**Methodological Appendix**

This document complements the methodological part of the paper “Do Concepts Matter? Latin America and South America in Brazilian Foreign Policymakers”. In this appendix we address gaps in the analysis as well as expand discussions that had to be summarized in the main text due to space constraints. The remainder of this document is organized as follows: a) Variables and operationalization; b) Documents and OCR.

**1. Variables and operationalization**

In order to operationalize the variables related to the concepts of Latin America and South America, we have opted to employ a simple, direct, and parsimonious collection of terms. In this regard, in the case of future replications, please, use the following scheme:

|  |  |
| --- | --- |
| **Concept** | **Terms** |
| South America | "América do Sul" OR "South American" OR "Amérique du Sud" OR "América del Sur" OR "sul-americano" OR "sul-americanos" OR "sul-americana" OR "sul-americanas" OR "south american" OR "south americans" OR "sudamericano" OR "sudamericanos" OR "sudamericana" OR "sudamericanas" |
| Latin America | "América Latina" OR "Latin America" OR "Amérique Latine" OR "latino-americano" OR "latino-americanos" OR "latino-americana" OR "latino-americanas" OR "latin american" OR "latinoamericano" OR "latinoamericanos" OR "latinoamericana" OR "latinoamericanas" |

As can be seen, the terms convey the idea of both concepts in singular and plural, as well as in Portuguese, Spanish, English, and French. They all are between quotation marks as required by the software used in order to count the exact expressions. Moreover, we have used the logical Boolean operator OR because we wanted to account for all values related to the concepts.

**2. Documents and OCR**

All documents were available in pdf format and have been converted to txt in order to make queries. However, not all of them were editable. We have overcome this obstacle by using the technology Optical Character Recognition (OCR). We need to report this due to the fact that there are still plenty of misinterpretations of alphanumeric characters for non-English languages. That is also the reason why we have included all txt files together with the data frame and its codebook. The following chart lists the documents converted to OCR before they were converted to txt.

|  |  |
| --- | --- |
| **DOCUMENT**  **(PDF → OCR → TXT)** | **N OF SPEECHES** |
| 1997.2 | 27 |
| 1998.1 | 33 |
| 1998.2 | 31 |
| 1999.1 | 22 |
| 1999.2 | 32 |
| 2003.1 | 54 |
| 2003.2 | 222 |