THIS PAPER PRESENTS SEVERAL THEORETICAL PROPOSITIONS
ABOUT THE RELATIONSHIP BETWEEN LINGUISTIC DIVERSITY WITHIN A
NATION AND WITHIN ITS SUBPARTS OR REGIONS. THE PROPOSITIONS
DESCRIBED ARE EQUALLY VALID FOR THE RELATIONSHIPS BETWEEN ANY
AREAL UNIT AND ITS COMPONENT PARTS AND MAY THUS BE APPLIED TO
A CITY AND ITS DISTRICTS OR TO A PROVINCE AND ITS COUNTIES.
THE "A" INDEX (A QUANTITATIVE MEASURE OF MOTHER TONGUE
DIVERSITY PROPOSED BY GREENBERG IN 1956) GIVES THE
PROBABILITY OF RANDOMLY PAIRED INHABITANTS OF A GIVEN COUNTRY
HAVING DIFFERENT MOTHER TONGUES. THE ADVANTAGE OF THIS
MEASURE IS THAT IT PERMITS QUANTITATIVE DESCRIPTION OF THE
DEGREE OF MOTHER TONGUE DIVERSITY EXISTING IN A GIVEN AREA AS
WELL AS ALLOWING FOR COMPARISON BETWEEN AREAS. THE AUTHOR
FEELS IT CRUCIAL TO RECOGNIZE THAT ANY CONCENTRATION OF A
MINORITY LANGUAGE IN A REGION WILL BE MORE LIKELY TO LOWER
MOTHER TONGUE DIVERSITY IN THE REGION IF THE NATION HAS A
RELATIVELY HIGH LEVEL OF DIVERSITY TO BEGIN WITH. THESE
FINDINGS ALSO ILLUSTRATE HOW EXTREMELY DIVERSE NATIONS CAN
SURVIVE, IGNORING QUESTIONS OF BILINGUALISM AMONG THE NATIVE
SPEAKERS OF LESSER TONGUES AND THE EXISTENCE OF "LINGUA
FRANCAS." THIS PAPER IS TO BE PUBLISHED IN THE "ACTS OF THE
10TH INTERNATIONAL CONGRESS OF LINGUISTICS" BY THE ACADEMY OF
THE S.R. OF ROUMANIA IN 1968. (AMM)

"NATIONAL AND REGIONAL LANGUAGE DIVERSITY"*

by

Stanley Lieberson
Department of Sociology
University of Washington
Seattle, Washington 98105
U.S.A.

AL 000 997

NATIONAL AND REGIONAL LANGUAGE DIVERSITY[*]

by

Stanley Lieberson
University of Washington

This paper presents several theoretical propositions about the relationship between linguistic diversity within a nation and its subparts or regions. The propositions described below are equally valid for the relationships between any areal unit and its component parts. Thus they may be applied to a city and its districts or to a province and its counties.

The starting point is the "A" Index, a quantitative measure of mother tongue diversity proposed by Greenberg (1956). The index gives the probability of randomly paired inhabitants of a given country having different mother tongues. Its computation is rather simple; namely, after determining the proportions in a population with each mother tongue, "A" is equal to 1.00 minus the sum of squares of each proportion. An illustration of the computations is presented below in Table 1 for Switzerland. The value of "A" may be given a simple operational interpretation; namely, if each resident were to be paired with every other resident, "A" is the percentage of dyads in which a common mother tongue is absent. Thus "A" ranges from zero (in the situation where the entire population has the same mother tongue) to 1.00 (in the impossible situation where no two inhabitants have the same mother tongue).

The advantage of this measure is that it permits quantitative description of the degree of mother tongue diversity existing in a given area as well as allowing for comparison between areas. In addition, this measure has been extended

Table 1

Computation of "A" Index for Switzerland, 1960

Mother Tongue	Number	Proportion	Proportion Squared
German	3,765,203	.70	.49
French	1,025,450	.19	.03
Italian	514,306	.10	.01
Romanche . . .	49,823	.01	.00
Total	5,354,772	1.00	.53

$$A_t = 1.00 - .53 = .47$$

Note: Other mother tongues in Switzerland are not included in these calculations which are for illustrative purposes.

to permit quantitative determination of the degree of mutually shared mother tongues between two separate populations or groups (Lieberson, 1964). Clearly, the operational meaning given to the "A" Index is not intended to describe reality since it is unlikely that all residents will interact with one another with equal frequency. Rather, because of spatial and social segregation, residents will tend to interact with mother tongue compatriots more frequently than with those whose native language is different. Nevertheless, this measure provides the investigator with a quantitative index of the exact degree of diversity existing within the nation or unit under study.

The question is naturally raised of the relationship between mother tongue diversity on the national plane and its component subareas. In particular, will the regions which make up a nation have higher or lower mother tongue diversity than the nation? In order to answer this question, the use of symbols will be helpful. Let each mother tongue be expressed as a proportion of the nation's total population, such that $b_t + c_t + d_t \ldots + n_t = 1.00$. Further, let the alphabetic order denote the numerical position of each mother tongue in the nation such that b_t is the largest mother tongue, c_t is the second largest, etc. Let each mother tongue in a given region, i, be expressed as a proportion of the total population of the region. Thus $b_i + c_i + d_i \ldots + n_i = 1.00$ for region i. (Note that the ordering of languages need not apply to any particular region. Thus all we know is that d_t is greater than e_t, but d_i need not be greater than e_i). Further, unless noted to the contrary, the propositions presented below assume that the regional-national differences in mother tongue composition are proportionate to the initial positions of the mother tongue in the nation as a whole. That is, if d_t is smaller than d_i, the increment in d_i is drawn from the remaining mother tongues of the region in proportion to their size in the nation as a whole. This is clearly an arbitrary assumption that will not fit reality in many instances, but it does provide a means to deal with the entire problem of regional and national levels of mother tongue diversity.

There are two basic propositions which describe the general relationship between mother tongue diversity in a nation and its regions:

Mother tongue diversity in a region will be less than for the nation as a whole if the largest mother tongue of the country comprises a larger proportion of the regional population than of the national population. More formally,

$A_i < A_t$, if $b_i > b_t$ and if the remaining mother tongues contribute to the increase of b_i in proportion to their size in the nation.

On the other hand, mother tongue diversity in a given region will be greater than in the nation if the nation's largest mother tongue is also the region's largest tongue but by a smaller proportion of the latter population. If some other tongue is the largest native language in the region and its proportion of the region's population is even larger than is the nation's leading mother tongue's proportion for the nation, then regional diversity will be lower than diversity in the nation. More formally, $A_i > A_t$ if $b_i < b_t$ and $c_i < b_t$, $d_i < b_t$, ..., $n_i < b_t$. But $A_i < A_t$ when $b_i < b_t$, if $c_i > b_t$, or $d_i > b_t$, ..., or $n_i > b_t$. Keep in mind that these propositions depend on the gains or losses among other tongues being proportionate to their sizes.

There are several corollaries that should be noted before the implications of these propositions are considered:

If a language occupies a proportionately larger position in a region than in the nation as a whole, but this gain is at the expense of an even larger national language in the same region, all other factors held constant, diversity will be greater in the region than in the nation if $c_t > d_i$. To put it formally, if $c_t > c_i$, and $c_t > d_i > d_t$ such that $d_i = d_t + K$ and $c_i + K = c_t$, then $A_i > A_t$.

On the other hand, if a language occupies a proportionately larger position in a region than in the nation as a whole and if this gain is at the expense of languages which are smaller in the nation as a whole, all other factors being constant, diversity will be lower in the region than in the nation. To put it formally, if $c_t < c_i$ but $d_i < d_t$ such that $c_i = c_t + k$ and $d_i + k = d_t$, then $A_i < A_t$.

IMPLICATIONS

The implications of these propositions and processes are of considerable importance for the sociolinguistic understanding of the relationship between national and regional mother tongue diversity. When the largest mother tongue of a nation comprises a relatively small proportion of the total population, that is, when the nation is highly diverse, segregation of the mother tongues into different regions will tend to make the regional levels of diversity lower than in the nation. Clearly if the nation's largest mother tongue occupies an

even stronger proportionate position in a given region, then the region's diversity will be lower than that for the nation (making the assumptions noted earlier). But somewhat less obvious, a language which is numerically less important on the national scale can, under this set of conditions, be sufficiently concentrated into a region so as to also reduce the region's diversity in comparison to that of the nation.

This is to be compared with conditions when one language is the mother tongue of the vast majority of the nation's population. Under such circumstances, diversity in the regions will also often be less than that reached in the nation, but for different reasons. When there is low national diversity, it is unlikely for a region to have a sufficient concentration of one of the lesser national tongues such that c_i or d_i, for example, by itself exceeds the value of b_t. The concentration of a lesser tongue into a region at the expense of the nation's largest tongue will, more likely, raise that region's diversity over the national level. However, such concentrations will lower diversity in other areas by raising the level of b_i in many of the regions over b_t.

In short, A_i will often be lower than A_t in many regions if there is either a very high or a very low level of national diversity. High A_t nations will tend to have lower levels of diversity in regions where either the lesser national languages are concentrated or where b_i exceeds b_t. Linguistic homogeneity in a nation will also be accompanied by even greater regional homogeneity, but rarely will this be in a region with a concentration of national "minority" languages. Rather, lower levels of regional diversity will occur almost exclusively in those regions where b_i exceeds b_t. Thus the concentration of lesser tongues in a region will be much more likely to create a region with less linguistic diversity than the nation if A_t is high rather than low.

This does not mean that the regions of countries with a high A_t will have lower diversity than regions in countries with a low national diversity. But it does suggest that regions of the two types of countries will tend to be more alike than will the countries themselves. For example, using Greenberg's computations, the median regional "A" Index for Mexico is .19 and the median for the seven territories in the Caucasus is .33. By contrast, the "A" Indexes for Mexico and the Caucasus as a whole are much further apart, being .31 and .87 respectively. Likewise 11 out of 31 Mexican provinces have higher levels of diversity than the nation as a whole, but only one of the seven territories of the Caucasus exceeds the over-all level of diversity.

There is a contradiction between the political goals in many nations and those of native speakers of each mother tongue. Political, economic, and social benefits are usually believed to be derived from a reduced level of mother tongue diversity. It is not my intention to evaluate this common political argument, but rather to contrast this goal with the perspective of the native speakers of the various mother tongues represented in a nation.

If it is assumed that each person wishes to optimize
the use of his native tongue, then clearly segregation
will be desirable for them. For segregation of all forms -- by regions, by
cities, and within cities by quarters -- will increase the usage of one's native
language. On the other hand, the existence of segregation means that each
mother tongue can maximize its maintenance, but often at the cost of increas-
ing the area's level of mother tongue diversity. Thus, particularly in those
countries with a very large dominant mother tongue, the concentration of
speakers of a lesser mother tongue into a region will tend to raise the region's
level of diversity above that for the nation as a whole even though it strengthens
the position of the lesser tongue. By contrast, in more diverse nations, where
the largest mother tongue is not the native language of so large a proportion of
the population, segregation of lesser mother tongues will be more likely to make
regions less linguistically diverse than the nation as a whole.

It is crucial to recognize that any concentration of a minority language
in a region will be more likely to lower mothe · tongue diversity in the region if
the nation has a relatively high level of diversity to begin with. To put it another
way, any build-up of a minority language in a region will be more likely to raise
the region's "A" Index if the nation as a whole has a relatively low level of
diversity.

These findings also make it clear how extremely diverse nations can
survive, ignoring questions of bilingualism among the native speakers of lesser
tongues and the existence of lingua francas. It is relatively easier for regional
diversity to be lower than national diversity in countries where A_t is high, than
in nations where A_t is low. Thus one cannot assume that day-to-day inter-
personal relations in a nation with high linguistic diversity is radically different
from that in a moderately diverse nation. For the levels of national linguistic
diversity, as important as they are for understanding the political, economic,
and social context of language relations in the society, need not at all be closely
related to the levels of diversity on the local scene. The levels of mother ton-
gue diversity in the regions of India and Scotland, for example, are very likely
closer to one another than are the national levels of diversity in the two nations.

REFERENCES

Greenberg, Joseph H., "The Measurement of Linguistic Diversity," Language, Vol. 32, No. 1 (January-March, 1956), 109-115.

Lieberson, Stanley, "An Extension of Greenberg's Linguistic Diversity Measures," Language, Vol. 40, No. 4 (November, 1964), 526-531.