

Jonna Jantunen

UNDERSTANDING UNDERSTANDABILITY
Evaluating Understandability as a Part of the Linguistic
Accessibility of a Text and Its Translation

Faculty of Information Technology and Communication Science
Bachelor's Thesis
August 2020

ABSTRACT

Jonna Jantunen: Understanding Understandability: Evaluating Understandability as a Part of the Linguistic Accessibility of a Text and Its Translation
Bachelor's Thesis
Tampere University
Bachelor's Programme in Languages: Specialization in English Language
August 2020

In this thesis, I attempt to answer the question of what makes a text understandable by creating a set of criteria for the evaluation of linguistic accessibility and understandability. The aim is to create a language-independent set of criteria that can be used both to evaluate a text and as a set of guidelines to follow when writing accessible content. The main resources I used in the creation of the criteria are techniques for making a text easier to read from the Web Content Accessibility Guidelines 2.1 and the Selkomittari, a tool meant to evaluate whether a text qualifies as simplified Finnish or not. Simplified language shares some features with accessible language, and this thesis focuses on the Finnish–English language pair, making the Selkomittari a valuable resource.

I tested the criteria by evaluating the accessibility of a text found on the website of the Finnish Digital and Population Data Services Agency. It is a public sector website published on January 1, 2020 and should thus be following the EU Web Accessibility Directive. The analyzed text focuses on the registration of foreigners in the Population Information System, which means that the main audience of the text might not be fluent in any of the languages the text is available in. This increases the need for accessibility and understandability. I analyzed both the original Finnish text and its English translation.

I found no major differences in accessibility between the two languages. Instead, they both seem to share the same accessibility issues – most significantly, long sentences with too much information and nondescriptive headings. Both versions had few issues related to grammar, vocabulary, or the overall structure of the text. Overall, both texts met the criteria to some extent. However, this is not sufficient, as it does not take those with cognitive issues or poor language skills into account.

The criteria I created seemed usable as a tool to evaluate the accessibility of a text, at least in the data I used. The criteria were relevant, and I found all criteria applicable to the texts.

Keywords: accessibility, accessibility of a translation, linguistic accessibility, understandability, WCAG 2.1, web accessibility directive

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Jonna Jantunen: Understanding Understandability: Evaluating Understandability as a Part of the Linguistic Accessibility of a Text and Its Translation
Kandidaatintutkielma
Tampereen yliopisto
Kielten kandidaattiohjelma: Englannin opintosuunta
Elokuu 2020

Tässä kandidaatintutkielmassa pyrin vastaamaan kysymykseen siitä, mitä ymmärrettävä teksti tarkoittaa, luomalla kriteeristön kielellisen saavutettavuuden ja ymmärrettävyyden arviointiin. Tavoitteena oli luoda kieliriippumaton kriteeristö, jota voi käyttää sekä tekstin saavutettavuuden arvioimiseen että saavutettavan sisällön kirjoittamiseen. Kriteeristön luomisen apuna käytin pääasiassa Verkkosisällön saavutettavuusohjeiden (WCAG 2.1) menetelmiä helppolukuisen tekstin kirjoittamiseksi, sekä Selkomittaria. Selkomittari on työkalu, jonka avulla on tarkoitus arvioida sitä, onko teksti selkokieltä vai ei. Vaikka Selkomittari keskittyy selkosuomen arvioimiseen, osoittautui se erittäin hyödylliseksi työkaluksi oman kriteeristöni luomisessa, sillä selkokielessä ja saavutettavassa kielessä on samoja ominaisuuksia. Lisäksi analyysini keskittyy suomi–englanti-kielipariin, joten selkosuomeen pohjautuvasta materiaalista on hyötyä myös siinä suhteessa.

Testasin kriteeristön toimivuutta arvioimalla sen avulla Digi- ja väestötietoviraston verkkosivustolta löytyvän tekstin saavutettavuutta. Kyseinen verkkosivusto on julkaistu 1.1.2020, ja koska kyseessä on julkisen sektorin verkkosivu, täytyy sen noudattaa EU:n saavutettavuusdirektiivin ohjeistuksia saavutettavuudesta. Analysoimani teksti keskittyy ulkomaalaisen rekisteröimiseen väestötietojärjestelmään, eli sen kohdeyleisö ei välttämättä ymmärrä sujuvasti yhtäkään niistä kielistä, joilla teksti on saatavilla. Tämä lisää saavutettavan ja ymmärrettävän sisällön tarvetta. Analysoin sekä alkuperäistä suomenkielistä tekstiä että sen englanninkielistä käännöstä.

En löytänyt kieliversioiden väliltä saavutettavuuden kannalta merkittäviä eroavaisuuksia. Sen sijaan kieliversioilla on samoja saavutettavuusongelmia – erityisesti liian pitkät lauseet ja epäselvät otsikot osoittautuivat ongelmiksi molemmissa teksteissä. Kummastakaan kieliversiosta ei löytynyt merkittäviä ongelmia kielioppiin, sanastoon tai tekstin kokonaisrakenteeseen liittyen. Molemmat tekstit vastasivat kriteeristön saavutettavuusvaatimuksiin jollain tasolla. Tämä ei kuitenkaan riitä, sillä tekstien saavutettavuustaso ei ota huomioon sellaisia lukioita, joilla on kognitiivisia ongelmia tai heikko kielitaito.

Luomani kriteeristö vaikutti käytettävältä työkalulta tekstin saavutettavuuden arvioimiseen, ainakin analysoimani datan perusteella. Kriteerit olivat relevantteja ja ne kaikki sopivat analysoimani tekstin arvioimiseen. Teksteistä löytyi kaikki ne ominaisuudet, joita kriteeristöni arvioi.

Avainsanat: kielellinen saavutettavuus, käännöksen saavutettavuus, saavutettavuus, saavutettavuusdirektiivi, ymmärrettävyys, verkkosisällön saavutettavuusohjeet, WCAG 2.1

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

1	Introduction	1
2	Theoretical Background	2
2.1	Accessibility	2
2.2	Understandability	5
2.3	Web Accessibility Directive	6
2.4	Web Content Accessibility Guidelines (WCAG) 2.1	7
2.5	Simplified Language	9
3	Research Data and Method	10
3.1	Criteria for Evaluating Linguistic Accessibility of a Text	10
3.2	Research Data	13
3.3	Method	14
4	Analysis	16
4.1	Grammar and Spelling	16
4.2	Vocabulary	18
4.3	Structure	19
4.4	Amount of Information	21
5	Discussion	22
6	Conclusion	24
7	References	26
7.1	Research Data	26
7.2	List of References	26

1 Introduction

Accessibility is an extremely important part of creating equality and equal opportunities for everyone. Unfortunately, the importance of language is often disregarded or diminished when considering issues of accessibility. Instead, when designing accessible digital content, attention is focused on the technical solutions. Although the technical solutions are undoubtedly important, I would argue that the use of language plays an important part as well. Additionally, I would argue that translation is an important part of accessibility, as a text only available in one language is inherently only accessible and usable to those who understand that specific language.

Accessibility, equality, and equal access to information are current topics in public discussions, and a growing number of websites aim to be accessible. Often, the first thought that comes to mind when discussing web accessibility is compatibility with screen reader software. Screen-reader software reads aloud the contents of a web page to aid the visually impaired. However, I would argue that the use of clear and accessible language is equally important, if not even more so, as anyone reading a text will benefit from it being written in an accessible, easy-to-understand way. Currently, there is little research in linguistic accessibility, especially that of web content, and no research that would take the issue of translation into account when considering accessibility.

The issue of accessibility has only become more prominent with the European Union's directive (EU) 2016/2102 on the Accessibility of the Websites and Mobile Applications of Public Sector Bodies (henceforth "Web Accessibility Directive"). The directive aims to increase equality in a growingly digital society and improve the quality of digital services by creating a list of minimum requirements for an accessible public sector website (Valtiovarainministeriö, n.d.). The Web Accessibility Directive draws on the Web Content Accessibility Guidelines (WCAG) 2.1, a set of recommendations for making web content accessible. The WCAG 2.1 considers linguistic accessibility briefly, but most of the recommendations, even those regarding language, are focused on technical solutions. They do, however, provide a short list of techniques for accessible writing, which I have used as one of the main resources for my thesis.

The other main resource for my thesis is Leealaura Leskelä's book on simplified Finnish (2019). Accessible language does not necessarily have to be simplified, but it shares some

features with simplified language. Both accessible and simplified language aim to increase understandability by using a limited set of words and short, simple sentences, although simplified language is more restricted. In addition, my thesis is based on the Finnish–English language pair, making Leskelä’s book relevant to it. Leskelä dedicates a significant portion of her book to explaining the Selkomittari criteria. The Selkomittari is a tool used to measure whether a text qualifies as Simplified Finnish. The tool is mostly focused on the use of language and thus it proved to be a very useful resource despite its focus on Finnish.

The aim of this thesis is twofold. Firstly, my goal is to create a simple and easy-to-follow set of criteria that can be used both to evaluate the linguistic accessibility of a text, and as rules to follow when writing an accessible text. The criteria introduced in this thesis attempt to be as language independent as possible. Secondly, I am interested in whether the accessibility of a text changes when it is translated. To evaluate this, and to test the criteria, I analyzed a Finnish text and its English translation from the website of the Digital and Population Data Services Agency (henceforth referred to as DVV, an acronym derived from the Finnish name Digi- ja väestötietovirasto).

Next, in chapter 2, I explain the main concepts and resources of this thesis in more detail and discuss the different aspects of accessibility. I introduce the criteria and discuss their creation, as well as introduce the research data and method in chapter 3. I then present the main findings of my analysis (chapter 4) and discuss them in more detail (chapter 5). In chapter 6, I conclude my research with a summary and a discussion on the differences between the language versions and the potential reasons for those differences.

2 Theoretical Background

In this section, I introduce the terms and theories that are most central to my thesis. In chapter 2.1, I introduce the concept of accessibility and explain the importance of linguistic and cognitive accessibility. I discuss understandability and its importance to linguistic accessibility in chapter 2.2, and shortly introduce the Web Accessibility Directive in chapter 2.3. Finally, I introduce the two main resources I used as a basis for my own criteria – WCAG 2.1 (chapter 2.4) and principles of simplified language (chapter 2.5).

2.1 Accessibility

Accessibility is a broad concept – it covers everything from whether a person using a wheelchair can easily access a building to whether a blind person can use a computer relying

solely on their available senses (Leskelä 2019, 48–49). Accessibility is an extremely important concept, as accessible solutions ensure independence for many people (ibid., 48). In this thesis, the focus is on linguistic web accessibility. By this I refer to whether the linguistic material on a website – for example, user interface texts, blog posts, and videos – is presented in a way that everyone can understand, regardless of their physical or cognitive abilities. This is still a very broad concept, so for the context of this thesis I will focus on the written content on web pages and its linguistic accessibility, ignoring the accessibility of technical solutions. I am also not considering typographical elements in this thesis.

Linguistic accessibility plays an important part in whether information is accessible to everyone or not. Access to information is a part of a person's basic rights, and unequal access to information has been shown to play a significant part in social exclusion (Leskelä 2019, 56). Additionally, the inability to access or understand information can cause issues in several areas of life, as the information can be crucial in taking care of one's health and wellbeing, applying for a job or welfare, finding an apartment, and other ways of participating in society (ibid.; Freyhoff et al. 1998, 7). Even everyday activities such as following the news, buying tickets for public transport, or reading recipes when cooking, can be difficult or impossible for some people if the information is not available in an accessible format (Leskelä 2019, 56). It has also been shown that the accessibility and understandability of information affects learning, especially with non-native speakers of the language, and with those suffering from learning difficulties (ibid., 59–60).

This unequal access to information does not only concern those who need accessible solutions to understand the information, but it can cause further issues in society. For instance, at the time of writing, the entire world is struggling to slow down the spread of the global COVID-19 pandemic. If some people cannot understand or access crucial information about the virus, they might not know how to protect themselves and others, how to avoid spreading the virus further, or even know the symptoms of the disease, and know when and how to get help. Therefore, I would argue that accessible information benefits the entire society.

Linguistic accessibility can be understood as a part of cognitive accessibility. Cognitive accessibility has to do with cognitive functions, such as memory, attention, visual perception, and language (Crabb et al. 2019, 2; Francik 1999, 31; Small et al. 2005, 1794). These functions are not independent; for instance, language, memory and attention all affect one's

ability to process and understand information. However, in this thesis, I will focus on language and treat it mostly independently from the other functions.

According to Francik (1999, 31), the language-factor of cognitive accessibility consists of the ability to speak, listen, write and type, and read. Of these, the ability to read is central to linguistic accessibility and this thesis. Similarly, Crabb et al. (2019, 2) recognize communicational issues as factors affecting accessibility. In their model, the communicational factor is concerned with issues of hearing, talking, reading, and expression (ibid.). However, not everyone recognizes language as an integral part of cognitive accessibility. Small et al. (2005, 1794) focus on web navigation and, building on previous research, identify four cognitive components relevant to their topic: situational awareness, spatial ability, task-set switching, and anticipated system response. None of these components mention language, although I would argue that it plays an important role in successful web navigation as well. For example, if a user cannot understand what a link says, they cannot determine which link to select in order to achieve the desired result. Small et al. do, however, mention the user's ability and willingness to read instructions as one of the factors affecting navigation (ibid., 1796).

Cognitive accessibility is especially important to those who might have trouble understanding written information due to cognitive disabilities such as autism, dementia, attention deficit disorder (ADD), dyslexia, or learning difficulties (Leskelä 2019, 49; Papunet 2019; WebAIM 2018). Additionally, old age brings with itself cognitive changes, especially in attention and memory. Older people often experience difficulties in focusing, maintaining attention, and processing multiple tasks simultaneously, all of which affect the ability to understand and process information (Lewis 2008, 52). The changes in memory negatively affect the ability to acquire new information and retain that information (ibid., 53; Leskelä 2019, 100). In addition, many mental health disorders, such as depression, can cause difficulty focusing or processing information (Leskelä 2019, 101). There are also situations in everyone's life where they would benefit from linguistic accessibility, as such common issues as tiredness, headache, or a noisy environment, can temporarily hinder information acquisition and processing (ibid., 102).

In addition to the groups mentioned earlier, linguistic accessibility is important to those who do not speak the language fluently. For instance, new immigrants who need information about social security or other important issues would greatly benefit from accessible language

(Leskelä 2019, 109, 104–105). Additionally, accessible language benefits those who do not speak the language as their mother tongue, for instance, those who have first acquired a sign language, as studies have shown that they often lack skills in the spoken language dominant in the society they live in (ibid., 102).

Cognitive accessibility is often left with little to no attention when considering issues in accessibility. This is due to the difficulty in defining and evaluating it, as well as the wide range of disabilities that can cause issues in cognitive skills (Leskelä 2019, 49; Lewis 2008, 18). These disabilities can be difficult to diagnose and might vary from person to person (Lewis 2008, 18; Papunet 2019, WebAIM 2018). This means that not all solutions will work for everyone. For example, one person might benefit from images that explain the concept rather than text, while another might be hindered by the images. Cognitive accessibility is not easy to achieve, and it is recommended to offer multiple ways to understand the content, such as text, images and the possibility to have the text read aloud (Papunet 2019; WebAIM 2018).

This thesis focuses on linguistic accessibility of written text and thus only focuses on the ways to make a text easier to understand using language. Clear language that is easy to understand benefits most of the people who benefit from cognitive accessibility in general, and in fact most people with no cognitive disabilities would benefit from it too. It has been shown that the difficulties encountered by people with different cognitive disabilities are very similar to, if not the same as, the various usability issues that affect all users – the impact is simply more severe on those with cognitive disabilities (Lewis 2008, 18–19). In addition, accessible, easy-to-understand language often helps when trying to learn about a new, difficult topic, even with those who do not have any cognitive difficulties otherwise (Leskelä 2019, 57), as they can focus their attention to the topic at hand, instead of worrying about not understanding the difficult jargon.

2.2 Understandability

In the context of my thesis, the principle of understandability is central. Simply put, understandability refers to whether something, in this case a text, is understandable – how easily the reader can comprehend the main idea behind the text and draw conclusions from it (Francik 1999, 31; Rello et al. 2013, 203). Understandability is closely related to readability, which Rello et al. define as “the legibility of a text, that is, the ease with which text can be read” (2013, 203). The terms are often used interchangeably, as the readability of a text strongly correlates with the understandability of the text (ibid., 205). Naturally, if the text is

not readable, that is if it cannot be read, it also cannot be comprehended. In this thesis, I will focus on understandability and not consider elements that solely affect the readability of a text, such as typography and the layout of the text.

Understandability is not easy to define or measure, as what is understandable to one person might not make sense to another. The Web Accessibility Directive simply states that “information and the operation of the user interface must be understandable” without further defining the concept of understandability (2016/2102, 5). This leaves the reader, and the public sector personnel responsible for making their websites accessible, with the question of what understandable information and user interface operation is. As this thesis focuses on linguistic accessibility, I will not discuss the understandability of user interface operation further. However, I attempt to answer the question “what is understandable information”, at least to some extent and aim to create a set of criteria to help write texts that are understandable to as wide an audience as possible.

Existing research into the accessibility of websites often considers information accessible when the user can find and perceive it. However, as Kato and Hori point out (2006, 287), information should not be considered accessible solely on the basis of it being findable and perceivable, because if the user does not understand the information, it is not accessible to them. Of course, the information must be findable and perceivable before the user can begin to understand it, but even that is affected by understandability to some extent – the instructions on finding information as well as the navigation system need to be understandable in order for the user to find the information they are looking for (ibid.). As can be seen from this, the different dimensions of accessibility overlap, which makes defining and analyzing only one difficult.

2.3 Web Accessibility Directive

The Web Accessibility Directive (2016/2102) came into effect in December 2016. The directive is a part of the EU’s actions in ensuring equal access to information, regardless of the person’s possible disabilities (Web Accessibility Directive (EU) 2016/2102, 2). Thus, the main aim of the directive is to increase equality and the quality of digital services. The member states of the EU are responsible for creating national legislation that complies with the directive. In Finland, this legislation, the Act on the Provision of Digital Services (306/2019), came into effect on April 1, 2019, and, simply put, states that all Finnish public sector bodies’ web and mobile content must be accessible in the way the Web Accessibility

Directive mandates. The schedule for this change is the following: all public sector websites launched after September 23, 2018 had to be accessible by September 23, 2019, and websites launched before September 23, 2018 must be accessible by September 23, 2020 (Web Accessibility Directive (EU) 2016/2102, 14; Act on the Provision of Digital Services (306/2019)).

The directive is based on the European Standard (ETSI) EN 301 549 Accessibility Requirements Suitable for Public Procurement of ICT Products and Services in Europe, which in turn is based on the Web Content Accessibility Guidelines (WCAG) 2.1 (Leskelä 2019, 51; Web Accessibility Directive (EU) 2016/2102, 7–8). The directive, the European Standard EN 301 549, and the WCAG 2.1 all mention that information should be accessible and understandable, but they all lack clarifications of what understandability means and how it should be achieved.

The Web Accessibility Directive defines accessibility as consisting of four main principles: perceivability, operability, understandability, and robustness (2016/2102, 5). In practice, especially from the point of view of language and communications, this means that the information must be presented in a way that the user can perceive and understand, and in a way that is possible to be interpreted by different users and user agents such as assistive technologies (ibid.).

2.4 Web Content Accessibility Guidelines (WCAG) 2.1

The Web Content Accessibility Guidelines (WCAG) 2.1 is a set of criteria, guidelines, and recommendations for making web content more accessible. It was developed by the World Wide Web Consortium (W3C) in cooperation with multiple international individuals and organizations (W3C 2018b). Their main goal was to create “a single shared standard for web content accessibility that meets the needs of individuals, organizations, and governments internationally” (ibid.). Version 2.1 of the WCAG was published on June 5, 2018, approximately ten years after version 2.0 (W3C 2018a).

The WCAG 2.1 is divided into four sections: “Perceivable”, “Operable”, “Understandable”, and “Robust”. These are the same four categories that the Web Accessibility Directive calls the principles of accessibility. These principles are divided into a total of 13 guidelines, which are then further divided into success criteria (W3C 2018a). The WCAG includes three levels

of conformance: A, AA, and AAA. The Web Accessibility Directive requires public sector bodies to fulfill level AA guidelines (European Standard (ETSI) EN 301 549).

The third principle, “Understandable”, is the most relevant to the topic of this thesis. This principle consists of three guidelines: “Readable”, “Predictable”, and “Input assistance”. The success criteria under “Predictability” and “Input assistance” focus solely on technical criteria and are thus not relevant to the scope of this thesis. The only level AA success criteria under “Readability” is the criteria 3.1.2 “Language of parts”, and it states that “[t]he human language of each passage or phrase in the content can be programmatically determined except for proper names, technical terms, words of indeterminate language, and words or phrases that have become part of the vernacular of the immediately surrounding text” (W3C 2018a). All the success criteria related to linguistic accessibility, rather than technical accessibility, are level AAA and are thus not required by legislation. However, I would argue that these criteria should still be taken into consideration. While they might not be required by this legislation, they nevertheless point to the direction of what is considered successful in these guidelines. Thus, by taking the level AAA criteria into account, one can identify the aspects of understandability that the guidelines consider important.

In this thesis, I will focus on WCAG 2.1 success criteria 3.1.5 “Reading level”. It states that a text should be understandable to a person with a lower secondary education level, and if it is not, an alternative and more understandable version should be available (W3C 2018a). W3C provides techniques to help meet the criteria. I have used the technique G153 “Making the text easier to read” as a source material when creating my own criteria. Interestingly, this technique is concerned with the readability of the text, although what is considered readability in the technique is more similar to the definition of understandability by Rello et al. (2013, 203) than their definition of readability. This is further evidence of the terms being used interchangeably.

W3C admits to facing challenges in defining criteria relevant to cognitive and linguistic accessibility due to a tight schedule and “challenges in reaching consensus on testability, implementability, and international considerations of proposals” (W3C 2018a). These are all common challenges when considering cognitive accessibility, as discussed earlier. They also state that they will continue working in this area in the future versions of the WCAG (ibid.). It will be interesting to see if the next version will have more clearly defined recommendations for linguistic accessibility.

2.5 Simplified Language

Simplified language, sometimes called controlled language, refers to an artificially created variety of language aimed at those with cognitive disabilities or other difficulties understanding standard language (Kuhn 2012, 102; Leskelä 2019, 93). Simplified language builds on standard language – or rather takes from it, creating a clearer and more understandable version of standard language. Simplified language is a condensed version of standard language using only the parts that are the easiest to understand (Leskelä 2019, 95). This means, for example, using a smaller set of vocabulary and paying careful attention to difficult sentence structures (*ibid.*). The term ‘controlled language’ is often used to refer specifically to the restricted variety of language used in technical documentation and machine translation (Jenge et al. 2006, 208).

Simplified language, while accessible, is not the same as accessible language. Simplified language is accessible, but not all accessible language is simplified language – and does not have to be. In this thesis, I use the term accessible language to primarily refer to accessible standard language, not simplified language. Both accessible standard language and simplified language are somewhat restricted and planned (Leskelä 2019, 95). However, simplified language is more restricted than accessible standard language, so much so that it constitutes a variety of standard language (*ibid.*; Saavutettavasti.fi n.d.).

Simplified language is specifically targeted at those who have trouble understanding standard language due to, for instance, cognitive disabilities (Leskelä 2019, 94). Standard language, on the other hand, is targeted at the general audience – everyone with sufficient language skills to read the text (*ibid.*). Accessible language is a slightly restricted and modified version of standard language that takes those with difficulties understanding standard language into account but is not specifically targeted at them (*ibid.*). Using simplified language often requires more resources and specialized knowledge than using accessible standard language. Thus, even though the importance of simplified language should in no way be disregarded, I would argue that most websites should aim to use accessible language at least to some degree, as it extends their audience to include those with difficulty understanding standard language.

In her book on simplified language and accessibility, Leskelä (2019) focuses on simplified Finnish, but similar principles can be used in other languages as well. Leskelä also explains the need for linguistic accessibility and simplified language in detail, and these reasons and justifications are universal. Because of this, I found it to be an excellent source. Additionally,

a significant amount of the information the book presents can easily be applied in other languages too. While the more detailed explanations on how to use simplified Finnish are not as useful when trying to create a universal, language-independent set of criteria, they provide a solid base. While my goal was to create as language-independent set of criteria as possible, I am most familiar with Finnish and English and thus the criteria is bound to be influenced by those languages. Additionally, I tested my criteria with Finnish and English texts.

A significant portion of Leskelä's book is dedicated to going through the criteria of the Selkomittari and explaining them in more detail (2019, 109–210). The Selkomittari is a tool created by Selkokeskus, the Finnish center for simplified language, to measure the understandability and readability of a text. It consists of 21 sections and a total of 106 statements, all of which are given either a plus (+) indicating that the statement is true, a minus (-) indicating that the statement is false, or a zero if the statement in question is not relevant in the text under evaluation (Selkokeskus 2018, 3). Each section is then given a grade between 1 (worst) and 3 (best) based on the number of pluses and minuses under that section (ibid.). The higher the average score of all sections, the closer to good simplified Finnish the text is. I used the Selkomittari as a basis for my criteria, which I will introduce in chapter 3.1.

3 Research Data and Method

3.1 Criteria for Evaluating the Linguistic Accessibility of a Text

The main reason I created my own set of criteria for accessible writing was that most of the current instructions and criteria I was able to find seemed vague and lacking. The WCAG 2.1 offers the techniques introduced earlier but, despite the effort to make them language independent, they are mostly focused on writing in English. Additionally, these techniques themselves are not very accessible. They are not written consistently and are somewhat repetitive, the recommendation to use short, common words being mentioned in multiple instructions in slightly different ways. Finally, the techniques are not the easiest to find, unless one knows what one is looking for.

Another reason I wanted to create my own criteria is the fact that the WCAG 2.1 techniques do not include some techniques that I consider important. For instance, the WCAG 2.1 techniques do not mention the amount or order of information at all, even though those are important points to consider when aiming for an accessible text (Leivo 2019; Leskelä 2019, 68, 118–121). Instead, the techniques are focused on word- and sentence-level rules. To

complement the WCAG 2.1 techniques, I drew from the Selkomittari, which includes criteria related to the order and amount of information. As mentioned, the Selkomittari is specific to the Finnish language, so I could not use all of the criteria without modification. However, they did provide a good basis for creating language-independent criteria as well.

The Selkomittari is quite extensive, and I wanted to keep my criteria short, so I had to decide which of the Selkomittari criteria were the most relevant. The WCAG 2.1 technique G153 on making the text easier to read, on the other hand, provides a very limited amount of instructions. I started with the WCAG 2.1 criteria and compared it to the Selkomittari criteria, selecting the ones included in both as the basis for my criteria. When necessary, I attempted to modify the criterion to make it as language independent as possible. As discussed earlier, the WCAG 2.1 techniques are somewhat repetitive, so I combined the ones concerned with similar topics into single criteria.

After that, I went through the Selkomittari criteria in more detail. I excluded all of the criteria that were exclusively related to simplified language, as my goal was not to create criteria for simplified language. I also excluded criteria 83–106, as those have to do with the visual layout of the text and are thus also outside the scope of my thesis. Finally, I excluded the criteria that were specific to the Finnish language and not extendable to other languages. As the Selkomittari is so extensive, it is bound to have some repetitiveness. Therefore, my goal was to select only one criterion from each section of the Selkomittari. This goal was not fully realized, as some of the sections are more extensive than others and thus include more relevant criteria. I also attempted to combine any similar criteria with each other, as I did with the WCAG 2.1.

I wrote my criteria in the imperative form, making them easy to use as a set of rules to follow when writing accessible text. However, they can also be used to evaluate the accessibility of a text. I attempted to make the criteria accessible by having them adhere to themselves. My goal was to create a set of criteria that anyone creating written web content can use to make their writing more accessible. Some of the criteria include examples to help understand the criteria. The examples are of Finnish and English, as those are the two languages I myself am familiar with. As my goal was to create a set of criteria that is as language independent as possible, some of the criteria might seem vague. I attempted to avoid any language-specific rules and instead provide a general guideline that can be adjusted to each language's specific grammar and vocabulary. The criteria are introduced below in table 1.

Table 1. Criteria for Evaluating the Linguistic Accessibility of a Text

#	Criteria
1	Follow the grammar and spelling rules of the language or variant you are using.
2	Tell the reader everything that they need to know, but nothing more. If necessary, direct the reader elsewhere for additional information.
3	Use descriptive headings. Make sure the heading matches the content of the paragraphs below it.
4	Divide the text into logical sections. Use subheadings to help the reader understand the structure of the text.
5	Structure the text logically. For example, follow the order in which things happen or are done.
6	Cover only one idea per paragraph. Start a new paragraph whenever you introduce a new topic or viewpoint.
7	Tell the most important thing first.
8	Use lists instead of paragraphs that contain long series of words or phrases separated by commas.
9	Use the active voice unless there is a specific reason to use the passive voice (for example, in situations where the subject is unknown or irrelevant).
10	Use the simplest sentence structure possible. In English and Finnish, this usually means following the Subject-Verb-Object structure.
11	Use verb tenses consistently.
12	Use short sentences. Cover only one point per sentence.
13	Use common, everyday words. Avoid professional jargon and other terms with specialized meanings. Replace long or unfamiliar words with shorter, more common words.
14	Explain words that might be hard to understand if it is necessary to use them.
15	Avoid using abbreviations. Explain all uncommon abbreviations and acronyms the first time you use them.
16	Make clear pronoun references. Repeat the subject if it makes the text easier to understand.

17	Use verbs to express what is being done. For example, in English, instead of saying “They performed an investigation” say “They investigated”.
18	Add cohesion by indicating logical relationships between phrases, sentences, paragraphs, or sections of the text. For example, in English, add cohesive devices such as “additionally”, “but”, or “however”.
19	Use names, terms, and labels consistently.

3.2 Research Data

After creating the criteria, I tested it with an existing public sector website. I decided to use content available on the DVV’s website. I chose this specific website since it is a public sector website that was released on January 1, 2020 and is thus legally required to follow the Web Accessibility Directive. Additionally, the DVV’s website includes information that is important for everyone in Finnish society, such as information on marriage, death, childbirth, and moving. This means that the information they present should be accessible to people with and without disabilities. Most of the information is also relevant for immigrants and residents who do not speak Finnish or Swedish, which increases the importance of accessible and clear English translations. The website offers information in both Finnish and English, which was an important criterion when selecting a website to test the criteria on, because my goal was to test it on both languages. In this case, the importance of the English translation is also relatively high, as over 7% of Finland’s population has something other than Finnish, Swedish or Sami as their native language (Statistics Finland n.d.).

As per the Web Accessibility Directive (2016/2102, 11–12), the DVV offers an accessibility statement which, at the time of writing, states that the website mostly fulfills the level A and AA accessibility recommendations of WCAG 2.1 (DVV 2020a; DVV 2020b). As noted earlier, the recommendations related to linguistic accessibility are all of level AAA (W3C 2018a) and thus not required by law. The DVV does not mention language in their list of features that are not yet accessible (DVV 2020a; DVV 2020b), but this could be because the linguistic accessibility is not a legal requirement. At the time of writing, the accessibility statement was the same in both Finnish and English, which should mean that there are no differences in accessibility between the language versions. However, since the accessibility statement does not mention language, it is not possible to determine whether the linguistic accessibility is the same from the accessibility statement alone.

Selecting a web page for analysis turned out to be more difficult than I had anticipated, since multiple pages on the DVV's website are currently missing English translations. An example of a missing translation is a page titled "Suomeen muutto" ("Moving to Finland"), which includes important information on what a person moving to Finland should know and do. This especially is a page that should be available in English, because a person moving to Finland might not understand any Finnish or Swedish. Although it would be best if all of the information on the website was available in English, I would argue that pages that include information clearly addressed to immigrants should be the first ones to be translated into English. Additionally, the English version should be made as accessible as possible, since not all immigrants will be fluent in English. Presenting the information in a clear and easy-to-understand way would increase the chances of everyone understanding the important information available on the website.

In the end, I was able to find a page that both had a comprehensive English translation and included information I deemed especially important to have available in English due to the main audience of the text. The page I selected is titled "Ulkomaalaisen rekisteröinti väestötietojärjestelmään" and its English translation is titled "Registration of a foreigner in the Population Information System". The page includes information about personal identity codes, registration of personal data, and municipality of residence. The information regarding personal identity codes is especially important to have in an accessible form, since a personal identity code is often necessary to live and operate in Finnish society.

3.3 Method

I started the analysis by copying the text from both the Finnish and English page into a Word-document to avoid any updates to the web page affecting my analysis. I did this on May 1, 2020 and used these copies for the analysis. As I focused my analysis solely on the linguistic material, the font or other typographical and visual elements of the web page were not relevant to the analysis. Therefore, copying the text to a separate document did not affect the results of the analysis.

During the analysis, I went through my criteria one by one and thought about ways to analyze each. Some of the criteria were countable, for example in the case of criteria 8, I could count the instances where lists were and were not used instead of paragraphs containing long series of words or phrases separated by commas. Others were more difficult to analyze, for instance whether the text was structured logically. In these cases, I relied on my own evaluation of

whether the criterion was fulfilled or not. Due to this, the analysis is somewhat subjective and does not provide a completely objective evaluation on the accessibility of the text. The questions I used as a basis for my evaluation are introduced in table 2 below.

#	Questions
1	Were there any grammar mistakes? Did the English text use only one variant?
2	Were there points where I would have needed more information? Were there points where a person who is not familiar with Finnish culture and government would have needed more information? Was there information not directly related to the topic?
3	Did the paragraphs contain the information that I anticipated they would, based on their headings?
4	Was information on the same topic in the same place, for example, under the same subheading?
5	Were there any inconsistencies in the order of information? For example, did I always have all the information I needed to understand the text?
6	Were there paragraphs with more than one main idea or viewpoint?
7	Did the beginning of each section introduce its main ideas before delving into details?
8	Were there any instances where long series of words or phrases separated by commas were used? Could bulleted lists have been used instead?
9	Was the passive voice used? Was it justified, for example, was it used to shift the focus away from the subject? Was it still clear to the reader what they are expected to do?
10	Was the Subject-Verb-Object structure followed? If not, why? Could the sentence have been formed using the Subject-Verb-Object structure or some other simpler structure?
11	Were there any instances where the verb tense was different from the rest of the text? Was it justified, for example, was the past tense used instead of the present tense because the text described past events?
12	Were there sentences with more than one main idea?
13	Were jargon or uncommon words used? Could these have been replaced with shorter or more common words without changing the meaning of the text?
14	If jargon or difficult words were used, were they explained?
15	Were abbreviations or acronyms used? Were they common? If not, were they explained?

16	Was it always clear what the text is referring to? How often was the subject repeated instead of using a pronoun?
17	Were there any instances where nominalized expressions were used? Could these have been replaced with more common verb structures?
18	Were cohesive devices used? Was the logical relationship between phrases, sentences, paragraphs, and sections of the text always clear?
19	Was one thing always referred to with the same name, term, or label?

After I had decided how to analyze each criterion, I went through each text once, noting any obvious accessibility issues. After that, I went through the texts while paying careful attention to one criterion at a time and taking note of any sections relevant to that criterion, whether positive or negative. I will introduce my findings and discuss the results of my analysis in more detail in the next chapter.

4 Analysis

The analysis showed that all the criteria were relevant in the texts and there were no criteria with no findings. In this chapter, I will introduce the main findings from both the Finnish and the English text. As some of the criteria overlap and interact with each other, I have divided this chapter into sections in which I discuss the different aspects of the text and their effects on accessibility rather than focusing on single criteria. The examples I present in this section are mostly from the English text and, unless otherwise stated, the same problems are present in the Finnish version as well.

4.1 Grammar and Spelling

As grammar rules can vary significantly between languages, I decided to not include any strict grammar rules in my criteria. Instead, criterion 1 simply states that the text should follow the grammar and spelling rules of the language or variant the text is written in. However, the criteria still include a couple of grammar-related criteria, most significantly criteria number 9, 10, and 16. Despite the variance in grammar rules, I determined that the use of active voice and simple sentence structure, as well as clear pronoun references, are aspects that could increase the accessibility of a text in many languages.

Overall, both texts have very few issues regarding grammar and spelling. The English text has one missing article, as can be seen in example 1 below, but the rest of the text is

grammatically correct. The Finnish text, on the other hand, has a couple of grammar mistakes and one misspelled word – *kotikuntaa* instead of *kotikuntaan*. The Finnish text also has one grammatically incorrect sentence, introduced in example 2 below. The problem here is that the Finnish noun *kotikunta* ‘municipality of residence’ inflects differently after verbs *olla* ‘to have’ and *saada* ‘to get’. In this sentence, the noun is only inflected after *saada*, making it asymmetrical and thus incorrect.

- (1) You may get a municipality of residence without personal visit.
- (2) Edellä mainitun lisäksi saat kotikunnan, jos Suomessa asuvalla perheenjäsenelläsi on tai hän saa kotikunnan. Jos saat kotikunnan, muutkin Suomessa asuvat perheenjäsenesi saavat kotikunnan.

The rest of the grammar issues in the Finnish text are related to lists. Finnish has strict rules to punctuating lists, and those are not always followed in the text. In Finnish, the list items that are not complete sentences are not supposed to be separated with commas and there should be a period after the last item. However, this rule is not followed consistently in the text, as some lists are missing the final period. Additionally, one of the lists separates items with commas, as can be seen in example 3 below. In this case, the purpose of the commas and the *ja* ‘and’ between the last two items is to indicate that all the listed requirements must be fulfilled in order to get a municipality of residence without a personal visit. However, it is not grammatically correct, and it would be more accessible to state that all the requirements need to be fulfilled already in the introductory sentence.

- (3) Voit saada kotikunnan ilman henkilökohtaista asiointia Digi- ja väestötietovirastossa jos:
 - olet EU-maan, Sveitsin tai Liechtensteinin kansalainen tai tällaisen henkilön perheenjäsen,
 - olet rekisteröinyt EU-kansalaisen oleskeluoikeutesi Maahanmuuttovirastossa tai sinulle on myönnetty perheenjäsenen oleskelukortti,
 - olet jo saanut suomalaisen henkilötunnuksen esimerkiksi oleskeluoikeuden rekisteröinnin yhteydessä, ja
 - muutat Suomeen muualta kuin Ruotsista, Norjasta, Tanskasta tai Islannista. Pohjoismaiden kansalaisten tulee käydä henkilökohtaisesti Digi- ja väestötietovirastossa saadakseen kotikunnan.

Both texts use the passive voice fairly often, but it is always justified (criterion 9). The passive voice is used mostly in situations where the subject is unknown or irrelevant, and the author wants to keep the focus on the reader instead of the authorities behind the actions. Additionally, the verb tenses are used consistently (criterion 11) and references are made clear by repeating the subject often (criterion 16) throughout both texts.

Generally, both texts use the simplest sentence structures possible (criterion 10). The English text has one sentence with a somewhat complicated structure, as can be seen in example 4.

The Finnish text has one similar issue, introduced in example 5 below. In this example, the phrase *voi olla tarpeen* ‘might be necessary’ is unnecessarily complicated. It would be clearer to say, for example *voit tarvita* ‘you might need’.

- (4) The preconditions for registration will then be examined by the authority.
- (5) Joidenkin asioiden hoitamiseen henkilötunnus voi olla tarpeen, vaikka et asuisikaan Suomessa (To take care of some matters, a personal identity code might be necessary even if you do not live in Finland [my translation]).

Overall, the sentences are clear and easy-to-follow, although some are long and include a lot of information. This will be discussed in more detail in section 4.3. Finally, both texts performed well in the evaluation of criterion 17, “Use verbs to express what is being done”. There is only one expression in the Finnish text that could have been modified to match the criterion better: *esittää pyynnön* ‘present a request’ could simply be *pyytää* ‘to request’.

4.2 Vocabulary

As discussed, the vocabulary used in a text affects its accessibility significantly. An accessible text uses short, common words (criterion 13), explains any uncommon words that are necessary to use (criterion 14) and avoids abbreviations (criterion 15). Additionally, the use of terms should be consistent throughout the text, using only one term to refer to one thing (criterion 19). The texts I analyzed managed to fulfill these criteria well, with only a few points that would benefit from editing to increase accessibility.

Neither of the texts use abbreviations, with the exception of a single *esim.* ‘e.g.’ in the Finnish text. This is a very common abbreviation and the use is thus justified. Additionally, the vocabulary used in both texts is common and easy to understand, except for *Population Information System*. The system itself is not explicitly explained in the texts, but it is mentioned that a person must be registered in it to receive a personal identity code. In this context, this explanation is most likely enough. In addition, both texts mention *legalised documents*, but the meaning of *legalised* is not explained, as is visible in example 6. Whether *legalised* is considered a common word probably depends on the reader. Finally, the Finnish text uses the phrase *avioliitonomaisissa olosuhteissa* ‘marriage-like conditions’, which is not explained. The English translation, on the other hand, simply uses the phrase *as a couple*. Example 7 below shows this in context.

- (6) the legalised, original and translated birth and marriage certificates and other official documents containing information that you would like the Digital and Population Data Services Agency to record in the Population Information System.

(7) FI: avopuolisosi, jos olette asuneet yhdessä avioliitonomaisissa olosuhteissa vähintään kaksi vuotta (your cohabiting partner, if you have lived together in marriage-like circumstances for at least two years [my translation])

EN: your cohabiting partner, if you have lived together as a couple for at least two years

For the most part, the terms and other words are used consistently throughout the texts. The English text uses *personal identification code* instead of *personal identity code* in one sentence, but otherwise both texts are consistent in the use of terms. However, both texts have some variation in the use of verbs, as the Finnish text uses both the verbs *tallentaa* ‘to save’ and *rekisteröidä* ‘to register’ when referring to the registration of information to the Population Information System. The English text also uses the verbs *to register* and *to record* in the same context. Additionally, both texts mention registering both the person and the person’s personal data into the Population Information System, which might leave the reader confused. Examples 8 and 9 below demonstrate these inconsistencies.

- (8) FI: Jotta saisit henkilötunnuksen, henkilötietosi on tallennettava Suomen väestötietojärjestelmään.
EN: Your personal data must be recorded in the Finnish Population Information System before a personal identification code can be given to you.
- (9) FI: Jos sinut rekisteröidään väestötietojärjestelmään, seuraavat tiedot rekisteröidään aina.
EN: If you are registered in the Population Information System, the following information is always registered.

4.3 Structure

The structure of a text also significantly affects its accessibility – is the text cohesive (criterion 18), does it follow a logical structure (criteria 4, 5 and 7) and do the headings help navigate the text (criterion 3)? Additionally, I have included criterion 8, the use of lists instead of long series of words or phrases separated by commas, in this section. I consider sentence structure more a matter of grammar and have thus discussed in section 4.1.

The overall cohesion in the texts is good – both texts include cohesive devices such as *then*, *in addition*, *also*, and *but* that make the text easy to follow. The texts are divided into logical sections and follow a logical order with the most important points introduced first. However, the list of requirements for a municipality of residence under the subheading *Who can get a personal identity code?* is slightly illogical, as it is not directly related to the heading.

However, it is related to a previously mentioned criteria for getting a personal identity code. It is therefore justified to have the list there, but as the list is followed by more lists related to getting a municipality of residence, it might make more sense to include all of those in their own section with a separate subheading.

The most significant issues with the structure of the texts are the long list items under the heading *Personal identity code*, one of which is introduced in example 10 below. The entire section consists of one list with three long list items. Each item even includes multiple points of information relevant to the reader, which I will discuss more in section 4.4. I would argue that this specific list makes the texts harder to follow since the items are so long and have so much information. Usually lists are used to introduce multiple short ideas that all relate to each other, as is done in the other lists included in these texts.

- (10) **If you have a residence permit or a residence card issued by the Finnish Immigration Service, or if the Finnish Immigration Service has registered your right of residence**, you have usually been assigned a personal identity code. The code can then be found on the reverse side of the residence permit card. The same applies to other situations in which the Finnish Immigration Service has made some other positive decision concerning your stay in Finland. If you then want to register information about your family in the Population Information System or want to have a municipality of residence in Finland, contact the Digital and Population Data Services Agency and request registration or a municipality of residence. The Digital and Population Data Services Agency will then examine the preconditions for registration and your right to a municipality of residence.

As mentioned earlier, the texts are divided into logical sections and each section has its own heading. However, some of the headings are not very accessible. The main heading of the text, *Registration of a foreigner in the Population Information System*, is not the most accessible choice for this text, because most of the text is actually concerned with getting a personal identity code and a municipality of residence, which require the person's registration in the Population Information System. The registration of information in the Population Information System is simply mentioned as a requirement to get the personal identity code and the system itself is not discussed in more detail.

The heading *Personal identity code* is used twice in the text – first in a section explaining what the personal identity code is and what it is used for, then in a section explaining how to get a personal identity code. It is not accessible to repeat a heading, unless the contents under the headings are the same or at least similar, and I would argue that both of these could be made more accessible by adding a couple of words, such as *What is* and *How to get*. Additionally, the heading *Do the following*, while technically descriptive is not accessible as the reader cannot immediately determine why they should do the following, based on the heading alone. Instead, it would be better to include the reason in the heading, for example *Do the following to get a personal identity code*. Finally, the last heading of both texts, *Registration of the municipality of residence*, does not match the contents below it. Instead, this section is about getting a municipality of residence if one's family member gets one.

4.4 Amount of Information

As discussed earlier, the amount and order of information can significantly affect the accessibility and understandability of a text. The order of information was already discussed in the previous section. Therefore, here I focus on whether the reader gets enough information (criterion 2) and whether the information is divided into small enough sections for the reader to easily comprehend (criteria 6 and 12).

For the most part, the texts provide enough information on the topics of the personal identity code and the municipality of residence. The reader is most likely capable of applying for a personal identity code and a municipality of residence based on these texts and there is no unnecessary information. However, the texts have multiple vague sentences, in which words like *sometimes* and *usually* are used, such as the one in example 11 below. In this example, the reader is not told in which situations they might need a personal identity code when they do not live in Finland. Instead, the text uses the phrases *may need* and *some matters*, possibly leaving the reader wondering when and for what they might need a personal identity code.

- (11) Even if you do not live in Finland, you may need a personal identity code to be able to take care of some matters.

Additionally, in example 12, the conditions on which the reader can be assigned a municipality of residence are not elaborated on further. However, these conditions are explained later in the texts, and I would argue that it would make the text more accessible if the fact that the conditions are discussed later would be mentioned in this context as well. Similarly, in example 13, the reader is instructed to contact the DVV, but is not told how to do so or where to find the contact information.

- (12) If you live in Finland, you can be assigned a municipality of residence in Finland on certain conditions.
- (13) If you then want to register information about your family in the Population Information System or want to have a municipality of residence in Finland, contact the Digital and Population Data Services Agency and request registration or a municipality of residence.

While the texts as a whole mostly provide the reader with enough information, they did have issues with the amount of information on sentence- and paragraph-levels. As a lot of the information is presented in bulleted lists, the amount of information on the level of a paragraph is mostly appropriate. However, the long list items under the second *Personal identity code* heading do cause an accessibility issue. The first two items, one of which is introduced in example 10 above, introduce two different ideas in one list item. The one in example 10 first explains that the reader might already have a personal identity code and then

explains how to register more information or a municipality of residence. Although these two ideas are related, I would argue that instead of one item on a bulleted list, it would be more accessible to divide this section into multiple short paragraphs that discuss only one idea each.

In addition to having too much information per paragraph, the text also has long sentences that include multiple points, such as the ones in example 10. Long sentences are often hard to follow, especially for those with cognitive or learning disabilities. This is the criterion that the texts have the most issues with. Example 14 below introduces one such issue: it has multiple different points of information about both the reader and the residence permit in one sentence. Long sentences are quite common in Finnish, and in the analyzed texts the Finnish one did have more issues with this criterion than the English one. Example 15 below introduces one case where the English version is more accessible than the Finnish one due to shorter sentences.

- (14) You are a citizen of another country and you have a temporary (B) residence permit, which is valid and entitles you to stay for at least one year and, considering your circumstances as a whole, you intend to stay in Finland permanently.
- (15) FI: Jos tarvitset henkilötunnuksen verotukseen, voit esittää pyynnön myös verotoimistossa, mutta tällöin et voi saada kotikuntaa eikä esimerkiksi perhettäsi koskevia tietoja voida rekisteröidä (If you need a personal identity code for tax purposes, you can also request it at the tax office, but in that case you cannot be assigned a municipality of residence and it will not be possible to register information about your family [my translation]).
EN: If you need a personal identity code for tax purposes, you can also request it at the tax office. However, in that case, you cannot be assigned a municipality of residence and it will not be possible to register information about your family.

5 Discussion

Overall, the texts met the criterion to some extent and, therefore, there is reason to believe that a person with no or minor cognitive disabilities and adequate language skills should be able to find and comprehend the necessary information. However, this is not sufficient, as it does not take those with cognitive issues or poor language skills into account. As discussed earlier, the information presented on these pages is important for anyone moving to Finland and wanting to apply for a personal identity code, which is often necessary to be able to live and operate in society. Therefore, it is important to have this information available in an accessible form.

Even though the texts met the criteria to some extent, they are not completely accessible from the point of view of linguistic and cognitive accessibility. Both texts have only a few issues regarding grammar, vocabulary, and structure, but especially the amount of information and headings require changes to make the texts linguistically accessible. Additionally, shortening

some of the sentences and correcting the punctuation of lists, especially in the Finnish version, would increase accessibility.

The few grammar and spelling issues found in the texts are minor and do not cause any serious accessibility issues. Additionally, most of them are mistakes that are easy to make and hard to notice when proofreading. For instance, the spelling mistake in the Finnish text is most likely a typing error, as the word is missing its last letter. Additionally, as the misspelled form *kotikuntaa* is also a grammatically correct inflection, any spellchecking software used would most likely not have marked it as incorrect, making it more difficult to notice before publication. However, even though the mistakes are not severe, they should be corrected to ensure accessibility and understandability. Similarly, for the most part, the texts are structured logically and in a way that increases accessibility. The main issue with structure is the long list items, as discussed earlier regarding example 10.

Descriptive headings help the reader navigate the text and find the information they are looking for. Headings are especially important on websites with multiple pages full of information. In the analyzed texts, some of the headings are descriptive enough and match the contents of the paragraphs below them, but some need editing. More descriptive headings might help the readers find the correct page and the information they are looking for. My criteria do not include aspects of typography and layout, but it is worth noting that the Finnish text does not use heading styles consistently and the first three headings use the same font and size as the body of the text, making the headings harder to notice and thus hindering navigating the text and finding necessary information.

The amount of information is a difficult topic, as it is not easy to determine what the readers already know and what needs to be explained. While it is important to explain all uncommon terms and carefully explain the necessary steps to take to achieve the desired results, it is also important to not include too much information – an overload of information can have a negative effect on the accessibility of the text. Of course, the audience of the text determines the amount of information that is necessary. For instance, when writing for experts of a certain field, one can safely assume that the readers are familiar with the basic concepts of that field. However, when a text is aimed at large groups of people with different cognitive abilities and language skills, the question of what is necessary on the one hand and what is too much on the other hand, becomes central and simultaneously increasingly difficult to answer.

As discussed earlier, the analyzed texts tend to both present too much information per sentence or paragraph and include too little information on certain matters. The ambiguity of phrases such as *usually* and *in some cases* can leave the reader confused, and, in the worst case, they might not understand what they need to do. Accessibility could be increased by avoiding vague phrases and including a list of related pages with more information on the topics that were not covered in detail in the text at hand. In example 13 presented above, the reader is told to contact the DVV, but not how to do so. However, the contact information is listed later at the bottom of the page, in a section not included in my analysis. Similarly, links to other pages with more information might benefit the reader.

One of the research questions in this thesis was whether there is a difference in accessibility between the language versions. Additionally, I was interested in what might cause that possible difference. Overall, the accessibility of both texts is very similar – they both face the same issues and have the same strengths. For example, both use short, common words and both require work on the headings. This is most likely due to the translation process: the Finnish text is the original and the English one is then created based on the Finnish text. Usually, a translator follows the style and structure of the original version, and, based on the similarities, I assume this is the case for these texts as well. However, the Finnish text does have more long sentences full of information, while the English one is more accessible in this aspect. This is most likely because long sentences with multiple relative clauses are more common in Finnish than in English. Similarly, the Finnish version has more grammar issues due to the stricter rules for punctuating lists.

6 Conclusion

In this thesis, I have examined the increasingly important topic of accessibility with a focus on linguistic accessibility, a topic often left with little to no attention. The WCAG 2.1, one of the most common sources for accessibility issues, simply states that a text must be understandable, but the concept of understandability is not defined in more detail (W3C 2018b). Similarly, the Web Accessibility Directive states that the information and user interface should be understandable but does not explain what understandability means (2016/2102). To answer the question of what understandability means in the context of accessibility, I have created a set of criteria for evaluating the linguistic accessibility of a text. The criteria are heavily based on the WCAG 2.1 technique for making the text easier to read and the Selkomittari.

To test the criteria, I have analyzed a text found on the DVV's website. The DVV offers information to everyone living in Finland. Thus, the audience of their texts consist of individuals with varying cognitive abilities and language skills, making accessibility a central issue on their website. The analyzed text's main audience is immigrants, making the accessibility of the English translation extremely important, as they can have varying levels of English skills. I have analyzed both the Finnish and English texts and concluded that the texts exhibit very similar accessibility issues. They both have issues especially in headings and the amount of information presented on sentence and paragraph-level. The overall accessibility of the texts is mediocre, and the texts would benefit from editing to increase the accessibility, especially to those with cognitive disabilities or poor language skills. The criteria seemed usable as a tool to evaluate the accessibility of a text, at least in the data I used. The criteria were relevant, and I found all criteria applicable to the texts.

I have attempted to make the criteria as language independent as possible, but they are bound to have some bias towards English and Finnish, as those are the languages I am familiar with and the main sources for my criteria are focused on those two languages. Additionally, the criteria were tested with Finnish and English texts. Therefore, it would be important to perform further analysis and testing with different languages to see whether the criteria work with languages other than English and Finnish. In addition, in this thesis I have only analyzed one text – an informative text on a public sector website – and it would be crucial to perform further testing with different types of texts to confirm the usefulness of the criteria. Finally, testing the usability and accessibility of texts created utilizing the criteria on the main audience of accessible texts would be beneficial. Testing on those with cognitive disabilities could prove especially useful.

Although this thesis has focused on the linguistic accessibility of a text, the written content of the text does not exist in a vacuum. Therefore, when analyzing the accessibility of a text, one should always take the context of the text into account and consider the different aspects of accessibility together. This means that in addition to considering the linguistic accessibility of a text, one should pay attention to the typography, the visual layout of the text, and the technical solutions, to make the text accessible to as wide an audience as possible.

7 References

7.1 Research Data

DVV = Digital and Population Data Services Agency. n.d. "Ulkomaalaisen rekisteröinti väestötietojärjestelmään." Accessed May 1, 2020. <https://dvv.fi/ulkomaalaisen-rekisterointi-vaestotietojarjestelmaan>.

DVV = Digital and Population Data Services Agency. n.d. "Registration of a Foreigner in the Population Information System." Accessed May 1, 2020. <https://dvv.fi/en/foreigner-registration>.

7.2 List of References

Act on the Provision of Digital Services (396/2019) = Laki digitaalisten palvelujen tarjoamisesta (369/2016). <https://www.finlex.fi/fi/laki/alkup/2019/20190306>.

Crabb, Michael, Michael Heron, Rhianne Jones, Mike Armstrong, Hayley Reid, and Amy Wilson. 2019. "Developing Accessible Services: Understanding Current Knowledge and Areas for Future Support." *CHI '19: Proceeding of the 2019 CHI Conference on Human Factors in Computing Systems*: 1–12. <https://doi.org/10.1145/3290605.3300446>.

Directive (EU) 2016/2102 of the European Parliament and of the Council on the Accessibility of the Websites and Mobile Applications of Public Sector Bodies (Text with EEA relevance), 2016 O.J. (L 327), 1–15.

DVV = Digital and Population Data Services Agency. 2020a. "Accessibility of the Dvv.fi online service." Accessed April 22, 2020. <https://dvv.fi/en/accessibility>.

DVV = Digi- ja väestötietovirasto. 2020b. "Dvv.fi-verkkopalvelun saavutettavuus." Accessed April 15, 2020. <https://dvv.fi/saavutettavuus>.

European Standard (ETSI) EN 301549:2015 on the Accessibility Requirements Suitable for Public Procurement of ICT Products and Services in Europe, 2015. 1–134. https://ec.europa.eu/eip/ageing/standards/ict-and-communication/accessibility-and-design-all/en-3015492015_en.

Francik, Ellen. 1999. *Telecommunications Problems and Design Strategies for People with Cognitive Disabilities: Annotated Bibliography and Research Recommendations*. Oakland: World Institute on Disability.

Freyhoff, Geert, Gerhard Hess, Linda Kerr, Elizabeth Menzel, Bron Tronbacke, and Kathy Van Der Veken. 1998. "Make It Simple: European Guidelines for the Production of Easy-to-Read Information for People with Learning Disability: for Authors, Editors, Information Providers, Translators and Other Interested Persons." ILSMH European Association. <http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1271&context=gladnetcollect>.

Jenge, Constantin, Sven Hartrumpf, Hermann Helbig, and Rainer Osswald. 2006. "Automatic Control of Simple Language in Web Pages." *International Conference on*

- Computers for Handicapped Persons*: 207–214. Berlin: Springer.
https://doi.org/10.1007/11788713_31.
- Kato, Takashi, and Masahiro Hori. 2006. “Beyond Perceivability’: Critical Requirements for Universal Design of Information.” *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*: 287–288.
<https://doi.org/10.1145/1168987.1169058>.
- Kuhn, Tobias. 2012. “The Understandability of OWL Statements in Controlled English.” *Semantic Web*, 4 (1): 101–115. <https://doi.org/10.3233/SW-2012-0063>.
- Leivo, Tiina. 2019. ”Vieraskynä: Poimi selkokielestä parhaat palat verkkokirjoittamiseksi.” *Saavutettava.fi*. Accessed May 15, 2020.
<https://saavutettava.fi/2019/08/31/vieraskyna-poimi-selkokielesta-parhaat-palat-verkkokirjoittamiseksi/>.
- Leskelä, Leela. 2019. *Selkokieli: saavutettavan kielen opas*. Helsinki: Kehitysvammaliitto ry.
- Lewis, Clayton. 2008. “Cognitive and Learning Impairments.” In *Web Accessibility: A Foundation for Research*, edited by Simon Harper and Yeliz Yesilada, 15–23. London: Springer-Verlag.
- Papunet. 2019. ”Kognitiiviset ja kielelliset vaikeudet.” Accessed April 15, 2020.
<https://papunet.net/saavutettavuus/kognitiiviset-ja-kielelliset-vaikeudet>.
- Rello, Luz, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013. “Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia.” *INTERACT 4*: 203–219.
https://doi.org/10.1007/978-3-642-40498-6_15.
- Small, Jeon, Pamela Schallau, Karen Brown, and Richard Appleyard. 2005. “Web Accessibility for People with Cognitive Disabilities.” *CHI '05 Extended Abstracts on Human Factors in Computing Systems*: 1793–1796.
<https://doi.org/10.1145/1056808.1057024>.
- Saavutettavasti.fi. n.d. ”Pikaohje selkeään yleiskieleen.” Accessed May 15, 2020.
<https://www.saavutettavasti.fi/yleiset-ohjeet/selkea-kieli/>.
- Selkokeskus. 2018. ”Selkeästi kaikille: Selkomittari.” Selkokielen mittari.
https://selkokeskus.fi/wp-content/uploads/2018/10/SELKOMITTARI_2018_11.10.18.pdf.
- Statistics Finland. n.d. ”Vieraskieliset.” Accessed February 11, 2020.
<https://www.stat.fi/tup/maahanmuutto/maahanmuuttajat-vaestossa/vieraskieliset.html>.
- Valtiovarainministeriö. n.d. ”Saavutettavuus.” Accessed February 11, 2020.
<https://vm.fi/saavutettavuusdirektiivi>.
- W3C. 2018a. “Web Content Accessibility Guidelines (WCAG) 2.1.”
<https://www.w3.org/TR/WCAG21/>.

W3C. 2018b. “Web Content Accessibility Guidelines (WCAG) Overview.” Accessed April 13, 2020. <https://www.w3.org/WAI/standards-guidelines/wcag/>.

W3C. 2019. “Techniques for WCAG 2.1.” Accessed April 8, 2020. <https://www.w3.org/WAI/WCAG21/Techniques/>.

WebAIM. 2018. “Cognitive.” Accessed April 15, 2020. <https://webaim.org/articles/cognitive/>.