The Global Leader Personality Dataset, 1946-2022 : Codebook

Mahesh Acharya *

*Department of Political Science, University of Alabama

1 Introduction

This data project generates the personality scores for over 10,000 global leaders spanning a period of 75 years (1946–2024). The corpus of speeches is taken from Baturo, Dasandi and Mikhaylov (2017). I organize these speeches by year, and bootstrap them for 100 times at sentence level before processing through psycholinguistic dictionaries – LIWC and MRCPD. Such a step generates counts and proportions of word usage across all linguistic categories of the relevant dictionaries. These counts and proportions are plugged into pre-trained Personality Recognizer Model (Mairesse et al., 2007) in line with similar previous work on US legislators (Ramey, Klingler and Hollibaugh, 2017). The model returns the Big Five (openness, conscientiousness, extraversion, agreeableness, and emotional stability) personality scores via SMOreg (Support Vector Machine Regression). SMOreg can be modeled as follows.

$$y = w.x + b \tag{1}$$

Here, y as a dependent variable is a $n \times 1$ vector of personality traits of a speaker to be predicted, and x as an independent variable is a $n \times k$ vector of the psycholinguistic features generated from LIWC and MRCPD. b is the bias term (intercept), and w is a $k \times 1$ vector of coefficients. SMOreg looks intuitively similar to linear regression but the main goal with former model is to minimize the sum of squared coefficients $\sum_{n=1}^{k} w_i^2$ (i.e. w as small as possible) as opposed to sum of squared errors in the latter. Such minimization in SMOreg is however subject to a pre-specified precision or tolerance margin ϵ that indicates a constrained optimization opposite to the unconstrained optimization of Ordinary Least Squares (OLS) method(Smola and Schölkopf, 2004).

To summarize, the word usage across different categories of LIWC and MRCPD predicts the Big Five personality traits of the speaker. The model outputs the personality scores of a leader on a scale of 1 to 7 with the higher score indicating a higher level of a relevant trait. For example, I find former US President Barack Obama highly extroverted as indicated by

his consistently high scores on the extraversion trait compared to the average leader.

2 Variables

Year: year, the speech is delivered at United Nations General Assembly (UNGA) meeting

Leader: speaker at a particular UNGA session in a year (for example: ARG_01_1946 means a speaker from Argentina in 1946 in the first UNGA session. Note that the speaker could be any top official from the relevant country such as President, Prime Minister, Foreign Minister, Ambassador, etc.)

Openn_SD: Standard Deviation for Openness variable

Consc_SD: Standard Deviation for Conscientiousness variable

Extra_SD: Standard Deviation for Extraversion variable

Agree_SD: Standard Deviation for Agreeableness variable

Emoti_SD: Standard Deviation for Emotional Stability variable

Openn_SE: Standard Error for Openness variable

Consc_SE: Standard Error for Conscientiousness variable

Extra_SE: Standard Error for Extraversion variable

Agree_SE: Standard Error for Agreeableness variable

Emoti_SE: Standard Error for Emotional Stability variable

Openness: Mean of the bootstrapped Openness scores

Conscientiousness: Mean of bootstrapped Conscientiousness scores

Extraversion: Mean of bootstrapped Extraversion scores

Agreeableness: Mean of bootstrapped Agreeableness scores

Emotional Stability: Mean of bootstrapped Emotional Stability scores

Openn_LB: Lower bound value of Openness variable

Consc_LB: Lower bound value of Conscientiousness variable

Extra_LB: Lower bound value of Extraversion variable

Agree_LB: Lower bound value of Agreeableness variable

Emoti_LB: Lower bound value of Emotional Stability variable

Openn_UB: Upper bound value of Openness variable

Consc_UB: Upper bound value of Conscientiousness variable

Extra_UB: Upper bound value of Extraversion variable

Agree_UB: Upper bound value of Agreeableness variable

Emoti_UB: Upper bound value of Emotional Stability variable

3 Comparison Group Cut Points

Low = If the score of leader under study is more than one standard deviation below the mean for the sample of leaders

Moderate or similar to average leader = If the score of leader under study falls around the mean for the sample i.e. neither one standard deviation above nor below the mean

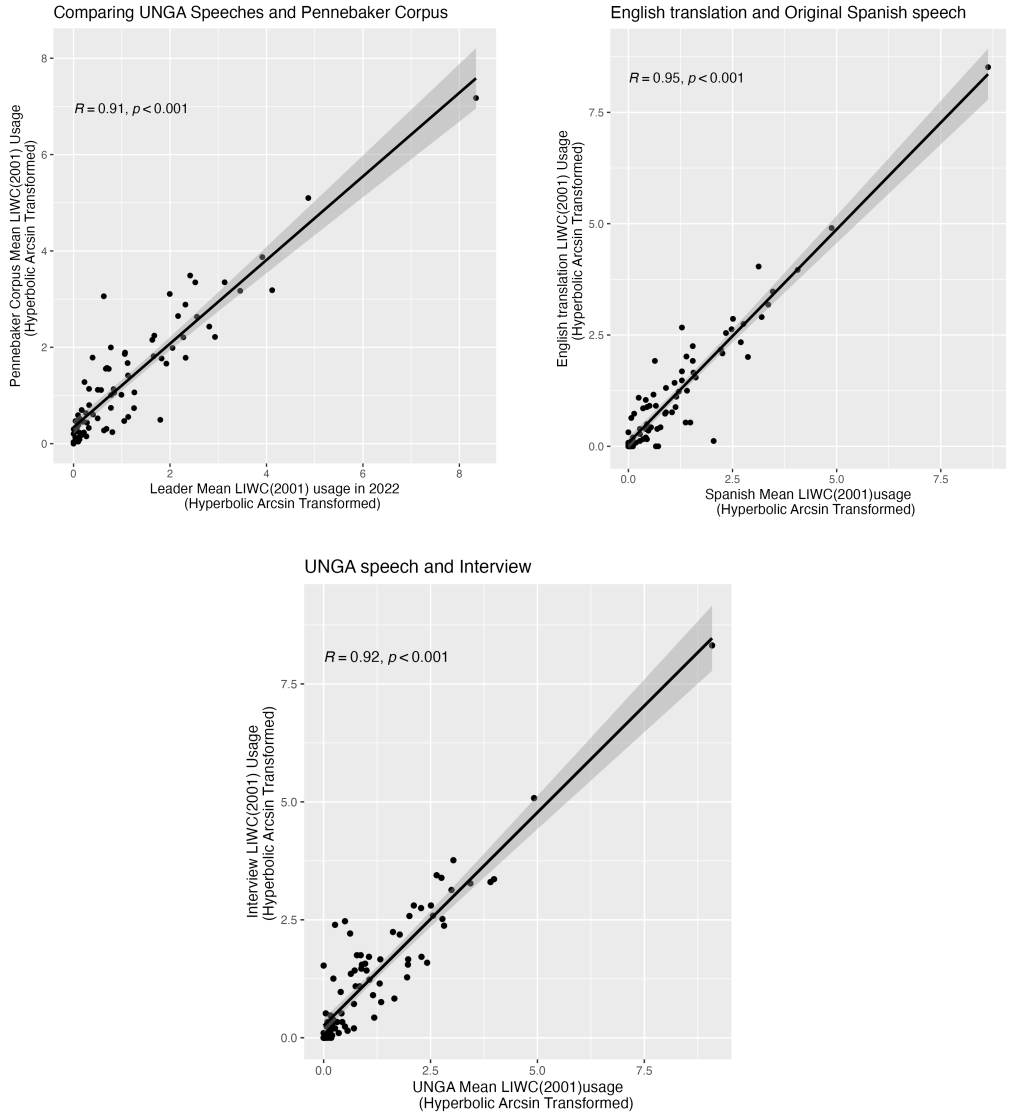
High = If the score of leader under study is more than one standard deviation above the mean for the sample of leaders

Table 1: Comparison Group Cut Points

Personality Traits	**Heads of State/Government**	**All Leaders**
Openness	$Mean = 3.61$	$Mean = 3.72$
	$Low < 3.20$	$Low < 3.32$
	$High > 4.02$	$High > 4.12$
Conscientiousness	$Mean = 3.42$	$Mean = 3.53$
	$Low < 2.98$	$Low < 3.09$
	$High > 3.86$	$High > 3.97$
Extraversion	$Mean = 3.82$	$Mean = 3.67$
	$Low < 3.39$	$Low < 3.24$
	$High > 4.25$	$High > 4.1$
Agreeableness	$Mean = 3.59$	$Mean = 3.6$
	$Low < 3.24$	$Low < 3.34$
	$High > 3.94$	$High > 4.02$
Emotional Stability	$Mean = 3.10$	$Mean = 3.19$
	$Low < 2.71$	$Low < 2.81$
	$High > 3.49$	$High > 3.57$

4 Addressing Concerns

Figure 1: Addressing Compatibility, Translation, and Speechwriter Concerns



References

Baturo, Alexander, Niheer Dasandi and Slava Mikhaylov. 2017. "Understanding State Preferences With Text As Data: Introducing the UN General Debate Corpus." *Research and Politics* pp. 1–9.

Mairesse, Francois, Marilyn A. Walker, Matthias R. Mehl and Roger K. Moore. 2007. "Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text'." *Journal of Artificial Intelligence Research* 30(1):457–500.

Ramey, Adam J., Jonathan D. Klingler and Gary E. Hollibaugh, Jr. 2017. *More Than a Feeling: Personality, Polarization, and the Transformation of the US Congress Chicago and*. London: The University of Chicago Press.

Smola, Alex J. and Bernhard Schölkopf. 2004. "A tutorial on support vector regression." *Statistics and computing* 14(3):199–222.