

Jung-ran Park

Human language

Resources from linguistics and beyond

Language, whether it be written, spoken, or signed, is what defines us as human beings. Intellectual activity, cognition, and all the products that flow from such are indeed based on the unique ability of *homo sapiens sapiens* to acquire and use a native language. Put another way, the essence of human intellectual and cultural heritage is made possible through the core medium of human language expressed through other media such as paper, audiovisual recordings, microform, and digital media, through which instrumentation knowledge, information, and culture is passed along and perpetuated both globally in real time and across the generations to time *in nitum*. In this sense, linguistics, the discipline dealing with language, is widely considered to be a meta-discipline.¹

Globalization and the advancement of Web technologies have foregrounded multilingual, multicultural, and multidisciplinary contexts and disciplines. These global and Web contexts place a high demand on language-related resources. This article aims at introducing and reviewing language-related Internet sites covering computational linguistics, which is closely interconnected with library and information science, computer science, and engineering, as well as linguistics *per se*, which itself is interconnected with various disciplines. These sites encompass language data covering field notes, lexical resources, written and spoken corpora, and language fonts and software, together with second-language learning resources, linguist-mediated digitization activities for preserving endangered human cultures and languages, e-books and e-journals, and more.

Meta-sites

The ACL NLP/CL Universe. Hosted by the Association for Computational Linguistics (ACL), this site has been devoted to natural language processing and computational linguistics since 1995. It is a comprehensive listing covering introductory materials on computational linguistics, various resources (bibliography, journals, papers, dictionaries, corpora, and natural language tools), software encompassing knowledge representation and information retrieval, subject-specific resources such as speech processing, discourse, semantics, machine translation, and natural language understanding. It also includes listings of academic departments, organizations, conferences, and research labs. The Browse the Universe interface allows users to navigate the site to interdisciplinary domains on language, computation, cognition, and information. *Access:* <http://tangra.si.umich.edu/clair/universe-rk/html/u/db/acl/>.

Ethnologue. Hosted by SIL International, this site is a veritable guide to the world's approximately 6,500 languages and cultures, providing a bounty of sociolinguistic and demographic data in addition to linguistic information. Special attention is given to lesser-known and studied languages. This site is one of the most comprehensive sites of language resources available, owing

Jung-ran Park is an assistant professor in the College of Information Science and Technology at Drexel University, e-mail: Jung-ran.park@cis.drexel.edu. Appreciation is expressed to research assistant Sangjoon Park for his help in gathering language resources on the Web.

© 2005 Jung-ran Park

to a database over 50 years in the making. Features include a massive bibliography, language maps, an online bookstore, and a broad array of software tools and computer resources, some available for free download. *Access:* <http://www.ethnologue.com/>.

Foreign Language Resources. Run under the aegis of Roger Williams University, this site is mainly centered on the major European languages, with links to newspapers, dictionaries, databases, professional organizations, and other Web resources. Provides links to the major comprehensive language and linguistic sites. *Access:* <http://library.rwu.edu/subjectguides/foreignlang.html>.

iLove Languages. Formerly the Human-Languages Page, redesigned by the same author, Tyler Chambers, iLove Languages is an excellent catalogue of resources on individual languages in relation to language learning and education. Includes links to translating dictionaries, native literature, language schools, and so on. *Access:* <http://www.ilovelanguages.com/>.

• **The Linguistic Data Consortium (LDC).** Hosted by the University of Pennsylvania, the LDC touts a membership more than 100-strong comprising universities, private companies and government research labs. This is probably the preeminent site for a wide array of speech, natural language and text databases together with many natural speech corpora and lexicons. Both English and foreign language corpora are represented. Included are a wide array of data, tools and standards, all easily navigable. Many corpora available for free download; others are restricted to members of the consortium. Researchers and scholars working in the area of computer-based linguistic technologies and natural language processing would be well served by checking this site out. *Access:* <http://www ldc.upenn.edu/>

The LINGUIST List. This site provides an academic forum for linguistic issues and for exchanging linguistic information. This is the list that in essence out-puts the discipline with the infrastructure necessary for viability in the digital information universe. The list

claims over 20,000 subscribers worldwide. In addition, it is the best maintained (with very frequent updates) of any such site on the Web, and its extensive resources cover all branches of linguistics.



It functions as the principal channel for the activities of the various linguistic communities and acts as a gateway to open language archives covering endangered languages and cultures, language processing tools, primary sources, and more. *Access:* <http://www.linguistlist.org/>.

Linguistic Resources on the Internet.

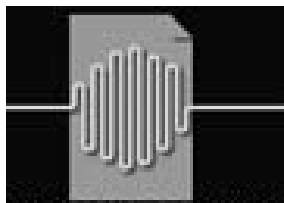
As indicated in the heading, this site, provided by the Summer Institute of Linguistics (SIL), offers extensive and authoritative linguistic resources organized into the following linguistic topics: speech and phonetics, morphology, grammar and syntax, text analysis and corpus linguistics, semantics and semiotics, lexicography and dictionaries, languages and language families, language rights and pedagogical resources. Topics are further categorized into research and research projects. *Access:* <http://www.sil.org/linguistics/topical.html>.



*Partners in
Language Development*

• **OLAC: Open Language Archives Community.** The compass of these 31 archives is international in scope and provides excellent and extensive primary sources in relation to language, culture, and open source language tools. These archives can be categorized into three subject domains: archives that concern preservation of indigenous and endangered languages and cultures (mostly composed of ethnographic resources such as audio-recordings of interviews with text transcriptions, naturally occurring discourse, ritual speech, songs, etc.); several large-scale archives composed of mostly open source tools dealing with hu-

man language technology, covering electronic dictionaries, electronic textual databases and multimedia and multi-modal databases that integrate speech, text and gesture and that in turn are linked to audio-visual media and natural language processing software such as parsers and speech recognizers; and archives of documentary material of over 8,000 languages and dialects worldwide together with material on linguistic and ESL (English as Second Language) studies. *Access:* <http://www.language-archives.org/>



Speech on the Web. This site is devoted mainly to areas of phonetics and the speech sciences. Numerous links are provided to meetings and workshops, dictionaries, electronic journals, and publishers. Computational linguistics, natural language processing, and artificial intelligence are linked insofar as they relate to phonetics and the speech sciences. The site has a very basic but easy-to-navigate layout. As the disclaimer at the site addresses, there is currently a backlog in new link additions. *Access:* http://fonsg3.let.uva.nl/Other_pages.html.

Yamada Language Guides. Run under the aegis of the University of Oregon, this is the main competition to iLove Languages in content and style. A comprehensive guide to information on worldwide languages, the site includes useful and in-depth annotated listings of language-related news groups and mailing lists. An outstanding feature is the provision of fonts for different languages. The virtual language lab is of some use. *Access:* <http://babel.uoregon.edu/yamada/guides.html>

Language Processing Tools and Software

Fonts in Cyberspace. As mentioned earlier, SIL provides an extensive guide list to language fonts containing over 400 sources for 123 languages. Provides links to various font archives as well as commercial fonts. *Access:* <http://www.sil.org/computing/fonts/index.htm>.

Linguistics Computing Resources on the Internet. SIL provides linguistic computing resources organized by topical categories. For example, under software tools, users can find a variety of language processing tools covering fonts, multilingual resources, speech analysis, text analysis, translation, and so on. *Access:* <http://www.sil.org/linguistics/computing.html>.

Natural Language Software Registry. This site is a superb compendium of the sources and capabilities of the range of natural language processing software available on the Web and secondarily of other natural language resources that are available. With the latest edition of the registry comes excellent added functionalities, including the provision of the capacity for menu-guided queries in addition to the previously highly structured listings and descriptions of software products. *Access:* <http://registry.dfki.de/>.

Software. Provided through the LINGUIST site, a broad range of language processing software is presented, together with extremely useful annotated descriptions. The following are some of the categories of software to be found at the site: computer-aided translation, fieldwork, lexicons, parsers, taggers, transcriptions, and speech analysis. An easily navigable resource for those concerned mainly with software products and resources. *Access:* <http://linguistlist.org/sp/Software.html>.

Yamada Language Center: Font Archive. Provides an extensive array of non-English fonts. *Access:* <http://babel.uoregon.edu/yamada/fonts.html>.

Corpora and Lexicon

Corpus Linguistics. Similar to but not quite as extensive as the above site. Includes links to text sites (comprising corpora, newspapers, and news sites) in English and a range of mostly European languages (but also Mandarin Chinese, Malay, and Hebrew), a section for learner corpora, software encompassing taggers and products for text analysis, online taggers and theses, and a bibliography. *Access:* <http://www.athel.com/corpus.html>.

Corpus Resources. Provides links to resources sectioned into corpora comprising an array of European- and Asian-based languages (and Pidgin and Creole sites), word lists, text archives, POS taggers and parsers, and others. A well-annotated and frequently updated site *Access:* <http://pioneer.chula.ac.th/~awirote/ling/corpuslst.htm>.

Dictionaries. Provided via the LINGUIST site, this site encompasses a large list of dictionaries comprising monolingual, bilingual and multilingual resources. It also leads users to other dictionary metasites. *Access:* <http://linguistlist.org/sp/Dict.html>

Lexigraf Web page. Provides information on the multilingual science lexicography project currently taking place at Aristotle University (Thessaloniki) together with the resources and tools being developed in conjunction with the project. *Access:* <http://egnatia.ee.auth.gr/~yhat/yiannis/>.

Links to Corpus Linguistics & Related Sites. For the corpus-based researcher, a wide array of sites are listed here in a very basic but easy to navigate layout. Included are links to all the major corpus linguistic sites and projects (with a section devoted to resources in Polish), bibliographies, corpus and linguistics online courses, tutorials, and glossaries. Also included are links to downloadable software for corpus work, links to computational technology and language technology sites, an online library, and a listing of online journals and newspapers. Frequently updated. *Access:* <http://www.staff.amu.edu.pl/~przemka/corplink.html>.

SIGLEX. This is the site of a special interest group on the lexicon for the Association for Computational Linguistics. As the name suggests, the site is mainly centered on issues and links related to lexical issues and is an excellent resource for researchers and scholars in this area. Divided into two main sections, online resources and corpora archival links, the site is not nearly as extensive in listings as others. *Access:* <http://www.clres.com/siglex.html>.

WordNet. A product of the Cognitive Science Laboratory at Princeton University,

this site bills itself as a lexical database for the English language and is meant to be easily downloadable. It can also be used online with easy functionality. According to the site, it is organized based on current psycholinguistic theories of human lexical memory, and, as such, is divided into synonym sets covering the major parts of speech, each set covering one underlying lexical concept. An excellent resource for English corpus based researchers. *Access:* <http://wordnet.princeton.edu/>

Online Journals/Papers/Books

The Internet TESL Journal. A monthly Web journal for teachers of English as a Second Language, this site covers articles, research papers, lesson plans, classroom handouts, teaching ideas, and associated links. A very good resource for up-to-date material in this field. *Access:* <http://iteslj.org/>.

Journal of Language and Linguistics. This is an online journal covering theoretical and applied topics in linguistics, language studies, and language learning. *Access:* <http://www.jllonline.net/>.

Linguistics Journals and Newsletters on the Web. A substantial number of e-journals are provided here, some of which are available free for download. However, there were also several broken links encountered by the author. *Access:* <http://www.ciil.org/virlib/Univ.rochesterlist%20of%20Journals%20on%20the%20Web.htm>.

Survey of the State of the Art in Human Language Technology. This is an online book of approximately 600 pages dealing with issues related to language technology. Published in 1996. *Access:* <http://cslu.cse.ogi.edu/HLTsurvey/>.

Associations/Organizations

Listed below are sites not touched on in earlier sections:

The Consortium for Lexical Research. *Access:* <http://clr.nmsu.edu/Tools/CLR/>.

CSLU: Center for Spoken Language Understanding. *Access:* <http://cslu.cse.ogi.edu/>.

(continued on page 228)

ACRL's ad revenues, a significant portion of the Association's revenues, have dropped dramatically,

The dues of the organization have not increased since 1990/1 (this dues increase was approved by membership on the 1989 ballot).

ACRL has the lowest dues of any ALA divisions and is the only division that provides two publications as a requisite of membership.

In ACRL, the cost of providing basic member services is \$75/member/year.

Overall, expenses are going up and revenues are going down, with the result that the Association is rapidly spending down its operating reserve.

The association needs additional revenues to forward the strategic plan.

Be it resolved, that the ACRL Board establish the annual membership dues as follows, subject to subsequent approval by the ACRL membership as required under the ACRL bylaws:

Annual dues for personal members shall be set at \$45 for one year (2006) and \$55 thereafter, except that the annual dues for personal members who are full-time students or retirees shall be set at \$35.

Approved an increase in the ACRL Organizational Member dues to \$110.00 in the next membership renewal cycle, FY06. *W*

(The Must List continued from page 205)

ies, television shows, just about anything in popular culture) that they love each week.

2. Frances Maloy. Creativity as Leadership Strategy in Times of Change, *College & Research Libraries News* 65, no. 8 (September 2004): 444.

3. Asian American cinema resources, www.lib.washington.edu/subject/communications/bi/com495/.

4. UW Libraries Resources by Subject, www.lib.washington.edu/subject/.

5. The *Another HYSTERIC Librarian for Freedom* button that I purchased at the ALA Store serves as a response to Attorney General John Ashcroft's comments about librarians' hysteria over the Patriot Act.

6. The September Project, www.theseptemberproject.org/ *W*

(Internet Resources continued from page 176)

EAGLES On Line: Expert Advisory Group on Language Engineering Standards. Access: <http://www.ilc.cnr.it/EAGLES96/home.html>.

European Language Resources Association (ELRA). Access: <http://www.elra.info/>.

Linguistic Society of America (LSA). Access: <http://www.lsadc.org/>.

Discussion Lists and Reference Service

Ask a Linguist. LINGUIST list also provides a reference service to users with a panel of 60 professional linguists available for any inquiries about linguistics. This service is very similar to the reference service in many libraries through Ask a Librarian. Access:

<http://linguistlist.org/ask-ling/index.html>

Mailing Lists. LINGUIST provides over 100 listservs. Access: <http://linguistlist.org/lists/get-lists.html>.

Notes

1. Jung-ran Park, 2004. Language-related Open Archives: Impact on Scholarly Communities and Academic Librarianship, *E-JASL: The Electronic Journal of Academic and Special Librarianship* 5, no. 2-3, (2004). http://southernlibrarianship.icaap.org/content/v05n02/park_j01.htm; and Steven Bird and Gary Simons, Seven Dimensions of Portability for Language Documentation and Description. *Language* 79, no. 3 (2003): 557-582. *W*