Five Dimensions of Collaboration: Toward a Critical Theory of Coordination and Interoperability in Language Documentation

Akiemi Glenn

University of Hawai‘i at Mānoa

In the literature on best practices of language documentation, “collaboration” has emerged as an important concept. While collaboration between scholars is not usually the norm in linguistics, a theory of language documentation must grapple with its theoretical orientation to collaboration. By reviewing the practices of researchers in other disciplines, this paper identifies five aspects of academic collaboration—coordination, distribution of labor, standards for interoperation, authorship and authority, and feedback—that have special bearing on the enterprise of language documentation. I investigate these as a starting point for linguists and our collaborators to consider critically what collaboration means for a documentation project and for the discipline of linguistics.

1. INTRODUCTION. Documentary linguistics is engaged in an evolving discourse on best practices, methods, and ethics in an effort to establish a theory of language documentation. A developing consensus in the literature is that a language documentation should be comprehensive, data-centered, and usable by diverse audiences of academics, speech community members, and others. Writers have argued that comprehensiveness requires an orientation to many issues that are beyond the standard training and expertise of most linguists, and that any successful documentation project will draw on the multiple strengths of interdisciplinary collaboration as well as collaboration with the speech community under study. As a result, “collaboration” has surfaced as a key term in the literature on best practices in language documentation and is used to describe both the interactions between scholars and between researchers and communities.

Possibly because of past neglect of the issue in linguistics and academia more broadly, the overwhelming orientation of literature addressing collaboration is concerned with the practical and ethical dimensions of such work with communities. (See Yamada 2007; Warner, Luna, and Butler 2007; Dwyer 2006; and many others.) This inclination is welcomed and most needed, but much less discussion has been devoted to the notion of collaboration between scholars engaged in language documentation enterprises. At the moment, collaboration is not for the most part the norm in linguistics, and for linguists coming to

\[\text{1 This paper owes much to Bird and Simons’s (2003) “Seven dimensions of portability for language documentation and description” in both its title and spirit as a conversation piece for a coalescent discipline.}\]
a documentation project, the methods of effective collaboration may be difficult to identify. A theory of language documentation must also grapple with its theoretical orientation to collaboration. While the nature of collaboration between researchers and communities is best determined by the situations and individual cultures involved, there should be a theoretical space within the discipline of language documentation to begin to explore best practices for researchers working together. Despite our individual interests and personal backgrounds, researchers share the culture of academia, and may use that to identify ways to work together to enrich our research and to produce more meaningful documentations for future generations of researchers and language community members. In turn, more effective interdisciplinary collaboration will likely facilitate more effective researcher-community collaboration and richer products of documentation.

My aim here is to engage some key trends in collaboration between researchers in the sciences and humanities in order to identify five aspects of academic collaboration that have special bearing on the enterprise of language documentation. This paper will also investigate these five dimensions as a basis for linguists and our collaborators to consider critically what collaboration means for a language documentation project and for the discipline of linguistics. By presenting these five dimensions, I will suggest that a true theory of coordination of our efforts and interoperatability of the products of our documentation will require linguists to reorient our assumptions about the value of “lone-wolf” linguistics and the power of the relationships between researchers. I will also emphasize the advantage in engaging and perhaps altering the social and cultural underpinnings of linguistics as a discipline.

2. COLLABORATION IN THE ARTS AND SCIENCES. Precedents for collaboration in the sciences are well known. Physicists, biologists, and engineers often work as and are funded as large research teams. As a result, it is not uncommon for significant publications in those disciplines to be multi-authored. Funding agencies expect and encourage researchers to work in collaborative teams both within and across disciplines. Major funding entities like the National Science Foundation in the United States actively solicit collaborative proposals (Gaughan and Bozeman 2002), and have commissioned studies of the most effective practices and key benefits of these kinds of relationships between investigators (Cummings and Kiesler 2005). Often grants in the physical sciences allow principal investigators to hire graduate student research assistants, who then have the opportunity to work cooperatively and can perpetuate collaborative research when they have students of their own (Thagard 1997). Researchers themselves appear to conceptualize their collaboration in terms of the higher quality products of such research (Melin 2000). In the sciences, co-authored publications appear to have more of an impact on their respective disciplines when viewed by citation frequencies than single-authored papers (Narin and Withlow 1990), and in many disciplines collaboration appears to increase the productivity of individual researchers (Lee and Bozeman 2005).

While collaboration is more highly valued in the sciences than in the humanities, negotiations of research collaboration in the sciences are not always easy, and the relationships that develop in such contexts are not necessarily non-hierarchical, or unfraught with personal politics. In fact, a significant portion of writing on collaboration in scientific fields
identifies problems with the practice and highlights the difficulties of drawing lab workers into sharing materials or results, or of goading innovators into sharing their methods (Cohen 1995, Locke 1999). The nature of the sciences and their objects of study facilitate the communal approach to problem solving. The emphasis on collecting empirical data makes division of labor possible, and the emphasis on laboratory work facilitates the re-creation of procedures in locations separated by great distance. Modern sciences involve a great deal of networking, collaboration, and sharing of knowledge, facilitated by technologies such as the Internet. Those who investigate how knowledge is produced in the sciences have thus concluded that science is significantly a form of social practice (Hine 2005). As a result of these social practices, participants develop relationships that are transitive (Newman 2001, Bird et al. 2009). That is, they facilitate future collaborations with others within a network of collaborators.

The humanities, by contrast, demonstrate less robust or institutionalized traditions of collaboration and have developed alternative methods of rewarding, and in some cases punishing, collaborative work. For linguistics and language documentation these divergent views of collaboration can be especially challenging to navigate, as our broad discipline utilizes methodologies and theoretical stances that make it difficult to classify linguistics definitively among the sciences or among the humanities.

2.1 COLLABORATION IN THE COMPUTER AND INFORMATION SCIENCES. Language documentation’s emphasis on producing, managing, and archiving data has required some forays into the realms of computer and information science, and while models for collaborative teams in physics, for example, might be useful to review, these disciplines provide more immediately useful contrasts to our own field.

The field of software and Internet development is highly associated, even in popular culture, with dispersed collaborative networks. Famously, the open-source LINUX operating system was developed by a distributed network of programmers and stakeholders whose diverse expertise created a vast network of testers and problem solvers always ready to tackle new bugs or klutzy frameworks to the immediate benefit of the operating system (Fanderclai 2004). And, as Web 2.0 has demonstrated, there is an increasing orientation to “distributed cognition” as a facilitator of knowledge production and collection, evidenced in phenomena like Wikipedia (www.wikipedia.org) and Project Gutenberg’s online collection of copyright-free e-books (www.gutenberg.org/).

Library and information science more generally reflects this trend, especially mediated by digital networks. MIT’s D-space initiative is one exemplar of this new orientation as it seeks to connect research institutions and individual researchers through portals to open-access archives of research products. Both software development and information networks facilitate “bottom-up” approaches (David 2004) to knowledge sharing that privilege the needs of the producers and consumers of the products.

2.2 COLLABORATION IN THE HUMANITIES. The culture of the humanities, by contrast, has traditionally been oriented away from the kinds of collaboration that are so common in the physical and computer sciences. Research in the humanities has a tendency to glorify work produced by individuals and “great minds.” In large measure, establishing
a career in a given field has meant developing a repertoire of personal intellectual property (Fanderclai 2004).

Research in the humanities is also valued differently from physical science research across disciplines and in popular culture. Where researchers in the physical or information sciences are often viewed as contributors to larger society through discoveries with medical applications, Internet or software innovations, or in pop-science, the work of the humanities is often viewed outside of academia as lacking in urgency, real-world consequences, or applications (Davidson and Goldberg 2004).

The result is that while sciences are oriented toward problem-centered collaboration to produce rapid and comprehensive results and have the incentive to “get things right,” researchers in the humanities often do not feel accountable for their research or results beyond the confines of their, often niche, disciplines. There may be reasons why non-academics can more often name eminent physicists and geneticists than influential researchers in the humanities, let alone seemingly esoteric disciplines like linguistics. One of these reasons may be that the sciences have benefitted from eloquent spokespeople who are able to translate the applicability of research findings into terms that are accessible for more general audiences. Our most famous modern linguist, Noam Chomsky, is well known in popular culture, but not for his contributions to linguistic theory.

Objects of inquiry in the humanities have not traditionally been particularly suited to the kinds of division of labor that are found in science collaborations or to the kinds of time-consuming data collection that could be conducted by student researchers (Thagard 1997, Melin 2000, Soller 2004). Though some funding agencies in the humanities have in recent years sought to support interdisciplinary teams, institutions still devalue co-authored works in tenure review and evaluation of researchers, thereby reinforcing the notion that individually authored work is more important (Jackson 2004). In situations in which humanists engage in collaboration, for lead researchers the point is frequently to get the work done while giving novice collaborators experience in a new area of the discipline (Fanderclai 2004), not to make use of junior collaborators’ new skills or innovative techniques. As a result, the orientation away from collaboration is calcified through the generations of scholars: young researchers never worked collaboratively with advisors and do not collaborate with their advisees (Thagard 1997).

The discipline of linguistics in some ways seems to reflect the split between the sciences and the humanities, taking on some cultural aspects of the sciences, with their emphasis on primary data, and some cultural aspects of the humanities, with their qualitative methodologies. In subdisciplines like psycholinguistics and cognitive linguistics that work closely with other scientific disciplines, multi-authored works and collaborative research are common, while in more humanist enterprises such as theory-centered phonology or syntax, single authoring and research are the norm. But with the emergence of a subdiscipline like documentary linguistics, whose ostensible goal is the thorough recording and presentation of the structures and linguistic practices of speech communities, especially endangered ones, the need for collaboration and the benefits to the methodology and products derived from it are striking and merit considerable attention.

3. COLLABORATION IN LANGUAGE DOCUMENTATION. Himmelmann (1998) has argued that the data-driven orientation of what he terms “documentary linguistics” estab-
lishes that a language documentation should be of use to a diverse array of potentially interested parties. To achieve this, in principle, documenters should be fluent speakers of the language and know the linguistic and cultural practices of the community under study, as well as have a depth of knowledge of linguistic theory and broad range of methods for data collection and analysis. Himmelmann concedes that the wish list is a tall order to ask of one researcher: “These demands will only rarely be met by a single individual. Hence, the compilation of a high-quality language documentation generally requires interdisciplinary cooperation as well as close cooperation with members of the speech community” (1998:171). Himmelmann (2006) lists interdisciplinary collaboration as one of the important elements of documentary linguistics that distinguish it from the larger discipline of linguistics.

Funding initiatives like the VolkswagenStiftung’s Dokumentation Bedrohter Sprachen project (DoBeS) have emphasized interdisciplinary collaboration as they have supported language documentation projects (http://www.mpi.nl/DOBES/). DoBeS projects also require that at least one of the researchers on the funded team is a German citizen, which often means that the resulting collaboration is international. Based on her experience in working on the DoBeS-funded Kuikuro project, Bruna Franchetto (2006) suggests that interdisciplinary collaboration is the most effective way to address the needs of the variety of people who will have uses for a language documentation. Such a collaboration, in her view, must produce well-structured lexical and textual databases that include detailed theme-specific ethnographic information as part of the meta-data and exploit digital links among components, subcomponents, and their contents. She suggests that attention to native metalinguistic concepts and cross-disciplinary issues may additionally provide crucial data for understanding the social, cultural, and linguistic systems of a community.

Many language documentation theorists likewise welcome the demise of the “lone-wolf” or “lone-ranger” linguist archetype in favor of more comprehensive documentations carried out by teams (Austin 2007). Others see the “lone-ranger” as a relic of a colonial past, harkening back to the days of the intrepid explorer in the service of Empire, or at best butterfly collectors who take from the communities they study and have no ethical imperative to give anything back (Dwyer 2006).

The usefulness of collaboration is evident in the physical sciences and in documentary linguistics, but institutional barriers for the latter preclude many potentially rewarding collaborations. Primarily, the institutional culture of linguistics, with its foundation in the humanities, frequently makes collaboration an afterthought. As documentary linguistics is further conceptualized as a subdiscipline of linguistics, it will have to develop its own view of collaboration in a way that addresses the data-driven needs of the enterprise. I suggest that by looking to discussions of collaboration in the sciences and humanities, documentary linguistics might be better able to identify several key issues around which to organize a discussion of best practices, ethics, and the value of collaboration that may also facilitate a re-thinking of how collaboration is viewed in linguistics more generally.

4. FIVE DIMENSIONS OF COLLABORATION IN LANGUAGE DOCUMENTATION. By focusing on the precedents of collaboration in the sciences, I do not intend to suggest that those collaborative structures are non-hierarchical, apolitical, or always mutually satisfying for participants. In reviewing the practices of collaboration in other disciplines, there
are five recurring aspects of collaboration, however, that would seem to pertain to developing and executing language documentation projects. These five interdependent dimensions are not the only considerations in negotiating a cooperative documentation project, but they might serve as orientation points to facilitate further discussions and to help conceptualize the benefits that collaboration may have for our emerging discipline. While there is obviously a need to continue to develop best practices for collaboration between researchers in language documentation, that is beyond the scope of the present paper. Instead the discussion is meant to raise questions for the discipline as language documenters consider the value and dynamics of working together.

The five dimensions of academic collaboration presented here—coordination, distribution of labor, standards for interoperation, authorship and authority, and feedback—are meant to interact in various ways with the five phases of a documentation project: planning, fieldwork, analysis, archiving, and dissemination of end products. The dimensions are not presented to correspond with particular stages, though some will be more salient in certain parts of a documentation project than in others. For example, attention to coordination, distribution of labor, and standards will be particularly relevant to the planning phase. Coordination and issues of authorship, though likely negotiated in the planning, will also be important in the dissemination of final products. Feedback from collaborators may be on-going, as some aspect of each of the dimensions will necessarily be involved in all stages. By focusing attention on these aspects of academic collaboration in the conceptualization and execution of documentation projects, documentary linguistics has the potential not only to enrich the data we produce, but also to reinvigorate the set of social practices of our science to make our research more effective and meaningful on multiple fronts.

1. Coordination.

Peter Austin (2007) has argued that project management is one of the core skills that linguists require to carry out language documentation, as they presumably manage native speaker consultants in the field and coordinate the work of participants in different disciplines. Successful projects that rely on distributed cognition and the expertise of diverse stakeholders also require a central hub, a person or institution who possesses the knowledge and skills to be able to synthesize and organize the variety of input. The larger and more comprehensive the project, the more free-form collaboration becomes unwieldy. Fanderclai (2004:315) says that the coordinator must be “driven by his or her fascination with the subject matter to follow through and [must] know how to attract interested and talented people to the project.” A central coordinator allows for a clarity of research goals and efforts and a modicum of objectivity in assessing whether goals are being met with the methods at hand. A coordinator also serves to clarify the nature of the particular collaboration: “there has to be a distinction between fruitful chats over coffee and systematized collaboration with publications as one result” (Melin 2000:33). The role of the coordinator takes on added significance when considering the distribution of labor and how distributed work may be managed.

2. Distribution of labor.

Problem-based approaches to collaboration, as encountered in the physical sciences, provide several models of collaborative relationships. Coordination also allows the divi-
sion of labor to be made explicit through the management of virtual workspaces. Cablitz, Ringersma, and Kemp-Snijders (2007) outline how their DoBeS-funded project to create a multi-media lexicon of Marquesan allowed them to create the LEXUS program both to manage lexical data in a relational database and to distribute responsibilities through the medium of the program’s virtual workspace, thus allowing researchers to create a more complex and comprehensive project than any one of them would have been able to create alone.

Thagard (1997) identifies four genres of collaboration that make use of different methods of labor division: employer/employee, teacher/apprentice, peer-similar (working within the same discipline) and peer-different (working interdisciplinarily). The employer/employee relationship between a primary researcher and a student assistant might also mirror that of linguist and native speaker consultant, whereby the collaborators have unequal investments in the outcome of the research, while the teacher/apprentice relationship facilitates the training of junior researchers in cooperative research. Peer collaboration necessitates a different kind of negotiation in labor distribution. For peer-different collaborations, participants likely will contribute to the labor pool what their training, expertise, and interests dictate. Peer-similar collaborators may likewise have different strengths and expertise, but the main point is that for a collaboration to be successful and equitable participants must have a clear understanding of the work that is expected of them and their co-participants.

3. Standards for interoperation.

Simons (2007) defines interoperability as “the ability for two or more systems to exchange information or services and for each to make satisfactory use of what is exchanged.” Bird and Simons (2003) presented portability, the ability of different software applications and data resources to interact, as a central issue in language documentation. Groups like the Open Language Archives Community, or OLAC (http://www.language-archives.org/), rely on the diverse expertise and interests of their collaborators to develop standards for language resources. GOLD, the General Ontology for Linguistics Description (http://www.linguistics-ontology.org), is another community that works toward developing standards for language resources. Standards are necessary because they facilitate the discoverability of resources and the interoperability of systems. Data that are not structured according to conventions are destined to become usable only by those who have knowledge of the “local” conventions of the linguist, transcriber, or other entity that developed them. A move toward a more collaborative model of language documentation will underscore the necessity of standards in data sharing and formatting.

While Simons and Bird’s emphasis is on standards and the interoperability and portability of data, there is a second facet of interoperability, that of interoperability of the humans working together. Whether OLAC or GOLD, or both, develop good sets of standards and best practices for archiving the interoperability of the language resources in question is ultimately determined by the human actors that create and interact with them and the choices humans make about whether to adhere to one set of standards over another or whether to engage their data in such a way at all. Some may choose to ignore a set of proposed standards for personal reasons: enmity against the creators, or simply the belief that one can do it better. For collaborators working across disciplines, a particular amount of
cross-discipline education may need to be taken into account so that each party has enough information about proposed standards, or divisions of labor, so they are comfortable working for and contributing to the collective good of the project. Interoperability is in a certain measure, then, a human problem and something that must be acknowledged and attended to in developing an effective collaboration.

4. Authorship and authority.

Some lingering cultural barriers to collaboration in the humanities are the issues of authorship and authority that are bound quite closely to notions of the single expert, a solitary scholar whose individual enterprise makes them the authority on the topic at hand. As a result, the order of names on a co-authored project is a serious issue, and may code seniority, expertise, and the amount of work put into the collaboration by each participant. In some collaborative relationships, participants are encouraged to reuse data for projects and publications on their own (Fanderclai 2004:312). Melin notes that since we live in a linear world, even if the authors did the same amount of work for a given co-authored paper, someone has to come first. “[T]he way that authors set their names on articles may not always reflect the real situation [of authorship] which can lead to miscalculations” (Melin 2000:32).

The order of names on a publication may seem to be a small issue, but it points to larger problems in the way that that intellectual property is regarded in academia and in the humanities especially. Inspired by the open-source mores of software and Internet development communities, researchers in the humanities and sciences are rethinking the nature of copyright and intellectual property, and slowly but significantly, open-access publications are spreading in academia. The journal Language Documentation & Conservation, now in its third year, makes all of its content available through licenses provided by Creative Commons (http://creativecommons.org), an international network of stakeholders who work to make content more easily accessible and usable outside the strictures of traditional copyright. For the journal’s founders, the decision to make content available through Creative Commons licenses was just as logistical as it was principled. In its first issue linguist and legal scholar Paul Newman (2007) noted that attention to issues of communication is just as important to the enterprise of language documentation as fieldwork. Issues about authorship and authority are in flux as stakeholders in documentation projects work and rework their relationships against the backdrop of the academy’s own negotiations about intellectual property. Still, these issues are important to consider in embarking on any collaboration.

2 Some institutions and journals do advocate for the alphabetical ordering of collaborators’ surnames. This practice, however, may have unanticipated outcomes in some disciplines. Einav and Yariv (2006) find that in economics, authors whose names begin with letters that appear earlier in the alphabet are more likely to receive tenure in top departments and other recognitions before colleagues with names that begin with later letters. While this “alphabetical discrimination” may not be the case in other disciplines, this issue in economics suggests that even efforts to represent the equality of collaborator contributions may be mitigated by traditions that persist in recognizing ordering of authors as code for their levels of involvement or expertise.
5. Feedback.

Though it is arguable that language documentation is more product-oriented than other sub-disciplines of linguistics with its emphasis on the production of a body of data that can be used by scholars and speech communities, there has been little discussion of how to obtain feedback from end-users of those products. Collaborators on computer technology, however, frequently consider feedback at several points of development. Writing about the model of LINUX development, Fanderclai asks researchers in the humanities to imagine “a research project where all of the source materials—the methods, the raw data, the rough drafts, and the tentative conclusions—are released via the Internet as soon as they are created” (Fanderclai 2004:317). She argues that making the data available at several stages provides opportunities for feedback from diverse stakeholders and has the potential to eventually strengthen the final analysis as it is being developed.

In a collaboration, continuous evaluation of the meaning, value, and significance is just as important as feedback on the data and analysis. Dwyer’s second principle for ethics in language documentation with communities also applies to co-working with researchers: “The research relationship must be consultative, continuously negotiated, and respectful” (2006:38). The continuous negotiation need not only be on an interpersonal level among the participants. Universities and other institutions will also have to consider feedback on collaborative language documentation. Though in general they have been slow to develop reward systems that recognize and value collaborative work, universities stand to give some of the most important feedback, in terms of tenure and promotion, if there is to be hope for a change in the culture that underrates collaboration.

5. COLLABORATION IN THE CULTURE OF DOCUMENTARY LINGUISTICS.

5.1. THE DATA MADE US DO IT. The culture of the humanities has oriented researchers away from collaboration, but the comprehensive data-driven culture of language documentation can bring us back to it. The need to create comprehensive documents requires expertise that is beyond the training of rank-and-file linguists and in turn demands that language documenters must look beyond the confines of sub-discipline, discipline, and nation to develop projects that are useable by an array of stakeholders. Because of this focus on the data, language documentation is a new sector of the humanities and thus requires new structures of interaction among researchers and in the planning process at institutional levels (Davidson and Goldberg 2004).

5.2 FROM LONE WOLF TO THE WHOLE PACK: WHAT COLLABORATION COULD MEAN FOR LINGUISTICS. The loudest calls for the death of the “lone wolf” have issued from linguists working on language documentation. From there, the appeal has begun to slowly ripple throughout the field. We already may observe that language documentation’s attitude toward making primary data the center of analysis has influenced other sub-disciplines of the field. When language documentation theorists celebrate the death of the “lone-wolf” linguist, they are by implication also celebrating a new way of conceptualizing linguistic research, and thus heralding a new academic culture in linguistics.

In many senses a language documentation is never finished. There is always more data to discover, except in cases of extinct languages, and even then there are new ways
to approach both data and analysis. Such a realization can be intimidating to new scholars. How will we ever collect enough data to produce a suitable documentation? Viewed another way, however, the same statement is inspiring. “A lone researcher can’t keep up. Even a small group of researchers can’t keep up. We need more ways to collaborate on a much larger scale” (Fanderclai 2004:311). Instead of thinking of researchers in competition, our field must start to consider alliances in real and meaningful ways. I have tried to use the examples of the physical sciences and their collaborative traditions as a foil to the business-as-usual non-collaboration in the humanities. By identifying five dimensions of collaboration that have bearing on language documentation I hope to have turned attention to possibilities in the work of documenting endangered and under-described languages of the world. There may be more than enough to go around, and in working through collaboration we may be exposed to new methods and questions that we would never consider on our own.
REFERENCES


Akiemi Glenn
akiemi@hawaii.edu