

Spanish in the United States



Studies in Anthropological Linguistics

6

Editors

Florian Coulmas

Jacob L. Mey

Mouton de Gruyter
Berlin · New York

Spanish in the United States

Linguistic Contact and Diversity

Edited by

Ana Roca and John M. Lipski

Mouton de Gruyter
Berlin · New York 1993

Mouton de Gruyter (formerly Mouton, The Hague)
is a Division of Walter de Gruyter & Co., Berlin

⊗ Printed on acid-free paper which falls within the guidelines of the
ANSI to ensure permanence and durability.

Library of Congress Cataloging-in-Publication Data

Spanish in the United States / edited by Ana Roca and John M. Lipski.
p. cm. — (Studies in anthropological linguistics ; 6)

Includes bibliographical references and index.

Contents: Linguistic heterogeneity, civil strife, and per capita gross
national product in inter-polity perspective / Joshua A. Fishman — A
reconsideration of the notion of loan translation in the analysis of
U.S Spanish / Ricardo Otheguy — The dialectics of Spanish language
loyalty and maintenance on the US-Mexico border / Margarita Hi-
dalgo — Spanish clitics in a contact situation / Manuel J. Gutiérrez
and Carmen Silva-Corvalán — Language choice in Hispanic-back-
ground junior high school students in Miami / Barbara Zurer Pearson
and Arlene McGee — Literacy stories / René Cisneros and Betty
Leone — Language maintenance institutions of the Isleño dialect of
Spanish / Felice Anne Coles — Convergent conceptualizations as pre-
dictors of degree of contact in US Spanish / Ricardo Otheguy and
Ofelia García — Creoloid phenomena in the Spanish of transitional
bilinguals / John M. Lipski — Diversification and Pan-Latinity /
Frances R. Aparicio — Oral proficiency testing and the bilingual
speaker / David Barnwell.

ISBN 3-11-013204-4 (cloth : alk. paper)

ISBN 3-11-016572-4 (pbk.)

1. Spanish language — 2. Social aspects — United States. 2. Socio-
linguistics. 3. Languages in contact — United States. 4. Bilingualism
— United States. I. Roca, Ana. II. Lipski, John M. III. Series.

PC4826.A42 1993

306.4'4'0973—dc20

93-14956

CIP

Deutsche Bibliothek — Cataloging-in-Publication Data

Spanish in the United States : linguistic contact and diversity / ed. by
Ana Roca and John M. Lipski. — Berlin ; New York : Mouton de Gruy-
ter, 1999

(Studies in anthropological linguistics ; 6)

ISBN 3-11-016572-4

NE: Roca, Ana [Hrsg.]; GT

© Copyright 1993 by Walter de Gruyter & Co., D-10785 Berlin.

All rights reserved, including those of translation into foreign languages. No part of this book may be
reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy,
recording or any information storage and retrieval system, without permission in writing from the
publisher.

Typesetting and Printing: Arthur Collignon GmbH, Berlin —

Binding: Lüderitz & Bauer GmbH, Berlin. — Printed in Germany.

Acknowledgements

We wish to take this opportunity to thank the Florida Endowment for the Humanities and Florida International University for their generous support of the Ninth National Conference on Spanish in the United States, held in Miami in 1988. The support of the university and of the Florida Endowment for the Humanities enabled Professor Ana Roca to organize a national forum titled "English as the Official Language: Advance or Retreat?" as part of the conference.

We would like to give our sincerest thanks to the collection's contributors, an excellent group of scholars with whom it has been a pleasure to interact. We thank both U. S. Hispanic bilingual students as well as non-native speakers of Spanish who make the effort and dedicate the time it takes to become more proficient and literate in the second most important language in their country.

Contents

Acknowledgements	V
Introduction	1
<i>Joshua A. Fishman</i>	
Linguistic heterogeneity, civil strife and per capita gross national product in inter-polity perspective	9
<i>Ricardo Otheguy</i>	
A reconsideration of the notion of loan translation in the analysis of U.S. Spanish	21
<i>Margarita Hidalgo</i>	
The dialectics of Spanish language loyalty and maintenance on the U.S.-Mexico border: A two-generation study	47
<i>Manuel J. Gutiérrez and Carmen Silva-Corvalán</i>	
Spanish clitics in a contact situation	75
<i>Barbara Zurer Pearson and Arlene McGee</i>	
Language choice in Hispanic-background junior high school students in Miami: A 1988 update	91
<i>René Cisneros and Elizabeth A. Leone</i>	
Literacy stories: Features of unplanned oral discourse	103
<i>Felice Anne Coles</i>	
Language maintenance institutions of the <i>Isleño</i> dialect of Spanish	121
<i>Ricardo Otheguy and Ofelia Garcia</i>	
Convergent conceptualizations as predictors of degree of contact in U.S. Spanish	135
<i>John M. Lipski</i>	
Creoloid phenomena in the Spanish of transitional bilinguals ..	155

Frances R. Apacicio

Diversification and Pan-Latinity: Projections for the teaching of Spanish to bilinguals	183
---	-----

David Barnwell

Oral proficiency testing and the bilingual speaker	199
--	-----

Index	211
-------------	-----

Introduction

The idea for this collection stems from our research interest in the linguistic aspects of Spanish as used in the United States today. Tied to our academic interest has been our professional participation in *El Español en los Estados Unidos*, a national conference in linguistics begun by Professor Lucía Elías-Olivares in 1979 and first held at the University of Illinois-Chicago Circle.

El Español en los Estados Unidos has become an annual event, hosted by universities around the country, including the University of Texas (Austin), the University of New Mexico (Albuquerque), the University of Iowa (Iowa City), Indiana University (Bloomington), Hunter College of the City University of New York, Florida International University (Miami), the University of Arizona (Tucson), and the University of Southern California (Los Angeles). *El Español en los Estados Unidos* has dealt with linguistic aspects of Spanish in the United States, as well as with selected cross-disciplinary issues. This meeting provides scholars with a professional forum for the exchange of ideas, as well as the presentation of completed studies and research in progress. Hard data, valuable insights, incisive discussions, and the dissemination of research findings (Amastae—Elías-Olivares 1982; Elías-Olivares 1983; Elías-Olivares et al 1985; Wherritt—Garcia 1988; Bergen 1990; Roca—Lipski 1992), are among the many achievements of this conference.

The present volume makes an additional contribution to the growing research bibliography on linguistic aspects of Spanish in the United States.¹ Some of the contributors are established scholars in the field, while others are entering the field with promising studies.

As the Hispanic population continues to expand, so has the body and scope of linguistic research on the Spanish spoken in the United States. The earliest studies concentrated on rural Mexican American Spanish in the Southwest. The parameters of the field have been extended in response to changing urban and demographic realities, and to paradigm shifts in linguistics itself. Attention is no longer confined to the traditional domains of phonology, morphology, syntax, and semantics. Today, a growing number of linguists are concerned with the intricate and complex issues relating language to society, education, ethnicity, the media, and politics.

2 *Introduction*

What has become clear in the last twenty years is that the study of Spanish in the United States cannot and should not be divorced from a cross-disciplinary understanding of its context in time, place, and history.

The Hispanic population continues to be the largest and fastest growing linguistic minority in the United States. We can no longer approach linguistic research along narrow lines, thinking only, for example, in terms of the Spanish spoken by Mexican Americans in the five Southwestern states where the majority reside (California, Texas, Arizona, Colorado, and New Mexico). Likewise, we cannot continue to limit the study of Puerto Rican Spanish to New York, nor of Cuban Spanish to Florida.

With ever-increasing numbers of Hispanics arriving in the U. S. from many different nations, the demographic profile today is far more complicated than twenty, or even ten years ago. Just like the Latino population itself, the Spanish spoken in this country is not homogeneous. Spanish speakers in the United States come from different generations, nationalities, and cultures. The population also reflects varied levels of schooling, experience, economic power, and status in the work force. Linguistically, this population includes a wide range of receptive and productive skills in English and Spanish.

Spanish was first used in what is now the United States as early as the 1530s, preceding the English-speaking settlers of New England. In much of the Southwest, Spanish has been used continuously since then, before and after this region was annexed to the United States. The first large-scale influx of Spanish speakers from outside the modern United States occurred at the beginning of the 20th century, as a result of the turmoil caused by the Mexican Revolution. Many found work in mines, agriculture, or in the railroad industry. However, when the depression of the 1930s hit the country and unemployment was at its peak, many Mexican Americans, including some who were U. S. citizens, were deported to Mexico.

Puerto Rico became a part of the United States at the close of the Spanish-American War in 1898, and Puerto Ricans became U. S. citizens right before the U. S. entered World War I. Partly as a result of poor economic conditions on the island — increasingly under more U. S. control — the emigration flow reached its peak after World War II. From 1945 to 1955 over 50,000 Puerto Ricans would leave the island annually to come to the United States. Although the majority of them reside in New York, there are Puerto Ricans in many other parts of the Northeast, as well as in other regions of the country.²

Although there were Cubans living in places like Key West, Tampa, and New York since the latter part of the 1800s, it was not until after Fidel Castro took over the government in 1959 that large waves of exiles began to arrive in South Florida in the early sixties. With over a million Cubans in the United States, this population forms the largest Hispanic group in the U. S. after Mexican Americans and Puerto Ricans. In comparison to these last two, though, much less has been published on the language situation of Cuban Americans, most likely because it is a much more recent phenomenon.

Today, serious problems caused by overpopulation and poverty in Mexico, as well as the devastation and economic instabilities resulting from recent civil wars in Central America (as in Nicaragua and El Salvador, for example), have become social forces which have helped to increase the number of Hispanic immigrants entering the United States. The availability of legal and illegal seasonal agricultural work and other types of work, has served as economic bait for people who left their homeland for political or economic reasons. Additionally, the dream of a better life in "el Norte", where there already exists a growing population of Hispanics with whom immigrants could identify and communicate with, has offered psychological incentives for coming here. These and other factors have played a role in increasing legal and illegal immigration of Mexicans and other Latin Americans.

At the same time, research is also expanding to include the study of linguistic aspects of the Spanish of the multiple subgroups which make up the more varied and complex Hispanic population, which now includes an increasing number of Spanish speakers from the Caribbean and Central America. Most recently, for example, the U. S. receives regular arrivals of "balseros" from Cuba, who are so-named because they sail to South Florida in small makeshift rafts, or *balsas*. Nicaraguans, Salvadorans, Guatemalans, Dominicans, and Colombians are among other groups which increasingly enter the U. S. with and without visas. Although Mexican Americans make up over 63 percent of the Hispanic population in the United States the most evident increase in the Hispanic population growth rate between 1982 and 1987 was due to the Central and South American immigrant groups coming into the country (Bureau of the Census 1987).

Our collection opens with an insightful essay by Joshua A. Fishman called "Linguistic Heterogeneity, Civil Strife and Per Capita Gross National Product in Inter-Polity Perspective". In his essay (also the keynote address of the conference), Professor Fishman summarizes previous stud-

ies and discredits some of the claims of major negative consequences that are commonly said to follow from linguistic heterogeneity and which we commonly find in the popular media and in political discussions: the claims that linguistic heterogeneity leads to or exacerbates civil strife and lowers national productivity. Professor Fishman's essay examines methodological considerations and available data in the literature as he compares popular thinking with sociolinguistic research in this connection.

Ricardo Otheguy's paper examines the notion of loan translation in the analysis of U. S. Spanish. He argues that loan translation is a defective construct which hinders rather than illuminates discussions on the process of language contact.

Margarita Hidalgo's paper is an excellent sociolinguistic, two-generation study on the question of linguistic assimilation vs. linguistic loyalty and maintenance in Chula Vista, a border city in Southern California, which has now become the second fastest growing city of San Diego County.

In "Spanish Clitics in a Contact Situation", Manuel J. Gutiérrez and Carmen Silva-Corvalán examine clitics in Spanish in order to take a look at transfer phenomena in the bilingual context of the city of Los Angeles, California.

Barbara Pearson and Arlene McGee's paper reports on language choice in Hispanic background junior high school students in Miami. In their report, Pearson and McGee offer a review of the literature and raise issues regarding the differences in the degree of proficiency and use of Spanish by generation and by domain.

In "The *Isleño* Dialect of Spanish: Institutions for Language Maintenance", Felice Ann Coles examines an isolated and dying language spoken by only a few residents of St. Bernard Parish, Louisiana. Coles reports on the institutional language maintenance efforts which the *Isleños* have coordinated as a way of "revitalizing" their sense of ethnicity.

In their incisive paper, Ricardo Otheguy and Ofelia García discuss "Conceptualizations as Predictors of Degree of Contact in U. S. Spanish". They examine the idea of conceptual uniqueness of languages in contact phenomena, arguing that conceptualizations that are generally found among English-speakers in North America, but not among Spanish-speakers in Latin America, have been diffused into the culture of Latinos in the United States (page 135).

In "Creoloid Phenomena in the Spanish of Transitional Bilinguals", John M. Lipski underlines the problems related to the term *semi speakers*

and explains why explains why he opts for the more neutral term *transitional bilinguals* (TB). Lipski points out the lack of even a rough estimate of the proportional of TB Spanish speakers in the U. S., either in the schools or in society as a whole, and describes a combination of the features which would give a reasonable prediction of TB status.

In "Diversification and Pan-Latinity: Projections for the Teaching of Spanish to Bilinguals", Frances R. Aparicio regards not only how the profile of the Hispanic student in the United States has been rapidly changing, but how this fact underlines the need to re-examine some of the controversial pedagogical questions in the teaching of Spanish to U. S. Hispanic bilingual students of diverse cultural and linguistic backgrounds.

David Barnwell's essay, titled "Oral Proficiency Testing and the Bilingual Speaker", addresses questions regarding the use of the ACTFL Oral Proficiency interview to measure the language proficiency of U. S. Hispanic bilinguals.

The final paper, "Literacy Stories: Features of Oral Discourse", by René Cisneros and Elizabeth A. Leone, analyzes a broad range of features of spoken, unplanned discourse, in order to gain insight into the ways U. S. Spanish language users actually speak.

In putting together this volume, we aim to provide the reader with linguistic discussions and questions which reflect the current research trends in the field. To borrow a phrase coined by linguistics professor Garland Bills, we now must speak of "the many faces of U. S. Spanish". In order to acquire a truer picture of Hispanics in the United States today, in order to better understand the intricacies of the language in contact with English, we must have a more encompassing perspective of the cultural and linguistic varieties of the Spanish-speaking people in both urban and rural America. It is our hope that the perspectives presented in the volume's discussions will stimulate interested scholars to carry out research which remains to be done.

Notes

1. A tangible reflection of research on Spanish in the United States over the past quarter century is the growing bibliography of seminal monographs and anthologies, including *Bilingualism in the Barrio* (Fishman, Cooper, Ma, et al.

- 1971), *El lenguaje de los chicanos* (Chavez, Cohen, and Beltramo, eds. 1975), *Studies in Southwest Spanish* (Ornstein and Bowen, eds. 1976), *Chicano sociolinguistics* (Peñalosa 1980), *Spanish in the United States; Sociolinguistic aspects* (Amastae and Elías-Olivares, eds. 1982), *Chicano discourse* (Sánchez 1983), *Spanish in the United States; Beyond the Southwest* (Elías-Olivares, ed. 1983), *Spanish and English of United States Hispanics: A critical annotated bibliography* (Teschner, Bills, and Craddock, eds. 1985), *Spanish language use and public life in the U. S. A.* (Elías-Olivares, et al. 1985), *Research issues and problems in United States Spanish* (Ornstein, Green, and Márquez, eds. 1988), *Spanish in the U. S.: The language of Latinos* (Wherritt and García, eds. 1989), *Spanish in the United States: Sociolinguistic aspects* (Bergen 1990). Although not exclusively about Spanish in the U. S., *Bilingualism and language contact: Spanish, English, and Native American languages* (Barkin and Brandt, eds. 1982), provides insightful studies. Two additional collections include excellent overviews and articles on the subject. These are *language diversity: Problem or resource? A social and educational perspective on language minorities in the United States* (McKay and Wong, eds. 1988) and *Sociolinguistics of the Spanish-speaking world: Iberia, Latin America, United States* (Klee and Ramos-García, eds. 1991).
2. For an excellent overview of “The Language Situation of Puerto Ricans”, see Ana Celia Zentella’s essay of the same title in *Language diversity*. Also, for an equally informative and useful overview on “The Language Situation of Cuban Americans”, see Ofelia García’s essay in the same collection and Ana Roca’s response on the issue of “Language Maintenance and Language Shift in the Cuban American Community of Miami”, in *Language planning* (Marshall, ed. 1991). *Language diversity* also includes a valuable overview essay by Guadalupe Valdés on the language situation of Mexican Americans.

References

- Amastae, Jon — Lucía Elías-Olivares
 1982 *Spanish in the United States: Sociolinguistic aspects*. Cambridge: Cambridge University Press.
- Barkin, Florence — Elizabeth A. Brandt — Jacob Ornstein-Galio (eds.)
 1982 *Bilingualism and language contact: Spanish, English, and Native American languages*. New York: Teachers College Press, Columbia University.
- Barnach-Calbó Martínez, Ernesto
 1980 *La lengua española en los Estados Unidos*. Madrid: Oficina de Educación Iberoamericana.

Bergen, John

- 1990 *Spanish in the United States: Sociolinguistic aspects* Washington: Georgetown University Press.

Elías-Olivares, Lucía (ed.)

- 1983 *Spanish in the U. S. setting: Beyond the Southwest*. Rosslyn, VA: Clearinghouse for the National Association for Bilingual Education.

Elías-Olivares, Lucía — Elizabeth A. Leone — René Cisneros — John Gutiérrez (eds.)

- 1985 *Spanish use and public life in the USA*. (Contributions to the Sociology of Language 35.) Berlin: Mouton de Gruyter.

Fishman, Joshua A. — Robert L. Cooper — Roxana Ma

- 1971 *Bilingualism in the Barrio*. (Language Science Monographs 7). Bloomington: Indiana University Press.

García, Ofelia

- 1988 "The language situation of Cuban Americans", in: Sandra Lee McKay — Sau-ling Cynthia Wong (eds.), 166 — 192.

Hernández-Chavez, Eduardo — Andrew D. Cohen — Anthony D. Beltramo

- 1975 *El Lenguaje de los Chicanos: Regional and social characteristics used by Mexican Americans*. Arlington, VA: Center for Applied Linguistics.

Klee, Carol — Luis A. Ramos-García (eds.)

- 1991 *Sociolinguistics of the Spanish-speaking world: Iberia, Latin America, United States*. Tempe, Arizona: Bilingual Review Press/Editorial Bilingüe.

McKay, Sandra Lee — Sau-ling Cynthia Wong, (eds.)

- 1988 *Language diversity: Problem or resource? A social and educational perspective on language minorities in the United States*. New York: Newbury House.

Ornstein, Jacob — J. Donald Bowen

- 1976 *Studies in Southwest Spanish*. Rowley, MA: Newbury House.

Ornstein, Jacob — George K. Green — Dennis J. Bixler-Márquez

- 1988 *Research issues and problems in United States Spanish*. Pan American University at Brownsville, Texas, in cooperation with the University of Texas at El Paso.

Peñalosa, Fernando

- 1980 *Chicano sociolinguistics*. Rowley, MA: Newbury House.

Roca, Ana

- 1991 "Language maintenance and language shift in the Cuban American community in Miami: The 1990s and beyond", in: David F. Marshall (ed.), *Language planning: Focusschrift in honor of Joshua A. Fishman*. Amsterdam: John Benjamins.

Sánchez, Rosaura

- 1983 *Chicano discourse: Socio-historic perspectives*. Rowley, MA: Newbury House.

8 Introduction

Teschner, Richard – Garland Bills – Jerry Craddock (eds.)

- 1985 *Spanish and English of United States Hispanos: A critical, annotated, linguistic bibliography*. Arlington, VA: Center for Applied Linguistics.

Valdés, Guadalupe

- 1988 The Language Situation of Mexican Americans, in: Sandra Lee McKay – Sau-ling Cynthia Wong (eds.), 111 – 139.

Wherritt, Irene – Ofelia García (eds.)

- 1989 *US Spanish: the language of Latinos*. (*International Journal of the Sociology of Language* 79.)

Zentella, Ana Celia

- 1988 “The language situation of Puerto Rican’s”, in: Sandra Lee McKay – Sau-ling Cynthia Wong (eds.), 140 – 165.

Linguistic heterogeneity, civil strife and per capita gross national product in inter-polity perspective

Joshua A. Fishman

Among the major charges against linguistic heterogeneity that are encountered in the popular press in political discussions are the claims that it leads to or exacerbates civil strife, on the one hand, and that it lowers national productivity, on the other hand, in both cases because linguistic heterogeneity presumably counteracts rationality, civility, sensitive communication and the smooth operations of government and industry alike. These are charges to which some of my own recent research pertains¹ and it may prove instructive to compare popular thinking and sociolinguistic scholarship in this connection.

Each of the foregoing charges can be translated into a formal inter-polity hypothesis, namely:

(a) The greater the degree of linguistic heterogeneity in a country ("degree of linguistic heterogeneity" being operationalized as the proportion of the population claiming as its own the major mother tongue of any given country, the smaller that proportion the greater the degree of linguistic heterogeneity, and, correspondingly, the larger that proportion the smaller the degree of linguistic heterogeneity), the greater the frequency and severity of civil strife in that country, and, similarly:

(b) the greater the degree of linguistic heterogeneity in a country, the lower the per capita gross national product in that country.

Methodological considerations

Until quite recently it would have been virtually impossible to do conclusive, worldwide, empirical research on hypotheses such as the above because of the large number of countries (= polities) and the large number of additional variables that need to be examined in order to rigorously test these hypotheses. There are approximately 170 polities in

the world today and if these were simply to be compared two at a time, in order to determine whether the linguistically more homogeneous one differs significantly (with respect to severity/frequency of civil strife and/or with respect to per capita gross national product) from the linguistically less heterogeneous one, over 25,000 individual comparison would have to be made. Obviously, it would be both inordinately difficult to undertake and then to make sense out of so many comparisons.

Actually, however, the methodological problem indicated above for two variables and 170 countries is compounded many times over if we realize that in order to test our hypotheses we also need to simultaneously consider many, many other variable that are descriptive of the countries of the world *in addition to* the two that we are focusing upon. What we really want to know is whether linguistically more heterogeneous and less heterogeneous polities differ in connection with civil strife and per capita gross national product *over and above* (independently of) the differences between such countries due to any and all other factors to which civil strife and per capita gross national product may be indirectly related.

Civil strife, e. g., should be considered too when we are looking into the relationship between degree of linguistic heterogeneity and per capita gross national product. This is necessary so that we can tell whether any encountered relationship between degree of linguistic heterogeneity and per capita gross national product, whatever that may, is masked by or even due to the relationship between civil strife and per capita gross national product. And the same is also true, of course, with respect to degree of religious heterogeneity, degree of racial heterogeneity, proportion of the annual budget allocated to military expenses, etc., etc. *Only if we can also consider all other possibly contributing variables can we tell whether linguistic heterogeneity per se really makes an independent (i. e. a non-redundant) contribution to per capita gross national product.* However, there are an almost endless number of such other possibly contributing variables (indeed, political scientists have perfected 230-some different dimensions (238 to be exact), all in all, for describing countries) and all of these need to be utilized simultaneously, together with linguistic heterogeneity, when attempting to account for inter-polity differences in civil strife or in per capita gross national product.

Thus, our task is to compare all countries simultaneously on all variables simultaneously if we really want to find out whether degree of linguistic heterogeneity is *a truly independent (necessary, non-redundant) correlate* of either civil strife or per capita gross national product. The price of bananas and the number of gloves sold on a particular day may

correlate substantially. However, only if we include all other variables that might also possibly influence the cost of bananas on a particular day (e. g., average daily temperature in the banana groves, daily transportation costs between the groves and the markets, labor costs in the groves and in the markets, etc., etc.) can we safely avoid coming to the specious conclusion that the number of gloves sold is really genuinely (that is: independently) related to the cost of bananas. Does this sound like an impossibly tall order: to analyze hundreds of countries and hundreds of variables simultaneously? Do we have the necessary data in order to do that and do we have the necessary methods by means of which to do that?

Data and analytic methods

Fortunately, the variety and even the quality of the data we need has been provided by the cross-polity databanks that American political scientists, both in government and in academia, have prepared, and repeatedly revised and expanded during the past quarter century. These databanks provide sifted, corrected and continually updated data on all the countries of the world in conjunction with over one hundred thirty different economic, political, social, cultural, historical, geographic and demographic variables. This data is not perfect, but it is the best available today anywhere in the world and since quite a bit of American economic, political and military planning and policy is based on this data it must at least be reasonably good on the whole and may even be quite a bit better than that.

The analytic methods to do what needs to be done have been provided by statisticians and computer specialists who have relatively recently perfected approaches (primarily cumulative multiple correlation and factor analysis) that make it relatively easy, on the one hand, to examine huge amounts of multivariate data and, on the other hand, to parsimoniously zero in on the relatively few variables in any large data-set that are really the only independent (and, therefore, the only crucial) variables in explaining or in accounting for the variation in any given criterion variable. My co-workers and I are, I believe, the first to put both the exhaustive data-sets and the new statistical analytic methods to joint use in conjunction with determining the role of inter-polity variation in

linguistic heterogeneity in so far as accounting for cross-polity variation in civil strife and in per capita gross national product.

Perhaps an apology is in order for the above brief detour into methodological issues, which no matter how brief it may be to the specialist inevitably seems overly long and insufficiently understandable to the non-specialist. My concern is basically related to the usual scholarly preoccupation that findings, interpretations and conclusions rest upon foundations that are as firm as possible. It is also related to an attempt to get away from a contrasted approach which may be referred to as “the favorite country approach”. We are all good, journalists, politicians and academics alike, at arguing from “preferred cases”, that predictably provide *negative* answers to the questions we have initially posed about linguistic heterogeneity’s possibly harmful consequences by examining them only in connection with Switzerland (where linguistic heterogeneity results in neither heightened civil strife nor in lowered per capita gross national product) or that answer the same questions in the *positive* by referring only to India, without even pausing to consider the many other dimensions (besides linguistic heterogeneity) on which Switzerland and India differ substantially and differ in ways that are directly related to civil strife and/or per capita gross national product. It is to escape from this more usual approach of arguing from “preferred (and biased) cases” that I have gone out of my way to study all variables and all countries simultaneously in order to clarify the true (i. e., the independent) relationship between linguistic heterogeneity, civil strife and per capita gross national product.²

Civil strife³

Political scientists have kept records on civil strife in all of the countries of the world in four different ways: (a) magnitude and frequency of conspiracy against the established government, (b) magnitude and frequency of internal warfare due to revolution, sedition or secession, (c) magnitude and frequency of internal turmoil (riots, strikes, protests) and (d) a composite average of the above three.⁴ The latter is the only measure of civil strife that will be discussed in this paper, although essentially identical results obtain from analyses of the contribution of linguistic

heterogeneity to civil strife as measured by each of the other measures of civil strife as well.

A cumulative multiple correlation analysis which begins by aiming over 230 variables at the composite civil strife criterion variable reveals that the differences in composite-magnitude-and-frequency of civil strife across all of the countries of the world are both highly predictable and parsimoniously predictable. Out of 230 — some possible predictors of this criterion only a certain 13 make truly independent contributions that, taken together, yield the optimal multiple correlation of .82. Particularly noteworthy for our purposes is the fact that *linguistic heterogeneity is not a member of this optimal subset of predictors*. Of course, linguistic heterogeneity does have a certain correlation with (.21) with civil strife and, therefore, if we were to disregard, for a moment, its redundancy with other variables it would, at most, explain only 4% (or .21 squared) of the total variance in civil strife. However, when linguistic heterogeneity is confronted with all other possible predictors of civil strife it is eclipsed entirely and it becomes clear that it itself explains nothing at all in connection with civil strife that is not explained better (i. e., less redundantly or more independently) by other variables. What are the variables that displace linguistic heterogeneity from any consideration as an independent predictor of civil strife? They are variables such as (a) short term deprivation, (b) persistent deprivation, (c) the absence of coercive potential by the government and (d) the presence of anomic groups that are organized to exert pressure for their own particular benefit. Just a few variables such as these immediately come to the fore to form the relatively small optimal subset of 13 indispensable predictors that, when taken together, account for 68% (= .82 squared) of the worldwide inter-polity variance in composite civil strife. A cumulative multiple correlation of .82 is quite an impressive accomplishment, but, of course, this still leaves about 32% of the variance in civil strife to be accounted for by variable not yet utilized (i. e., by variables not yet in the total 230 — some variable data set) or by better measures of the variables that *have* been utilized in this data-set. In either case, it is relatively certain that linguistic heterogeneity will not be among them because we have already utilized it in two different measurement modes⁵ and the results have been identical: linguistic heterogeneity simply does not appear in the optimal subset of independent variables needed for the maximal and most parsimonious prediction of civil strife. We may have the conviction that in one country or another linguistic heterogeneity *does* make an important contribution to the explanation of civil strife, and perhaps that is indeed so. Civil strife

in each country is overdetermined by its own historical circumstances, but in the world as a whole, across all countries and across all historical circumstances, once deprivation, central coercive potential and the presence of organized anomic groups are taken into consideration, it is entirely unwarranted to posit linguistic heterogeneity itself as a co-cause of civil strife. Whatever minor and redundant importance it may have is entirely attributable to its redundancy with the stronger independent predictors that do wind up in the optimal subset of predictors of civil strife.

Per capita gross national product⁶

Turning now to our second criterion variable, we proceed to ask whether worldwide inter-polity variance in linguistic heterogeneity makes an independent contribution to the explanation or prediction of worldwide inter-polity variance in per capita gross national product. Once again it becomes evident that the worldwide inter-polity variance in our second criterion variable is highly predictable on the basis of a relatively small subset of optimal predictors. Out of our original pool of 233 available variables only a certain 10 are needed in order to yield a cumulative prediction of .90 and, once again, linguistic heterogeneity is not a member of this optimal subset. Linguistic heterogeneity naturally has its own correlation with per capita gross national product, a correlation of $-.32$, which, at best (i. e., were it to be non-redundant) would indicate that linguistic heterogeneity could explain only 10% ($= .32$ squared) of the worldwide inter-polity variation in per capita gross national product. However, linguistic heterogeneity is far from being a non-redundant predictor of per capita gross national product. When linguistic heterogeneity is confronted with all other possible predictors of per capita gross national product it is eclipsed entirely, indicating that in itself it really explains nothing at all about per capita gross national product.

The cross-polity variables that do wind up in the optimal subset of predictors of worldwide inter-polity variation in per capita gross national product are primarily (a) governmental modernization, (b) per capita newspaper circulation, (c) presence of a multi-party and parliamentary-republican government, and (d) industrial rather than agricultural concentration of the workforce. All in all, these commulatively account for 81% ($= .90$ squared) of the total worldwide inter-polity variation in per