

The current state and future challenges of digital humanities in Latin America

Adriana Picoral¹ Tony Berber Sardinha²

¹University of Arizona
Tucson, AZ, USA
adrianaps@arizona.edu

²São Paulo Catholic University
São Paulo, SP, Brazil
tonyberber@gmail.com



Center for Latin American Studies, Ohio State University, 9/24/2021



- 1 Introduction
- 2 Pronominal choice in Brazilian Twitter
- 3 Dissemination of misinformation in Twitter
- 4 The style of Machado de Assis
- 5 Representations of Brazil in Google Books



Acknowledgments

- Thank you to the Center for Latin American Studies.
- Thank you to the Dept. of Spanish and Portuguese.
- The research reported here was supported by: CNPq (Brasília, Brazil), grants #25/2020, #02/2021, #306994/2017-8, #407788/2018-2; PiPeq PUCSP Aux-PI 13589, AuxPI 6880, PUCSP PIPRINT 18065

Plano de Incentivo à Pesquisa

PIPEq
PUC-SP



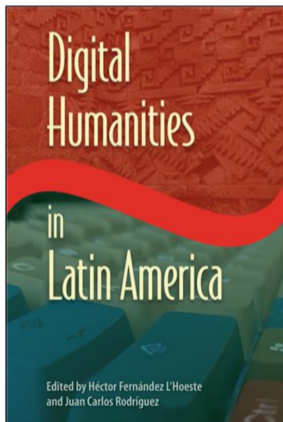
Download slides

Slides:

<https://bit.ly/3o1NP8v>



Digital Humanities in Latin America



- Brazil
 - Museu da Língua Portuguesa
 - Laboratório em Rede de Humanidades Digitais (IBICT)
 - Laboratório de Humanidades Digitais (FGV)
 - Laboratório Virtual de Humanidades Digitais (USP)
 - Portal multimodal/multilíngue para o avanço da ciência aberta nas Humanidades (cienciaaberta.org) (PUCSP)
 - Observatório de Linguagem (PUCSP)
 - CEPRIL, Centro de Pesquisas e Recursos em Linguagem (PUCSP)
- Argentina
 - Asociación Argentina de Humanidades Digitales



Digital Humanities



- Crear bases de datos con recursos digitales relevantes para las Humanidades. Esto incluye la captura, estructuración, documentación, preservación y disseminación de los datos.
- Desarrollar metodologías que permitan generar nuevos elementos derivados de estos datos.
- Generar investigación y conocimiento para incrementar nuestra comprensión en las Humanidades.

(Galina Russell, 2011, p. 3)



Transition

Adriana

Digital Humanities

- What is Digital Humanities? No single definition
- Humanities + Computational Methods
- Focus of the talk: overview of different methods in language studies
- Examples provided in this talk:
 - fine-grained study in pronominal variation
 - exploratory study in cross-linguistic misinformation “spread”
 - multidimensional analysis of Machado de Assis
 - representations of Brazil in Google Books



Language Variation in Brazilian Portuguese Tweets

- Pronominal variation in spoken Portuguese has been extensively studied by the field of sociolinguistics (see for example Vianna, & dos Santos Lopes, 2012 for first person pronoun studies and Scherre, Andrade & Catão, 2021 for second person pronoun).
- Binary choice:
 - *tu* and *você* (second person singular) – *vossa mercê* (your mercy)
 - *nós* and *a gente* (first person plural) – innovative pronoun *a gente* (literal translation meaning “the people”) ongoing process of change in Brazilian Portuguese (Zilles, 2005)



Language Variation in Brazilian Portuguese Tweets

- Pronominal variation in Brazilian Portuguese is affected by both linguistic and extra-linguistic factors (e.g., dialect or region)
- Research Question: Are spoken language choices reflected in social media language? ¹
- Seven metropolitan regions in Brazil
- 2,478 tokens for first person (N=1,679)
- 2,753 tokens for second person (N=1,999)

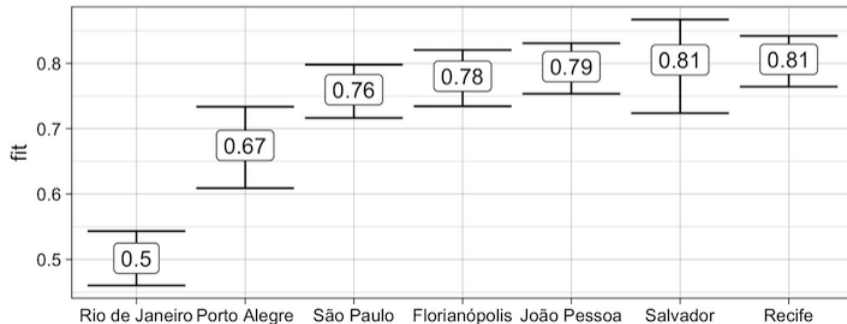


¹study in collaboration: GIRL <http://girlanguages.com>



Language Variation in Brazilian Portuguese Tweets

Probability estimates for `a gente` use (vs. `nós`)
across regions in Brazil



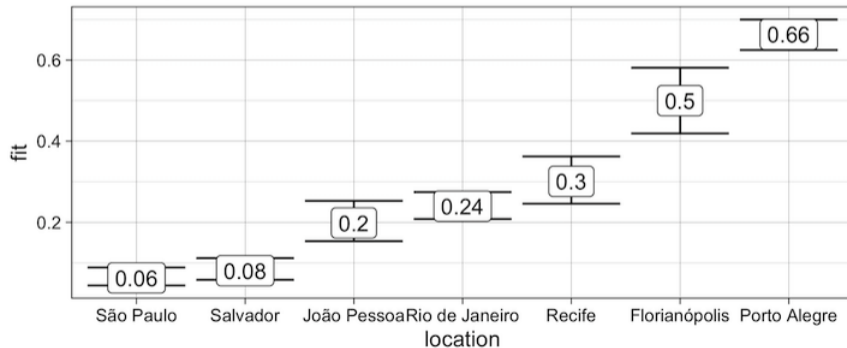
Language Variation in Brazilian Portuguese Tweets

- What is different about Rio de Janeiro?
 - A number of factors may explain the permanence of “nós” (we) in Rio de Janeiro, including the use of third singular agreement (e.g. “nós vai” – we goes) to resolve temporal ambiguities
 - 70% of “nós” (we) tokens are combined with a third singular agreement verb form in Rio de Janeiro.
 - This combination of the standard first plural pronoun + a third singular verb form might be:
 - an index of local identity
 - a result of the less monitored discourse found in Twitter in comparison to other written data (McCulloch, 2019)
 - a mockery strategy



Language Variation in Brazilian Portuguese Tweets

Probability estimates for `tu` use (vs. `você`)
across regions in Brazil



Language Variation in Brazilian Portuguese Tweets

- These results match previous literature, with slight differences in probabilities

	Our Results	Scherre et al. 2021; 2019
Porto Alegre	.66	.91
Florianópolis	.50	.76
Rio de Janeiro	.24	.27
Recife	.30	.14
João Pessoa	.20	.04
Salvador	.08	.01
São Paulo	.06	.00



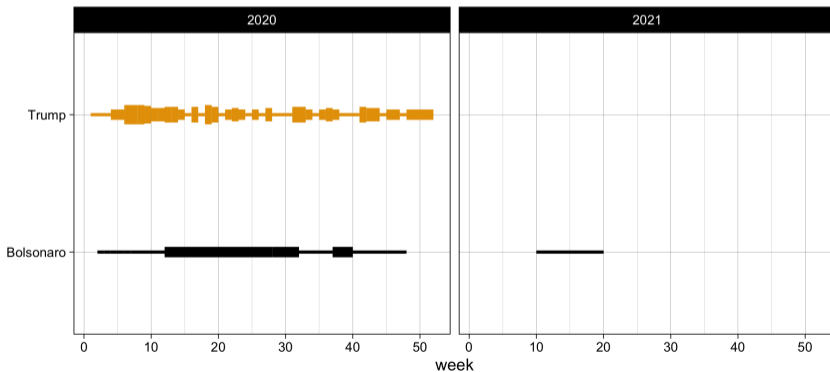
Cross-linguistic misinformation

- Fake news spread in social media is a growing concern across the globe (Pereira Galhardi et al., 2020)
- Similarities across different languages?
- Can we see a pattern of misinformation being “spread” from one language to another?
- Tweets from Trump and Bolsonaro (all Tweets)
- Baseline = tweets before 2020
- Methods: exploratory visualization, key word analysis (log-likelihood)



Language Variation in Brazilian Portuguese Tweets

Number of tweets containing the term chloroquine/cloroquina across weeks



Cross-linguistic misinformation

Trump's tweets on July 6 2020

- “Treatment with hydroxychloroquine cut the death rate significantly in sick patients hospitalized with COVID-19 – and without heart-related side-effects, according to a new study published by Henry Ford Health System. In a large-scale retrospective analysis... of 2,541 patients hospitalized between March 10 and May 2, 2020 across the system's six hospitals, the study found 13% of those treated with hydroxychloroquine alone died compared to 26.4% not treated with hydroxychloroquine.” @HenryFordNews

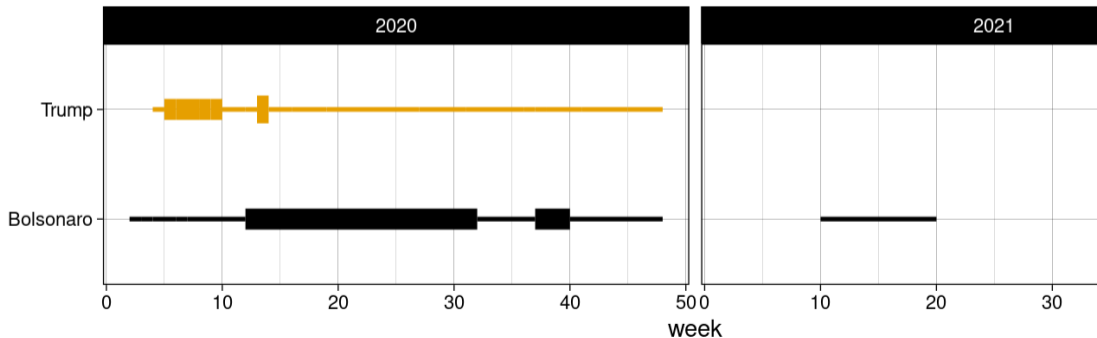
Bolsonaro's tweet on July 18 2020

- @FoxNews mostra estudos sobre a eficácia da Hidroxicloroquina no combate ao Coronavírus. <https://t.co/C0qbQmqDez>



Language Variation in Brazilian Portuguese Tweets

Number of tweets containing thanks and obrigado across weeks



Cross-linguistic misinformation

Trump's tweet on February 7 2020

- Very true, thank you Ted & John. I love Texas! <https://t.co/NoWTheat7Z>

Bolsonaro's tweet on October 3 2020

- Muito obrigado meu AGRESTE. Brasil acima de tudo, Deus acima de todos!. No YouTube: <https://t.co/brpzLpltNZ> <https://t.co/ljPS1sJ8dN>



Cross-linguistic misinformation

- Key word analysis (log likelihood) used to reveal which words were statistically more frequent in 2020/2021 using tweets from prior years as a reference corpus
- Other examples of term pairs on an interactive interface: <https://bit.ly/osu-demo>
- Next steps:
 - Linear regression for effect of previous tweets by Trump on tweets by Bolsonaro
 - Multimedia analysis (using Google's APIs)

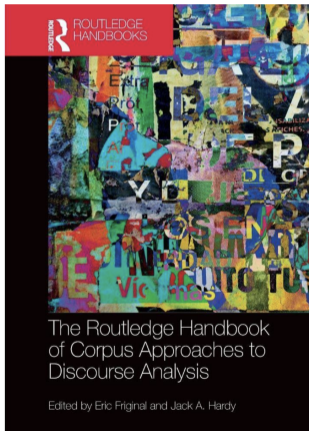


Multi-Dimensional Analysis of Style

What are the major dimensions of style for Machado de Assis?

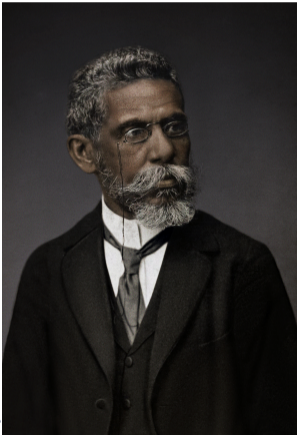
Joint work with Carlos Kauffmann

Reference



Kauffmann, C., & Berber Sardinha, T. (2021). Brazilian Portuguese literary style. In E. Friginal & J. Hardy (Eds.), *The Routledge Handbook of Corpus Approaches to Discourse Analysis* (pp. 354-375). Abingdon: Routledge.

Joaquim Maria Machado de Assis



- Born 1839 in Rio de Janeiro, died 1908
- Imperial (1889) and slavery Brazil (1888)
- Working class, washerwoman and wall painter
- Mixed race, white Portuguese mother and mixed race father
- Considered the greatest black writer in Western literature (Harold Bloom)
- Not seen as a black person in Brazil, usu. portrayed as white (Flynn, Calvo-González, & de Souza, 2013)
- Founder and president of the Academia Brasileira de Letras (ABL)
- Prolific writer, 840+ works, among novels, short stories, plays, crônicas, poems and sonnet volumes



Dimensions of Machado's style

- 1 Abstract style vs Oral involved style
- 2 Narrative style
- 3 Cogitative, hypothetical style
- 4 General background information vs Specific context-dependent reference
- 5 Thought presentation vs elaborated description



CLIMA Corpus: Corpus Literário de Machado de Assis

Register	Texts	Word tokens	Time period
Short stories	76	316,462	1870–1906
Novels	9	543,059	1872–1908
Total	85	859,531	36 years

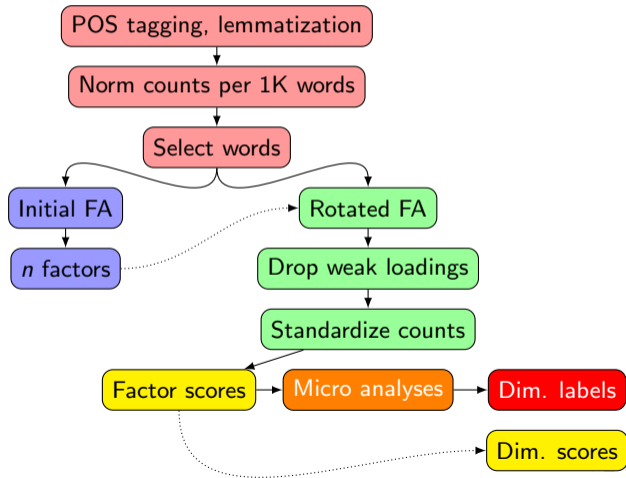
- **Short stories:** A Cartomante; A Causa Secreta; A Chinela Turca; A Desejada das Gentes; A Igreja do Diabo; A Mulher de Preto; A Parasita Azul; A Segunda Vida; A Senhora Galvão; A Sereníssima República; Adão e Eva; Anedota Pecuniária; Anedota do Cabriolet; As Academias de Sião; As Bodas de Luís Duarte; Aurora sem Dia; Cantiga de Esponsais; Capítulo dos Chapéus; Confissões de uma viúva moça; Conto Alexandrino; Conto de Escola; D. Benedita; D. Paula; Entre Santos; Ernesto de Tal; Eterno!; Evolução; Ex Cathedra; Frei Simão; Fulano; Galeria Póstuma; Ideias do Canário; Linha Reta e Linha Curva; Luís Soares; Lágrimas de Xerxes; Manuscrito de um Sacristão; Marcha Fúnebre; Maria Cora; Mariana; Miss Dollar; Missa do Galo; Na Arca; Noite de Almirante; O Alienista; O Anel de Polícrates; O Caso da Vara; O Cônego ou Metafísica do Estilo; O Dicionário; O Diplomático; O Empréstimo; O Enfermeiro; O Espelho; O Lapso; O Relógio de Ouro; O Segredo de Augusta; O Segredo do Bonzo; Pai Contra Mãe; Papéis Velhos; Ponto de Vista; Primas de Sapucaia!; Píldes e Orestes; Singular Ocorrência; Suje-se Gordo!; Teoria do Medalhão; Trio em Lá Menor; Um Apólogo; Um Capitão de Voluntários; Um Erradio; Um Homem Célebre; Uma Senhora; Uma Visita de Alcibíades; Um Férias; Uns Braços; Verba Testamentária; Viver!; Último Capítulo.
- **Novels:** A Mão e a Luva; Dom Casmurro; Esaú e Jacó; Helena; Iaiá Garcia; Memorial de Aires; Memórias Póstumas de Brás Cubas; Quincas Borba; Ressurreição.

Multi-dimensional analysis

- Introduced by Biber (1986).
- Seeks to describe the 'fundamental parameters of textual variation' (Biber, 1986, p. 410)
- '[a] linguistic dimension is determined on the basis of a consistent co-occurrence pattern among features. That is, when a group of features consistently co-occur in texts, those features define a linguistic dimension.' (Biber, 1988, p. 13)
- 'parameters [that] ... define continuums of variation rather than discrete poles' (Biber, 1988, p. 9)
- 'Each dimension comprises an independent group of co-occurring linguistic features, and each co-occurrence pattern can be interpreted in functional terms.' (Biber, 1988, p. 14)
- Major assumption: '**cooccurrence reflects shared function**'. (Biber, 1989, p. 7; my emphasis)



Procedures: MD Analysis



Functional dimension 1

- Abstract style
 - nominalizations (.71), abstract nouns (.7), past participles (.47), (nouns in subject position .35), (indefinite articles .38), (adjectives in attributive position .36).
 - Abstract discourse is signaled by a range of nominal features such as nominalizations, abstract nouns, and nouns in subject position, in addition to past participles and indefinite articles, all of which enable the expression of abstraction in different ways.
- Oral involved style
 - clause-linking coordinating conjunctions (-.73), additive coordinating conjunctions (-.69), discourse markers (-.58), (communication verbs -.36).
 - Oral involved discourse is built around clausal devices, such as clause-level conjunctions, discourse markers and communication verbs.



Functional dimension 1 (pos): Abstract style

Conto Alexandrino (1884), short story

Para conciliar os interesses da **ciência** com os impulsos da **piedade**, os réus não eram **escalpelados** à vista uns dos outros, mas sucessivamente. Quando vinham aos dois ou aos três, não ficavam em lugar donde os que esperavam pudessem ouvir os gritos do paciente, embora os gritos fossem muitas vezes **abafados** por meio de aparelhos; mas se eram **abafados**, não eram **suprimidos**, e em certos casos, o próprio objeto da **experiência exigia** que a **emissão** da voz fosse franca. Às vezes as **operações** eram simultâneas; mas então faziam-se em lugares **distanciados**.

Features

abstract nouns and nominalizations (ciência [science], piedade [piety], operações [operations], etc.), and past participles (escalpelados [scalped], abafados [muffled], etc.).



Functional dimension 1 (neg): Oral involved style

Adão e Eva (1896), short story

E falava a maligna, falava à toa, sem parar, contente e pródiga da língua; **mas** o diabo interrompeu-a:

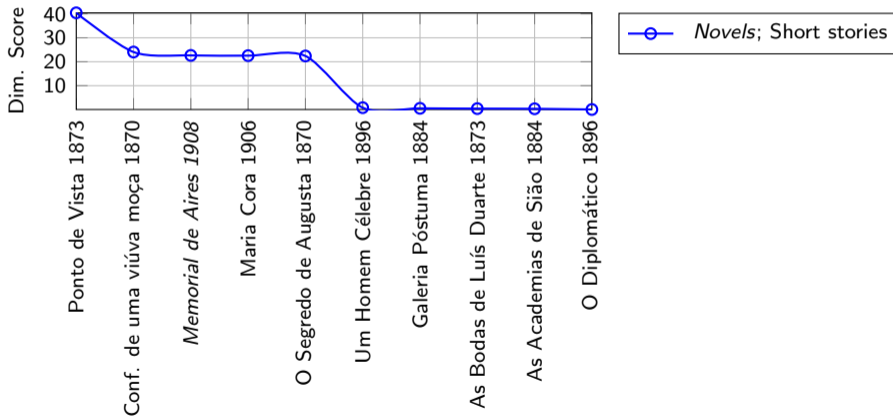
- Nada disso, **nem** ao ar, **nem** ao mar, **nem** à terra, **mas** tão-somente ao jardim de delícias, onde estão vivendo Adão e Eva. (...)
- **Deveras? Então** vou; farei tudo o que quiseres, meu senhor e pai. Anda, diz depressa o que queres que faça. Que morda o calcanhar de Eva? Morderei...

Features

conjunctions (e [and], mas [but], nem [neither]), and discourse markers (então [so], deveras [really]).



Dim.1: Abstract style vs Oral involved style



Dim.1: Abstract style vs Oral involved style

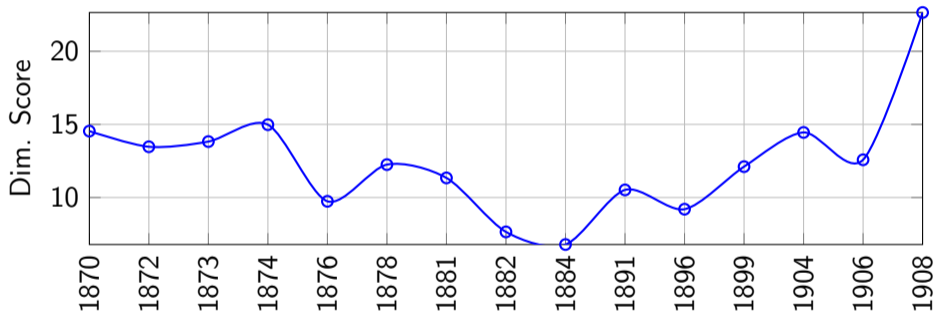


Figure: Mean dim. 1 scores (By year; $R^2=23.0\%$)



Dim.1: Abstract style vs Oral involved style

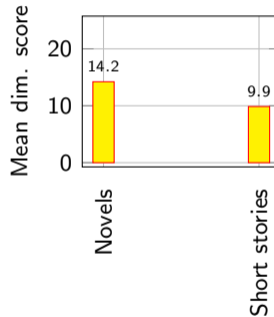


Figure: Mean dim. 1 scores (By register; $R^2=4.3\%$)



Functional dimension 2

- Narrative style
 - third-person verb forms (.92), third-person object pronouns (.7), rare object pronouns (.65), perfect preterit (.53), (imperfect preterit .37), (action verbs .31), (mental verbs .33).

Narrative style

Narrative discourse is put in play basically through third person verb forms, usually in the past tense, and verbs of action and cognition,



Functional dimension 2 (pos): Narrative style

Noite de Almirante (1884), short story

Deolindo Venta-Grande (era uma alcunha de bordo) **saiu** ao Arsenal de Marinha e **enfiou** pela Rua de Bragança. **Batiam** três horas da tarde. **Era** a fina flor dos marujos e, demais, **levava** um grande ar de felicidade nos olhos. A corveta **dele voltou** de uma longa viagem de instrução, e Deolindo **veio** à terra tão depressa **alcançou** licença. Os companheiros **disseram-lhe**, rindo (...).

Features

third-person verb forms in the perfect and imperfect preterit (era [was], saiu [left], voltou [returned]); action verbs in the preterit (saiu [left], veio [came], alcançou [reached]), third person object pronoun (lhe [him]).



Dim.2: Narrative style

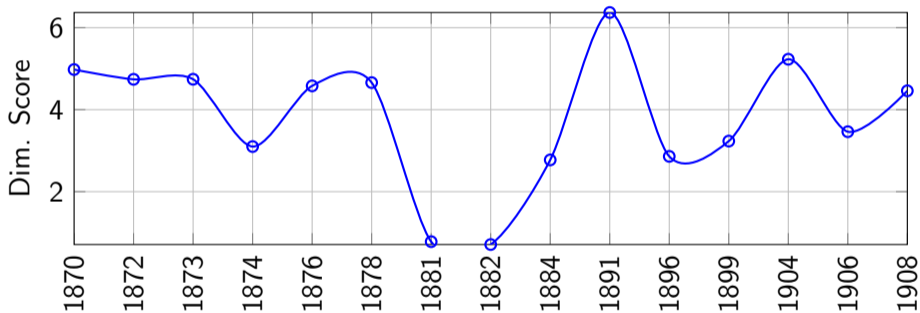


Figure: Mean dim. 2 scores (By year; $R^2=25.1\%$)



Functional dimension 2 (pos): Narrative style

Title	Register	Year	Score
Memórias Póstumas de Brás Cubas	Novel	1881	.78
Uma Visita de Alcibíades	Short story	1882	.11
Verba Testamentária	Short story	1882	-.28
O Espelho	Short story	1882	-.77
O Segredo do Bonzo	Short story	1882	-1.97
Teoria do Medalhão	Short story	1882	-3.64
A Sereníssima República	Short story	1882	-5.19



Functional dimension 3

- Cogitative, hypothetical discourse
 - 'ser' and 'estar' (to be) verbs (.85), existence verbs (.76), adjectives in predicative position (.58), communication verbs (.38), modal verbs (.37), demonstrative determiners and demonstrative pronouns (.35).

Cogitative, hypothetical discourse

Cogitative, hypothetical discourse, which expresses nuanced speech and thought associated with the narrator or the characters. The features enabling this discourse are essentially verb forms of the verb 'to be' (i.e. both 'ser' and 'estar' in Portuguese), in addition to forms of existence, modal and communication verbs. Adjectives in predicative position are also employed to complement the relational verbs and create this hypothetical, speculative effect.



Functional dimension 3 (pos): Cogitative, hypothetical discourse

O Segredo de Augusta (1870), short story

– Papai já acordou? **pergunta** Adelaide à sua mãe.

– Não, **responde esta** sem levantar os olhos do livro. (...)

As mãos compridas e bem feitas pareciam criadas para os afagos de amor. (...)

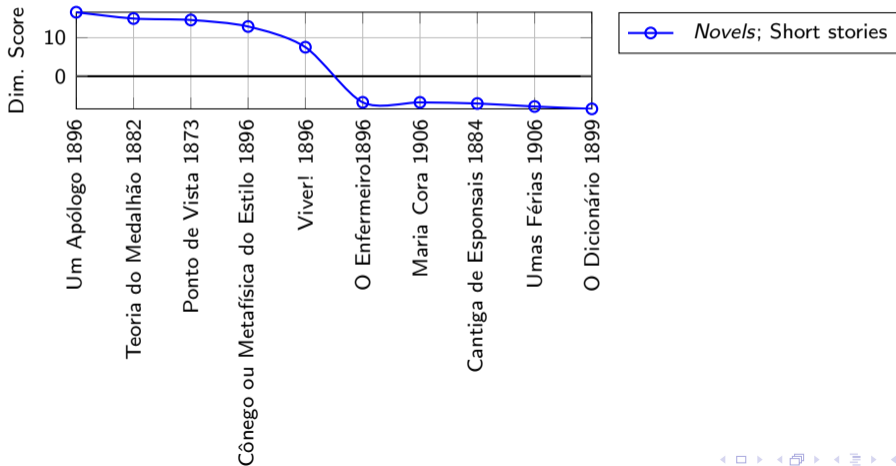
Todavia, **era** bem **capaz** de apaixonar um homem, sobretudo se ele **fosse** poeta, e **gostasse** das virgens de quinze anos, até porque **era** um pouco **pálida**, e os poetas em todos os tempos tiveram sempre queda para as criaturas **descoradas**.

Features

relational 'to be' (era [was], fosse [were, subjunctive], gostasse [liked, subjunctive]), verbs of communication (pergunta [asks], responde [replies]), modal (pareciam [seemed]), adjective in predicative position (pálida [pale], capaz [able], descoradas [colorless]), and demonstrative pronoun (esta [this]).



Dim.3: Cogitative, hypothetical discourse



Dim.3: Cogitative, hypothetical discourse

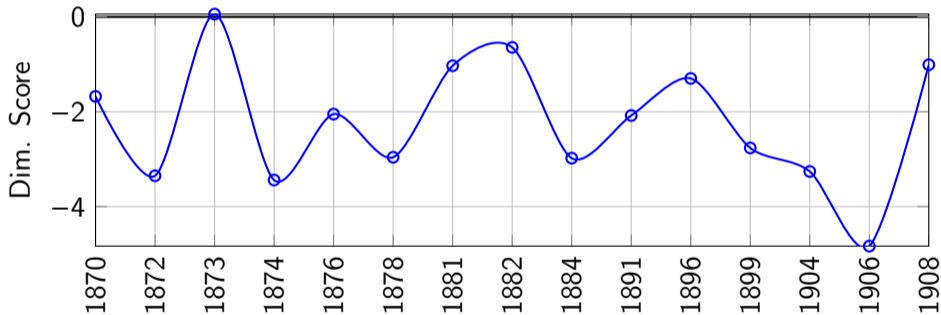


Figure: Mean dim. 3 scores (By year; $R^2=8.6\%$)



Functional dimension 4

- General background information discourse
 - adverbials (.64), imperfect preterit (.55), infinitive clauses preceded by preposition (.53), action verbs (.36), (adjectives in predicative position .35).
 - General background information is generally provided through the imperfect preterit tense, which enables the author to describe recurrent actions or states in the past that help the reader to understand the current narrative flow. Some of these actions and states appear in infinite clauses whose verbs are governed by prepositions. Adverbials in turn enable the sequence and/or emphasis of these descriptions.
- Specific context-dependent reference
 - possessive pronouns (-.55), nouns in subject position (-.44), (communication verbs -.31), (demonstrative determiners and demonstrative pronouns -.31), (discourse markers -.31).
 - Specific context-dependent reference is established largely through deictic features such as possessive pronouns and demonstratives. Nouns in subject position establish topical reference by introducing or referring back to individuals, concepts and objects, which often work as the reference for the possessive pronouns and demonstratives.



Functional dimension 4 (pos): Background or contextual information

Pai contra Mãe (1906), short story

Há meio século, os escravos **fugiam** com frequência. **Eram** muitos, e nem todos **gostavam** da escravidão. Sucedia ocasionalmente apanharem pancada, e nem todos **gostavam de apanhar** pancada. Grande parte era apenas repreendida; **havia** alguém de casa que **servia** de padrinho, e o mesmo dono não **era mau**; **além disso**, o sentimento da propriedade **moderava** a ação, porque dinheiro também dói. A fuga **repetia-se**, entretanto. Casos houve, ainda que raros, em que o escravo de contrabando, apenas **comprado** no Valongo, deitava a **correr**, **sem conhecer** as ruas da cidade. Dos que **seguiam para casa**, não raro, apenas ladinos, **pediam** ao senhor que lhes **marcasse** aluguel, e **iam ganhá-lo** fora, **quitandando**.

Features

imperfect (eram [were], fugiam [escaped], gostavam [liked], havia [there were], etc.), adverbial (além disso [besides]), infinitive clauses preceded by preposition (gostavam de apanhar [liked to get beaten up], sem conhecer [without knowing], etc.), action verbs (seguiam [went], correr [run], etc.).



Functional dimension 4 (neg): Specific context-dependent reference

Viver! (1896), short story

PROMETEU: Prometeu é o **meu** nome.

AHASVERUS: Tu Prometeu?

PROMETEU: E qual foi o **meu** crime? Fiz de lodo e água os primeiros homens, e depois, compadecido, roubei para eles o fogo do céu. Tal foi o **meu** crime. **Júpiter**, que então regia o Olimpo, condenou-me ao mais cruel suplício. Anda, sobe comigo a **este** rochedo.

Features

possessive pronouns (meu [my]), noun in subject position (Júpiter [Jupiter]), demonstrative determiner (este [this]).



Dim.4: Background info. vs Context-dependent reference

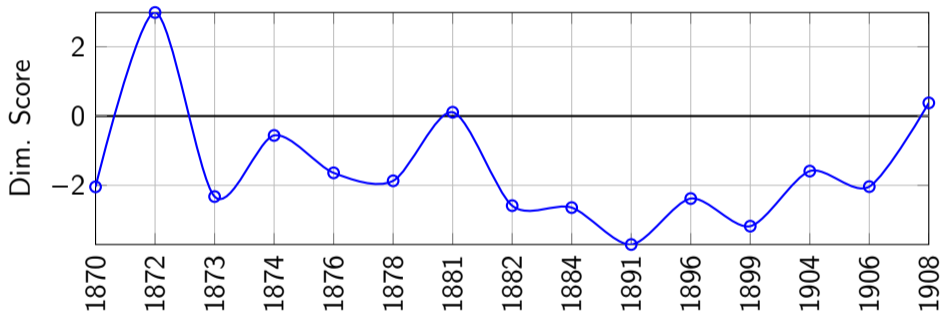


Figure: Mean dim. 4 scores (By year; $R^2=18.3\%$)



Functional dimension 5

- Thought presentation
 - mental verbs (.69), first-person verb forms (.66), subordinating conjunctions (.53), negation (.46), (possessive pronouns .37), (communication verbs .35), (imperfect preterit .35).
 - The characters' thoughts are often presented through the repeated use of mental and communication verbs, first person verb forms, subordinating conjunctions and negation.
- Elaborated description
 - adjectives in attributive position (-.52), indefinite articles (-.39), (definite articles -.33).
 - Description is elaborated through adjectives (in attributive position), and articles (definite and indefinite).



Functional dimension 5 (pos): Thought presentation

Dom Casmurro (1899), novel

O **meu** fim evidente **era** atar as duas pontas da vida, e restaurar na velhice a adolescência. Pois, senhor, **não** consegui recompor o que foi nem o que fui. Em tudo, **se** o rosto é igual, a fisionomia é diferente. **Se** só me **faltassem** os outros, vá; um homem consola-se mais ou menos das pessoas que perde; mais **falto** eu mesmo, e esta lacuna é tudo. (...) – **Sei que** você fez promessa... mas uma promessa assim... **não sei...** **Creio que**, bem pensado... Você que **acha**, prima Justina? – Eu? – Verdade é **que** cada um **sabe** melhor de si, **continuou** tio Cosme; Deus é que **sabe** de todos.

Features

mental verbs (acha [think], sabe [know], etc.), first-person verb forms (falto [am missing], sei [know], creio [believe]), subordinating conjunctions (se [if], que [that]), negation (não [not]), possessive pronoun (meu [meu]), communication verb (continuou [continued]), and the imperfect preterit (era [was]).



Functional dimension 5 (neg): Elaborated description

O Dicionário (1899), short story

Como era calvo desde **verdes** anos, decretou Bernardão que todos **os** seus súditos fossem igualmente calvos, ou por natureza ou por navalha, e fundou esse ato em **uma** razão de ordem **política**, a saber, que **a** unidade **moral** do Estado pedia **a** conformidade **exterior** **das** cabeças. Outro ato em que revelou **igual** sabedoria, foi o que ordenou que todos **os** sapatos **do** pé **esquerdo** tivessem **um pequeno** talho **no** lugar **correspondente** **ao** dedo mínimo, dando assim **aos** seus súditos **o** ensejo de se parecerem com ele, que padecia de **um** calo.

Features

adjectives in attributive position (verdes [green], política [political], moral [moral], etc.), indefinite articles (uma [a-feminine], um [a-masculine]), and definite articles (o, a, os, as [the]).



Dim.5: Thought presentation vs Elaborated Description

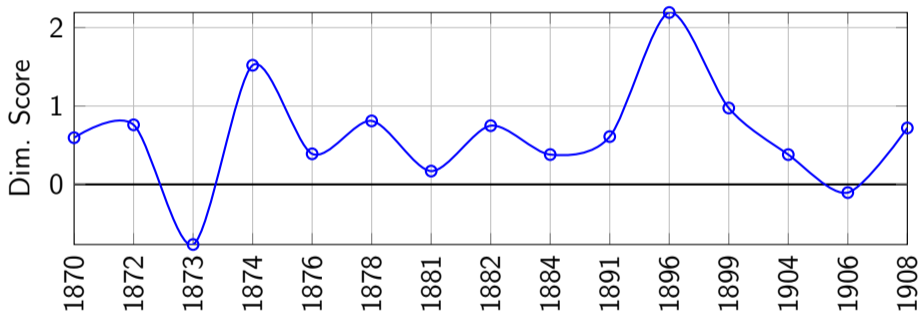


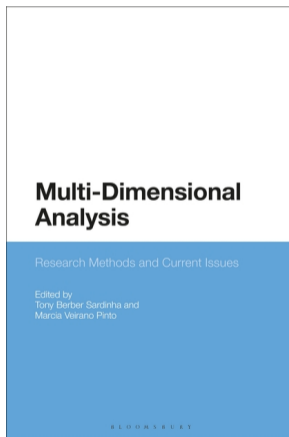
Figure: Mean dim. 5 scores (By year; $R^2=22.4\%$)



Transition

Representations of Brazil in Google Books

Multi-dimensional analysis: Research methods and current issues (2019)



Roland Barthes's Panzani Ad analysis



Italianicity (Barthes, 1964)



Discourse and representation



(Baker & McEnery, 2015, p. 5)

ways of looking at the world, of constructing objects and concepts in certain ways, of representing reality

(Burr, 1995, p. 48)

the set of meanings, [...] representations, [...] statements and so on that in some way together produce a particular version of events

Topic and discourse



(Hall, 1992, p. 92)

a group of statements which provide a language for talking about – i.e. a way of **representing** – a particular kind of knowledge about a topic. When statements about a topic are made within a particular discourse, **the discourse makes it possible to construct the topic in a certain way. It also limits the other ways in which the topic can be constructed.**
(my emphasis)



Michael Stubbs's study on language and culture

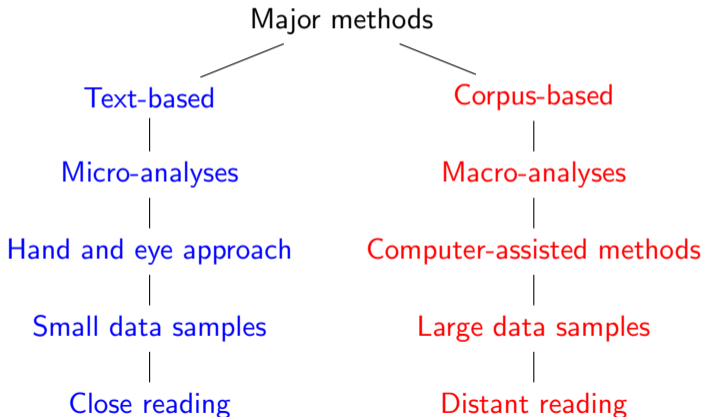


Stubbs (1996) Corpus of British news

- France: cycling, restaurants, rugby and wine
- Germany: banks, Daimler Benz, Hitler and Nazis
- Japan: emperor, exports, industrial, scandal, shame and trade
- British: army, government, empire
- English: language, history, heritage
- Scottish: devolution, highlands, salmon



Discourse analysis methods



Grammatical vs lexical MD analysis

	Grammatical	Lexical
Goal	Underlying parameters of functional variation	Underlying parameters of lexical variation
Method	Multivariate statistics	Multivariate statistics
Linguistic features	Mostly grammatical	Lexical
Primary interpretation	Function-based	Discourse-based



Google N-gram corpus for this investigation

Years	1800 – 2008
Time span	209 years
Tokens	466,666,156,981
Pages	2,429,890,866
Publications	4,520,002
Google Books Data	99.6%

(Berber Sardinha, 2019, 2020)



N-grams

N-gram: Character string (word) sequence

- 1-gram: Brazilian
- 2-gram: **Brazilian music**
- 3-gram: Brazilian music genres
- 4-gram: Brazilian music genres in
- 5-gram: Brazilian music genres in the

⋮

Bigram	Year	Frequency count	Publication count
American people	1825	177	77
American people	1826	284	72
American people	1827	143	56



Brazilian Dim. 1: The economy and politics



Loadings

economic (1.00), development (.99), industrial (.97), experience (.90), economy (.90), policy (.89), military (.89), political (.86), firms (.77), case (.77), export (.76), politics (.76), agriculture (.75), foreign (.74), firm (.73), authorities (.69), market (.69), exports (.68), industry (.67), public (.60), cult (.55), trade (.53), national (.51), slave (.49), government (.49), sugar (.44), society (.42), gold (.42), army (.41), independence (.37), capital (.35), nation (.34), iron (.33).

Brazilian Dim. 1: The economy and politics

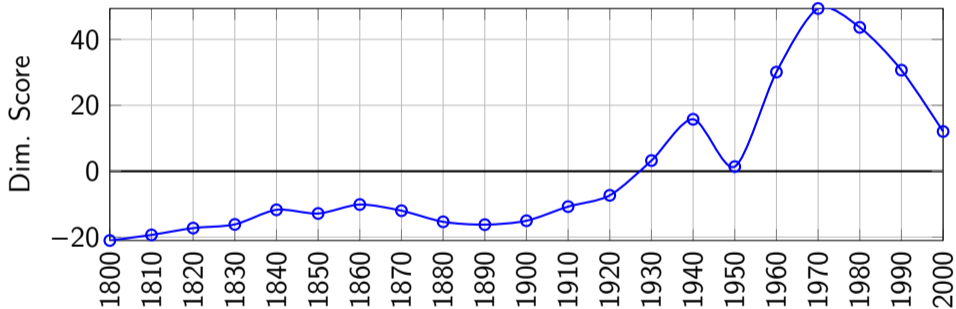


Figure: Mean dim. 1 scores (By year interval; $R^2=87.0\%$)



Brazilian Dim. 2: Traditional art, sciences, the people, and the land



Loadings

poet (.83), composer (.81), folk (.79), life (.74), writers (.74), family (.70), literature (.68), writer (.68), novel (.67), plant (.62), student (.60), population (.57), historian (.55), child (.55), author (.51), town (.48), social (.47), border (.47), people (.47), history (.46), culture (.42), woman (.40), law (.35), land (.33).

Brazilian Dim. 2: Traditional art, sciences, the people, and the land

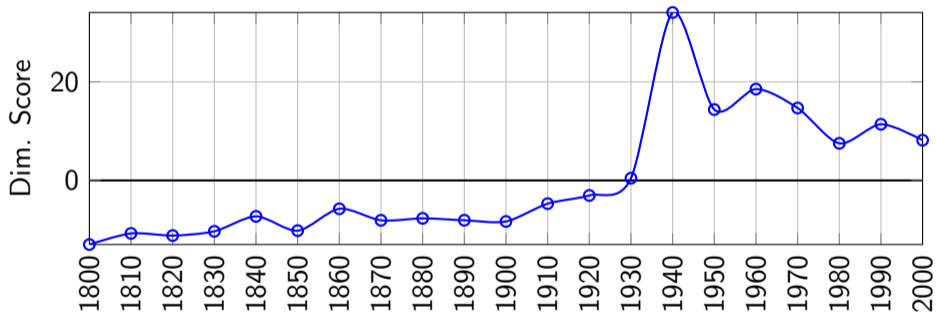


Figure: Mean dim. 2 scores (By year interval; $R^2=61.6\%$)



Brazilian Dim. 3: Raw materials and the landscape



Loadings

rubber (.77), highlands (.77), states (.76), port (.75), product (.72), ports (.68), highland (.66), coffee (.66), citizens (.65), plateau (.61), ship (.61), state (.56), ore (.56), frontier (.54), territory (.53), products (.51), citizen (.51), ships (.47), jungle (.47), production (.45), ores (.43), coast (.43), diamonds (.41), cotton (.41), soil (.39), waters (.37).

Brazilian Dim. 3: Raw materials and the landscape

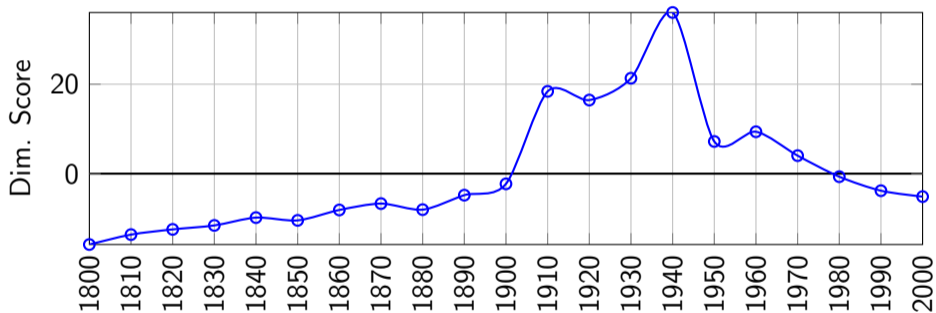


Figure: Mean dim. 3 scores (By year interval; $R^2=73.2\%$)



Brazilian Dim. 4: Art, women, men, religion, the environment



Loadings

rain (.92), religion (.92), popular (.79), women (.76), elite (.72), music (.61), cultural (.59), companies (.58), film (.53), art (.48), cinema (.46), cities (.41), men (.36), man (.30).



Brazilian Dim. 4: New artistic forms, women and men, religion, and the environment

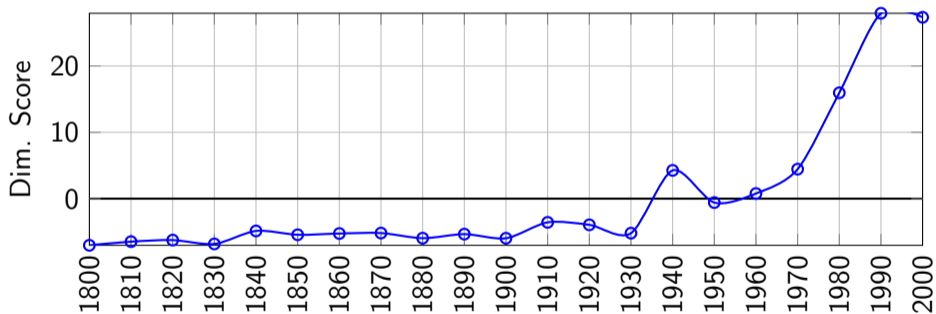


Figure: Mean dim. 4 scores (By year interval; $R^2=92.1\%$)



Brazilian Dim. 5: The monarchy, steam transportation, and the wilderness



Loadings

steamer (.75), province (.65), squadron (.64),
 flag (.62), fleet (.58), forests (.57), forest (.55),
 towns (.55), tea (.48), subjects (.46), merchant
 (.44), officer (.44), steam (.44), line (.43),
 mine (.41), minister (.39), vessels (.38), planter
 (.38), vessel (.36), species (.34), subject (.33),
 agent (.31), river (.31).

Brazilian Dim. 5: The monarchy, steam transportation, and the wilderness

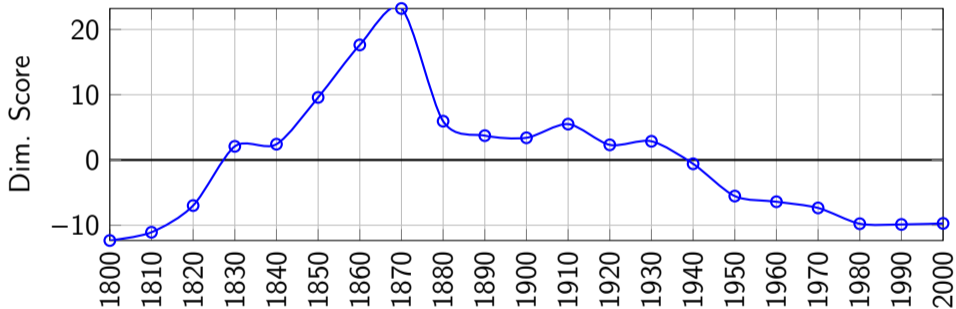
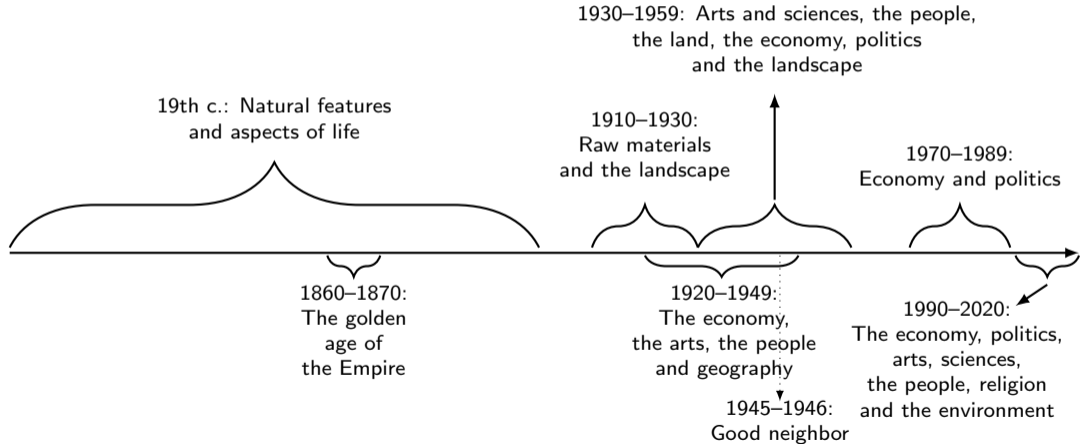


Figure: Mean dim. 5 scores (By year interval; $R^2=62.3\%$)



Timeline (Berber Sardinha, 2020)



Images, myths, and metaphors: Brazil

- 1 'The miracle nation': economic growth ('the China of the 1960s and 1970s')



Images, myths, and metaphors: Brazil

- 1 'The miracle nation': economic growth ('the China of the 1960s and 1970s')
- 2 'Good neighbor nation': a distant culture disseminated for the benefit of American interests.



Images, myths, and metaphors: Brazil

- 1 'The miracle nation': economic growth ('the China of the 1960s and 1970s')
- 2 'Good neighbor nation': a distant culture disseminated for the benefit of American interests.
- 3 'God's nation': Edenic motif, idyllic landscape, tropical paradise, 'land blessed by God', exoticism, grandeur, natural riches



Images, myths, and metaphors: Brazil

- ① 'The miracle nation': economic growth ('the China of the 1960s and 1970s')
- ② 'Good neighbor nation': a distant culture disseminated for the benefit of American interests.
- ③ 'God's nation': Edenic motif, idyllic landscape, tropical paradise, 'land blessed by God', exoticism, grandeur, natural riches
- ④ 'Artistic nation': high brow, popular culture, cultural landscape metaphor



Images, myths, and metaphors: Brazil

- 1 'The miracle nation': economic growth ('the China of the 1960s and 1970s')
- 2 'Good neighbor nation': a distant culture disseminated for the benefit of American interests.
- 3 'God's nation': Edenic motif, idyllic landscape, tropical paradise, 'land blessed by God', exoticism, grandeur, natural riches
- 4 'Artistic nation': high brow, popular culture, cultural landscape metaphor
- 5 'Imperial nation': Monarchical period; the sea and waterways as images of the country



Images, myths, and metaphors: United States

- 1 'The powerful nation': exceptionalism, prominence, singularity



Images, myths, and metaphors: United States

- 1 'The powerful nation': exceptionalism, prominence, singularity
- 2 'The rich nation': multi-faceted, complex



Images, myths, and metaphors: United States

- ① 'The powerful nation': exceptionalism, prominence, singularity
- ② 'The rich nation': multi-faceted, complex
- ③ 'The chosen nation': sacred destiny, the chosen people, American dream; 'nature's nation': natural wonders



Images, myths, and metaphors: United States

- ① 'The powerful nation': exceptionalism, prominence, singularity
- ② 'The rich nation': multi-faceted, complex
- ③ 'The chosen nation': sacred destiny, the chosen people, American dream; 'nature's nation': natural wonders
- ④ 'Freedom and expansion': natural right to occupy



Images, myths, and metaphors: United States

- 1 'The powerful nation': exceptionalism, prominence, singularity
- 2 'The rich nation': multi-faceted, complex
- 3 'The chosen nation': sacred destiny, the chosen people, American dream; 'nature's nation': natural wonders
- 4 'Freedom and expansion': natural right to occupy
- 5 'The innocent nation': innocence, the wonders of the new world, nature's nation, freedom from British rule



Comparison of broad representations

Notion	US	Brazil
Power	✓	
Freedom	✓	
Empire		✓
God chosen	✓	✓
Miracle		✓
Innocence	✓	✓
Naturally beautiful	✓	✓
Artistic richness	✓	✓
Neighbor		✓



References I

- Baker, P., & McEnery, T. (2015). Introduction. In P. Baker & T. McEnery (Eds.), *Corpora and discourse studies: Integrating discourse and corpora* (p. 1-20). Palgrave Macmillan: Basingstoke.
- Barthes, R. (1964). Rhetoric of the image. In R. Barthes (Ed.), *Image, music, text* (p. 32-61). Fontana: London.
- Berber Sardinha, T. (2019). Using multidimensional analysis to detect representations of national identity. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis* (p. 231-258). London: Routledge.
- Berber Sardinha, T. (2020). A historical characterization of American and Brazilian cultures based on lexical representations. *Corpora*, 15(2), 183-212.
- Biber, D. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language*, 62(2), 384-414.



References II

- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. doi: 10.1017/cbo9780511621024
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3-43. doi: 10.1515/ling.1989.27.1.3
- Burr, V. (1995). *An introduction to social constructionism*. Routledge: London.
- Flynn, A., Calvo-González, E., & de Souza, M. M. (2013). Whiter shades of pale: “coloring in” machado de assis and race in contemporary brazil. *Latin American Research Review*, 48, 3–24. doi: 10.1353/lar.2013.0046
- Galina Russell, I. (2011). ¿Que són las humanidades digitales? *Revista Digital Universitaria*, 12(7), 3-8.
- Hall, S. (1992). *Formations of modernity*. Polity Press: Cambridge.



References III

Stubbs, M. (1996). *Text and corpus analysis – computer-assisted studies of language and culture*. Oxford: Blackwell.

