Applying Accessibility-Oriented Controlled Language (CL) Rules to Improve Appropriateness of Text Alternatives for Images: an Exploratory Study

Silvia Rodríguez Vázquez1,2, Pierrette Bouillon1, Anton Bolfing3
1Multilingual Information Processing Department (TIM/ISSCO)
Faculty of Translation and Interpreting, University of Geneva, Switzerland
40, Bd. du Pont d’Arve, CH-1211 Geneva 4

2Cod.eX Research Group
Faculty of Translation and Interpreting, University of Salamanca, Spain
6-16, Francisco Vitoria, 37008 Salamanca

3Access for All Foundation, Switzerland
Dörlistrasse 10, CH-8057 Zürich
E-mail: Silvia.Rodriguez@unige.ch, Pierre.Bouillon@unige.ch, anton.bolfing@access-for-all

Abstract
At present, inappropriate text alternatives for images in the Web continue to pose web accessibility barriers for people with special needs. Although research efforts have been devoted to define how to write text equivalents for visual content in websites, existing guidelines often lack direct linguistic-oriented recommendations. Similarly, most web accessibility evaluation tools just provide users with an automated functionality to check the presence of text alternatives within the <img> element, rather than a platform to verify their content. This paper presents an overview of the findings from an exploratory study carried out to investigate if the appropriateness level of text alternatives for images in French can be improved when applying controlled language (CL) rules. Results gathered suggest that using accessibility-oriented alt style rules can have a significant impact on text alternatives’ appropriateness. Although more data would be needed to draw further conclusions about our proposal, this preliminary study already offers an interest insight into the potential use of CL checkers such as Acrolinx for language-based web accessibility evaluation.

Keywords: image text alternatives, appropriateness, accessibility-oriented CL rules

1. Introduction and Background
Over the last decades, the Word Wide Web has become a complex multimodal platform, transforming the access to information experience, traditionally channelled through mainly text-only websites, into an enriched human-computer interaction practice. Content is now often rendered through multiple sensory modalities (auditory, tactile or visual) that sometimes challenge the message transfer and, thus, its accessibility, for people with special needs. Hence, the Web Accessibility Initiative (WAI) from the World Wide Web Consortium (W3C) recommends through its Web Content Accessibility Guidelines (WCAG 2.0), recently approved as an ISO/IEC International Standard (ISO/IEC 40500: 2012), to provide text alternatives for any non-text content, so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language (Guideline 1.1, W3C, 2014).

For the present research, the focus is specifically on images (including, inter alia, graphics, charts, pictures, icons, buttons and interactive areas), for which text alternatives are provided through an alt attribute within the HTML <img> element, and should serve their equivalent purpose. While being a widely extended recommendation, whose importance is acknowledged by most web professionals’ communities, studies have shown that compliance is still lower than expected (Access for all, 2011; Hanson and Richards, 2013). The latter might be derived from lack of awareness among content authors, as well as the existence of multiple sources of both official and non-official sets of guidelines on how to write text alternatives, which often remain too abstract and lack linguistic-oriented guidance for each specific language. In Rodríguez Vázquez and Torres del Rey (2012), an innovative approach was proposed to evaluate the accessibility degree of text alternatives for images by means of a state-of-the-art controlled language (CL) checker, which could also be potentially implemented as an automated tool for authoring support.

In the light of the aforementioned proposal, this paper aims at estimating the impact of applying human-oriented controlled-language (HOCL) rules to text alternatives of non-accessible images in order to obtain appropriate text equivalents for images in French.

2. Related Work
Significant research efforts have been devoted to study existing accessibility evaluation methods (AEM) and putting forward new ones depending on the context to be considered (Brajnik, 2008). Five main categories have been detailed description is needed. However, it is not supported by many of the major browsers and its implementation in HTML 5 is still under review (W3C, 2013).
defined in the literature: i) inspection methods: an auditor checks a webpage for its accessibility; ii) automated testing: an auditor uses an automated accessibility tool to check accessibility conformance of a webpage; iii) screening techniques: an auditor uses a webpage simulating some of the conditions that are typical for people with disabilities (for instance, artificially reducing some sensory or motor capabilities); iv) subjective assessments: an auditor asks a group of users to use a website autonomously and send back their opinions; and v) user testing: real users are asked to navigate through a webpage or site while being observed by accessibility auditors (Brajnik et al., 2012). The inspection method, and most concretely WCAG 2.0 conformance review, has proven to be the most popular one, where auditors follow a list of prescribed criteria to check the webpage(s) against (ibid). According to WCAG 2.0, those criteria can be machine testable or reliably human testable, with evaluators needing to reach at least an 80% agreement on the audit results in the case of the latter. Evidence suggest, however, that such level of agreement is rarely attained, and that more concrete guidance on how to interpret WCAG 2.0 criteria could lead to a higher degree of reliability (ibid). In this sense, when narrowing accessibility conformance review to the particular case of images, a study of the literature shows that considerable research attempts have been already made to offer that kind of guidance.

On one hand, guidelines for image description have been proposed both by scholars and official bodies to enhance human verification of text alternatives. From both sides, the general tendency has been to offer general recommendations on when to use the alt attribute – depending on the nature of the image (informative vs. decorative) –, its context within the webpage, and its purpose (Korpela, 2010; W3C, 2012). Guidance on linguistic appropriateness has been generally limited to subjective techniques; for instance, care about spelling and grammar; avoid special characters; use normal prose; and make text simple, succinct and accurate (Nyong, 2009; WebAIM, 2010). The Technical Specification ISO/IEC TS 20071-11:2012 offers a more detailed question-guided procedure for providing informative text alternatives, but language-oriented hints are barely referred to. In Craven (2006), Tercedor Sánchez (2009a) and Tercedor Sánchez et al. (2009b), attention has been drawn to the need of avoiding redundant expressions such as “image of”, minimizing orality markers such as “in the image we can see”, or using perceptual content with care (for instance, colours or forms); however, no automatic checking solution has been offered so far.

On the other hand, studies on machine checking of image text equivalents have mainly focused on image OCR or text pattern recognition techniques (e.g., dictionary-based word search, file type abbreviations, HTML code, number of characters) to automatically identify what should not be present in appropriate text alternatives (Hu and Bagga, 2003, Olsen et al., 2010). The size of the image has also been widely used as a reference to detect non-accessible images (Bigham et al., 2006), classifying as informative (and thus in need of alt) those bigger than 10 x 10 pixels, and then automatically giving them text alternatives based on web content analysis, OCR and human labelling (ibid). Besides evaluation and repair solutions, guidance tools, such as the one proposed by Tang (2012) help authors to assess the importance of the information communicated by an image through a customizable set of questions. Nonetheless, to the best of our knowledge, no in-depth linguistic analysis in terms of style, syntax and terminology adequateness has been carried out until present.

Our research has attempted to bridge the gap between human and machine verification of text alternatives through an analysis of accessible linguistic patterns for image text equivalents. This served as ground basis for the development of an initial set of accessibility oriented CL rules and their evaluation. Before presenting our study findings, we illustrate in section 3 how to render images accessible by providing an appropriate text alternative.

3. Appropriateness of Text Alternatives for Images: an Overview

Writing appropriate text alternatives to obtain accessible images in the Web has been sometimes defined as an art in itself (Tang, 2012; Bigham et al., 2006). Their quality level can be influenced not only by the accessibility and language skills of the person who adds the text alternative to the image, but also by the context in which it appears within a given webpage. The value of the alt attribute of an <img> element should be considered as an image textual replacement, rather than as image description (Korpela, 2010). Consequently, it is crucial to understand the meaning of the image in context and interpret it correctly, as we illustrate in the following example.

Figure 1: Example of inappropriate text alternative for an image with an embedded link.

The text alternative for the image marked with a red arrow in Figure 1 is “Facebook Logo”. Although it might seem appropriate at first glance, it is worth noting that the image has an embedded link to the Facebook page of the Faculty
of Translation and Interpreting (FTI), University of Geneva (UNIGE). In this case, the Facebook logo could be considered either as a decorative image for which we could introduce an empty alt (alt=""),2 given that there is a “Find us on Facebook” textual link underneath, or as a functional image that would need a non-empty alt such as “Facebook page of FTI-UNIGE”.

In the study presented in the following sections, a similar judgement exercise has been carried out by sighted and blind accessibility experts in order to test our hypothesis, which states that applying CL rules to inappropriate text equivalents could contribute to a higher degree of image accessibility.

4. Experimental Plan

In this section, we describe how we selected the data used in our study, which CL rules were defined and applied, and how was the evaluation performed.

4.1 Material Selection

The identification of accessible and non-accessible linguistic patterns frequently used in text alternatives was obtained through a comprehensive review of existing guidelines found in the literature on how to create accessible text alternatives and a detailed analysis of a corpus of websites, which served as our main training data set.

4.1.1. Training Web Corpus

The web document collection initially consisted of 100 websites chosen for a web accessibility evaluation study, commissioned by the Swiss authorities and conducted by the Swiss Access for all Foundation (Access for all, 2011), for which two AEM were implemented: conformance review and user testing. The 100 websites were mostly Swiss, including governmental and cantonal sites, as well as those of the ten biggest cities of the country. Others were related to public transport, educational and media services. For the purposes of this study, the web document collection size was simplified as follows: first, websites in languages other than French were dismissed, as well as two global websites, namely Facebook and Wikipedia. The Wayback Machine (Internet Archive, 2014) was used to retrieve all pages within a distance of one link from the homepage (also included). From the reduced sample (57 websites), five more websites were eliminated due to the impossibility to locate them in the Internet Archive.3 The final corpus consisted of 52 websites, accounting for 1,938 webpages.

4.1.2. Training Image Set

The experimental web corpus files (in HTML format) were converted into XLIFF (version 1.2). A set of regular expressions were then applied to the XLIFF file set in order to extract the following information: <img> elements with an alt attribute, <img> elements with no alt attribute, and <img> elements with an empty alt attribute. From the resulting image set, we filtered 1) images with the same source and the same attribute value used in the same page context, and 2) images automatically added by the Internet Archive during the websites retrieval. Since the study of the null alt appropriateness, as well as the existence or non-existence of an alt attribute, were beyond the scope of this research, we limited our training image set to a total of 7,945 unique text alternatives (see Table 1).

<table>
<thead>
<tr>
<th>Initial set</th>
<th>No alt</th>
<th>Empty alt</th>
<th>Non-empty alt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>50,633</td>
<td>12,238</td>
<td>29,154</td>
</tr>
<tr>
<td>Filtered set</td>
<td>12,778</td>
<td>1,267</td>
<td>7,945</td>
</tr>
</tbody>
</table>

Table 1: Total number of images and alt variants before and after the training image set filtering process.

The final training image set was explored as a source of good and bad examples of text alternatives, taking as a baseline the results from the web accessibility study lead by the Access for All Foundation (2011). Evidence found on both subsets was compared and enlarged with information gathered during the extensive literature review on appropriateness of text alternatives. Following an inductive approach, we used AntConc, a freeware concordancer software programme, to perform a semi-automatic analysis on the training image set with a view to obtaining representative examples of appropriate and inappropriate text alternatives that shared similar linguistic patterns. This analysis helped us to identify a preliminary set of language-based rules to be applied when assessing text alternatives for their appropriateness.

4.2 Definition of Accessibility-oriented Style Rules for Text Alternatives Verification

Our preliminary set of rules covered general punctuation, spelling and grammar verification, and a special accessibility-oriented alt style validation package. As in previous studies (Hu and Bagga, 2003, Olsen et al., 2010), a complementary subset of rules was also defined under the style rule set with a view to detecting undescriptive elements within image text alternatives. Table 2 summarizes the accessibility-oriented alt style rule package inferred from the literature and the web corpus analysis.

<table>
<thead>
<tr>
<th>Id</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Avoid_undescriptive_words</td>
</tr>
<tr>
<td>2</td>
<td>Avoid_file_name_and_format</td>
</tr>
<tr>
<td>3</td>
<td>Avoid_non_alphabetical_characters_concatenated</td>
</tr>
</tbody>
</table>

2 If the alt text is empty or set to null (i.e. alt=""), it means for assistive technology that the image can be safely ignored. Having no alt attribute, considered as a non-accessible technique, could lead users to think they are missing important content in the page. Also, when there is no alt, some screen readers read the file name of the image, which can be confusing to listen to (WebAIM, 2010).

3 The web accessibility study carried out by the Access for all Foundation took place between March and June 2011. For the last five websites dismissed, no complete version had been stored in the Internet Archive during that four months-period time.
4.2.1. Selected Rules for Evaluation

For the purposes of this exploratory study, we have formalised 12 of the 30 language-based guidelines into controlled language (CL) rules using Acrolinx, a state-of-the-art controlled language programme which includes both a rule developing and testing platform. Acrolinx rules are written following an error description formalism that allows the “specification of error phenomena using regular expressions over linguistic objects, represented as feature structures. These feature structures denote the linguistic annotations provided by different underlying NLP components, such as POS and morphological taggers” (Bredenkamp et al., 2000). Table 3 shows a brief explanation of the rules selected for the study, as well as examples of inappropriate text alternatives flagged by the tool.

Table 2: Preliminary accessibility-oriented alt style rule package.

<table>
<thead>
<tr>
<th>Id</th>
<th>Explanation &amp; example of alt flagged</th>
</tr>
</thead>
</table>
| 1  | Isolated undescriptive words or word sets should be avoided. Some examples include: title, web, page, online, click, spacer, read more, image, home.  
Alt flagged: alt="accueil" |
| 2  | Image file name and format should not be repeated in the alt text.  
Alt flagged: alt="bookcover12-pag.png" |
| 3  | Strings of non-alphabetical characters (such as numbers or punctuation signs) combined with letters that are not presented in a meaningful way should be avoided.  
Alt flagged: alt="formcont5_ssr" |
| 17 | Images that provide structure to the web document should not be described in the alt. If so, they should reflect the macrostructure of the web content (e.g., alt="‘New blog entry”).  
Alt flagged: alt="ligne horizontale” |
| 18 | Text alternatives of functional images which allow the user to perform specific actions should describe them using a verb in infinitive or imperative form (e.g. alt="Print this page”).  
Alt flagged: alt="Version pour impression” |
| 19 | When a navigational symbol is included, the sense of the navigation should be specified, as well as the items through which the user is browsing (e.g., alt="Next article")  
Alt flagged: alt="flèche” |
| 20 | If social media logos have an embedded link, a description of the related action should be provided (e.g., alt= “Share this page on Facebook”).  
Alt flagged: alt="page Facebook” |
| 21 | Isolated abbreviations and acronyms should not be included as the only element in the text alternative.  
Alt flagged: alt="mp3” |
| 24 | Text alternatives that are too short to be informative and too long to be retrieved by assistive technology should be avoided. For this rule, we have set at a maximum of 100 characters.  
Alt flagged: alt="‘Link to FTI website” |
| 26 | Redundant expressions containing words like logo, insignia, or emblem should be used with care, especially when the image has an embedded link (e.g., alt="Home page of Artionet website”).  
Alt flagged: alt="Logo Artionet” |
| 27 | Indicating that an image has an embedded link with an expression such as “link to” is not necessary, since assistive technology already gives this information to the user.  
Alt flagged: alt="Link to FTI website” |
| 28 | Succinctness implies the omission of redundant expressions such as “image of”, “picture of”, “icon”, “photo illustrating…”  
Alt flagged: alt="Une image de l’accident” |

Table 3: Rule Id and explanation of the rules selected for the study with examples of alt values flagged in French.

4.2.2. Evaluation Image Set

A web corpus of 12 websites was built to assess the impact of the 12 rules selected from our preliminary accessibility-oriented alt style rule package. The new web
4.3 Exploratory Evaluation study

The goal of this evaluation study was to investigate if text alternatives’ appropriateness in terms of accessibility improves or not when applying the selected rule set. The method chosen was a human comparative assessment of alt values by both sighted and blind users, as per ISO/IEC TS 20071-11:2012. The study took place in mid-March 2014.

4.3.1. Study Design and Environment

From the evaluation image set, accounting for 1,731 unique non empty alt values, a sample of 110 text alternatives flagged after running the Batch Checker were selected for the study. Changes proposed by Acrolinx were manually applied, thus creating 110 pairs of original and edited text alternatives (220 alt values in total, of less than 100 characters each).

SurveyMonkey was the evaluation environment that best served our study purposes. The survey was written in French and consisted of a first section covering demographic questions, a second section where the 110 images and their two text alternative proposals were presented, and a third section including a brief post-task questionnaire.

Participants were asked to assess text alternatives’ appropriateness in context, based on a five-point grading system (first clearly better, first slightly better, about equal, second slightly better, second clearly better). To this end, each text alternative proposal had an embedded link to the page in which their corresponding image appeared. At the top of each page, we introduced a “Skip to image for evaluation” link that jumped the user down to an anchor at the beginning of the image that needed to be assessed. This technique was used to facilitate a quick and easy location of the image, especially for blind users. Contrary to this participant group, sighted users were visually presented with the image as well (see Figure 2). Original and edited text alternative proposals, together with their respective webpages, were randomly presented as “first text alternative” and “second text alternative” throughout the survey in order to avoid research bias. An optional comments box was also introduced after each assessment question for participants to include any potential remarks about the choice made or a text alternative proposal of their own, if deemed necessary.

4.3.2. Participants

Four users participated in the evaluation: two were sighted users and two were blind users (N=4, aged between 28 and 39, \( \bar{x} = 34.7, sd = 4.99 \), all males). Both non-sighted participants were regular users of the same screen readers: JAWS and NVDA. All participants were accessibility experts who have been working in web accessibility projects for more than 4 years (\( \bar{x} = 5.5, sd = 3.1 \)) and self-rated their knowledge on appropriateness of text alternatives for images in the Web as advanced. The four participants were recruited by the Access for All Foundation and were of Swiss origin, fluent both in German and French.

5. Data Analysis and Results

At this stage of our research, Acrolinx rules allow the scanning of webpages for error candidates, but no style suggestions are shown to the user yet. Consequently, as mentioned in section 3, the appropriateness level achieved in the new text alternative (rewritten without the error(s) flagged by Acrolinx) is subject to the editor’s accessibility background. Therefore, for the purposes of this exploratory study, when participants suggested a third different text alternative proposal of their own in which the rule that was being tested was not violated, we inferred that the rule had a positive impact.

<table>
<thead>
<tr>
<th></th>
<th>BU1</th>
<th>BU2</th>
<th>SU1</th>
<th>SU2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw better</td>
<td>6</td>
<td>21</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>%</td>
<td>5%</td>
<td>19%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>Edited better</td>
<td>92</td>
<td>82</td>
<td>85</td>
<td>97</td>
</tr>
<tr>
<td>%</td>
<td>84%</td>
<td>75%</td>
<td>77%</td>
<td>88%</td>
</tr>
<tr>
<td>No impact</td>
<td>12</td>
<td>7</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>%</td>
<td>11%</td>
<td>6%</td>
<td>14%</td>
<td>1%</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Figure 2: Evaluation environment.
We hypothesize that tent is meaningful. They results reveal significant impact on text alternative, to the reduced, among other reasons.d slightly better” and “second clearly appeared to agree that the more complete the is set at 75 as a source of disagreement. Although the optimum length Our in future studies.

Interestingly, this positive tendency towards choosing the selected rule set for the exploratory study has results was significant the difference between “raw better” and “edited better” under the “edited better” category, while “No alternative (p<0.001).

Table 5: Evaluation results per rule.

<table>
<thead>
<tr>
<th>Rule</th>
<th>alt verified</th>
<th>No impact</th>
<th>%</th>
<th>Raw better</th>
<th>%</th>
<th>Edited better</th>
<th>%</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoid_alt_too_long</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>40%</td>
<td>3</td>
<td>60%</td>
<td>0.82</td>
</tr>
<tr>
<td>Avoid_description_of_structural images</td>
<td>10</td>
<td>-</td>
<td>1</td>
<td>10%</td>
<td>9</td>
<td>90%</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>Avoid_file_name_and_format</td>
<td>5</td>
<td>-</td>
<td>0</td>
<td>0%</td>
<td>5</td>
<td>100%</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Avoid_non_alphabetical_characters</td>
<td>10</td>
<td>-</td>
<td>0</td>
<td>0%</td>
<td>10</td>
<td>100%</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Avoid_redundant_expression_image</td>
<td>20</td>
<td>2</td>
<td>20%</td>
<td>10%</td>
<td>16</td>
<td>80%</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Avoid_redundant_expression_link</td>
<td>10</td>
<td>-</td>
<td>1</td>
<td>10%</td>
<td>9</td>
<td>90%</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Avoid_redundant_expression_logo</td>
<td>10</td>
<td>-</td>
<td>1</td>
<td>10%</td>
<td>9</td>
<td>90%</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Avoid_stand-alone</td>
<td>5</td>
<td>1</td>
<td>20%</td>
<td>1</td>
<td>20%</td>
<td>3</td>
<td>60%</td>
<td>0.64</td>
</tr>
<tr>
<td>abbreviations_or_acronyms</td>
<td>10</td>
<td>1</td>
<td>10%</td>
<td>2</td>
<td>20%</td>
<td>7</td>
<td>70%</td>
<td>0.18</td>
</tr>
<tr>
<td>Avoid_undescriptive_words</td>
<td>10</td>
<td>1</td>
<td>10%</td>
<td>2</td>
<td>20%</td>
<td>7</td>
<td>70%</td>
<td></td>
</tr>
<tr>
<td>Describe_social_media_actions</td>
<td>10</td>
<td>1</td>
<td>10%</td>
<td>1</td>
<td>10%</td>
<td>8</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>Expand_description_of_navigation</td>
<td>5</td>
<td>1</td>
<td>20%</td>
<td>0</td>
<td>0%</td>
<td>4</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>items_function</td>
<td>10</td>
<td>2</td>
<td>20%</td>
<td>2</td>
<td>20%</td>
<td>6</td>
<td>60%</td>
<td></td>
</tr>
</tbody>
</table>

We applied the same coding principle if they had included a comment indicating that there was still room for improvement as regards the text alternative they chose as more appropriate without contravening the rule being evaluated.

Table 4 offers an overview of the evaluation results per participants: blind users (BU) and sighted users (SU). We have grouped the values “first clearly better” and “first slightly better” to obtain a general “raw better” category, and merged “second slightly better” and “second clearly better” under the “edited better” category, while “No impact” observations refer to “about equal” responses. Although evidence found indicates that participants perceived the edited version as more appropriate in more than 75% of the cases, we used McNemar’s test to assess if the difference between “raw better” and “edited better” results was significant. As shown in table 4, results reveal that the selected rule set for the exploratory study had a statistically significant impact on text alternative appropriateness (p<0.001).

Table 6: Interrater agreement within and between blind and sighted user groups.

<table>
<thead>
<tr>
<th></th>
<th>Raw better</th>
<th>Edited better</th>
<th>Total</th>
<th>Int. agr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blind users</td>
<td>2 (2%)</td>
<td>85 (77%)</td>
<td>87</td>
<td>0.7909</td>
</tr>
<tr>
<td>Sighted users</td>
<td>5 (5%)</td>
<td>93 (85%)</td>
<td>98</td>
<td>0.8909</td>
</tr>
<tr>
<td>Two groups</td>
<td>1 (1%)</td>
<td>79 (72%)</td>
<td>80</td>
<td>0.7272</td>
</tr>
</tbody>
</table>

Finally, the interrater reliability values observed between groups and users from the same group type indicate a high degree of agreement among participants (see Table 6). To compute this reliability test, we included “about equal” results under the “edited better” category. This decision is based on a closer examination of each “about equal”
response, during which we found two main different patterns: either no comment was left for the evaluator (from what we assumed that the rule had not decreased the appropriateness level of the text alternative), or new text alternative suggestions were put forward without violating the selected rules.

6. Discussion and Conclusions
As part of the exploratory study, participants were requested to fill in a short post-task questionnaire. When asked about the criteria on which they had based their evaluation, blind users mentioned their own personal experience as the main foundation for their choices, although one of them reported to have taken into account the context in which each image appeared as well. While this was also the basis for one of the two sighted users’ decisions, this group considered their professional experience as a sufficient source of knowledge on text alternative appropriateness. None of the participants indicated to have looked at language correctness in terms of spelling and grammar, which lead us to believe that web accessibility experts give more importance to text alternatives appropriateness from a functional perspective, rather than from a strictly linguistic point of view. Nonetheless, in spite of having selected the image context as one of the criteria on which to ground their evaluation, 75% of the participants (N=3 out of 4) estimated that they had checked the page in which the image appeared only in 25% of the cases before making their judgement. This could explain some responses where there was a clear disagreement between users. Another factor that might have influenced participants’ decisions during the evaluation is their previous knowledge about the web accessibility level of the websites chosen for the study. While two participants reported that they had not visited the selected webpages until the time of the evaluation, one blind user said he knew 25% of the pages and one sighted user indicated that he had browsed 50% of them prior to the study.

All in all, the present paper illustrates how the application of CL rules specifically designed for image text alternatives verification can lead to an improvement of their appropriateness level and thus contribute to a higher web accessibility degree. The study also shed light on the potential effectiveness of using Acrolinx as a validation tool for language-related accessibility barriers, whose rules proved to have a positive impact on text alternatives adequateness, regardless of the type of user performing the evaluation.

7. Limitations and Future Work
Due to time and mobility constraints, a subjective assessment was chosen as the AEM for this exploratory study. Although it was designed following recommendations included in ISO/IEC TS 20071-11:2012, blind participants experienced certain difficulties while checking some images in context, thus needing to devote more time to the task. We hope to follow a user testing approach in future studies, as well as to increase the number of participants and rules being tested. Despite these limitations, evidence found reveals that controlled language could be used as an interesting resource for text alternatives verification support. Further research would be needed, however, to reinforce this statement. As a first step, we plan to carry out a second exploratory study where users will be given the set of rules and the text alternatives flagged by Acrolinx, and asked to suggest their own improvement proposal. Results of both exploratory studies will be then compared with the aim of strengthening the rules’ precision and adding pertinent suggestions. Also, within this research project framework, we will complement each rule with help resources for the user, explaining the reason why a certain error has been flagged. This would make of Acrolinx a potential evaluation, guidance and repair tool for both experts and non-experts in image accessibility.

8. Acknowledgements
We would like to extend our gratitude to all members of the Access for All Foundation for their ongoing support in providing us with access to their studies material and human resources, as well as to the four participants who took part in the evaluation. Another special thank you should be given to Emilio Rodríguez Vázquez de Aldana, who donated his time and experience to the preparation of this study.

9. References


