EMERGING TECHNOLOGIES
Tag Clouds in the Blogosphere: Electronic Literacy and Social Networking

Robert Godwin-Jones
Virginia Comonwealth University

Electronic literacy today is a moving target. How and why we read and write online are evolving at the fast pace of Internet time. One of the most striking developments in the past few years has been how new social networking phenomena on the Web like community tagging, shared bookmarking, and blogs have created convergences between consumers and creators, between reading and writing, between public and private spaces. Blogs invite us to write responses to items we have read, to move from observer to participant. Shared tagging invites us to analyze texts and sum up their distinctiveness in keywords. Writing online may involve coding or scripting, as we try to add distinctiveness in formatting or interactive functionality to our texts, blurring the lines between writing and programming. Web browsing and reading must be supplemented by abilities in sorting, navigation, and critical thinking. Integration of other media into texts complicates further the notion of literacy. We will examine in this column some of the ways in which these developments are reflected in new tools, services, and approaches to finding, creating, and transforming texts on the Web, and what this might mean for language learning.

Discovery: Tagging and the Semantic Web

One of the challenges we face in using the Web, whether as language learners or instructors, is in finding the resources appropriate to our needs. We know there is a wealth of information and opportunity on the Web, authentic texts in all languages, on-line communities of learners and practitioners, wonderfully inviting Web sites spotlighting cultural practices, vibrant exchanges of views on all subjects under the sun, and all manner of opportunities for reading and writing – if only we could find them. New methods of finding and identifying Web resources involve fundamental skills of analysis, contextualization, and conceptualization, not to mention reading and writing themselves. You can't "tag" a Web resource without being able to extract salient points the author makes, considering how to summarize in keywords what's important, and placing that text in the context of others.

Of course, the traditional and most-widely used means of finding texts, or other Web resources, is to perform a search, most often using Google. With the vastness of the Web today (one report indicates Google indexes over 8 billion Web pages) and the proliferation of junk, googling can be a hit or miss proposition. An alternative to searching is browsing by classification, as in the original Yahoo model. This is an area in which librarians and professional organizations have contributed mightily by evaluating, collecting and annotating categories of texts and resources. A site/service such as Merlot offers expert-based reviewing and ranking of Web sites, including excellent collections of language learning sites. Communities of practice such as Webheads also contribute. On the other hand, many sites that purport to be site collectors are simply commercial endeavors or just place-holders for advertising. As such sites proliferate, students more than ever need skills in critical thinking to be able to sift and evaluate.

One of the proposed solutions to the chaos of the Web, going back to a suggestion from Tim Berners-Lee, the creator of the World Wide Web, is the implementation of what has been called the Semantic Web, a system in which meaningful information about Web texts can be extracted automatically from Web pages and collected by intelligent "agents". Agents are computer programs launched from a server which function autonomously over a period of time, similar to the crawling programs used by search engines to discover and catalog Web pages. By adding meaning to information, the Semantic Web holds the promise of powerful opportunities for creating educational content through combining resources from many sources, using human or machine means, to build a variety of customized learning resources.

The challenge of the fulfillment of this vision is its reliance on 1) the inclusion of meta-data and 2) an established set of ontologies which explain terms and relations in a given subject area. The ontologies allow agents to make sense of the resource's meta-data. Creating such ontologies is not an easy process, nor one on which consensus is likely to be easy to reach. A recent development that might be of help is the creation of a Web ontology language, OWL, a markup language for publishing and sharing ontologies on the Web. The second technical requirement for the Semantic Web is wide-spread use of meta-data -- this has been a tough sell to Web authors. Although meta-data systems such as the Dublin Core and IMS-LOM have been in place for some time, they are by no means universally used (even by search engines). The meta-data specification most often associated with the Semantic Web is RDF (Resource Framework Discovery). RDF describes resources in XML and is meant to be used in situations in which the information needs to be processed by applications, rather than to be displayed to people. The recently proposed RDF/A specification streamlines considerably the creation of RDF by allowing it to be directly embedded into the HTML of a page (added as a simple tag attribute) rather than contained in a separate file or in the page header.

The promise of the Semantic Web is evident in the experimental "semantic browser", Magpie, an add-on to Internet Explorer or Mozilla/Firefox, which associates words and phrases in a Web text with available ontologies and keeps track of key terms in dynamically created "collectors". The unique feature of Magpie is that it does not require manually annotated texts but searches and collects based on keywords in the appropriate ontology. Another alternative browser, Conzilla, is a "concept browser" which presents information in the form of context maps. W3C's Amaya is an experimental browser that leverages the combination of ontologies and RDF; it makes use of a W3C project called Annotea, which features shared annotations stored on a central server. An implementation of the kind of text mining and collecting envisioned by the Semantic Web can be seen in the daily news analysis (Europe Media Monitor) available from the Joint Research Centre of the EU. It searches out articles written in a variety of languages in a given subject area, extracts and stores references to places, people, and organizations and generates a geographical map (highlighting mentioned locations) and a set of commented links. As more keywords are used in different news clusters, the system learns over time which entities are associated with one another.

While the Semantic Web has been mostly of scholarly interest and not widely discussed outside of academic and techie circles, another effort to create order out of chaos on the Web has proven to be explosively popular. Community tagging is a bottom-up, grass-roots phenomenon, in which users classify resources with searchable keywords. The tags are free-form labels chosen by the user, not selected from a controlled vocabulary. The first wide-spread use was on flickr, a site which offers photo-sharing services. Users of flickr are able to add their own tags to any photo. Users can also aggregate pictures into photosets, create public or private groups, and easily add flickr-stored photos to a blog. In the past two years there have been a number of sites and services which make use of this kind of open tagging system. Some of the better-known are del.icio.us, a bookmarking service, Technorati, a blog cataloging site, and digg, a gathering place for tech fans. These sites create clickable "tag clouds" for resources, groupings of tags arranged alphabetically, with the most used or popular keywords highlighted through being shown in a larger font. Figure 1 below shows the most popular tags on flickr from the middle of March, 2006.

One should note that the tags represented here are all in English, but on some sites (particularly on Technorati) other languages are also used. There are tagging sites which cater to other languages such as French (BlogMarks) and Japanese (Livemark). Many such sites make use of RSS (Really Simple Syndication) to notify interested users of changes and new developments. In flickr, RSS feeds can be attached to individual tags, or to photos and discussions. In addition to RSS, flickr and other social networking sites typically offer functions such as search (for users and tags), comments (and comment trails), and APIs (application program interfaces) for posting to or from the tools, used especially in combination with blogs. An interesting use of RSS in combination with tagging is at the Flashcard

exchange, where, for example, one can view or subscribe to all flashcards posted for learning Spanish (or other languages).

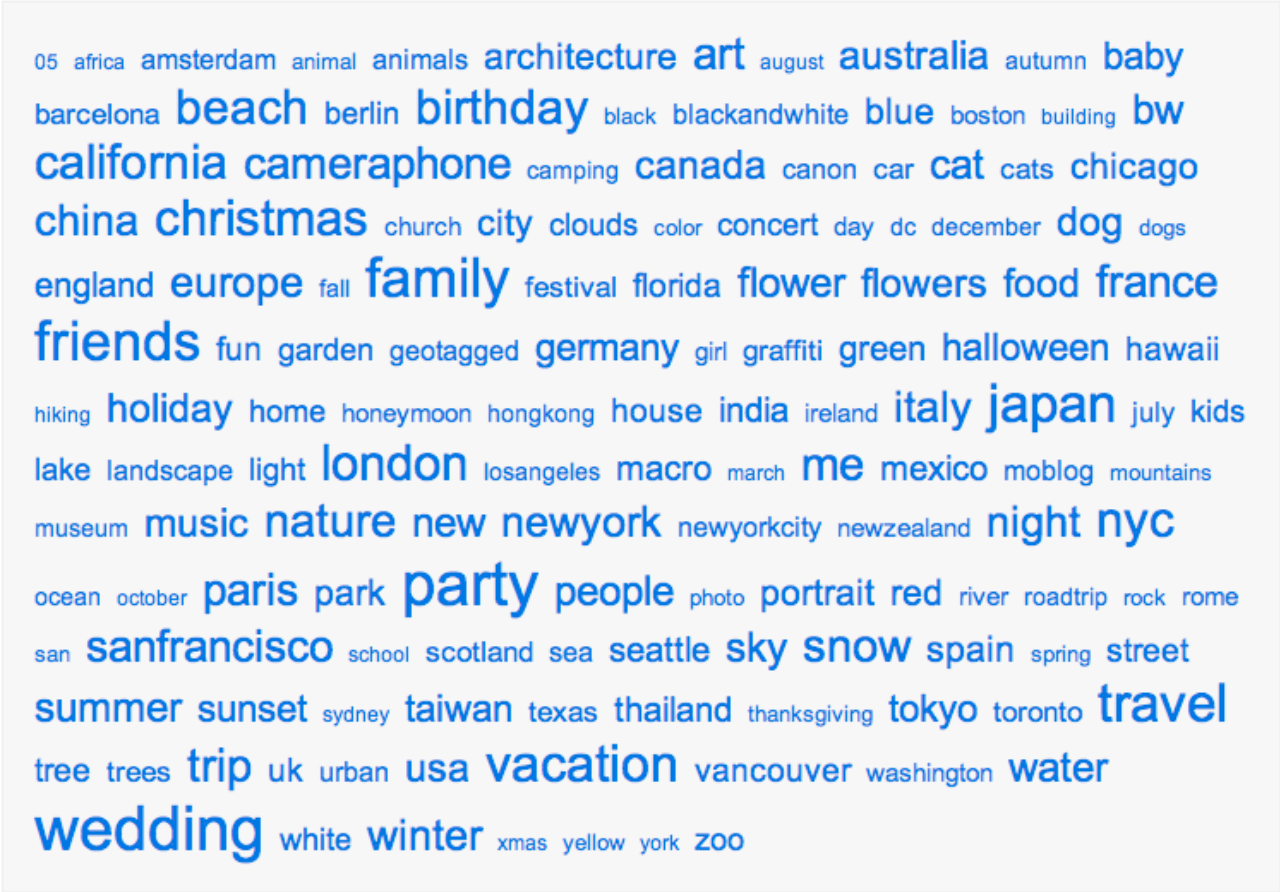

Figure 1. Tag cloud from flickr

The tagging process is by no means simply technical – a way of categorizing resources – it also has a strong social dimension as users of the site find common interests and create on-line communities. It represents another example of the fuzziness separating consumers and creators on the Web today. A contribution to a tagging site, seen by other users, may cause additional tags or comments to be added, automatically building and updating and thus ultimately defining a resource. Instead of one person making a judgment about a blog entry, photo, or other resource, a consensual classification is created. In effect, a text or object identifies itself over time. This creation of "folksonomies", as they have been called, can be seen as a democratic implementation of the Semantic Web. The idea of users becoming creators is one of the key concepts behind what some refer to as Web 2.0. It also involves the kind of social networking and "collective filtering" that can be seen on sites such as amazon, ebay, or netflix, in which users' reviews and comments build a self-generating database of information. The emphasis is on the Web as a gathering place in which users both benefit and contribute. Of course, in the process a lot of reading and writing is being done in discussion forums. Feedback or comment forms are part of all social or community networking sites.

Creating: Blogs and on-line writing

Just as bookmarking through del.icio.us has moved from an individual and private to a shared, public process, blogs too move writing into the public domain, the blogosphere. Blogs by their nature and page structure encourage feedback and represent both a reading and a writing activity. In the best of cases, this

kind of online writing stimulates debate, furthers critical analysis, and encourages articulation of ideas and opinions. In form, most blogs look very similar, with new entries at the top of the page, followed by feedback and accompanied by links on the side of the page. Although all blogs are public, some tend to be outwardly directed and others more inward and introspective. There have been some interesting discussions of blogs by composition professionals, focusing on this private/public dichotomy. While language instructors may use blogs for reading/writing practice in the target language, most educational uses of blogs have involved course blogs in which the instructor leads a discussion on course-related topics. Blogs offer interesting opportunities for collaborative projects, debates, or interactive travel logs. They provide an opportunity for students to write in a public sphere (as compared to closed discussion forums) and in a more coherent and organized way than in chat or instant messaging. A major benefit of using blogs is to provide an environment in which students engage with the topic and with one another, developing skills of persuasion and argumentation.

Blogs have experienced phenomenal growth in recent years. It has been estimated that some 70,00 new blogs are created every day (almost 1 a second). Technorati in February, 2006, was indexing 28 million blogs. There is now also an astonishing variety of blogging software and services, with most bloggers electing to have their blogs hosted rather than running their own server software. Popular blogging software such as Blogger, WordPress or Moveable Type make setting up and operating a blog not much more difficult than creating a PowerPoint presentation. Some blogging software specializes in particular uses, such as photo blogging (Buzznet), mobile blogging (Blogplanet), audio (Audioblog), video (vidblogs), or news (GrokSoup). A number of sites specialize in setting up blogs for educational use: Blogs2Teach, weblogs4schools, and edublogs. The eslblogs site offers free blogs for ESL students and instructors. Many blog enthusiasts use the Technorati site as a means to identify blogs of potential interest. Technorati recently added the capability of using tags to label not only individual blog posts, but also entire blogs.

One of the interesting new developments in the blogosphere is the integration of blogs with other tools and services. Drupal combines a blog with a powerful content management system, used famously in the Howard Dean US presidential campaign. Elgg combines blogs, e-portfolios, and social networking, and also includes many other functions such as file repositories, community tags, and podcasting. Both services represent a new breed of open source, group software with the emphasis both on individualizing resources and on creating a welcoming social network. Both systems can be substantially customized and extended; Elgg, for example, can be integrated into moodle, the open source learning management system (LMS). Some instructors are finding that using a course blog offers a possible alternative to a traditional LMS such as Blackboard or WebCT. It is possible to create a more student-centered learning environment using blogs, particularly if students create blogs that they control and whose content they own. Student blogs can be linked to a course site (or blog), even to a conventional LMS. The difference, especially to LMS discussion forums, is that through their own blogs students connect not only with their school communities, but also with other communities (social, professional, family, hobbies), including ones which may be important for them after graduation. A student blog, in addition to serving as a social and educational tool, can also function nicely as a personal portfolio. It should be mentioned that Blackboard has promised to add blogs to its set of tools, and WebCT already has done so. However, the blogs in WebCT are still trapped within the proprietary system and do not include the essential blog feature of RSS feeds. There are several services recently established, WordPress MU (for multi-user) and Lyceum, with the express purpose of allowing creation of multiple individual blogs which can be easily grouped together.

Delivery: Adding Value to Electronic Texts

Blogs offer reading (and writing) practice in everyday, informal language. A different language register can be experienced through working with literary texts in electronic formats. Electronic texts abound on the Internet, although e-books, texts readable on small, portable devices, have failed to catch on in a big

way. There was a discussion recently on Slashdot about why that might be the case. The consensus response points to a variety of factors: 1) the practicality of books over e-books, 2) the absence of a good, portable, easily readable device for reading e-books, and 3) the scarcity of content in the format needed. What complicates the last issue is also the problem of digital rights management (DRM), restrictions on the use of the e-book that are built in to its delivery format. While e-books in different languages may be of interest to language learners, the key factor in using electronic texts in language learning is not their portability but the extent to which they are accompanied by comprehension aids or other add-ons. Static texts on a screen hold little if any benefit over print versions, although the availability of many texts in different languages can be a significant resource to language teachers and learners. Electronic texts add value if they incorporate features such as glosses, notes, multimedia annotations, or translations. Unfortunately, there are relatively few freely-available texts which add such comprehension help. This seems surprising, considering how relatively easy it has become to create an annotated text, as demonstrated by John Kundert-Gibbs in his on-line Chaucer edition. There are, of course, sites which fulfill this mission well, such as the Perseus Project, the Edda Project, or the Digital Dante Project. For more on the latter, see the extensive discussion in the LLT "On the Net" column in the January, 2006, issue.

In fact, much can be done in terms of presenting texts to students for language learning. In addition to glosses (textual or graphical), audio clips can accompany a text, as well as links to external reference sources such as dictionaries or encyclopedias. Comprehension questions can be included to test students' understanding of what they have read. Indeed, adaptive electronic texts could be made which deliver the text differently depending on student responses, sending a student back for remediation, for example, if a set of questions is not answered correctly. There are tools which can create this kind of interactivity automatically through pull-down menus. CourseGenie transforms Word documents into interactive Web pages. The LessonBuilder by SoftChalk allows for pop-up annotations, in-line questions, and activities. These and similar tools make it easy to create interactive texts and also offer many formatting options, usually by including multiple style sheets or page templates from which to select.

One of the other alternatives for creating interactive text pages is to encode the text in XML, then to use one of the variety of available tools to display it in a user-chosen manner in a Web browser, converted on the fly to HTML. This is, for instance, how the University of Virginia Library delivers its texts in different formats, including HTML, PDF and mobile device formats, from the same document. The texts are encoded in XML, using the guidelines of the TEI, the text encoding initiative. Other text collections, such as Oklahoma State's Electronic Publishing Center and The University of Texas's Dumble Survey, also use this approach. One of the essential tools in enabling these transformations in XSLT, Extensible Style Language Transformations, which is most often used to convert XML to HTML and other formats. In comparison to CSS (Cascading Style Sheets), XSLT is quite complex, but also very powerful. Google recently posted the source code for AJAXSLT, an implementation of XSLT in Javascript, used to create powerful server-client applications like Google Maps, Gmail, and Google Earth. XSLT could be used in language learning to parse a text, and through a Web service, search an on-line dictionary, automatically linking in definitions. This would allow authentic texts to be automatically glossed for student use. An experimental version of such an interface can be seen in an online version of a German fairy tale, in which a script seeks out glosses based on searching (in order): 1) a (local) story word list, 2) a (local) intermediate German-English glossary, and 3) an (online) German-English dictionary. The advantage of having glosses loaded locally in the browser memory (through the use of JavaScript) is to speed the look-up process, as well as to allow for offline use. This site uses a traditional server-based CGI script, but could be done better (and run faster) in XSLT. Having texts in XML makes it much easier to add additional Web-based services, as well as to extract information automatically from the text that could be used in other ways.

In addition to building structural tags in a text using XML, it would be interesting to enable community tagging of texts as part of a language learning process. This might involve groups of students creating word groups or simply finding appropriate keywords to describe a text or parts of a text. This would offer options for collaborative projects, classes or groups together tagging a text or a collection of texts. Online forums or blogs could be used to discuss the results of the tagging, perhaps in the context of other views/interpretations of the text. In the vision of a future Semantic Web, one could easily see a role for language learning, as texts such as blog posts, newspaper clippings, journal entries, literary texts, are collected by agents, based either on community tags or hierarchical taxonomies (or both). This might involve identifying not only the kind and purpose of the text, but also its language level in terms of vocabulary and style, its intended audience and its popularity (or lack thereof). As a result students could work with different groupings of texts, including much more variety than is the case in the typical language textbook or reader, leading to the development of reading and writing skills in a variety of registers.

The development of multiple literacies is needed in an environment in which there are no clear boundaries between text and other media. Digital media students today incorporate sound and moving images as much as they do text, and often feature non-linear browsing and interactivity. One of the key features of the evolving online world is that it offers an ever-shifting blend of individualization and community involvement. Working with online media is quite different in this respect from reading a book. One need only consider the experience of using a news-feed collector which accumulates RSS feeds from multiple sources. The subject may change from paragraph to paragraph and, depending on the student's interest level, the text feed might be scanned, the entire post might be read, or the student might go to the Web site to see the text in context, to read more, or to write a comment. There is a clear social dimension to electronic literacy; reading and writing on-line are often collaborative activities. As educators we not only need to facilitate literacy skills in this new environment, we also need to be creating language learning media or applications which mirror the kind of online world students experience -- student-centered with collaborative opportunities, allowing plenty of space for creative and reflective processes.

RESOURCE LIST

Literacy and reading on-line

- Electronic Literacy By David Reinking
- Beyond the Digital Divide? Investigations of Internet Access and Agency in a Mobile Era Special issue of C*urrents in Electronic Literacy*
- Exploring New Literacies in Online Peer-Learning Environments By Cynthia C. Choi and Hsiang-ju Ho
- Reading Online Electronic Journal of the International Reading Association
- 21st Century Literacies Good links on new literacies; also lessons
- Beyond Classroom Boundaries: Constructivist Teaching with the Internet By Amelia E. El-Hindi
- Constructing Our Identities through Online Images By Gail E. Hawisher
- Digital Discussions: La Esperanza in the Shared Virtual Classroom By Judi Moreillon
- Defining New Literacies in Curricular Practice By Ladislaus M. Semali
- Graphic Novels for Multiple Literacies By Gretchen E. Schwarz
- A Face-to-Face Graduate Class Goes Online: Challenges and Successes By Jan Turbill
- Internet Project: Preparing Students for New Literacies in a Global Village By Donald J. Leu, Jr.
- Learning on the Web: A Content Literacy Perspective By John E. McEneaney

The Semantic Web and Community Tagging

- The Educational Semantic Web Special Issue of the Journal of Interactive Media in Education
- del.icio.us Example of tagging (for shared bookmarks)
- The Semantic Web In Breadth By Aaron Swartz
- How the Semantic Web works From w3C
- OWL (Web Ontology Language From W3C
- Ontology Development 101: A Guide to Creating Your First Ontology By Natalya F. Noy and Deborah L. McGuinness
- A Semantic Web Primer for Object-Oriented Software Developers From W3C
- RDF/A Primer 1.0 From W3C
- eLGGG Learning Landscape "Open source learning landscape platform"
- Drupal Open source content management platform
- Magpie Semantic browser
- Conzilla A concept browser
- CIPHER project For creating collaborative cultural heritage sites
- Europe Media Monitor Collector of news articles
- EMM NewsExplorer News collector in multiple languages
- Annotea An Open RDF Infrastructure for Shared Web Annotations
- digg Collaborative site collector
- 'Tagging' gives Web a human meaning From news.com
- Scoop Software for content management and annotating
- flickr Popular photo-sharing site
- RawSugar Tagging service
- Rojo Tagging service
- Tags & Folksonomies - What are they, and why should you care? From threadwatch
- Folksonomy From wikipedia
- Collaborative tagging From wikipedia
- Consensus Web Filters From Kevin Kelly
- reddit Web filter site
- Web 2.0 needs Data 2.0 From tagschema.com
- Tag Clouds From wikipedia
- Flashcards in Spanish From the Flashcard Exchange
- Tag mania sweeps the Web From InfoWorld
- shadows Tagging service
- Folksonomies - Cooperative Classification and Communication Through Shared Metadata By Adam Mathes
- Social Bookmarking Tools (I): A General Review From D-Lib Magazine
- What is Web 2.0 From O'Reilly
- Webheads ESL Community of Practice

Blogs and on-line writing

- Web Watch: Writing forthe Web By Susan Deysher
- Bridging the Composition Divide: Blog Pedagogy and the Potential for Agonistic Classrooms By Janice Wendi Fernheimer and Thomas J. Nelson
- The Digital Divide and its Discontents By Matthew Payne
- Best of the Blogs From Deutsche Welle
- edu.blogs Ewan McIntosh's blog (Scottish language teacher)
- Blogs From Vance Stevens
- Blogs 2 Teach UK service for teachers
- Blogging Tools From the e-learning centre
- Weblogs in Education
- Blogs as Course Management Systems From The Electric Lyceum
- eLGG and Moodle integration
- WordPress MU Multi-user version of the popular blogging software

Electronic Texts

- Literature on the Web By Donna Neutze
- eBooks - What's Holding You Back? Discussion on slashdot
- Google AJAXSLT Google Ajax library for transforming text using XSL-T
- Careo Campus Repository of Educational Objects
- Creating Interactive Glosses and Sounds: Chaucer's The Canterbury Tales By John Kundert-Gibbs
- Design And Evaluation Of The User Interface Of Foreign Language Multimedia Software: A Cognitive Approach By Jan Plass (in LLT)
- XSL stylesheets for TEI XML For transforming TEI formatted text to pdf
- The Linguist's Shoebox Tools for working with texts
- What's the Big Deal about XML? By Bill Walker
- Perseus Project Electronic classical texts
- COHSE The Conceptual Open Hypermedia Project
- Electronic Publishing Center Oklahoma State Library
- Electronic Text Collections in Western European Literature From the University of Virginia