Amazonian Narratives in Presidential Discourse (1985 to 2020)

Henrique Sposito & Livio Silva-Muller

11/4/2021

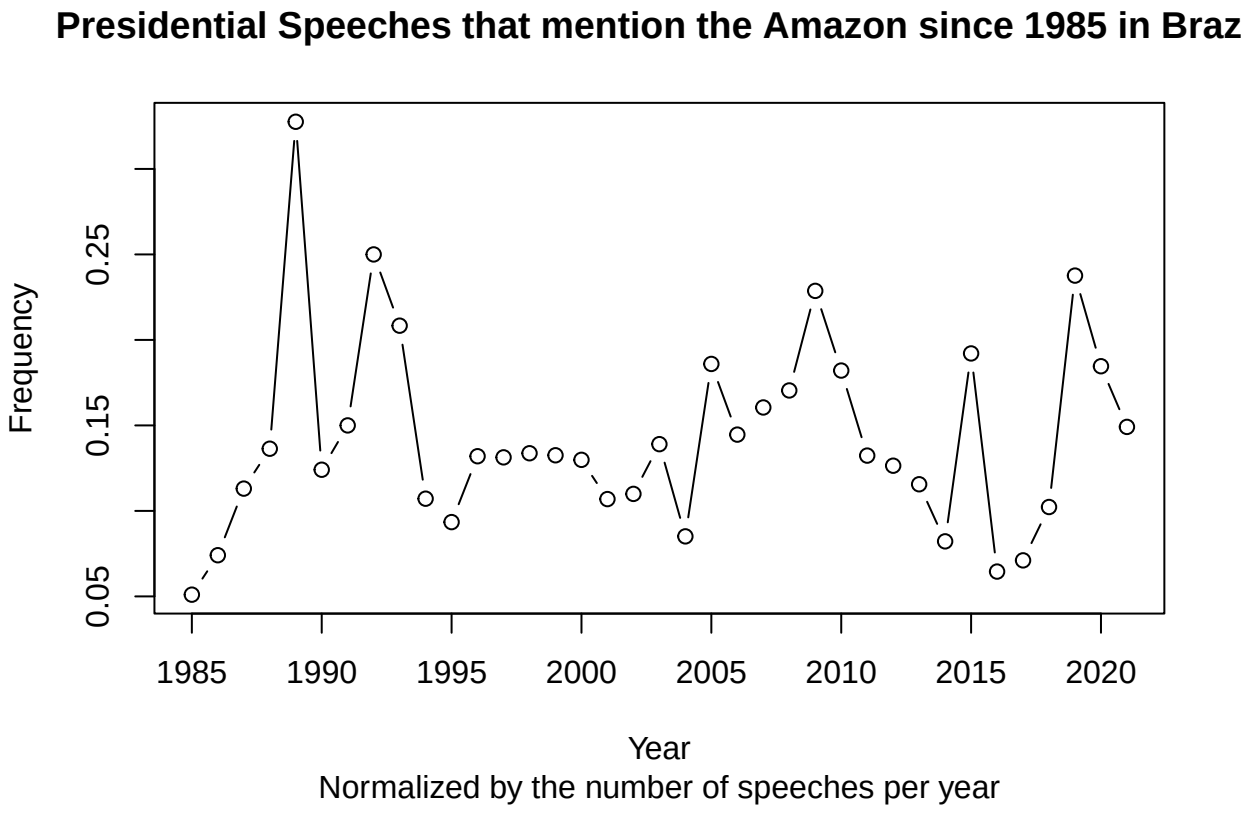
R Markdown

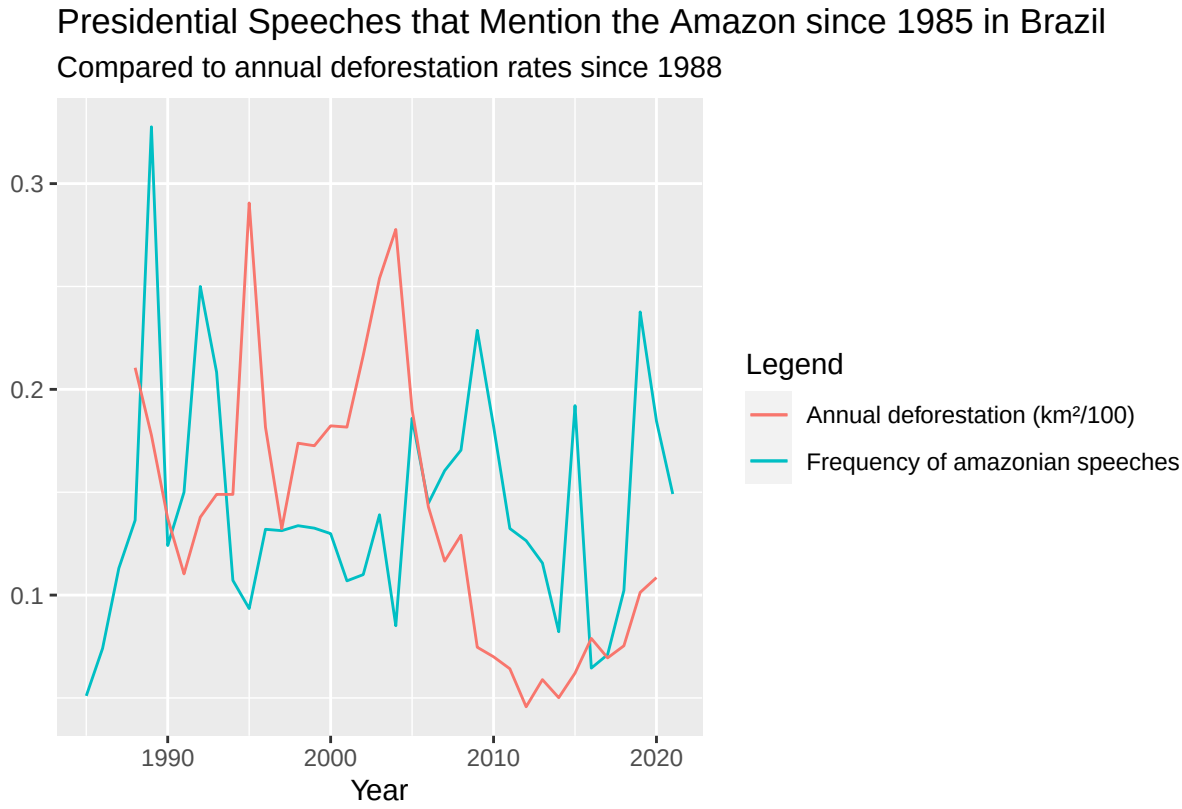
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Including Plots

You can also embed plots, for example:





requencies normalized by the number of speeches per year in dataset

```
# Location was added with poldis::extraxt_location(), still needs to be improved...
# To see the variable summary, please run: `summary(as.factor(Amazon_speeches$location))`
# Let's code the Amazonian states, the other Brazilian states, and speeches delivered outside Brazil.
Amazon_lc <- mutate(Amazon_speeches, stage = case_when(grepl("Amazonas|Para|Roraima|Acre|Amapa|Rondonia
# Just a comparison with the proportion of speeches per stage on the broader dataset
stages <- mutate(BR_Presidential_Speeches, stage = case_when(grepl("Amazonas|Para|Roraima|Acre|Amapa|Ro
stage_rate <- data.frame(Stage = names(summary(as.factor(Amazon_lc$stage))),
                         Amazon_speeches = paste0(round(as.numeric(summary(as.factor(Amazon_lc$stage))/
                         All_speeches = paste0(round(as.numeric(summary(as.factor(stages$stage))/6088)*
stage_rate
```

```
##                  Stage Amazon_speeches All_speeches
## 1     Amazonian_States         16.26 %       9.74 %
## 2        International         18.82 %      21.45 %
## 3 Non_Amazonian_States         53.19 %      58.77 %
## 4       Non_Identified         11.73 %      10.04 %
```

```
# Get 30 most frequent words overall
amazon_dtm  <- DocumentTermMatrix(amazon_corpus)
amazon_dtmm <- as.matrix(amazon_dtm)
amazon_f <- data.frame(term = names(colSums(amazon_dtmm)),
                       freq = colSums(amazon_dtmm))
amazon_f <- amazon_f[order(amazon_f$freq, decreasing = T),]
rownames(amazon_f) <- NULL
amazon_f30 <- data.frame(head(amazon_f, 30))
amazon_f30
```

```
##               term freq
## 1             pais  444
## 2  desenvolvimento  342
## 3            mundo  274
## 4              rio  259
## 5             meio  251
## 6           grande  248
## 7           paises  226
## 8         nacional  209
## 9          energia  200
## 10            povo  186
## 11           ainda  181
## 12        programa  178
## 13         estados  177
## 14      brasileiro  175
## 15           sobre  171
## 16      importante  169
## 17             vez  169
## 18            area  168
## 19         milhoes  166
## 20        floresta  165
## 21          apenas  164
## 22        ambiente  160
## 23      integracao  160
## 24           tempo  158
## 25       republica  156
## 26             mil  148
## 27      cooperacao  142
## 28         questao  142
## 29            toda  142
## 30           maior  139
```

```
# Let's see if bigrams help us more here
tamazon <- tidytext::tidy(amazon_corpus) %>%
  unnest_tokens(bigram, text , token = "ngrams", n = 2) %>%
  dplyr::count(bigram, sort = TRUE) %>%
  ungroup()
tamazon
```

```
## # A tibble: 63,080 x 2
##    bigram                          n
##    <chr>                       <int>
##  1 meio ambiente                 152
##  2 forcas armadas                 71
##  3 zona franca                    71
##  4 visite site                    66
##  5 muitas vezes                   62
##  6 palavra it                     54
##  7 site secretaria                53
##  8 desenvolvimento sustentavel    51
##  9 lula silva                     48
```

```
## 10 inacio lula                    47
## # ... with 63,070 more rows
```

```
# Let's get the 30 most frequent term by speaker
amazon_presid <- aggregate(Amazon_speeches$text, list(Amazon_speeches$president), paste, collapse = " ")
  rename(doc_id = "Group.1", text = "x")
amazon_presid_t <- tm::VCorpus(DataframeSource(amazon_presid)) %>%
  tidy() %>%
  unnest_tokens(word, text) %>%
  count(id, word, sort = TRUE)
Sarney <- head(dplyr::filter(amazon_presid_t, id == "Sarney"), 30) %>%
  mutate(Sarney = word) %>% select(Sarney)
Collor <- head(dplyr::filter(amazon_presid_t, id == "Collor"), 30) %>%
  mutate(Collor = word) %>% select(Collor)
Itamar <- head(dplyr::filter(amazon_presid_t, id == "Itamar"), 30) %>%
  mutate(Itamar = word) %>% select(Itamar)
FHC <- head(dplyr::filter(amazon_presid_t, id == "FHC"), 30) %>%  mutate(FHC = word) %>% select(FHC)
Lula <- head(dplyr::filter(amazon_presid_t, id == "Lula"), 30) %>%  mutate(Lula = word) %>% select(Lula)
Dilma <- head(dplyr::filter(amazon_presid_t, id == "Dilma"), 30) %>%  mutate(Dilma = word) %>% select(Di
Temer <- head(dplyr::filter(amazon_presid_t, id == "Temer"), 30) %>%  mutate(Temer = word) %>% select(Te
Bolsonaro <- head(dplyr::filter(amazon_presid_t, id == "Bolsonaro"), 30) %>%
  mutate(Bolsonaro = word) %>% select(Bolsonaro)
f_speaker <- cbind(Sarney, Collor, Itamar, FHC, Lula, Dilma, Temer, Bolsonaro) %>%
  kableExtra::kbl(caption = "Top 30 Words Per President") %>%
  kableExtra::kable_classic(full_width = F, html_font = "Times New Roman")
f_speaker
```

```
library(topicmodels)

# Fit a simple LDA model
# Please note that we randomly set a K = 10 (10 cluster = topics)
amazon_lda <- topicmodels::LDA(amazon_dtm, k = 10, control = list(seed = 1234))

# Get the top 10 words per topic
amazon_lda10 <- tidy(amazon_lda, matrix = "beta") %>%
  arrange(desc(beta)) %>%
  group_by(topic) %>%
  filter(row_number() %in% 1:10) %>%
  arrange(topic)
topten_topic <- aggregate(amazon_lda10$term, list(amazon_lda10$topic), paste, collapse=", ") %>%
  rename(Topic = "Group.1", Terms = "x") %>%
  kableExtra::kbl(caption = "Top 10 Words per Topic for Presidential Speeches Mentioning the Amazon") %
  kableExtra::kable_classic(full_width = F, html_font = "Times New Roman")
topten_topic
```

```
# Let's get the gamma (percent of topic in document for speakers)
amazon_lda_presid <- tidy(amazon_lda, matrix = "gamma") %>%
  arrange(desc(gamma)) %>%
  group_by(document) %>%
  filter(row_number() %in% 1:2) %>% # get the top 3 topics per speaker
  arrange(document) %>%
  mutate(gamma = round(gamma, 3)) %>%
```

Table 1: Top 30 Words Per President

Sarney	Collor	Itamar	FHC	Lula	Dilma	Temer
grande	meio	iniciativa	rio	pais	pais	seguran
pais	milhoes	integracao	energia	desenvolvimento	conferencia	forcas
desenvolvimento	ambiente	paises	it	mundo	mulheres	estados
area	floresta	cooperacao	desenvolvimento	milhoes	desmatamento	armada
cooperacao	defesa	mercosul	integracao	apenas	cumprimento	desmata
paises	recursos	tratado	cardoso	republica	jose	meio
nacional	brasileiros	grupo	meio	coisa	mundo	pais
sobre	problema	america	pais	desmatamento	presentes	publica
meio	programa	desenvolvimento	ainda	brasileiro	programa	tempo
rio	ambiental	economico	tempo	pessoas	desenvolvimento	brasilei
ambiente	brasileiro	amazonico	palavra	mil	deputados	exemplo
programa	desenvolvimento	amazonicos	fazendo	nacional	federal	interess
povo	exemplo	comercio	grande	rio	nordeste	nordest
atraves	questao	futuro	venezuela	secretaria	maior	presenc
natureza	trabalho	parlamento	povo	silva	paises	banco
mundo	estados	andino	sobre	paises	regioes	intelige
oportunidade	hectares	assembleia	nacional	estados	estados	moment
sarney	lado	grande	madeira	meio	grande	ambient
vez	mundo	livre	toda	voce	meio	area
ambiental	preservacao	politica	realmente	importante	nacional	banda
brasileiro	republica	protecao	itacoatiara	programa	cidades	disse
brasileiros	somente	abertura	importante	site	entao	jose
futuro	acailandia	area	presenca	visite	interior	larga
projeto	ainda	brasileiros	feito	ainda	milhoes	questao
maior	dessa	chile	producao	brasileira	questao	rondoni
patrimonio	evitar	colombia	vez	povo	sempre	coisa
politica	jornalista	conjunto	assim	grande	senador	espirito
ainda	pais	democracia	maneira	energia	estar	extraor
mil	planeta	economica	questao	vezes	importante	federais
grandes	sobre	guiana	sempre	entao	banco	federal

Table 2: Top 10 Words per Topic for Presidential Speeches Mentioning the Amazon

Topic	Terms
1	pais, rio, mil, entao, povo, mundo, programa, cidade, grande, voce
2	estados, meio, floresta, momento, federal, governadores, questao, vezes, toda, grande
3	paises, pais, conferencia, mundo, tempo, grande, cooperacao, tudo, sobre, brasileiro
4	rio, energia, madeira, porto, fazendo, itacoatiara, rondonia, producao, venezuela, pais
5	presenca, integracao, forcas, nacional, armadas, energia, defesa, grande, desenvolvimento, importante
6	desmatamento, pais, milhoes, mundo, paises, desenvolvimento, ainda, producao, area, hectares
7	pais, apenas, mundo, povo, floresta, energia, franca, coisa, desenvolvimento, coisas
8	jose, deputados, silva, republica, presentes, joao, federal, lula, secretaria, nacional
9	desenvolvimento, nacional, sobre, area, paises, ambiental, meio, ambiente, cooperacao, sustentavel
10	desenvolvimento, pais, cooperacao, integracao, meio, programa, ambiente, protecao, paises, defesa

Table 3: Top 2 Topics Per Speaker

president	topic	gamma
Bolsonaro	2	1.000
Bolsonaro	9	0.999
Collor	9	0.999
Collor	2	0.999
Dilma	8	0.999
Dilma	4	0.999
FHC	2	1.000
FHC	10	1.000
Itamar	5	1.000
Itamar	5	0.998
Lula	6	1.000
Lula	7	1.000
Sarney	10	1.000
Sarney	3	1.000
Temer	8	0.998
Temer	4	0.997

```
  distinct() %>%
  rename (president = document) %>%
  kableExtra::kbl(caption = "Top 2 Topics Per Speaker") %>%
  kableExtra::kable_classic(full_width = F, html_font = "Times New Roman")
amazon_lda_presid # Only Collor associated with topic cluster that mentions "ambiente"...
```

```
# Create the dictionary

sovereignty <- ("amazonia e brasileira|amazonia e nossa|soberania|interesse estrangeiro|ocupar|forcas a
development <- ("estrada|rodovia|hidroeletrica|desenvolv|balbina|itaipu|incentivos fiscais|integrar|int
conservation <- ("preserv|conserv|determinacao|povos indigenas|indigenas|direitos humanos|areas demarca
anti_environmentalism <- ("preservamos demais|nao vamos demarcar|demarcar menos|agricultura|producao de
```

Table 4: Narratives per President in Brazil

president	sovereignty	development	conservation	anti_environmentalism	other
Bolsonaro	100	58	109	9	23
Collor	23	13	44	1	4
Dilma	67	53	71	11	41
FHC	217	389	213	5	32
Itamar	7	24	9	0	0
Lula	249	312	274	31	129
Sarney	196	162	148	2	11
Temer	23	6	17	1	10

Table 5: Narratives per President in Brazil (normalized)

president	sovereignty	development	conservation	anti_environmentalism	other
Bolsonaro	0.22	0.13	0.24	0.02	0.05
Collor	0.15	0.09	0.29	0.01	0.03
Dilma	0.07	0.06	0.07	0.01	0.04
FHC	0.15	0.27	0.15	0.00	0.02
Itamar	0.12	0.43	0.16	0.00	0.00
Lula	0.12	0.15	0.13	0.02	0.06
Sarney	0.34	0.28	0.25	0.00	0.02
Temer	0.06	0.02	0.04	0.00	0.03

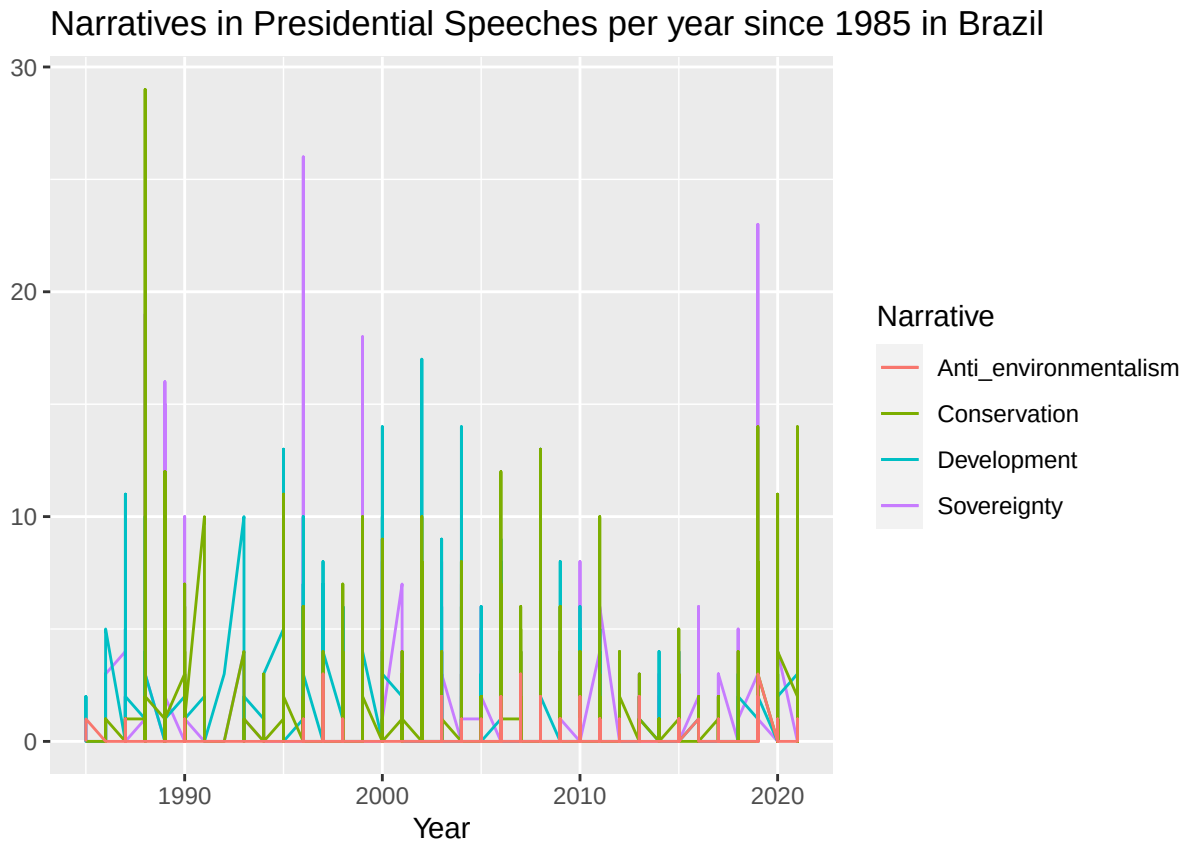


Table 6: Narratives per President in Brazil

stage	sovereignty	development	conservation	anti_environmentalism	other
Amazonian_States	186	288	193	10	32
International	127	190	121	7	34
Non_Amazonian_States	488	448	445	33	154
Non_Identified	81	91	126	10	30

Table 7: Narratives per President in Brazil (normalized)

stage	sovereignty	development	conservation	anti_environmentalism	other
Amazonian_States	1.33	2.06	1.38	0.07	0.23
International	0.78	1.17	0.75	0.04	0.21
Non_Amazonian_States	1.07	0.98	0.97	0.07	0.34
Non_Identified	0.80	0.90	1.25	0.10	0.30