

# Identifying family ties among politicians

## Challenges of information extraction evaluation

Diana Santos, Suemi Higuchi, Cláudia Freitas

d.s.m.santos@ilos.uio.no, Suemi.Higuchi@fgv.br, claudiafreitas@puc-rio.br

21 March 2022



## Contents

- Family ties in Brazilian politics: how can we assess this phenomenon?
- DHBB and AC/DC: corpus-based information extraction
- Tricky evaluation issues

- something that is common around the world (the Kennedys...)
- but considered by some as specially prevalent in Brazil
- several previous studies
  - Conniff (2003,2006)
  - Oliveira et al. (2017)
  - Schoenster (2014)
  - Higuchi (2021), Higuchi et al. (2019, 2022)

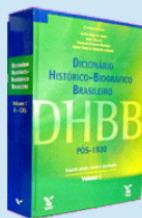
In this presentation we are looking more closely at the work which used DHBB to assess this.

## Dicionário Histórico-Biográfico Brasileiro

- Resource: *Dicionário Histórico-Biográfico Brasileiro* (DHBB);
- Biographical and thematic entries;
- Domain: Contemporary political history of Brazil (1930-);
- A reference work for historians, political scientists, journalists, etc.

1984

Print version



2001

Database format available on internet



2018

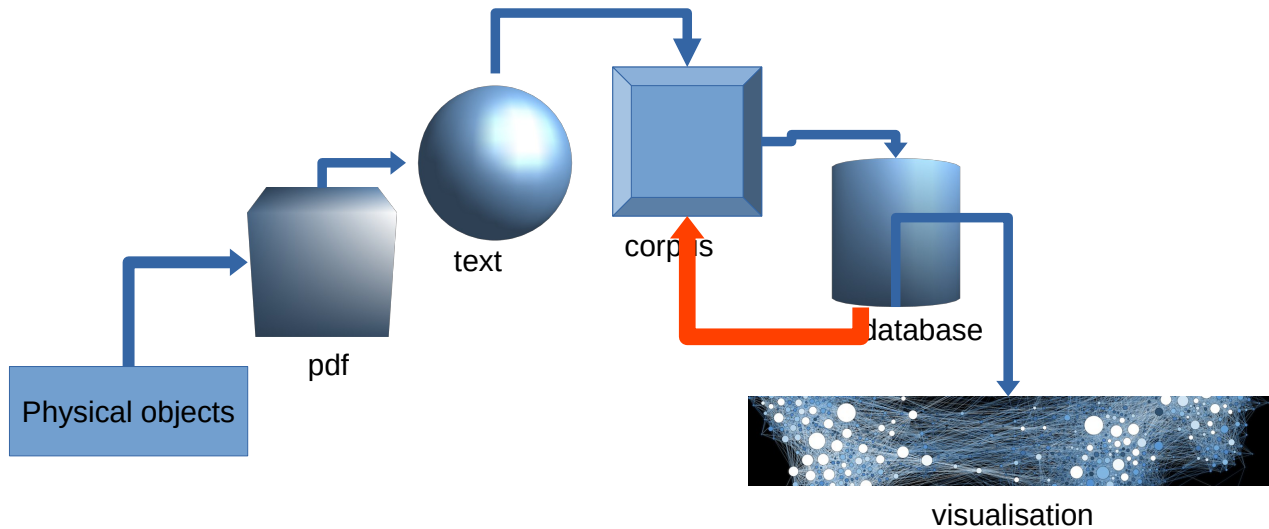
Corpus version available from Linguateca



DHN 2019 Conference

Follow this model of corpus-based Digital Humanities:

## Different phases and objects in DH



## DHBB as a corpus of encyclopedic entries

In addition to all other syntactic and semantic information that AC/DC corpora are endowed with, DHBB features

- Kind of entry (biographical or event)
- Gender of the biographee
- Family relationships (DHBB's contribution to the AC/DC "family")
- Grounding of politicians
- Marking "cargos" (political roles)
- (planned) Professions

## Our practical question

How to extract (and then later add to the corpus) the family information among politicians?

- family relations: relatively trivial
- among politicians: what is a politician?
- what is the information about the politicians we want?

And, after extraction, how much is much?

## Our contribution

We formalized five different levels for evaluation of information extraction of family ties.

- 1 did the patterns find valid family relations?
- 2 did the patterns find family relations between politicians?
- 3 did the patterns find family relations between politicians which were possible to identify in DHBB?
- 4 did the patterns extract family relations between politicians?
- 5 did the patterns extract family relations between politicians which were possible to identify in the DHBB?

## Did the patterns find valid family relations?

**YES:** **Seu pai foi agricultor** familiar na região de Rolândia

**NO:** Em novembro, a comissão central da CNBB publicou um manifesto em que, ao mesmo tempo que condenava a subversão, solidarizava-se com «**nossos irmãos bispos**, padres e leigos, no momento em que são vítimas da incompreensão e da injustiça no cumprimento de um trabalho apostólico autêntico» .

Precision: .985

## Did the patterns find family relations between politicians?

**YES:** **Seu pai foi deputado** federal por Rondônia de 1991 a 1995 .

**NO:** Foi casado com Maria Leonor de Castro Bastos desde janeiro de 1966, com quem teve **uma filha, Marcela Bastos**, além de 2 netos .

Precision: .605 .765

## Did the patterns find family relations between politicians which were possible to identify in DHBB?

**YES:** Com a redemocratização do país em 1945, enquanto os Ramos foram para o Partido Social Democrático (PSD) , os Konder participaram da organização da União Democrática Nacional (UDN) , cuja **seção estadual Irineu, pai de Jorge Bornhausen**, viria a presidir por mais de uma vez .

**NO:** Durante os trabalhos da CPI, Frota Aguiar interrogou como testemunha **Lutero Vargas, filho do chefe** da nação e presidente da seção carioca do PTB .

Precision: .595 .765

## Did the patterns extract family relations between politicians?

Now we are looking just at the patterns, not at the whole sentence.

**YES:** **Sua mãe foi deputada** federal, também pelo Maranhão.

**NO:** **Seu pai foi líder** político e prefeito de sua cidade natal em 1955 .

Precision: .405 .705

# Did the patterns extract family relations between politicians which were possible to identify in the DHBB?

**YES:** Entre seus familiares, encontravam-se seu primo Mauro Borges, deputado federal e governador de Goiás, e **seu tio, Pedro Ludovico Teixeira**, que foi integrante da Aliança Liberal e participou da Revolução de 1930, passando, a partir de então, a exercer um domínio político no estado de Goiás que durou quase 40 anos: como interventor (1930-1935; 1937-1945) , governador (1935-1937; 1951-1954) , constituinte (1946) e senador (1946-1951; 1955-1969) .

**NO: Seus dois irmãos atuaram na política:** José Silveira, como deputado federal pelo Paraná de 1959 a 1963, e Roberto Silveira, como governador do estado do Rio de Janeiro de 1959 a 1961 .

Precision: .375 .690

## Further examples

Of course, not all cases of possessives refer to the biographee:

**SIM SIM SIM NÃO NÃO** Um dos casos que analisou e que se chocou diretamente com o governo de Rosinha Garotinho, que contava com **seu marido, o ex-governador Anthony Garotinho**, como Secretário de Segurança, foi a morte do comerciante chinês Chan Kim Chang .

- There is a lot that needs to be specified when one evaluates
- All the five measures are valid, but measure different things
- There is further work to be done, not everything is well defined yet:
  - What to do about first and last names? Are all to be required?
  - How to assess concordances where more than one family relationship is described?
  - How to confirm the correct referent of the possessive pronoun?
- How to measure the (cumulative) extent of family relations? (further work)
  - % of the biographees who have a family relation described in their entry?
  - number of family relationships / number of mentioned politicians?