Language and Conflict

Tristan James Mabry

International Political Science Review 2011 32: 189 originally published online 28 February 2011
DOI: 10.1177/0192512110379489

International Political Science Review
32(2) 189–207

DOI: 10.1177/0192512110379489
ips.sagepub.com

Language and Conflict

Tristan James Mabry

Abstract

Language divides are common components of group conflict, a phenomenon reflected widely in theories of nationalism. This article evaluates measures developed by David Laitin and James Fearon in the minorities at risk dataset claiming to quantify language difference and concludes they are deeply flawed. The introduction outlines language divides vis-a-vis conflict. A theoretical analysis in the second section argues against rational choice analyses of language politics; in the third section a sociolinguistic matrix shows that these fractional measures represent language ancestry but nothing else (morphology, syntax, lexicon, orthography, status). Theoretical implications and alternative methods are considered in the fourth section followed by a summary conclusion.

Keywords

ethnic conflict, language, methodology, nationalism, rational choice

Introduction

The relationship between language and conflict parallels the relationship between ethnicity and conflict. To the extent that an ethnolinguistic population is synonymous with an ethnocultural group, general models of modernity and nationalism typically parse cultures according to languages. Hence, the Russian culture and the Russian language, for example, are considered inextricable. This model is framed theoretically within the modernist paradigm of nations and nationalism (Smith, 1998; cf. Roshwald, 2006). Credited most often to the works of Ernest Gellner (1964, 1983, 1987) and Benedict Anderson (1983, 1998), this model includes structural and constructivist arguments framed against *longue durée* processes of urbanization and mass communication. Thus, modernization homogenizes low cultures into high cultures that are institutionalized in states as *national* cultures. If rival cultures clash, however, then ethnic groups may mobilize members in a group conflict.

David Laitin (2000a: 102), one of the very few American political scientists persistently asking penetrating questions about language, considers the modernist paradigm of nationalism a 'blunt theory'. It is his position that the power of language (along with most other markers of ethnicity

Corresponding author:
Tristan James Mabry, Naval Postgraduate School, 1 University Circle, Herrmann Hall #312, Monterey, CA 93943, USA
[email: tjmabry@nps.edu]

and culture) to engender conflict is greatly over-estimated, especially in regard to the mobilization of groups to commit large-scale violence, such as insurgency or civil war (Fearon and Laitin, 2003). He is therefore critical of authors and analysts who routinely connect language division to violent ethnic conflict. Yet at the same time, Laitin (like Gellner, Anderson, and Adrian Hastings) endorses language as a proxy for culture, and by extension a proxy for national membership. He (2007, p. 59) argues this is sensible because 'language has special attributes that make it especially amenable' to his model of nation formation, a phenomenon he attributes to rational individuals calculating whether or not to join a mass movement 'for purely instrumental reasons' that need not change one's 'deep-seated identity.' This is where Laitin diverges sharply from the canon of nationalism theory, which generally posits that the *deepest* identity – an identity so inextricable with the self that its protection is worth dying for – is, in fact, a national identity.

To falsify a link between language and conflict, Laitin sought data that would enable a large-scale statistical analysis, a method he views as a critical component of the best political science. This, of course, is a debatable position (Green and Shapiro, 1994; Monroe, 2005), especially in regard to the role of language, which, as argued below, is exceptionally resistant to categorization and quantification. Moreover, if large-N analysis of language community relations were useful to disciplines that routinely research language, such as sociolinguistics or anthropology, it seems more than a little surprising that nobody before Laitin came up with the idea. It is suggested here that the method was not applied elsewhere because it is incapable of uncovering anything of value, particularly when compared with the abundance of research from other disciplines that already tells us what we need to know about language divides.

Thus, this article does not endorse quantitative analysis of language politics, but rather evaluates quantitative variables ascribed to characteristics of language communities, variables developed by David Laitin and James Fearon and eventually incorporated into the minorities at risk (MAR) dataset as ancestral language scores (ALS). In sum, such variables are rejected as incapable of representing that which is nominally under consideration: the relative importance of *language difference* in relation to cultural-cum-national conflict. Therefore, any conclusions drawn from analyses that employ such data are questionable.

What then is language difference? First, language difference may be disaggregated across a very broad spectrum of characteristics. In the analog of comparing two people, there may be similarities or differences according to physical traits such as height, weight, coloring, gender, allergies, immunities and so on, and many more conceptual distinctions such as 'personality' or 'beliefs.' In the case of languages, it is possible to observe similarities or differences according to structural traits such as whether a subject precedes a verb and object (or some other ordering), but also such distinctions as the range of phonemes, patterns of intonation, number of words (and for what things and ideas), the presence or absence of writing and the many variations of alphabets or scripts, and whether a language pair is more or less related to a common ancestral language. It is important to acknowledge that a measure of language distance is reasonably objective, though as an instrument of measuring language difference the measure appears insignificant. Imagine a meteorologist who produces a weather forecast based on quantifiable and objective data: so far, so good. On closer inspection, his or her data includes points measuring air temperature, but does not include measurements of barometric pressure, precipitation, relative humidity, cloud cover, wind speed, wind direction or dew point. Hence, the data is an objective measure of temperature, but it does not describe the weather, let alone a climate.

The second degree of language 'difference' is socially and politically subjective: what matters in the course of ethnic mobilization or mass behavior is the perception of inter-group commonality

or estrangement, and the relevance or irrelevance of language divides. In any polyglot population, which in practice means any country's population since absolute homogeneity is fiction, language serves as an immediate and obvious marker of identity. While it is in no way the only marker of group membership (others including phenotypical differences, social customs, religion and so on), it is important to recall that 'a specific characteristic of a nation often becomes the rallying point in a national struggle and, in such event, is described as indivisible from the nation itself' (Connor, 1994: 105). Although this is frequently the case with language, defining this critical characteristic is an exercise of interpretation. This article proceeds with a theoretical analysis of language and its relation to group conflict, including the contrast between structural constructivist accounts and rational choice theory, particularly as employed by Laitin. The third section, Language Communities, disaggregates his language measures and evaluates their utility according to sociolinguistic practices, and finds them flawed. The theoretical implications of rejecting Laitin's approach to the study of language, group identity, and ethnic conflict are considered along with possible alternative methods, followed by a summary conclusion.

Theoretical Analysis

Safran (2004) notes that Laitin is in good company when marginalizing the role of language difference in catalysing conflict, including qualitative authors who do not share his methods. Historical accounts by Hobsbawm (1990) and Brass (1991) both emphasize access to and distribution of scarce resources in an economic arena of communal competition; here language is employed instrumentally by the controlling elites but is not itself a source of group antagonism.

Yet the proponents of language as a critical component of conflict are abundant because the logic of language difference is simple and compelling: ethnic differences lead to ethnic conflict; language differences equal ethnic differences; therefore, language differences are associated with ethnic conflict. In the broader literature of ethnic conflict, nationalism and independence movements, this outcome is often anticipated, though the mechanics of the process differ. For Gellner (1983), arguably the most influential theorist of nationalism across the social sciences, the spread of modernization and urbanization pits any number of low cultures – and their mother tongues – in a life-or-death competition for status as a high culture. The triumphant high culture has its own preferred language, which is then protected by its own nation-state. In one form or another, it is this modernist paradigm of nations and nationalism that informs the received theory of group antagonism in the process of state formation or reformation (Barbour and Carmichael, 2000; Fishman, 1973; O'Leary, 1998; O'Reilly, 2001; Smith, 1998). In a complementary model, Anderson (1983) argues a shared language, particularly in print, is a necessary condition for membership of, or exclusion from, the imagined national community: different language communities are essentially different nations. More recently, Kaufman (2001) shows that a language can serve as a critical symbol representing the mythological narrative, that is, the social and political identity, of an entire population. Hence, when a language is imbued with this kind of symbolic significance, any threat to the health of that language can trigger a massive mobilization.

This literature represents a broader theoretical discussion of language and the importance of its role – primary, secondary, tertiary, or even irrelevant – in the formation of nations and the emergence of nationalist conflict. Yet Laitin's position on language also targets a much larger body of research from across the spectrum of comparative politics, including the lion's share of case studies produced by country or area specialists. These are the authors who address language discord as a catalyst of group conflict, such as the contests of Standard Arabic versus Berber (Morocco) or

French (Algeria); English versus Spanish (USA), French (Canada), Kannada (Karnataka, India), or Afrikaans (Apartheid South Africa); Mandarin versus Tibetan (China) or Taiwanese (Taiwan); Russian versus Estonian/Latvian/Lithuanian (Baltics), or Ukrainian (Ukraine); Urdu versus Bengali (Bangladesh), Sindhi, Saraiki or Pashto (Pakistan); Tamil versus Hindi (India), or Sinhalese (Sri Lanka), and so on This is not to mention languages mostly forgotten that nonetheless retain some measure of political potency, for example, Gaelic, Inuktitut, Welsh or Sanskrit. It is important to note, however, that Laitin does *not* claim that language division is politically unimportant. He finds that language is a powerful though instrumental tool of ethnic minority political operatives. Yet this point is essentially reduced to a caveat in a much larger body of work arguing there is no clear relationship between contentious language politics and *violent rebellions*. The inference is that language divides are not necessarily something to worry about.

Laitin is the leading practitioner of rational choice and game theory methodology to test prominent theories of comparative politics that posit ethnicity, culture and language are critical explanatory components of modernization, state development, national identity and especially communal conflict. This line of research dates to 1988 when Laitin (1988: 299; 1993) developed a typology of language conflicts based on the premise that the best way forward was to 'translate the monographic literature and other historical survey sources into preference functions that are empirically valid.'

He is regarded as a harsh critic of area studies (Munck and Snyder, 2007: 638) because he (2000a: 130) argues for the myopic premises of 'even the best and most informed case studies may be wrong,' a point later developed in a direct attack on erstwhile area-specific expertise (Laitin, 2005). Put another way, even the very best that area studies has to offer is often critically flawed. Laitin's view was cemented during his tenure as President of the Comparative Politics section of the American Political Science Association (APSA), from 1993 to 1995, due in no small part to a series of polemics on methodology he published as 'Letters from the President' in the section's newsletter.[1] Yet it is important to remember that it was not always thus; far from it. His first two books, *Politics, Language, and Thought: the Somali Experience* (1977) and *Hegemony and Culture: Politics and Religious Change among the Yoruba* (1986), offer a wealth of ethnographic observations, presented with narrative methods, and steeped in the ideas of anthropology, linguistics, social theory and philosophy (including Clifford Geertz, Benjamin Lee Whorf, Antonio Gramsci and Ludwig Wittgenstein, respectively). Thus, his later adoption and advocacy of game theory and statistical methods surprised (and sometimes alarmed) many of his contemporaries. In 2005, a symposium at the annual APSA meeting produced a range of papers on Laitin's methodological metamorphosis: Kanchan Chandra sees the phenomenon as a method-driven 'movement away from generalizing about outcomes to generalizing about mechanisms,' Ted Hopf observes that Laitin's relative use of ethnography compared with statistical analysis has inverted 'from equality to subordination to absence,' and Ashutosh Varshney concludes that Laitin's dismissal of single country studies 'amounts to denouncing his old scholarly self' (Hopf et al., 2006: 10, 17, 25).

Nonetheless, his field experience – in Somalia, Nigeria, Spain, Estonia and elsewhere – informs his personal evaluations regarding the relative ease of minority language *mobilization* because 'language is so intimately connected to group identity' (Laitin, 2000a: 113). But for Laitin, the approach of area studies and in-depth case analysis inevitably fails to account for a generalizable collective action conundrum. When a group identity is threatened, it is in the collective interest of the group to refuse any and all state attempts to assimilate the minority into the majority (or a subjugated majority into the state-sanctioned culture of a dominant minority). However, 'it would be individually rational for any particular member of the minority to assimilate,' especially in regard

to language since assimilation is possible with relative 'ease' (Laitin, 2000a: 113). Laitin (2000a: 113, n. 122) states that his 'research career has been devoted to this dilemma.'

A major product of this research was *Identity in Formation* (1998a), a robust exposition of how and why ethnic Russians may or may not choose to integrate or assimilate into the titular societies of four newly independent states: Estonia, Latvia, Kazakhstan and the Ukraine. In each case, he argues, these 'beached diasporas' would benefit materially by assimilating, by learning the language of a one-time Soviet minority and even sending their children to state schools where the language of education is not Russian.

In the western republics (Estonia, Latvia, the Ukraine), most Russians choose to remain, but in Kazakhstan many Russians choose to emigrate rather than assimilate. Laitin (1998a: 174) attributes this decision, which 'cannot be explained by a strict accounting of costs and benefits,' to intractable Russian racism and perceptions of Kazakh inferiority.[2] For Slavic Russians, 'learning a Turkic language and assimilating *downward* in a status hierarchy is a psychological impossibility' (Laitin, 1998a: 175). Hence, Laitin (1998a: 260) claims that 'incorporating status variables into a rational-choice framework will enrich that framework, enabling researchers to theorize more realistically about social and cultural change.' While more realistic, it is also methodologically troubling, at least to some reviewers. The anthropologist (and Provost of Brown University) David Kertzer (1999: 125) applauds Laitin for acknowledging the important effects of fiscally irrational beliefs that nonetheless 'allow for a better understanding of the course of social, political, and cultural change.' Yet he chides Laitin for attempting to 'salvage rational choice theory by freeing it from its material basis,' a move that undermines 'the very assumptions on which rational choice theory is based.'[3] Even if separated from a strictly material base, and reframed simply as the 'preferences of the actors,' Kaufman (2001: 204) demonstrates that those actors' preferences 'are rooted in their attitudes,' which derive primarily from what he calls 'symbolic politics,' that is, the politics of collective *historical myths*, rather than the politics of individual gain. In other words, subordinate language communities 'do not give up their linguistic heritage without a fight, even if the payoff is significant' (Safran, 2004: 2).

The question of individual rational actors choosing against their own economic interests to assimilate assumes, of course, that assimilation is an option. This requires setting aside evidence of racism. In the case of beached Russians in Kazakhstan, it was the minority that refused to join the majority. Yet arguably the converse is more common, that is, a majority that refuses to accept a minority. The notion of ethnic Uighurs in China successfully assimilating into Chinese society, for example, is dubious: no matter how fluent in Mandarin, Uighurs remain phenotypically distinct from Han Chinese and are subject to formal and informal discrimination (Amnesty International, 2004; Human Rights Watch, 2005; Zingg, 2002). The same can be said of Russians in Kazakhstan. These physical traits are what Gellner called *entropy-resistant*: his hypothetical example was 'a certain number of individuals who are, by an accident of heredity, pigmentationally blue.' If the blues are concentrated at either the top or the bottom of the social scale, and if their blueness persists over generations, this physical trait will 'constitute a very serious problem for industrial society' (Gellner, 1983: 65). (The African-American population of the USA is an obvious and immediate example.) The problem is the need of a modern state for seamless communication and social cohesion among its erstwhile homogenous citizens, whose unity is the *raison d'être* of the nation-state. This problem, as experienced by the Uighurs in Xinjiang and the Russians in Kazakhstan, respectively, presents two choices: fight or flight. On this point, at least, Gellner and Laitin should concur. This is particularly notable since Laitin (1998b: 137) is especially critical of Gellner's 'deeply flawed' structuralist theory of nationalism, arguing that his 'functionalism runs mad.'

Language Communities

Laitin's (2000a) 'Language conflict and violence: the straw that strengthens the camel's back' charges that a fundamental principle of the modernist paradigm, viz. that divergent language groups in the same state are expected to battle for supremacy, is not only wrong, but probably backwards. This argument is based on his modeling of language communities. This model is presented as an answer to his own rhetorical question: 'what is a language community?' (Laitin, 2000b). He argues that a pair of characteristics shapes a language community. The first is political: it is the *cohesion* of a language community, that is, the degree to which it is concentrated or diffuse, and is categorized according to a pair of language policy regimes. The second is linguistic: it is the relative *similarity* or *dissimilarity* of actual languages (including elements of speech, script or structure) that are used by two or more language communities in the same state.

Laitin (2000b: 153) rejects theories of nationalism, democracy and civil war that rely on what he views as 'vague and unspecified notions of cultural (or linguistic) heterogeneity.' Regarding this last point, he considers more defined and specific notions that rely on some index of ethnolinguistic fractionalization: there are now many, and all vary somewhat depending on the initial coding of 'ethnicity' or 'religion.' It has been more than three decades since Taylor and Hudson (1972) parsed ethnolinguistic scores from the *Atlas Narodov Mira* (1964) data, but only in recent years have really great strides been taken toward more refined techniques (Alesina et al., 2003; Brady and Kaplan, 2000; Campos and Kuzeyev, 2007; Cederman and Girardin, 2007; Posner, 2004). In regard to coding ethnicity or language, Laitin (2000b: 143) rejects earlier efforts that assume individuals 'map one-to-one onto ethnic or linguistic groups.' Instead, he proposes a coding of language cohesion that does not quite answer the question 'what is a language community?' but does offer a typology of language regimes. A language regime is either (1) *rationalized* or (2) *multilingual*. In either case, citizens are understood to share a range of languages that determine different language *repertoires*.

Laitin's adoption of the term 'rationalization' pointedly follows Max Weber's account of modern bureaucratization, that is, consistent rules that aid organization and efficiency within and across institutions. Laitin enumerates three methods of state language rationalization:

(1) the official adoption of an *acquired* lingua franca 'that is not associated as the mother tongue of a significant language-group living in that state' (for example, French in Senegal);
(2) the official adoption of a dominant *majority's* language (for example, Spanish in Spain);
(3) the official adoption of a dominant *minority's* language (for example, Afrikaans in Apartheid South Africa; Mandarin in Taiwan[4]).

In each case the outcome is a 'single language for educational and administrative communications,' that is, the official language of the state and therefore the only language to benefit from the unparalleled support of state institutions (Laitin, 2000b: 151).

The second type of language regime is multilingual, though there are two species of this genus. One variety (Laitin labels it M2) is typical of multinational federations wherein an official language is recognized by each region, so regional populations need know only one language, and there is no need to know the languages of adjacent regions. In other words, these are multilingual states with monolingual citizens. Laitin (2000b: 153) offers Belgium and Switzerland as archetypes. The other variety (called M1) is more complex and much harder to quantify. These are multilingual states with multilingual citizens, such as India. A native of Mumbai speaks Marathi

but must also learn Hindi and English; if they moved to Hyderabad in the state of Karnataka, they would be expected to learn Kannada (and their children would be required to learn it in public schools).[5]

For both kinds of multilingual states, which are of special relevance to research on nationalism and ethnic conflict, Laitin (2000b: 152) codes the degree to which a group of multilingual citizens actually constitutes a cohesive language community according to three measures: (1) whether *centrality* is higher or lower, depending on whether there is *one* language that is shared by more or fewer multilingual citizens; (2) the absence or presence of a 'normative rule' for who is expected to learn which language(s); and (3) a higher or lower incidence of language redundancy, because multilingual citizens from different regions will tend to share knowledge of one or more languages in their repertoire. This trio of measures may be analyzed as variables that determine whether a language community is coherent or fractured. The outcome of this analysis is then employed as a proxy for 'ethnic and cultural diversity.' As a more precise measure than earlier attempts to quantify diversity, Laitin (2000b: 143) offers this refined methodology to 'allow for useful statistical tests of received theory,' namely those which would test theories of nationalism, democracy and conflict by testing predictions based on higher or lower levels of heterogeneity in a state's population. At the time, this was the most thorough consideration of the problems and possibilities for measuring the cohesiveness of a language community.

Yet as a measure for quantifying heterogeneity, language communities still present a number of problems when employed as a proxy for ethnicity. By disaggregating one from the other, recent work by Anderson and Paskeviciute (2006) used a number of indexes to show that *linguistic* diversity is an unreliable predictor of *ethnic* diversity. They (2006: 788) claimed to demonstrate that 'the overall correlation between ethnic and linguistic heterogeneity is positive' but 'modest.'[6] In other words, the term 'ethnolinguistic' itself is not always appropriate. This does not, however, detract from the importance or utility of studying language in cases of divided multilingual societies.

Language Divides: Distance Versus Difference

Laitin endorses language as an empirically viable variable of ethnicity and culture. Though extraordinarily difficult to quantify, Laitin's position on the language proxy is akin to Churchill's estimation of democracy: it is the worst possible choice except for all others. Quantification problems aside, the use of language as a litmus test of communal identity is both theoretically and methodologically sensible. Theoretically, Laitin is again in accord with Gellner regarding language as a proxy for culture. A major difference between them, however, is that Gellner treats language and culture as essentially synonymous, arguing that language '*is* culture' (Gellner, 1964: 195). In contrast, Laitin (1998a: 368) argues that language is simply a preferred way to operationalize culture: 'I treat language as a proxy for culture and linguistic assimilation as an indicator of cultural assimilation.' This is sensible because formal and informal language adoption and (to a lesser extent) language behavior may be observed, and changes over time detected, so language is 'particularly kind to social scientists seeking a window on identity shift' (Laitin, 1998a: 368). The trouble, of course, is what to observe and how to measure characteristics of language groups.

There are many difficulties with language data generally, most of which are acknowledged forthrightly by Laitin. The most irksome is that 'people are notoriously bad reporters of their linguistic repertoires and behaviors. Census and survey reports on language abilities and language use are egregiously untrustworthy' (Laitin, 1998a: 368). This problem could be solved through other means of data collection, such as ethnographic observations of inter-personal communication. This

approach is applied typically in case studies or small-N comparative work, such as Laitin's four post-Soviet cases in *Identity in Formation*. Indeed, Rogers Smith faults most of Laitin's conclusions in this work because his 'rational choice model of identity formation contributes surprisingly little to understanding the processes of identity formation.' Smith (2004: 305–306) applauds the 'explanatory power' of Laitin's 'evidence of group attitudes' but is confounded because 'he gives no real theoretical account' of these attitudes. Yet marshaling such contextualized evidence at a large-N scale is entirely unlikely since the resources required would be fantastic.

Having defined and adopted a typology of language communities, and prescribed measures of relative community cohesion, Laitin then searches for a quantifiable determination of how similar or dissimilar is the relationship between or among one or more languages or dialects. A point of clarification: the distinction between a language and a dialect is political. Linguists define *language* 'in different ways according to different theories' (Matthews, 1997). *Ethnologue* – the standard reference of language classifications – separates languages and dialects depending on whether a pair of language communities share or do not share the same subjective 'ethnolinguistic identity' (Gordon, 2005). Even so, there are many examples of linguists and sociolinguists attempting to distinguish specific characteristics that separate distinct tongues. Laitin (2000b: 148) notes that these earlier methods, dating to the 1950s and Joseph Greenberg's (1956) use of word lists to estimate when language communities divided,[7] were ultimately rejected by linguists because related languages and dialects can diverge widely across word choice, pronunciation, structure and even separate forms for speech and text.

This does not, however, mean that these variables – vocabulary, grammar, phonics, script – are irrelevant to measuring language *difference*. These are entirely relevant and perfectly valid variables of comparison. The trouble is there are too many relationships between and among language pairs or groups, and there remain too many open questions of subjectivity and scale, especially regarding speech, understanding and acquisition. Accordingly, linguists from Denmark, Norway and Sweden are especially interested in measuring the degree to which people from different ethnolinguistic communities can understand some of each other's language *without study*. This phenomenon, called receptive multilingualism, is the focus of a five-year (2006–2011) research project, Linguistic Determinants of Mutual Intelligibility in Scandinavia, funded by the Netherlands Organisation for Scientific Research.[8] The disappointing implication of this active research is that there is no single, reliable and generalizable *measure* of language *difference*.

Hence, Laitin's decision to abandon the search for a useful measure of language *difference* is not unreasonable. Instead, he turns to a measure of language *distance*, that is, the genetic relationship of languages that share a common ancestor. Just as zoologists employ a Linnaean taxonomy to determine which animals are members of the same phylum or class, linguists employ a taxonomy of language families, such as Indo-European (German, Russian, Farsi) or Altaic (Mongolian, Turkish, Chechen). This is the technique employed by *Ethnologue* to classify each language.[9] Using these language family trees in *Ethnologue*, Laitin scored the relative distance, that is, the number of shared branches, connecting or separating antagonists in the MAR dataset.[10] This operationalization ultimately yielded a number of variables that Laitin employed in his analysis of language dis-/similarity and conflict detailed below. Before considering his findings, the utility of his method must be considered first.

The perspective of evolutionary biology is helpful in illustrating why ancestral distance is a very poor proxy of current difference. Consider the relationships between and among sharks, dolphins and dogs. On a phylogenetic tree, that is, a 'tree of life', one common ancestor shared by all three (and our own species) is the progenitor of the gnasthostomata branch of the group vertebrata, of the

Table 1. Laitin's Model of 'Language Similarity and Rebellion'. Language Distance and Rebellion (1945–1995): Comparison of Means

Language ancestry of minority versus dominant group: (Laitin and Fearon tripartite variable LANGSIM)	Mean value of rebellion (scale 0–7): (MAR variable REBELLION)*	*n*
All cases	2.49**	244
1. Different language family	2.06	111
2. Same language family	2.62	93
3. Same language	3.40	40

Source: Adapted by the author from Laitin (2000a, 103).
Notes: *The value of the scale moves from 0 (no rebellion) through guerrilla activity up to 7 (protracted civil war).
**Bivariate correlation of LANGSIM to REBELLION is positive (0.1359 at $p = 0.03$).

group craniata, of the group chordata, and so on back in time. The gnathostomes, as they are called, were dominant in the Middle Devonian age some 380 million years ago. About 20 million years later the sarcopterygii branched off and ultimately diverged into all four-limbed creatures, including dogs and (before their return to the sea) dolphins. According to the ancestral distance scheme, dogs and dolphins are *more related*, and therefore *more similar*, than sharks and dolphins. The many, many differences separating dogs and dolphins (fur, tails, panting, and so on) are ignored in this classification, as are the many similarities of sharks and dolphins (fins, swimming, piscivory, and so on). The only thing that matters here is when the lineage split.

In sum, language distance as a measure of language similarity is severely limited. Nonetheless, Laitin's method demanded a significant amount of data. Language distance data *is* available, and in the absence of any other method to measure language difference (and therefore any other data), it is this misleading measure of ancestry that underpins Laitin's analyses of language similarity and conflict. Laitin (2000b: 149) does acknowledge that distance is a 'rough and ready measure of language difference.' Yet despite this marginal mea culpa, language *distance* is then freely conflated with language *difference*, a move that is both significant and deleterious.

Nonetheless, 'Language conflict and violence: the straw that strengthens the camel's back' upended an accepted generalization about language community relations. Laitin did this by citing 'powerful evidence' to support findings that are 'quite stunning.' The very first finding presented is this: 'The greater the language difference between the language of the majority and that of the dominant group, the *lower* is the probability of violence' (Laitin, 2000a: 99). This sentence, in which difference alludes to his measure of distance, is presented plainly in the introduction of the article, and without qualification. The 'rough and ready' becomes tried and true.

Confusingly, language distance is also here called 'linguistic similarity' (Laitin, 2000a: 103) (see Table 1). His conclusion is based on a tripartite variable called LANGSIM (though LANG*DIS* would be more accurate) that codes MAR minorities and majorities according to whether the feuding groups: (1) speak entirely unrelated languages from distinct families; (2) speak distinct languages from the same family; or (3) speak the same language. (It is important to note here, for reasons discussed below, that all of these languages are vernaculars, that is, mother tongues, but they are not necessarily read, written or printed.) These three categories are used to separate three collections of conflicts. A MAR variable called REBELLION gauges conflict along an eight-point scale, from zero, meaning no rebellion, to seven, signifying protracted civil war. For the three collections of conflicts, the mean value of REBELLION is presented in a table that shows more 'similar' languages are more likely to engage in rebellion.

There are three reasons to challenge this claim. The first is that LANGSIM describes almost nothing about the actual differences between languages, such as the relative status of one to the other, whether there is a degree of mutual intelligibility, whether one is more easily acquired, or whether they share a writing system (assuming both languages have a writing system in the first place). Second, despite celebrating the connection of language 'similarity' and rebellion in the introduction, the actual evidence is presented with a major caveat (emphasis in original): 'without introducing controls, *the data show that greater linguistic similarity raises the probability of violence*' (Laitin, 2000a: 103). However, after controlling for demographic factors, such as whether a minority in question has an urban base, Laitin (2000a: 104) admits that language distance has '*no explanatory power.*' Third, there is no small number of prominent counterfactuals, including cases of ethnic conflict between language groups with no genetic relationship such as Kurds (Indo-European) versus Turks (Altaic) in Turkey, or Arabs (Afro-Semitic) in Iraq; Uighur (Altaic) versus Mandarin (Sino-Tibetan) in China; Chechen (Altaic) versus Russian (Indo-European) in Chechnya; Quecha (Quechan) versus Spanish (Indo-European) in Bolivia and Peru; Mayan (Mayan) versus Spanish (Indo-European) in Chiapas; Fur (Nilo-Saharan) versus Arabic (Afro-Semitic) in Darfur; Azerbaijani (Altaic) versus Armenian (Indo-European) in Nagorno-Karabakh and so on.

Finally, even if all methodological and empirical problems with the 'straw that strengthens the camel's back' are dropped, there remains a theoretical lacuna. Challenging the expectation that greater degrees of linguistic difference should raise the probability of conflict (and vice versa) is perhaps a duel with a straw man. The modernist paradigm of ethnonational conflict anticipates ethnolinguistic identification and ethnonational mobilization, but makes no real distinction between languages that are by most accounts very similar (for example, the mitosis of Serbo-Croatian) and languages that are extremely different (for example, Uighur and Chinese). The proposition that *more or less similar* languages mean more or less conflict is theorized neither by structural functionalists like Gellner, nor by rational choice practitioners such as Laitin.

Language Divides: MAR Variables' Ancestral Language Scores

Language distance, as a quantifiable variable, evolved far beyond the tripartite variable LANGSIM and came to occupy a new niche in the MAR dataset. The same methodology that produced LANGSIM was later used to produce a new variable called LANGFAM. Rather than a tripartite variable of language families, a more refined coding assigned a value to each language in a conflict pair from 1 to 20: a minority in a specific state is coded as 20 if they shared the same language as the majority; a minority from an entirely unrelated language family is scored as one. In between, values are assigned according to how many branches of the linguistic family tree are shared before they diverge across separate lines of descent. (See Table 2.)

A problem here is that the distances between English–French and English–Russian receive the same score. Genetically, the number of branches separating English from either language is the same. It is not unreasonable to suggest, however, that most Anglophones have a far easier time understanding French, in either its spoken or written forms, than Russian: in the spoken form, because of greater cultural familiarity and a great number of cognates; in the written form, if for no other reason, because French and English *share the same Latin alphabet* rather than Cyrillic. Skeptics may suggest that the scoring is not problematic, since an average American who speaks neither French nor Russian would, at any given moment, fail to comprehend either language: hence, this incomprehension could be coded equally. But this misses a glaring point about text: any literate Anglophone with no foreign language training can glean some meaning from French (from

Table 2. Language Distance Scoring – Ancestral Language Variable LANGFAM: Language Distance of English Versus Selected Sample Languages (Scoring Range: 1 (least related) to 20 (same language) – Points of Genetic Divergence are **bold**)

Language	Family (F = 1)	F+1 (Branch)	F+2	F+3	F+4	F+5	F+6	F+7	F+8	F+9	Score
English	Indo-European	Germanic	West	English	English						20
Scots	Indo-European	Germanic	West	English	**Scots**						5
German	Indo-European	Germanic	West	**High German**	German	Middle German	East Middle German	German: Standard			4
French	Indo-European	**Italic**	Romance	Italo-Western	Western	Gallo-Iberian	Gallo-Romance	Gallo-Rhaetian	Oïl	French	2
Russian	Indo-European	**Slavic**	East	Russian							2
Turkish	**Altaic**										1
Chinese: Mandarin	**Sino-Tibetan**										1

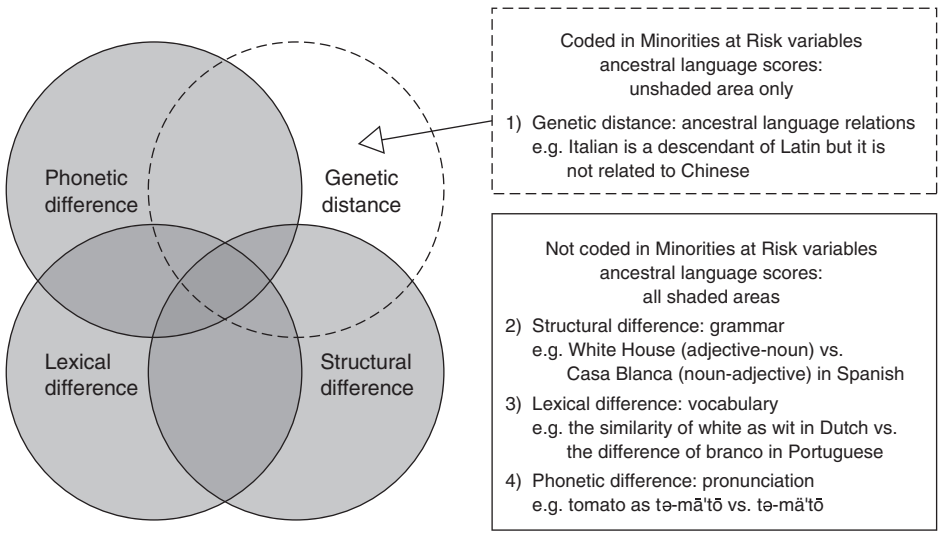

Figure 1. Four Domains of Language Distinction: Distance versus Difference

such cognates as *bleu*, *intelligence* or *musique*) while any language in a Cyrillic script is an absolute non-starter. Even more importantly, this objection also misses the point of language contact *over time* and the relative probability of successful state-sponsored assimilation. In this instance, the most useful measure is a score for relative *difficulty of acquisition*, such as how many classroom hours are required before the average student achieves functional capability across four domains of foreign language skill: speaking, understanding, reading and writing. Yet this kind of measure is possible only in *dyads*. Chinese is notoriously difficult for non-Chinese to learn, but is Chinese harder or easier to learn if your native language is French, Zulu, Kurdish, Mongolian, Basque, Tamil, Icelandic or Malay? In any one of these pairings, is listening comprehension easier than reading comprehension or vice versa?

Nonetheless, this troublesome variable, LANGFAM, is the first in a MAR suite collectively called ALS; the other four variables in the suite all rely on the same technique of simply scoring language distance.[11] The MAR codebook notes that the Laitin and Fearon variables were approved by the MAR Advisory Board as 'a more objective measure of group identity "distance"' but without acknowledging that the measure itself is both fractional and coarse.[12] It is coarse because it cannot account for actual differences between languages. It is fractional because it omits *three other categories of language characteristics* that are studied routinely when comparing specific languages (see Figure 1). Structural differences address syntax, that is, how words are ordered in speech.[13] Lexical differences are observed in vocabulary, including cognates (shared word origins) and borrowed words. Phonetic differences are in simple terms matters of pronunciation, though there are many human sounds that are found in some languages but not others, such as tonal languages or click languages. There is no effective way to quantify these three characteristics in a manner consistent for any two languages. Nonetheless, they are *essential* to determining actual language differences.

For example, according to the LANGFAM scheme, Maltese and the vernacular of Algeria (a species of Arabic spoken in the western Maghreb) are more related than even English and Scots.

Maltese and Algerian are both classified as Afro-Asiatic/Semitic/Central/South/Arabic languages, with a significant LANGFAM score of five. The differences between Maltese and Algerian, however, are substantial: Maltese 'is *descended* from Maghrebi Arabic but has borrowed heavily from Italian; it is a separately developed form with *different* syntax and phonology' (Gordon, 2005). Moreover, Maltese is a high status official language printed in a Latin script. In Algeria, spoken Arabic is not written at all; the official language is Standard Arabic written in the Arabic script. *None* of these differences are represented by the ALS variables. While omitting these measures from MAR is explicable because the differences themselves are practically immeasurable, it is misleading to present the suite of language distance variables to political scientists who are neither familiar with nor warned about the very great problems of conflating genetic distance with actual differences. The consequence of incorporating such data *prima facie* into any subsequent statistical analysis addressing ethnolinguistic fractionalization is, at best, questionable research.

Again, it is important to note that Laitin (2000b: 148–149) acknowledges a number of problems with *Ethnologue*'s coding of language distance, and concedes that language distance is *not* a reliable measure of mutual intelligibility: 'structural differences are not a good proxy for communicative differences ... "linguistic distance only in part determines communications" difficulties.' Noting that about one-third of Post-Soviet Russians in non-Slavic Estonia and Georgia were studying the titular languages of the republics, he suggested that '"language distance" does not act as an impediment to Russians learning the language of their new home.' In this case, Laitin's (1991) use of the term 'language distance' is correct, though his point is incomplete. While a large number of Russians may be studying new languages, this says nothing about the expected return on investment of studying a more-or-less similar language, what Etzioni (2008: 123) calls the 'labor to fluency ratio.'

Most critically, the measure cannot account for subjective language *status*: in regard to language planning and rationalization in a polyglot polity, sociolinguists have long argued that a critical quality that 'outranks intelligibility as a criterion for the choice' *is status* (Fasold, 1984: 36).[14] An influential political scientist also raised this point back in 1985. Writing on 'the conflict generating power of linguistic issues,' Donald Horowitz (1985: 219–224) argued that language is 'a symbol of domination' in a contest 'that is entirely relative to the status of others.' Subjective claims about language status may be informed by calculated payoffs, but they are also essentially normative, emotional, non-rational and 'not quantifiable.'

Theoretical Implications

At the beginning of this article it was noted that Laitin is critical of social science that elevates the importance of ethnic, linguistic or cultural divides as determinants of conflict. Instead, he argues that the material interests of individuals are more logical determinants of collective ethnic action or inaction, and that civil wars correlate with macro-characteristics of countries, not cultures. Hence, in Laitin's view, theories of ethnicity and nationalism, especially when applied by sociologists or area specialists, are exaggerated accounts of phenomena that are entirely explicable only when explained by the individual preferences of rational actors.

Yet the conventional wisdom of other large-N research regarding ethnicity and conflict is shifting. In the wake of 'Ethnicity, insurgency, and civil war', a number of enterprising researchers developed novel methods for answering a question posed by Fearon and Laitin, that is, what is the relationship of ethnic heterogeneity to conflict? In concert with Fearon and Laitin, Anderson and Paskeviciute (2005) suggest that ethnic heterogeneity may strengthen civil society as distinct

groups mobilize in the public sphere to represent their interests in a democratic polyarchy. This effect, however, is dependent on a degree of tolerance necessary to maintain such a dialogue. Laitin (2000a: 99) noted himself that 'those interested in peace should encourage the open expression of language grievances and the subsequent political bargaining over official languages and language of education.' Of course, this assumes there is some interest in making or keeping peace with people from a different community who currently share the same state.

But after cross-referencing a number of datasets, Anderson and Paskeviciute's multilevel (by individual and by country) analysis indicates this is not a secure assumption. They parse ethnic and linguistic heterogeneity with a conventional Hirschman–Herfindahl fractionalization index for the former and a linguistic diversity index for the latter. This second measure, compiled in *Ethnologue* (also known as Greenberg's diversity index), is the probability that any pair of people picked at random in the same country will or will not share the same tongue (Gordon, 2005). In this case, the number of speakers in a specific language community is measured as a proportion of a country's population. The political effects of ethnic versus linguistic heterogeneity are distinct, a finding theorized by the authors as follows: (1) 'the expression and reception of dissimilar views also requires that people trust one another'; (2) 'trust is in shorter supply in more heterogeneous societies'; (3) 'it is linguistic rather than ethnic heterogeneity' that lowers trust (Anderson and Paskeviciute, 2006: 799).

To be fair, this is a study of heterogeneity as a factor aiding or impeding the development of civil society, and is not meant to determine the probability of armed rebellion, per se. However, using a heteroskedastic model, Blimes (2006: 545) demonstrated an *indirect* effect of ethnic fractionalization on the specific probability of civil war:

> Fearon and Laitin's (2003) argument that conditions that favor insurgency lead to an increased risk of civil war is compelling … However, in a sample of countries that have low levels of ethnic fractionalization, most of their hypotheses do not hold up. *Ethnic conflict theory anticipates these results*, suggesting that when there are factors that make civil war more likely to occur, ethnic groups provide natural cleavages for society to fracture along and can help overcome collective action problems that may hinder a civil war from occurring in countries that do not have such ethnic groups.

In short, 'we do not know as much as we may think we do' about insurgency and civil war (Blimes, 2006: 545).

The kind of work pursued by Anderson, Paskeviciute, and Blimes indicates that the limits of large-N analysis in the study of language communities and conflicts are not yet in sight. Yet it is also important to recognize, again, that this is not an analysis of language difference. While the operationalization of all four dimensions of every language pair – phonetic, genetic, lexical and structural – remains out of reach (for reasons methodological, epistemological, fiscal and computational), the most important measure of language difference is *not* linguistic: it is *social*. The subjectively perceived *status* of rival languages is the most critical measure of ethnolinguistic mobilization. A low status language is politically unimportant, unless it aspires to higher status. A high status language is politically unimportant unless it is challenged, either by a rival high status language or by an ambitious low status language.

In many cases language status can be determined by opinion polling, especially in developed countries or in otherwise stable societies. Of course, survey data is entirely ill advised in deeply divided societies (Darfur, Sri Lanka and so on) or in multiethnic/multinational states where

questions about ethnic identity are an invitation to deportation (Uighurs or Tibetans in China, Kurds in Turkey, Chechens in Russia and so on). Nonetheless, even in areas where survey work is unsafe or logistically improbable, mainstream methods of social anthropology and especially sociolinguistics address this question as a matter of routine. There is already a wealth of data presented in these disciplines of enormous utility to political science, though again the methodology does not lend itself to large-N analysis for the simple reason that its collection is definitively small-N, as in a single country (or community) case study.

However, there is one characteristic of language that is a necessary condition for High Culture status that *is* also quantifiable: *print*. The importance of writing, printing and universal literacy is a central part of Gellner's path to industrialization, urbanization and ultimately nationalization, though the specific association of print with national identity is most commonly attributed to Anderson. His model of 'print-capitalism' links the mass-production methods of modern publishers to market forces: once a language is printed, the supply of printed material must grow to meet the demand of more and more readers, who consequently discover they are linked to others like themselves in an imagined community. But what if a group is unable to print material in its own language, either because the language is banned or because it is not (or is no longer) written? In some cases a language may have lost its print culture only to reclaim it later (Finnish, Gaelic); in others its loss may be viewed with ambivalence (Kashmiri); and in many cases there is no standardized written form of a widely used vernacular (the many Moro languages of Mindanao, or the many Roma varieties of Europe). In each case, the absence, presence, or loss of print is relatively easy to operationalize in a simple coding of 1, 2, 3. This is a proposition to be developed outside the constraints of this article.

Conclusion: Laitin, Language and Conflict

Laitin's corpus of research on ethnicity and language in relation to state and communal conflict is influential for two reasons: (1) there are relatively few political scientists who ask questions and seek answers specifically about language; and (2) he was the first and most prolific proponent of rational choice and game theory to challenge decades of political sociology scholarship warning of ethnic rifts as anathema to state stability. In his pursuit of more and better data, Laitin's determination to operationalize linguistic difference is both admirable and unfortunate. It is admirable because, as linguists and sociolinguists know all too well, quantifying a living language – including all of its lexical, structural, phonetic, genetic, social and political characteristics – is like trying to define the shape of smoke. Comparing one puff to another is a job for a statistical Sisyphus. It is unfortunate because Laitin's selection of ancestral language distance as a proxy for all other components of language difference is misleading and may easily distort the findings of researchers asking important questions about deeply divided societies or ethnic conflict.

In conclusion, the use of language as a proxy for culture in the study of ethnic conflict remains a valuable, even essential component of comparative political analysis. An attempt to quantify language difference according to language ancestry, as developed by David Laitin and James Fearon and incorporated in the MAR dataset, has produced data that is definitively fractional and therefore of little value. Furthermore, the most important political characteristic of any language community in contact with another is the relative social and political *status* of their two (or more) languages. Determining the subjective perspective of status is labor intensive, contextual, and often inter-subjective, but the payoff is potentially invaluable.

Acknowledgements

The author wishes to express his sincere thanks to Brendan O'Leary for his many helpful comments and suggestions, as well as John Gould, Elizabeth Mercurio, Marc Ross and Robert Vitalis. The author is also grateful to the Ethnicity, Nationalism and Migration section of the International Studies Association for its encouragement and support in developing this project.

Disclaimer

The opinions and analysis presented in this article are those of the author, and are not the official views of the Department of the Navy, the Department of Defense, or the US Government.

Notes

1. Back issues of the APSA-CP: Newsletter of the American Political Science Association Organized Section in Comparative Politics are available online at www.nd.edu/~apsacp/backissues.html
2. Despite describing this racism in detail, including mention of the Russian racial epithet denoting Kazaks as lice, Laitin does not actually use the word 'racism.'
3. Rational choice theorists could protest that the 'material basis' of the theory is a chimera, since the approach is more accurately a maximization of preferences. Still, in this case Laitin cites the irrational decision to embrace Russian chauvinism and eschew economic gain.
4. Taiwanese, formally classified as the Min Nan variety of Chinese, is the mother tongue of the majority in Taiwan. It is not mutually intelligible with Mandarin Chinese. For the majority, Mandarin is an acquired language; for the minority, who are descendants of the defeated Nationalist (KMT) forces that retreated to Taiwan from the mainland in 1949, Mandarin is their first language.
5. In fact, many educated workers migrating to work in hi-tech Hyderabad resent the fact that their children are expected to study in what they see as a low status language, Kannada, and instead send their children to unauthorized private schools that use English as the medium of instruction. The state government of Karnataka is closing many of these schools.
6. Their measure of Pearson's r is 0.38; $p = 0.011$.
7. Greenberg called his method glottochronology; for contemporaneous comparison, a similar method, called lexicostatistics, employs statistical analysis of cognates and borrowed words.
8. A project description is available – in English – on the website of Charlotte Gooskens at the University of Groningen (http://www.let.rug.nl/~gooskens/). She devotes much of her research to the challenge of determining mutual intelligibility.
9. *Ethnologue* denotes creoles and pidgins, but does not distinguish between languages and dialects for the reasons mentioned above. The ancestry of more than 6900 languages is available on the *Ethnologue* website (http://www.ethnologue.com).
10. Minorities at Risk Project (2005) College Park, MD: Center for International Development and Conflict Management. Available online at www.cidcm.umd.edu/inscr/mar
11. Minorities at Risk, Dataset Users Manual 030703, pp. 15–16. Available online at http://www.cidcm.umd.edu/mar/margene/mar-codebook_040903.pdf
12. Note that the MAR definition of Ancestral Language Scores, however, correctly identifies the variables as measures of language distance, not difference. Current members include: Victor Asal, Jóhanna K. Birnir, Dawn M. Brancati, Mary Caprioli, Jonathan Fox, Ted Robert Gurr, John Ishiyama, Patrick James, Erin K. Jenne, Michael Johns, Will H. Moore, Daniel Posner, Stephen M. Saideman, Monica Duffy Toft, Peter Trumbore and Stefan Wolff.
13. Interestingly, this is one measure of language difference that can be coded from *Ethnologue* data. Essentially, all languages use subjects (S), objects (O) and verbs (V) but the standard order of these

varies across six different combinations: SOV/OSV/SVO/OVS/VSO/VOS. This, however, is but one of many structural differences that divide syntactical practice.

14. Cited in Millar (2005: 3). In political science, the interplay of language status and language utility is profiled by Albaugh (2007) in comparative research on language of instruction policy in a number of African states.

References

Albaugh EA (2007) Language choice in education: a politics of persuasion. *The Journal of Modern African Studies* 45(1): 1–32.

Alesina A, Devleeschauwer A, Easterly W, Kurlat S and Wacziarg R (2003) Fractionalization. *Journal of Economic Growth* 8(2): 155–194.

Amnesty International (2004) *People's Republic of China: Uighurs Fleeing Persecution as China Wages its 'War on Terror'*. London: Amnesty International.

Anderson BR (1983) *Imagined Communities: Reflections on the Origins and Spread of Nationalism*. London: Verso Editions.

Anderson BR (1998) *The Spectre of Comparisons: Nationalism, Southeast Asia, and the World*. London: Verso.

Anderson CJ and Paskeviciute A (2005) Macro-politics and micro-behavior: mainstream politics and the frequency of political discussion in contemporary democracies. In: Zuckerman AS (ed.) *The Social Logic of Politics*. Philadelphia: Temple University Press, 228–248.

Anderson CJ and Paskeviciute A (2006) How ethnic and linguistic heterogeneity influence the prospects for civil society: a comparative study of citizenship behavior. *The Journal of Politics* 68(4): 783–802.

Barbour S and Carmichael C (eds) (2000) *Language and Nationalism in Europe*. Oxford: Oxford University Press.

Blimes RJ (2006) The indirect effect of ethnic heterogeneity on the likelihood of civil war onset. *Journal of Conflict Resolution* 50(4): 536–547.

Brady HE and Kaplan CS (2000) Categorically wrong? Nominal versus group measures of ethnic identity. *Studies in Comparative International Development* 35(3): 56–91.

Campos NF and Kuzeyev VS (2007) On the dynamics of ethnic fractionalization. *American Journal of Political Science* 51(3): 620–639.

Cederman LE and Girardin L (2007) Beyond fractionalization: mapping ethnicity onto nationalist insurgencies. *American Political Science Review* 101(1): 173–186.

Connor W (1994) *Ethnonationalism: the Quest for Understanding*. Princeton, NJ: Princeton University Press.

Etzioni A (2008) A global, community building language? *International Studies Perspectives* 9(2): 113–127.

Fasold RW (1984) *The sociolinguistics of society*. Oxford: Blackwell.

Fearon JD and Laitin DD (2003) Ethnicity, insurgency, and civil war. *American Political Science Review* 97(1): 75–90.

Fishman JA (1973) *Language and Nationalism: Two Integrative Essays*. Rowley, MA: Newbury House.

Gellner E (1964) *Thought and Change*. Chicago, IL: University of Chicago Press.

Gellner E (1983) *Nations and Nationalism*. Oxford: Blackwell.

Gellner E (1987) *Culture, Identity, and Politics*. Cambridge: Cambridge University Press.

Gordon RG, Jr (ed.) (2005) *Ethnologue: Languages of the World*, 15th edn. Dallas, TX: SIL International.

Green DP and Shapiro I (1994) *Pathologies of Rational Choice Theory: A Critique of Applications in Political Science*. New Haven, CT: Yale University Press.

Greenberg JH (1956) The measurement of linguistic diversity. *Language* 32: 109–115.

Hopf T, Herrera YM, Chandra K, Woodruff D, Varshney A and Laitin D (2006) Symposium: ethnography meets rational choice: David Laitin, for example. *Qualitative Methods* 4(1): 2–33.

Horowitz DL (1985) *Ethnic Groups in Conflict*, 2nd edn. Berkeley, CA: University of California Press.
Human Rights Watch (2005) Devastating blows: religious repression of Uighurs in Xinjiang. *Human Rights Watch* 17(2): 1–115.
Kaufman SJ (2001) *Modern Hatreds: The Symbolic Politics of Ethnic War*. New York: Cornell University Press.
Kertzer DI (1999) On rational choice and beached diasporas. *International Studies Review* 1(3): 121–125.
Laitin DD (1977) *Politics, Language, and Thought: The Somali Experience*. Chicago, IL: University of Chicago Press.
Laitin DD (1986) *Hegemony and Culture: Politics and Religious Change among the Yoruba*. Chicago, IL: University of Chicago Press.
Laitin DD (1988) Language games. *Comparative Politics* 20(3): 289–302.
Laitin DD (1991) The national uprisings in the Soviet Union. *World Politics* 44(1): 139–177.
Laitin DD (1993) The game theory of language regimes. *International Political Science Review* 14(3).
Laitin DD (1998a) *Identity in Formation: The Russian-speaking Populations in the Near Abroad*. Ithaca, NY: Cornell University Press.
Laitin DD (1998b) Nationalism and language. In: Hall JA (ed.) *The State of the Nation: Ernest Gellner and the theory of nationalism*. New York: Cambridge University Press, 135–157.
Laitin DD (2000a) Language conflict and violence: the straw that strengthens the camel's back. *Archives Européennes de Sociologie* XLI(1): 97–137.
Laitin DD (2000b) What is a language community? *American Journal of Political Science* 44: 142–155.
Laitin DD (2005) The Perestroikan challenge to social science. In: Monroe KR (ed.) *Perestroika! The Raucous Rebellion in Political Science*. New Haven, CT: Yale University Press, 115–135.
Laitin DD (2007) *Nations, States, and Violence*. Oxford: Oxford University Press.
Matthews PH (1997) *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.
Miklukho-Maklai Ethnological Institute (1964) *Atlas Narodov Mira*. Moscow: Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union.
Millar RM (2005) *Language, Nation, and Power*. Basingstoke: Palgrave Macmillan.
Monroe KR (2005) *Perestroika! The Raucous Rebellion in Political Science*. New Haven, CT: Yale University Press.
Munck GL and Snyder R (2007) *Passion, Craft, and Method in Comparative Politics*. Baltimore, MD: Johns Hopkins University Press.
O'Leary B (1998) Ernest Gellner's diagnoses of nationalism. In: Hall JA (ed.) *The State of the Nation: Ernest Gellner and the Theory of Nationalism*. New York: Cambridge University Press, 40–88.
O'Reilly C (ed.) (2001) *Language, Ethnicity, and the State*, Vol. 2. New York: Palgrave.
Posner DN (2004) Measuring ethnic fractionalization in Africa. *American Journal of Political Science* 48(4): 849–863.
Roshwald A (2006) *The Endurance of Nationalism: Ancient Roots and Modern Dilemmas*. Cambridge: Cambridge University Press.
Safran W (2004) Introduction: the political aspects of language. *Nationalism and Ethnic Politics* 10(1): 1–14.
Smith AD (1998) *Nationalism and Modernism: A Critical Survey of Recent Theories of Nations and Nationalism*. London: Routledge.
Smith R (2004) Identities, interests, and the future of political science. *Perspectives on Politics* 2(2): 301–312.
Taylor CL and Hudson MC (1972) *World Handbook of Political and Social Indicators*. New Haven, CT: Yale University Press.
Zingg E (2002) Ethnic Uighurs face language dilemma in China. *Agence France Press,* 3 November.

Biographical Note

Tristan James Mabry is a Research Assistant Professor in the Department of National Security Affairs at the Naval Postgraduate School in Monterey, California. His research addresses the comparative politics of ethnicity, identity, and separatism across Eurasia. He is the co-editor, with Brendan O'Leary and John McGarry, of *Divided Nations and European Integration* (University of Pennsylvania Press, forthcoming 2011). He is currently completing a book on the relationship between nationalism and Islam based on field research in Iraq, Pakistan, India, Indonesia, the Philippines, and elsewhere.

Disclaimer

The views expressed herein are those of the author alone and do not necessarily reflect the views of the United States Government, Department of Defense, or Department of the Navy.