Boxuan Yu

Professor Michael Hout

Statistics for Social Research

Spring 2023

Linguistic Affiliation and National Identity

Introduction

In contemporary society, nationalism had become a term that is constantly been mentioned throughout media and even ordinary discourse among people. Many voices tend to associate nationalism with extreme patriotism that exhibit excessive enthusiasm in promoting the interest of a specific nation community. From many cases, we can see that the comment on such phenomenon is negative for their potential harmful impact to groups outside a specific nation. However, although nation and nationalism are never some unfamiliar terms to many people, we might be still not yet managed to reveal the myth of them in terms of how they are formed and operated. According to Michael Hechter, nation is a "relatively large group of genetically unrelated people with high solidarity" and nationalism is then a form of "collective action that seeks to render the boundaries of the nation congruent with those of the government unit", which ultimately lead to the establishment of "nation-state" (2000, p.9-11). In Hechter's ideas, nation-state exists as "the principal form of political organization in the modern world" (Internal Colonialism, p. 9). Therefore, such notion can certainly shuffle some contemporary conceptions that merely associate nationalism with aggressive behaviors and thus present us a new focus of nationalism: how its root, nation, was consolidated and grew into a power that can shape the dynamics of modern politics and social networks?

A Nation-Building Thesis

Based on the existing literatures, scholars in the related fields of research commonly share the idea that the solidarity among people who are identified with a same nationhood is

consolidated through the standardization of language (Hechter, 2000, p.48; Anderson, p.147). In other words, using same while standardized language provides people with same national identity a linkage. Such linkage is not merely a one-time self-identification but, as Benedict Anderson illustrated, a function that is capable in reproducing the sense of linkage, or affiliation, as well as the perception on distinction with other population at an "imagined" space (Anderson, p.37). Apart from the theoretical framework, we can also absorb some knowledge on the significance of language to nation-building and nation-consolidating agenda through empirical study. During the French Revolution, which was interpreted as one of the first cases of modern nationalist revolutions, a key feature of the newly created "French nation" and French nationalism is to uphold a standardized language as the official one for all citizens regardless of people's regional diversity (Adrey, 124-126; Wimmer, 57).

Thus, I would propose a hypothesis based on the existing theories as well as related historic accounts that linguistic preference would affect one's subjective notion on national belonging. Meanwhile, as reminded by Michael Hechter's definition on nation, I also intend to discover whether genetic relationship, or lineage, might impose a positive impact on individual's affirmation of national identity. In other words, I suggest that if one has an ancestor that upholds a specific national identity, such individual might be even more willing to admit a national belonging sentiment.

Data and Measurement

To better understand how the use of language and the subjective notion on national belonging are related, I reach out to the data archive of ISSP, International Social Survey Program, which is an eminent cross-national collaboration program over the research upon various topics under the discipline of social science. Subsequently, I picked the dataset of National Identity Survey gathered in 2013. The survey was conducted at over 30 countries with numerous categories of questions indicated to discover the thoughts of respondents on issues like their subjectively perceived connection to their states as well as the weighted pre-conditions for such perception. The questionnaire of the survey is also modified in different regions in order to adapt the distinctive social context. For instance, respondents

in the United States will only receive questionnaire asking their sense of connection with their located region, a specific county or city, and state, the United States, instead of their prediction on people's general tendency. Meanwhile, questions on related policy in a country is also selected for different regions to help understand each regional cases thoroughly.

For my study, I select three discrete variables from the results of three questions listed in the survey, the degree of the sense of connection to one's situated state, the opinion on speaking a specific language as an important criterion for being truly being specific nationality, and the opinion on one's lineage for the above concern. Among these three, the sense of connection is considered as the dependent variable so that it will hypothetically be influenced by linguistic preference and lineage status, as two independent variables. For these three discrete variables, they each contain several values based on one's response in the questionnaire:

1 – Very important (Very close [to your state])
2 – Fairly important (Fairly close)
3 – Not very important (Not very close)
4 – Not important at all (Not close at all)
~~5 – Can't choose~~
~~8 – Don't know~~
~~9 – Refused~~

Based on the options shown, I will explain briefly how they demonstrate the information that is related to the three variables I've mentioned. For the dependent variable, if one selects "1 – Very close [to your state]", that will imply the respondent's subjective recognition on the relatively tight connection to state, and thus the national identity as we've discussed how nation and state is closely connected previously. For the other two variables, if respondents select "1 – Very important", they are then subjectively admitting the high significance of speaking a specific language or having domestic ancestral bond as the criteria for truly obtaining sense of belonging to the specific national identity. However,

since the last choices, “Can’t choose”, “Don’t know”, “Refused” are not applicable to my numerical measurement, I decide to delete the samples in the data that choose any of these choices under any of the three questions examined. Then, 40,466 observations out of 45,297 observations are valid in my study, with percentage distributions as shown in Figure 1. In convenience, I name the three variables based on the Q&A with three names respectively: “national belonging”, “linguistic affiliation”, “lineage affiliation”. Then, I will use three regression models to analyze some relationships between each variable.

In Model 1, I use “national belonging” as dependent variable while “linguistic affiliation”, a key factor in my anticipation, as the only independent one to process regression. From Figure 2, we can see that all values included in “linguistic affiliation”, as a categorical variable, are statistically significant. With the coefficients calculated, we also see a trend of increase from “Language 2” to “Language 4”, which represent “Fairly important” and “Not important at all” in the survey. This indicates that by more subjectively neglecting the importance of specific linguistic preference in a state, respondents display an increasing tendency in perceiving less connection with one’s national identity, as reflected by Figure 3. Also, as to clarify, since the original dataset displayed the values in the way that higher value indicates less degree of perceived importance (or connection), the coefficients here also suggests that with coefficients become numerically high, one possesses a less sense of national belonging. In this sense, the value that is more approaching to 1 would imply that one is feeling more connected to the state.

Apart from Model 1, I then add a new variable, lineage affiliation to the regression function to test the multivariate association. Based on the numerical results in Figure 4 and graphic projection in Figure 5, we can see that either linguistic affiliation or linguistic affiliation present a similar effect on one’s subjective notion on national belonging. More specifically, the increase in numerical values of both independent variables would lead to the increase in one’s conscious distance to national identity. Meanwhile, such effect is even stronger by each unit when the value of linguistic and lineage affiliation variable changes from 2 to 4.

Finally, I decide to input a new interaction variable to the Model 2 for examining whether there is an interaction relation between the independent variables or not. According to the results shown in Figure 6, we can see that most scenario of interaction based on the two categorical variables are statistically significant, except coefficients relate "Language 4" which itself is less significant than how it is displayed in previous two models. Nevertheless, based on the t-values in most cases, we can certainly admit that there's an interaction effect between linguistic affiliation and linguistic affiliation in terms of the effect on one's subjective perception of national belonging. However, with the help of Figure 7, we can also easily identify that by simultaneously rejecting the importance of linguistic affiliation and lineage affiliation, one display the most least attachment to the state. Furthermore, by ignoring the case of "Lang4:Lin2" that is statistically insignificant, we can see that the interaction effect is always getting stronger when one thinks language is not important for national belonging (increase in numerical values of linguistic affiliation variable), holding the opinion on lineage affiliation unchanged. But we don't see a general pattern of the interaction effect on lineage affiliation when we hold linguistic affiliation unchanged.

Discussion and Conclusion

Based on the data and the three models presented, we can reject the null-hypotheses that there are no there's no causal relationships between our dependent variable, subjective notion of national belonging, and two independent variables, linguistic affiliation as well as lineage affiliation of individuals respectively. In a detailed explanation, we can say that in most cases, if one think linguistic preference or lineage is significant in perceiving national belonging, one is more likely to have a strong self-awareness on connection with the nation-state, holding another independent variable unchanged. Meanwhile, we also confirm that if one admits the high relevance of linguistic preference and lineage simultaneously, this individual has the most relative strong bond with the specific national identity. Vice versa, one display least connection to nationhood when this individual rejecting the importance of the two independent variables at the same time.

Thus, with the above statistical summary, we can briefly entail that the theories and empirical study on cases that can be traced to centuries ago on the relationship between national identity-building, or nation-building, and language preference still have their place in the contemporary society, given the fact that the data used for this research is gathered in 2013. Moreover, the ancestral bond is also suggested to be important in studying how high degree of solidarity can be reached within a nation. However, I would certainly agree that there's more variables in the study of nation-building and this study have a vast room for improvement. Besides, there are also other concerns apart from the lack of variables.

Firstly, thinking of the comparison between Model 3 and previous two models. We can see that there's a decrease in the statistical significance of "Language 4", though it is still significant. At the same time, the only insignificant variable, "Lang4:Lin2" is also related to the "Language 4". One potential explanation to this circumstance is that the number of cases examined relate to "Lang 4" might not be statistically sufficient based on Central Limit Theorem. Retrieve back to the distribution of the answers, the amount of respondents who select "Linguistic Affiliation 4" is indeed the least compared to other options. Numerically, there is only 1465 cases out of 40466 applicable cases, as merely 3.6% of the entire sample. Therefore, I offer two perspectives of thoughts on it. The foremost one is that there is still room for expanding the range of respondents. Besides, it is also plausible to say that the majority of citizens around the globe, or at least in the countries where this survey took place, retain a relatively strong connection to the nationhood consciously. Thus, it is hard to certain extent for the examination on the scenario of reversed conditions as it might be fundamentally hard to find sufficient amount of people who feel distant from the nationhood.

Secondly, regarding the general methodology, this dataset as well as the starting point of the research contain another flaw. In contemporary society, with the increase in social mobility, we can witness tremendous cases of immigration that facilitate the linguistic diversity of many states. For instance, in the United States, we can no longer, especially in recent few years, say that using English is sufficient to share an imagined space with most other citizens as population who mostly use Spanish, French, or even Chinese are

increasing dramatically. However, whether linguistic diversity diminish the sense of national belonging or not is statistically not yet been collected yet, indicating a potential track for future empirical study.

Figure 1. Percentage Distribution of the Answers Under the Three Questions

	1	2	3	4
National Belonging	44.35	43.14	9.87	2.64
Linguistic Affiliation	59.17	28.60	8.61	3.62
Lineage Affiliation	31.88	27.98	25.42	14.72

Figure 2. Model 1 - Numerical Subjective Association between Linguistic Affiliation and National Identity

```
Call:
lm(formula = national.belonging ~ linguistic.affiliation, data = memo)

Residuals:
    Min      1Q  Median      3Q     Max
-0.9495 -0.5990  0.1567  0.4010  2.4010

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
Intercept  1.599031   0.004770  335.24   <2e-16 ***
Language 2 0.244238   0.008355   29.23   <2e-16 ***
Language 3 0.306851   0.013381   22.93   <2e-16 ***
Language 4 0.350457   0.019864   17.64   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.738 on 40462 degrees of freedom
Multiple R-squared:  0.03155,     Adjusted R-squared:  0.03148
F-statistic: 439.4 on 3 and 40462 DF,  p-value: < 2.2e-16
```

Figure 3. Model 1 in Graph

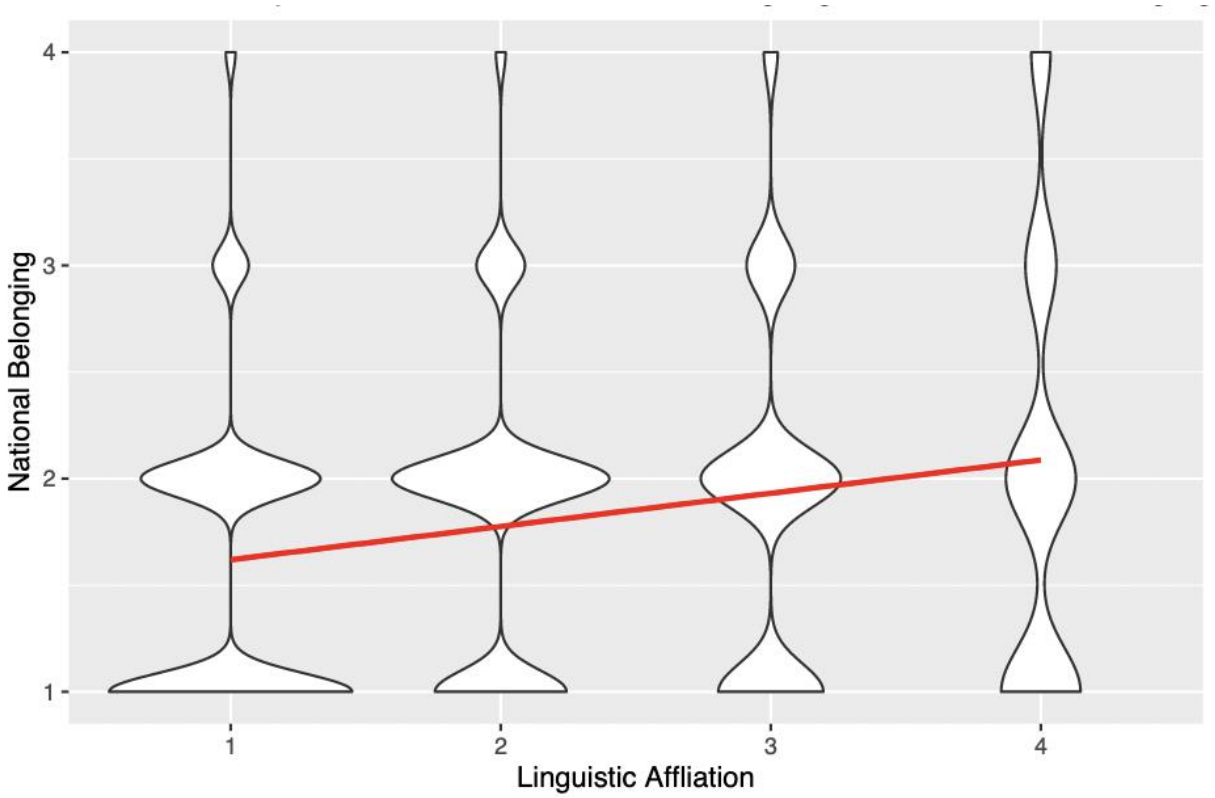


Figure 4. Model 2 – Numerical Association between linguistic Affiliation, Lineage Affiliation, and National Belonging

```
Call:
lm(formula = national.belonging ~ linguistic.affiliation + lineage.affiliation, data =
memo)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01602 -0.66538  0.08894  0.33462  2.50794

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
Intercept  1.492058   0.006606  225.85   <2e-16 ***
Language 2 0.185378   0.008690   21.33   <2e-16 ***
Language 3 0.247641   0.013562   18.26   <2e-16 ***
Language 4 0.290342   0.020003   14.52   <2e-16 ***
Lineage 2  0.173318   0.009801   17.68   <2e-16 ***
Lineage 3  0.189626   0.010050   18.87   <2e-16 ***
Lineage 4  0.233621   0.011869   19.68   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7329 on 40459 degrees of freedom
Multiple R-squared:  0.04498,     Adjusted R-squared:  0.04484
F-statistic: 317.6 on 6 and 40459 DF,  p-value: < 2.2e-16
```

Figure 5. Model 2 – Plotted Coefficients

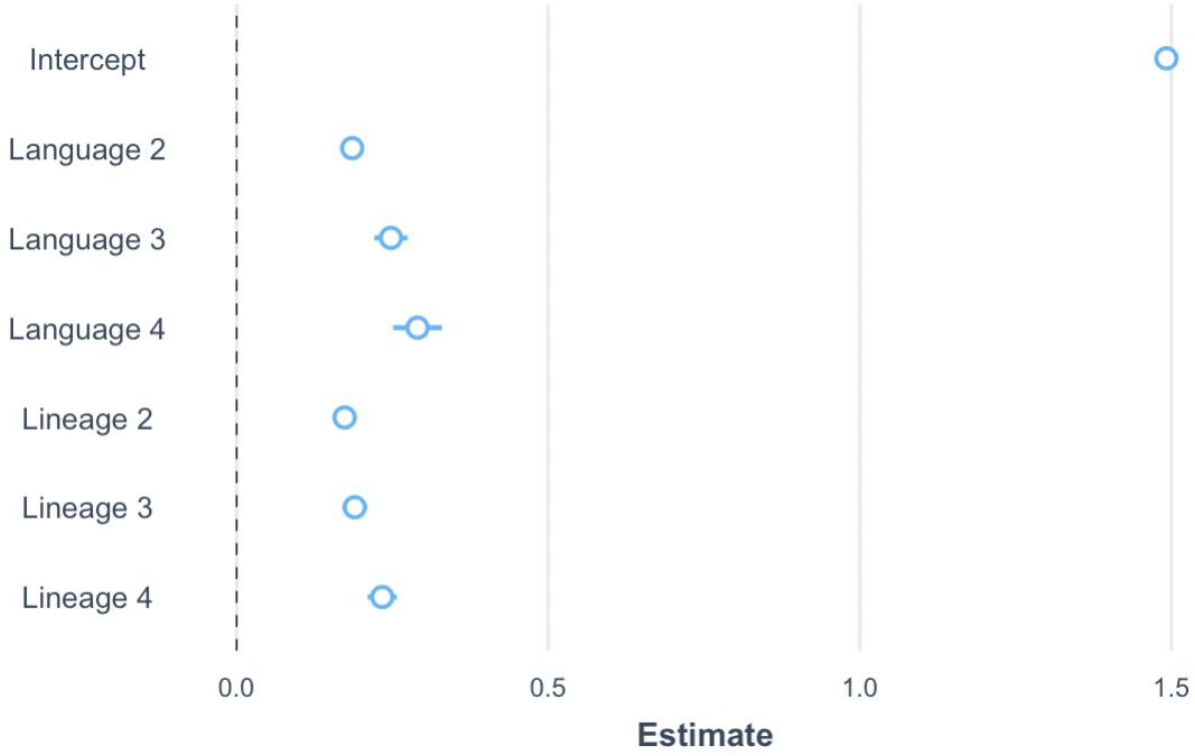


Figure 6. Model 3 – Interaction between Linguistic Affiliation and Lineage Affiliation

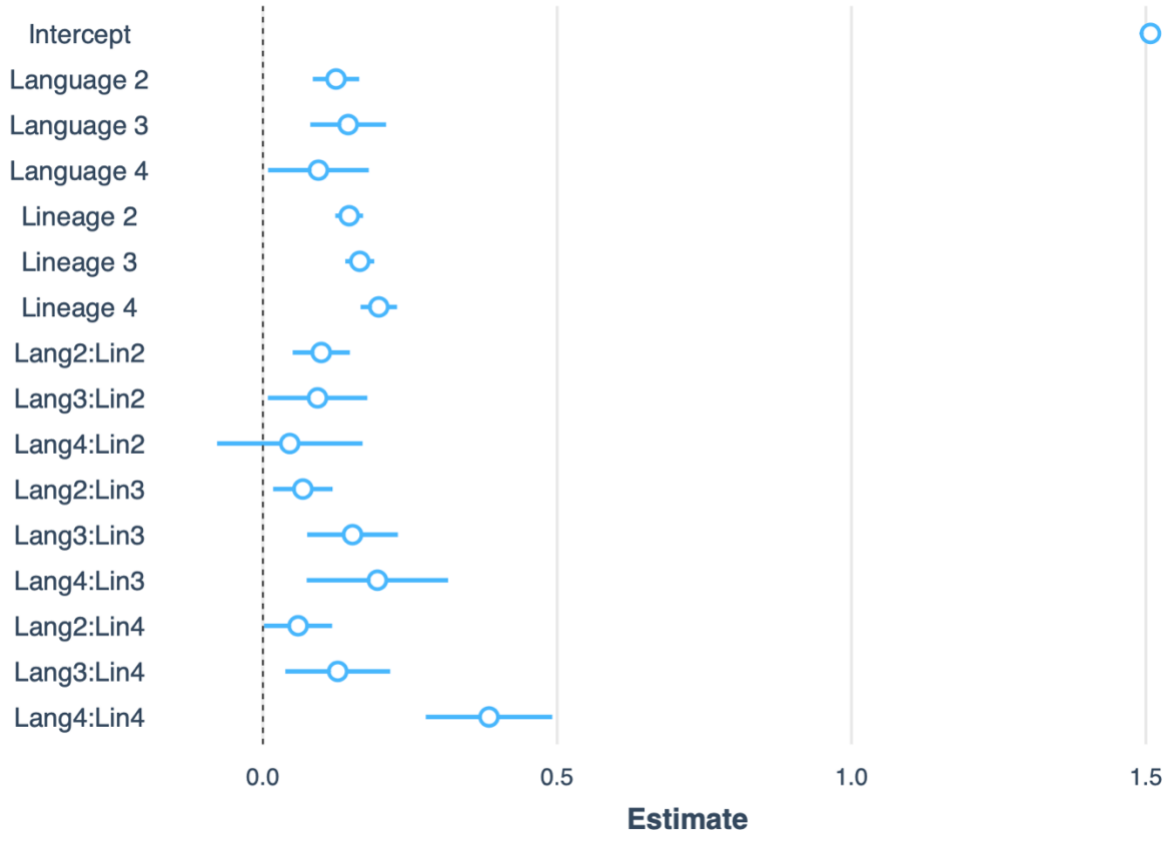
```
Call:
lm(formula = national.belonging ~ linguistic.affiliation + lineage.affiliation +
linguistic.affiliation:lineage.affiliation, data = memo)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1829 -0.6543  0.1115  0.3457  2.4922

Coefficients:
           Estimate Std. Error t value Pr(>|t|)
Intercept  1.507801   0.007119 211.785  < 2e-16 ***
Language 2 0.124125   0.020106   6.173 6.74e-10 ***
Language 3 0.144790   0.032856   4.407 1.05e-05 ***
Language 4 0.094275   0.043651   2.160 0.030799 *
Lineage 2  0.146493   0.012059  12.148  < 2e-16 ***
Lineage 3  0.164519   0.012556  13.103  < 2e-16 ***
Lineage 4  0.196795   0.015791  12.462  < 2e-16 ***
Lang2:Lin2 0.099105   0.024807   3.995 6.48e-05 ***
Lang3:Lin2 0.092773   0.043084   2.153 0.031299 *
Lang4:Lin2 0.045548   0.063014   0.723 0.469793
Lang2:Lin3 0.067859   0.025800   2.630 0.008538 **
Lang3:Lin3 0.152182   0.039298   3.873 0.000108 ***
Lang4:Lin3 0.194443   0.061259   3.174 0.001504 **
Lang2:Lin4 0.059776   0.029553   2.023 0.043114 *
Lang3:Lin4 0.127085   0.045458   2.796 0.005182 **
Lang4:Lin4 0.384015   0.054804   7.007 2.47e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7321 on 40450 degrees of freedom
Multiple R-squared:  0.04728,     Adjusted R-squared:  0.04692
F-statistic: 133.8 on 15 and 40450 DF,  p-value: < 2.2e-16
```

Figure 7. Model 3 – Plotted Coefficients



Works Cited

Adrey, Jean-Bernard. *Language, Nation and State in French Linguistic Nationalism: History, Developments and Perspectives*, Palgrave Mcmillan, 2009.

Anderson, Bennedict. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, Verso, 1991.

Hechter, Michael. *Internal Colonialism: the Celtic Fringe in British National Development*, Transaction Publishers, 1999.

Hechter, Michael. *Containing Nationalism*, Oxford University Press, 2000.

Wimmer, Andreas. *Nationalist Exclusion and Ethnic Conflict : Shadows of Modernity*, Cambridge University Press, 2002.