Linguistics and Political Science: 
A Strategy for Interdisciplinary and Ethical Research 
Methodology on Language Endangerment 
and Political Conflict

Shobhana Chelliah  
Department of Linguistics, University of North Texas

James Meernik  
Department of Political Science, University of North Texas

Kimi King  
Department of Political Science, University of North Texas

We propose that linguists and political scientists develop an interdisciplinary and ethical research strategy for studying the relationships between language endangerment and political conflict. A leading cause of language endangerment is political violence driven by outside actors who expropriate land, extract resources, and displace individuals, many of whom reside in communities that speak endangered languages. Most language documentation projects, however, do not address the political landscape that causes the conflict, whether it is history, language policy, conflict over natural resources and ethno-religious identities, or absent and co-opted governmental institutions experienced by the communities in question. At the same time, political scientists have developed models to explain and predict the political conflict and violence that threaten entire communities and can also explain why indigenous communities are particularly at risk of being harmed by this type of violence. We suggest that an interdisciplinary strategy that combines some of the large N data analysis strengths of political science with the qualitative, community-driven research of linguists can best help scholars understand the determinants of language loss; conduct such research ethically, and help utilize the fruits of this research to support and empower endangered language communities.

1. Introduction  
A child is separated from its family when evacuated from pitched battle; a family must flee when targeted by ethnic rivals; a community assimilates to a cultural majority group to gain acceptance. A common thread linking these scenarios is the socio-political instability and violence that leads to language endangerment as powerful groups undermine the autonomy and culture of those who stand in the way of their political goals (Bradley & Bradley 2019). We propose that linguists and political scientists develop an interdisciplinary and ethical research strategy for studying the relationships between language endangerment and political conflict. While one
of the leading causes of language endangerment is instability brought about by such political conflict, most language documentation projects do not address the political landscape that causes the instability, whether it is history, language policy, conflict over natural resources and ethno-religious identities, or absent and co-opted governmental institutions experienced by the communities in question. It is clear that when conflict and instability occur in a society, endangered language communities suffer as conflict actors seize their land, kidnap their people, and undermine community traditions and practices. At the same time, political scientists have developed models to explain and predict political conflict and violence that also threatens entire communities. While political scientists recognize the critical role of language in these conflicts, they often access language-related factors via data and reporting from elites rather than those experiencing the violence. Databases with personal accounts from experiencers – their narratives of life in a conflict zone – are rare, and when available, data are often accessible only in translation in a world language. In this article we set out a methodology for collaboration between linguists and political scientists to generate a more comprehensive understanding of the political conditions that lead to language endangerment and loss. We advocate for a holistic methodology that makes use of data analytics used in political science with the knowledge linguists gain from in-depth and up-close studies of endangered languages. Furthermore, our strategy is anchored in and guided by ethical concerns for the communities studied as well as the need for researchers to utilize the fruits of their knowledge to support and empower members of these endangered language communities.

Our paper is organized as follows. First, we review what we know regarding current practices of data collection and research methodologies to better understand language endangerment and political instability and conflict (LEPI) in documentary linguistics and political science. We are greatly aided in this effort by a National Science Foundation funded conference [#1624346] (October 2018) that brought together linguists and political scientists to identify, understand, and improve upon our common interests and methods while also critically thinking about the ethics and best practices of our disciplines. Second, we discuss the methodological practices and shared challenges both disciplines face, and we examine how a synthesized research agenda would improve our collective understanding of language endangerment in communities affected by political violence and the ethical considerations that should guide the research process. We also discuss how this methodology could be deployed in an ongoing study of language loss in Colombia. Our conclusions are grounded in establishing the broader impact of social justice and peace which can be furthered by access and preservation of knowledge from community members’ conflict narratives.

2. Language documentation and political conflict As subjects in their respective fields, language endangerment and political conflict are closely tied together through the common theme of better understanding the causes and consequences of language loss within communities. Language documentation (LD) by linguists creates, preserves, and provides access to a record of the linguistic practices that are characteristic of a speech community. That record could have many purposes, the most
common ones being language description, language pedagogy, and culture and language revitalization. The standards and processes for language documentation have developed in tandem with developing technologies and archiving availability for long term preservation, access, and online dissemination of documentary materials. Additionally, methods have been shaped by community-directed or initiated projects which have determined the content of the documentation and the method in which it is collected, used, and archived.

Knowledge regarding the causes and consequences of political violence and instability has evolved substantially from Thucydides (fifth century B.C.) and Hans Morgenthau (1962a, 1962b, 1962c) to the many scholars who have published thousands of articles since World War II. The study of conflict has been one of the principal foci of political science in general, particularly in the field of international relations. These topics have ranged from the balance of power among nations and arms races, to rational choice models and diversionary use of force theory (for excellent summaries for the state of the discipline, see Midlarsky 2009, Mason & Mitchell 2016). Within the last three decades, there has been an exponential expansion of conflict studies to include human security and human rights to prevent post-conflict countries from returning to intrastate conflicts by providing a sustainable peace (Lie et al. 2007; David 2017).

As we continue in the following sections to describe the various and often-used research methodologies of this field of research, we cannot emphasize strongly enough that our characterization of various general trends is not always reflective of all research. The task of categorizing all of this research, even from just the last 30 years, would be a monumental undertaking requiring a small army of research assistants. Thus, there are multiple exceptions to all of the points and generalizations we make. Our goal is to start a dialogue for how such research is generally done and how the methodologies we propose can help us answer more and different valuable questions.

2.1 Linguistics and language documentation: Current practices  In the past 20 years, developments in language documentation (LD) methodology have focused on best practices in audio and video recording, providing access to source recordings through technologies for transcription, translation, and language archiving (McDonnell et al. 2018). Through partnerships – those between communities, speakers, and language documentarians – LD has gained a strong user-centered focus with language vitality as a central goal (Bischoff & Jany 2018). Through collaborations between language documenters and other disciplines (e.g., ethnomusicologists (Grant 2016), botanists (McClatchey 2011), astronomers (Holbrook 2011)), LD has shown potential uses for and gained insights on improving documentation in various fields.

There is no discussion to date in the LD literature on methods for collecting data on political conflict or instability similar to collecting information on botany or astronomy. Training for data collection in the field focuses on typology and data analysis and on practical issues such as data management and archiving, recording methods, data collection and collaboration, how to relate to others in the field, managing physical and mental health, time management, and safety.
Review of geopolitical factors is not standard preparation for a linguistic field trip. At best, traditional LD methods provide, as accidental byproducts, insights into social networks, disputes, changing understanding of land rights, and so on. More can be done and done systematically (Shulist 2018; Urla 2012). Factors relating to political instability may be incorporated into LD projects as part of ethnographic description or social network analysis. For example, in a longitudinal investigation of language change in the villages of the Lower Fungom region of the Cameroonian Grassfields, Di Carlo and Good (2014, 2017) discuss extreme linguistic diversity and the development and maintenance of language ecologies in the Lower Fungom, which encompasses 13 villages and includes the use of seven to nine distinct “languages”. There are no monolingual speakers. For the most part, these are not related or mutually intelligible. For the approximately 12,000 speakers of those seven to nine languages who live in this rural economy, multilingualism is the norm. Each village is associated with a language. In fact, without a language, a dwelling group cannot be considered a village. Di Carlo and Good consider how linguistic identity promotes or inhibits political stability given this multilingual scenario and interactions between the groups. They find that the local political system seems to serve as a buffer against language endangerment. The status quo for the languages at play is also supported by a state-level system which is neutral to the linguistic ecology in the villages. This lack of attention by the state provides some safety for the continued existence of the languages. These discoveries about the Lower Fungom have been made as a result of language documentation work which incorporated the study of language ideologies, networks, and stability – not the norm for typology study.

Another example of the orthogonal study of language endangerment during documentation work comes from Emiliana Cruz (2019) who describes a method she has used for LD that also provides insights on political conflict and its repercussions for language. Cruz describes “Share while Walking” where she walks with a community member for the primary purpose of discussing land shapes and toponyms. She has found that on these walks, speakers often also share their feelings about living in the village, about boundary disputes, water privatization, mining concessions, local conflicts, and problems due to the gun trade and organized crime. Cruz highlights the need to understand the interplay between language stability and social stability where the researcher has to engage with experiencers at a local level. “Share while Walking” works only with deep trust between the researcher and speaker, and it may be that the researcher needs to be part of the community, as was the case for Cruz. We discuss the ethics of public access to materials gained through this method in §3.

Typically, documentary and field linguists tend to avoid direct conversation about political conflict to avoid danger to the community, the speakers, and the researcher. When collecting narratives, linguists most often collect traditional folktales, conversations that do not embarrass community members or contain problematic gossip. However, since a stated goal of LD is to support language communities in their language revitalization efforts, it is then incumbent on the language documenter to understand the causes of language endangerment even where the documentation may require extraordinary precautions for the safety of all concerned and the ethical cre-
ation of such documentation. We hold that avoidance of dealing with this sensitive aspect of the speaker’s lives would in itself be unethical. As we discussed at the LEPI conference, several questions remain as to how political conflict can be safely and ethically documented. How traumatic or dangerous might it be to tell, document, preserve, and disseminate personal stories and narratives revealing political instability and repressive policies? Would it be respectful to use these collections as data for grammatical description or linguistic and political science theorization?

2.2 Political science: Current practices and processes Since the first large-scale, systematic, and data-driven effort to explain international conflict – the Correlates of War (COW) project (founded by J. David Singer at the University of Michigan in the 1960s) – the study of conflict has mostly assumed a quantitative focus. Many noteworthy efforts to advance our understanding of international wars, and now largely intrastate wars (e.g., civil wars, which have been the predominant form of major conflict in the world since the end of the Cold War), have followed this model. To be sure, there are classic works of international relations that address conflict and are qualitative in nature, such as Waltz’s Theory of International Politics (1979) and Gilpin’s War and Change in World Politics (1981), as well as more recent outstanding qualitative work over the last three decades (Brown 2019). In part because of the ease with which one can access and analyze data sets like the Correlates of War, quantitative research has proven to be more popular. There have been ongoing efforts to maintain and expand the COW data with additional information, such as the National Material Capabilities data (Correlates of War Project 1816–2012), the Issue Correlates of War data (Hensel & Mitchell 1816–2001), Bruce Bueno de Mesquita’s hugely influential rational choice model and data to predict war,¹ and the EUGENE software program, which was designed to bring many of these data sets together into one easy to use program. The major conflict data collection initiatives are now most fully realized on the web sites maintained by two Scandinavian institutions. The Peace Research Institute of Oslo (n.d.) and Uppsala University (Sweden) Conflict Data Program (n.d.) bring large numbers of these data sets together and have further substantially decreased the costs of doing research in this area.

Paradoxically, the very features that make these databases so accessible in political science research are the same ones that have made a more contextualized and localized analysis of conflict violence comparatively more difficult because of the manner in which the data are organized. A substantial majority of data sets in this field of research (especially like those listed above) use the nation-state year as the unit of analysis with each observation or row of data providing information about one country and its conflict per year. The data typically range for extended periods of time – decades, even multiple centuries – giving thousands of observations about snapshots of conflicts among and within nations. Databases organized in this manner dominate the international relations field and, to a lesser extent, the field of comparative politics. This aggregation of data and the field’s view of political conflict in any one country is

¹Please see https://as.nyu.edu/content/nyu-as/as/faculty/bruce-bueno-de-mesquita.html (Accessed 2020-05-16.)
thus from a fairly high level of observation. These data sets and the scholars who have labored for decades to provide this rigorously-checked, extremely useful information have substantially advanced our study of international and intrastate conflict. Our discussion is not criticism, because these data sets were designed to answer questions about the behavior of nation-states, not communities – and they have been an enormous success. We suggest, however, that if we seek to answer questions regarding the causes and consequences of violence within countries, we need to develop data and information that is organized around appropriate and relevant units of analysis for our studies. If we wish to study the impact of political violence on language endangerment, for example, political science scholars must train their sights on subnational or local data, including both quantitative and qualitative data.

The major data sets in the study of conflict – cross-national time series data as it is often called – have helped us understand a variety of processes at the international and national level that restrain or facilitate the movement toward conflict. We know that democracies rarely, if ever, wage war on one another; that there is inconsistent evidence regarding whether various types of international systems (i.e., bipolar as during the Cold War; multipolar as during the period between the end of the Napoleonic wars and World War I) affect conflict; and we know a great deal about the impact of alliances, credibility, arms races, and a myriad of factors about the likelihood of war. Nonetheless, these data and the studies they have spawned are not always as helpful as we might like. Such databases are often better suited for uncovering correlations rather than causal processes (e.g., virtually everyone agrees that democracies do not fight each other, but there is considerable disagreement over why this is). Such macro-level data sets are also not equipped to explain variation in violence within countries (i.e., most wars are fought over particular areas within a country, while large parts of the nation may see little if any conflict violence, such as the war between the Ukraine and Russia that takes place almost exclusively in the east of Ukraine). The good news is that work has already begun on developing data at the subnational level that can help us answer questions about the causes and consequences of political violence within states.

There have been several major data gathering enterprises that have endeavored to focus within countries on collecting data on “events”. These databases, and the resulting studies, use the conflict event as the unit of analysis (or some aggregation of it by discrete period of time, week or month). For example, an event might be a battle between government forces and rebels or a rebel massacre of civilians. Such data are most often derived from newspapers as the cost of machine-coding media stories is fairly cheap. Such data allow scholars to see when and where violent events take place, which actors were involved, the number of casualties, and other geo-location information. Data sets like the Armed Conflict Location and Event data (ACLED) (Raleigh et al. 2010) and the Social Conflict Analysis Data (SCAD) (Salehyan et al. 2012) provide a great deal of information that allows researchers to study conflict processes at the local or even micro-level of analysis. This allows us to examine why there are spatial differences in conflict locations or how the strategic moves of conflict actors influence subsequent behavior. Such databases have been a boon to studies of intrastate
war. The readily available data, most of which is still relatively novel in the field of intrastate conflict studies, has made it very attractive for scholars and especially graduate students who are looking to study the causes and consequences of subnational conflict. This, in turn, has reduced the need to go into the field to gather rich contextual information that may be important for better understanding conflicts. Because the efficiency, time, and monetary gains resulting from analyzing conflict from the comfort of one’s secure office are substantial (field research typically requires major grants to fund such work), most conflict scholars are fairly removed physically and culturally from the people and places they study. While rigorous research may not always require such deep knowledge, our thorough understanding of conflict dynamics at the subnational level, and especially as it relates to language-based conflict and language endangerment, demands more in-depth, more local, and greater contextual, and often qualitative, approaches that we find are the hallmark of linguistic studies of language endangerment.

Ultimately, our principal goal is to suggest a research strategy for documenting and collecting data on language endangerment and political violence that satisfies multiple criteria. By utilizing both traditional research methods in political science, such as large N data (i.e., ranging from hundreds to tens of thousands of cases) bases and statistical analyses and the more qualitative methods of linguistics, we strive to develop a holistic methodology for simultaneously studying political violence and language endangerment. Our strategy is to borrow the best practices from each discipline in order to develop a more thorough understanding of language endangerment, advance knowledge, and support the endangered language communities. We are encouraged by the fact that there is an increasing interest in interdisciplinary research on language endangerment and political violence (e.g., Anderson & Paskeviciute 2006; Bormann et al. 2017; Laitin 2000; Mabry 2010; Medeiros 2017).

3. Process and ethics, from initial data collection to archiving and dissemination of data   As we work with communities whose marginalization is quite profound, we recognize that their very existence vis a vis government and society is tenuous. Their culture and traditions and the resilience that often comes from contending with antithetical social and economic forces can provide these communities with tools to contend with such challenges. But there is no mistaking that both their social communities and the languages that bind them together are being eroded, if not attacked, by these forces, which creates a precarious situation for the communities. It is our responsibility as scholars to first, as physicians are commanded, “do no harm”. The magnitude and potential consequences of our actions can have both direct and ripple effects that may cause further harm to these communities. It is equally becoming clear that our viewpoints are limited by our experiences as scholars from industrialized nations and that we must turn to communities themselves to help shape the direction of research (Getty 2010; Hart 2010). We hold that it is absolutely essential that scholars consider the ethical implications of every step in the research process as we seek to understand the relationships among political conflict, violence, and language endangerment. We will also argue that scholars have an ethical responsibil-
Linguistics and Political Science: A Strategy for Interdisciplinary and Ethical Research…

It is important to gain knowledge in a participatory way with the communities they study and that researchers should use the knowledge gained from fieldwork to give back and assist the community in a mutually agreed upon manner. We describe below some key ethical considerations that scholars should consider when documenting stories of conflict. This section concludes by outlining the critical importance of working with communities to understand more organically the structures and processes they face which contribute to language endangerment.

3.1 Research design

Because we wish to both document and preserve endangered languages and understand the role of political violence in causing language endangerment, our interdisciplinary strategy for data collection must be comprehensive, both in breadth and depth of the information that is gathered. That is, we must collect data at all levels of analysis – the individual, the community, and the nation – to understand how political violence in the community and within the state brings about the conditions that lead to language endangerment as well as the other problems such violence poses to these communities. Individual conflict narratives are a key, perhaps the key, to understanding how these various forces work together to the detriment of these communities. We recognize that individuals’ recollections and narratives of the violence they have experienced over many years form the foundation of our understanding of language endangerment and political violence. Their stories, narratives, and histories are not only critical in developing a more fine-grained and nuanced appreciation of how conflict violence has affected their lives and languages, it is also important to document these stories for future generations. Narratives of life in a conflict zone are also vital in language documentation as such pivotal events in the history of the community provide unique language patterns, vocabulary, and discourse patterns that can increase the depth of language documentation. Documenting these narratives now, while people are still alive and able to speak, is vital. Just as in the case of language loss, once these speakers are gone, their stories are gone with them. Conflicts can recur, and when they do, we will need to bank on lessons learned from past experiences. Likewise, as other research on transitional justice within aging populations has found, invaluable information passes with the persons who have experienced the violence (King & Meernik 2017). A key first step in conducting holistic and ethical research is to conduct the widest possible survey of evidence at all levels of analysis.

We must begin by identifying relevant data and knowledge at the community level that will provide us with the fullest sources of information about local conditions, and at the national level from government, international, NGO, and other publicly available sources. The readily available statistical data will allow us to measure levels of poverty, health, education, crime, displacement, inequality, and many other factors that provide a picture of the systemic challenges faced by individuals living in these communities. Knowledge of the “on the ground” facts as experienced by documentary linguists helps us evaluate the accuracy of those publicly available sources and explore more in-depth the challenges which endangered language groups face.
Similarly, even though linguists tend to focus more on micro-level factors having an impact on LD, national-level data are also critical, especially survey data, because these help us explain how the political climate, elections, institutional structures and regimes, and public opinion influence government responses. Knowing how the government responds to conflict, language loss, and other challenges facing communities is vital for better understanding the attendant threats and risks. For example, we know that support for the Colombian peace referendum in 2016 (which was narrowly defeated) varied positively with municipal level violence – those who experienced the most war were the most desirous of peace. Those living in the urban centers of Colombia were less likely to indicate support for peace, which creates additional challenges for marginalized communities in the future. Those who control the resources in government and the major cities are not nearly so committed to building peace which has consequences for language endangerment and the potential for continued violence. Understanding that the root causes of conflict and the likelihood of the conflict recurrence allows scholars and practitioners to have a more accurate picture of the threats to linguistic integrity. Generally, however, there are few, if any, ethical concerns the researcher must be cognizant of and practice with regard to these large national databases. While, to be sure, there may be measurement and extent of coverage issues when using these data, their usage does not place the ethical burden on the researcher who utilizes such data.

The ethical concerns we emphasize are those that arise when the researcher enters the community to study it. Data collection must be conducted in a manner that prioritizes the safety and security of those individuals we are studying above all else. We return to the subject of ethics and professional responsibility in more depth below, but for now, we note challenges that are especially problematic for research on political violence. There are numerous problems (not the least of which is individual, community, and researcher safety) in collecting data about events (especially political) in communities afflicted with violence. Because in some of these communities the conflict actors may be present, or just over the horizon, it is absolutely imperative that whatever narratives of conflict violence are communicated, such encounters must allow for the maximum security possible that does not call attention to the encounter. Whatever notes are taken or recordings made of these encounters must be housed in such a manner that privacy and anonymity are protected at all times. Protocols must be in place for where and for what period of time the original recordings may be kept and whether and how they should be destroyed. Language documentation protocols for dealing with these personal narratives might be different than for other narratives in the corpus. Everything from the arrival of the researchers and their living situation in the community, the methods used to record narratives, data storage, and data availability must be scrutinized for potential lapses in privacy and anonymity. Experience has shown that it is possible to conduct ethical as well as scientifically rigorous research on personal experiences with conflict violence (Arjona 2016; Daly 2016; King & Meernik 2017; Stepakoff et al. 2014; Stepakoff et al. 2015; Stover 2005).
3.2 Obtaining permission and minimizing risk  The development of the research design – from the objectives and planning to collecting and analyzing data – involves human subjects who deserve ethical treatment and concern for their well-being, especially in conflict ridden regions, where many persons have been exposed to extreme violence and trauma. The need for Institutional Review Board (IRB) approval and in-country permissions should begin a process that raises, considers, addresses, and continuously re-evaluates ethical and practical challenges of doing this interdisciplinary research. While IRBs can provide much general guidance regarding knowledge of the communities, countries, and regions, there are a myriad of challenges to conducting research in marginalized communities. It is here where linguistic studies that consider the unique circumstances that come with field work are highly salient. The IRB can assist, for example, with designing questions that reduce the likelihood of causing an emotional trigger, but only the researchers and practitioners are likely to know how such questions may be received within the community under study. Scholars need to consider not only how their question may affect the individual’s emotional state but also how various responses are likely to be perceived in the community. Will the very act of speaking with others, particularly about conflict issues, be perceived as an act of dissension, if not rebellion, by certain actors in or near the community? If individual community members identify particular problems or challenges facing the community or point to certain actors as responsible for their precarious situation, will sharing that information affect the security and wellbeing of the individual, their families, or their personal and professional affiliations? Despite our very best efforts to maintain the anonymity of our study subjects, it is always possible, especially in small communities, that others will overhear, or perhaps the individual will tell family members what they said, or perhaps even what questions were asked. All this information may arouse the suspicions of those with less than honorable intentions toward the community.

Persons who speak to outsiders in politically unstable environments may face repercussions depending on the information that is provided and the context in which it occurs (King & Meernik 2017; Meernik & King 2019; Crowley 2007). It is not just the mere speaking to others that can subject individuals to human security threats but also the substance of what is spoken about. Those who speak out about problems may be more vulnerable (especially in language endangered communities), and these persons may be more likely to be targeted for speaking out. Their vulnerability may stem from social isolation, lack of family ties, physical health, and psychological challenges among others because it draws attention from opponents who would prefer that their political, and potentially criminal, actions are not exposed. Therefore, researchers must consider the consequences of their field work on the communities they are studying. It is vital to thoroughly review all potential risks facing the community (as well as the scholar herself) from the questions posed to individuals. This evaluation must necessarily be done in consultation with those within the community, who have the best understanding of the threats they face, as well as with experts on the impact of violence on mental health.
We suggest that research must not be done “at any cost” – that is imposing the necessity of doing the research when the risks are too high that it poses direct threats to those who are telling their stories. Better that the study be postponed or revised to reduce the element of security risk. While the IRB can assist in this process (and certainly consideration of the risks to vulnerable populations are explicitly part of the IRB process), it is incumbent on the researcher to use all of her specialized knowledge to give extremely careful consideration into the risk assessment as well as provide potential, workable solutions for ameliorating risk and, most importantly, have protocols in place to minimize threats to human security that accompany the research process. Such physical risk is different from the potential for re-traumatization which may occur from individuals actually telling their story. We address that below under fieldwork, and now turn there.

3.3 Fieldwork  Political scientists and psychologists often collect data and information from war crimes victims, refugees, and other vulnerable populations by utilizing some type of survey. To provide standardized and comparable measures across respondents, the questions on the survey may be objective and close-ended (e.g., “have you ever experienced one of these types of conflict violence (torture, displacement, loss of property, etc.)?”). Political scientists also provide survey respondents with more open-ended opportunities to relate the conflict narrative and other stories that they wish to tell, which is more similar to linguistic research which may focus on individual level (Cukor-Avila 2005). Creating such opportunities, whether through open-ended questions or simply giving the individual the chance to spontaneously offer her own observations, is important for gaining the trust of these victims and witnesses (such as the “share while walking” method discussed earlier), and it may provide unanticipated insights. It also allows the respondent to narrate their own story of what is important to them. This process also allows for a deeper understanding of their experiences, even if the true import of what we hear may only be fully grasped later.

Such an opportunity to provide open-ended information both serves as a collection point for much-needed data about individual and collective experiences and provides those involved with opportunities to speak more expansively on their experiences (see King & Meernik 2017; Meernik & King 2019; Stepakoff et al. Henry 2014; Stepakoff et al. 2015; Cody et al. 2014). In our own study at the International Criminal Tribunal for the Former Yugoslavia (ICTY), tribunal witnesses were asked broad and open-ended questions after a multiple choice survey, and witnesses used the opportunity to share their personal experiences from the war, share grief and memories of loved ones lost, and provide their impression of what it means to “do justice” and to provide fairness. While the questions were prompts, they directed the respondents’ attention to share their experience about testifying for an extended period of time (while some spoke for less than 15 minutes, other interviews lasted almost three hours). Some participants even expressed that it allowed them perspective and closure on the process. Allowing the space for open-ended discussions about the broader impact of what is being studied through directed topics can
only be achieved by such a dialogue that cannot be addressed by closed-ended survey responses (King & Meernik 2017). We encourage researchers to build in opportunities for such open-ended engagement to provide more spontaneous narratives which can provide invaluable insight. Following current trends in language vitality surveys (Rice & Linn 2017), it would also be useful to provide training through seminar-style lectures and discussion on how shared stories are used for peace and reconciliation or healing of communities. These seminars can provide training in survey creation so that communities or individuals can themselves create memory banks featuring those aspects of conflict experience they find most salient.

As one caveat to this, we note the critical importance of having persons who are trained to handle the method of open-ended discussions. Trauma survivors face difficulties in recalling their experiences and this can contribute to re-traumatisation in the “truth telling” process. Among the paradigms about truth-telling and trauma, victim-centered justice in political science is relatively new, and there can be “a vast gulf between the occurrence of trauma and the capacity of victims to publicly narrate their experience” (Niezen 2020: 4). There is a substantial debate about whether truth-telling (telling your story, recounting the past experience, providing a historical record) contributes to a catharsis or whether it can trigger past traumatic memories (Hamber 2009; Hayner 2010; Byrne 2004). Trauma psychologists indicate that there may be cathartic empowerment in truth-telling (Herman 1992, 2003; Laub 1992), and depending on the mechanism where witnesses tell their stories (truth commissions, tribunals, etc.), transitional justice scholars in law and political science have questioned whether there may be healing power in truth-telling (Gibson 2006; Brounéus 2010; Mendeloff 2009; Barria & Roper 2005; Cody et al. 2014). Telling your story to a trained psycho-social professional is very different from communicating trauma you have experienced to judges seeking truth and justice from you, lawyers cross-examining you, linguists wanting to document you, and political scientists who want to understand you. Testimony in this context has very different causes and consequences, and its treatment – if it is to become restorative or reparative for the individual or larger society – needs much more development theoretically, ethically, and practically to ensure that best practices for psycho-social well-being are incorporated.

3.4 Data management While in the field and after returning to the university, scholars in both linguistics and political science face challenges in ensuring the security of data containing sensitive information that might place a study participant in jeopardy. There are some common and some unique ethical concerns for the two disciplines. In both fields, researchers are working with people who may unintentionally or due to a research stimulus reveal sensitive facts about themselves. Post-conflict research involves issues regarding physical and psychological trauma, security, and protection

Because of this, in the ICTY study, the persons doing the interviews were trained and experienced social workers who could identify when, and if, respondents were becoming distressed about difficult topics (King & Meernik 2017). For this reason, we suggest that all research should incorporate concerns about well-being and protecting those we study from additional harm that might come from sharing their histories with us.
of witnesses, and these must be held in strict confidentiality. For example, those who talk about specific instances of abuse during conflict could open themselves up to retribution. In conflicts as complex and multi-sided as the ones we study in Colombia and Northeastern India, it can be quite difficult to distinguish from which sectors threats might be emanating. Hence, even when scholars are familiar with the conflict dynamics, the intricacies of each individual’s life, and the dangers they confront, researchers must at all times maintain vigilance so as not to reveal names, dates, and places. On the other hand, there are also individuals who want to speak out and seek to use the researchers as a means to disseminate false information or a point-of-view not in keeping with the majority. The researchers would need to be able to assess the veracity and intentions behind the accounts. Open-access archiving of language documentation projects would need a rubric by which to weigh the possible risks of exposure of these types of narratives to the public access. Possibly, unlike traditional narratives or other first-person narratives, politically sensitive narratives could automatically come under embargo.

The nature of data is also a critical question that needs to be assessed. There is a body of literature discussing ownership of the language and research in the field of language documentation. Language archiving is still relatively new, so it still remains to be seen how ownership of data squares with archiving and presentation of materials in a shared public space. Who owns the data or story and what can be done with it? Can it and should it transfer easily from collection for language documentation to big data analysis for political scientists?

In 2019, ICPSR-QDR held a Data Sharing Workshop at American Political Science Association in Washington, DC to discuss how to optimize open sharing of research with human participants. This is of equal concern to linguists. As funding agencies, such as the National Science Foundation, require data and results to be made available in open-access repositories, we must establish how to provide access, but at the same time, protecting subjects and researchers becomes important. Data management plans and consent agreements must also reflect a new reality that, with big data repositories, personal narratives can become data. Narratives in a language archive may easily become decontextualized and used for all manner of data mining, e.g., natural language processing for sentiment. In addition, traditional consent forms for language documentation might need specific permissions to use narratives for analyzing political instability and giving voice to traumatic experiences.

4. Applying the research and ethical strategy In this penultimate section we describe how we will pursue this ethically-driven research strategy in studying language endangerment and political conflict among indigenous groups in Colombia. We examine conflict violence and language endangerment in Colombia for several reasons. First, there has been a great deal of both data-driven quantitative research and more qualitative and contextual research on conflict, often by the same scholars. Second, Colombia is a data-rich environment where scholars will find that there is substantial data on socioeconomics, conflict violation and human rights violations, and event data based on the actions of the conflict actors. Our goal in this research is to un-
understand what risk factors differentiate communities (i.e., Colombian municipalities) where endangered languages are present from those where no such endangered languages are present. Subsequently, our interdisciplinary research goals are to both document the conflict experiences of these communities and the endangered languages found in them.

Some background on Colombia is in order. Colombia recently concluded a decades-long civil war with guerillas of the FARC (Revolutionary Armed Forces of Colombia) movement. However, it also suffered systemic and horrific violence in earlier years (e.g., the period known simply as “La Violencia” in the late 1940s and 1950s, where conservative and liberal militias and armed groups attacked each other) with other rebel groups (e.g., the M19; ELN, ELP guerilla movements), with paramilitary groups that were often aligned with the state in its brutal battles against leftist insurgents, and with criminal actors, such as the infamous Medellin and Cali cocaine cartels in the 1980s and early 1990s as well as other drug cartels today. In short, Colombia has substantial experience with various types of political violence (e.g., Arjona 2016; Daly 2016; Gallego 2018; Matanock & Garcia-Sanchez 2017; Steele 2017) and criminal violence (Lessing 2015; Yashar 2018), to name but a few. These authors use a variety of approaches, but many of them include both contextual information gained from firsthand observation (e.g., Arjona 2016; Daly 2016; Lessing 2015).

The goal of our study of political violence and language endangerment in Colombia is to better understand which characteristics or factors are associated with language endangerment and, in particular, the role played by political violence in influencing the loss of language speakers. We suggest four steps in an interdisciplinary research process to document and explain language endangerment and political violence, assuming the researcher(s) have already thought through the ethical concerns we raised in the previous section: 1) thorough review of all written and quantitative evidence and the use of statistical analysis to explain general patterns of language loss; 2) in-country study and gathering of qualitative and quantitative data from interviews and surveys; 3) holistic review of all the evidence gathered and production of reports, articles, and books to document the major findings; and 4) support of community through the application of research findings. We address each point below.

4.1 Review and analysis of existing research  

Research processes often begin with the review of previous research, which we assume all well-trained scholars utilize. We suggest a review of all of the data available for a given country at the most decentralized unit of analysis possible. In some countries this may be the “state”, “province”, or “department”, while in other nations, like Colombia, it is possible to obtain extensive data at the municipal level. The municipality is more akin to a county in the United States than a city, although Colombian municipalities often consist of a large city, such as Bogota or Medellin. In Colombia, like in most countries, the first place to look is the national government organization in charge of collecting census data, but there are many other sources of data available from scholars who develop their own indicators or curate indicators from other web sites. These sources may

3Data on Colombia can be found at https://www.dane.gov.co/index.php/en/ (Accessed 2020-01-02.)
concern only Colombia or contain statistical data on a wide universe of countries. Increasingly, conflict scholars are utilizing events data, as we described earlier, which provides a very fine-grained level of detail on which actors were involved in a conflict event, what the event concerned, who was involved, and when it took place and where. Such data, especially when combined with more traditional socioeconomic and demographic factors, brings multiple benefits.

First, and most obviously, these data can be used to provide an initial analysis of the specific characteristics that distinguish communities of language endangerment in comparison to other municipalities in Colombia. These initial analyses with quantitative data help us test for the relevancy and statistical significance of those factors we suspect will be linked to language endangerment, such as the level of conflict violence, poverty, and the degree of marginalization of the community. Second, through the use of these data, we can identify a representative sample of communities that can be studied more intensively through interviews and qualitative research. More specifically, through statistical analysis, we can identify communities that are very similar on a core set of features (to ensure we are comparing apples to apples) but who differ on a small number of features that are of special theoretical interest to us, such as which types of conflict actors were most often present in the community or the extent of environmental damage occurring in the community. We argue that understanding these more macro-level characteristics associated with language loss, before beginning field work, is as crucial as a review of the literature. One must be aware of the relative impact of a host of characteristics related to language loss and conflict violence before going into the field to better understand the language and cultural habits of the communities under study (see, for example, Taff et al. 2018).

4.2 Fieldwork and in-country study  This step in the research process will be more familiar to linguists who spend a great deal of time in the communities they study and less so for political scientists who typically do not spend significant amounts of time in the field collecting data directly from individuals. In research on language endangerment, there are well specified protocols for studying and documenting endangered languages. Language documentation projects typically involve both community and non-community documenters working in tandem to determine what genres of speech to document, who to record, what purpose the corpus will fulfill, where the corpus will be archived, and who will have access to that corpus. Rather than having an academic research project set the parameters for the documentation, researchers aspire to create documentation that is useful for community needs, whether that be for cultural-affirming activities, language education, or other purposes. It is through field study, usually living with, and creating long-term relationships with, speakers and families, that linguists are able to understand and incorporate the community needs into research methodology and products.

When exploring the linkages between political violence and language endangerment, it is critical to ensure that many of the terms often used by conflict scholars are understood in the same manner by community members. For example, individuals may have very different definitions and everyday usages of terms such as “conflict
actor”, “human rights violation”, “justice”, “reconciliation”, “truth”, “order”, “security”, and so forth. Researchers need also to understand how speakers use the terms “language” and “dialect” and “correct” and “incorrect” language. Before making use of such terms in the final analysis, scholars must ensure through focus groups and other discussions that they understand how community members will understand the meaning of these critical terms. Scholars who take the time and trouble to understand how communities understand and use these terms will help ensure that there is sound construct validity in the research design and will gain a different vantage point from which to understand conflict violence and language loss. This understanding will also aid in evaluating the accuracy of large databases, such as the Ethnologue, which are used in theoretical discussion in political science and linguistics but which may have outdated or inaccurate countings of languages, numbers of speakers, domains of use, rates of bilingualism, dialect varieties, literacy rates, and religious affiliations.

For our research in Colombia, we are focusing on the impact of political violence on language endangerment, which also raises a host of ethical concerns we discussed earlier but which also depends on the researcher understanding how locals have experienced conflict violence. Another vital issue will be how the people themselves determine which violence is related to the armed struggle amongst the Colombian army, the ELN and FARC guerilla movements, and the paramilitaries, which is related more to drug trafficking and other criminal activities, and which might be driven by personal and idiosyncratic rationales. We will want to determine what constitutes conflict violence and how community members investigate and decide which groups were responsible for which acts of violence. Only by engaging in community-based fieldwork is it possible for us to learn the answers to these questions. Thus, a rigorous and compelling research design depends on fieldwork to provide proof of concept tests regarding conflict violence and language endangerment. Knowledge of community practices and epistemology are essential for knowing how community members know what they know and for understanding a history of the community that is generally not found in scholarly materials. Linguists are accustomed to such work and its benefits, while conflict scholars are beginning to appreciate the extent to which such local analyses can help us test and refine our more generalized models of conflict.

4.3 Holistic review of the evidence The third step is fairly broad and is already typical of research in both fields, but the holistic review of the evidence that we encourage does not always occur. We recognize as scholars with feet in both the linguistic and political science worlds, as well as the quantitative and qualitative domains, that this 360-degree review of the evidence is a goal rather than reality. The space occupied jointly by all four groups in a concentric circle design is not terribly large. Thus, it is all the more vital that conflict scholars within political science understand that they must have a more intimate knowledge of the workings of these marginalized communities to explain conflict dynamics and to fully appreciate the risks that these communities live with as they exist in these dangerous zones. It is all the more essential that linguists holistically understand conflict dynamics, the goals that drive the behavior of conflict actors that put communities at risk, and how violence is explained and
attributed. Knowledge of these factors can help ensure that our research brings no harm to these communities and also allows for developing the most thorough and robust accounts of political violence and language endangerment.

We suggest a pyramid strategy of analyzing these data. First, scholars should consider the most general or macro-level characteristics of the unit of analysis chosen for study – the base of the pyramid. The Colombian endangered language communities and the individuals that comprise them are our ultimate unit of analysis, but we begin at the national, state/province/departmental, and municipal levels of analysis to utilize as much data as possible regarding the correlates of language loss communities. Our goal is to identify as many statistically significant correlates of language loss communities to determine which types of characteristics and trends are most likely to give scholars the most rigorous and comprehensive explanatory power. Second, we analyze the data attained through completion of surveys and other types of conflict narratives in the endangered language communities to observe those factors identified by the community members in accounting for language loss. We would support community-devised surveys which would identify factors and frame questions appropriate to the community (Rice & Linn 2017), taking into consideration the efforts the community is already engaged in for language revitalization through language and culture awareness and teaching campaigns and existing peace and reconciliation activities. Third, we will need to analyze both the contextual data we have gathered from our video and audio recordings in which community members speak their languages and our observations of communities from field notes. Only by making use of the most comprehensive data from all levels of analysis can we as scholars develop the most holistic and accurate account of why some languages are in danger while other small language communities are not. Our three goals are to better understand the factors that lead to language endangerment; document the conflict experiences of these communities; and document their endangered languages.

4.4 Ethics, responsibilities, and giving something back  We suggest there is another normative interest political science and linguistics should share and that should constitute the final stage in the research process – the growing recognition that scholars should not just always “take” from these communities but should also give something back. As the authors of a UNESCO report on language endangerment assert:

Any research in endangered language communities must be reciprocal and collaborative. Reciprocity here entails researchers not only offering their services as a quid pro quo for what they receive from the speech community, but being more actively involved with the community in designing, implementing, and evaluating their research projects (UNESCO 2003: 3).

Collaboration between scholars and a community may include the preservation and access to their language in a format usable to the community for language revitalization. It might be assistance in using the data and documentation we collect to help a community establish links to past practices or shared notions of origin as through a
Museum of Memory or the building of a place to store or display their recollections. There are numerous possibilities. Critics might argue that we risk tainting our findings if we become too close to those with whom we are working. We suggest instead, however, that much like a doctor, our first and most fundamental goal must be to do no harm to those we are studying. It is a potentially toxic, one-sided relationship if scholars gather data from those who must re-live and re-tell their trauma and yet those same scholars provide nothing in return to the community for this enormously consequential sharing of memories. We must seriously question what harm is coming to these communities from their repeat interactions with the outside world? We suggest that identifying early on and always working toward an equitable exchange is key to this research as well as to the more general goal of preserving endangered languages for the benefit of the community. We believe that the benefits to the community, as well as to the scholarly community, are substantial and that to omit such work from the holistic research process may cause significant damage to the community studied and reduce the prospects of successful collaborations with endangered language communities in the future.

There is much scholars can and should do to learn from and preserve these endangered languages and their communities. Concerns regarding bias that might result from close collaboration and identification with individuals the scholar is studying are real, but there are options for helping communities and doing good research. First, we see that there is an imperative to share the knowledge gained from such studies and in the most relevant format for the community. Providing this information to communities in a mechanism that works best for their needs should be the minimum baseline from which studies begin. Second, when it is important to help these communities in other ways when they are facing endangerment or extinction, researchers must work to identify the most appropriate types of support for the community and for preserving the neutrality or objectivity of the researchers. For example, if there are vital community needs or requests that the researcher may wish to address, such support work could be managed through a non-profit organization established by the scholars or a larger scholarly community to provide support for these communities in the form of supplies, advice, and other benefits. Strict divisions would need to be in place so that decisions and funds pertaining to research are not mixed with those regarding community improvement projects, which necessitates more intentional and advance planning by the researcher. Such support is vital for these communities, and we believe it is also rewarding for the scholar for both personal and professional reasons.

5. Conclusion  We have endeavored to show that linguists’ and political scientists’ normative and methodological commitment to the preservation of language and conflict narratives in these marginalized communities share a desire to better understand the experiences of these communities. We have also argued that this means there is a need for the development of comprehensive, systematic, and ethical data collection processes in these communities. By recognizing our shared aims and values and by availing ourselves of the knowledge and wisdom each field has acquired, we enhance
the scope, depth, and breadth of our understanding of the languages and conflict narratives of these communities.

We argue that linguists and political scientists not only share a goal of understanding the political and social phenomena that affect these communities, we also both seek to preserve the knowledge contained in these communities, whether it is to document an endangered language or to learn from community members’ conflict narratives. Ultimately, both fields are interested in the preservation of knowledge and language, which, in turn, is heavily dependent on peace and stability in these communities so that individuals can share their knowledge of endangered languages and political violence.

To understand why and how linguists and political scientists should more intentionally collaborate in the study of communities suffering from language endangerment and political violence, we note first that both disciplines already share common interests. First and foremost is a desire to learn from these marginalized communities. Each field studies these communities because they sit at the crossroads of political change and cultural preservation. Political change and, in particular, conflict and violence, affect individuals and communities and hence increase the risks of language loss as community members are killed, flee, or even join some of the fighting forces as a form of protection against worse violence. Linguists seek to document the endangered languages that are often (certainly not always) spoken in these communities. Political scientists look to explain the causes and consequences of this violence that results in language endangerment or has contributed to language loss. Linguists who study endangered languages are there to document and understand these languages by recording the stories and lived experiences of native speakers to better capture the language inside the culture where it exists with all of its complexity and nuance. Likewise, political scientists who study these at-risk communities are there to study the behaviors and beliefs of conflict actors who have contributed to violence, as well as to collect data on public opinion and community experiences that have been affected by political violence in many of these very same communities. Thus, scholars in linguistics and political conflict often (not exclusively): 1) focus on marginalized communities and populations; 2) have a strong interest in recording narratives from these individuals that implicate dangerous and sensitive topics; and 3) study topics that are time sensitive, especially as language loss and the potential for continued violence create pressures to conduct such research before it becomes impossible.

Second, whether explicitly acknowledged or not, both fields share an underlying if unspoken normative goal of advancing peace. While political science often decries the intrusion of normative goals into objective analyses, our conflict literature has often been influenced by the goal of providing practical policy and other types of analysis and advice that can improve the architecture of peace.⁴ As Nils Petter Gleditsch (founding editor of the Journal of Peace Research and of the Peace Research Institute Oslo [PRIO]), Nordkvell and Strand (2014) phrased it, “Peace research was born at the intersection of peace activism and the emergence of modern social science” (146). Or, as Kenneth Boulding, one of the founding members of the Peace Science

⁴See, for example, Richmond et al. (n.d.).
Society (the preeminent organization of academics in the world that studies conflict), wrote, “Peace research, however, has always been normative, in the sense that it has been practiced by people who are deeply conscious of the pathologies of conflict” (Boulding 1978: 343). Hence, there are good arguments to be made that political scientists have a strong interest in the maintenance of peace both as a normative goal and as a means to an end to conduct research. Linguists share these normative and practical motivations as well. Indeed, peace is absolutely vital to language preservation as conflict violence only serves to heighten the risk of language loss. We both aim to document the languages and experiences of those countries and populations who are caught in the conflict trap.

In fact, the notion of socially engaged and ethical research is taking hold, especially in fields like linguistics and anthropology in which fieldwork is done (Jacob 2013; Nevins 2013; Piller 2016; Roche 2019). While sharing knowledge gained through fieldwork with the communities in which such studies take place seems to be the minimum that scholars could provide to signify a just and equal partnership, more can be done. These communities may have many other needs ranging from assistance with farming, security vis a vis local armed actors, and health care to perhaps just being left alone if that is what they seek. One can certainly make a strong moral case that scholars as fellow humans should establish partnerships with such communities that provide the type of assistance endangered language communities may seek with the neutrality and objectivity of the data collection process. The key would seem to be entering into this partnership as equals with mutual respect and a desire to learn from and help sustain such communities.

Finally, in addition to these shared normative and methodological concerns, political scientists and linguists have common interests in developing accurate and appropriate methodologies for acquiring knowledge and data about people whose lives are being documented through the data collection and archiving. Both disciplines have a shared commitment to adhering to ethical practices and approaches for studying populations and gathering data from and about them. This commitment necessarily should recognize and respect the interests and needs of these communities as well as their vulnerability to potential repercussions, the risk of future political violence, and re-traumatization for having participated in research. These are critical tasks for both political scientists and linguists in an era where conflict continues apace in the world, languages continue to be endangered (often because of these conflicts) on an increasing basis, and there appears to be less global interest by states and the international community in supporting and protecting communities at greatest risk.
References


Rice, Keren & Mary S. Linn. 2017. Language surveys & wellbeing. (Presented at 5th International Conference on Language Documentation & Conservation (ICLDC), Hawai‘i, 2–5 March 2013.) (http://hdl.handle.net/10125/42070)


Salehyan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, & Jennifer Williams. 2012. Social conflict in Africa: A


Uppsala Conflict Data Program (UCDP), Department of Peace and Conflict Research. n.d. Multiple years and data sets. (https://ucdp.uu.se/) (Accessed 2020-05-16.)


Shobhana Chelliah  
shobhana.chelliah@unt.edu