

Digital Impoliteness Corpus (CID-1)

Corpus composition

Created during the initial phase of the Digital Impoliteness project, the Digital Impoliteness Corpus (CID-1) brings together comments about public events involving three members of Brazil's Chamber of Deputies and a judge of Brazilian's Supreme Federal Court. The YouTube dataset presented here covers eight highly engaged videos on their respective subjects, two for each event. After extraction, the comments were classified for the presence of verbal attacks. The 6,512 comments selected on that basis make up this version of the corpus.

Congresswoman Sâmia Bomfim's videos discuss Luciano Zucco's question to her during the MST parliamentary inquiry: whether she wanted "medicine or a hamburger." The videos about congresswoman Érika Hilton show her response in Parliament to a statement denying that trans women are women. Congresswoman Duda Salabert's videos cover General Augusto Heleno addressing her as "sir" during the joint inquiry into the events of 8 January. Those about Judge Cármen Lúcia concern her vote in the Superior Electoral Court (TSE) ruling that made former president Jair Bolsonaro ineligible to run for office.

The comments classified as attacks were organized by video and automatically translated into English using Google Translate. The aligned corpus pairs each Portuguese comment with its English translation and records the public figure and source video. Table 1 gives the extracted and selected counts by video.

Table 1 – Composition of CID-1

Public figure	Event	YouTube video	Comments extracted from YouTube	Comments classified as attacks (CID)
Sâmia Bomfim	MST parliamentary inquiry: Luciano Zucco asked Sâmia whether she wanted "medicine or a hamburger."	CyzPttpAssw	458	200
		WCtSd5X9qJM	33	22
Érika Hilton	During a parliamentary debate, the congresswoman responded to a statement denying that trans women are women.	_N9gWlln4T0	5,971	809
		EzodFAnVTtc	691	213
Duda Salabert	During the joint parliamentary inquiry into the events of 8 January, General Augusto Heleno addressed the congresswoman as "sir."	g2sRbZnyB_4	6,386	3,609
		Vmxlro-oM6Y	1,585	1,107
Cármen Lúcia	The justice cast the decisive vote in the TSE ruling that made Jair Bolsonaro ineligible to run for office.	FUzWy4E6-gE	1,573	336
		cslGnM5P0n0	708	216
Total	Four events and eight videos		17,405	6,512

Source: Digital Impoliteness project.

Searching by public figure and video

In Sketch Engine, each comment carries an identifier for the public figure (*doc.target*) and the source video (*doc.video_id*). You can therefore select attacks directed at one of the four women or limit the search to one of the eight videos. The video IDs match the links in Table 1.

To create a subcorpus in Sketch Engine, open the subcorpus creation screen, similar to Figure 1, choose “Subcorpus from text types”, name the subset, and select the desired ID under *doc.video_id*. You can combine this filter with *doc.target* before saving.

Figure 1 – Screenshot of the subcorpus creation screen in Sketch Engine

Create subcorpus

Subcorpus name *

☒ Subcorpus from text types
☐ Subcorpus from concordance

expand all collapse all

doc.corpus ▼

doc.target ▲

camen_lucia
duda_salabert
erika_hilton
samia_bonfim

doc.relation ▼

doc.video_id ▲

CyzPtlpAssw
EzodFAnVTlc
FUzWy4E6-gE
Vmxfro-qM6Y
WCISd5X9qJM
_N9gWln4T0
cslGnMSP0n0
g2sRbZnyB_4

doc.source_row ▼

File ID ▼

File name ▼

Accessing CID-1

Once access has been granted by the project team, CID-1 can be consulted in Sketch Engine in Portuguese, in English translation, or as an aligned Portuguese-English corpus. To request access, email the address associated with your Sketch Engine account to impolidezdigital@gmail.com.

- [Corpus in Portuguese](#)
- [Corpus translated into English](#)
- [Aligned Portuguese-English corpus with video labels](#)

Publications based on CID-1

The publications below draw on CID-1 or subsets of its data.

Publications	Corpus subset used
OLIVEIRA, Ana Larissa; DRINÓCZI, Tímea; MIRANDA, Monique. <u>The new boundaries of the sayable in political discourse: the language of conflict as a challenge to pluralist societies</u> . In: Constitucionalismo, desigualdade e desinformação. Belo Horizonte: Editora Expert, 2024, p. 1147–1172.	<i>Subcorpora concerning Sâmia Bomfim and Érika Hilton</i> . This work examines YouTube comments about the two events, focusing on the circulation of misogynistic and transphobic attacks.
OLIVEIRA, Ana Larissa; MIRANDA, Monique; DRINÓCZI, Tímea; CUNHA, Gustavo Ximenes. <u>As novas fronteiras do dizível: a linguagem impolida contra figuras femininas das esferas política e judiciária</u> . Texto Livre, v. 18, e58111, 2025.	<i>Subcorpora concerning Sâmia Bomfim, Érika Hilton, and Justice Cármen Lúcia</i> . This article examines attacks and other expressions of impoliteness in YouTube and X/Twitter comments.
OLIVEIRA, Ana Larissa; DRINÓCZI, Tímea; SILVA, Janaína Lima Penalva; MIRANDA, Monique Vieira; MARQUES, João Pedro Cirino. <u>Linguagem Digital, Gênero e Democracia no Brasil</u> . Münster: Instituto de Comunicação da Universidade de Münster, 2025.	<i>Samples from the Sâmia Bomfim, Érika Hilton, and Duda Salabert subcorpora</i> . This educational book discusses attacks on the judiciary, women, and LGBT+ people, using examples and reflective activities.
MIRANDA, Monique; OLIVEIRA, Ana Larissa; DRINÓCZI, Tímea; PEIXOTO, Isabel. <u>Methodological Approaches and Data – A Brazilian Case Study</u> . In: DRINÓCZI, Tímea; OLIVEIRA, Ana Larissa (eds.). <u>Law, Linguistics and the Far Right</u> . London: Routledge, 2026.	<i>Subcorpora concerning Sâmia Bomfim, Érika Hilton, and Justice Cármen Lúcia</i> . This chapter describes the construction of part of CID-1 and the procedures used to study the comments.
OLIVEIRA, Ana Larissa; DRINÓCZI, Tímea; MIRANDA, Monique. <u>Online Verbal Attacks on Women in Power: A Brazilian Case Study</u> . In: DRINÓCZI, Tímea; OLIVEIRA, Ana Larissa (eds.). <u>Law, Linguistics and the Far Right</u> . Routledge, 2026.	Drawing on the subset described above, this chapter examines online verbal attacks on women in positions of power and their implications for democratic debate.
OLIVEIRA, Ana Larissa; MIRANDA, Monique; UNGER, Johann. The role of creative language in ostracizing transgender people online: a Brazilian case study. Accepted for publication in Zeitschrift für Diskursforschung.	<i>Duda Salabert subcorpus</i> . This article examines creative language in transphobic attacks against the congresswoman and other users' evaluations of those messages, using YouTube and X/Twitter data.