

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

<https://doi.org/10.23925/2237-759X2026V58e75751>

## REPRESENTATIONS OF AMERICAN IDENTITY: A MULTIDIMENSIONAL ANALYSIS OF ENGLISH GOOGLE BOOKS TRIGRAMS

## REPRESENTAÇÕES DA IDENTIDADE AMERICANA: UMA ANÁLISE MULTIDIMENSIONAL DE TRIGRAMAS EM INGLÊS NO GOOGLE BOOKS

Rafael Fonseca de ARAÚJO  
(Universidade Metropolitana de Santos / FATEC Baixada Santista)  
[prof.rafael\\_araujo@hotmail.com](mailto:prof.rafael_araujo@hotmail.com)

Tony Berber SARDINHA  
(PUC-SP)  
[tonycorpuslg@gmail.com](mailto:tonycorpuslg@gmail.com)

Maria Claudia Nunes DELFINO  
(FATEC da Praia Grande / PUC-SP)  
[claudia@fatecpg.com.br](mailto:claudia@fatecpg.com.br)

Joseph Glynn COLLENTINE  
(Northern Arizona University)  
[j.collentine@nau.edu](mailto:j.collentine@nau.edu)

William Michael LAKE  
(Georgia State University)  
[wlake1@gsu.edu](mailto:wlake1@gsu.edu)

**ABSTRACT:** This paper reports a Lexical Multidimensional Analysis (Berber Sardinha, 2014, 2019, 2020a, 2020b; Berber Sardinha & Fitzsimmons-Doolan, 2025) of English Google Books trigram data from publications spanning 21 decades (1800–2008). The study aims to compare social representations of the peoples of the Americas in publications written in English, focusing on recurrent noun phrase trigrams containing the collocations North American, Latin American, Central American, and South American. The findings indicate that regions outside North America are predominantly represented through a tripartite pattern involving economic opportunity, internal conflict, and natural abundance, reflecting enduring socio-historical perspectives in English-language discourse.

**KEYWORDS:** Multidimensional Analysis; lexical dimensions; Corpus Linguistics; Google Books, Social Representation

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

**RESUMO:** Este artigo apresenta uma Análise Multidimensional Lexical (Berber Sardinha, 2014, 2019, 2020a, 2020b; Berber Sardinha & Fitzsimmons-Doolan, 2025) de trigramas do Google Books, extraídos de publicações que abrangem 21 décadas (1800-2008). O estudo tem como objetivo comparar as representações sociais dos povos das Américas em publicações em língua inglesa, com foco em trigramas de sintagmas nominais recorrentes contendo as colocações *North American*, *Latin American*, *Central American* e *South American*. Os resultados indicam que regiões fora da América do Norte são representadas por um padrão tripartite envolvendo oportunidade econômica, conflito interno e abundância natural, refletindo perspectivas sócio-históricas duradouras no discurso em língua inglesa.

**PALAVRAS-CHAVE:** Análise Multidimensional; Dimensões Lexicais; Linguística de Corpus; Google Books; Representação Social.

## **Introduction**

Much has been written about stereotypes. Moscovici (2001) emphasizes that representations serve two main functions: (a) they conventionalize objects, people, and events by assigning them a definite form, categorizing them, and establishing them as models shared by a group; and (b) they are prescriptive, combining structure and tradition in ways that shape thought before opinions are consciously formed.

Building on this understanding of how representations shape perception, previous research has explored how stereotypes manifest in different contexts. While Berber Sardinha (2014, 2019) has used MDA methods to examine variations in the use of terms such as *American* and *Brazilian*, less is known about stereotypes associated with specific geographic regions through lexical MD Analysis. This study conducts an exploratory factor analysis of recurrent concepts associated with North America, Central America, Latin America, and South America.

## **1. Literature Review**

The interdisciplinary study of social representation arises from various intellectual domains, such as science, religion, and mythology, which contribute to the transmission of culturally inherited epistemologies (Moscovici, 1988). Unlike earlier conceptions of culture, Durkheim (1938) describes it as an emergent web of representations that encompass deeply embedded values, beliefs, and symbolic systems of a natural collectivity, such as the tribal societies he studied extensively. Society binds individuals to it, shaping their reality,

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

particularly in traditional settings. Culture may thus be understood as a symbolic response to social uncertainty.

Moscovici (1988) defines social representation as:

"contents of everyday thinking and the stock of ideas that gives coherence to our religious beliefs, political ideas and the connections we create as spontaneously as we breathe. They make it possible for us to classify persons and objects, to compare and explain behaviors and to objectify them as part of our social setting. While representations are often to be located in the minds of men and women, they can just as often be found 'in the world', and as such examined separately." (Moscovici, 1988, 214)

Individuals both shape and are shaped by society, as social representations—or stereotypes—function as widely held yet oversimplified images or ideas about particular groups or concepts. In line with Moscovici's definition, Jodelet (1989) argues that representations function as categories of practical knowledge that facilitate communication and understanding of social, material, and ideological contexts. They manifest as cognitive elements, such as concepts and images, which are socially elaborated and contribute to the construction of perceived reality. Social representations help individuals classify, interpret, and engage with their environment, guiding decision-making and influencing how groups are positioned in relation to one another. For instance, the term *American* may carry positive connotations in some countries but negative connotations in others, depending on cultural and historical contexts.

Building on the theoretical framework of social representations (Moscovici, 1988, 2001; Jodelet, 1989), which link language to the interpretation of social reality, this study examines Pan-American social representations as reflected in a corpus of Google Books publications spanning the 1800s to the 2000s. The study follows previous research on national representations in corpus linguistics, particularly the work of Stubbs (1996), et al. (2013), and Berber Sardinha (2014, 2019).

Stubbs (1996) argues that corpus-based studies can identify recurrent phrases, which play a key role in establishing national stereotypes. He adds that "naming or labeling are powerful ideological tools" (p. 185). By analyzing language patterns, corpus-based approaches can uncover labels that reflect dominant perceptions of specific groups. Stubbs' focus on *naming and labeling* aligns with Jodelet's (1989) idea that naming distinct aspects of reality enables individuals to classify, interpret, and position themselves within social hierarchies (e.g., insider/outsider, high class/low class, colonizer/colonized).

Baker et. al.(2013a) examined changes in the representation of LGBTQIQA+ individuals in The Daily Mail and The Sun, using corpus

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

linguistics to track how homophobic discourse and negative stereotypes were articulated in British newspapers between 2001 and 2014. He argues that language plays a central role in shaping public perceptions, reinforcing or challenging prejudice. Baker et al. (2013b) further explore how newspapers construct representations through cumulative discourse, in which ambiguous or biased word usage gradually permeates societal consciousness, corroborating Stubbs' (1996) findings on repeated linguistic patterns.

Berber Sardinha (2014, 2019) applies similar corpus-based methodologies to examine national representations of *Americans* and *Brazilians* in English-language Google Books publications from the 1800s to the 2000s. He employs Multidimensional Analysis (MDA) to identify lexical dimensions that reveal the most salient associations with these national identities over time. Sardinha (2014) proposes that "the degree to which such phrases are repeated is arguably a significant indication of their weight in society" (p. 3). Similarly, this study aims to identify lexical dimensions that reflect Pan-American representations in Google Books and explore how terms like *American* and *Latin American* have evolved across decades.

The present study follows a Multidimensional Analysis (MDA) framework (Biber, 1988 et al.), an exploratory factor analysis technique designed to map linguistic associations in large corpora (Berber Sardinha, 2004). Originally developed to compare spoken and written registers in English (e.g., Biber, 1988), MDA employs computational tools and multivariate statistical techniques to identify underlying patterns of lexical-grammatical co-occurrence, revealing key linguistic dimensions of variation. By examining these dimensions, the method provides a structured approach to analyzing register variation, highlighting both the similarities and differences among registers in terms of their linguistic characteristics.

Biber (2001) emphasizes that this corpus-based approach treats texts as the primary unit of analysis, operating on the assumption that different registers display distinctive co-occurrence patterns. These patterns define linguistic spaces, which in turn reflect the communicative functions underlying variation (Zuppari, 2014). In his seminal work, Biber (1988) identifies five major dimensions of variation in English, including contrasts between interaction and information, narrative and non-narrative concerns, abstract versus non-abstract information.

Biber's method was adapted by Berber Sardinha (2014) in order to identify groups of lexical associations by analyzing lists of Google Books bigrams, also understood as dimensions of variation, in this case, lexical dimensions revealed by lexical variation (or "semantic" fields).

While the original Multidimensional Analysis proposed by Biber (1988) models variation through the co-occurrence of lexico-

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

grammatical features in order to characterize communicative functions and registers, subsequent work has demonstrated that similar multidimensional principles can be applied exclusively at the lexical level. In this lexical adaptation, variation is not interpreted in functional terms, but rather in terms of semantic patterning and discursive organization encoded in recurrent lexical associations. Lexical Multidimensional Analysis (Berber Sardinha, 2014, 2019) thus represents a methodological extension of grammatical MDA, preserving its statistical logic while shifting the interpretive focus from register functions to social representations and discourse-level constructs.

The Lexical Multidimensional Analysis (LMDA), proposed by Berber Sardinha (2014), is a variant of functional Multidimensional Analysis, aiming to capture semantic groupings and conceptual trends through lexical features present in the texts. The original Multidimensional Analysis proposed by Biber (1988) explains variation through the correlation of lexico-grammatical features (e.g., nouns, verbs, adjectives, adverbs, relative pronouns) with the aim of characterizing different discursive functions or registers. The LMDA used in the present study aims to explain variation by identifying collocates to characterize novel or emerging uses of semantic fields helping to establish discourse themes.

Berber Sardinha (2019, 234) highlighted that in lexical MD Analysis the interpretation of the dimensions takes into consideration the lexical features of a corpus:

using primarily (but not solely) such properties as semantic preference (Stubbs, 2007), lexical sets (Sinclair and Jones 1974-96), word/semantic/lexical fields (Trier 1931; Lehrer 1974), aboutness (Philips, 1989; Scott, 2000; Yablo 2016), topics (Berber Sardinha 1997), and subject matter (Schütze 1998).

For example, analyzing lexical variation in Applied Linguistics journals, Berber Sardinha (2016) identified dimensions of lexical variation in a diachronic analysis from the main historic periods on the *TESOL Quarterly* journal based on a corpus made up by the journal publications from 1967 to 2014.

Besides explaining the textual variation through lexical cooccurrence, the Lexical Multidimensional Analysis can also identify social representations. For example, Berber Sardinha (2014; 2019) identified five lexical dimensions of Brazilian and North American social representations based on 200 years of publications in Google Books. The lexical dimensions related to social representations in the Brazilian culture were: 'Economy and politics'; 'Traditional arts, sciences, the people and the land'; 'Raw material and landscape'; 'New artistic ways, women and men, religion and environment'; 'Monarchy, steam transportation and jungle'. The lexical dimensions in the North

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

American culture were 'Superpower versus developing country'; 'Developing nation'; 'The American people'; 'War and liberty'; 'Sophisticated versus primitive society'. In addition to identifying the social dimensions, he also showed how these social representations varied through time. The first dimension related to Brazil is anchored in the 1970s and 1980s, reflecting two contrasting economic periods. The 1970s are characterized by the so-called "economic miracle," a time of rapid growth and economic expansion.

In contrast, the 1980s mark a period of deep financial crisis, defined by soaring external debt and hyperinflation. The first North American dimension presents two contrasting poles, reflecting distinct historical periods. The 1940s to 1960s mark the height of "American power," driven by the United States' pivotal role in World War II and its emergence as a global superpower. In contrast, the early 1800s to 1810s emphasize a more regionally focused status, highlighting the nation's influence within a localized geopolitical context.

The study proposed here aims at contributing to the study of large-scale representations from a corpus perspective by applying a novel methodological approach to describe, through a Multidimensional Analysis perspective, the scope of lexical variation found in Google Books publications in English, focusing at the bigrams North American, Central American, Latin American and South American from a 200-year period. Thus, it sets out to identify the major social representations within the studied corpus. Although grounded in the multidimensional framework proposed by Biber (1988), this study operationalizes the model exclusively at the lexical level, following the principles of Lexical Multidimensional Analysis as formulated by Berber Sardinha and Fitzsimmons-Doolan (2025).

Building on previous applications of Lexical Multidimensional Analysis to national cultures, the present study extends this approach to Pan-American regions. The analysis focuses on noun phrase trigrams anchored by the adjective-noun sequences *North American*, *Central American*, *Latin American*, and *South American*, which function as recurrent and conventionalized discursive labels in English-language publications used to index, compare, and contrast regions within the Americas. These trigrams provide an empirically grounded entry point for examining how large-scale regional representations are constructed over time.

Within this framework, *American identity* is not as a homogeneous or essentialized category, but as a relational construct that emerges from systematic lexical contrasts between North America and other American regions. By comparing the co-occurrence patterns associated with these four regionally anchored bigrams across a 200-year period, the study seeks to identify the major social representations through which English-language publications have

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

articulated broader understandings of the Americas and their internal distinctions.

As with previous studies relying on Google Books data, issues of corpus representativeness and technical limitations must be acknowledged. Prior research has noted biases related to sampling, time periods, genre distribution, and language varieties, as well as OCR-related errors in historical texts (Pechenick et al., 2015). Despite these limitations, Google Books remains a valuable resource for large-scale diachronic analyses, particularly when the focus is on relative patterns of lexical variation rather than on absolute frequency counts.

## **2. Method**

The identification of lexical dimensions of variation (Berber Sardinha 2014, 2019, 2020); Fonseca de Araújo & Berber Sardinha 2018) for each sub-dataset which formed the dataset was performed as follows. Trigram collocates were selected from the Google Ngrams: <http://storage.googleapis.com/books/ngrams/books/dataset/v2.html>

All chosen trigrams began with the following four bigrams: "North American"; "Latin American"; "Central American"; and "South American". via a Unix Shell script, specially developed for this study, which downloaded and converted to text format all the lists of trigrams for each one of the four datasets from Google Books provided by the Google Ngram Viewer tool.

The dataset is made of four sub-datasets (North American, Latin American, Central American and South American), that is, lists of trigrams found in publication in Google Books corpus in English. Google Books corpus in English is about half trillion words. Since fewer books were published before 1800, we looked into 99.6% of the total, that is 21 decades with approximately 4.5 million publications, almost 2.25 billion pages, which result in nearly 500 billion tokens.

The dataset was compiled and processed by a script designed specifically for this study, through data extraction of the trigrams formed by the bigrams 'North American, Latin American, Central American and South American' followed by the next word. Part-of-speech tagged trigrams were eliminated (since they are the duplication of untagged trigrams). Only the trigrams in which the third word was written in lowercase letters and had at least three letters (thus avoiding proper names, punctuation, and nonwords) were maintained. Trigrams containing words from a stoplist were deleted.

To account for variations in corpus size over time, trigram frequencies were normalized per million words for each year. Next, for each decade, the 150 most frequent trigrams were identified, and their occurrence rates were further normalized per thousand words published in that decade. From this refined set, the 50 most frequent

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

third words in trigrams were selected (e.g., in North American, the word culture).

To assess how lexical patterns evolved over time, Z-scores were calculated for each variable, standardizing the factor values across individual years. These scores were then averaged by decade, providing a clearer view of long-term shifts in thematic dimensions. To statistically evaluate these trends, one-way ANOVA tests were conducted, determining whether decade-level changes in factor scores were significant. The  $R^2$  values presented in Table 3.2 quantify the proportion of variance explained by these temporal shifts, highlighting the extent to which lexical representations fluctuated over time. These statistical measures, alongside graphical representations, offer a robust framework for understanding diachronic variation in lexical dimensions.

This study employs Lexical Multidimensional Analysis to examine lexical dimensions of variation (Berber Sardinha, 2014), focusing on how social representations of North, Latin, Central, and South Americans evolved over time. These representations are constructed through semantic preference (Hoey, 2005), where words group together based on shared meaning. The analysis examines frequent trigrams in Google Books publications (1800–2008), retrieved via the Google Ngrams Viewer.

The dataset was processed in SPSS 23 and subjected to factor analysis, where trigrams were treated as variables and years as cases. Using Principal Axis Factoring (PAF) (Gómez, 2013), eigenvalues were plotted on a scree plot to determine where factors ceased to contribute significantly to variance (Figures 1–4). Multicollinearity checks established five factors for North American, six for Latin American, five for Central American, and seven for South American datasets. Variables with commonalities below 0.20 were removed.

A rotated factorial extraction using the Promax method clarified the factor structure. Variables with factor loadings below 0.30 or high cross-loadings were eliminated. Interfactorial correlations assessed factor independence.

Each factor was then interpreted based on the lexical items loading onto it, identifying recurring thematic dimensions. These factors will be analyzed qualitatively in the next section to assess their linguistic and cultural significance over time. A factor matrix was generated for each dataset, eliminating variables with weights below 0.30 or higher weights in other factors. Factor analysis revealed five factors for *North America*, six for *Latin America*, five for *Central America*, and seven for *South America*. These results will be analyzed qualitatively in the next section to interpret factors in terms of lexical dimensions.

To evaluate the significance of lexical shifts over time, ANOVA tests were conducted on each dataset, with all results reaching

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

statistical significance ( $p < 0.001$ ). The analysis confirms that decade-level changes account for a substantial proportion of variance in lexical dimensions, as indicated by the  $R^2$  values. While the specific values vary across factors, the overall trend suggests meaningful diachronic fluctuations. Given the consistent statistical significance, a full table is unnecessary here but remains available for scrutiny by referees. These findings reinforce the robustness of the observed lexical patterns and support further qualitative interpretation.

### 3. Lexical Dimensions of Variation

As stated above, four analyses were performed: *Central American* + \_\_, *Latin American* + \_\_, *North American* + \_\_\_\_ and *South American* + \_\_\_\_\_, revealing the existence of 22 factors in the data (5 for *Central*, 6 for *Latin*, 5 for *North* and 7 for *South*).

#### 3.1. – Lexical Dimensions for the North American Trigrams

Dimensions from the *North American* trigram collocates were based upon the interpretation of the variables and their respective loadings, on each factor, as seen in the following table.

Table 1: Factorial pattern for *North American*

| Factors                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | % of variance |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| <b>Factor 1: American Society and Culture</b><br>women (1.07), native (1.06), men (1.04), culture (1.03), market (1.01), plate (.99), counterparts (.98), society (.95), auto (.95), population (.93), tour (.93), children (.91), cities (.91), experiences (.91), standard (.89), universities (.88), markets (.87), economy (.87), model (.85), literature (.84), free (.82), counterpart (.80), countries (.79), art (.79), serial (.76), city (.74), history (.72), land (.71), deserts (.69), companies (.65), region (.64), slave (.63), standards (.59), life (.58), desert (.58), prairie (.38), ice (.31), soil (.30) | 23.3%         |
| <b>Factor 2: Economic Power</b><br>imperialism (.87), capital (.85), interests (.81), interest (.75), blastomycosis (.64), nation (.61), oil (.59), business (.51), waters (.51), public (.50), people (.50), mainland (.49), student (.43), text (.38), type (.38), prairies (.35), wilderness (.33)                                                                                                                                                                                                                                                                                                                           | 7.1%          |
| <b>Factor 3: Fauna</b><br>form (.78), forms (.74), species (.72), fresh (.66), province (.65), fauna (.53), birds (.52), bird (.51), genus (.51), ant (.39), port (.37), review (.36), shore (.32), water (.31), mammals (.31), animal (.30)                                                                                                                                                                                                                                                                                                                                                                                    | 3.8%          |
| <b>Factor 4: Native People and Colonial Possessions</b><br>savage (.51), possessions (.50), languages (.49), tribes (.39), sea (.38), coasts (.32), lake (.30), tribe (.40), plantations (.31)                                                                                                                                                                                                                                                                                                                                                                                                                                  | 2.5%          |
| <b>Factor 5: Aborigines and Geography</b><br>coast (.50), races (.48), aborigines (.46), origin (.44), race (.37), flora (.35), tree (.35)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 1.8%          |

Fonte: the authors.

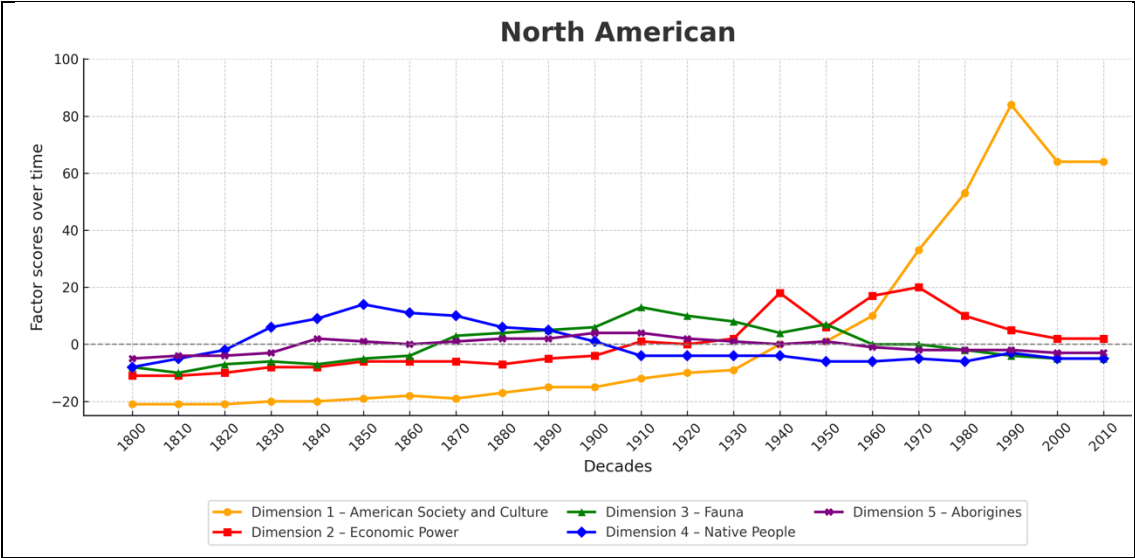
Notes: - Variance explained: 38.5%.  
- Bartlett's Test of Sphericity:  $\chi^2 = 503.65, p < 0.001$

Table 2: *North American* Factor Correlation Matrix

|          | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|----------|----------|----------|----------|----------|
| Factor 2 | 0.55     |          |          |          |
| Factor 3 | 0.00     | 0.35     |          |          |
| Factor 4 | -0.36    | -0.46    | -0.19    |          |
| Factor 5 | 0.09     | 0.33     | 0.54     | -0.18    |

Fonte: the authors.

Figure 1 – Factor loading scores over time for *North American*



Fonte: the authors.

Dimension 1, *American Society and Culture* shows two main themes. First, *North America* appears to be associated with political topics such as women (1.07), native (1.06), history (0.72), slave (0.63), and land (0.71), reflecting discussions on gender rights, Indigenous sovereignty, historical narratives, systemic inequality, and territorial disputes, with notable emphasis in the 1990s. Second, While native (1.06) and women (1.07) have high loadings, the presence of words such as market (1.01), auto (0.95), economy (0.87), universities (0.88), standards (0.59), and companies (0.65) suggests that North American culture is also strongly associated with economic and institutional themes.

These associations may reflect North America's role not only as a center of cultural influence but also as a global economic force, driven by industries such as automotive and technology, as well as trade agreements like NAFTA and broader globalization trends. This factor,

therefore, captures a multifaceted representation of North America, intertwining political issues related to gender and Indigenous identity with economic expansion, industrial development, and institutional frameworks.

Factor 2, *Economic Power*, exhibits a strong correlation (0.55) with Factor 1, *American Society and Culture*, as shown in the correlation matrix. This suggests that North America's economic dominance is an extension of its broader societal structures rather than a separate phenomenon. While the original researchers emphasize imperialism, capital, and Cold War influence, the lexical data—imperialism (.87), capital (.85), economy (.87), business (.51), oil (.59)—indicates a broader theme of economic expansion and resource control. Rather than an independent dimension, Factor 2 appears as a specialized subset of Factor 1, reinforcing North America's interconnected economic and institutional influence.

Factor 3, *Fauna*, and Factor 5, *Aborigines and Geography*, exhibit a strong correlation (0.54) in the correlation matrix, highlighting the close relationship between North America's biodiversity and geographic landscape. Factor 3—species (.72), fauna (.53), birds (.52), genus (.51), mammals (.31)—focuses on biological classification, while Factor 5—coast (.50), aborigines (.46), origin (.44), flora (.35), tree (.35)—emphasizes land, indigenous presence, and natural environments. This suggests that discussions of North American fauna are not purely scientific but also tied to geography, indigenous ecological knowledge, and environmental narratives, reinforcing how biodiversity is studied in relation to human history and geographic landscapes rather than in isolation.

Factor 4, *Native People and Colonial Possessions*, reflects a historical and sociopolitical discourse on indigenous identity, colonial expansion, and territorial control, as seen in savage (.51), possessions (.50), languages (.49), and tribes (.39). Notably, it has a strong negative correlation (-0.461) with Factor 2 (Economic Power) in the correlation matrix, suggesting a contrasting historical trajectory—while Factor 2 emphasizes economic dominance and expansion, Factor 4 highlights colonial dispossession and indigenous struggles, reinforcing competing discourses on North America's historical development.

### **3.2 – Lexical Dimensions for *Latin American Trigram***

The dimensions from the *Latin American trigram* collocates were based upon the variables interpretation and their respective loadings, in each factor, which can be found on the table below.

Table 3: Factorial pattern for Latin American

| Factors                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | % of variance |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| <b>Factor 1: <i>Political Revolution and Economic Development, positive pole:</i></b><br>common (1.22), university (1.11), integration (1.08), universities (1.04), revolution (1.03), parties (1.02), communist (1.00), scene (.99), development (.98), economic (.98), nationalism (.97), society (.94), economy (.93), area (0.92), political (0.90), politics (0.88), military (.83), student (.83), population (.81), continent (.80), agriculture (.80), national (.77), group (.77), capital (.74), students (.71), people (.68), government (.67), peoples (.67), governments (.67), world (.66), social (.64), revolutions (.64), labor (.61), nation (.58), export (.58), urban (.58), constitutions (.57), business (.57), exports (.56), affairs (.55), nations (.55), states (.54), life (.54), intellectuals (.54), state (.53), leaders (.51), societies (.51), statesmen (.51), leader (.50), diplomats (.50), intellectual (.50), delegates (.49), dictators (.47), policy (.46), member (.46), capitals (.44), problem (.43), studies (.43), bishop (.42), independence (.41), culture (.41), economist (.39), work (.38), region (.38), view (.37), land (.37), delegations (.37), friends (.37), diplomat (.36), ports (.36), foreign (.33) | 38.6          |
| <b>Factor 2: <i>Immigrants, Cultures and Women</i></b><br>immigrants (1.15), literary (1.11), writer (1.06), cultures (1.04), women (1.02), cultural (1.01), fiction (.97), literature (.97), immigrant (.96), writers (.96), novel (.95), case (.92), men (.91), context (.91), liberation (.87), art (.84), experience (.77), economies (.74), culture (.72), intellectual (.72), region (.71), film (.71), markets (.69), studies (.69), intellectuals (.69), societies (.64), migrants (.63), market (.60), bishops (.60), music (.59), theologians (.57), history (.56), church (.56), debt (.53), theology (.49), social (.48), relations (.45), leaders (.45), democracy (.38), civil (.35), debtors (.32), military (.31), independence (.30)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | 12.6          |
| <b>Factor 3: <i>Worker and Republics, positive pole:</i></b><br>worker (.74), republics (.69), republic (.68), lands (.62), ties (.59), neighbors (.59), neighbor (.56), dictator (.55), public (.55), delegates (.54), statesmen (.52), constitution (.52), powers (.49), story (.49), capitals (.48), diplomats (.48), delegate (.47), periodicals (.46), cities (.46), friends (.44), dictators (.44), field (.43), power (.42), produce (.41), constitutions (.41), states (.41), state (.41), life (.40), news (.39), music (.39), problem (.37), land (.37), delegations (.36), affairs (.33), civilization (.30)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 5.1           |
| <b>Factor 3, <i>Worker and Republics, negative pole</i></b><br>tariff (-.44), integration (.40), communist (-.35), product (-.34), university (-.33)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |               |
| <b>Factor 4: <i>Importing:</i></b><br>tariff (1.15), product (1.14), import (1.11), producer (1.11), products (.70), imports (.69), ports (.63), export (.53), students (.53), student (.50), exports (.49), trade (.48), foreign (.35), neighbors (.32)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 3.8           |
| <b>Factor 5: <i>Businesspeople:</i></b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 3.0           |

|                                                                                                                                                                                                                                 |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| buyer (.99), merchants (.99), importer (.99), merchant (.97), custom (.71), business (.60), field (.57), papers (.49), trade (.48), port (.47), power (.40), view (.34)                                                         |     |
| <b>Factor 6, Debts and Religions, positive pole:</b><br>debtor (.84), debtors (.78), debt (.64), church (.57), theology (.55), theologians (.55), relations (.39), bishops (.35), member (.34), foreign (.34), liberation (.32) | 2.7 |
| <b>Factor 6, negative pole:</b><br>immigrant (-.55), men (-.38), migrants (-.42), civil (-.36), immigrants (-.34), cultural (-.32)                                                                                              |     |
|                                                                                                                                                                                                                                 |     |

Fonte: the authors.

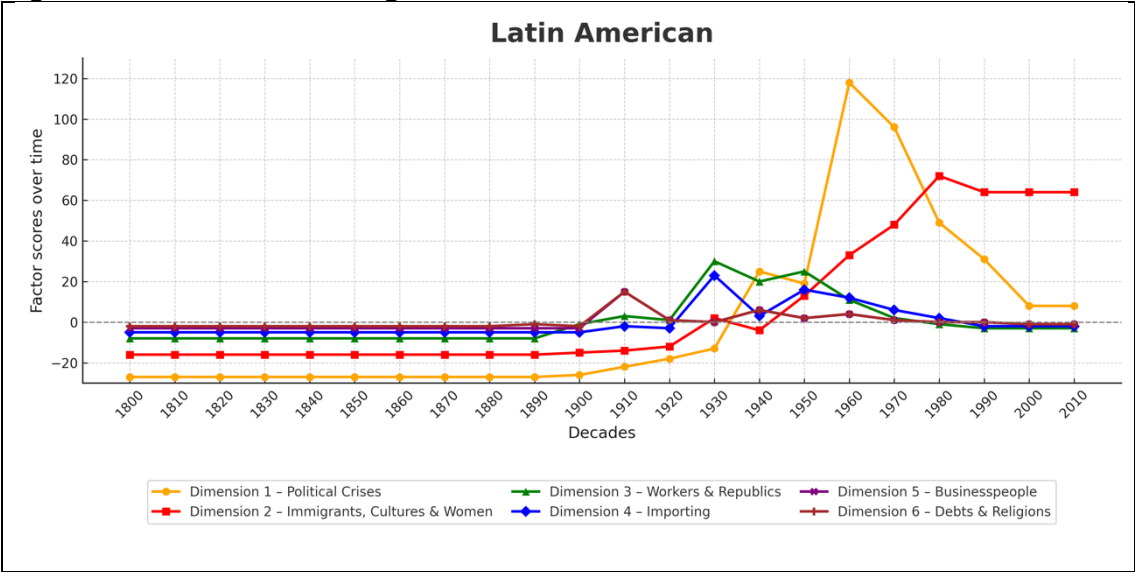
Notes: - Variance explained: 65.9%.  
 - Bartlett's Test of Sphericity:  $\chi^2 = 161.00$ ,  $p < 0.001$

Table 4: *Latin American* Factor Correlation Matrix

|          | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 |
|----------|----------|----------|----------|----------|----------|
| Factor 2 | 0.53     |          |          |          |          |
| Factor 3 | 0.57     | 0.19     |          |          |          |
| Factor 4 | 0.44     | 0.13     | 0.62     |          |          |
| Factor 5 | 0.15     | 0.03     | 0.25     | 0.23     |          |
| Factor 6 | 0.41     | 0.50     | 0.15     | 0.08     | 0.01     |

Fonte: the authors.

Figure 2 – Factor loading scores over time for *Latin American*



Fonte: the authors.

The first dimension reflects concerns over *Political Revolution and Economic Development* (Factor 1), which stands as the dominant discourse on Latin America. The key lexical items, revolution (1.03) and development (0.98), highlight the region's struggles with governance, political change, and economic policies. The discourse focuses on nation-building, ideological shifts, and economic restructuring, aligning with peaks in the 1940s and 1960s, periods

marked by labor populism, economic reforms, and shifting trade policies. This factor is closely linked to *Workers & Republics* (Factor 3, positive) and *Importing* (Factor 4) due to their high correlations (0.57, 0.62, 0.44), suggesting a broader theme connecting political and economic systems, labor movements, and trade policies. The economic dimensions of political revolutions are reflected in terms such as worker (0.74), republics (0.69), and tariff (1.15), illustrating how labor and trade policies were embedded in governance.

Another dimension emerges from *Immigrants, Cultures & Women* (Factor 2) and *Debts & Religions* (Factor 6, Positive), which are correlated (0.50) and form a separate theme centered on social identity, migration, and religious institutions. The first major theme, migration and cultural identity, is marked by immigrants (1.15), literary (1.11), and women (1.02), reflecting discourse on diaspora, gender, and cultural production. The second, religion and economic struggles, is seen in debtor (0.84), debtors (0.78), and church (0.57), highlighting the intersection of financial crises and religious influence, particularly in the 1980s and 1990s.

Finally, *Businesspeople* (Factor 5) represents a specialized economic theme that does not strongly align with governance or labor discourse (correlations  $\leq 0.25$ ). Discussions on merchants (0.99), importers (0.99), and trade (0.48) were largely independent, peaking in the 1910s, when economic ties between Latin America and the United States intensified.

### 3.3 – Lexical Dimensions for Central American Trigram

Dimensions from the *Central American* trigram collocates were based upon the variables interpretation and their respective loadings, in each factor, as shown in Table 5.

Table 5: Factorial patterns for Central American

| Factors                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   | % of variance |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------|
| <b>Factor 1: Political Crisis</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 20.9%         |
| crisis (.99), policy (.98), conflicts (.97), policies (.96), conflict (.96), issues (.93), revolution (.90), problem (.90), issue (.86), situation (.85), refugees (.84), leaders (.83), refugee (.82), war (.79), revolutions (.78), governments (.77), nations (.75), neighbors (.72), government (.72), nation (.72), political (.71), peace (.66), society (.66), relations (.65), region (.65), country (.61), leader (.61), economies (.60), countries (.55), exports (.54), isthmus (.53), history (.48), unity (.44), politics (.44), dictators (.41), people (.41), capitals (.41), affairs (.35), population (.33), peoples (.31), dictator (.31), export (.30) |               |
| <b>Factor 2: Civilizations and Colonies</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |               |
| republics (.82), republic (.78), lands (.73), civilizations (.69), port (.67), highlands (.66), coast (.61), form (.59), jungle (.59), capitals (.57), court (.56), civilization (.55), species (.55),                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |               |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |       |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| highland (.52), conference (.52), treaties (.52), waters (.52), colonies (.51), art (.50), coffee (.48), land (.46), origin (.44), type (.43), forms (.43), peoples (.41), ports (.41), capital (.41), sculpture (.39), affairs (.39), state (.39), regions (.38), treaty (.38), group (.38), banana (.37), volcano (.36), countries (.36), states (.36), area (.35), mahogany (.33), affair (.33), mainland (.33), bird (.32), rubber (.32), dictators (.32), plant (.31), birds (.31), provinces (.30) | 10.0% |
| <b>Factor 3: Economy, Common Market and Nature</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                       | 4.8%  |
| integration (.97), economic (.95), common (.88), cultures (.80), market (.62), export (.62), culture (.61), economies (.58), exports (.55), area (.54), birds (.46), land (.45), high (.42), mainland (.41), trade (.40), population (.39), national (.39), banana (.36), sea (.34), countries (.33)                                                                                                                                                                                                     |       |
| <b>Factor 4: Independence, Politics and Geography</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                    | 3.5%  |
| confederation (1.03), volcanoes (.99), independence (.99), federation (.80), territory (.80), union (.70), states (.64), confederacy (.60), coasts (.59), state (.57), politics (.50), city (.40), coast (.36), tree (.31), provinces (.30), forms (.30)                                                                                                                                                                                                                                                 |       |
| <b>Factor 5: Migration</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | 3.4%  |
| immigrant (.90), migrants (.87), immigrants (.86), men (.76), civil (.75), women (.60), migrant (.55), forest (.49), population (.37), peace (.36), forests (.32), refugee (.34), refugees (.32)                                                                                                                                                                                                                                                                                                         |       |

Fonte: the authors.

Notes: - Variance explained: 42.5%.

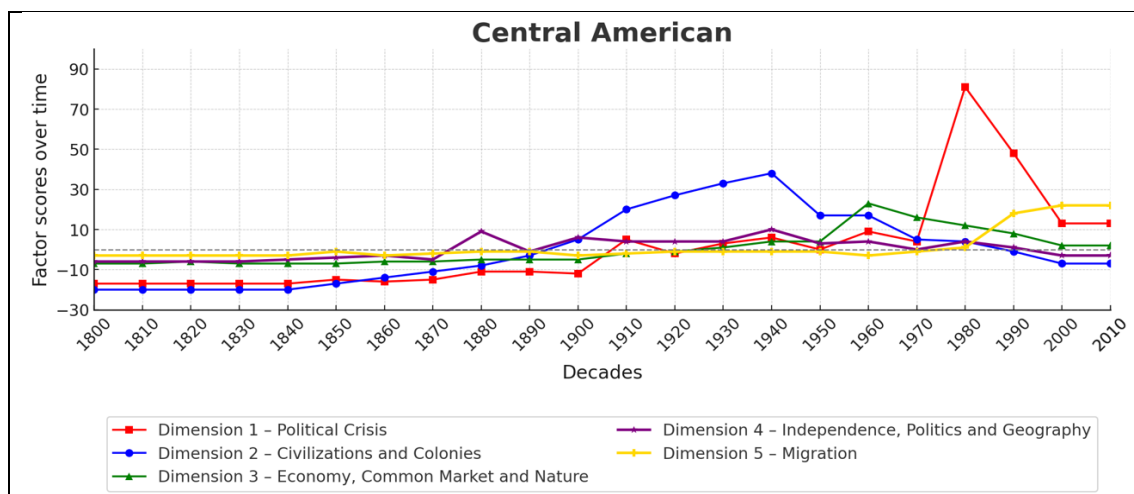
- Bartlett's Test of Sphericity:  $\chi^2 = 85.65$ ,  $p < 0.001$

Table 6: *Central American* Factor Correlation Matrix

|          | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|----------|----------|----------|----------|----------|
| Factor 2 | 0.30     |          |          |          |
| Factor 3 | 0.45     | 0.33     |          |          |
| Factor 4 | 0.30     | 0.46     | 0.15     |          |
| Factor 5 | 0.25     | -0.18    | 0.26     | -0.19    |

Fonte: the authors.

Figure 3 – Factor loading scores over time for *Central American*



Fonte: the authors.

The dimension accounting for most of the variance reflects concerns over *Political Crisis* (Factor 1), which stands as the dominant discourse on *Central America*. The key lexical items, *crisis* (0.99) and *policy* (0.98), highlight the region's struggles with political instability, conflicts, and governance issues. The discourse focuses on wars, revolutions, refugee movements, and government policies, aligning with peaks in the 1980s, a period marked by civil wars, U.S. intervention, and Cold War geopolitics.

The second factor *Civilizations and Colonies* (Factor 2) is closely linked to *Economy, Common Market & Nature* (Factor 3) and *Independence, Politics & Geography* (Factor 4) due to their moderate correlations (0.45, 0.30), suggesting that discussions of economic integration and political independence were entangled with broader regional crises. The economic dimensions of political instability are reflected in terms such as *economies* (0.60), *exports* (0.54), and *market* (0.62), demonstrating how trade and economic policies were deeply affected by political conditions.

Another dimension emerges from *Civilizations & Colonies* (Factor 2) and *Independence, Politics & Geography* (Factor 4), which are strongly correlated (0.46), reflecting how historical, geographic, and political identity shaped discourse. The first major theme, civilizations and colonial legacies, is marked by *civilizations* (0.69), *colonies* (0.51), and *republics* (0.82), reflecting discussions on pre-Columbian societies and colonial rule, peaking in the 1940s. The second, independence and geography, is seen in *independence* (0.99), *confederation* (1.03), and *volcanoes* (0.99), highlighting the geopolitical landscape of Central America, also peaking in the 1940s.

Finally, *Migration* (Factor 5) represents a distinct socio-political theme, weakly correlated with other factors ( $\leq 0.26$ ). Discussions on *immigrants* (0.90), *migrants* (0.87), and *refugees* (0.84) suggest a focus on population movements, displacement, and demographic shifts, peaking in the 2000s, when migration flows toward the U.S. intensified.

### **3.4 – South American Trigrams Lexical Dimensions**

Dimensions from *South American* trigram collocates were based upon the interpretation of the variables and their respective loadings, in each factor, which can be found on the table below.

Table 7: Factorial patterns for South American

| <b>Factors</b>                                                                                                                                                                                                                                                                                                                                                                                                                                           | <b>% of variance</b> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|
| <b><i>Factor 1: Neighboring Nations and Jungle</i></b><br>neighbor (.104), neighbors (.91), nations (.87), nation (.86), route (.85), papers (.72), news (.71), trypanosomiasis (.69), capital (.61), governments (.56), jungle (.55), product (.51), tour (.49), service (.49), market (.48), history (.44), lands (.41), mainland (.38), markets (.36), city (.34), trip (.33), dictators (.33), blastomycosis (.33)                                   | 16.0%                |
| <b><i>Factor 2: Business</i></b><br>trade (.87), business (.78), commerce (.75), port (.74), ports (.74), states (.68), republic (.67), republics (.67), state (.63), expedition (.63), cities (.57), waters (.57), markets (.56), market (.56), people (.55), city (.54), merchant (.54), peoples (.48), public (.48), town (.40), government (.40), policy (.39), politics (.37), wool (.33), region (.31), history (.31), types (.31), question (.30) | 8.6%                 |
| <b><i>Factor 3: Cultures and Nature</i></b><br>cultures (.80), rain (.74), rodent (.72), culture (.70), plate (.62), mainland (.62), mammals (.59), land (.58), group (.57), mammal (.55), (jungle (.53), origin (.48), fish (.45), tour (.45), monkey (.43), sea (.41), native (.39), dictator (.39), myth (.35), men (.33), monkeys (.32), population (.32), blastomycosis (.32)                                                                       | 4.4%                 |
| <b><i>Factor 4: Geography and Fauna</i></b><br>bird (.75), form (.68), forms (.66), species (.60), birds (.59), fauna (.51), forests (.51), coast (.51), genus (.49), pampas (.49), forest (.48), races (.42), ostrich (.42), tribes (.39), type (.39), monkeys (.39), types (.39), coasts (.37), tribe (.36), origin (.34)                                                                                                                              | 3.2%                 |
| <b><i>Factor 5: Mission Affairs</i></b><br>affairs (.97), mission (.83), privateer (.71), cause (.70), insurgents (.65), patriot (.63), independence (.61), question (.61), fish (.42), patriots (.34)                                                                                                                                                                                                                                                   | 2.5%                 |
| <b><i>Factor 6: Travels and Landscape</i></b><br>rivers (.76), river (.67), savage (.60), plains (.53), horse (.46), forest (.45), forests (.43), tree (.41), shore (.40), plain (.37), ores (.33), mines (.31), traveler (.30)                                                                                                                                                                                                                          | 1.7%                 |
| <b><i>Factor 7: Mining and Emancipation of the Provinces</i></b><br>station (.66), emancipation (.51), liberty (.49), independence (.47), mining (.44), patriots (.42), mines (.38), savages (.34), provinces (.34), volcanos (.34), plants (.33), bird (.33), governments (.32), commerce (.31)                                                                                                                                                         | 1.6%                 |

**Fonte:** the authors.

Notes: - Variance explained: 37.9%.

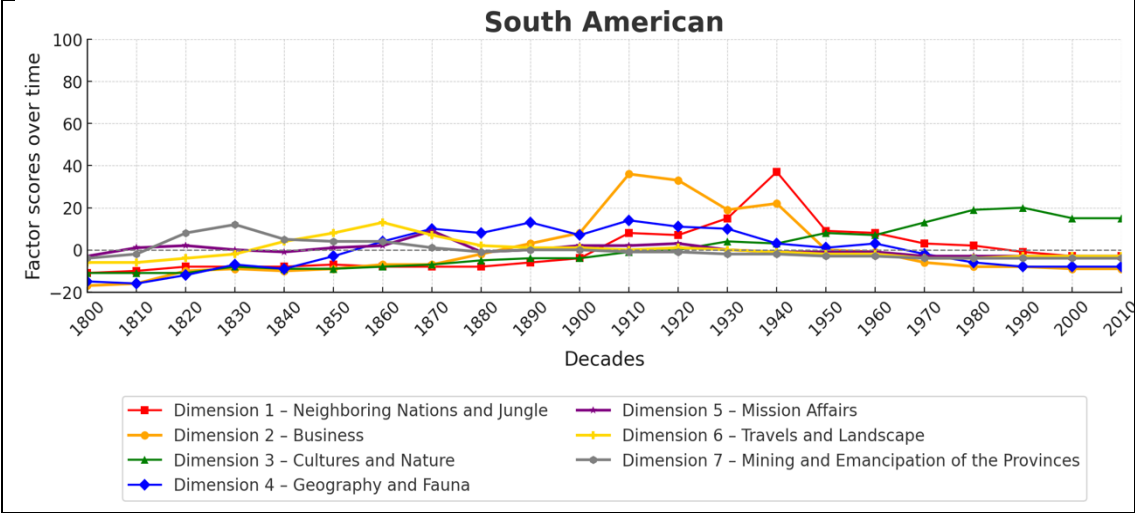
- Bartlett's Test of Sphericity:  $\chi^2 = 138.16$ ,  $p < 0.001$

Table 8: *South American* Factor Correlation Matrix

|          | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 |
|----------|----------|----------|----------|----------|----------|----------|
| Factor 2 | 0.55     |          |          |          |          |          |
| Factor 3 | 0.35     | 0.06     |          |          |          |          |
| Factor 4 | 0.34     | 0.57     | 0.09     |          |          |          |
| Factor 5 | 0.07     | 0.26     | -0.16    | 0.16     |          |          |
| Factor 6 | -0.15    | 0.04     | -0.25    | 0.20     | 0.04     |          |
| Factor 7 | -0.06    | 0.03     | -0.38    | -0.14    | 0.27     | 0.12     |

Fonte: the authors.

Figure 4 - Factor loading scores over time for *South American*



Fonte: the authors.

The dimension accounting for most of the variance reflects concerns over neighboring nations, economy, and trade (Factor 1 & Factor 2), which stands as the dominant discourse on South America. The key lexical items, *neighbor* (1.04), *nations* (.87), *trade* (.87), and *business* (.78), highlight discussions surrounding regional diplomacy, commerce, and economic integration. The discourse focuses on maritime trade, market expansion, and treaties, aligning with peaks in the 1910s and 1940s, periods marked by economic treaties, resource trade, and regional governance debates.

The second dimension entails cultures, nature, and geography (Factor 3 & Factor 4), is closely linked to the aforementioned dimension denoting neighboring nations, economy, and trade (Factor 1 & 2) due to moderate correlations (0.34, 0.35), suggesting that discussions of indigenous cultures and biodiversity were shaped within broader economic and geopolitical contexts. The cultural and environmental identity of South America is evident in terms such as *cultures* (.80), *rain* (.74), *fauna* (.51), and *tribes* (.39), reflecting discourse on indigenous traditions, biodiversity, and land use policies, peaking in the 1990s with increased attention to indigenous rights and ecological conservation.

Another dimension emerges from *Mission Affairs* (Factor 5) and *Mining & Emancipation* (Factor 7), which are moderately correlated (0.27), reflecting how religious missions and independence movements were intertwined with economic structures. The first major theme, political and religious movements, is marked by *mission* (.83), *independence* (.47), and *patriot* (.63), suggesting discourse on missionary activity and post-colonial governance. The second, resource extraction, is seen in *mining* (.44) and *commerce* (.31), pointing to the economic restructuring of newly independent states, peaking in the 1830s and 1870s.

Finally, *Travel and Landscape* (Factor 6) remains a distinct theme, weakly correlated with other factors ( $\leq 0.20$ ). Discussions on *rivers* (.76), *traveler* (.30), and *ores* (.33) suggest a focus on exploration, geographic discovery, and economic expansion, peaking in the 1860s.

#### 4. Discussion

The present study aimed to measure perceptions of distinct regions of the Americas using Lexical Multidimensional Analysis (LMDA) to analyze the concepts that most frequently co-occurred with the noun phrase trigrams anchored by bigrams *North American*, *Latin American*, *Central American*, and *South American* in English-language texts. By examining the socio-historical panorama of the Americas, we gain insight into how these regions have been represented over time, shaped by varying contexts in which the inhabitants of each region play a central role.

Before turning to region-specific findings, a cross-regional overview reveals that representations across the four datasets are recurrently structured around economic, political, cultural, and geographic dimensions, albeit with region-specific configurations and emphases. In Central America, lexical associations reflect an interplay of political crises, ancient history, natural resources, and migration. Latin America, on the other hand, is primarily associated with progressive politics, economic structures, and cultural production. South America is depicted as a distant, exotic land characterized by wildlife, disease, and paramilitary activity. In contrast, North America's dominant themes suggest a position of power, as its associated concepts cluster around economic influence and governance. These patterns reflect historical realities and evolving discourses.

The most frequent trigrams in Google Books publications provide insights into how Anglo-American discourse has shaped perceptions of the Americas. Although these representations may fluctuate depending on English-speaking populations worldwide, the data largely capture how North American and British authors conceptualized these regions.

The following sections explore thematic continuities and changes over time.

Some conceptual associations remained stable over time. Notably, references to South America's religious missions showed little diachronic variation, reinforcing its historical portrayal as a land of missionary work. Similarly, lexical items related to Latin American literary and intellectual production—such as literary, writer, and culture—persisted across time, reflecting a long-standing narrative about the region's intellectual contributions.

Despite these constants, significant diachronic variation emerged in four major thematic domains: culture, geography, politics, and economy. Collocates associated with North and Latin America emphasize culture, while North and South America share references to fauna and geography, especially in early 20th-century texts. This shared emphasis on natural landscapes aligns with prior scholarship on environmental discourse (Walls, 2009). Additionally, North American discourse often linked Central America to geographic concerns and colonial possessions, reinforcing a historical view of the region as a politically and economically dependent entity. The relative absence of geographic terms in Latin American discourse may suggest that the region is conceptualized more in political than geographic terms.

In the North American dataset, factors such as society, power, and native people emerged as distinct themes. The strong lexical associations between American society and economic dominance suggest that North America is portrayed as both a cultural and economic hegemon. This is further illustrated by peaks in economic discourse during the 1970s, corresponding to shifts in U.S. economic policies and global influence. While Factors 1 and 2 are moderately correlated ( $r = 0.55$ ), this value indicates related but analytically distinct construct rather than a hierarchical relationship, with economic power representing a specific dimension within a broader representational space. A similar theme emerges in the Latin American dataset, where the dimensions political revolution and economic development highlight the region's historical engagement with governance, labor movements, and trade policies. The following passage exemplifies the economic discourse surrounding Latin America:

- (1) "He has considered extremely risky to extend longer terms to Latin American merchants." (Barrett & Moreno-Lacalle, 2010, p. 60).

In comparing Latin and South America, shared topics include political crises, economic instability, and business. Publications from the 20th century frequently highlight economic uncertainty, reinforcing stereotypes about Latin America's financial instability. The following excerpt illustrates these concerns:

- (2) "The deep financial and economic crisis that struck Latin America in the 1980s simultaneously underlined the extent to which significant U.S. economic interests and the health of the international financial system depend on the continued economic viability of major Latin American debtor countries." (Middlebrook & Rico, 1986, p. 3).

The Latin American and Central American datasets share topics related to migration and political instability. The mass exodus from these regions to the United States, especially in the 1990s, is reflected in discourse surrounding migration. The following excerpt underscores the cultural integration of Latin American immigrants in the U.S.:

- (3) "However, Latin American immigrants share at least as many aspects of U.S. political culture—experience with representative political systems, a liberal tradition, etc.—as they do differences—authoritarianism, etc." (Jones-Correa, 1994, p. 10).

North America's unique dimensions include references to women's societal roles and Indigenous peoples. These themes are largely absent in other datasets, suggesting a more developed discourse on gender and Indigenous identity within North America. In contrast, Latin American discourse highlights labor and economic marginalization, particularly regarding workers and their access to consumer goods:

- (4) "There is little doubt that the average Latin American worker would like to own an automobile and would like to see his earnings increased to a level that would permit him to buy one." (The Inter-American Quarterly, 1941, p. 1).

Central and South America share thematic links concerning nature and independence. However, Central American discourse uniquely emphasizes trade agreements, while South America is associated with mining and resource extraction. The following passage illustrates early discourse on Central American economic integration:

- (5) "One author, however, holds that 'the first concrete step in the organization of the Common Market can be said to have been taken in 1952, with the formation in Tegucigalpa of the Committee on Central American Economic Cooperation.'" (Cochrane, 1969, p. 51).

The economic discourse surrounding Central America is also reflected in literary critiques, such as Pablo Neruda's (1950) "The United Fruit Company," which critiques the economic exploitation of the region:

- (6) "Disembarks, / ravaging coffee and / fruits / for its ships that spirit away / our submerged lands' treasures / like serving trays." (Bleiker, 1999, p. 1129).

Latin America's cultural production remains an important theme, particularly regarding literary contributions. The emergence of Nobel laureates in literature from the region in the 20th century aligns with increased references to literature, as seen in the following excerpt:

- (7) *"It is not often observed that the current status of Latin American literary studies closely corresponds to the arrival of new theoretical paradigms in humanistic research."* (Campa, 1999, p. i).

North America's self-perception is also notable, with the term counterpart appearing frequently in American Society and Culture (Dimension 1). This suggests that North Americans conceptualize their role in relation to others rather than in isolation. This global perspective became more pronounced after the founding of the United Nations, reinforcing the theme of international engagement:

- (8) "The preceding chapter situates the economic, social, and political development of these lands within a historical framework defined by European (and later by Japanese and North American) imperialism."

South America's discourse similarly evolved, with increased references to trade, diplomacy, and geopolitical influence. The themes of travel and landscape (Factor 6) peaked following Alexander von Humboldt's explorations in the 19th century, illustrating how scientific and literary narratives shaped perceptions of the continent (Bunkse, 1981).

Overall, the lexical dimensions of variation presented in this study reveal persistent historical narratives and evolving regional perceptions.  $R^2$  values varied across factors, ranging from moderate to high levels of explanatory power, indicating differing degrees of diachronic sensitivity. While certain stereotypes remain unchanged, others shift in response to political, economic, and cultural developments. These findings highlight the intersection of language and social representation, demonstrating how corpus-based approaches can provide valuable insights into diachronic shifts in regional discourse.

## 5. Conclusion

This study provides insights into how English-language discourse frames the Americas, often reinforcing distinctions between North America and the rest of the continent. While this classification aligns with historical and political realities, it also oversimplifies the cultural and economic diversity of the regions. As Narloch and Teixeira (2011) suggest, treating Latin America as a single entity is as imprecise as

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

grouping diverse African populations under one label. Future research should explore how such categorizations shape global perceptions.

Additionally, metadata limitations in Google Ngrams remain a challenge (Koplenig, 2017). This study controlled for such issues by qualitatively analyzing trigram collocates, yet questions remain regarding the influence of English as a second language on representations of the Americas. Future studies should investigate whether these patterns differ across native and non-native English-speaking authors.

Finally, cross-linguistic comparisons could offer further insights. Google Ngrams includes datasets in other languages, providing opportunities to analyze how non-English discourses construct regional identities. Such research would deepen our understanding of cross-cultural differences in perceptions of the Americas.

## References

BAKER, P.; GABRIELATOS, C.; KHOSRAVINIK, M.; KRZYŻANOWSKI, M.; MCENERY, T.; WODAK, R. *Corpus linguistics and sexuality: a critical analysis of discourses of sexuality in British newspapers*. London: Routledge, 2013a.

BAKER, P., GABRIELATOS, C., MC ENERY, T. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press, 2013b.

BERBER SARDINHA, T. *Linguística de Corpus*. Barueri, SP: Editora Manole, 2004

BERBER SARDINHA, T. On Being American and Brazilian in Google Books. Paper presented at American Association for Corpus Linguistics Conference (AACL 2014), Flagstaff, AZ, 2014.

BERBER SARDINHA, T. Corpus linguistics and history: Lexical dimensions in TESOL Quarterly. Paper presented at American Association for Corpus Linguistics Conference (AACL 2016), Ames, IA, 2016.

BERBER SARDINHA, T. Using Multi-Dimensional Analysis to Detect Representations of National Cultures. In: Berber Sardinha, T. & Veirano Pinto, M. *Multi-dimensional Analysis: Research, Methods and Current Issues*. London: Bloomsbury Publishing, 2019

BERBER SARDINHA, T.; Fitzsimmons-Doolan, S. *Lexical Multidimensional Analysis*. Cambridge: Cambridge University Press, 2025.

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

BARRETT, J.; MORENO-LACALLE, J. Report of the Second Pan American Commercial Conference (1919). Washington, D.C.: Pan American Union, 1919. Fac-símile reprint. Whitefish, MT: Kessinger Publishing, 2010.

BIBER, D. Variation Across Speech and Writing. Cambridge: Cambridge University Press, 1988.

BLEIKER, R. Plablo Neruda and the struggle for political memory. *Third World Quarterly*, 20 (6), 1129-1142. Mexico Press, 1999.

BUNKSE, E. V. Humboldt and aesthetic tradition in geography. *Geographical Review*, 127-146, 1981.

CAMPA, R. Latin American literary studies in the age of globalization. Pittsburgh: University of Pittsburgh Press, 1999.

COCHRANE, James D. Central American economic integration. New York: Praeger, 1969.

DELFINO, M. C. N.; FONSECA DE ARAÚJO, R.; BERBER SARDINHA, T. (2018). Revista Brasileira de Linguística Aplicada: Multi-dimensões temáticas. In: Linguística de corpus: perspectivas [digital resource] / (Orgs.) Maria José Bocorny Finatto, Rozane Rodrigues Rebechi, Simone Sarmento, Ana Eliza Pereira Bocorny. — Porto Alegre: Instituto de Letras - UFRGS, 2018.

DURKHEIM, E. The Rules of the Sociological Method. New York: Free Press, 1938.

GÓMEZ, P. C. Statistical methods in language and linguistic research. Sheffield: Equinox, 2013.

HOEY, M. Lexical Priming. A new theory of words and language. London: Routledge, 2005.

JODELET, D. Representations sociales: un domaine en expansion. Em, D. Jodelet (Org.) Les Representations Sociales. Paris: Presses Universitaires de France, 1989.

JONES-CORREA, Michael. Between two nations: the political predicament of Latinos in New York City. Ithaca: Cornell University Press, 1994.

ARAUJO, Rafael Fonseca de; SARDINHA, Tony Berber; DELFINO, Maria Claudia Nunes; COLLENTINE, Joseph Glynn; LAKE, William Michael. Representations of American identity: a multidimensional analysis of english Google books trigrams. *Revista Intercâmbio*, v.LVIII, e75751, 2026. São Paulo: LAEL/PUCSP. ISSN 2237-759X

KOPLINIG, A. The impact of lacking metadata for the measurement of cultural and linguistic change using the Google Ngram data sets—Reconstructing the composition of the German corpus in times of WWII. *Digital Scholarship in the Humanities*, 32(1), 169-188, 2017.

MIDDLEBROOK, Kevin J.; RICO, Carlos. The politics of economic crisis in Latin America. New York: Oxford University Press, 1986

MOSCOVICI, S. Notes towards a Description of Social Representations. *European Journal of Social Psychology*, 18, 211-250, 1988.

MOSCOVICI, S., & DUVEEN, G. Social representations: Explorations in Social Psychology. Washington Square, New York: New York University press, 2001.

NARLOCH, L., & TEIXEIRA, D. Guia Politicamente Incorreto da América Latina. São Paulo: Leya, 2011.

PECHENICK, E.A., DANFORTH, C. M., DODDS, P.S. Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution. *PLoS ONE* 10(10): e0137041, 2011.  
<https://doi.org/10.1371/journal.pone.0137041>

STUBBS, M. Text and corpus analysis. Chapter 1. [S.l.]: [s.n.], 1996. Disponível em: <https://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1996-text-corpus-ch-1.pdf>. Acesso em: 09 nov. 2025.

THE INTER-AMERICAN QUARTERLY. Washington, D.C.: [s.n.], 1940–1941.

WALLS, L. D. The passage to cosmos: Alexander von Humboldt and the shaping of America. University of Chicago Press, 2009.

Recebido em: 08/11/2025  
Aprovado em: 09/03/2026



Esta obra está licenciado com uma Licença [Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/) que permite o uso irrestrito, distribuição e reprodução em qualquer meio, desde que a obra original seja devidamente citada