Writing to be read: Readability indices for Open Educational Resources

Griff Richards[1,2]

[1]Conseil scolaire francophone de la Colombie-Britannique,
180 – 10200 Shellbridge Way
Richmond, B.C., CANADA V6X 2W7

[2]Simon Fraser University Surrey

griff@sfu.ca

Abstract. Much of the text of pen resources such as Wikipedia is written at a college level of readability, thus posing an access barrier to the general public. Reading levels are important when considering wiki approaches for a K-12 OER community, and even more so when developing for minority language schools. In Canada, many students learning in a minority language setting lack the environmental influences that promote the acquisition of a large fluent vocabulary. This paper discusses plans to automatically index pages in the eVrest wiki by readability, and to gather corroborating usage data on the impact of this approach. Readability might prove a pragmatic adjunct to semantic density when tagging learning objects.

Keywords: Open Educational Resources, learning objects, semantic density, readability, wiki

1 Introduction

Canada's constitution recognizes two official languages. English is dominant in most of Canada other than the Province of Quebec, while French is dominant in Quebec and in historical pockets of French settlement. Where numbers permit, each language group has a constitutional right to educate their children in their mother tongue. Anglophone public schools are found in Quebec, and Francophone public schools are found sprinkled across the rest of Canada. Approximately 142,000 francophone students access their right to French minority language instruction. Most of these are clustered in eastern Ontario or in Canada's eastern maritime provinces, the former French Acadia. In the Canada's west and north about 15,000 students are enrolled in francophone minority schools and of those, only 4,000 are in British Columbia, an area the size of Western Europe.

In Canada's west, almost 80 percent of "francophone" families are "inter-langagière", having only one francophone parent with French as their mother tongue.. Indeed, one regional francophone school near Vancouver boasts 47 different

languages. Not surprisingly, many students in minority language schools struggle to become proficient in their first official language. Francophone minority students often excel in English achievement scores even though they have not been educated in that language.

As the vocabulary and reading level of many francophone minority students falls below that of francophone learners of the same age in Quebec [1] it is difficult to find learning resources written at an appropriate reading level for many curriculum areas. The shortage is particularly felt in secondary math and sciences where texts typically have little regard for the range of reading difficulties found in secondary students [2]. The shortage is exacerbated by the need to align resources with regional (Provincial) curriculum guides..

2 The eVrest context

Écoles virtuelles, ressources éducatives, et strategies en téléapprentissage (eVrest) is a knowledge exchange project to encourage the sharing of resources and on-line education know-how among the minority school boards across Canada. eVrest has established a community wiki for on-going collaboration and exchange of resources. By moving content out of the secure course management systems to a community wiki, we hope to make the content becomes accessible to all learners of the eVrest community.

In response to the specific needs of francophone minority students, the eVrest wiki is considering two innovative features:

1. The automatic analysis and indexing of all postings in terms of readability (in English, French and potentially Spanish),
2. The soliciting of user reviews so that submitted materials can be rated as to usefulness and accessibility by the various levels of users.

Ultimately, eVrest foresees the evolution of a large collection of content suitable for a wide range of reading levels. In this regard, readability indexing provides two advantages – users seeking materials of an appropriate level can add the reading level to their search criteria, and contributors can be careful in writing materials at the appropriate level for these learners. A series of guidelines will also be developed to aid contributors in this process. This approach has confidence that even advanced scientific concepts can b explained in accessible language. It also paves the way for inclusion of learner-generated content once quality assurance processes can be established.

3 Readability

Computation of readability indexing is quite simple, and a number of rubrics are available (see http://juicystudio.com/services/readability for an in depth discussion and demonstration). The Flesch Index [3] is quite common and is used in MS Word for English texts. Formula 1 [4] provides the calculation for a sample of 100 words. It simply factors the average words per sentence (W) and the number of polysyllabic

words (S) and subtracts them from a constant to returns a value between 0 (very easy) and 100 (very hard).

$$\text{Flesch Index English} = \text{Constant } 206.84 - (0.85 \times W) - (1.02 \times S) \quad (\mathbf{1})$$

eVrest is looking at the Flesch Index because the constants have been calibrated for French by Kandel and Moles [5]. Formula 2 [4] provides the corrected weightings.

$$\text{Flesch Index French} = 209 - (0.68 \times W) - (1.15 \times S) \quad (\mathbf{2})$$

A variant of the Flesch Index, the Flesch-Kincaid returns a value roughly equivalent to the grade level. For example 6.4 would indicate 6 years and 4 months of education of the average reader who will find the passage comfortable to read. Automatic calculation of a reading index would require preliminary analysis of the text to best guess whether a passage was written predominantly in English, French, or fell outside the guidelines for indexing.

4 Discussion

There are a number of ways to designate the intended audience for a particular learning resource. The EUN's Calibrate Project [6] used age of the learner population in order to bridge across the nomenclature of several educational systems. This is a valuable meta-datum, but it does not capture the variability in the student population. The rule of thumb for homogeneous language groups has always been that the reading level would extend plus or minus three years. A normal Grade Eight class might be expected to have a normal distribution of reading levels ranging from 5 to 11. The variance is much wider in minority classrooms, and negatively skewed. That said, readers with difficulty are not unintelligent. They are capable of understanding intellectual and scientific concepts if presented at the right level. In the context of the eVrest community, they are often polyglots speaking two or three languages fluently.

The addition of a metadata field for readability is considered because it may prove useful in the context of eVrest, and it is not a technically difficult to accomplish as the calculation can be made each time a wiki page is submitted. Provision of a web form to scan text or links could also make it easy for submitters to contribute the reading level of external links and learning objects. This could also be deployed as a web service and if found to be useful, considered for use in other learning object discovery and exchange settings. It would be a boon to other learners of additional languages.

Readability indices can also provide some direction to authors in terms of the appropriate level of submissions. However one needs to be careful in that writing mechanistically to achieve a readability level can result in unnatural gibberish. Written language reflects a vast complexity of grammar, vocabulary, jargon, and other linguistic conventions that are not captured in a simple readability index. For this reason that readability indices should remain only one guide to finding

appropriate learning resources, and not take on undue precedence over other metadata.

Readability is a simple concept that helps gage the reading difficulty of a text passage. It should not be confused with **semantic density** which "measures an LO's effectiveness as compared to its size or duration." Although semantic density is a defined Learning Object Metadata (LOM) field [7] it remains meaningful only within a community of practice. In reality, there is no true measure of how well a learning object transmits its conceptual message, and semantic density remains an elusive meta-data field that is seldom used. If reading level proves to be both a reliable and useful measure, then consideration might be given to its inclusion in metadata specifications.

The next step in eVrest will be to implant the Readability indices for French and English texts and collect reviews from learners as to their appreciation of the resources accessed. The author sees this as one initial step toward the eventual development of personal language portfolios that will promote the language development of each learner through careful monitoring of their phonological awareness and vocabulary acquisition. Precise and personal attention to supporting the French language is a core value in the preservation of minority language culture. Finding and accessing appropriate learning resources is a major step towards achieving that goal.

Acknowledgements eVrest is in part supported by the Canada Council on Learning Knowledge Exchange Program. Dr. Richards' work is in part supported by Canada's Natural Sciences and Engineering Research Council through the LORNET Research Network. The author wishes to acknowledge the technical assistance of Manon Ruel.

References

1. Dion, S.: "Comment amélioré le recrutement et la retention des élèves dans les ecoles francophones?" Notes pour une allocution de l'honorable Stéphane Dion, 4 Oct 2003, Toronto. Last accessed 19 June 2007 from http://www.pco-bcp.gc.ca/default.asp?Language=F&Page=archive&sub-speeches&Doc=20031004_f.htm
2. Allington, R. : Reading so that all can learn. Keynote presentation to Interactive Innovations Conference, Vancouver, 01 March 2007.
3. Flesch, R. F.: A New Readability Yardstick. Journal of Applied Psychology, 32: 221-233. (1948).
4. Cossette, C. : Les iages démaquillées ou L'iconique : comment lire et écrire des images fonctionelles pour l'enseignment, le journalisme et la publicité. Quebec : Éditions Riguil internationals. Pp 422 – 442. (1982). Last accessed 8 July 2007 from http://comviz.com.ulaval.ca/module1/1.4_sensimagique.php.
5. Kandel, L. and Moles, A. : Application de l'Indice de Flesch à la langue français. Cahiers d'Etudes de Radio-Television, 1958. 19: p. 253-274 – cited at http://www.utexas.edu/research/accessibility/resource/readability/manual/about-English.html
6. EUN : The EUN Learning Resource Exchange Metadata Application Profile Version 3.0, June 2007. Last accessed 11 July 2007 at http://fire.eun.org/LRE-AP-3.0.pdf
7. IEEE IEEE 1484.12.1-2002 : Learning Object Metadata Standard for Learning Objects