Int. J. Human-Computer Studies 86 (2016) 1–10

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Measuring user rated language quality: Development and validation of the user interface Language Quality Survey (LQS)☆

Javier A. Bargas-Avila*, Florian Brühlmann

Google/YouTube User Experience Research, Brandschenkestrasse 110, 8002 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form
24 August 2015
Accepted 28 August 2015
Communicated by E. Motta
Available online 10 September 2015

Keywords:
User interface
Language
Text
Translation
Internationalization
Localization
L10n
I18n

ABSTRACT

Written text plays a special role in user interfaces. Key information in interaction elements and content are mostly conveyed through text. The global context, where software has to run in multiple geographical and cultural regions, requires software developers to translate their interfaces into many different languages. This translation process is prone to errors – therefore the question of how language quality can be measured is important. This paper presents the development of a questionnaire to measure user interface language quality (LQS). After a first validation of the instrument with 843 participants, a final set of 10 items remained, which was tested again ($N = 690$). The survey showed a high internal consistency (Cronbach's α) of .82, acceptable discriminatory power coefficients (.34–.47), as well as a moderate average homogeneity of .36. The LQS also showed moderate correlation to UMUX, an established usability metric (convergent validity), and it successfully distinguished high and low language quality (discriminative validity). The application to three different products (YouTube, Google Analytics, Google AdWords) revealed similar key statistics, providing evidence that this survey is product-independent. Meanwhile, the survey has been translated and applied to more than 60 languages.

1. Introduction

Key information in interaction elements and content within user interfaces are mostly conveyed through text. Graphical user interfaces have evolved substantially when compared to text-based user interfaces, but they still rely heavily on language to communicate with users. Therefore language plays a crucial role in Human–Computer Interaction. Single words can make the difference between failure or success.

The importance of language within a user interface (UI) becomes clear when text elements are removed. Fig. 1 shows three screenshots of the video-sharing site YouTube. The first (a) shows the original, the second (b) shows the website, but with all text elements removed, while on the third (c) all graphic elements are deleted. The illustration shows how the textless version is stripped of the most useful information: it is almost impossible to predict and choose which video to watch and navigation becomes impossible.

Text used in interfaces is highly dependent on cultural and regional aspects. For example, instructional text such as a tutorial could be worded informally for the US, but such an informal wording might be very inappropriate in other cultures. Hence it is important to consider not only mere correctness of translation of text but also style and tone aspects in the specific cultural context. Beside translation of text, interface elements such as icons and pictures should also be considered in the process of localization. Worldwide, there are about 200 languages that are spoken by at least 3 million people (Lewis et al., 2013). Companies with worldwide reach need to localize their products to make sure they can be used by everyone. For instance, Google search currently supports more than 140, Facebook more than 60, and YouTube more than 60 languages.

Websites and user interfaces are generally developed in one source language and translated afterwards by professional linguists. The process of translation is prone to errors and might introduce a number of problems that are not present in the source user interface. For example, the word auto can be translated to French as *automatique* (automatic) or *automobile* (car), which obviously has a completely different meaning. Another problem arises from words that behave as a verb when placed in a button or as a noun if part of a label (Leiva and Alabau, 2014). For example, the word *access* can stand for “you have access” (as a label) or “you can request access” (as a button). This *word sense disambiguation problem* (Muntés Mulero and Paladini Adell, 2012) arises often in UI translations. Further, possible pitfalls are gender, prepositions without context (Muntés Mulero and Paladini Adell, 2012) or other characteristics of the source text that might influence the translation process (Dilts,

☆This paper has been recommended for acceptance by E. Motta.

* Corresponding author.

E-mail addresses: javier.bargas@me.com (J.A. Bargas-Avila), florian.bruehlmann@gmail.com (F. Brühlmann).

http://dx.doi.org/10.1016/j.ijhcs.2015.08.010
1071-5819/© 2015 Elsevier Ltd. All rights reserved.

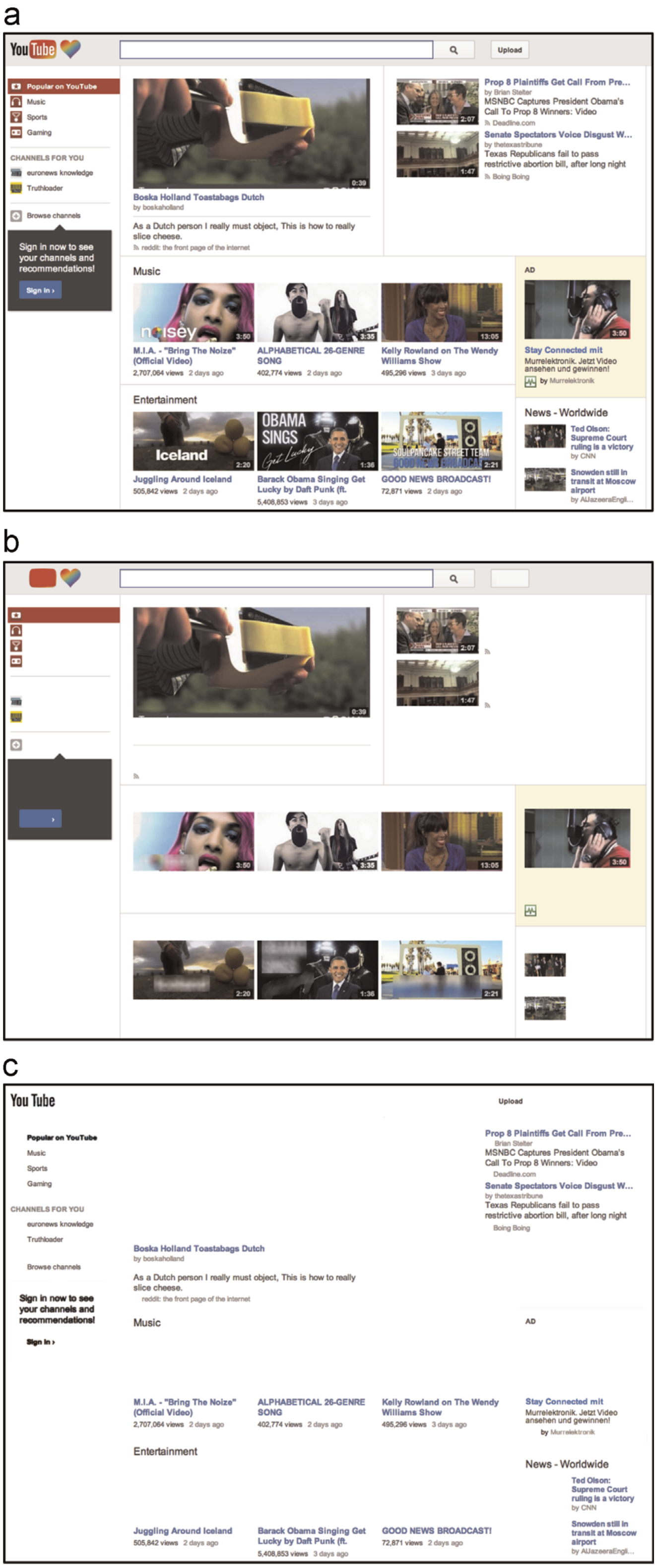
a
YouTube
Upload
Popular on YouTube
Music
Sports
Gaming
CHANNELS FOR YOU
euronews knowledge
Truthloader
Browse channels
Sign in now to see your channels and recommendations!
Sign in ›
Boska Holland Toastabags Dutch
by boskaholland
As a Dutch person I really must object, This is how to really slice cheese.
reddit: the front page of the internet
Prop 8 Plaintiffs Get Call From Pre...
by Brian Stelter
MSNBC Captures President Obama's Call To Prop 8 Winners: Video
Deadline.com
Senate Spectators Voice Disgust W...
by thetexastribune
Texas Republicans fail to pass restrictive abortion bill, after long night
Boing Boing
Music
M.I.A. - "Bring The Noize" (Official Video)
2,707,064 views 2 days ago
ALPHABETICAL 26-GENRE SONG
402,774 views 2 days ago
Kelly Rowland on The Wendy Williams Show
495,296 views 3 days ago
AD
Stay Connected mit Murrelektronik. Jetzt Video ansehen und gewinnen!
by Murrelektronik
Entertainment
Juggling Around Iceland
505,842 views 2 days ago
Barack Obama Singing Get Lucky by Daft Punk (ft.
5,408,853 views 3 days ago
GOOD NEWS BROADCAST!
72,871 views 2 days ago
News - Worldwide
Ted Olson: Supreme Court ruling is a victory
by CNN
Snowden still in transit at Moscow airport
by AlJazeeraEngli...
b
c

Fig. 1. Example of how UIs look when text or graphics are removed.

2001). Such mistranslations might not only negatively affect trustworthiness and brand perception, but also the acceptance of the website and its perceived usefulness (Sun, 2001).

As companies scale their products to multiple languages, the need for quality metrics increases: How can product managers learn more about the quality of a translation in an interface when they might not even speak the language themselves? In this paper, a method is presented that delivers metrics about language quality by asking users to rate the language of the user interfaces in a survey.

2. Theoretical background

Schriver (1989) distinguishes three different classes of text quality evaluation: (1) text-focused, (2) expert-judgement-focused, and (3) reader-focused. These three classes express different levels on how explicit the feedback from the target audience is: "… text-focused methods (…) never use direct reader response; experts – through their experience – provide surrogate reader feedback; and reader-focused methods make explicit use of audience response." Schriver (1989, p. 241).

2.1. Text-focused evaluation

Text-focused methods operate by having a person or a computer examine a text and assess text quality by applying rules and guidelines that define what good text quality is. These methods include readability formulae (e.g., Fry, 1968; Kincaid et al., 1975) and user models (e.g., Blackmon et al., 2005; Chi et al., 2001) which can be applied by software that would allow automation of certain aspects of evaluation. Such automized analysis is inexpensive and can spot certain obvious classes of error such as misspellings or provide general statistics about number of complex or passive sentences that could reduce readability. But in general, these provide little information about the overall performance of the text (whole-text level) or whether the text meets the needs of readers.

2.2. Expert-judgement-focused evaluation

Expert reviews involve a systematic screening of the text corpus by professional linguists. The major advantage of this method is that in-depth valuable feedback, which is based on expert knowledge, is produced. A drawback of this method can arise if evaluators are too close to the text or product that is examined, therefore making it harder to mentally take the users perspective when evaluating the language (Schriver, 1989). Also, this method is quite expensive to scale for products that are translated into many different languages.

2.3. Reader-focused evaluation

Schriver (1989) distinguishes two classes of reader feedback methods: (1) concurrent tests that evaluate the behaviors of readers in real-time, and (2) retrospective tests that are usually applied shortly after the reader has finished reading the text or after a certain time period. Concurrent methods include performance testing and thinking-aloud methods, while retrospective methods involve comprehension tests and surveys. Retrospective user testing is useful for revising existing text (Schriver, 1989).

Reader-focused methods have the advantage of giving information on global aspects of text quality and information about how the audience may respond to the text (Schriver, 1989). While retrospective methods such as surveys have disadvantages over concurrent methods (e.g., thinking-aloud or performance testing) because they rely on the use of memory, a survey during or after the interaction with a software might be a relatively reliable method to measure text quality. An empirical comparison of expert-focused and reader-focused methods of text evaluation showed that mutual agreement on problems in a text among experts is usually relatively low and contributed to a large set of false-alarms – problems that the readers did not report (Lentz and de Jong, 1997). This study also showed that experts experience difficulties with predicting the