INDIGENOUS LANGUAGES SHAPING MULTILINGUAL INTERFACES

TE TAKA KEEGAN
Department of Computer Science
University of Waikato, Aotearoa/New Zealand

Abstract. This paper reports on an investigation into the indigenous language usage of two bilingual/multilingual digital libraries. Results show that the indigenous language was significantly used by clients and indicate why clients chose to use the indigenous language. Feedback from clients has suggested how the interface should be improved to assist both indigenous and non-indigenous language usage. These results serve as an example of how indigenous languages are shaping multilingual interfaces.

1. Introduction

As the amount of indigenous language interfaces being made available in multi-lingual online resources increases it is important to consider the role that indigenous languages have in online environments. Three important questions arise; is an indigenous language used significantly in a multi-lingual interface, if it is used *why* is it used, and, what changes are made in the interface to accommodate the indigenous language that enhance the interface as a whole?

The paper reports some findings from a comprehensive investigation into the indigenous language usage of two digital libraries that make content available in an indigenous language.[1] The indigenous languages analysed in these online contexts are te reo Māori[2] of Aotearoa[3] and the Hawaiian language (of Hawai'i). Both languages have suffered as a result of colonisation to such an extent that it was predicted that they would not survive into the 21st century as living languages (Benton, 1978, 1981). Recent initiatives from the people of both countries have sought to address the language loss and information technology has been identified as one avenue to assist in language revitalisation (P. Keegan, 1996; Warschauer, 1998).

Very little research has been written on indigenous language usage of multilingual interfaces. However Cunliffe (2004) and others (Cunliffe, et al, 2002) have suggested that web presence will become important to language survival and Dyson, Hendricks & Grant (2006) have stated that information technology is being used by indigenous

---

[1] For a full report see T.T. Keegan (2007)

[2] Te reo Māori: the Māori language

[3] Aotearoa/New Zealand is a modern phrase used to describe the country commonly known as New Zealand. For the sake of brevity Aotearoa/New Zealand will be shortened to Aotearoa.

peoples. This paper adds to the indigenous language usage in multilingual interface literature by using both transaction log analysis and user interviews to evaluate the Māori and Hawaiian language usage in two digital library systems.

1.1. THE TWO DIGITAL LIBRARY SYSTEMS ANALYSED

Two online digital libraries (DL) systems were evaluated. Both systems were built using the Greenstone digital library software (www.greenstone.org), both systems made available large collections of legacy newspapers primarily written in an indigenous language and both systems made the material available through bilingual or multilingual interfaces.

The Māori Niupepa DL (www.nzdl.org/niupepa) serves a collection of historic Māori newspapers published between 1842 and 1933 (Apperley, et al, 2002). It is a significant source of historic New Zealand texts—just under 18,000 newspaper pages, collected from 35 separate periodicals. Approximately 54% of the newspapers were written only in te reo Māori, 44% are written in parallel reo Māori and English texts, and the remaining 2% were written solely in English. Three facilities are provided to access the Niupepa documents: full text search, browse by individual title and browse by date. The default language of the interface was alternated in 2005 between te reo Māori and English. The language of the interface may be switched to English or Māori by a button on the home page or changed to any one of 43 other interface languages by accessing the Preferences Page.

Feedback from clients suggested that the Māori Niupepa DL was used by te reo Māori clients primarily for undertaking tribal research, research on te reo Māori and/or on personal or family information. Research on historic practices, assistance with translations and recreational reading were also reported as reasons for using this digital library. The range of industries accessing the Māori Niupepa DL appeared to be mostly limited to academic institutions and tribal organisations with personal researchers also using the resource.

The Hawaiian Nūpepa DL (www.nupepa.org) serves a collection of historic Hawaiian newspapers published between 1834 and 1927. The collection contains over 53,000 newspaper pages, collected from 46 separate periodicals. The newspaper texts are predominantly written in the Hawaiian language; less than 1% is written in English. Three facilities are also provided to access the Hawaiian Nūpepa documents: full text search, browse by individual title and browse by date. The language of the interface is set by default to the Hawaiian language, but may be alternated between English and Hawaiian by clicking a button that is available on all pages of the Nūpepa interface.

Feedback from Hawaiian Nūpepa clients suggested that this digital library was used extensively in the Hawaiian language education institutions, both at secondary and tertiary levels, often as a prescribed teaching resource, for studies on Hawaiian language, history and culture. In addition the Hawaiian Nūpepa DL has been used regularly as a source of written material for Hawaiian language publications.

2. Methodology

Two methods were used to analyse indigenous language usage of the two digital library systems. Transaction Log Analysis (TLA) was used to provide raw statistical data about the number of clients using the DLs in the indigenous language. Client feedback was also generated by using web questionnaires and undertaking user interviews to determine why the indigenous language was being used and how the interface could be improved to better support indigenous language usage.

2.1. TRANSACTION LOG ANALYSIS

Transaction logs were recorded in the year 2005. A transaction log is a detailed record (log) of all interactions (transactions) between a system and clients of that system (Jansen, 2006). Two transactions logs were analysed; first a record of all web browser requests to the Māori Niupepa DL in 2005, and second a record of all web browser requests to the Hawaiian Nūpepa DL in 2005.

Once the transaction logs were collected they were prepared into datasets by removing all the unwanted data. Two types of data were removed, redundant CGI arguments and extraneous requests. The extraneous requests removed included requests from web robots, requests recorded incorrectly (i.e. incorrect or missing CGI arguments) and requests from system developers. A full description of this methodology is given by T.T. Keegan (2007). The resultant datasets contained (as far as could be determined) valid requests from actual clients of the Māori Niupepa and Hawaiian Nūpepa digital libraries. The Māori Niupepa dataset consisted of 185,001 requests and the Hawaiian Nūpepa dataset consisted of 211,215 requests.

2.2. CLIENT FEEDBACK

Two techniques were also used to obtain client feedback; a web questionnaire and client interviews.

The web questionnaire was placed on the Māori Niupepa website in 2006 for a total of 250 days. The questionnaire was designed to be as unobtrusive as possible. It would only appear to those clients who had enabled cookies on their web browsers and who requested the home page a second time. It was a pop-up window that could be simply ignored and full access to the Niupepa DL could continue. Once a client had completed and submitted the questionnaire it would not re-appear to that client. In total, 60 responses were received to the web questionnaire.

Interviews were conducted with clients or potential clients of the Māori Niupepa and Hawaiian Nūpepa digital libraries. Those interviewed comprised 3 clients who used the Māori Niupepa interface in te reo Māori, 8 clients who used the Hawaiian Nūpepa interface in the Hawaiian language and 19 clients who used the Māori Niupepa interface in various languages (e.g. German, Mandarin) but mostly English. While the number of clients interviewed was too small to draw definitive conclusions the interviews were able to provide some important insights about issues with indigenous language usage of digital libraries.

3. Was the indigenous language used?

Before we can begin to determine if indigenous languages are shaping multi-lingual interfaces we must first establish that multilingual interfaces are being used in an indigenous language. The TLA detailed how much the two interfaces in this study were being used in the indigenous language.

3.1. INDIGENOUS USAGE OF MĀORI NIUPEPA DIGITAL LIBRARY

Transaction Log Analysis of the 185,001 filtered Niupepa requests from 2005 showed that the interface was set to te reo Māori in 37,936 requests and set to English in 147,065 requests. The percentages of these figures are displayed in Figure 1.
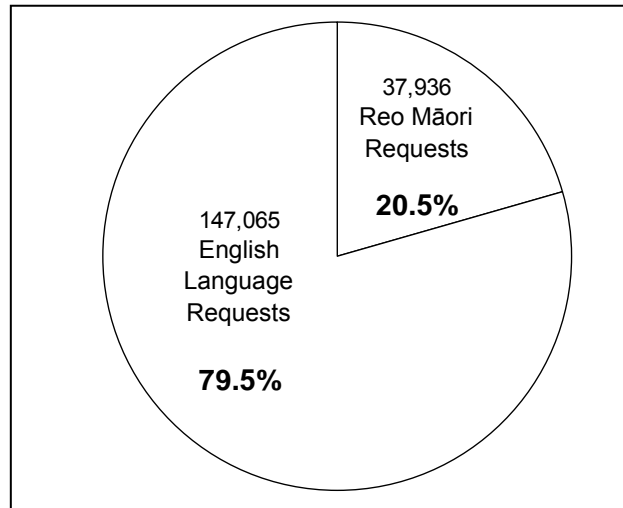
Figure 1. Language of Requests to Māori Niupepa DL in 2005.

3.2. INDIGENOUS USAGE OF HAWAIIAN NŪPEPA DIGITAL LIBRARY

Transaction Log Analysis of the 286,383 filtered Nūpepa requests from 2005 showed that the interface was set to Hawaiian in 112,786 requests and set to English in 98,429 requests. The percentages of these figures are displayed in Figure 2.
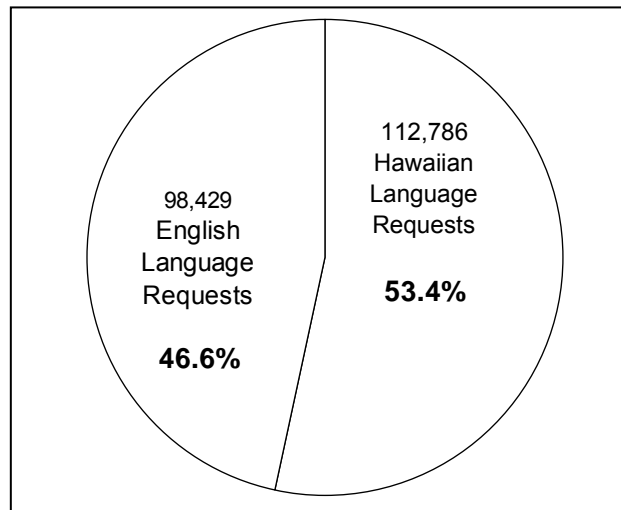
Figure 2. Language of Requests to Māori Niupepa DL in 2005.

3.3. COMMENTS ON THESE RESULTS

Is the 20.5% usage of te reo Māori in the Niupepa DL considered significant? A report published by Te Puni Kōkiri (2007) suggests that 27% of Māori people are able to speak Māori in day-to-day conversations and that there is a strong correlation between the speaking of te reo Māori and the understanding and reading of te reo Māori. The report went on to suggest that 66.7% of Māori have the use of a computer with Internet access, a rate that can be considered comparable to non-Māori. Māori comprise 14.6% of Aotearoa's resident population (Statistics New Zealand, 2006) and if only 27% of this group are able to read te reo Māori it can be surmised that less than 4% of potential Aotearoa clients have the capability to access the interface through the medium of te reo Māori. Given the low number of potential reo Māori clients (less than 4%) the comparatively high usage of the interface in te reo Māori (20.5%) indicates that this interface was used significantly in the indigenous language.

At 53.4% the Hawaiian language usage of Hawaiian Nūpepa DL can obviously be considered significant, but why is this so much higher than in the Māori Niupepa DL? There appear to be two major factors that contribute to this higher rate of indigenous language usage. First the Hawaiian DL makes available virtually no content in the non-indigenous (English) language whereas the Māori Niupepa DL makes available almost 2/3s of it's content in either a parallel English/Māori text or with English abstracts which describe what is written in te reo Māori. Thus the Hawaiian Nūpepa DL is of little use to clients who are not literate in the Hawaiian language whereas the Māori Niupepa DL does provide most of its content in a form suitable for clients who are not literate in te reo Māori. The second factor is that the Hawaiian Nūpepa's default language was constantly set to Hawaiian whereas the Māori Niupepa's default language was set to te reo Māori only 60% of the time in 2005. Bilingual clients (like many other users of online systems) are often unwilling to change system preferences. Research has

shown (T. T. Keegan & Cunningham, 2005) that an online resource whose default interface language is set to a particular language will have a higher rate of usage in that language.

4. Why was the indigenous language used?

To understand why an indigenous language was used an understanding must first be gained as to why the resource itself was used by indigenous language clients. The Introduction Section of this paper has mentioned the background and purpose of clients using the Māori Niupepa and Hawaiian Nūpepa digital libraries. However the web questionnaires and user interviews also highlighted how important the content of the electronic resource was.

Clients reported using the content to assist with discussions on history, cultural practices, land tenure and genealogy. Some of the content had even been presented in courts of law. Indigenous language clients lauded the quality and simplicity of the older style of indigenous language writings and suggested that making this available was important to the revival of the traditional indigenous language.

Some comments by indigenous language clients included:

'A great source of written information pertaining to the lives of my tupuna [*ancestors*], iwi [*tribe*], hapu [*sub-tribe*] and whanau [*family*]. In addition, the papers themselves provide interesting accounts of the social and political context of Maori communities at that time. This is incredibly valuable.'

'Kare i tua atu hei patengi kai ma Hinengaro. Ko te mihi ki a Nehe ma i waiho mai e ratou enei tu korero hei whakamaro i toku reo ake kia rite ki ta ratou i korero ai …'
[*There is no greater database to nourish the mind. Salutations to those of the past who left behind these words to strengthen my own (indigenous) language so that it can become similar to the language spoken by them…]*

Thus, the content attracted many indigenous language clients to use these two digital libraries. Why then did these clients, almost all of whom were bilingual, choose to use these electronic resources in the indigenous language, as opposed to the more commonly spoken English language?

Feedback from clients suggested that the indigenous language was used for three primary reasons:

- the content was primarily in the indigenous language and it was cognitively less challenging to operate the interface in the same (indigenous) language.

- as the default language of the interface was normally set to the indigenous language, clients using the digital library would often continue to use this language if the functions of the interface were easily understood.

- clients wanted to support the use of the indigenous languages. They strongly stated they wanted to use the indigenous language in as many avenues as possible and digital libraries with multi-lingual interfaces provide another avenue to do this.

Indigenous language clients also mentioned that if the indigenous language version of the interface was not clear they would switch to the English language version to understand how the interface worked. Once the interface was understood it was then usually switched back to the indigenous language for the reasons listed above.

5. How can the interface usability be improved to assist indigenous language usage?

Analysis of both the TLA and user feedback highlighted usability issues with the interface design of both digital libraries. These issues are to be addressed by site developers of both the Māori Niupepa DL and Hawaiian Nūpepa DL and serve as an example of how analysis of indigenous language usage can lead to improvements in the design and usability of online systems.

Before listing the usability issues discovered it is worth noting that most of the critiquing appeared to be from non-indigenous clients; indigenous clients appeared less concerned with usability issues and/or more reluctant to mention problems they encountered. There are perhaps three potential reasons why indigenous clients appeared unwilling to mention interface problems: indigenous clients often had more interest in the texts which led them to pay less attention to the flaws in the interface; indigenous clients usually had less experience with online interfaces and, consequently, did not have high expectations as to how easy an interface should be too use; perhaps the unwillingness is part of the indigenous culture where belittling a person or their work is not seen as constructive.

Despite a noticeable reluctance by indigenous language clients to speak negatively about the interface usability issues were identified and have been grouped into four headings as listed in the following four subsections.

5.1. TRANSLATIONS

Often indigenous language online resources are created by translating a non-indigenous language resource into the indigenous language. This was how the te reo Māori version of the Niupepa DL interface was created. Feedback from clients has highlighted issues to be aware of when translating resources into an indigenous language.

Translations should be unambiguous, accurate and relevant to the context. The te reo Māori translation for the link "About English Abstracts" was translated as "He Rāpopototanga Reo Pākehā". This phrase could be misconstrued to mean "a summary in English". TLA indicated the reo Māori version of this link was requested at an unusually high rate, which suggests confusion and, thus, ambiguity. Inaccurate translations can influence language usage in digital libraries.

The site should be translated completely. Some multilingual users mentioned that if they noted a web site partially translated in one language they would switch the interface language and use the resource in the dominant language.

Different language versions of multilingual websites should not be direct translations of each other. TLA indicated that clients who used the site in the indigenous language were most interested in indigenous language documents. Clients who used the site in the non-indigenous language were most interested in the non-indigenous language documents. Consequently, an indigenous language version of the interface should focus on highlighting and directing clients to the indigenous language content in the first instance. Alternatively, a non-indigenous language version of the interface should highlight and direct clients to the non-indigenous language content in the first instance.

5.2. CONTENT DESCRIPTION AND FORMAT

The different types of content should be clearly described. Observation of clients using the Māori Niupepa DL indicated that they were often unsure if they were reading newspaper documents or English abstracts about the newspaper documents. They were also unaware that the English abstracts only summarised te reo Māori documents. Clients were observed searching the English abstracts for English language texts, only to be disappointed when the only results returned were links to te reo Māori documents.

The abstracted text documents need disclaimers. Many clients were unaware that the extracted text documents had been generated by an OCR process that was not 100% accurate. While errors in the extracted texts of documents were sometimes noted and commented on by clients, many clients were unaware that the facsimile image available of the document was an exact replica of the original page and thus contained no extraction errors.

A suitable file format for the display of images online is the pdf format. Images stored in pdf format with a high resolution can easily be enlarged to assist with reading. The pdf format also allowed the document pages to be easily printed.

A list of titles predominantly in one language should also include translations for the benefit of clients who do not understand that language. Even with the interface set to English, a listing of periodical names was perceived as foreign by clients who did not understand the indigenous language. Parallel translations of the titles could assist those clients.

5.3. QUERY ACTIVITY

The number of documents that are searchable should be made known. Many clients assumed that all digital library documents were searched when a query was submitted. However, while 98.6% of documents were searched by the Māori Niupepa search engine only 6.4% of documents were searched by the Hawaiian Nūpepa search engine.

Make as many documents searchable as possible. Both TLA and the user interviews indicated that an important feature of the digital libraries was the ability to undertake full text searching. Clients of the Hawaiian Nūpepa DL expressed an

eagerness for more document pages to become available for full text searching as quickly as possible.

Query parameters and query operators should function correctly. The query parameters and query operators appeared to act incorrectly or inconsistently in one of the digital libraries. Consequently, clients could not rely on these to work correctly and their use was subsequently abandoned.

Search results should contain preview images or summary information about the returned documents. A preview image of the result, especially when text surrounding the query term was included, allowed clients to decide quickly if the document was relevant to the query. Results could be scanned much more quickly and the search activity was more efficient when previews of the returned documents were available. Clients also suggested that when a returned document was displayed following a query, the document should be automatically scrolled to the query term position and the query term should be highlighted.

The user interviews revealed how surprisingly poor digital library clients were at spelling. A spell checking function incorporated into the query term input box would improve the submission of queries. If the spell checker worked with both the indigenous and non-indigenous languages it would be even more useful.

5.4. NAVIGATION AND LAYOUT

Navigation should be clear, logical and consistent. Users commented that at times they were unsure where they were in the digital library and how they could return to a previous screen. Buttons and icons should afford their purpose. User feedback suggested that clients were not aware of the purpose of some buttons and icons simply by looking at them, with different view points offered by indigenous and non-indigenous clients.

The screen layout should be consistent with common practice. Users mentioned that navigation would be easier if the buttons were located in a similar position to that used on other web sites; the home button at the top left and the help button at the top right. The button locations should be consistent throughout the site.

A language toggle button on every page of the interface would assist clients to interact with the digital library bilingually i.e. use the DL in both languages in the one online session. Bilingual usage was quite common in the Hawaiian Nūpepa DL where over a quarter of all usage sessions involved using the interface in the both the Hawaiian language *and* the English language.

6. Summary –the influence of the indigenous language

For indigenous languages to have influence in multilingual interfaces the interfaces must first be built and used in an indigenous language. Analysis of two such systems, the Māori Niupepa DL and the Hawaiian Nūpepa DL, has shown that these systems were used significantly in the indigenous language in 2005. These systems were used in the indigenous language because this language matched the content stored in the DL,

because it was the default language setting of the DL interface, and because the clients has a desire to use their indigenous language in as many avenues as possible.

Feedback from indigenous language clients raised a number of usability issues with the interface which were grouped into headings of; translations, content description and format, query activity, and navigation and layout. The Māori Niupepa DL and Hawaiian Nūpepa DL developers have committed to making changes to their interfaces, based on the issues raised, so that their systems become more usable to the indigenous language clients. These changes, where appropriate, will also be incorporated into the non-indigenous languages versions of the interfaces and serve as an example of how indigenous languages can shape all language versions of a multilingual interface.

References

Apperley, M. D., Keegan, T. T., Cunningham, S. J., & Witten, I. H. (2002). Delivering The Māori Newspapers on the Internet. In J. Curnow, N. Hopa & J. McRae (Eds.), *Rere Atu Taku Manu! Discovering History Language and Politics In The Māori Language Newspapers.* (pp. 211-236). Auckland: Auckland University Press.

Benton, R. A. (1978). *Can the Maori Language Survive?* Wellington: New Zealand Council for Educational Research.

Benton, R. A. (1981). *The Flight of The Amokura.* Wellington: New Zealand Council for Educational Research.

Cunliffe, D., Jones, H., Jarvis, M., Egan, K., & Huws, R. (2002). Information Architecture for Bilingual Web Sites *Journal of the American Society for Information Science and Technology, 53*(10), 866-873.

Cunliffe, D. (2004). Promoting Minority Language Use on Bilingual Websites. *Mercator Media Forum, 7*(1), 42-53.

Dyson, L. E., Hendicks, M., & Grant, S. (2006). *Information Technology and Indigenous People.* Hershey PA: Information Science Publishing.

Jansen, B. (2006). Search log analysis: What it is, what's been done, how to do it. *Library and Information Science Research, 28*(3), 407-432.

Keegan, P. (1996, 14-17 January). *Initiatives in Maori Networking.* Paper presented at the New Zealand Computers in Education Society, Hamilton, New Zealand.

Keegan, T. T. (2007). *Indigenous Language Usage in a Digital Library: He Hautoa Kia Ora Tonu Ai.* PhD thesis, University of Waikato, Hamilton, New Zealand.

Keegan, T. T., & Cunningham, S. J. (2005, May 26-28, 2005). *What Happens If We Switch The Default Language of a Website.* Paper presented at the First International Conference on Web Information Systems and Technologies, Miami, Florida.

Statistics New Zealand. (2006). 2006 Census Data.   Retrieved January 2007, 2007, from http://www.stats.govt.nz/census/2006-census-data/

Te Puni Kōkiri. (2007). Survey on The Health of the Māori Language - Final Report [Electronic Version], 190. Retrieved January 2008 from http://www.tpk.govt.nz/publications/subject/.

Warschauer, M. (1998). Technology and Indigenous Language Revitalization: Analyzing the Experience of Hawai'i. . *Canadian Modern Language Review, 55*(1), 140-161.