (sub-section 3.2.1) or assessing evaluation methods (sub-section 3.2.2). The second section (3.3) focuses on research that has combined usability and accessibility by reporting the findings of a systematic review on the topic and then presenting some studies that examine the relationship between accessibility and usability problems (sub-sections 3.3.1 and 3.3.2). In the third section (3.4), we report the work of those academics who have analysed the role of localisers and language service providers in the implementation of accessibility and who have tried to combine accessibility and localisation assessment (sub-sections 3.4.1 and 3.4.2). Finally, section 3.5 focuses on relevant studies that examine different studies that involve university websites: a systematic review (sub-section 3.5.1), accessibility (sub-section 3.5.2), policies (sub-section 3.5.3), user interface (sub-section 3.5.4), and internationalisation (sub-section 3.5.5).

3.2 Accessibility evaluation

Since this study focuses on visually impaired people as target users, we deemed it relevant to review prior research studies with blind individuals or blind accessibility experts.

3.2.1 Perceived Web accessibility

Aizpurua and other scholars have investigated the question of perceived accessibility, with a focus on blind users. In a 2013 study (Aizpurua, Arrue, and Vigo 2013), the role of expectations on perceived Web accessibility was explored through an experiment with 11 blind users. The participants were interviewed about their previous experiences and expectations on restaurant websites and were then asked to perform three tasks on each website (following a within-subject design): familiarise themselves with the website, find the menu and make a reservation. The four tested websites were selected because of their accessibility level (assessed

by the authors through automatic testing and barrier walk-through) and their style (innovative vs traditional). The initial findings of this study suggest that "previous experiences, either real (at restaurants) or virtual (with similar websites) may determine the general expectations participants had with regard to the selected websites" (ibid.). The three authors published another paper on the same subject two years later (Aizpurua, Arrue, and Vigo 2015), where they expanded on the factors that can influence perceived accessibility. Specifically, they identify past experiences, prejudices, evoked memories, unmet expectations and confidence as fundamental aspects to consider when assessing accessibility. While this seems to imply that conforming to accessibility guidelines is not sufficient to guarantee accessibility, the authors concede that the identified subjective factors cannot be covered by technical guidelines with standardisation purposes (ibid.). Both studies were relevant to us inasmuch as they made us to reflect on the role of subjective factors in user testing. The experimental design and, in particular, the post-task and post-test interviews (see sub-section 4.6.3) were influenced by these considerations.

3.2.2 Evaluation methods

The topic of perceived Web accessibility has been further investigated by Aizpurua in a study on the implications of user testing with blind users (Aizpurua et al. 2014). The study focused particularly on the sources of bias that may have an influence on the validity and reliability of the evaluation results. In fact, in the experiment with the four restaurant websites, the users' perceived accessibility level did not always correspond to the level resulting from the evaluation of guidelines conformance. By analysing the results, the authors identified sources of bias related to the wording of instructions, to the level of user expertise, to the setting of the experiment and to the way findings are reported that need to be controlled in order to avoid an impact on the validity and reliability of the results. They provide suggestions on how to control them and conclude that user testing can be a valid and useful method to assess accessibility (ibid.). In section 4.2 of the *Methodology* chapter, we will describe how we attempted to take these recommendations into account for our study.

In 2008, Brajnik tested conformance review against barrier walk-through in an experiment where 12 participants were asked to perform two accessibility evaluations on the two selected websites (within-subject design). The results of the study show a clear prevalence of violations identified on both websites using the barrier walk-through method, a difference that, according to the author, might be due to the highest level of fine-tuning of barriers over guideline checkpoints. Nonetheless, the number of identified violations was not the only factor

considered by the researcher. For example, conformance review scored slightly better (10%) in terms of reliability. While Brajnik acknowledges that some aspects of the experimental plan could be improved (such as tool support and instructions), he believes that barrier-walkthrough has proved to be a method that leads to more relevant results. The findings of this study are one of the reasons why we opted for an evaluation method involving the users, instead of checking against guidelines violations.

In both the aforementioned studies, guideline conformance was used as a benchmark for the evaluation of another method (user testing and barrier walkthrough), but other scholars (Power et al. 2012) have focused on the WCAG guidelines by investigating the complex relationship between problems encountered by screen reader users and the guidelines. The study involved a task-based user evaluation with blind users and a think aloud protocol. The 32 participants were asked to rate the severity of each problem on a four-point scale adapted from Nielsen's usability problems rating scale (Nielsen 1994, 103). The evaluation of a total of 16 websites revealed that the number of problems experienced by users while browsing WCAG 2.0-compliant websites was not significantly lower than non-compliant websites, while the difference was significant when comparing WCAG 1.0-compliant websites to non-conformant ones. In particular, only 49.6% of problems were addressed by directly relevant WCAG 2.0 success criteria. In the conclusions, the authors highlight the fact that blind users reported problems when they encountered unexpected content or when they could not find content on a website and these problems are not covered by WCAG 2.0. The authors advocate for a shift from the problem-based paradigm for Web accessibility to a more heuristic approach such as the one adopted by usability research based on user data, in other words, one that focuses on the use of the Web by people with disabilities, instead of just the problems they encounter (ibid.). This is precisely the approach to accessibility that we embraced in our study, where the navigation patterns of screen reader users play a central role.

3.3 Accessibility and usability

As mentioned in the *Research foundation* chapter (section 2.5), many scholars have studied the relationship between accessibility and usability. In the following sub-sections, some of their work is analysed and compared.

3.3.1 Overlapping problems

In 2007, Petrie and Kheir published a study on the topic, where they investigated the nature of the relationship of the problems encountered by non-disabled (sighted) people and blind screen reader users, and tried to find out whether problems that are encountered by both blind and sighted people are more severe for the blind people than the sighted people. Two websites of mobile phone companies were selected and six sighted and six blind participants were asked to perform 7 tasks on each website. The severity of the problems they encountered was rated on Nielsen's four-point scale (Nielsen 1994, 103) by both the participants and the researchers. The importance of problems was also assessed according to accessibility and usability guidelines. Results showed that the problems encountered by the blind and sighted participants constituted intersecting sets with an overlap of about 14%. Even though it was lower than expected, the authors highlight the fact that visual impairment is one of the disabilities that has a greater impact on the interaction with the Web and that the problems encountered by other groups of disabled people could have a greater overlap. As for the problems encountered by both groups of participants, only limited support was found for the hypothesis that they would be rated more severely by blind people, as one of the websites presented no significant difference. Finally, they found a strong agreement between participants' ratings but no significant relationship between the ratings given by participants or researchers and those provided by the guidelines (ibid.). The study by Petrie and Kheir is relevant for our thesis because it examines the relationship between usability for non-disabled people and accessibility for disabled people, and shows there is an overlap. We take this a step further and examine usability for disabled people, considering them as target users.

3.3.2 Perceived accessibility and user experience

The relationship between Web accessibility and usability has more recently been explored by Aizpurua, Harper and Vigo (2016) in a study where they compared perceived accessibility and conformance to guidelines to user experience attributes. In fact, during the experiment with the four restaurant websites (see sub-section 3.2.1) they also collected user experience scores by means of questionnaires and semi-structured interviews. Specifically, they used existing tools such as *Attracdiff 2*, a questionnaire that consists of a set of 23 word pairs reflecting opposite adjectives that can be rated on a 7-point scale to assess perceptions of users, and the Emotion Word Prompt List (EWPL) by Petrie and Precious, that consists of 11 emotional words that can be rated through 7-point Likert items. The authors have the run analyses of statistical

correlation and found that accessible websites were related to positive emotional reactions, while non-accessible ones were correlated the negative ones. While these outcomes were not surprising, they provide "empirical evidence indicating that perceived accessibility and user experience could be understood as interchangeable qualities for blind users" (ibid. 23). These findings were very interesting for our study because they show how the user experience is strongly impacted by the perception of accessibility and validates our understanding of usability and accessibility as interconnected qualities that can be studied together.

Although all the studies reviewed considered usability and screen reader users, none of them seems to have considered multilingualism as an important factor. We are aware of studies that have combined multilingualism and usability, such as Miraz, Ali, and Excell (2013) and Andreu-Vall and Marcos (2012). However, to the best of our knowledge, no specific guidelines exist to assess the usability of multilingual websites from an accessibility perspective.

3.4 Accessibility and localisation

The relationship between accessibility and localisation has also been investigated by different scholars, although in a more scattered manner, and some students have tried ot address both of them simultaneously in their master's dissertations.

3.4.1 Localisation workflow and accessibility

An article by Rodríguez Vázquez and Torres Del Rey (2014) highlighted the responsibility of all agents in the localisation process – as opposed to just the Web developers – for the implementation of accessibility best practices. Web translators, they argued, have a considerable impact on the accessibility of a monolingual website that is being translated and "[f]or a website to be correctly localised, the new language versions should, at least, be as accessible as the original product" (ibid. 35). This recommendation was crucial for our research because it highlighted the importance of advocating for accessibility – in our case usability for people with visual impairment – in multilingual websites such as the ones we have tested.

The place of Web accessibility requirements in the localisation workflow has been further investigated by Rodríguez Vázquez and O'Brien (2017) in a qualitative study where they interviewed fifteen representatives of six leading language service providers. The findings suggested that accessibility is not yet considered by default in the localisation process and that there was still a lack of awareness about the importance of this topic among localisation actors. With this study, we hope to catch the attention of the people responsible for the creation and

maintenance of multilingual Web content in Switzerland, so that they become more aware of the usability challenges that visually impaired users have to face when navigating on multilingual websites.

Finally, we deem it relevant to mention the empirical investigation on the achievement of appropriate text alternatives for images conducted by Rodríguez Vázquez (2016) because it focuses on an aspect that is still seldom examined by researchers. Her study, aimed at "[determining] the extent to which localisers are capable of assuring that an acceptable level of image accessibility" (ibid. 177), combined descriptive and experimental research strategies and included a controlled localisation experiment whose output was then subjected to evaluation by screen reader users. Interestingly, the author highlighted the lack of a standardised procedure for assessing the accessibility level of multilingual websites, something that we noticed as well with regards to the usability level of multilingual websites for people with visual impairment. In this sense, both studies try to introduce the element of multilingualism in a research field that rarely considers this aspect. In the next sections, two master's theses that try to combine accessibility and localisation evaluation in a multilingual context will be presented.

3.4.2 Assessing accessibility and localisation

In the framework of bilingual Canada, LeBlanc (2018) created a quality assessment grid that included translation, localisation and accessibility criteria. She then applied it to six bilingual (English and French) municipality websites and found that Web accessibility principles were generally ignored and that the grammar rules and semantics of the minority language were often neglected (ibid. iv). This last finding was particularly interesting for our research because a similar situation could be found in multilingual Switzerland, where German, French and Italian have equal official status, but German was declared as the primary language by 62,8% of the population in 2016, making French (22,9%) and Italian (8,2%) *de facto* minority languages[26], as shown in the Figure 3.1:

[26] https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/langues-religions/langues.html Last access: 11 June 2018

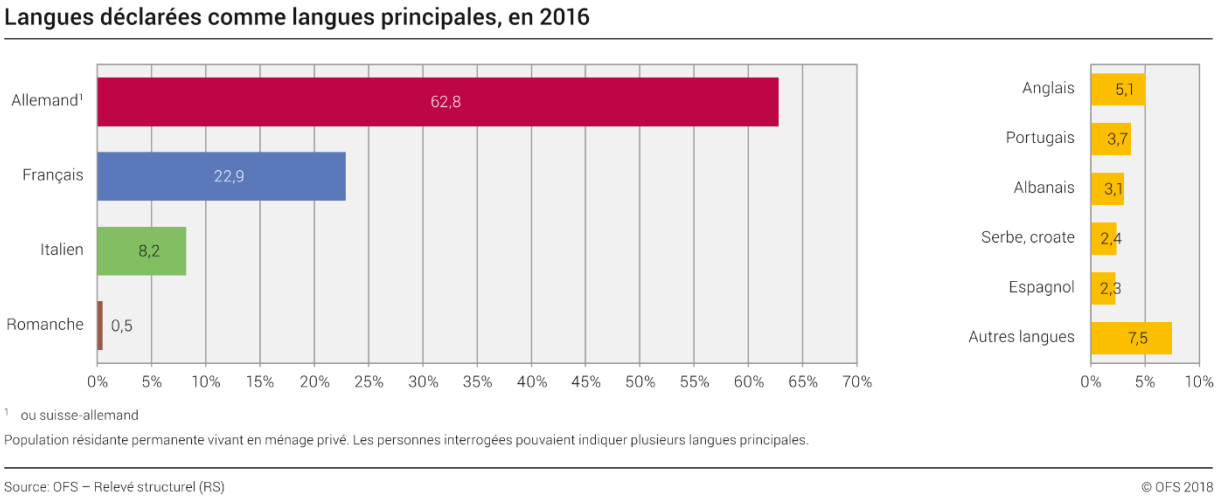


Figure 3.1 Language declared as primary language by the Swiss population in 2016 (Office fédéral de la statistique 2018)

Quazzico (2016) also focused on localisation and accessibility assessment, but in this case, in the context of Switzerland. She conducted a study that combined automatic and manual testing on the websites of a Swiss broadcasting company. The author compared the 2011 and 2015 versions of the *Radio Télévision suisse* (RTS) and the *Radiotelevisione svizzera* (RSI) websites to find out whether there had been any improvement in terms of localisation or accessibility. The results of her tests showed that there was indeed an improvement, but the aspect that was most interesting for the present study is that of the language attribute of the pages (see sub-section 5.4.1), that was missing in 2011 but was was correctly coded in 2015. As in the case of Quazzico (2016), our study has a strong focus on language. However, the studies differ in terms of evaluation methods and the type of studied websites, and while she adopted a technical perspective, we examine, among other aspects, the reaction of the users to the lack (or presence) of the right language attribute.

3.5 Higher education institutions

As mentioned in the introductory chapter, access to education is key for people with disabilities, hence the need to build accessible university portals. Given that this type of websites were the ones finally selected for your study (see section 4.5.1 – Selection of websites), it seems crucial to explore if there has been prior work focusing on this genre of websites from an accessibility, usability or localisation angle. The first sub-section (3.5.1) summarises the findings of a systematic review of usability studies. The second sub-section (3.5.2) describes a study that examined the accessibility of numerous Swiss websites, including 20 universities. In the third sub-section (3.5.3), we describe a research study that examines the Web accessibility policies of twenty selected universities. The fourth sub-section (3.5.4) introduces a study that

compares the traditional graphical user interface with an enhanced one applied to the website of the University of Basel. Sub-section 3.5.5 focuses on the strategies adopted for the translation of university web portals.

3.5.1 Systematic review

In 2017, a systematic review of studies that deal with the usability of university websites undertaken by Yerlikaya and Onay Durdu (2017). In this review, they investigate, among other things, the distribution of the number of publications by years, the most common evaluation and data analysis methods, the user profile distribution and the most commonly addressed usability issues in the studies. This 'map' of previous studies was useful since it provided a picture of the work done so far, the trends observed and the aspects that were underrepresented. We followed the trends by choosing to carry out a summative study (that is, as study on websites that are already being used, as opposed to websites that are being designed), a user-based usability study (as opposed to expert-based or data-driven), and using a questionnaire and a descriptive approach to analysis. But we also selected methods often used in other fields, such as observation and qualitative analysis, and our participants had an unusually varied background. According to the authors (ibid.) nearly half of the 53 examined studies listed accessibility as a usability issue, but only four investigated the issue in depth, as we aim to do.

3.5.2 Accessibility of Swiss universities' websites

The 2016 Swiss Accessibility Study (Étude Accessibility 2016 en Suisse) is a comprehensive list of Swiss websites. Now at its fourth edition , the study is published every three to five years by the Foundation Access for All. This foundation has been active in the field of accessible technology since 2000 and is a competence centre financed by donations, sponsors, and the services it provides, such as certifications and consultancy . The study tested the accessibility of 98 Swiss websites that belong to the public administration, companies owned by the State, cantons, the ten largest cities, universities, news providers and online shops, as well as 15 mobile applications. The websites were tested between February and June 2016 by three accessibility experts, one of whom is completely blind. Windows 7 and 10, Mozilla Firefox (the latest version available at the time) and JAWS 17 were used to carry out the tests. The main accomplishment of this study is that it gives the reader a clear outlook on the situation of web accessibility in Switzerland, something that we decided to exploit, as described in the *Methodology* chapter (section 4.5.1). Such a comprehensive approach was nonetheless bound to sacrifice other aspects: as previously mentioned, given the number and size of the tested

websites, only four pages for each website were thoroughly analysed. For our study, we decided to test only two websites in more detail.

3.5.3 Accessibility policies

Brajnik and Graca (2018) wrote an article based on Graca's master's thesis on accessibility policies for higher education institutions. The paper analyses existing accessibility policies of twenty universities in Europe, in the USA, Canada and Australia and examines which accessibility-related goals should be included in policies and how they should be formulated. To assess the quality of the policies they analysed, three aspects were considered: completeness, concreteness and understandability. The ultimate goal is fulfilled inasmuch as the authors outline and explain eight components that a Web accessibility policy should include, in the hope that universities interested in implementing accessibility will follow these best practices (ibid.). This study covers some aspects, such as policies, that are complementary to our area of interest, but that we did not explore. Nonetheless, the state of the art's chapter of Graca's thesis (2018) is useful to us because she examines the legislation in terms of digital accessibility (ibid.). Unfortunately, while the introduction to the general framework is relevant to us, the analysis of accessibility legislation for higher education only considers English-speaking universities and does not cover Switzerland. Similarly, the study has not yet been followed up with an experiement with real end users to test the accessibility level of the examined university portals.

3.5.4 Accessibility and user interface

In the context of Switzerland, a study was conducted on the website of the University of Basel in order to test an alternative to the original graphical user interface (GUI) (Leuthold, Bargas-Avila, and Opwis 2007). The authors developed nine guidelines to create a text-only user interface that they called 'enhanced text user interface' (ETI) and applied them to the tested website. They then asked 39 blind participants to perform two tasks on three different interfaces: the actual website, partially compliant with the WCAG, the fully compliant version of the same website, and the ETI version. They did not find any significant difference between the partially and fully WCAG-compliant versions of the GUI interface, but they did observe better performances with the ETI interface for the three examined usability factors of 'efficiency', 'errors' and 'satisfaction' in one of the two tasks. The strength of this research study was the application of the user-centred design method according to the ISO 13407

standard[27] and the quantitative approach that differentiates it from most usability lab studies with blind users. Limitations include only two tasks to test the interfaces and the lack of two more usability factors (learnability and memorability), both of which were due to time constraints, but would deserve further investigation in the future (ibid. 268). Our study adopts a more holistic approach and considers overall usability, instead of single factors.

3.5.5 Localisation and internationalisation

A 2012 study by Fernández Costales focused on the translation of university websites and addressed questions such as 'into which languages are online contents mostly translated?', 'are European language policies on multilingualism being effectively applied?' and 'are institutional websites localised for a particular market or rather globalised for the international audience?' While his study investigates university websites in the European Union (and therefore excluded Switzerland from the analysis), Fernández Costales raised an interesting point about the adaptation level of university websites:

> it cannot be stated that university websites are localised, since there are almost no examples of sites specifically adapted to a particular locale or market. It is possible to find some Spanish universities translated into Chinese or some British institutions addressing specific groups of international students, but this is not a general rule and the adaptation level does not allow us to consider these as examples of localisation. On the contrary, the common trend is toward an internationalised version of the website in which contents and information are provided to a global audience. In this regard, the role of English as the international lingua franca can be easily confirmed in the case of European university websites (Fernández Costales 2012, 56).

This remark was cause for reflection when selecting the websites for testing (see section 4.5.1 of the *Methodology* chapter). Besides, the findings by Fernández Costales (ibid.) suggest that "it cannot be stated that there is a direct link between the number of official languages [of a country] and the translation of the sites" (ibid. 56-57). While in some countries like Spain, most universities based in regions with co-official languages (Catalonia, Galicia and the Basque Country) have versions in Spanish, English and the corresponding language, other European countries like Belgium "show a clear segregation strategy when avoiding two of the three official languages in university websites.: all the institutions in Flanders avoid using French, and the universities in the francophone area forget about including contents in Dutch" (ibid.). Whilst browsing numerous Swiss websites, including university websites, we noticed that not

[27] Valid until 2010, then revised by ISO 9241-210: https://www.iso.org/standard/52075.html Last access: 10 August 2018

all official languages are always present and that English is sometimes used as *lingua franca*, a tendency that Fernández Costales also found in his study (ibid.).

4 Methodology

4.1 Overview

In this chapter we are going to take a closer look at the methodology that we briefly mentioned in the *Introduction* chapter. We will explain how the goals and research questions have influenced the design of the study and describe why we chose a specific combination of methods (sections 4.2 and 4.3). Section 4.4 provides a detailed description of the call for participation, the recruitment process and the partipants' profile (sub-sections 4.4.1, 4.4.2 and 4.4.3). Section 4.5 illustrates the stimuli that we used for the test, with a focus on the selection and inspections of the test websites (sub-sections 4.5.1 and 4.5.2). Finally, section 4.6 thoroughly examines the different stages of the experiment, by presenting the test's scenarios and tasks on both websites (sub-section 4.6.1), the methodology test (sub-section 4.6.2), the test session's procedure (sub-section 4.6.3) and the technical setting of the test sessions (4.6.4).

4.2 Usability study: goals and research questions

As Lazar states, "choosing which method to use is a highly context-dependent issue related to a variety of factors including the primary purpose of the study, time constraints, funding, the participant pool, and the researchers' experience." (Lazar, Feng, and Hochheiser 2017, 25). In the *Introduction* (section 1.3), we illustrated the general purpose of the present work, which is to explore the relationship between the usability, accessibility and localisation of websites by investigating how screen reader users navigate multilingual websites. Within the framework of this thesis, an exploratory usability study was conducted in order to achieve two objectives: (1) observe the interaction of screen reader users with partially localised websites, with a particular focus on language navigation patterns and (2) evaluate the impact of language selection by the user and the implementation of different localisation strategies on the usability of those websites.

In regards to the first objective, we have formulated the following research question:

R1: What are the language navigation patterns of screen reader users browsing on partially localised websites?

By language navigation patterns, we are referring to the combination of preferences, choices, needs and actions, as well as the impact of those on the language in which the browsing takes place.

As for the second objective, we formulated one research question for each variable. The first one (**R2**) considers the impact of language on the usability of partially localised websites. Specifically, we want to investigate whether partially localised websites present specific language-related usability issues. In the two hypotheses formulated to investigate this research question, we contend that the usability issues encountered by screen reader users on the original version of a website could differ from, and be less frequent than, those encountered on the partially localised versions. The other research question (**R3**) considers the impact of the degree of localisation on the usability of the studied multilingual websites. They read as follows:

R2: Are there specific usability issues related to partially localised websites that hamper an accessible browsing experience for screen reader users?

> **H1**: Usability issues vary depending on the navigation language chosen by the user.
>
> **H2**: There are less usability issues in the original language version than in the localised version(s) of a multilingual website.

R3: Does the use of different localisation strategies have an impact on the overall usability of multilingual websites for screen reader users?

Our study shared some characteristics with the classic usability test mentioned in the *Research foundations* chapter (section 2.3 - *Usability*). The five aspects listed by Dumas and Redish (1999, 22) have all been addressed in the present study: (1) the goal of the test was not to directly improve the usability of a product, because it was not commissioned by the websites' owners, but more to collect data on the language navigation patterns and increase awareness on how usability can be improved for disabled users; (2) the participants were representative of a group of potential users, namely screen reader users; (3) participants were asked to perform realistic tasks, in the context of a real-life scenario; (4) what they did was observed and noted and what they said was audiotaped; (5) the data was then analysed to identify the main problems.

The next section describes how the experiment was designed, which variables were measured and how we addressed the drawbacks of the chosen methods.

4.3 Experimental Design

In experiments, independent variables refer to the factors that the researchers are interested in studying or the possible "cause" of the change in the dependent variable, while dependent variables refer to the outcome or effect that the researchers are interested in. The primary goal of researchers is to study the relationship between dependent variables and independent variables and how changes in independent variables induce changes in dependent

variables (Lazar, Feng, and Hochheiser 2017, 30). To answer the research questions, we designed an experiment with two independent variables, namely language and localisation degree, and one dependent variable, which was the usability level.

Since our experiment investigated the impact on usability of two independent variables of different nature, we opted for a split-plot design (see Table 4.1). A between group-design was chosen to investigate the independent variable of language and verify the hypothesis of research question **R2** (groups 1-3 vs 2-4)**;** while a within-group approach, which effectively isolates individual differences, was adopted to compare the usability of the two websites and answer research question **R3** (1-2 vs 3-4) (Lazar, Feng, and Hochheiser 2017, 49–57). Table 4.1 shows that the language variable had two levels, German (DE) and not German (NON-DE) (see section 5.2 - *Data analysis approach*) and the localisation degree also had two levels, named after the websites that exemplify each category (Website A for 'reorganised' and Website B for 'imperfect mirror') (see sub section 4.5.1).

		Language variables	
		DE	NON-DE
Localisation degree	Website A	1	2
	Website B	3	4

Table 4.1 Split-plot design that shows the two variables: Localisation degree and Language

The two main drawbacks of the within-group design are usually fatigue and the learning effect. We tried to control the first setback by adopting two complementary strategies: instead of giving the participants a document to read and fill in, which would require even more interaction with the keyboard and screen reader, the researcher read the instructions and the questions aloud; secondly, the experiment session was designed to last less than two hours and a break of five to ten minutes was suggested between the two stages of the study (see section 4.6.3 - *Procedure*), as recommended by Lazar, Feng, and Hochheiser (ibid.). In addition, the fact that participants were not required to travel to an unknown location, thus avoiding a potential source of stress, was also indirectly aimed at controlling fatigue. The second drawback, the learning effect, is generally more conspicuous when the experiment compares, for example, the navigation effectiveness of two menus within a website, because the participant would gain a significant amount of knowledge on the website architecture during the first task. Our experiment was not particularly vulnerable to this phenomenon, because the content of the two websites was structured in a different manner. Nonetheless, both the order of the website and the order of 2 out of the 3 tasks designed (see sub-section 4.4.2) were counterbalanced. As in Petrie and Kheir (2007, 400) and Aizpurua, Harper, and Vigo (2016,

10), the reason why one of the tasks was always kept in the first position is that it was the more general one and it formed a meaningful scenario that began simply and then progressed to more complex tasks (see sub-section 4.6.1 for a detailed description of the tasks). Another strategy suggested by Lazar, Feng, and Hochheiser (2017, 55) to control the learning effect is to provide training before the session. While training on how to use the websites would have been counter-productive for our test, we encouraged the test users to take a couple of minutes to become acquainted with the homepage, instead of starting immediately.

4.4 Participants

Given the fact that this study is part of a master's thesis and due to time constraints, not obtaining a high number of participants was a risk that we had to accept. Lazar states that for research focusing on users with disabilities, it is generally acceptable to have 5 to 10 test users, and many research studies in the proceedings of the ASSETS Conference on Computers and Accessibility had 15 or fewer participants (Lazar, Feng, and Hochheiser 2017, 504–5).

4.4.1 Call for participation

A call for participation including a description of the scope of the study and a link to a recruitment form was drafted in three languages (DE - FR - IT) and sent via e-mail to associations and usability experts all around Switzerland. A 'snowball' sampling approach (Oates 2005, 98) was used for the recruitment process: some recipients were recommended to the researcher by Markus Riesch, Head of e-Accessibility of the Swiss Confederation, and Corinne Doret Baertschi, General Secretary of the association 'Ecoute-voir' based in Romandie, while other recipients were located on the Internet. A full list of the associations can be found in *Appendix A*.

To avoid responses from subjects that were not suitable for the test, first we had to identify the selection criteria. The target group of the experiment was defined according to the research questions: we wanted to find out how screen reader users navigate multilingual websites. First of all, it was important that our participants relied upon the screen reader to access the Web. We thought about introducing exclusion criterion, such as 'only completely blind people' or 'people who do not use a braille display', but we did not do so because neither a little bit of residual sight nor the use of a braille display would exclude the use of a screen reader as the main tool of navigation. Moreover, it would further limit the potential pool of participants, which we already believed would be small. In the call for participants we settled for 'blind or

visually impaired individuals who are familiar with screen reader software'. Another requirement was 'being a screen reader user for at least one year' as a way of avoiding complete beginners, whose results might have been strongly influenced by the lack of familiarity with this type of assistive technology. Secondly, since we were going to test Swiss websites that were available in German, French and/or Italian (see sub-section 4.5.2 – *Description of test websites*), we needed native speakers of those languages, so we added this requirement to the call for participation ('being a native or near-native speaker of German, French or Italian'). We accepted 'near-native' speakers because we anticipated a potential problem with Swiss-German dialect speakers: we know by experience that some of them consider standard German as their second language (even though it is the official language used in schools, in the media and by the public administration) and we did not want them to think they could not take part. We considered looking exclusively for participants with a Swiss passport or living in Switzerland, so as to ensure they would automatically speak one of the official languages, but there were a few problems with this approach: (1) it would include native speakers of Romansh, a national language but not an official one[28], which was outside the scope of our study; (2) it would include immigrants with a different native language and exclude native speakers who live abroad or just over the border. The last requirement stated in the call for participants was 'being aged 18 or over'. No superior limit was set in the hope of obtaining a sample as representative as possible of the screen reader community, considering the small size of the study.

Once we had established the participation requirements, we drafted an email that mentioned such criteria, the context and the aim of the study, and a brief description of the test session with its estimated time. Interested candidates were asked to click on a link to a recruitment form that included a screening questionnaire and an informed consent form (see section 4.4.2 – *Recruitment form*). Finally, the readers were encouraged to share the call for participants and they were given the email address of the researcher in case they had any questions. The text of the call for participation can be found in *Appendix B*.

4.4.2 Recruitment process

The call for participation contained a link to a recruitment form. Initially, we had planned to use Google Forms, but a quick accessibility test performed by a screen reader user showed that there were some issues. We therefore turned to a tool that reportedly allowed users to build

[28] https://www.admin.ch/opc/en/classified-compilation/19995395/index.html#a4 and https://www.admin.ch/opc/en/classified-compilation/19995395/index.html#a70 Last access: 22th June 2018

Section 508-compliant[29], more accessible forms, namely SurveyGizmo[30]. The final version was tested by the same person who carried out the methodology test (see section 4.6.2 – *Methodology test*) and it proved accessible. Moreover, a free SurveyGizmo account allowed us to open three different forms and collect up to 100 responses, which was enough for our study. There were some limitations on the type of questions (for example, the option 'skip/disqualify' based on the answer was not available), but we managed to resolve these issues.

The recruitment form was divided into three sections. The first one was a screening questionnaire, where participants were asked to confirm that they corresponded to the user profile we were looking for. When asking to confirm that they were 18 or older, we introduced a distinction between under and over 40 years old, in order to get an idea of the respondents' age range before the actual test without appearing too indiscreet at an early step. We also asked to confirm that they had been using a screen reader for one year or more, and that they were native or near-native speakers of German, French or Italian. Finally, we asked where they would be available to meet the researcher to carry out the study. The second section contained a detailed description of the test session. Participants were informed that they were going to be asked to perform several tasks and report any frustrating experience, as well as to express their opinions and answer a few general demographic questions. They were also reassured of the fact that the data collected would be treated confidentially, that there were no potential risks anticipated and that they could withdraw from participating at any time. They were informed that participation was voluntary and the researcher would be taking notes and audio-taping the session. Finally, we briefly mentioned that the results might have 'a positive impact on the whole screen reader community by helping to improve awareness about the issue of universal usability when translating a website'. The third section was a consent form in the shape of a series of statements to which it was possible to agree or disagree individually. On this topic, Saldanha and O'Brien (2013, 48) write

> […] the adequacy of one 'I Consent' or 'I Agree' button is questionable. How can we be certain that the participant has actually read the information? While it is impossible to know this for certain, one recommendation is to break the consent down into specific and separate statements and to 'force' participants to agree to each one in turn (ibid.).

Since it was an online form and our target participants were visually impaired, we opted for asking them to enter their full name and the day's date instead of signing. Finally, we asked

[29] see section 1.1.2 – *Legal basis*
[30] https://www.surveygizmo.com/

them to provide their email address in order to be able to contact them. We repeated the same question twice to prevent spelling mistakes. The recruitment form can be found in *Appendix C*.

The recruitment form, available in German, French and Italian, was opened 20 times from the 27 April to the 31 May 2018. Some responses were partial, others where double. The researcher contacted every interested person who complied with the requirements, provided a valid e-mail address and consented to take part. In the end, a total of ten participants partook in the usability study, which took place over a period of one month, from the 8 May to the 7 June 2018.

Each participant was assigned an individual code for anonymity purposes that was composed of two digits and two letters. The letters represented the language code, while the numbers the order of the sessions. Since some of the participant were almost bilingual and we sent a multilingual e-mail containing links to three different language versions of recruitment forms, the language in which the respondents completed the recruitment form was sometimes different from the communication language used with the researcher, the instructions language chosen during the session or their preferred navigation language. However, the codes were assigned to participants according to the language they indicated as their mother tongue. The counterbalanced order was based on those codes, to guarantee that any learning effect was equally compensated among all native speakers, as shown in Table 4.2.

Code	Tasks order
01DE	A-1-2a-2b / B-1-2a-2b
02DE	B-1-2b-2a / A-1-2b-2a
03DE	A-1-2b-2a / B-1-2b-2a
04DE	B-1-2a-2b / A-1-2a-2b
05DE	A-1-2a-2b / B-1-2a-2b
01FR	A-1-2a-2b / B-1-2a-2b
02FR	B-1-2b-2a / A-1-2b-2a
03FR	A-1-2b-2a / B-1-2b-2a
01IT	A-1-2a-2b / B-1-2a-2b
02IT	B-1-2b-2a / A-1-2b-2a

Table 4.2 Participants' codes and counterbalanced order

4.4.3 Participants' profile

The following paragraphs describe the profile data that we gathered through the post-test interview (see section 4.6.3 – *Procedure*).

4.4.3.1 Demographic profile

The questionnaire at the end of the interview (see section 4.6.3 - *Procedure*) revealed that the mean age of participants was 45.1 years old (SD=14.3), with the youngest being 25 and the

oldest 63. The education background was quite varied and sometimes difficult to classify, as the education system has changed substantially over the past 30 years. Seven participants (N=7) declared themselves as holding a university degree or an equivalent certificate, while three (N=3) indicated that they held a secondary school diploma or had completed a traineeship. All the participants were either students (20%) or professionals (80%). Three people (N=3) had a job as accessibility experts or testers, three (N=3) as regular employees, one (N=1) as an intern and one (N=1) as an IT expert.

4.4.3.2 Expertise

The people who declared themselves as accessibility experts were consistent and said they were experts on the subject in the question about familiarity with accessibility. The other participants (N=7, 70%) reported knowing what accessibility is and having some experience with it. All of the people interviewed reported using a computer to browse the Internet at least once a day and declared that they had been using a screen reader for more than 5 years. We introduced these questions to take into account the recommendations by Aizpurua et al. (2014) that were described in the *Related work* chapter (sub-section 3.2.2).

4.4.3.3 Language skills

Half of the participants (N=5) indicated German as their mother tongue, 30% (N=3) indicated French and 20% indicated Italian (N=2). When asked for a self-assessment of their language level according to the European Framework[31], 60% of the participants (N=6) stated that they had a C1 or C2 level of passive knowledge of at least one other language. All participants stated they had at least an A1 or A2 level of passive knowledge of at least one language other than their mother tongue. The following table (Table 4.3) provides an overview of the participants' skills. The language competence section uses the code 'M' for mother tongue, 'C' for a C1 or C2 level, 'B' for a B1 or B2 level and 'A' for an A1 or A2 level.

[31] https://rm.coe.int/168045bb52 Last access: 30 April 2018

	Internet				SR			Accessibility				Languages					
	less 1/week	1-2/week	More 2/week	Daily	1/2 yrs	3-5 yrs	5+ yrs	None	Little	Some	A lot	German	French	Italian	English	Russian	Spanish
01DE												M	A	C	B		
02DE												M	C	C	C		A
03DE												M	A	B	C		
04DE												M	B		C		A
05DE												M	B	A	B		
01FR												B	M		A	B	
02FR												A	M	A	C		
03FR												B	M		B		
01IT													A	M	A		
02IT												A	C	M	C		B

Table 4.3 Summary of the study participants' skills

4.5 Usability study materials

The best way to achieve the goals illustrated in section 4.2 of the present chapter was to select two websites that had been partially localised using different strategies. Switzerland seemed particularly suited for this purpose because the country has four linguistic regions and three official languages.

4.5.1 Selection of websites

For the selection of websites, we took as a reference the 2016 Swiss Accessibility Study (Access for all 2016) described in sub-section 3.5.2 – *Accessibility of Swiss universities' websites*, as it seemed particularly suitable to be used as a starting point for the present work.

The results of the test are presented in the Access for All report as shown in Figure 4.1.

Figure 4.1 Example of the presentation of the test results in the Accessibility Study (Access for all 2016, 59)

(1) An overall mark is expressed on a scale of 1 to 5 stars, where 5 is the best mark. The websites that scored 1 or 2 are considered fundamentally inaccessible for people with disabilities, while websites that scored 5 can be considered relatively easy to use, even though they might not completely conform to the highest level of accessibility guidelines.

(2) A table shows the accessibility profile, where the evaluation criteria are classified into 12 categories that focus on: non-textual content, keyboard accessibility, logical order, semantic structure, multimedia, display adaptability, contrast, understandability, coherence, compatibility, assistance and PDF accessibility.

(3) The scenario is described at the bottom, next to the experience report. The first one is an ordinary activity or a goal that any user of the website might be interested in. The second one is a description of the navigation experience.

Since most websites contained numerous web pages, sometimes reaching the thousands, the Study only tested four pages for each website. In addition to the homepage and the contact page, at least two other pages were selected according to a pre-defined scenario in order to be tested against an accessibility checklist created by the Foundation and based on the WCAG 2.0. The scenario-oriented testing method was an effective alternative to make up for the small sample, so we decided to use a similar approach for our study (as described in sub-section 4.5.1). Another overlooked aspect was that of languages: the topic is barely mentioned throughout the study (only twice in different experience reports, pages 107 and 133). However, Pascale Gazareth touches upon this aspect in her article *À qui profite la meilleure accessibilité de l'internet?* ('Who benefits from an improved Web accessibility?') (Access for All 2016).

She argues that the implementation of simplified language would help immigrants and people from other Swiss linguistic regions understand the content of a website that is not available in their native language (Access for all 2016).

While this is a fascinating and reasonable theory, for the purpose of our study we were more interested in finding out how the experts dealt with multilingual websites. When examining the report, our hypothesis was that the main working language was German, but we wondered how they tested websites that did not have a German version (such as, for instance, the University of Geneva website). We contacted Dr Anton Bolfing, in charge of research and development at the Foundation, and he replied to our message confirming our hypothesis. In addition, he explained that most of the 61 success criteria they used are independent from human language and focus instead on tagging and programmatic semantics. Those criteria that are dependent on human language (such as 1.1 and 1.2 'are there any meaningful alternatives for non-text and time-based contents?', 1.3.3 'textual references to sensor properties', 2.4.4 'link purpose' and 2.4.6 'meaningful and clear headings and labels') were tested on non-German sites by people with basic knowledge in the relevant languages. Specifically, two of the testers were bilingual Italian-German or trilingual (French too) and all of the staff had learned basic French in school. Their linguistic competence was considered sufficient to assess the meaningfulness of headings, link purposes and the like, especially because the study was not focused on multilingualism or internationalisation. As a tip, he suggested looking at the screenshot that accompanies every evaluation to identify the testing language.

We collected the list of 98 websites in a table that contained their respective overall marks, and we excluded those that had scored 1 or 2 stars because they might have been completely inaccessible for our test users, undermining the experiment itself. Similarly, we ruled out those that scored 5 stars because they were likely to be too good to observe many issues. We then proceeded to a manual inspection of the 54 websites that had scored 3 or 4 stars to determine which ones were at least bilingual in two of the three Swiss official languages. For all of the inspected sites, it was relatively easy for a regular sighted user to find the language selector at the top or at the bottom of the homepage. At this stage, we did not check the localised versions of the websites, but only looked at the available languages.

The next step was to select the websites that were partially localised, so we developed a categorisation based on the levels of localisation analysed in the *Research foundation* chapter (sub-section 2.2.2 – *Levels of web localisation*). In fact, while analysing the proposals put forward by Yunker (2003), Singh and Pereira (2005) and Jiménez-Crespo (2012), we soon realised that we could not really apply them to Swiss websites. Yunker's approach is more

focused on the strategy of the company (start small vs re-design everything) (Yunker 2003), but it fell outside the scope of our study. Singh and Pereira (2005) use the concept of 'foreign subsidiaries', a partially or wholly owned company that is part of a larger corporation with headquarters in another country[32], which makes little sense when talking about different linguistic regions within the same country. In fact, we believe the challenges found when trying to apply the existing localisation categories relied on two major reasons. First of all, the peculiarity of the country: Swiss websites need to be translated to reach the country's entire population, but they might not need much cultural adaptation. As de Grandi (2017, 29) explains in her thesis about localisation in advertising, Switzerland's linguistic variety comes with small differences in the customs and habits of the inhabitants of the different linguistic regions, which are similar to that of Germany, France or Italy, but, as a whole, Switzerland shares traditions and elements that differentiate it from the neighbouring countries. Secondly, most scholars' classifications adopted an industry perspective, but most websites on the Web for All study were not, strictly speaking, companies. State departments, cantons and cities are part of the public administration, and universities are somewhere between the public and the private sector; schools are considered public institutions, but universities are not part of the compulsory education. Jiménez-Crespo's classification, on the other hand, offers some interesting elements, such as the concept of mirror websites (Jiménez-Crespo 2013) that we adapted for the purposes of this thesis. We finally identified two main levels of localised websites: (1) 'partially localised', where some of the content has been translated and/or adapted, but it is still possible to find source language elements in the target language version, or links in the target language version that open source language pages; and (2) 'fully localised', where virtually all content had been translated, with the possible exception of some PDF documents. Within the first level, we then distinguished between two localisation degrees: (a) 'reorganised', where the content has been summarised and the structure reorganised; and (b) 'imperfect mirror', where the content has been reduced in volume, but the structure remains the same. Table 4.3 summarises this classification.

Partially localised		Fully localised
Reorganised	Imperfect mirror	
Website A	Website B	x

Table 4.4 Recapitulative table of the localisation levels' classification adopted for the purpose of this thesis

[32] http://www.businessdictionary.com/definition/foreign-subsidiary-company.html Last access: 24 June 2018

A second manual inspection of the remaining websites allowed us to exclude fully localised sites because they fell outside the scope of our research study. Applying this third criterion left us with six websites: the canton of St. Gallen, the canton Zurich, the canton Zug, the city of Lucerne, the Zurich University of Applied Sciences and the Bern University of Applied Sciences. The website of St. Gallen (sg.ch[33]) was dismissed because it did not fit into any of the two localisation categories; its language selector offered many languages, but the user was redirected to a page that explained that the content was not available. The website of the city of Lucerne (stadtluzern.ch[34]) was excluded as well because it offered audio files in 11 languages, but no navigation possibilities in French or Italian. Eventually, we opted for the two university websites (zhaw.ch[35] and bfh.ch[36]) because, according to our classification, they belonged to two different localisation degrees within the 'partially localised' level.

As mentioned in section 3.5.3 – *Localisation and internationalisation*, Fernández Costales (2012) states that the adaptation level of university websites generally does not allow us to consider them as examples of localisation. If we consider, for instance, Dunne's definition of localisation (see section 2.2), the two chosen websites could be considered as not 'localised', because there is no "adaptation of non-textual content (from colors, icons and bitmaps, to packaging, form factors, etc.) […] to take into account the cultural, technical and regulatory requirements of that locale" (Dunne 2006, 4): the colours do not change in the different versions, the currency and date settings are the same across Switzerland and so are the regulatory requirements. But we argue that, since we adopted Jiménez-Crespo's definition of Web localisation, "interactive digital texts are modified to be used in different linguistic and sociocultural contexts, guided by expectations of the target audience and the specifications and degree requested by initiators" (see section 2.2), we can state that our test websites were indeed examples of localised websites.

4.5.2 Description of test websites

Before proceeding to select specific tasks to perform on the two websites, the researcher thoroughly browsed them, in order to understand the structure and the kind of information architecture they offered. Both websites display a homepage with a language selector at the top

[33] https://www.sg.ch/ Last access: 28 November 2017
[34] https://www.stadtluzern.ch/ Last access: 28 November 2017
[35] https://www.zhaw.ch/ Last access: 25 April 2018
[36] https://www.bfh.ch/ Last access: 25 April 2018

of the page and multiple links to other sections such as ‘studies’, ‘news’, ‘continuing education’, ‘library’, ‘jobs’ and a research field.

The website for Zurich University of Applied Sciences (ZHAW) – hereafter ‘website A’ – presents a simple and clean structure, and looks quite modern with its scroll-down design. The homepage has two internal tabs, ‘university’ (the default one) and ‘select schools’, and it contains the contact information at the bottom. The language selector offers four languages (German, English, French and Italian), but a quick look at the French and Italian versions is enough to understand that the content has been re-organised and has not been entirely translated. The user is informed by a message that says “*Nos contenus se déclinent, pour une grande partie, en allemand et en anglais. Vous trouverez néanmoins de nombreuses pages en français*” (‘our content is mainly available in German and English, but you will find numerous pages in French’) and a similar message is to be found in the Italian version. The university has eight departments: Applied Linguistics, Applied Psychology, Architecture, Design and Civil Engineering, Health Professions, Life Science and Facility Management, School of Engineering, School of Management and Law and Social Work. Only the Applied Linguistic department has translated most of its content into French and Italian. For these reasons, we classified it as ‘reorganised’.

The website for Bern University of Applied Sciences (BFH) – hereinafter ‘website B’ – is available in three languages, two of which (German and French) are very similar in content and structure, whereas the third one (English) is slightly different. We classified it as ‘imperfect mirror’ because we were only considering German and French. The contact information is available through a link at the top of the page and all versions contain a list of quick links and direct links to the departments: Architecture, Wood and Civil Engineering; Health Professions; School of Agricultural, Forest and Food Science; Bern University of the Arts; Social Work; Engineering and Information Technology; Business; Swiss Federal Institute of Sport.

4.6 Usability study environment

The researcher decided that going on-site to assist the participants and take observation notes would be the best option. In case of a low response rate from participants with disabilities in the local area, the test could be adapted to distributed research, in order to be carried out remotely (Lazar, Feng, and Hochheiser 2017, 504–5), but eventually this adaptation was not deemed necessary. Field testing has certain advantages over lab testing, such as a better understanding of the user’s workspace and tools, but it also has limitations, including the

inability to control the environment (e.g. distractions, lack of privacy) (Barnum 2011, 39–41). Given that the target users of the study are accustomed to a particular setting in terms of the use of assistive technology and their mobility would have required a bigger effort than usual, lab testing was excluded. The potential drawbacks of field testing were addressed by asking the participants to meet the researcher in a quiet room of their choice (home, workplace or public space). With regard to the context, Switzerland was the perfect setting to carry out the study for the following reasons: as a plurilingual country, it is to be expected that at least some of its websites are multilingual; the researcher was acquainted with the country's linguistic composition, having proficient knowledge of German, French and Italian; and it was geographically convenient for the purposes of conducting on-site tests.

It is important to mention that all documents were drafted in English and the relevant ones were translated into German, French and Italian. The researcher personally translated the material into her mother-tongue (Italian), and post-edited the translation output of DeepL[37] into French and German. All the texts intended for participants were revised by native or bilingual translators or university students.

4.6.1 Scenarios and tasks

In order to test the websites, we decided to guide the users' navigation by asking them to perform some tasks described in a scenario. We opted for a scenario-oriented method, similar to the one adopted for the 2016 Swiss Accessibility Study, because scenarios make the tasks more realistic and "take some of the artificiality out of the test" (Dumas and Redish 1999, 172–73) by identifying the objective and providing the users with any background information they would have in a real-life situation (ibid.). According to the same authors, usability testing is a sampling process, and it is not possible to test every single task. When sampling tasks, a researcher should select those that users will actually perform with the product and that could potentially cause usability problems (Dumas and Redish 1999, 160).

We established that three tasks for each website would give us enough data without requiring too much effort from the test users in terms of time or fatigue. This number of tasks was also in line with some similar studies we were aware of, such as Power et al. (2012); Aizpurua, Arrue, and Vigo (2013); Aizpurua et al. (2014); Vigo and Harper (2013). Given the fact that university Web portals tend to be more informative than interactive, the tasks we designed mainly involved looking for specific information. The choice of tasks was based on

[37] Online Machine translation tool: https://www.deepl.com/translator

the navigation of a sighted user (the researcher), but was later validated by a blind user during the methodology test (see sub-section 4.6.2). In order to introduce some variety, we defined three levels of difficulty: (i) easy, (ii) intermediate and (iii) difficult/impossible. We then used those categories, together with the scenario and our intuition about potential issues, to select the six tasks described in the following section.

4.6.1.1 Website A

When testing the ZHAW website, participants were asked to imagine that they were prospective students who lived in Zurich and were hesitating between two very different faculties, the School of Applied Languages and the School of Engineering.

The first task (**Task A-1**) was to find a generic ZHAW e-mail address contact and telephone number). Since both pieces of information were at the bottom of the homepage, the researcher anticipated no particular difficulty and classified the task as 'easy'. The second task (**Task A-2a**) was to find the date, time and location of the open days of the Bachelor of Applied Languages. This task presented some difficulties and was classified as 'intermediate' because it required visiting a number of sections before reaching the faculty's list of events, and while the path to get to the list was available in all four languages, the list itself was only in German. Finally, to introduce at least one interactive task, users were asked to look for the newsletter of the School of Engineering (**Task A-2b**). The path was quite complicated and only available in German and English, while the form to fill in to sign up for the newsletter was only in German. Consequently, this task was classified as 'difficult/impossible'.

4.6.1.2 Website B

For the BFH website, participants were asked to imagine that they were prospective students from Bienne, and that they were interested in the Bachelor offered by the Swiss Literary Institute (SLI). Unlike website A, this time the scenario focused on just one faculty, the Bern University of the Arts (HKB), to which the SLI is attached.

The first task (**Task B-1**) that was classified as 'easy' was to find the e-mail address and telephone number of the Swiss Literary Institute. This task was similar to the corresponding task on website A, but required more clicks to be completed, as the homepage of the BFH does not offer any contact information. For a sighted user, a minimum of two clicks is required to find the telephone number of the institute and the e-mail is only available through a link that opens directly into Outlook. The second task (**Task B-2a**) was to find the open days of the Bachelor in Literary Writing of the SLI. This task was introduced, once again, for interactivity because the quickest path would lead the user to an events section that contained filters with

scroll-down options. The piece of information itself was impossible to find, as there were no events in the interval required. The task, classified as 'difficult/impossible', was nonetheless interesting because it allowed the researcher to observe the reactions to missing information. The third task (**Task B-2b**) was to find the address and opening times of the library of the SLI. While the address was relatively easy to find for a sighted user, the opening time was not clearly indicated and had to be found in a descriptive paragraph. We therefore classified it as 'intermediate'.

4.6.2 Methodology test

Before starting the testing sessions, a methodology test was conducted remotely on the 25 April 2018 with a blind Italian-speaking translator based in Italy. The objective of this test was twofold: firstly, to determine whether the instructions were comprehensible for a blind user and the tasks were feasible; secondly, to verify that the estimated time was correct. During the meeting, which took place over the phone, the researcher read the instructions to the user, asked him to try and complete the tasks and to share his thoughts about both the tasks and the test websites.

The methodology test proved useful in achieving the first goal, because it allowed the researcher to improve the clarity of the instructions by making some minor adjustments, as well as to verify that the process ran smoothly. Unfortunately, due to time constraints, it was not possible to complete the whole interview, so the time estimation could not be verified.

4.6.3 Procedure

The test sessions started with the researcher arriving on site, greeting the participant and setting up the apparatus. Participants were asked to use their own laptops or computers. Whenever they owned more than one (for example, one for work and one for personal use), they were asked to use the one they would pick to browse the Web. The researcher used her laptop, a USB microphone and *Audacity*[38], a piece of software to record the test session. When everything was ready, the researcher started the recording and began reading the instruction aloud (*Appendix D*). Participants were told their skills were not being tested, as recommended by Aizpurua et al. (2014) (see sub-section 3.2.2), and they could ask the researcher to repeat at all times. They were also asked to report any frustrating experiences they might encounter. Then, the researcher read the instructions for the first task and let the participant try to complete

[38] https://www.audacityteam.org/

it. Every time they succeeded or gave up, she read the instructions for the next task. Participants were free to speak or remain silent, but they were encouraged to explain what was wrong each time the researcher noticed that they were stuck with a particular task. Their reactions and most of their navigation paths were observed and noted by the researcher.

The tasks performed on the two websites were followed by two identical semi-structured interviews aimed at collecting further data to complement observation notes. These post-task interviews were divided into three sections.The first section was focused on gathering technical data (see sub-section 4.6.4) such as information about the operating system, the browser, the assistive technology used during the testing session, as well as previous experience on the website.

The second section consisted in an adapted Computer System Usability Questionnaire (CSUQ). This specific questionnaire, aimed at measuring the usability of a system, was chosen over other popular usability questionnaires, such as SUS (System Usability Questionnaire), because it included some task-related statements that the other ones lacked. Rodríguez Vázquez, O'Brien, and Fitzpatrick (2017) used the same questionnaire for a study on the usability of post-editing environments, which was also a usability study with blind users. This questionnaire contains 19 statements to which the user rates satisfaction on a seven-point Likert scale. The results can then be regrouped into four categories: 'system usefulness', 'information quality', 'interface quality' and 'overall satisfaction' (Albert and Tullis 2013, 140–41). Since the original questionnaire was conceived to rate the usability of systems, it was necessary to adapt it for website rating and, more importantly, the abilities of the test users. After replacing the word 'system' with 'website', we replaced 'information' (which refers to online help, on-screen messages and other documentation for systems) with labels and menus. Statement 8 'I believe I became productive quickly using this website' was not applicable, as participants were not really required to 'produce' anything, so we changed it to 'I believe I could quickly learn how to use this website', on the lines of item 7 ('It was easy to learn to use this website'). We also modified statement 15 ('the organisation of information on the system screens was clear') to 'the structure of the website was clear' and replaced the word 'interface' with 'navigation' in statement 16 and 17, because it better describes the experience of a blind person browsing a web page. The adapted CSUQ can be found in *Appendix E*.

The third section began with a question on the level of frustration while navigating the website, with a rating that varied from 1 (not frustrating at all) to 7 (extremely frustrating). The next question asked participants to estimate the number of minutes they had lost because of the problems they encountered. Finally, test users were prompted to express any additional

comments on the tasks or website. At the end of the second post-task interview, the participants were asked a few demographic questions such as their age, highest level of education, and current occupation. This post-test interview also included questions about their expertise with screen readers and the Web, as well as their passive language skills and can be found in *Appendix F*. The recording of each testing session was later transcribed by the researcher using *Express scribe*[39].

4.6.4 Use of technology

Before proceeding to present our findings and discuss how they provide evidence to answer the research questions that we put forward in the previous chapters, we deemed it relevant to briefly describe the technical setting of the test sessions. We collected these data through the post-task interviews.

4.6.4.1 Website A

As illustrated in the following table (Table 4.5) nine participants out of ten (N=9) used Windows (N=5 Windows 7, N=3 Windows 10, N=1 Windows 8). Only one user (N=1) chose MacOS as an operating system. That same user (N=1= browsed the web with Safari, while the others were almost equally distributed between Internet Explorer (N=6) and Mozilla Firefox (N=4), with one user (N=1) switching browser in the middle of a session, as a coping strategy to solve a problem (see sub-section 5.4.2). JAWS was the preferred screen reader for almost all participants (N=9), except for the Mac user (N=1), who chose VoiceOver. The use of assistive technology was quite balanced, as 40% of the participant (N=4) used a braille display (to various degrees) during the test, while 60% (N=6) did not. Some users mentioned having one but did not use it for the test. Unlike the other factors considered, we noticed that a clear prevalence of braille display users (three out of four) were French-speaking users (N=3). One of them explained that they switched from the screen reader to the braille display whenever the screen reader accidentally began to read content with the wrong pronunciation. Finally, 70% of the participants (N=7) did not remember ever visiting the ZHAW website before, and the three (N=3) who did (all German-speaking), had only browsed it once or twice.

[39] https://www.nch.com.au/

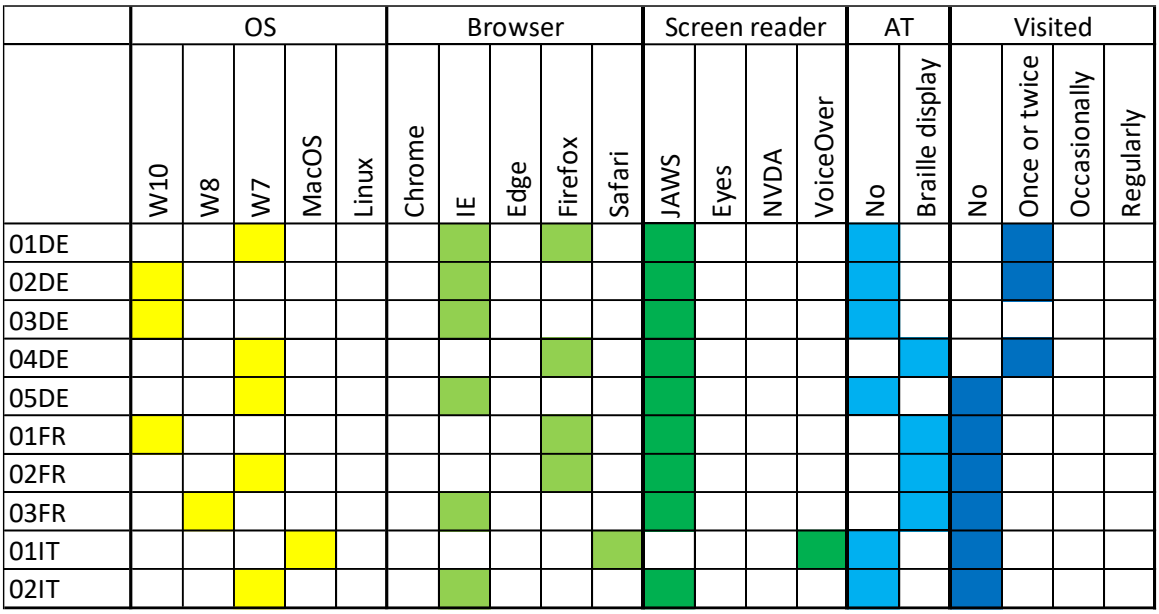


Table 4.5 Settings and previous experience of participants on website A

4.6.4.2 Website B

Table 4.6 illustrates the settings of the navigation on the BFH websites. The distribution of the operating system remains the same as for website A: 40% (N=4) Windows 7, 30% (N=3) Windows 10, 10% (N=1) Windows 8 and 10% (N=1) MacOS. Internet Explorer was the preferred choice for six out of ten participants (N=6) to navigate on website B, while three users (N=3) chose Firefox and one (N=1) used Safari. As for the previous website, nine participants (N=9) used JAWS and one (N=1) VoiceOver and four out of ten participants (N=4) used a braille display. Nobody remembered visiting the website before.

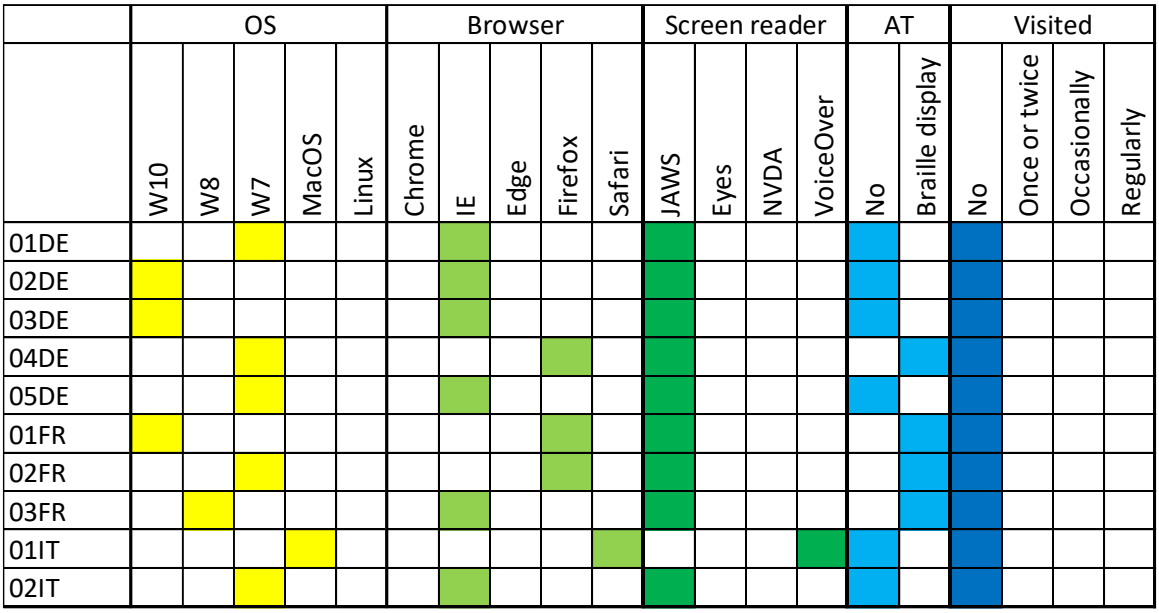


Table 4.6 Settings and previous experience of participants on website B

5 Findings and discussion

5.1 Overview

In this chapter, we present the data collected throughout the experimental study, discuss the main results obtained and analyse them to answer our research questions. Section 5.2 illustrates the mixed-method approach selected to analyse the data. Section 5.3 examines the first research question **R1** and its sub-questions by focusing on the language navigation patterns of the participants. In section 5.4, a thematic approach is adopted to investigate the type and quantity of issues encountered by participants, and answer research question **R2**. The third research question **R3** is examined in section 5.5, where we focus on usability and compare the two websites. Finally, section 5.6 briefly summarises the findings of the previous sections.

5.2 Data Analysis Approach

Throughout our usability study, we recorded a total of 20 navigation experiences by 10 participants on two websites. The experiment provided a rich dataset that contained different types of data. Our observation notes and the transcripts of what participants said while completing the tasks provided qualitative data that would help us investigate **R1** and **R2**. The post-task usability questionnaire and the questions at the end of the interview provided us with quantitative data that would help us answer **R3**. Finally, we collected demographic data to obtain a profile of the participants and asked some questions about skills and previous experience, in order to exclude potential biases, such as being a frequent user.

The questions about language skills were particularly relevant as language was one of the independent variables that we aimed to study. Given the low number of Italian-speaking participants, for the analysis of the results we decided to consider all the non-German native speakers as a group. This approach was deemed appropriate because: (1) the number of the two groups would be equal and (2) it made sense to compare the original version (German) to the localised versions (French, Italian and, occasionally, English).

A multi-method approach was deemed necessary to deal with the variety of data sources. Even though the small number of participants did not allow us to run an advanced statistical analysis, we opted for a statistical descriptive approach for our quantitative data that would allow us to present the results in a clear way. To analyse the qualitative data, we opted for a

thematic approach similar to the one used by Rodríguez Vázquez and O'Brien (2017) because it is a valid approach to deal with the data emerged from interviews, as it has been widely used in prior Web accessibility-related studies to "identify, analyze and report patterns within the data collected" (ibid. 245). This method consists of six stages: (1) familiarizing with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes and (6) producing the report (Braun and Clarke 2006, 87). These steps were followed with a certain degree of flexibility aimed at adapting the method to the variety of data sources. In the first stage, we transcribed not only what the participants said, but also our annotations on their reactions and navigation paths. This approach allowed us to identify 69 problematic situations that we later classified in an increasingly abstract way (stage 2). Afterwards, we identified the main themes according to our hypothesis (see section 4.2) and a second coder was asked to review a sample of the transcripts against the coding scheme defined by the first coder[40], a method used by Rodríguez Vázquez and O'Brien (2017, 246) in the above-mentioned study that relied on thematic analysis (see sub-section 3.4.1). A high level of agreement was reached by the two coders and differences in coding were discussed (stages 3 and 4). This exchange contributed to the fifth stage, that consisted in defining and naming themes. Section 5.4 presents our findings (stage 6).

5.3 Language navigation patterns

The findings reported in this section will shed light on the language navigation patterns adopted by the screen reader users who took part in the test sessions. We hope that our analysis will contribute to answer our first research question:

R1: What are the language navigation patterns of screen reader users browsing on partially localised websites?

To investigate this topic, we broke it down into four questions that consider different aspects:

R1.1: Which language did users instinctively choose?

R1.2: How did users initially switch to their preferred language?

R1.3: Why was the language changed during navigation?

R1.4: What were the language skills of the users who voluntarily browsed in a different language?

[40] Since it was not possible to have all the data coded by more than one person, having a small percentage of the data recoded can still provide a good indication of how reliable the coding has been (Saldanha and O'Brien 2013, 193).

5.3.1 Language preferences

Before taking a closer look at the language choices of the participants, it is important to recall what the options were. If we consider the location of the universities whose websites were tested, namely Bern and Zurich, it is reasonable to assume that the original and most complete version of both websites was German. Website A offered an English version with a similar structure to the original, but we did not consider it because English was outside the scope of our study. We classified the Italian and French localised versions of website A as 'reorganised' because the structure changed considerably and a message informed the user of the lack of some content (see sub-section 4.5.2 – *Description of test websites*). Website B, on the other hand, did not offer any Italian version, but only an English version (which, again, we did not consider) and a French version, which we classified as 'imperfect mirror' (see section 4.5.1 – *Selection of websites*).

For the abovementioned reasons, the language navigation pattern of native German-speaking participants (N=5) was very straightforward: when they inserted both addresses in the address bar at the top of the browser, they landed directly on the German homepages and did not change the language. The language navigation pattern of non-German native speakers was more diverse. French speakers were positively surprised when they landed correctly on the French homepage of website B. For example, participant 01FR said '*Ok, en français : magnifique ! Comment ça se fait ? C'est peut-etre mon adresse IP qui m'a mis directement en français ? C'est pas mal. Ca c'est un très bon point'.* Yet, all of them (N=3) changed the language to French when they landed on the German homepage of website A. This is hardly surprising, considering that none of them reported having more than a B level of German (see Table 4.2 - *Summary of the study participants' profile*). Finally, the two Italian-speaking participants (N=2) had different reactions when they landed on the German homepages of both websites. User 01IT, who reported an A level of English and French and no knowledge of German, switched to Italian when it was available (website A) and to English when Italian was not an option (website B). Participant 02IT, who reported a C level of French and English and an A level of German, switched to French on both websites. When asked whether they had noticed that Italian was available on website A and why they did not select it, they replied that they had noticed, but French came first [the 'fr' label was read before the 'it' label by the screen reader] and they were used to it [as they had been living in a francophone country for years]. Table 5.1 summarises the behaviour of participants when they landed on the homepage of each website.

		01DE	02DE	03DE	04DE	05DE	01FR	02FR	03FR	01IT	02IT
Website A	Landed on	DE	DE	DE	DE	DE	DE	DE	DE	DE	DE
	Switched to	x	x	x	x	x	FR	FR	FR	IT	FR
Website B	Landed on	DE	DE	DE	DE	DE	FR	FR	FR	DE	DE
	Switched to	x	x	x	x	x	x	x	x	EN	FR

Table 5.1 Language preference of participants by native tongue

To answer question **R1.1**, all participants but one (N=9) browsed the language version corresponding to their mother tongue. Whenever it was not available, they selected the version in a language of which they had at least some passive knowledge. One participant (N=1) selected a foreign language version (the language of the country where they had been living for years) even though their native tongue was available. While the small sample does not allow us to draw conclusions on the language navigation patterns of the whole screen reader community, our study provides clues that could guide further investigation on the language preferences of multilingual screen reader users. Also, the positive comments of the French native speakers suggest that perceived usability improves when the users land on their preferred language version. Localisers should be aware of this and implement the necessary techniques to make it happen. For instance, they may use IP tracking to select the geographical region and/or auto-detection of the user's browser language so that the website can be automatically served in that language (Miraz, Ali, and Excell 2013, 238), although other solutions may be possible as well.

5.3.2 Language selector

Another aspect of language navigation patterns that needed to be examined to answer question **R1.2** was how users switched to their preferred language. A previous exploratory study by Rodríguez Vázquez (2015) on the main difficulties faced by screen reader users on multilingual websites and their coping techniques identified the language selector as a potential accessibility barrier. In the tested websites, the most common issues with the language selector that Rodríguez Vázquez found, such as embedded links or flags without the appropriate text alternative, were not present. For website B, 80% of the participants (N=8) were satisfied with the language version they landed on, as it was their native tongue, so they did not attempt to change the language. The two Italian-speaking participants (N=2), who landed on the German homepage, easily found the language selector while they were exploring the homepage, before beginning with the actual tasks. They did so by letting the screen reader sequentially read the list of links. For website A, only 50% of participants (N=5) landed on the preferred version. Four out of five non-German-speaking users (N=4) found the language selector by simply exploring the homepage and reading the content sequentially. One participant (N=1) had to look

'fr' up in the screen reader's link list, a coping strategy that was also mentioned in the study by Rodríguez Vázquez (ibid.).

To answer question **R1.2**, in the context of this study, we observed two main strategies for changing the language at the beginning of the session: reading the list of links until the preferred language code is found or actively searching it in the list of links.

5.3.3 Changing the language as a coping strategy

In this sub-section we explore the circumstances that led some participants to change the language *during* the session and try to answer question **R1.3**. While German-speaking participants browsed in German all the time, the language navigation pattern of non-German native speakers was less straightforward. We do not consider here the instances when a page unexpectedly opened in another language, as we will look at those cases when trying to answer our second research question (see sub-section 5.4.1 – *Language related issues*). Now we focus instead on the instances in which the language was changed voluntarily. Specifically, we want to investigate *why* users have recurred to switching to a language other than their native tongue.

Out of five non-German native speakers, two (N=2) voluntarily switched to German in the middle of the session. Participant 01FR did it during task 2a on website A, after repeatedly finding himself on the German version while looking for information on the French version. Participant 01IT did it during task 2b on website A, when they guessed that what they were looking for was not available in Italian.

With only two examples, it is impossible to draw conclusions as to what pushes screen reader users to change the language as a coping strategy. Yet, we observed two potential reasons: (i) repeated unexpected changes between language versions; (ii) the impression that something is not available in one's native language. These findings also suggest that the localised versions of website A, where the content was reorganised and many sections were not translated, are sometimes insufficient to meet the needs of non-German native speakers, who are forced to try and browse the original version.

5.3.4 Language skills

When examining language skills to answer question **R1.4**, we noticed that there was no clear pattern to explain why some users decided to try with German and others did not. Participants 01FR and 01IT, who voluntarily switched to German, reported respectively a B level and no knowledge of the language, whereas participants 02FR, 03FR and 02IT, who did not voluntarily switch to German, reported a level that varied from A to B. Table 5.2 shows the

language skills of the non-German native speakers and whether they decided to switch to German during the session.

	Language skills							
	German	French	Italian	English	Russian	Spanish	Latin	Tried browsing in German?
01FR	B	M		A	B			Yes
02FR	A	M	A	C				No
03FR	B	M		B				No
01IT		A	M	A				Yes
02IT	A	C	M	C		B	C	No

Table 5.2 Language skills of non-German native speakers and their language selection during the session

These findings seem to suggest that an intermediate knowledge of a language does not guarantee that a screen reader user will try and browse that language version, and no knowledge at all does not guarantee that they will not try. A thorough investigation with more participants should be undertaken to verify these hypotheses.

5.4 Language-specific usability issues

As announced in section 5.2, we opted for a thematic analysis approach to identify patterns within the qualitative data collected. Specifically, we looked for recurring themes in the collection of problems encountered by the users while they were completing the tasks and coded for a specific research question (**R2**). We used an abductive approach, "whereby understanding is generated through an analysis that moves continuously to and fro between the more abstract level of concepts and the more concrete level of data" (Saldanha and O'Brien 2013, 63).

The themes reported in this section will serve to test the hypothesis related to research question **R2.** According to our hypothesis **H1** and **H2** users encounter different usability problems depending on the language version they have chosen to browse and the localised version of a multilingual website presents more usability issues for screen reader users than the original version.

From a general perspective, the classification of problems encountered by the participants shows that language-specific usability issues exist (12 out of 69 identified problems) and mainly affect users browsing one of the localised versions (11 out of 12 language-specific issues). Overall, we identified 'language-specific issues' (such as 'involuntary language change' and 'language code'), that were related to the fact that the website was translated and two major obstacles that would also affect monolingual websites: 'technical issues' (such as 'crashes',

'setting up filters', 'headings', 'internal search engine', 'issues with links') and 'lack of clarity' (such as 'terminology issues' and 'navigation issues'), which are summarised in the following diagram (Figure 5.1) and will be examined into more detail in the following sub-sections.

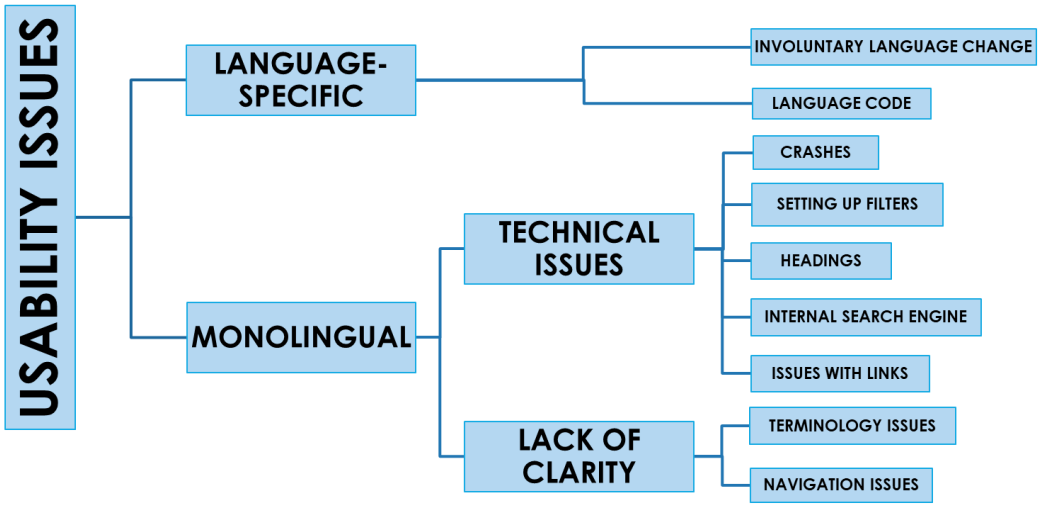


Figure 5.1 Tree diagram of the issues classification

5.4.1 Language-related issues

As mentioned above, this category included the problems that could be traced back to localisation and was broken down into two more specific issues.

Involuntary language change

In section 5.3.3 - *Changing the language as a coping strategy*, while trying to answer research question **R1.3**, we focused on the users who voluntarily changed the language during the test as a coping strategy. However, in this section, we examine the cases where users were faced with a change of language that they were not expecting. For instance, participant 01FR said: '*Ohlalà il m'a foutu en allemand, pourquoi ? Là je suis un peu mal barré, il faut que je retourne en français*'. Most participants showed signs of frustration and tried to go back to the language version they were browsing. When they did not manage to do that, they either switched to Google or give up. Only twice (out of ten times) participants 02FR and 02IT eventually decided to attempt to browse in a language they had not chosen. User 02IT, for example, landed on the German version of School of Engineering (website A) and said: '*Why is it switching to German? I cannot even find the other languages. I would like to change the language. It says DE active and EN. Clicking the space bar does not work. Never mind, I will try with German'.*

Language code

Only two participants (03DE and 01FR) experienced the problem of the screen reader using the wrong voice profile to read content. As we mentioned in sub-section 3.4.2 - *Key WCAG success criteria for Web localisation*, it is essential to identify the human language used in a page for the screen reader to select the correct pronunciation (WCAG criterion *3.1.1 Language of Page*) (Gutiérrez y Restrepo and Martínez Normand 2010). In website A, a German-speaking participant (03 DE) landed on a page written in English but pronounced (i.e read) as German. The user, aware of the fact that this could be due to the settings of the screen reader[41], checked whether the language recognition option was on. They even tried to re-start JAWS, but the situation did not change. A French-speaking user (01FR) found a section in website B that was coded as German even though the content was in French. They kept browsing but complained about it. We also noticed that both participants talked about 'code' and 'programming', suggesting a good knowledge of the principles behind the functioning of a screen reader. A closer look at their profile revealed that they were, respectively, an IT expert and an accessibility expert, which explained their knowledgeable comments.

5.4.2 Technical issues

With this category we wanted to capture and classify the technical problems encountered by the participants, in other words, every issue that appeared to involve the interaction between the website and the screen reader. It included crashes, difficulties in setting up the filters or opening the internal search engine, lack of headings and broken links.

Crashes

A total of five users (N=5) experienced a crash during the test. Two of the crashes, both on website A, were clearly screen reader issues: participant 03DE had to re-open the website, whereas participant 05DE, who was trying to go back to the browser after reading a PDF document simply waited and continued browsing. The three French-speaking participants (N=3), on the other hand, all experienced a crash that they could not quite explain. For example, participant FR01 said: '*Il y a eu un plantage, je ne sais pas où*'.

Setting up filters

[41] Screen readers can be set to always read content with a specified pronunciation or to automatically recognise the language code (lang or xml:lang attribute) of the pages.

Task B-2a was designed so that successful users would be required to set up a filter in order to find the open days of the Swiss literary institute. This task was included to add some interactivity and interactivity and challenge the participants with a potentially problematic goal. As we will further illustrate in section 5.5.1 - *Success rate*, not everybody managed to get to that section of the website. Only two of them (N=2) encountered a technical issue. Participant 01DE complained that '*This filter does not work properly. The page reloads and I lose the focus whenever I try to go down the list. […] When I proceed with 'tab' it freezes. It works, but it is not ideal'.* When they managed to set up the filters, but did not understand why they would not get any results, they said: '*I must have made a mistake while setting up the filters, the search must be wrong*'. In fact, the search was correct and had produced no results, but no message was displayed to explain that there were no results, so they only found an empty table.

Headings

In the section about WCAG success criteria for Web localisation (see sub-section 3.4.2), we mentioned headings as elements that have a great impact on the ease of navigation (Gutiérrez y Restrepo and Martínez Normand 2010). Two participants complained about the lack of such headings or the impossibility to jump from one to the other on website B. For instance, participant 01DE said: '*Headings are missing here. There are some headings but I cannot jump between them*'. The reaction varied from going back (for example, participant 03DE was disappointed by the lack of headings on a PDF, so they went back to the website) to keep browsing or resorting to a screen reader search.

Internal search engine

The search engine embedded in a website can prove a powerful tool for both sighted and screen reader users, because it allows them to bypass navigation and look directly for a specific section or element. This tool was present on both websites and used multiple times by most participants. However, users encountered different types of problems on the two websites. The internal search engine on website A was difficult to access. Six out of ten participants (N=6) could not open it at the first attempt (e.g. 01FR '*Ah merde, j'arrive pas à activer le champ de recherche. Ah voilà, il s'est ouvert, je ne sais pas comment*) or could not open it at all (e.g. 03DE '*The search button does not work'*). Participant 03FR was surprised by the fact that the shortcut to open the search engine was usual: '*Il y a pas une zone de recherche ? Normalement alt+E… Ohlalà! alt+5 un raccourci que je n'emploie pas […]. Si j'appuie sur "Suche" je ne peux pas introduire du texte'.* We observed various reactions to this problem: some went back

to home or re-opened the website, some tried again, others switched to Google or gave up. Website B, on the other hand, was generally opened more easily, but its results included matches from other websites as well. This caused participant 03FR to end up on the wrong website. As they did not realise it, the researcher decided to intervene and suggest going back to the original website. These findings suggest that the internal search engine is a useful tool for screen reader users, as they tend to be disappointed when it does not work. Yet, an internal search engine that works as a browser can be counterproductive and lead screen reader users to another website entirely, without them noticing.

Issues with links

More than half of the problems with links occurred on website B, while trying to complete task B-1 and find the contact information of the Swiss literary institute (5 out of 9 issues). Participants 01DE, 05DE, 02FR, 03FR and 01IT all struggled with the tab that needed to be selected to find the address of the Swiss Literary Institute. The following screenshot (Figure 5.2) illustrates the problem:

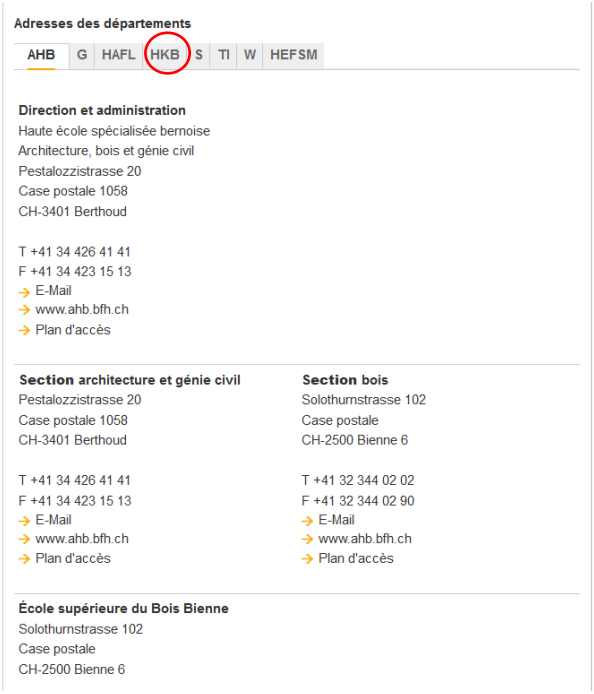


Figure 5.2 Screenshot of the contact section, with relevant tab (HKB) circled in red

After clicking on the HKB tab, participant 03FR said: '*C'est bizarre, j'arrive pas à l'ouvrir. […] Est-ce que le lien marche?*'. Another participant (02FR) tried twice and commented: '*Il marche pas la HKB. [tries again] Trouvé, il disait pas que c'était un lien sur la même page*'. From this last comment it would seem that the 'HKB' tab label did not sound like a link to the same page. It could either be a technical problem or lack of knowledge of the

participants. However, since 50% of the users (N=5) had the same problem, we believe it qualifies as usability problem and would need checking. Even if it complied to accessibility guidelines, there is a margin for improvement in terms of usability.

Two participants (01DE and 01IT) had some difficulties with the link to the events on website A. Participant 01DE said: '*I tried to click but the link does not work. It might be a PDF document, I should check whether it downloaded something. […] No it did not. [tries again] I do not know why this link does not work. It could be inaccessible for a screen reader. I will go somewhere else, then'*. Since most participants who clicked on the same link did not encounter this problem, one possible explanation is that there was a technical issue with a specific combination of screen reader / browser / operating system. To support this hypothesis, we observed that participant 01DE, an accessibility expert, tried switching browser (from Mozilla to Internet Explorer) and the link worked.

5.4.3 Lack of clarity

This category aimed at covering the issues that seemed to confound the users at an abstract level, make them feel lost or uncertain about what to do next. It included unclear terminology and doubts about the structure of the website.

Terminology issues

An appropriate choice of words and phrases in menus and sections is arguably one of the key-elements to increase clarity in a website. While terminology is a concept often associated with translation, we noticed that German speakers encountered more terminology issues than non-German speakers (3 out of 4 issues). Since we assumed that the German version is the original version, the problems cannot be traced back to localisation. For example, participant 05DE tried three promising links ('Aufnahmenerfahren', 'Voraussetzungen und Zulassungsbedingungen' and 'Termine und Fristen') while looking for the open days on website B, before saying '*I cannot find this event. The links here do not bring me to this information*' and deciding to google it. Whilst looking for the open days on website A, participant 03FR complained that none of their screen reader searches had given any results ('portes', 'journée' or 'événements') and commented '*Il y a un certain nombre de liens mais il faut être un peu un insider…*" just before giving up. The cited examples suggest that unmet terminology expectations can cause the users to adopt rather drastic measures, such as abandoning the website to try a new search on the browser or give up.

Navigation issues

As one of the participants pointed out, screen reader users lack the *vue d'ensemble* that sighted people can easily obtain when they explore a website for the first time. For this reason, the information architecture plays a key role for blind Web users. Almost all participants (N=9 out of 10) were at some point confused about navigation, said they were lost, landed on unexpected content or were overwhelmed with too much information. Participant 02FR, for instance, felt lost while looking for the opening times of the Swiss literary institute's library: '*Déjà je suis embêtée parce que Institut Littéraire Suisse je trouve pas. [Chercheuse: Là vous êtes dans quelle page ?] Si je le savais… […] Je suis dans contact. Je sais pas où chercher, si c'est un département…[Qu'est-ce que vous feriez ?] Je téléphonerais ou j'écrirais un mail à la bibliothèque*'. Participant 01DE expressed their doubts on website A as follows: *'[Chercheuse: Did you go back to home?] Yes, because I felt I was missing something in the navigation: the departments'*. The same participant (01DE) also complained about landing on unexpected content on website B: '*If you click here on the address, it opens Google maps. I did not know it would open a map!*'. Participant 01FR was very vocal in their disappointment about the number of links on some pages of website B: '*Bachelor en écriture littéraire: 50 liens. Oh la vache! Je vais me taper toute la liste? 56 liens! [clicks on link] 59 liens! Non je trouve pas. C'est de la merde, trop de liens!'*.

It is important to recall none of the participant was familiar with the websites, so we cannot rule out the possibility that the perceived usability would improve over time, if they kept browsing the websites. Indeed, some of the users commented that they would have needed more time to get acquainted with the websites and judge their usability. However, longitudinal measurements were outside the scope of our study and we adopted two strategies to address the issue raised by the participants. They were allowed to navigate freely at the beginning of the session and we modified statement 8 of the CSUQ to 'I believe I could quickly learn how to use this website', so that they had the chance to estimate whether continued and frequent use would improve usability.

5.4.4 Influence of language chosen by the user on type of usability issues

To test **H1** we examined the types of issues encountered by the participants on the German version and the non-German versions of both websites. Evidence suggest that all users experience lack of clarity and technical issues, whereas language-related issues are encountered almost exclusively by non-German native speakers. The following table (5.3) compares the

number of issues encountered by participants browsing the original version of both websites and the localised versions:

		German version	Localised versions
Monolingual issues	Technical issues	17	10
	Lack of clarity	14	16
Localisation issues	Language-related issues	1	11

Table 5.3 Number of issues encountered on both websites by German vs non-German native speakers by category

The collected data therefore verify our hypothesis and suggests that the type of usability issues in partially localised websites do vary according to the language version chosen by the users.

5.4.5 Number of total usability issues: original vs localised

Our second hypothesis focused on the number of usability issues encountered by participants browsing the original version of the website compared to the localised versions. Since we assumed that the language chosen by the user would have an impact on the perceived usability and our findings showed the existence of language-related usability issues specific to the localised versions, it was reasonable to assume that said localised versions would be subjected to a higher number of issues. In fact, we only observed a small difference (32 issues in the original versions vs 37 in the localised versions). These numbers support our hypothesis **H2** and will be complemented with the overall usability scores obtained per website, but such a small difference requires to be cautious in generalising the results.

5.5 Localisation levels and usability

The findings reported in this section serve to investigate the third research question:

R3: Does the use of different localisation strategies have an impact on the overall usability of multilingual websites for screen reader users?

Given the fact that the websites we tested have been localised in different ways (as illustrated in sub-section 4.5.1), to answer question **R3** we reported the success rate of the users on both websites and then compared the results of the CSUQ questionnaires, as well as the overall frustration scores and the perceived lost time.

5.5.1 Success rate

One of the most basic ways to measure usability is the success rate. This measure, especially when limited to binary values (yes/no), does not account for the ease of use, but only for the final result. Table 5.4 illustrates the success rate per task. The mean success rate for website A was 57% and for website B it was 43%. Overall, it appears that users were more successful on website A (localisation degree 'reorganised') than website B (localisation degree 'imperfect mirror). In the next sub-sections (5.5.2 and 5.5.3) these findings will be challenged by the scores of the perceived usability.

	Website A			Website B		
	1	2a	2b	1	2a	2b
01DE	Yes	Yes	No	Yes	Yes	No
02DE	Yes	Yes	No	No	No	Yes
03DE	Yes	Yes	No	Yes	No	Yes
04DE	Yes	Yes	Yes	Yes	Yes	Yes
05DE	Yes	Yes	Yes	Yes	Yes	Yes
01FR	Yes	No	No	Yes	No	No
02FR	Yes	No	No	Yes	No	No
03FR	Yes	No	No	No	No	No
01IT	Yes	No	No	No	No	No
02IT	Yes	No	No	No	No	No
	100%	50%	20%	60%	30%	40%

Table 5.4 Success rate per participant and per task

Task A-1

Task 1 on website A only required finding the contact information of the university and was classified as 'easy'. Indeed, all participants completed it, so the success rate was 100%. It would therefore appear that the localisation strategy of reorganising content was successful on the homepage.

Task A-2a

Task 2a on website A, that involved finding the date, time and location of the open days for a specific Bachelor offered by the School of Applied linguistics, was completed by 50% of the participants (N=5), with a success rate ranging from 100% among native German speakers and 0% among non-German native speakers. These results suggest that the localisation strategy adopted on this website ('reorganised') does not guarantee a high level of usability for users of a localised version once they reach a deeper level of the website, as it was the case for this task.

Task A-2b

Task 2b on website A consisted in finding and subscribing to the newsletter of the School of Engineering. Only 20% of participants (N=2) managed to complete the tasks and they were both native German speakers. These results indicate that it is difficult, even for users of the original version, to get to this specific section of the website.

Task B-1

Task 1 on website A required finding the contact information of a specific faculty and was classified as 'easy'. Yet, only six out of ten participants (N=6) successfully completed the task. Four of them (N=4) were native German speakers and two (N=2) were native French speakers. This task was problematic for many users, regardless of the language version, because of the technical problem described in section 5.4.2. If we consider the distribution by language of the success rate, we cannot state that there is a connection with the localisation strategy.

Task B-2a

Task 2a on website B involved finding the date, time and location of the open days for a specific Bachelor by using the filters of the event section. This task was conceived as impossible, as manual inspection by a sighted user had produced no results, so we had initially thought to consider successful the correct use of event's filters to find out that there were no scheduled open days. One user (N=1) achieved that. Unexpectedly, two users (N=2) found a link to an external website that advertised such event. We therefore considered that the success rate was 30% (N=3). All the other users (N=7) either did not find the event filter, or they found it and did not manage to set it up.

Task B-2b

Task 2b on website B, that consisted in finding the address and opening time of a specific library, was successfully completed by 40% of the participants (N=4), all of whom were native German speakers. These results suggest that the localised versions of this websites, classified as 'imperfect mirror', seem to fail to lead the user to this specific section of the website. We believe that one of the main issues was the location of the opening times in the middle of a paragraph (as described in sub-section 4.6.1.2), because some participants did find the address, but were then lost when it came to the opening times. Additionally, users that browsed in a

language other than their native tongue found it even more difficult to read a whole paragraph in a foreign language, rather than just the titles, so they tried to avoid it.

5.5.2 Usability measurement - CSUQ scores

The usability questionnaire is a validated method to measure the perceived usability of a system. As we illustrated in section 4.6.3, we chose the CSUQ and we adapted it for evaluating websites, taking into account the abilities of the evaluators. The questionnaire, that included 19 statements, can be found in *Appendix E*. Participants were asked to indicate their degree of agreements on a scale where 1 indicated strong disagreement and 7 strong agreement. Since all the statements were phrased positively, a higher score meant a higher level of usability.

With exception of statement 9 ("The website gave error messages that clearly told me what the problem was"), the mean scores of all statements were systematically somewhat higher for website B (degree of localisation: imperfect mirror). In Figures 5.3 and 5.4 the results are visualised in four categories: 'system usefulness' regroups the values of the scores from statements 1 to 8, 'information quality' of the scores from statements 9 to 15, 'navigation' represents the score from statements 16 to18 and 'overall satisfaction' from statement 19.

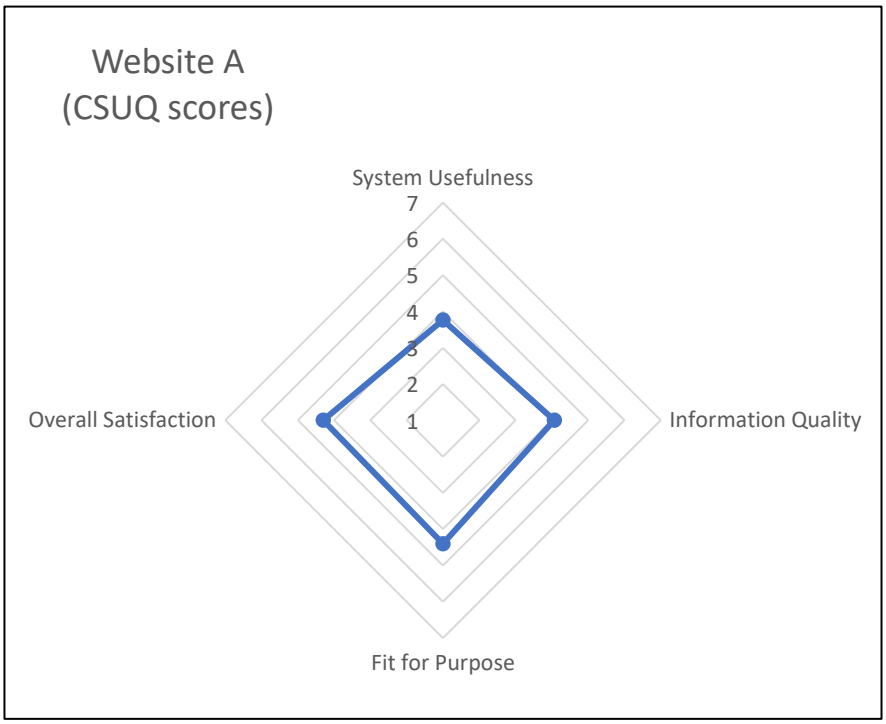


Figure 5.3 CSUQ scores of website A by subscale

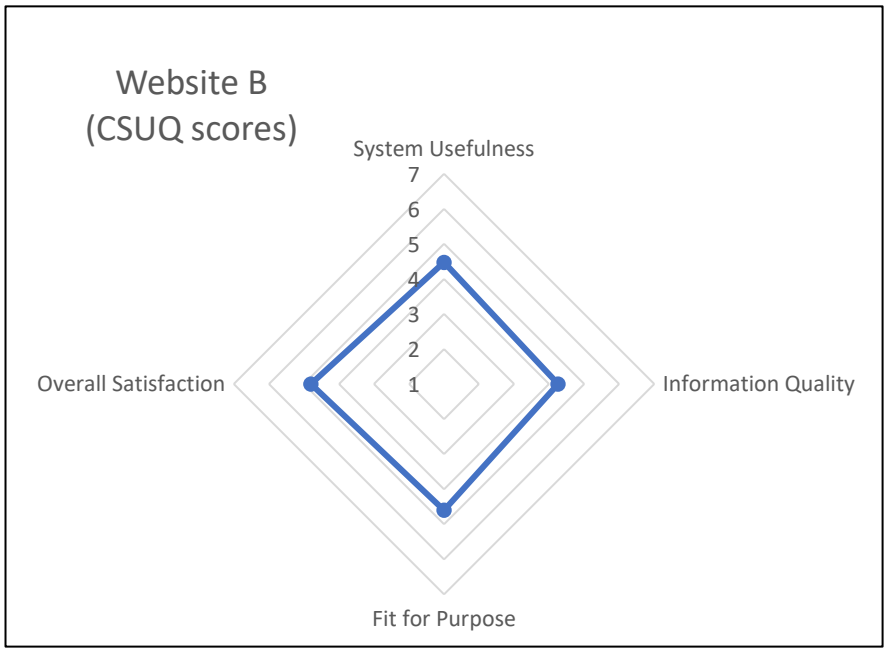


Figure 5.4 CSUQ scores of website B by subscale

Even though this is an exploratory qualitative study, these results have also been tested to calculate their statistical significance. The following table (5.5) shows that only the difference between website A and B for the subscale 'system usefulness' proved to be statistically significant. A closer look to individual items showed that statement 9 ('The website gave error messages that clearly told me what the problem was') had a negative impact on the significance of the overall scores. Indeed, this particular item appeared to be unclear for participants, who did not know how to rate it: they had not received any error message, but they still had to pick a number on the Likert scale, because a 'not applicable' (n/a) option was not available. Unfortunately, this issue was discovered once the tests had started and the n/a option could not be added. The mean score of this statement was included in the overall usability score, but the p-value would have been 0,002, if it had been excluded.

	Subscale								**Overall usability**	
	System usefulness		**Information quality**		**Navigation**		**Overall satisfaction**			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Website A	3.76	1.64	3.96	1.85	4.43	1.81	4.30	1.70	3.97	1.75
WebsiteB	4.48	1.64	4.11	1.62	4.70	1.36	4.80	1.48	4.39	1.57
p-value (t-test)	0.006		p > 0.05		p > 0.05		p > 0.05		p > 0.05	

Table 5.5 Recapitulative table of the mean usability scores by subscale and overall

These results show that the perceived usability level of website B was slightly but consistently better higher than website A. The small difference between the results was not

enough to state that the localisation degree of website B yielded a higher usability level than the localisation degree of website B. One possible explanation was that the 'imperfect mirror' localisation strategy used for website B was indeed better for screen reader users than the 'reorganised' strategy used for website A, but Italian-speaking users had rated website B poorly because it did not have an Italian version. However, this hypothesis was rejected by a closer look at the data, showing that Italian native speakers assigned better scores to website B as well. We also tried to compare the scores of non-German native speakers on website A and B between them but, again, the difference was not statistically significant. With the data at our disposal, we had to conclude that an 'imperfect mirror' localisation strategy only yields better result as far as the usability category 'usefulness' is concerned.

5.5.3 Overall frustration

After rating their agreement to the statements of the CSUQ questionnaire, participants were asked to indicate, from a general point of view, how frustrating it was to browse each website, based on a Likert scale from 1 (not at all frustrating) to 7 (extremely frustrating). We understand that the higher the level of frustation it is, the lower the level of usability. The mean score of frustration on website A was 3.5 (SD=1.8), while on website B it was 3.3 (SD=1.3). These results are consistent with the CSUQ scores, because they indicate a marginally better usability level for website B than website A.

5.5.4 Perceived lost time

Time was not the primary concern of our study, and the fact that the experiments were not carried out in a controlled environment such as a laboratory made it difficult to obtain precise measurements. One of the tests, for example, was carried out in a shared workplace with some background noise and one or two brief interruptions. Also, four participants received a phone call during the session and, even though they ignored it or picked up just to say they were going to call back, the validity of time measurements was somewhat compromised. However, as we will illustrate in sub-section 6.3.1, this increased the ecological validity of the study.

Instead of measuring the time needed to complete every single task and attempt to compare it objectively, we asked participants to estimate how much time they had lost overall, in each session. While this is a subjective and less reliable measurement, we believe it is interesting because it is complementary to the frustration perception and reveals another aspect of usability: efficiency. Indeed, during the test, several participants commented that they dislike wasting time and sometimes give up or call when they cannot find what they are looking for.

The following tables (5.6 and 5.7) show the perceived lost time next to the total time spent to carry out the tasks for both websites. The third column shows the percentage of lost time out of the total time.

A	TOTAL	Perceived lost time	Percentage
01DE	00:27:55	00:03:00	11%
02DE	00:31:54	00:15:00	47%
03DE	00:35:38	00:15:00	42%
04DE	00:17:01	00:10:00	59%
05DE	00:25:28	00:05:00	20%
01FR	00:21:37	00:05:00	23%
02FR	00:16:41	00:10:00	60%
03FR	00:33:41	00:30:00	89%
01IT	00:26:35	00:20:00	75%
02IT	00:35:20	00:10:00	28%
Mean	00:27:11	00:12:18	45%

Table 5.6 Perceived lost time on website A by participant (in min.)

B	TOTAL	Perceived lost time	Percentage
01DE	00:34:25	00:05:00	15%
02DE	00:14:13	00:03:00	21%
03DE	00:28:04	00:30:00	107%
04DE	00:13:30	00:05:00	37%
05DE	00:29:23	00:05:00	17%
01FR	00:18:38	00:10:00	54%
02FR	00:15:21	00:07:00	46%
03FR	00:24:54	00:30:00	120%
01IT	00:23:22	00:10:00	43%
02IT	00:42:56	00:12:30	29%
Mean	00:24:29	00:11:45	49%

Table 5.7 Perceived lost time on website B by participant (in min.)

As the tasks were similar but not equal on both websites, the total time needed to complete (or attempt to complete) the tasks was not relevant *per se*, but it was in relation to the perceived lost time. Participants spent an average of 27'11'' on website A and 24'29'' on website B. They estimated that they lost, in average, 12'18'' or 45% of the time on website A and 11'45'' or 49% of the time on website B.

These findings are not consistent with the usability questionnaire and the overall frustration question, as they show that participants believed they had lost more time on website B than website A, suggesting a lower usability level for website B. Yet, it is interesting to notice two outliers that probably turned the results around: participants 03DE and 03FR estimated that they had lost more time than they actually spent completing the tasks on website B. A high level of frustration might have been the cause for this overestimation, but these results represent two

peaks among generally low results, that would be compatible with the hypothesis that website B (localisation degree: ‘imperfect mirror’) is perceived as slightly more usable than website A (localisation degree: ‘reorganised’).

5.6 Summary of the findings

In this section we briefly summarise the results analysed in the previous sections.

The language navigation patterns were investigated in section 5.3. Results showed that the vast majority of the participants selected the website version corresponding to their native tongue whenever it was available and that users were positive surprised when they landed directly on the preferred version. To change the language, most of them read the content of the homepage until they heard the language code of their preferred language. One of them actively searched for the preferred language. On the other hand, changing the language to browse the website in a foreign language was not a very popular choice and was only used as a coping strategy when other attempts had failed. Finally, no apparent correlation was found between this voluntary change of language and language skills.

Section 5.4 focused usability issues and how they varied across language versions. We observed a clear difference in the distribution of issues between the original and the localised versions of both websites: while the issues classified as ‘technical issues’ and ‘lack of clarity’ were almost equally present across versions, there was a definite prevalence of ‘language-related issues’ in the localised versions. Also, the total number of usability issues encountered by the users browsing the localised versions was somewhat higher than the original version. These findings highlight the importance of considering usability for disabled users when localising websites, as the number of usability issues might significantly increase, if it is not done properly.

The third section (5.5) compared the usability of the two tested websites, which have been localised to different degrees. Firstly, the success rate indicated that participants were generally more successful on website A than website B. However, the tasks performed on the two websites were different, so it is wise not to focus uniquely on these results and try and contextualise them, as we have attempted to do by describing the single tasks. The CSUQ scores represent a more reliable way of comparing the websites, as they aim at measuring overall usability. The scores were consistently but only slightly higher for website B than website A, but the results were not statistically significant and therefore not conclusive. Frustration scores also indicated a slightly higher level of usability on website B. Finally, perceived lost time

measurements seemed to be contradictory, as the percentage of lost time was higher for website B (indicating a lower efficiency). However, the absolute lost time is lower and we identified two scores that were clearly the result of overestimation.

The study proved useful to answer our research questions, starting from a more general level (language preferences and navigation patterns) to an increasingly specific level (language-related usability problems and impact of the localisation strategy on the usability level). Throughout our analysis we have raised some issues that should be taken into account when localising a website in a country such as Switzerland. In the next chapter (*6. Conclusion*) we will recall our steps and highlight the aspects of our findings that should be considered by web localisers.

6 Conclusions

6.1 Overview of the research

This master's thesis examined the impact of multilingualism and partial localisation strategies on the web navigation experience of screen reader users and was composed of two main parts: a theoretical framework and an exploratory study.

Theoretical framework

After the general introduction, Chapter 2 reviewed the literature to date in order to provide the theoretical foundations on the topics addressed by this thesis. The chapter presented the main definitions of the concepts of localisation, usability and accessibility and the theoretical perspective that was adopted in this study. It also illustrated the major theories that were developed around the relationship between these concepts. Chapter 3 highlighted the prior work that was reviewed before designing the present study, allowing us to bring forward the relevance of our study and its contribution to the existing research. The number and profile of the participants (when applicable), the results, the aspects covered, the advantages and drawbacks were all carefully examined in order to select the best method to answer our research questions, as described in the *Methodology* chapter.

Exploratory study

Chapter 4 illustrated the methodology adopted to conduct an exploratory study that investigated the navigation patterns of visually impaired users on multilingual websites, examined language-related usability issues and studied the impact of different translation strategies on the perceived usability of two university websites. In Chapter 5, we presented the findings and tried to explain what we observed and measured.

6.2 Main findings

Unsurprisingly, we found that most screen reader users preferred to browse in their mother tongue. We also noticed that they were used to recognising the code of their preferred language when they heard it, as opposed to the language name in full, which is often recommended as best practice from a localisation perspective, and that they were not keen on

changing the navigation language to a foreign language, unless they were really stuck on a task - and sometimes not even then.

We also observed that users encounter more usability issues on localised versions than on the original version and, specifically, that they encounter more language-related issues. Double-checking language attributes while localising would be an effective way to drastically reduce this kind of issues. Ideally, all usability issues that affect the navigation experience of visually impaired users should be addressed and solved, but our classification also offers multiple ways of prioritising. Most frequent issues, for example, could be addressed first. Alternatively, issues that cause users to give up could be given priority. However, this topic should be examined in more depth in further studies.

Finally, we were hoping to determine whether one of the identified localisation strategies had a better impact on overall usability that the other. Unfortunately, the results were not conclusive enough for us to recommend one localisation degree over the other. Nonetheless, it appears that neither partial localisation strategy is satisfactory for screen reader users because, as a group, the users who browsed any of the localised versions consistently assigned lower scores than those who browsed the original version. Perhaps, the main characteristics of partially localised websites (see section 4.5.1), such as the lack of some content and the links that point to content in a different language without any warning, add a confounding element for screen reader users that makes their navigation experience less satisfactory. To verify this hypothesis, an approach similar to the one described in section 3.5.2 by Leuthold, Bargas-Avila, and Opwis (2007) could be used to test an artificially fully localised version of the same websites and compare the results.

6.3 Limitations and future work

We have already acknowledged some of the limitations of the study whilst describing the methodology or discussing the findings, but this section aims at collecting them and recalling how they have been addressed or what could be improved in a future study.

The inclusion of more participants, for instance, would have given the study more statistical power (Lazar, Feng, and Hochheiser 2017, 459), but the available resources in terms of time and interested candidates did not allow it. Besides, qualitative studies with 5 to 10 participants are not uncommon in the fields of accessibility and usability (see section 4.4). A similar study that aims for statistical significance should involve at least 5 participants for each native language (a total of 15 people). Also, the study environment varied from user to user and

might have introduced some biases in the form of distractions. As highlighted in sub-section 6.3.1, this variety represents an advantage as it is closer to real life than any laboratory could ever be. Nonetheless, we tried to counterbalance the effect of a noisy environment by letting the participants free to choose whether they wanted to wear earphones or not. Time was another difficult aspect to control. We tried to keep the probability of major changes to the websites to a minimum by running the test sessions within one month. A second full time researcher would have been useful to double-code the whole interview analysis, even though we argue that a small percentage of the data recoded can still provide a good indication of reliability (see section 5.2). We have already acknowledged the issue with one of the statements, that could have been solved by modifying the wording or introducing a 'not applicable' option. The drawback of this last solution would be an increase in the lack of available data for the analysis, because participant might be tempted to use the n/a option whenever in doubt. As mentioned previously, the results of the study cannot be applied to a larger population because a sample of convenience was used instead of a random sample (Simon 2011). Additionally, the variety of data sources, that strengthened the internal validity of the study (see sub-section 6.3.1), became a challenge when it came to the analysis. Traditional methods had to be adapted and combined to analyse both quantitative and qualitative data within the framework of the research questions. Finally, our findings were specific to a country (Switzerland), a set of languages (German as a source language and French and Italian as target languages) and a particular type of website (university websites). However, this specificity allowed us to focus on an under researched area and make our contribution to the new Swiss project with a focus on barrier-free communication in higher education settings (see section 1.2 – *Motivation*).

Future studies that do not aim at replicating the results could introduce a control group composed by sighted users to compare the usability or the navigation patterns of the two groups. If conducted in a laboratory, the same study could focus on time on tasks, as it would be easier to measure and more reliable. A longitudinal study could either focus on the development of the tested websites or examine the effects of the continuous use of the websites on the perceived usability. The same methodology could be applied to other websites on the Access for All list or from another country entirely, as long as it is multilingual. Finally, as one of the issues raised by participants was the lack of time to explore the whole website before having to judge its usability, instead of comparing two websites, a deeper investigation of just one website could provide more clues about the usability of a larger portion of the website.

6.3.1 Validity

In this section, we address the quality criteria that need to be met in order to ensure that our study makes some contribution to the existing research: validity, reliability and generalisability (Saldanha and O'Brien 2013, 27). At a basic level the validity of our results depends on the extent to which the data that are collated and analysed can contribute to answering the research questions. Oates (2006, 131) stated that "[an] experiment has good internal validity if the measurements […] are indeed due to [the] manipulation of the independent variable and not to any other factors". In fact, this criterion applies to 'true experiments', that aim at identifying causal relationships and where participants are randomly assigned to different conditions, while our study can be classified as 'quasi-experiment'(Lazar, Feng, and Hochheiser 2017, 45). However, we designed our experiment in order to collect our data from multiple sources and measure as many variables as possible[42]. To increase the internal validity of our study, we verified that our participants had a similar background in terms of expertise (they all reported surfing the Internet daily and have been using a screen reader for more than 5 years) and previous experience on the websites (nobody was a frequent user or even remembered visiting them more than once or twice). We also asked a second coder to review a sample of the thematic analysis we carried out to answer the second research question (see section 5.2 – *Data analysis approach*) and used a validated questionnaire to collect the data to answer the third research question (see section 5.5.2 - *Usability measurement - CSUQ scores*).

Reliability refers to the extent to which the results are replicable by other researchers investigating the same research question, using the same stimuli and methods at different time. To increase reliability, we discussed the data collection and analysis methods in a transparent way and made our theoretical standpoint as clear as possible (Saldanha and O'Brien 2013, 35).

Generalisation, or external validity, refers to the extent to which the results are not unique to a particular set of circumstances, but allow to make claims about the larger population (Oates 2005, 132; Saldanha and O'Brien 2013, 36). Given the very specific target users that we decided to study, and the lack of randomisation due to the small pool of potential participants, we cannot claim that our results apply to the general population or even the whole screen reader community. Finally, we argue that the strength of our qualitative research study comes from its ecological validity, that depends on whether the effect is representative of what happens in real

[42] This method was inspired by Morado Vázquez (2012, 122), who also conducted a quasi-experiment in the field of Translation Studies.

life (Crano, Brewer, and Lac 2014, 110). Since it is reasonable to think that our target population does not usually browse the internet in a distraction-free laboratory, the fact that we did not control the physical environment where the participants carried out the test adds to the ecological validity of the study.

While there are endless possibilities to improve or modify the present study, we believe that our findings have made a contribution to the fields of usability and localisation by considering usability for a specific category of people who access the Web in a non-visual way and by introducing the aspect of multilingualism, which is often overlooked. We argue that the needs of people with disabilities are often not sufficiently met in usability practice and research and that addressing accessibility, usability and inclusion together can lead more effectively to a more accessible, usable, and inclusive Web for everyone[43]. With our work, we hope to raise awareness about the necessity of thinking about all users when localising a website and pave the way for future studies in this direction.

[43] https://www.w3.org/WAI/intro/usable Last access: 15 July 2018

7 Bibliography

Abou-Zhara, Shadi. 2008. 'Web Accessibility Evaluation'. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 79–106. Human–Computer Interaction Series. London: Springer-Verlag.

Access for all. 2016. 'Étude Accessibility 2016 En Suisse' https://www.access-for-all.ch/images/Accessibilty_Studie/2016/EtudeAccessibility2016enSuisse.pdf Last access: 17 November 2017.

Aizpurua, Amaia, Myriam Arrue, Simon Harper, and Markel Vigo. 2014. 'Are Users the Gold Standard for Accessibility Evaluation?' In *Proceedings of the 11th Web for All Conference*. 13:1-13:4. ACM Press.

Aizpurua, Amaia, Myriam Arrue, and Markel Vigo. 2013. 'Uncovering the Role of Expectations on Perceived Web Accessibility'. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 74. ACM Press.

Aizpurua, Amaia, Myriam Arrue, and Markel Vigo. 2015. 'Prejudices, Memories, Expectations and Confidence Influence Experienced Accessibility on the Web'. *Computers in Human Behavior* 51 (October): 152–60.

Aizpurua, Amaia, Simon Harper, and Markel Vigo. 2016. 'Exploring the Relationship between Web Accessibility and User Experience'. *International Journal of Human-Computer Studies* 91 (July): 13–23.

Albert, William, and Thomas Tullis. 2013. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Second edition. Interactive Technologies. Waltham: Elsevier Science.

Andreu-Vall, Mar, and Mari-Carmen Marcos. 2012. 'Evaluación de sitios web multilingües: metodología y herramienta heurística'. *El Profesional de la Información* 21 (3): 254–60.

Barnum, Carol M. 2011. Usability Testing Essentials: Ready, Set...Test! Burlington: Elsevier Science.

Barreto, Armando. 2008. ‘Visual Impairments’. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 3–14. Human–Computer Interaction Series. London: Springer-Verlag.

Bourne, Rupert R. A., Seth R. Flaxman, Tasanee Braithwaite, Maria V. Cicinelli, Aditi Das, Jost B. Jonas, Jill Keeffe, et al. 2017. ‘Magnitude, Temporal Trends, and Projections of the Global Prevalence of Blindness and Distance and near Vision Impairment: A Systematic Review and Meta-Analysis’. *The Lancet Global Health* 5 (9): e888–97.

Brajnik, Giorgio, and Sanela Graca. 2018. ‘On Accessibility Policies for Higher Education Institutions’. In *Proceedings of the 15th International Cross-Disciplinary Conference on Web Accessibility*. W4A '18. Lyon, France. New York, USA: ACM. DOI: 10.1145/3192714.3192833

Brajnik, Giorgio. 2008. ‘A Comparative Test of Web Accessibility Evaluation Methods’. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility.* 113–20. ACM Press.

Braun, Virginia, and Victoria Clarke. 2006. ‘Using Thematic Analysis in Psychology’. *Qualitative Research in Psychology* 3 (2): 77–101.

Cook, Albert M., and Janice Miller Polgar. 2014. Assistive Technologies- E-Book: Principles and Practice. Fourth edition. St. Louis: Elsevier Health Sciences.

Crano, William D., Marilynn B. Brewer, and Andrew Lac. 2014. Principles and Methods of Social Research. Third edition. New York: Routledge.

Cunningham, Katie. 2012. Accessibility Handbook: Making 508 Websites for Everyone. Beijing: O’Reilly.

de Grandi, Ludovica Lea. 2017. ‘Localization in Advertising: Case Study of Switzerland’. MA disseration. University of Geneva. Geneva, Switzerland.

Dumas, Joseph S., and Janice C. Redish. 1999. A Practical Guide to Usability Testing. Revised edition. Human/Computer Interaction. Portland, USA: Intellect.

Dunne, Keiran J. 2006. Perspectives on Localization. American Translators Association Scholarly Monograph Series. Amsterdam/Philadelphia: J. Benjamins Publishing Company.

Edwards, Alistair D.N. 2008. ‘Assistive Technologies’. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 141–62. Human–Computer Interaction Series. London: Springer-Verlag.

Fernández Costales, Alberto. 2012. ‘The Internationalization of Institutional Websites: The Case of Universities in the European Union’. Edited by Anthony Pym and David Orrego-Carmona. Tarragona: Translation Research Projects 4, 51–60.

Graca, Sanela. 2017. ‘Comparison of Web Accessibility Policies in Universities’. MA dissertation. Università degli Studi di Udine. Udine, Italy.

Gutiérrez y Restrepo, Emmanuelle, and Loïc Martínez Normand. 2010. ‘Localization and Web Accessibility’. *Tradumàtica: Tecnologies de La Traducció*, no. 8 (July): 1.

Hanson, Vicki L., John T. Richards, and Cal Swart. 2008. ‘Browser Augmentation’. In *Web Accessibility. A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 215–30. Human–Computer Interaction Series. London: Springer-Verlag.

Harper, Simon, and Yeliz Yesilada. 2008. Web Accessibility: A Foundation for Research. Human–Computer Interaction Series. London: Springer-Verlag.

Jiménez-Crespo, Miguel A. 2013. Translation and Web Localization. Abingdon: Routledge.

Lazar, Jonathan, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. Research Methods in Human-Computer Interaction. Second edition. Cambridge, MA, USA: Elsevier Science & Technology Books.

LeBlanc, Louisane. 2018. ‘Sites Web municipaux dans un Canada bilingue : évaluation de la qualité de la traduction, de la localisation et de l’accessibilité’. MA dissertation. Université Concordia. Montréal, Canada.

Leuthold, Stefan, Javier A. Bargas-Avila, and Klaus Opwis. 2007. ‘Beyond Web Content Accessibility Guidelines: Design of Enhanced Text User Interfaces for Blind Internet Users’. *International Journal of Human-Computer Studies* 66 (4): 257–70.

Lewis, James R. 1995. ‘IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use’. *International Journal of Human-Computer Interaction* 7 (1): 57–78.

Lula, Mark. 2011. ‘Web Content Accessibility of the Government Websites in Switzerland’. Seminar thesis. Université de Fribourg. Fribourg, Switzerland.

Lunn, Darren, Simon Harper, and Sean Bechhofer. 2011. ‘Identifying Behavioral Strategies of Visually Impaired Users to Improve Access to Web Content’. *ACM Transactions on Accessible Computing* 3 (4): 1–35.

Miraz, Mahdi H, Maaruf Ali, and Peter Excell. 2013. ‘Multilingual Website Usability Analysis Based on an International User Survey’. In *Proceedings of the Fourth International Conferences on Internet Technologies and Applications* (ITA 13), 236–44.

Morado Vázquez, Lucía. 2012. ‘An Empirical Study on the Influence of Tranlsation Suggestions’ Provenance Metadata’. PhD thesis. University of Limerick. Limerick, Ireland.

Nielsen, Jakob. 1994. Usability Engineering. San Francisco: Elsevier.

Oates, Briony J. 2005. Researching Information Systems and Computing. Oxford: SAGE.

Petrie, Helen, and Omar Kheir. 2007. ‘The Relationship between Accessibility and Usability of Websites’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 397–406. ACM.

Petrie, Helen, Andreas Savva, and Christopher Power. 2015. ‘Towards a Unified Definition of Web Accessibility’. In *Proceedings of the 12th Web for All Conference*, 35:1–35:13. W4A ’15. New York, USA: ACM.

Power, Christopher, André Freire, Helen Petrie, and David Swallow. 2012. ‘Guidelines Are Only Half of the Story: Accessibility Problems Encountered by Blind Users on the Web’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 433–442. CHI ’12. New York, NY, USA: ACM.

Pym, Anthony. 2004. The Moving Text: Localization, Translation, and Distribution. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Pym, Anthony. 2011. ‘Website Localization’. In *The Oxford Handbook of Translation Studies*, edited by Kirsten Malmkjær and Kevin Windle, 410–24. Oxford, England: Oxford University Press.

Quazzico, Federica. 2016. ‘Localisation et accessibilité du Web à partir de l’étude de cas de la RTS et la RSI’. MA dissertation. University of Geneva. Geneva, Switzerland.

Quesenbery, Whitney. 2008. ‘The Five Dimensions of Usability’. In *Content and Complexity: Information Design in Technical Communication*, edited by Michael J. Albers and Mary Beth Mazur, 81–102. Mahwah: Laurence Erlbaum Associates, Inc.

Rodríguez Vázquez, Silvia. 2015. ‘Exploring Current Accessibility Challenges in the Multilingual Web for Visually-Impaired Users.’ In *Proceedings of the 1st International Workshop on Multilingual Web Access* (MWA), 871–73. New York: ACM Press.

Rodriguez Vazquez, Silvia. 2016. ‘Assuring Accessibility during Web Localisation: An Empirical Investigation on the Achievement of Appropriate Text Alternatives for Images’. PhD dissertation. University of Geneva. Geneva, Switzerland. University of Salamanca. Salamanca, Spain.

Rodríguez Vázquez, Silvia, Sharon O’Brien, and Dónal Fitzpatrick. 2017. ‘Usability of Web-Based MT Post-Editing Environments for Screen Reader Users’. In *MT Summit XVI*, 13–25. Nagoya, Japan.

Rodríguez Vázquez, Silvia, and Sharon O’Brien. 2017. ‘Bringing Accessibility into the Multilingual Web Production Chain’. In *Universal Access in Human–Computer Interaction. Design and Development Approaches and Methods*, edited by Margherita Antona and Constantine Stephanidis, 10277:238–57. Cham: Springer International Publishing.

Rodríguez Vázquez, Silvia, and Jesus Torres Del Rey. 2014. ‘Fostering Accessibility through Web Localization’. *MultiLingual Magazine* 25 (7): 32–37.

Rubin, Jeff, Dana Chisnell, and Jared Spool. 2008. Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. Second edition. Indianapolis: Wiley Publishing.

Saldanha, Gabriela, and Sharon O'Brien. 2013. Research Methodologies in Translation Studies. Abingdon: Routledge.

Sandrini, Peter. 2008. 'Localization and Translation'. *MuTra Journal* 2: 1–25.

Simon, Marilyn K., and Jim Goes. 2011. 'Assumptions, Limitations and Delimitations'. In *Dissertation and Scholarly Research: Recipes for Success*. Seattle, WA, USA: Dissertation Success, LLC.

Singh, Nitish, and Arun Pereira. 2005. The Culturally Customized Web Site: Customizing Web Sites for the Global Marketplace. London: Routledge.

Theofanos, Mary Frances, and Janice C. Redish. 2003. 'Guidelines for Accessible and Usable Web Sites: Observing Users Who Work With Screen Readers'. *Interactions 10* (6): 38–51.

Vigo, Markel, and Simon Harper. 2013. 'Coping Tactics Employed by Visually Disabled Users on the Web'. *International Journal of Human-Computer Studies* 71 (11): 1013–25.

Yerlikaya, Zehra, and Pınar Onay Durdu. 2017. 'Usability of University Websites: A Systematic Review'. In *Universal Access in Human–Computer Interaction. Design and Development Approaches and Methods*, edited by Margherita Antona and Constantine Stephanidis, 277–87. Cham: Springer International Publishing.

Yesilada, Yeliz, Giorgio Brajnik, Markel Vigo, and Simon Harper. 2012. 'Understanding Web Accessibility and Its Drivers'. In *Proceedings of the International Cross-Disciplinary Conference on Web Accessibility*, 19:1-19:9. W4A '12. Lyon, France: ACM. https://doi.org/10.1145/2207016.2207027.

Yunker, John. 2003. Beyond Borders: Web Globalization Strategies. Indianapolis, Ind: New Riders.

Appendices

Appendix A - List of associations

The call for participation was sent to the following associations:

German-speaking Switzerland

- Sehbehindertenhilfe Basel
- Blinden- und Behinderten-zentrum Bern
- Schweizerischer Blinden- und Sehbehindertenverband SBV
- Blindenschule Zollikofen

French-speaking Switzerland

- Écoute-voir
- Association pour le Bien des Aveugles et malvoyants ABA
- Centre de Compétences en Accessibilité CCA

France

- Association Valentin Haüy

Italian-speaking Switzerland

- Unitas - Associazione ciechi e ipovedenti della Svizzera italiana

Italy

- Unione Italiana Ciechi e Ipovedenti

Appendix B - Call for participation

Hello,

I am a student of translation at the University of Geneva and am looking for screen reader users willing to participate in a web usability study I am conducting for my master's thesis.

The study aims to compare the experiences of screen reader users with different native languages when navigating the website of a Swiss university.

Sound interesting? Here is all you need to know!

We are looking for blind or visually impaired individuals who are familiar with screen reader software. If you wish to take part, you will use your own laptop and your usual screen reader to navigate two websites and complete some tasks according to instructions. A researcher will be present to assist you. She will ask you to report any problems you may encounter and record your answers with an audio device. You will also be asked to complete a questionnaire about your experience navigating the websites and some basic information about yourself. Taking part in this research study is voluntary and will take a maximum of 2 hours.

The conditions to take part in the study are:

- **being aged 18 or over**
- **being a screen reader user for at least one year**
- **being a native or near-native speaker of German, French or Italian**

If you match this profile and would like to participate, please let us know by filling in a short consent form that you will find at the following link (active until the 31st of May): [link]

Please feel free to contact us at the following address [elisa.casalegno@etu.unige.ch] if you have any questions regarding the study and feel free to mention the study to anyone who might be interested.

Thank you and kind regards,

Elisa Casalegno

Appendix C - Recruitment form

Thank you for your interest in the study and for taking the time to fill in this form.
In the first section you will be asked to confirm that your profile corresponds to the user profile we are looking for. In the second section you will find an explanation of how the study works. In the third section you will be asked to confirm that you want to take part in it.

Please provide your email address* _____

Please confirm your email address by typing it in again* _____

First section

1. **I am aged***
   - o under 18 years old
   - o between 18 and 40 years old
   - o over 40 years old
2. **I have been using a screen reader for one year or more***
   - o True
   - o False
3. **I am a native or near-native speaker of German, French or Italian***
   - o True
   - o False
4. **I could meet the researcher to carry out the study*** (It is possible to select more than one option)
   - ☐ in Geneva
   - ☐ in the Vaud area
   - ☐ in Bern
   - ☐ in the Westschweiz area
   - ☐ in the Zurich area
   - ☐ in Ticino
   - ☐ other (Please specify)

Second section

This study aims to compare the experiences of screen reader users with different native languages when navigating the website of a Swiss university. It is part of a research project for a master's thesis at the University of Geneva. It will be conducted on site, that is, the researcher is going to meet you at a location of your choice, in April or May 2018.

You will be asked to perform some tasks on two websites and to report any frustrating experience that you encounter. The researcher will be reading out the instructions and you will be asked some questions about your overall experience and your opinions. Finally, you will be asked some general demographic questions about yourself.

We are not going to be testing your abilities, but instead we are trying to find out how usable these websites are. The researcher will be taking notes and audio-record the session.

The data collected will be treated confidentially and anonymously by the research team.

We anticipate no potential risk associated with the study and you may withdraw participation at any time without repercussion.

Participation in this research study is voluntary, but we hope the results will have a positive impact on the whole screen reader community by helping to improve awareness about the issue of universal usability when translating a website.

Third section

Please confirm whether you agree with the following statements

I have read and understood the information provided about the study*

- □ Yes
- □ No

I confirm that I am over 18 years old*

- □ Yes
- □ No

I am aware that I will be asked to complete some tasks and express my opinions*

- □ Yes
- □ No

I am aware that my voice will be recorded*

- □ Yes
- □ No

I have understood that I can withdraw participation at any time*

- □ Yes
- □ No

I confirm that my involvement in the study is voluntary*

- □ Yes
- □ No

I agree to the confidential treatment of my data*

- □ Yes
- □ No

I consent to take part in this research project*

- □ Yes
- □ No

Instead of signing, please type today's date and your full name* _____

[End] = **Thank you for your time! We will contact you within 2 days.**

Appendix D - Test instructions

Introduction

The study is going to be structured as follows: I will give you instructions to perform three tasks on a website and ask you to report any frustrating experience that you encounter.
I will be reading out the instructions and you can interrupt me or ask me to repeat at any time.
At the end, I will ask you some questions about your overall experience and your opinions.
Then we will repeat the same steps for a second website.
Finally, I will ask you some general demographics questions about yourself.

Please remember that we are not testing your skills, but instead we are trying to find out how usable these websites are.
I will be taking notes and audiotape the interview.
Is everything clear so far?

Website A

I'd like to ask you to find some information by navigating the ZHAW website (that is the Zurich University of Applied Sciences). Please keep in mind two important things:

(1) We are not testing your skills: please behave as you would do if you were alone. You can navigate freely, switch between pages and click on links whenever you want.

(2) Every time you feel frustrated (for instance something doesn't work in the way you expected, the screen reader crashes or the content is not available), please stop for a moment and tell me. I will ask you what the problem is and how you solved it. Do you have any questions?

Please open the web site "www.zhaw.ch" and take a couple of minutes to explore the homepage.

Let's begin. Please imagine that you are a prospective student who lives in Zürich. You are hesitating between two very different types of studies.

1) First of all you would like to find the contact information of the university, so that you can ask some questions about the accessibility of the facilities. Please try to find the email address and telephone number of the ZHAW.

2a) Someone told you that the Institute of Translation and Interpreting offers a very interesting bachelor's programme and you would like to know more about it. Imagine you want to attend one of the open days for the IUED bachelor (that is the Institute of Translation and Interpreting). Try to find all the available dates in the summer and their time and location.

2b) You've already attended the open day at the School of engineering and they mentioned another event that is going to take place over the summer. To keep updated, you want to subscribe to the newsletter of the School of engineering.

[Post-task interview]

Website B

I'd like you to find some information on the BFH website (that is the Bern University of Applied Science). Please keep in mind two important things:

(1) We are not testing your skills: please behave as you would do if you were alone. You can navigate freely, switch between pages and click on links whenever you want.

(2) Every time you feel frustrated (for instance something doesn't work in the way you expected, the screen reader crashes or the content is not available), please stop for a moment and tell me. I will ask you what the problem is and how you solved it. Do you have any questions?

Please open the web site "bfh.ch" and take a couple of minutes to explore the homepage.

Let's begin. Please imagine that you are a prospective student who lives in Biel/Bienne. Someone told you that the Swiss literary institute (that is part of the Bern University of the Arts HKB) offers a very interesting bachelor's programme and you would like to know more about it.

1) First of all, you would like to find the contact information of the university, so that you can ask some questions about the accessibility of the facilities. Please try to find the email address and telephone number of the Swiss literary institute.

2a) Now imagine you want to attend one of the open days for the bachelor in Literary Writing. Try to find all the available dates in the summer and their time and location.

2b) You would like to know whether the library of the Swiss literary institute is close to where you live and whether it's open at the weekends. Please look for its address and opening times.

[Post-task interview]

[Post-test interview]

Appendix E - Post-task interview

Site _____

Which operating system did you use to complete the task?

- Windows 10
- Windows 8
- Windows 7
- Mac
- Linux
- Other _____

Which browser did you use to complete the task?

- Chrome
- Internet explorer
- Microsoft edge
- Mozilla Firefox
- Safari
- Other _____

Which screen reader did you use to complete the task?

- JAWS
- ZoomText
- Window-Eyes
- NVDA
- VoiceOver
- Other _____

Did you use any other assistive technology to complete the task?

- No
- Yes, a braille keyboard
- Other _____

Have you ever visited this website prior to today?

- No
- Yes, I've visited it once or twice
- Yes, I visit it from time to time (less than once a month)
- Yes, I visit it regularly (more than once a month)

Please indicate your degree of agreement with the following statements

1. Strongly disagree
2. Disagree
3. Somewhat disagree
4. Neither agree nor disagree
5. Somewhat agree
6. Agree
7. Strongly agree

1. Overall, I am satisfied with how easy it is to use this website
2. It was simple to use this website

3. I could effectively complete the tasks and scenarios using this website
4. I was able to complete the tasks and scenarios quickly using this website
5. I was able to efficiently complete the tasks and scenarios using this website
6. I felt comfortable using this website
7. It was easy to learn to use this website
8. I believe I could quickly learn how to use this website
9. The website gave error messages that clearly told me what the problem was
10. Whenever I made a mistake using the website, I could recover easily and quickly
11. Labels and menus of this website were clear
12. It was easy to find the information I needed
13. Labels and menus of the website were easy to understand
14. Labels and menus were effective in helping me complete the tasks and scenarios
15. The structure of the website was clear
16. The navigation of this website was pleasant
17. I liked navigating this website
18. This website has all the information I expect it to have
19. Overall, I am satisfied with this website

On a scale of 1 to 7, where 1 is 'not at all frustrating' and 7 is 'extremely frustrating', how frustrating was navigating this website for you overall?

How many minutes do you think you lost because of the problems you encountered?

_____

Is there anything else that you would like to add concerning the websites or the tasks you conducted?

_____

_____

_____

_____

_____

Appendix F - Post-test interview

Please indicate your age _____

Which is the highest level of education you have completed?

- Secondary school
- Bachelor's degree
- Master's degree
- Phd
- Other _____

What is your current occupation (if any)?

- Student
- Employee
- Accessibility/Usability expert
- Unemployed
- Other _____

How often do you use a computer to navigate the internet?

- Less than once a week
- Once or twice a week
- More than twice a week
- Daily

How long have you been using a screen reader?

- 1-2 years
- 3-5 years
- More than 5 years

How familiar are you with web accessibility?

- I've never heard of it
- I've heard of it but don't know what that is
- I know what it is, and I have limited experience with it
- I work in the field

Other than your mother tongue, please indicate in which of the following languages you can very easily read and understand almost any form of complex text (C1-C2 European level)

- German
- French
- Italian
- English
- None
- Other _____

Please indicate in which of the following languages you can read and understand texts written in everyday and contemporary language (B1-B2 European level)

- German
- French
- Italian
- English
- None
- Other _____

Please indicate in which of the following languages you can read and understand very short and simple texts (A1-A2 European level)

- German
- French
- Italian
- English
- None
- Other _____