

# APÊNDICE - Apêndice Metodológico

*Felipe Rocha*

## 1. Sobre o Documento

Conforme discutido no projeto, a pesquisa empírica dependerá da avaliação do dicionário-modelo criado por Vilela e Neiva (2011) e de sua adaptação para o Corpus da tese. O intuito é garantir que o esquema de codificação seja representativo da realidade textual analisada. Para tanto, uma série de medidas vêm sendo tomadas, as quais o presente documento informa. Desse modo, o objetivo desse apêndice é mostrar as conclusões já obtidas, discutir os possíveis efeitos que elas terão na montagem do dicionário final e assegurar níveis de replicação e de transparência em mínimos detalhes.

De acordo com o que foi previsto no cronograma da tese, a primeira fase da pesquisa empírica ainda não teve seu prazo encerrado e está em pleno desenvolvimento. Consequentemente, é válido registrar que esse documento não apresenta resultados conclusivos, apenas achados cuja natureza é exploratória e tentativa. *Insights* a serem compartilhados com os examinadores da banca.

Com fins de facilitar a visualização do exato local onde atualmente estamos na pesquisa empírica, a figura 1 apresenta em cor verde as etapas já concluídas, em amarelo, as que estão sendo realizadas e, em vermelho, as que ainda serão iniciadas:

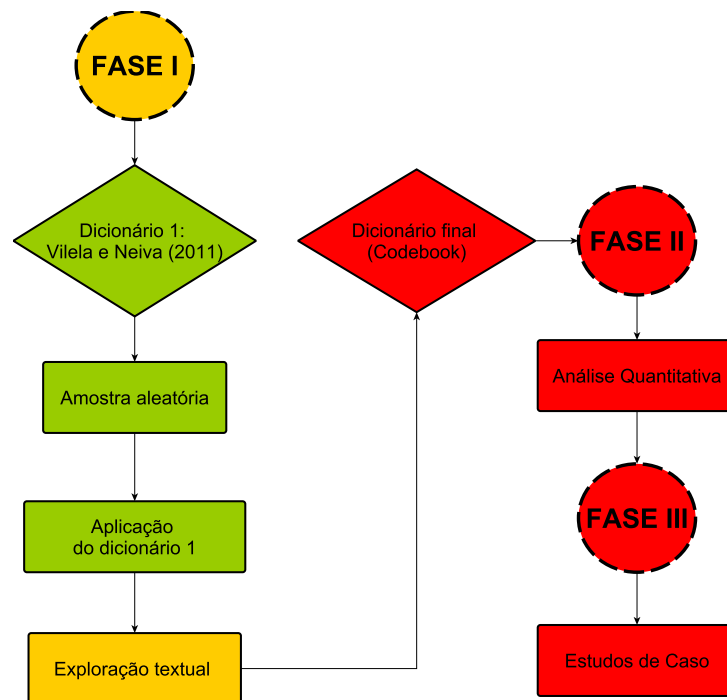


Figure 1: Fases Metodológicas

Através da figura é possível perceber a existência de etapas concluídas, em desenvolvimento e ainda não iniciadas. Por sua vez, esse documento apresenta a montagem do dicionário inicial, a construção da amostra aleatória, a aplicação do dicionário e alguns achados provenientes da exploração textual. Antes de discutir cada uma desses pontos, é necessário fazer duas ressalvas:

1. Evitamos inserir códigos computacionais em abundância, já que isso poderia servir de obstáculo para a compreensão das partes mais importantes do documento. Os códigos que, por ventura, aparecerão são

essenciais e estarão acompanhados de comentários explicativos cuja finalidade é didática. Eles serão representados, no início da frase, pelo símbolo “#”. Os interessados em uma versão completa dos códigos, favor, entrar em contato com: rocha.felipeferreira@gmail.com

2. Os achados incluídos possuem natureza exploratória. Por essa razão, não teceremos comentários específicos sobre o que cada gráfico e/ou tabela significa, já que isso pouco contribui com o objetivo primário do documento. Na última seção desse apêndice, entretanto, aqueles achados cujo impacto no esquema final de codificação da tese foi relevante serão discutidos de forma mais detalhada.

## 2. Pacotes Utilizados

Para realizar as análises foram necessários alguns pacotes adicionais aos já existentes no R e no ambiente do RStudio. Como essas são informações importantes para a replicação dos resultados, seguem, detalhadamente, os pacotes adicionais utilizados:

*# O comando "library" foi usado para carregar cada um dos pacotes.*

```
library(quanteda)
library(readtext)
library(tidyverse)
library(lubridate)
library(ggthemes)
library(topicmodels)
```

## 3. Corpus e suas Variáveis

Na parte metodológica do projeto de tese já foram apresentadas as principais informações descritivas do Corpus utilizado para a análise. Portanto, o que será feito agora é discutir melhor sobre cada uma das suas variáveis e níveis de mensuração.

Conforme já mencionado, há um total de 1700 discursos pronunciados pelos Presidentes e Chanceleres brasileiros entre 1995 e 2016. O banco de dados é formado por um amplo conjunto de variáveis. Para facilitar a visualização das mesmas, as discussões serão feitas em blocos e com base em 10 linhas, selecionadas aleatoriamente:

*# O comando abaixo foi utilizado para selecionar 10 linhas  
# aleatórias do Corpus e, então, plotá-los em forma de tabela*

```
set.seed(123)
variaveis_corpus <- sample_n(tese_corpus_dfm, 10)
knitr::kable(variaveis_corpus[, 1:3], row.names = FALSE,
              caption = "Primeiro bloco de variáveis do Corpus")
```

Table 1: Primeiro bloco de variáveis do Corpus

Text	Types	Tokens
MRE_C.Lafer_14.08.2002_pt_sao.paulo_brasil_america_jornal.gazeta.mercantil.txt	457	924
PRES_LULA_13.09.2006_pt_brasilia_brasil_america_ibas.txt	391	1029
MRE_L.F.Lampreia_22.09.1995_pt_washington_eua_america_conselho.empresarial.br.eua.txt	712	1788
PRES_LULA_20.04.2008_pt_acra_gana_africa_embrapa.txt	766	2203
PRES_LULA_25.09.2007_pt_nova.york_eua_america_onu.txt	801	1893
MRE_A.Patriota_29.01.2012_esp_NA_chile_america_jornal.elmercurio.txt	822	2545

Text	Types	Tokens
PRES_FHC_09.03.2001_pt_brasilia_brasil_america_orgem.gra.cruz.txt	283	628
PRES_LULA_20.11.2009_pt_salvador_brasil_america_palestina.txt	511	1068
PRES_FHC_13.07.1998_pt_brasilia_brasil_america_kofi.anan.txt	466	1296
MRE_M.Vieira_18.05.2016_pt_brasilia_brasil_america_transmissao.cargo.txt	671	1566

As primeiras três colunas apresentadas mostram as seguintes informações:

- **Text:** Trata-se do nome que descreve cada um dos textos em suas especificidades;
- **Types:** É a quantidade de termos únicos dos discursos (não se contabiliza a repetição). Por exemplo, na frase: “Viu uma lua no céu e viu outra lua no mar”, existem 8 types;
- **Tokens:** Quantidade de termos nos discursos (contabiliza-se a repetição). Por exemplo, na frase: “Viu uma lua no céu e viu outra lua no mar”, existem 11 tokens;

```
knitr::kable(variaveis_corpus[, 4:12], row.names = FALSE,
              caption = "Segundo bloco de variáveis do Corpus")
```

Table 2: Segundo bloco de variáveis do Corpus

Sentences	cargo	emissor	data	idioma	cidade	pais	continente	sobre
32	MRE	Lafer	2002-08-14	pt	sao.paulo	brasil	america	jornal.gazeta.mercantil
26	PRES	Lula	2006-09-13	pt	brasilia	brasil	america	ibas
43	MRE	Lampreia	1995-09-22	pt	washington	eua	america	conselho.empresarial.br.eua
86	PRES	Lula	2008-04-20	pt	acra	gana	africa	embrapa
98	PRES	Lula	2007-09-25	pt	nova.york	eua	america	onu
83	MRE	Patriota	2012-01-29	esp	NA	chile	america	jornal.elmercurio
21	PRES	FHC	2001-03-09	pt	brasilia	brasil	america	orgem.gra.cruz
48	PRES	Lula	2009-11-20	pt	salvador	brasil	america	palestina
36	PRES	FHC	1998-07-13	pt	brasilia	brasil	america	kofi.anan
53	MRE	Vieira	2016-05-18	pt	brasilia	brasil	america	transmissao.cargo

- **Sentences:** Quantidade de frases/sentenças em cada discurso;
- **Cargo:** Se Chanceler ou Presidente da República;
- **Emissor:** Nome do emissor do discurso;
- **Data:** Data de pronunciamento do discurso (em formato ANO/MÊS/DIA);
- **Idioma:** Idioma em que o pronunciamento está escrito;
- **Cidade:** Cidade de pronunciamento do discurso. Em casos omissos: NA;
- **País:** País de pronunciamento do discurso;
- **Continente:** Continente do pronunciamento;
- **Sobre:** Informações para rápida identificação da natureza do discurso;

```
knitr::kable(variaveis_corpus[, 13:19], row.names = FALSE,
              caption = "Terceiro bloco de variáveis do Corpus")
```

Table 3: Terceiro bloco de variáveis do Corpus

ano	lugar_disc	pares	mandato_presid	partido_pres	diplo_carreira	grau_diplo_pres
2002	doméstico	FHC e Lafer	FHC2	Não se aplica	0	Não se aplica
2006	doméstico	Lula e Amorim	Lula1	PT	Não se aplica	Alto
1995	internacional	FHC e Lampreia	FHC1	Não se aplica	1	Não se aplica
2008	internacional	Lula e Amorim	Lula2	PT	Não se aplica	Alto
2007	internacional	Lula e Amorim	Lula2	PT	Não se aplica	Alto
2012	internacional	Dilma e Patriota	Dilma1	Não se aplica	1	Não se aplica

ano	lugar_disc	pares	mandato_presid	partido_pres	diplo_carreira	grau_diplo_pres
2001	doméstico	FHC e Lafer	FHC2	PSDB	Não se aplica	Moderado
2009	doméstico	Lula e Amorim	Lula2	PT	Não se aplica	Alto
1998	doméstico	FHC e Lampreia	FHC1	PSDB	Não se aplica	Moderado
2016	doméstico	Temer e Serra	Dilma2	Não se aplica	1	Não se aplica

- **Ano:** Ano de pronunciamento do discurso;
- **Lugar\_disc:** Se o discurso foi enunciado em cenário internacional ou doméstico;
- **Pares:** Discursos pertencentes a respectivos pares de Presidente e Chanceler;
- **Mandato\_presid:** Informa em qual mandato presidencial o discurso aconteceu;
- **Partido\_pres:** Partido do Presidente;
- **Diplo\_carreira:** Se o emissor do discurso é diplomata de carreira ou não (em casos de Presidentes, será marcado que não se aplica);
- **Grau\_diplo\_pres:** Informa o grau de diplomacia presidencial. Essa é uma medida qualitativa, baseada no que aponta a literatura. Não é, portanto, uma estimativa numérica precisa. Nos casos dos Chanceleres, será marcado que não se aplica.

Em suma, essas foram as principais informações a respeito do Corpus, de sua estrutura e variáveis. Agora, é possível montar o dicionário inicial, conforme pensado por Vilela e Neiva (2011)

## 4. Montagem do Dicionário (VILELA e NEIVA (2011))

Conforme prevista na parte metodológica do projeto da tese, a montagem e operacionalização das categorias de análise começaram através de um dicionário pré-existente, já que isso aumenta a validade e confiabilidade da alocação temática. Assim, a primeira fase tem início com a incorporação do seguinte dicionário:

```
# Categoria: Instituições Internacionais
dicionario_fase1$instituições_internacionais
```

```
## [1] "união europ*ia"
## [2] "ue"
## [3] "comunidade dos países de língua portuguesa"
## [4] "cplp"
## [5] "organização das nações unidas"
## [6] "onu"
## [7] "banco mundial"
## [8] "organização dos estados americanos"
## [9] "oea"
## [10] "g7"
## [11] "g8"
## [12] "organização internacional do trabalho"
## [13] "oit"
## [14] "área de livre comércio das américas"
## [15] "alca"
## [16] "mercado comum do sul"
## [17] "mercosul"
## [18] "organização das nações unidas para a educação e cultura"
## [19] "unesco"
## [20] "fundo monetário internacional"
## [21] "fmi"
## [22] "fundo das nações unidas para a infância"
## [23] "unicef"
## [24] "organização mundial do comércio"
```

```
## [25] "omc"
## [26] "organização mundial da saúde"
## [27] "oms"
## [28] "instituições financeiras multilaterais"
## [29] "comunidade ibero-americana"
## [30] "governança global"
## [31] "banco interamericano de desenvolvimento"
## [32] "bid"
## [33] "organização das nações unidas para agricultura e alimentação"
## [34] "fao"
## [35] "programa das nações unidas para o desenvolvimento"
## [36] "pnud"
## [37] "associação latino-americana de integração"
## [38] "aladi"
```

```
# Categoria: Economia
dicionario_fase1$economia
```

```
## [1] "economia" "inflação"
## [3] "estabilidade financeira" "instabilidade financeira"
## [5] "financeira" "comércio"
## [7] "desenvolvimento econômico" "subsídio"
## [9] "protecionismo" "barreira não tarifária"
## [11] "tarifa" "investir"
## [13] "indústria" "produto interno bruto"
## [15] "pib" "agricultura"
## [17] "riqueza natural" "produção"
## [19] "produtivo" "força de trabalho"
## [21] "geração de renda" "geração de emprego"
## [23] "geração de postos de trabalho" "crise financeira"
## [25] "crise econômica" "credor"
## [27] "devedor" "privatizar"
## [29] "empresa" "exploração econômica"
## [31] "exploração capitalista" "monopólio"
## [33] "barreira comercial" "banco"
## [35] "super*vit" "déficit"
## [37] "orçamento" "exportação"
## [39] "importação" "agroindústria"
## [41] "agronegócio" "agropecuária"
## [43] "bndes" "mercado"
```

```
# Categoria: Desigualdade Social
dicionario_fase1$desigualdade_social
```

```
## [1] "desigualdade"
## [2] "igualdade"
## [3] "combate à fome"
## [4] "exclusão social"
## [5] "inclusão social"
## [6] "justiça social"
## [7] "injustiça social"
## [8] "pobre"
## [9] "rico"
## [10] "distribuição de renda"
## [11] "proteção social"
## [12] "indigência"
```

```
## [13] "indigente"
## [14] "miséria"
## [15] "favela"
## [16] "mst"
## [17] "segregação"
## [18] "programa social"
## [19] "bolsa família"
## [20] "bolsa escola"
## [21] "bolsa renda"
## [22] "bolsa alimentação"
## [23] "vulneravel"
## [24] "menos favorecidos"
## [25] "universalização da educação"
## [26] "universalização dos serviços públicos"
## [27] "contra a fome"
```

```
# Categoria: Meio Ambiente
dicionario_fase1$meio_ambiente
```

```
## [1] "proteção ambiental"      "questão ambiental"
## [3] "conservação ambiental"   "ambiental"
## [5] "aquecimento global"      "desaquecimento do planeta"
## [7] "quioto"                   "kyoto"
## [9] "meio ambiente"           "emissão de gás"
## [11] "efeito estufa"           "combustível renovável"
## [13] "biocombustível"          "energia renovável"
## [15] "energia limpa"           "desmatar"
## [17] "desmatamento"           "mudança de clima"
## [19] "poluição"                "carbono"
## [21] "co2"                      "ecologia"
## [23] "eco 92"                  "eco92"
## [25] "eco_92"                  "recurso hídrico"
## [27] "agenda 21"               "rio92"
## [29] "rio 92"                  "ecossistema"
## [31] "sustentável"             "poluente"
## [33] "amazônia"                "floresta"
## [35] "queimada"
```

```
# Categoria: Paz e Segurança Internacionais
dicionario_fase1$paz_seguranca_internacional
```

```
## [1] "paz"                      "narcotráfico"
## [3] "tráfico de drogas"       "conflitos de fronteira"
## [5] "conflitos fronteiriços"  "fronteira"
## [7] "terror militar"          "guerra"
## [9] "defesa internacional"    "defesa do território"
## [11] "conflito"                "arma"
## [13] "armas"                   "armada"
## [15] "desarmar"                "teste nuclear"
## [17] "guerra nuclear"          "bomba nuclear"
## [19] "segurança nacional"      "segurança internacional"
## [21] "conselho de segurança"
```

```
# Categoria: Direitos Humanos e Democracia
dicionario_fase1$direitos_humanos_democracia
```

```
## [1] "direitos"                "democracia"
```

```
## [3] "xenofobia"                "discriminação"
## [5] "racismo"                  "justiça"
## [7] "injustiças"               "minorias étnicas e sexual"
## [9] "minorias étnicas raciais" "emancipação do país"
## [11] "emancipação política"     "cidadania"
## [13] "exploração sexual"        "exploração do trabalho infantil"

# Categoria: Cooperação Internacional
dicionario_fase1$cooperacao_internacional

## [1] "cooperação" "parceria" "intercâmbio" "embrapa"
```

## 5. Amostra Aleatória

Tendo estruturado todo o dicionário inicial, o próximo passo foi criar uma amostra aleatória dos 1700 pronunciamentos. Nessa fase, como o dicionário de Vilela e Neiva está em português, serão selecionados apenas os discursos nesse idioma:

```
# Esse primeiro comando foi utilizado para selecionar apenas os discursos em português.

primeira_amostra_aleatoria <- corpus_subset(tese_corpus, idioma == "pt")

# Agora, construímos uma amostra aleatória de 314 discursos.

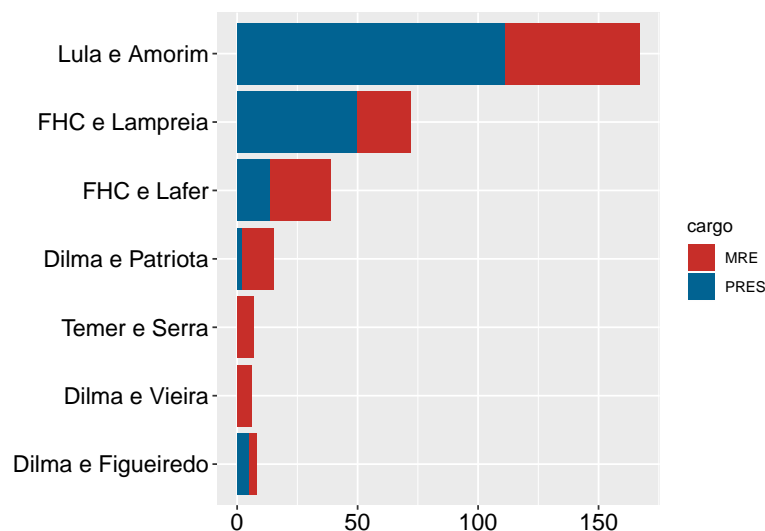
set.seed(1234)
primeira_amostra_aleatoria <- corpus_sample(primeira_amostra_aleatoria, 314)
```

A seleção de uma amostra de 314 discursos possui razões estatísticas. Ao se adotar um *confidence level* de 95% e um *confidence interval* de 5 na população total de discursos, a calculadora online de tamanho amostral aponta tal número de pronunciamentos como ideal.

A partir da obtenção dessa amostra aleatória, é necessário saber as suas características descritivas. Para isso, em primeiro lugar, segue a quantidade de discursos (n) para os Presidentes (PRES) e os Ministros de Relações Exteriores (MRE).

cargo	n
MRE	132
PRES	182

Pelo que se percebe, a quantidade de discursos alocada para cada um dos cargos não supervaloriza e nem subvaloriza nenhum dos atores em análise. Como há mais pronunciamentos presidenciais, naturalmente, a amostra coletou uma maior quantidade dos mesmos. De qualquer modo, tudo aleatoriamente. Outra informação a se observar é a quantidade de discursos por emissor e por par Presidente-Chanceler. Para ambas as informações, segue o gráfico abaixo.



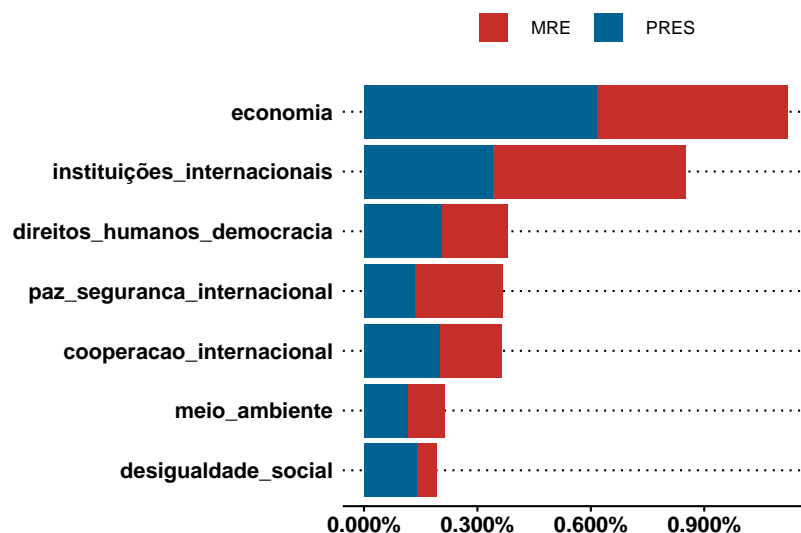
Conforme esperado, a amostra aleatória representa bem a proporção original de discursos totais de cada um dos emissores. Assim, Lula e Celso Amorim são os que mais possuem dados coletados e Temer e Serra são os que menos possuem pronunciamentos na amostra aleatória.

Com isso, é possível seguir em frente e aplicar o dicionário. Para evitar vieses, os resultados de frequência serão apresentados em escala relativa.

## 6. Aplicação Quantitativa do Dicionário

Até aqui, já passamos pela fase de construção do Dicionário inicial (Vilela e Neiva (2011)), já selecionamos a amostra aleatória a ser explorada e, agora, veremos como o dicionário se comporta ao ser aplicado. Essa, por sua vez, é a última fase já concluída da pesquisa. O que será visto é uma simulação em menor escala do que será realizado na tese, tão logo o dicionário final tenha sido fixado e aplicado. Portanto, os achados são temporários e apenas exploratórios.

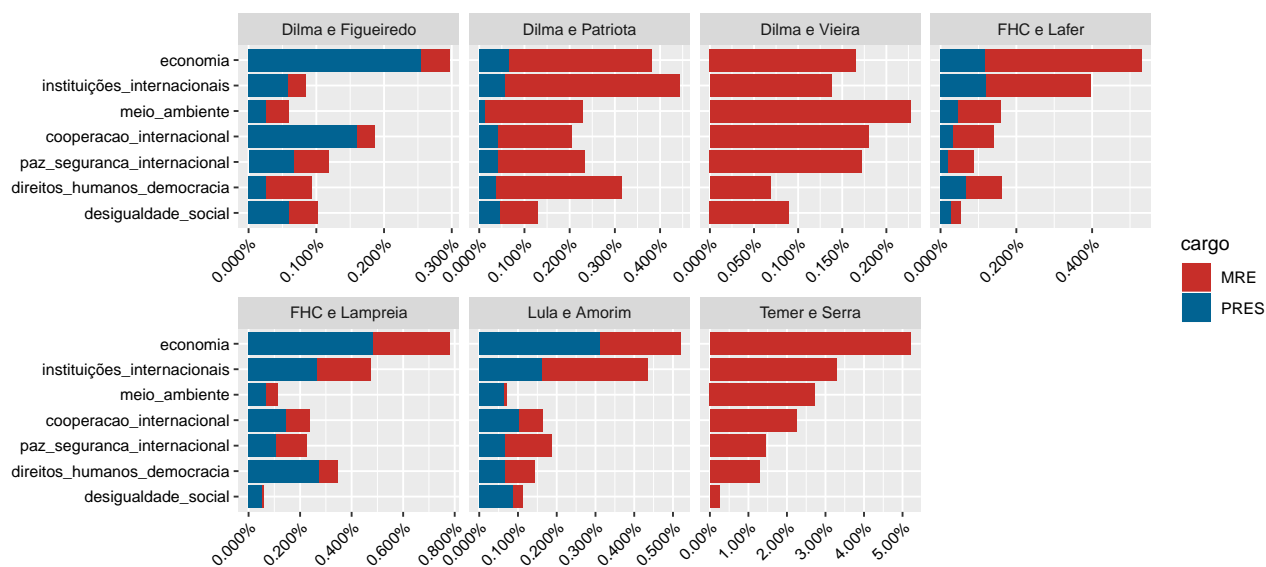
Em primeiro lugar, importa saber quais foram os temas mais e menos salientados por Presidentes e Chanceleres, conforme mostra o gráfico abaixo:



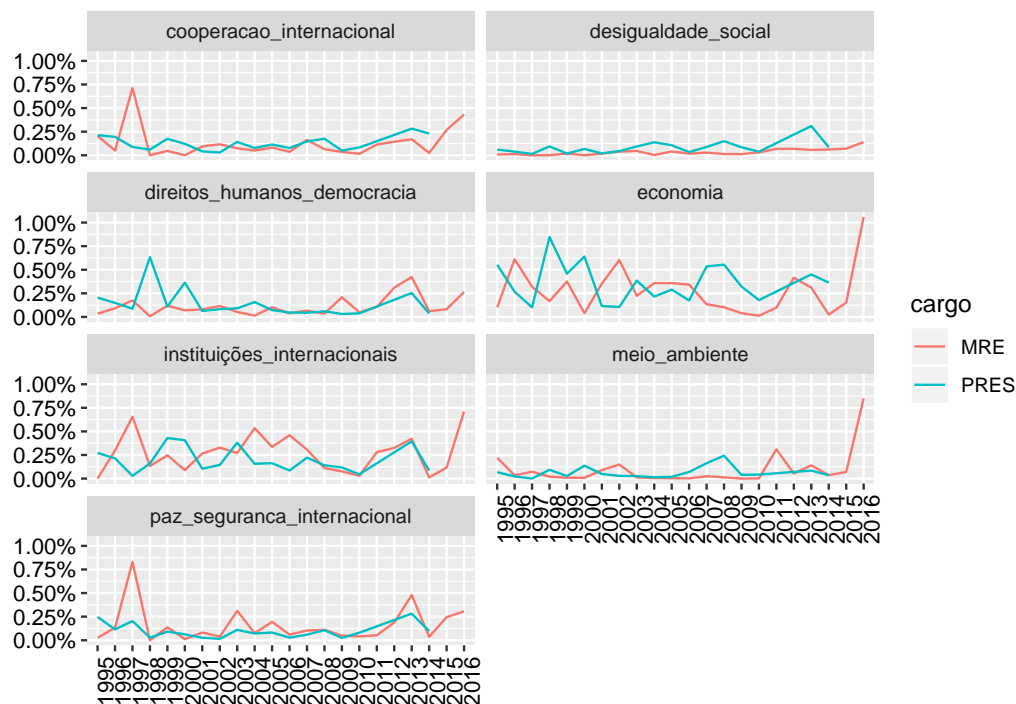
O gráfico informa a quantidade percentual de vezes em que os Presidentes e Chanceleres mencionaram as categorias em análise. Outra informação que pode ser obtida, através do explorar do dicionário inicial, é



verificar como cada um dos emissores em seus Pares mencionaram as categorias dos dicionários.



Por fim, é também possível comparar as ênfases temáticas em evolução temporal:



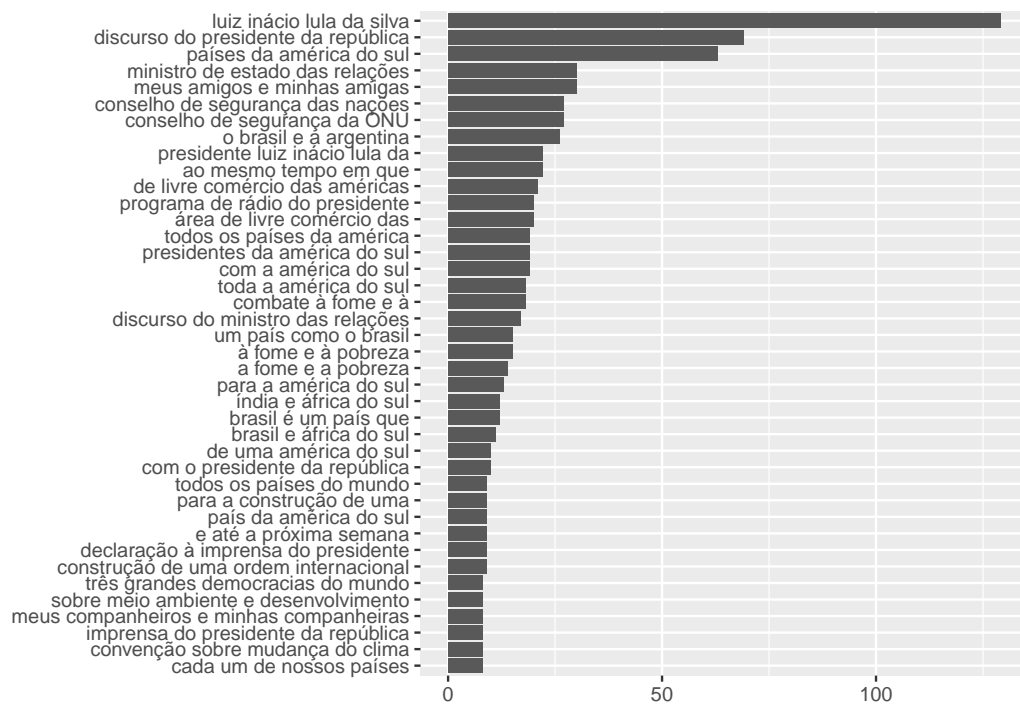
Tendo visto como o dicionário provê respostas numéricas para cada emissor e categoria, é possível partir para a apresentação dos primeiros resultados da fase de exploração textual. Conforme explicado, o objetivo é descobrir certos padrões e termos que precisam ser incorporados ou alterados na versão final do dicionário.

## 7. Exploração Textual [Primeiros Resultados]

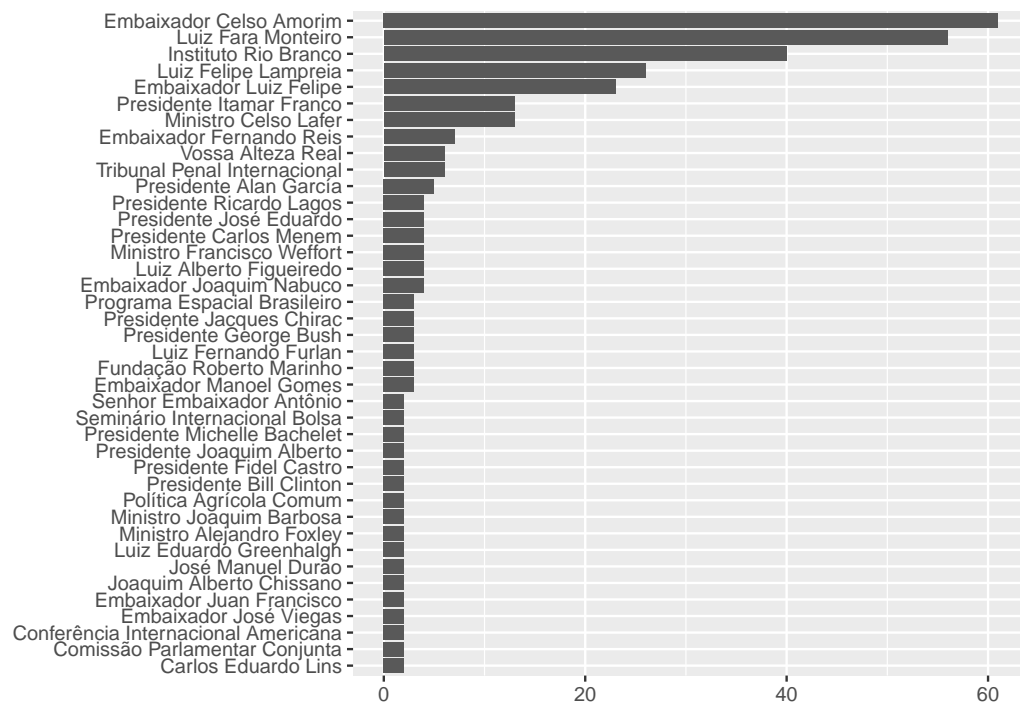
A fase de exploração textual foi dividida nas seguintes etapas:







Uma forma interessante de encontrar expressões importantes e específicas é através da análise de nomes próprios que aparecem no Corpus. Eles podem mostrar personalidades importantes, referências a lugares e instituições. O gráfico abaixo apresenta os 40 mais repetidos na amostra textual:



Para finalizar essa visão mais estrutural, vejamos como uma modelagem baseada em Topic Models categorizaria as palavras da amostra textual. Trata-se de uma modelagem automatizada em que cálculos de probabilidade são realizados com o intuito de alocar termos em categorias específicas:

##	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
##	[1,] "nabuc"	"globaliz"	"omc"	"amorim"	"gent"

##	[2,]	"voss"	"ambient"	"rod"	"acho"	"companheir"
##	[3,]	"excelent"	"competit"	"alca"	"vej"	"dinheir"
##	[4,]	"amizad"	"financeir"	"agrícol"	"dig"	"famíl"
##	[5,]	"cultural"	"estabil"	"subsídi"	"convers"	"acho"
##	[6,]	"franc"	"tecnológ"	"eua"	"algo"	"compr"
##	[7,]	"bilateral"	"crescent"	"amazôn"	"lá"	"acontec"
##	[8,]	"espírit"	"mod"	"conferent"	"venezuel"	"mil"
##	[9,]	"compartilh"	"dificultad"	"multilateral"	"sei"	"vou"
##	[10,]	"intercâmbi"	"capaz"	"amorim"	"pergunt"	"produz"
##		Topic 6	Topic 7	Topic 8		
##	[1,]	"itamaraty"	"onu"	"african"		
##	[2,]	"diplomac"	"hait"	"cúpul"		
##	[3,]	"branc"	"terror"	"árab"		
##	[4,]	"diplomat"	"milit"	"aliment"		
##	[5,]	"ministéri"	"chin"	"bilhõ"		
##	[6,]	"diplomát"	"conflit"	"chil"		
##	[7,]	"chancel"	"guerr"	"sul-american"		
##	[8,]	"atuaçã"	"cris"	"agricultur"		
##	[9,]	"servic"	"situaçã"	"agrícol"		
##	[10,]	"frent"	"eua"	"produçã"		

## 8. Reflexões para a versão final do dicionário

As mudanças listadas aqui possuem um caráter tentativo, reflexivo e, portanto, temporário. Elas são sugestões que precisam ser melhor aperfeiçoadas. Uma grande oportunidade para isso é com a banca de qualificação e as avaliações críticas dos examinadores. Em geral, até o momento, as modificações foram provenientes de três fontes principais: a) A divisão que o Itamaraty usa para agrupar determinados temas em seu site (na aba de Política Externa); b) Os resultados encontrados em pesquisa anterior publicados no artigo de Rocha (2017); e, c) Os achados trazidos nesse documento extraídos a partir da exploração do Corpus. Sem mais delongas, seguem as principais alterações para o dicionário final da tese:

- **Alterações derivadas da aba de Política Externa do site do Itamaraty:**
  - Inclusão da categoria “cultura” para contabilizar a quantidade de menções a arte, música, literatura, museus, audiovisual, teatro, artes cênicas e instituições culturais internacionais;
  - Inclusão da categoria “países e regiões” para contabilizar a quantidade de referências aos países e regiões do globo (de certo modo estaríamos mensurando também as relações bilaterais);
  - Na categoria “economia” seria interessante incluir também termos relacionados com comércio exterior, sistema internacional monetário e temas afins;
  - Na categoria “Instituições Internacionais” adicionar um conjunto de organismos regionais e multilaterais, dentre os quais: Unasul, CASA, Celac, CALC, OTCA, Cepal, GATT, OCDE, Aliança do Pacífico, Liga Árabe, OTAN, ASEAN, APEC, BRICS, IBAS, ASA, ASPA, FOCALAL, União Africana, UNAOC, Conferência Ibero-americana;
- **Alterações derivadas dos resultados encontrados em Rocha (2017):**
  - Adicionar categoria para contabilizar referências às menções sobre Política Externa como Política Pública;
  - Incluir categoria sobre ideologização, politização e partidarização da Política Externa Brasileira;
  - Prever uma categoria sobre diretrizes presidenciais, diplomacia presidencial, iniciativas e a própria figura presidencial;
  - Adicionar categoria com referências ao Itamaraty, funcionários, diplomacia, diplomatas e figuras ilustres;
  - Adicionar uma categoria sobre identidade internacional do Brasil;
- **Alterações derivadas do achados da exploração do Corpus da tese:**
  - Ao verificar os termos mais repetidos no Corpus, percebe-se a necessidade de:

- a) Adicionar mais termos à categoria “economia”, tais quais: dumping, anti-dumping, salarial, salário, rendimentos, liberalismo, neo-liberalismo, consenso de Washington, capitalismo, capital, marxismo, socialismo, micro-crédito, crédito, risco-brasil, microeconomia, macroeconomia, inflação, deflação.
- b) Adicionar termos referentes à TV e televisão na nova categoria “cultura”
- c) Adicionar uma categoria com menções à ciência e tecnologia e incluir termos como: cientista, científico, nanotecnologia;
- d) Adicionar uma categoria com menções à religião com termos sobre cristão, cristianismo, judeu, judaico, islã, budismo, etc;
- e) Adicionar na categoria “meio ambiente”: ecologia, terremoto, biodiversidade;
- **Ao verificar termos mais específicos do Corpus (tfidf):**
  - a) Atesta a importância de uma categoria para contabilizar países e regiões;
  - b) Reafirma a relevância de uma nova categoria para “cultura”
  - c) Há um padrão de menções a instituições políticas domésticas como parlamento, congresso, parlamentares, etc. Pensar em uma categoria;
  - d) Aparecem menções a carvão, energia nuclear, gasoduto, biocombustível que sugere a importância de subdividir meio ambiente em um ramo para energia;
  - e) Em desigualdade social seria interessante uma sub-categoria para contabilizar menções a minorias, gênero, homem, mulheres, juventude. Assim, seria interessante mudar o nome para “questões sociais”;
  - f) Na categoria “economia” adicionar termos referentes à moeda e monetário;
- **Ao verificar os resultados para a análise de expressões (Collocations):**
  - a) A relevância das menções a América do Sul precisa ser incorporada no dicionário;
  - b) Atesta a importância de categorias para menções aos Presidentes e aos diplomatas tradicionais;
  - c) Adicionar na nova categoria “cultura” as menções a rádio;
- **Na análise dos nomes próprios:**
  - a) Verifica-se demais a relevância de um conjunto enorme de personalidades governamentais. O problema é que não é possível exaurir as possibilidades em um dicionário. Portanto, uma categoria como essa não pode ser adicionada por falta de modos de operacionalização;
  - b) Dentro das menções ao Mercosul seria válido prever também as referências aos seus órgãos como Comissão Parlamentar Conjunta, Grupo Mercado Comum, Parlamento do Mercosul, etc.
  - c) Dentro da ONU, arranjar também sub-categorias para suas distintas agências;
  - d) Na categoria referente a aspectos sociais inserir: Programa Comunidade Solidária;
- **O modelo de Topic Models possui grandes limitações, mas, de um modo geral, atesta tudo que já foi inserido até aqui.**

Portanto, levando em conta cada uma dessas possibilidades de alteração, o dicionário final deve receber uma nova estrutura que reflita melhor o contexto textual. Nesse aspecto, uma importante atualização será a divisão das categorias em novas sub-categorias. Assim, por ora, pensamos na seguinte estrutura de categorização:

- Instituições Internacionais
  - Instituições Regionais
  - Instituições Multilaterais
- Países e Regiões
  - África
  - Ásia
  - América
  - Europa
  - Oceania
- Identidade Internacional do Brasil
- Itamaraty
- Cultura
- Presidentes e Diplomacia Presidencial
- Economia
- Meio Ambiente

- Conferências e Acordos
  - Recursos Energéticos
  - Desenvolvimento Sustentável (Sustentabilidade)
  - Geral
- Questões Sociais
- Cooperação Internacional
- Direitos Humanos e Democracia
- Paz e Segurança Internacional
- Natureza da Política Externa
  - Política Externa como Política Pública
  - Política Externa como Política de Estado
  - Ideologização da Política Externa