A paper prepared for:

Time and Again: Theoretical and Experimental Perspectives on Formal Linguistics. Papers in honor of D. Terence Langendoen.

Prepublication draft of 8 July 2006

Linguistics as a Community Activity: The Paradox of Freedom through Standards

Gary F. Simons
SIL International

Abstract

The Internet has given us a new playing field for global collaboration. It could transform the practice of linguistics through universal access to huge quantities of digital language documentation and description. But this transformation can happen only if certain aspects of community practice are formalized by defining and adhering to shared standards. After expanding on the vision for what linguistics could be like in the twenty-first century, this essay attempts to clarify the role of standards by considering two case studies of life with and without standards—using solar time versus standard time, and using language names versus language identifiers. The essay then develops two metaphors that seek to put standards in a positive light: "linguistics as community" and "development as freedom." The ultimate conclusion is that only by submitting to the constraints of shared standards will the community be free to develop the riches of knowledge it is seeking.

The man whom we honor with this volume is well-known for his contributions to formal linguistics in the usual sense of formal approaches to the description of language. In recent years, Professor Langendoen has also given attention to formalizing linguistics

more broadly—that is, to developing formalisms that will empower the practice of linguistics in community. My collaborations with him, first in the TEI (Text Encoding Initiative, `www.tei-c.org`) and then in the E-MELD project (Electronic Metastructures for Endangered Language Data, `www.emeld.org`), have centered around a vision for how the practice of linguistics could be transformed by the universal availability of digital language documentation and description that are encoded in a standardized and interoperable way. Time and again, however, we have encountered the less-than-enthusiastic response of colleagues who are not keen on being constrained by standards.

This essay addresses the place of standards in the formalization of community practice. After expanding on the vision for what linguistics could be like in the twenty-first century, the essay attempts to clarify the role of standards by considering two case studies of life with and without standards—using solar time versus standard time, and using language names versus language identifiers. The essay then develops two metaphors that seek to put standards in a positive light: "linguistics as community" and "development as freedom." The ultimate conclusion is that only by submitting to the constraints of shared standards will the community be free to develop the riches of knowledge it is seeking.

1. Linguistics in the twenty-first century

In the last two decades, some fundamental changes in the world have had a profound impact on the way business is conducted. These same changes are likely to have an equally profound effect on the way we do linguistics. In a recent book, *The World is Flat*, Thomas Friedman explains what is happening. Five centuries ago, when Columbus put his conviction that the world was round to the test by sailing west to reach "the countries of India," he thought he had reached part of the Indies, but in fact he had run into America. Friedman (2005:3–5) recounts how in 2004 he flew east to make his own voyage of discovery to Bangalore, the "Silicon Valley" of India. When he actually got to India, he was surprised to find parts of America—billboards touting American companies, software firms using American business techniques, people using American names and American accents at large call centers. His conclusion: the world must be flat.

Freidman goes on to describe "the ten forces that have flattened the world" (2005, chapter 2). The first of these was the fall of the Berlin wall in 1989, which almost overnight removed barriers that were keeping half of the world from collaborating with the other half. The second was the emergence of the World Wide Web in the mid 1990s and the shift of focus in the personal computing platform from the desktop to the Internet.

The other eight flatteners have been new approaches to collaboration on a global scale that the first two flatteners have made possible: work flow software, open-sourcing, outsourcing, offshoring, supply-chaining, insourcing, in-forming (e.g. Google), and the wireless technologies that make it possible to collaborate while on the move.

These ten flatteners have produced a Web-powered playing field for global collaboration. The flattening did not happen as soon as the Internet became available; rather, it took a decade or so for business processes to change in order to achieve the great productivity breakthroughs made possible by the new technologies. By the time the new processes were in place, three billion people (of China, India, Russia, Eastern Europe, Latin America, and Central Asia) who had been frozen out of the playing field only two decades ago, found themselves with the potential to plug into the field and play with the rest of the world. Friedman (2005:181–182) offers the following conclusion:

> It is this triple convergence—of new players, on a new playing field, developing new habits and processes for horizontal collaboration—that I believe is the most important force shaping global economics and politics in the early twenty-first century. … The scale of the global community that is soon going to be able to participate in all sorts of discovery and innovation is something the world has simply never seen before.

The triple convergence will shape far more than economics and politics; a global community of unprecedented scale is poised to participate in linguistic discovery and innovation as well.

Friedman (2005:9–10) summarizes the history of globalization in terms of three eras. In Globalization 1.0, which began with the voyage of Columbus, a handful of countries drove global integration as they sailed the seas to establish colonial empires. In Globalization 2.0, which began in the nineteenth century and was fueled by the Industrial Revolution with its falling transportation costs, multinational companies were the main force driving global integration. Now at the turn of the twenty-first century, fueled by the technologies of the Information Age, Globalization 3.0 brings us a new dynamic force—the power of individuals to collaborate and compete on a global scale.

This new era of the flattened playing field is having its effect on the academic world. In the previous era, the institutions of learning with global reach were the world-class universities. Just a generation ago, in order to access the information riches of the world, one had to physically enter the libraries and archives of such institutions. Today, the center of gravity has shifted from institution to individual. Whether you are a professor logging in from an Ivy League campus in North America or a peasant logging in from an Internet cafe somewhere in the developing world, you have access to the same indexed

collection of billions of information resources from around the globe. For the linguist, these resources include not just the secondary works of description and the tertiary works of theory that are typically found in libraries, but also the primary data that have traditionally been found in archives, plus a new form of primary data that is cropping up as hundreds (and soon thousands) of language communities start posting materials on the web in their own language.

This is the context for doing linguistics in the twenty-first century. All linguists will have finger-tip access to materials in and about thousands of languages, not just the materials they may have personally collected for a handful of languages. Linguists, educators, and native speakers from different corners of the globe will be able to form virtual communities of collaboration around particular languages of interest.

The field of linguistics is at a crossroads where it has a choice between two possible futures. The enormous quantity of materials on the new digital playing field could either hold the promise of unprecedented access to information, or, the specter of unparalleled frustration and confusion. The outcome will depend on how we choose to act—whether we act in community in order to define and follow common standards of practice that will make it possible to index and search the wealth of materials, or, act in isolation so as to proliferate idiosyncratic practices that will make it impossible to find and compare resources. In the end, achieving the ultimate vision of unprecedented access to information will require defining and adhering to standards.

2. Life without standards

The average linguist does not immediately warm up to the idea of standards. After all, one makes a mark in the world of academics by demonstrating the uniqueness of one's contribution, not by being like every one else. Thus we face an up hill battle in convincing linguists that standards are a good idea.

In fact, standards are an indispensable aspect of every day life. When we transact business in the marketplace, we are relying on federally set standards for weights and measures and for the regulation of money. When we switch on a light in the office, we are taking advantage of standards that specify the mating of bulbs with light sockets and of plugs with wall sockets, that govern the wiring of light fixtures to switches and the source of power outside the building, and that regulate the wider power grid for moving electricity from suppliers to consumers on a regional scale. When we perform a web search with Google, we are depending on a host of standards covering issues like physical connections between machines, transmission of data over physical lines,

addressing of internet nodes, encapsulation of information into a sequence of logical packets, reliable transmission and reassembly of those packets, formats of files, encoding of characters, and more. All the standards mentioned in this paragraph are things we simply take for granted. We will know that we have arrived in the twenty-first century of linguistics when the standards needed to achieve the vision described in section 1 are also taken for granted.

It is instructive to consider an example of life without standards. In medieval times (and earlier), the time was regulated by the position of the sun in relation to the individual's position on earth. Noon was defined as the moment when the sun was directly overhead. This standard worked fine as long as wind and muscle power constrained the distance that could be traveled in a day. But the advent of train travel in the nineteenth century changed all of that. In the most populated latitudes of North America, the earth rotates at a rate of twelve and a half miles a minute (Blaise 2000:30). Thus rail passengers could journey 100 miles in a couple hours, only to find that their pocket watches were eight minutes off when they arrived.

Today it is hard for us to imagine life without standard time, but just 150 years ago in North America there were 144 official times based on local solar noon (Blaise 2000:34). The rail network grew up in this context; each railroad set and published its schedules in terms of the official time of its headquarters, rather than of the city in which the train was stopping. Thus in a station that serviced more than one railroad, it was simultaneously a different time on each road (Blaise 2000:70). Rail passengers had to travel with a big book of time conversions in order to plan their connections, and it was still easy to miscalculate and miss the train. What's worse, sharing the same tracks among railroads employing different official times could lead to disastrous results—train wrecks were a daily occurrence (Blaise 2000:72).

These problems were finally solved in 1884 when the nations of the world gathered at the Prime Meridian Conference. In addition to establishing the zero meridian at Greenwich, this conference established the International Date Line and the system of universal time with 24 standard time zones stretching around the globe. For the first time in history, it was possible to answer the question "What time is it?" with a single global answer, rather than with a myriad of local answers.

3. Examples from linguistics

In the world of linguistics, an analogous act of international standardization is in the process of happening. For millennia, the standard means of identifying languages has

been by name. This has worked well when the people communicating are members of the same local community who use the same naming conventions. But languages do not have just one unique name—the preferred name for a language may change over time, the same language may have different names in different languages, and different outsiders may refer to the same language by different names (especially before a language name becomes standardized). Nor do languages have unambiguous names—different languages (in different parts of the world) may have the same name, or a name used for a language may refer to something that is not the language in other contexts.

When we move from the context of a closed local community to the context of an open global community where billions of once isolated information resources are being brought together into a single World Wide Web, we see that using names to identify languages is like using solar time to run the railroads. If we fail to modernize that standard, the passengers of the web will miss their language trains, or worse yet, witness an information wreck of global proportions.

For instance, *Ega* is the name of a language spoken in Côte d'Ivoire, but when searching for "Ega dictionary" in Google one quickly discovers that EGA is an acronym for Enhanced Graphics Adapter and Enterprise Grid Alliance. *Santa Cruz* is the name of a language spoken in Solomon Islands, but searching for "Santa Cruz dictionary" yields descriptions of a city in California and other places around the world that bear the same name. *She* is the name of a language in China, but searching for "She dictionary" uncovers the Woman-Speak Dictionary among other things.

Joseph Grimes anticipated this kind of problem over 30 years ago when he developed the database for managing the *Ethnologue,* a comprehensive reference work listing all known languages of the world—the most recent edition identifies 6,912 living languages (Gordon 2005). As Grimes explained at the time: "Each language is given a three-letter code on the order of international airport codes. This aids in equating languages across national boundaries, where the same language may be called by different names, and in distinguishing different languages called by the same name" (Grimes 1974:i). These three-letter codes became widely known when the *Ethnologue* was published on the web and became a *de facto* standard for groups like the Open Language Archives Community that needed unique identifiers for all known languages in order to catalog their collections of language resources (Simons 2000, 2002; Bird and Simons 2003).

In 2002 a subcommittee of the International Organization for Standardization (specifically TC37/SC2) formally invited SIL International to prepare a new standard that would reconcile the complete set of codes used in the *Ethnologue* with the approximately 400 codes already in use in the earlier ISO standard for language identification. In

addition, codes developed by Linguist List to handle ancient and constructed languages were to be incorporated. The result is a standard named ISO 639-3 that provides unique and unambiguous codes for identifying nearly 7,500 languages (SIL 2006). In 2005, the standard reached the status of Final Draft International Standard. As this essay is being prepared for publication, the final revision of the standard is being submitted to the subscribing national standards bodies for their vote on full adoption.

Two other examples of information standards for linguistics deserve mention. The first regards standardizing the encoding of characters in textual data. All digital information ultimately reduces to a sequence of binary numbers. In the early days of computing, there was no standard for how to represent writing in a data stream. Each hardware manufacturer developed a different scheme for mapping letters to numbers; it soon became apparent that the resulting impediment to information interchange between computers was not in any company's best interest. The leading manufacturers in the United States therefore got together to work out a common standard and in 1963 ASCII, or the American Standard Code for Information Interchange, was adopted (Brandel 1999). That standard says, for instance, that the number 65 will be used in a digital data stream to represent a capital A, 66 to represent B, and so on. It was elevated to the status of international standard, as ISO 646, in 1972. In all, only 95 printable characters were standardized (upper- and lower-case Roman letters, digits, punctuation, and a few other symbols), but this was enough to launch ubiquitous applications like email and the World Wide Web.

ASCII solved the character encoding problem for English, but what about the other languages of the world? In the ensuing decades, dozens of other character sets were formalized through national and international standards processes. With so many standards in use, however, it was not possible to correctly interpret the numbers into characters without knowing which standard had been followed to encode the information. With the advent of personal computers, linguists jumped on the bandwagon (e.g. Simons 1989) and developed clever system-specific solutions for building their own fonts that would redefine selected characters to meet the requirements of the language they were studying. However, the inevitable train wreck occurred whenever a linguist tried to share data with someone who used a different system, and even worse, when the original creator upgraded to a new system on which the special font solution no longer worked. Linguists began to learn the hard way that one must follow a common standard in order to ensure not only the interchange of information in the present, but also its survival far into the future (Simons 2006). Fortunately, the same problem was plaguing the software industry, so in the late 1980s a group of leading software companies got together to

define Unicode—a single standard for encoding the characters in all the major writing systems of the world. The first version was published in 1991 and by the time the next version appeared in 1993 it had attained the status of international standard (as ISO 10646). The current version standardizes the encoding of nearly one hundred thousand characters (Unicode Consortium 2003). Work is on-going to add to the inventory of writing systems and characters that are covered.

The final example I want to mention, GOLD (General Ontology for Linguistic Description), is the brainchild of Langendoen and two of his students (Farrar, Lewis, and Langendoen 2002; Farrar and Langendoen 2003). Inspired by the Semantic Web activity of the World Wide Web Consortium (Berners-Lee and others 2001; Hendler 2003), GOLD seeks to provide a formal description of the concepts (and relationships between concepts) that exist within the problem domain of linguistic description. By mapping the terms and markup vocabularies used in actual linguistic descriptions to their nearest equivalents in the linguistic ontology, it becomes possible to support smart searching for linguistic concepts across a large collection of descriptive materials (Langendoen and others 2002; Simons and others 2004). Each concept as defined in GOLD is like a standardized time zone—the manifestation of a concept in any given language is like the idiosyncratic solar noon of a particular locality, but by overlooking the slight differences in order to capture the greater similarities it becomes possible for data from disparate sources to operate in a coordinated fashion over the same information processing tracks. Though GOLD (at `www.linguistics-ontology.org`) is still in its infancy, it holds tremendous promise as a foundational part of the infrastructure for twenty-first century linguistics—but only if it can become firmly grounded in the community.

4. Metaphors to live by

The general reaction to the notion of standards within the linguistics community is not particularly enthusiastic. Typical reactions run the gamut from wariness to skepticism to hostility. The idea of conforming to standards, it would seem, is counter to the ideal of academic freedom. What can be done to change this perception?

In pondering this problem, I've been inspired by George Lakoff's work on metaphor (Lakoff and Johnson 1980) and its application to contemporary politics (Lakoff 1995, 1996). In the latter work, Lakoff points out that conservatives (with their "strict father" model) have gotten the upper hand over liberals (with their "nurturant parent" model) through the deft use of metaphor in the public discourse. How might metaphor be used to

inform the discourse about standards among linguists? What metaphors could linguists live by that would cast standards in a nurturant light rather than a strict one?

I've identified two. The first is "linguistics as community;" it highlights the role of standards in allowing the body of linguists to act in community. The second, "development as freedom," highlights the role of the shared constraints embodied in standards in freeing the community to develop the riches of knowledge it is seeking. These ideas are developed in the following two sections.

5. Linguistics as community

Science is "an inescapably cooperative, social activity"—so concludes chemist Henry H. Bauer (1992:52) in *Scientific Literacy and the Myth of the Scientific Method.* It is generally believed that the essence of science is the "scientific method," which involves making systematic observations, formulating hypotheses to explain what has been observed, and then validating or invalidating those hypotheses through further observation. Bauer observes that although the method has a place in science, it does not explain how science really works (Bauer 1992:19–41). The subfields of science differ in the extent to which they are data-driven or theory-driven, data-rich or data-poor, experimental or observational, quantitative or qualitative, and so on. They are consequently characterized by different methodologies.

Science is best understood, not in terms of a method, but in terms of social activity. Bauer argues that the unity of science derives from a shared commitment to certain kinds of cooperative action. In a chapter named "How science really works," he uses two metaphors to explain the cooperative activities that lie at the heart of science. These metaphors help to shed light on linguistics as a community activity and how standards fit into the picture.

The first metaphor is the jigsaw puzzle (Bauer 1992:42–44; after Polanyi 1962). Doing science is like putting together a large jigsaw puzzle with a group of people. One strategy might be to give each puzzler an equal share of the pieces and have them work independently, but this approach would be doomed to failure since few of the pieces given to any one puzzler would actually fit together. Alternatively, one could give each puzzler a copy of all the pieces, and eventually combine their separate, partial results. This approach at least has a chance of completing the puzzle, but would lead to large-scale duplication of effort and may not be significantly faster than a single puzzler working in ideal conditions.

The only effective way to put multiple puzzlers to work at once is to have them all work together on a single copy of the puzzle in sight of each other. In this way, as one puzzler fits in one more piece, all the others will see the resulting state of the puzzle and will potentially adjust their next step in response to the new state. These puzzlers take individual initiative in determining what part of the puzzle they will work on and how they will go about doing it. The result of the adjustment and self-coordination that occurs as they observe the outcomes achieved by their fellows is a joint achievement that could not be equaled by any single puzzler working in isolation.

When we apply this metaphor to the puzzle of human language, we see that it is a huge puzzle indeed—in fact, the number of pieces is unbounded (Langendoen and Postal 1984). There are tens of thousands of people around the world who are working professionally on this puzzle; countless others are also interested in observing and sometimes contributing. The Internet offers the potential—for the first time in history—for all of the puzzlers to see all the available pieces and to watch each other as they fit them together. The challenge is for the community to align its practices so as to make that potential into a reality, and standards are a key part of achieving such alignment. A standard like Unicode (for character encoding) is needed so that every puzzler who looks at a particular piece sees the same thing. A standard like ISO 639-3 (for language identification) is needed so that the puzzlers interested in a particular language can locate all the relevant pieces. A standard like GOLD (for identifying the concepts used in linguistic description) is needed so that the puzzlers interested in a particular linguistic phenomenon can locate all the relevant pieces. Without such standards, practice cannot be community-based, but will resemble the scenarios in which individual puzzlers work separately or in small groups.

The second metaphor is that of the filter (Bauer 1992:44–48). The process by which a scientific community generates scientific knowledge is like a process of putting ideas through a multistage filter. Each stage is manifest by social institutions that the scientific community has developed over time. The first stage is undergraduate and graduate training in which aspiring scientists learn to align their thinking and behavior with the norms of a particular scientific community. The second stage is research, or frontier science, in which a vast array of ideas get formed and tested. But the array of ideas is not unconstrained; the institution of grant funding (with its attendant mechanisms of proposal writing and peer review) serve to limit those ideas that actually get worked on. The results of research cannot contribute toward scientific knowledge until they enter the third stage, namely, the primary literature. Here the key institutions are conferences and journals and the primary mechanisms are peer review and editing. The review process

serves not only to filter out poor work but also to ensure that authors frame their ideas in light of established knowledge.

The fact that an idea is published does not make it scientific knowledge; it just makes it widely available. The next stage in the filter is for published ideas to be tested and used by others. In the process some ideas will ultimately be rejected by the community; others will be refined and extended and eventually make their way into the secondary literature of review articles, monographs, and graduate-level textbooks. After the passage of even more time and testing, we reach the final stage of textbook science in which the consensus of the community gets expressed as scientific knowledge in undergraduate textbooks. But even textbook knowledge is never completely correct, and it slowly changes over time as newer discoveries eventually make their way to the bottom of the knowledge filter.

The filter metaphor helps us to understand the rightful place of standards like ISO 639-3 and GOLD within the linguistics community. They are meant be a formalization of current textbook knowledge. As Langendoen (personal communication) is fond of saying with respect to GOLD, "We aren't trying to make the search engine into an expert linguist; we are just trying to bring it up to the level of an undergraduate student in linguistics." This means that researchers working on the frontier of a particular issue would always see something in a relevant standard that they would want to change, while the parts of the standard in which they are not expert would look okay to them. Standards will never align fully with frontier knowledge. They are necessarily slow to change, evolving only as knowledge works its way through the multistage filter to become the shared consensus of the community. In the meantime, even if a standard proves to be incorrect on a particular point, it has served its purpose in enabling the global community to organize and share its knowledge in a consistent and repeatable way.

6. Development as freedom

Freedom in the pursuit of knowledge is a high ideal within the halls of academe. As a doctrine of higher education, academic freedom is defined more narrowly as "the ability to teach, research, and write without fear of repercussion because the subject or conclusions are considered unacceptable by other faculty, administration, community organizations, state or local governments or religious groups" (Kant 2004). Nevertheless, the doctrine may at times be invoked more broadly to embody the notion of intellectual independence, which can in turn lie behind the eschewal of standards that would constrain one's approach to research and writing. But, in fact, such standards are

commonplace in academics—each institution involved in the knowledge filter (including every funding agency, journal, and book publisher) imposes standards that define its operation and thereby constrain participation.

Every community requires constraints on the behavior of its members. While the ability to do whatever one wishes may seem like the ultimate of freedom, in fact the resulting anarchy leads to anything but freedom in the end. As Jean-Jacques Rousseau, the great social theorist of the eighteenth century, concluded: "The mere impulse of appetite is slavery, while obedience to a law which we prescribe to ourselves is liberty" (Rousseau 1762, chapter 8).

This paradox of freedom through constraint should make sense to linguists, of all people, since this principle lies at the very heart of language. What is it that empowers the free interchange of ideas between two people who speak the same language? It is mutual conformance to a shared system of constraints. In any particular language, all the vocal sounds that are humanly possible are constrained to just a few score that are used in the languages. Sound patterns constrain the manner in which sounds combine to form the sequences that are actually possible in the language. Lexical associations of form to meaning constrain which minimal sound sequences actually mean something. Rules of grammar constrain the manner in which those minimal sequences combine to express larger thoughts. In fact, entire theories of language, such as Optimality Theory (Archangeli and Langendoen 1997), account for the surface forms in language in terms of constraint satisfaction. In the final analysis, it is the use of a shared system of constraints (or, a standard) that gives a speech community the freedom to communicate.

A shared system of constraints is what the linguistics community needs in order to develop the freedom it is ultimately looking for. The metaphor of "development as freedom" comes from the book of that title by the Nobel-prize winning economist Amartya Sen (1999). Whereas economists typically focus on issues like GNP growth or industrialization, Sen has distinguished himself by focusing on the social aspects of economic development. He argues that social development (including enhanced literacy, better health care, the empowerment of women, and the free flow of information) is the principal means of economic development. The expansion of such freedoms should also be the primary end of development, and as that end is achieved, it becomes the means for even further development.

What is the analogue of economic development for the linguistics community? The wealth of the community is its information. As linguists, we want to build an ever growing body of primary data encompassing every known human language. We want to analyze that documentary material to develop an ever growing body of secondary,

descriptive data (Himmelmann 1998). We then want to mine that body of data about individual languages to build an ever growing body of frontier knowledge about the workings of language in general, and ultimately to refine that body of frontier knowledge to build an ever growing body of textbook knowledge on language and linguistics. Maximal freedom and wealth for a linguist would be to have all such known data and knowledge at one's fingertips in a form that could be queried and compared and repurposed.

Another concept that is central to Sen's work is that of individual agency—the ability to set and pursue one's own goals and interests. He argues that "development consists of the removal of various types of unfreedoms that leave people with little choice and little opportunity of exercising their reasoned agency" (Sen 1999:xii). If the ultimate goal of linguists is to participate in putting together the enormous puzzle of human language, and if the optimal approach is to see all the pieces and the progress being made by all other linguists, then what are the unfreedoms that prevent linguists from exercising their agency within this enterprise? For some it is still a lack of access to the web infrastructure for global collaboration. This is a problem whose solution lies outside the scope of the linguistics community; fortunately, the world around us is tackling this problem. For the linguists who do have access to the web infrastructure, the unfreedoms are lack of access to information at all (because it is not in a digital form), inability to find relevant digital information (because it is not catalogued in a standardized way), and inability to compare information across languages (because it is not formatted in a standardized way). The antidote to these unfreedoms is the development of standards within the community—shared constraints that will empower individuals to both contribute to and benefit from the information riches of the community.

7. Conclusion

Following community-wide standards involves good news and bad news. The bad news is that there will be times when the details of a standard lag behind the best current knowledge and will thus prevent linguists from being able to express what they really want to record in their digital data. However, the good news is that when a plethora of idiosyncratic practices gives way to one universally-shared practice, linguists will gain the freedom to access the global riches of information about languages and linguistics. It is the promise of this future that is motivating many linguists to participate in the process of developing and refining the standards that will empower us to act in community. When that infrastructure is in place, a new breed of twenty-first century linguists using new

processes to collaborate with new players on a new global playing field will be able to achieve things we can only dream about today.

References

Archangeli, Diana and D. Terence Langendoen. 1997. *Optimality Theory: An overview.* Oxford: Blackwell.

Bauer, Henry H. 1992. *Scientific literacy and the myth of the scientific method.* Urbana and Chicago: University of Illinois Press.

Berners-Lee, Tin, James Hendler, and Ora Lassila. 2001. The Semantic Web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American* 284(5):34–43, May 2001. Online: http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2

Bird, Steven and Gary F. Simons. 2003. Extending Dublin Core metadata to support the description and discovery of language resources. *Computers and the Humanities* 37:375-388. Online preprint: http://arxiv.org/abs/cs.CL/0308022

Blaise, Clark. 2000. *Time lord: Sir Sandford Fleming and the creation of standard time.* New York: Pantheon Books.

Brandel, Mary. 1999. 1963: ASCII debuts. *Computerworld,* 12 April 1999. Online: http://web.archive.org/web/*/http://www.computerworld.com/news/1999/story/0,11280,35241,00.html

Farrar, Scott and D. Terence Langendoen. 2003. A linguistic ontology for the semantic web. *GLOT International* 7(3).97-100. Online: http://emeld.org/documents/GLOT-LinguisticOntology.pdf

Farrar, Scott, William D. Lewis, and D. Terence Langendoen. 2002. An ontology for linguistic annotation. *Semantic Web Meets Language Resources: Papers from the AAAI Workshop (Technical Report WS-02-16),* pages 11-19. Menlo Park, CA: AAAI Press. Online preprint: http://emeld.org/documents/AAAI-OntologyLinguisticAnnotation.pdf

Friedman, Thomas L. 2005. *The world is flat: A brief history of the twenty-first century*. New York: Farrar, Straus, and Giroux.

Gordon, Raymond G., Jr. (ed.). 2005. *Ethnologue: Languages of the world, Fifteenth edition.* Dallas: SIL International. Online version: http://www.ethnologue.com/

Grimes, Joseph E. 1974. *Word lists and languages*. Technical Report No. 2, Department of Modern Languages and Linguistics, Cornell University, Ithaca, NY.

Hendler, James. 2003. Science and the Semantic Web. *Science,* 24 January 2003, pages 520–521. Online: http://www.sciencemag.org/cgi/content/full/299/5606/520?ijkey=1BUgJQXW4nU7Q&keytype=ref&siteid=sci

Himmelmann, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36:165–191. Online preprint: http://corpus.linguistics.berkeley.edu/~ling240/himmelmann.pdf

Kant, Candace. 2004. AAUP and academic freedom: a history. *The Alliance,* December 2004, page 1. Nevada Faculty Alliance. Online: http://www.aaup.org/Com-a/Kantart.htm

Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, George. 1995. Metaphor, morality, and politics, Or, Why conservatives have left liberals in the dust. In *Webster's World of Cultural Democracy*. Seattle: The World Wide Web center of The Institute for Cultural Democracy. Online: http://www.wwcd.org/issues/Lakoff.html

Lakoff, George. 1996. *Moral politics: What conservatives know that liberals don't.* Chicago: University of Chicago Press.

Langendoen, D. Terence and Paul M. Postal. 1984. *The vastness of natural languages.* Oxford: Basil Blackwell.

Langendoen, D. Terence, Scott Farrar, and William D. Lewis. 2002. Bridging the markup gap: Smart search engines for language researchers. *Proceedings of the Workshop on Resources and Tools for Field Linguistics, Third International Conference on Language Resources and Evaluation (26-27 May 2002, Las Palmas, Canary Islands, Spain),* pages 24-1 to 24-9.

Polanyi, Michael. 1962. The republic of science: its political and economic theory. *Minerva* 1:54-73.

Rousseau, Jean-Jacques. 1762. *The social contract, or, principles of political right.* Translated by G. D. H. Cole. Web edition by the Constitution Society. Online: http://www.constitution.org/jjr/socon.htm

Sen, Amartya. 1999. *Development as freedom.* New York: Anchor Books.

SIL. 2006. *ISO/DIS 639-3* [the official web site of the ISO 639-3 Registration Authority]. Dallas: SIL International. Online: http://www.sil.org/iso639-3/

Simons, Gary F. 1989. Working with special characters. In Priscilla M. Kew and Gary F. Simons (eds.), *Laptop publishing for the field linguist: an approach based on Microsoft Word,* pages 103–118. Dallas: Summer Institute of Linguistics.

Simons, Gary F. 2000. Language identification in metadata descriptions of language archive holdings. *Proceedings of the Workshop on Web-based Language Documentation and Description (12-15 December 2002, Philadelphia, PA),* pages 274–282. Online: http://www.ldc.upenn.edu/exploration/expl2000/papers/simons/simons.htm

Simons, Gary F. 2002. SIL three-letter codes for identifying languages: Migrating from in-house standard to community standard. *Proceedings of the Workshop on Resources in Tools and Field Linguistic, Third International Conference on Language Resources and Evaluation (26-27 May 2002, Las Palmas, Canary Islands, Spain),* pages 22-1 to 22-8. Online preprint: http://www.sil.org/~simonsg/preprint/SIL language codes.pdf

Simons, Gary F. 2006. Ensuring that digital data last: The priority of archival form over working form and presentation form. *SIL Electronic Working Papers* 2006-003. Online: http://www.sil.org/silewp/abstract.asp?ref=2006-003

Simons, Gary F., William D. Lewis, Scott O. Farrar, D. Terence Langendoen, Brian Fitzsimons, and Hector Gonzalez. 2004. The semantics of markup: Mapping legacy markup schemas to a common semantics. In Graham Wilcock, Nancy Ide, and Laurent Romary (eds.), *Proceedings of the 4th workshop on NLP and XML (NLPXML-2004),* pages 25-32. Association for Computational Linguistics. Online preprint: http://emeld.org/documents/SOMFinal1col.pdf

Unicode Consortium. 2003. *The Unicode Standard, Version 4.0.* Boston, MA: Addison-Wesley. Online version: http://www.unicode.org/versions/Unicode4.0.0/