



Universiteit
Leiden
The Netherlands

Contributions to the computational processing of diachronic linguistic corpora

Cunha, E. Landulfo Teixeira Paradela

Citation

Cunha, E. L. T. P. (2020, March 19). *Contributions to the computational processing of diachronic linguistic corpora*. LOT dissertation series. LOT, Amsterdam. Retrieved from <https://hdl.handle.net/1887/133504>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/133504>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/133504> holds various files of this Leiden University dissertation.

Author: Cunha, E. Landulfo Teixeira Paradela

Title: Contributions to the computational processing of diachronic linguistic corpora

Issue date: 2020-03-19

APPENDIX A

Lexical heritage from past decades (data)

In Section 3.4.3, we mention the most common words in decade 2000s that, according to the algorithm proposed, were established in the Corpus of Historical American English (COHA) in particular previous decades. In this appendix, we display the lists containing these words (+ PoS tags), which are ordered according to their frequency in decade 2000s.

COHA employs PoS tags from the UCREL CLAWS7 Tagset. The complete description of these tags, including examples, is available at <http://ucrel.lancs.ac.uk/claws7tags.html>. The tags that occur in this table and in the other tables at Chapter 3 are:

- cc**: coordinating conjunction;
- cs**: subordinating conjunction;
- ii32**: general preposition (as part of a sequence);
- jj**: general adjective;
- jjr**: general comparative adjective;
- mc**: cardinal number, neutral for number;

mc2: plural cardinal number;
nn: common noun, neutral for number;
nn1: singular common noun;
nn122: singular common noun (as part of a sequence);
nn2: plural common noun;
nna: following noun of title;
nno: noun, neutral for number;
nnt1: temporal noun, singular;
nnt2: temporal noun, plural;
nnu: unit of measurement, neutral for number;
nnu2: plural unit of measurement;
ra: adverb, after nominal head;
rl: locative adverb;
rr: general adverb;
rt: quasi-nominal adverb of time;
to: infinitive marker;
uh: interjection;
vv0: base form of lexical verb;
vvd: past tense of lexical verb;
vvg: -ing participle of lexical verb;
vvgk: -ing participle catenative;
vvi: infinitive;
vvn: past participle of lexical verb;
vvz: -s form of lexical verb.

1850s	1860s	1870s	1880s
scientists_nn2	photo_nn1	phone_nn1	radio_nn1
experts_nn2	baseball_nn1	programs_nn2	skills_nn2
bike_nn1	photograph_vv0	photos_nn2	researchers_nn2
regional_jj	cigarette_nn1	hallway_nn1	golf_nn1
detective_nn1	options_nn2	long-term_jj	parking_nn1
focused_vvn	cops_nn2	makeup_nn1	ceo_nn1
tablespoons_nn2	typically_rr	downtown_jj	soviet_jj
strategies_nn2	protein_nn1	focus_vvi	networks_nn2
terrorists_nn2	telephone_nn1	classroom_nn1	techniques_nn2
users_nn2	concept_nn1	diabetes_nn1	ratings_nn2
shorts_nn2	dna_nn1	bacteria_nn2	nonetheless_rr
scientist_nn1	genetic_jj	driveway_nn1	korean_jj
technique_nn1	ethnic_jj	racial_jj	consultant_nn1
ongoing_jj	grabs_vvz	immune_jj	hockey_nn1
strategic_jj	backyard_nn1	colorful_jj	viewers_nn2
grill_nn1	cigarettes_nn2	taxi_nn1	olympics_nn2
aluminum_nn1	cd_nn1	scenario_nn1	residential_jj
parked_vvn	concepts_nn2	cultures_nn2	toxic_jj
fake_jj	spectacular_jj	palestinian_jj	subway_nn1
downtown_rl	terrain_nn1	diagnosed_vvn	meaningful_jj
focusing_vvg	focused_vvd	pinta_nn1	sneakers_nn2
photographer_nn1	specialist_nn1	gasoline_nn1	predictable_jj
treatments_nn2	shotgun_nn1	adolescents_nn2	championships_nn2
trailer_nn1	sensed_vvd	evaluation_nn1	binoculars_nn2
bicycle_nn1	worldwide_rl	victorian_jj	foyer_nn1
flipped_vvd	fingertips_nn2	semester_nn1	asset_nn1
gestured_vvd	aging_jj	bartender_nn1	dioxide_nn1
canyon_nn1	underwear_nn1	halloween_nnt1	initiatives_nn2
biology_nn1	clarity_nn1	cardboard_jj	coaching_nn1
ambulance_nn1	optimistic_jj	collaboration_nn1	heck_nn1
variables_nn2	technological_jj	headlights_nn2	orientation_nn1
institutional_jj	emotionally_rr	fingernails_nn2	overseas_rl
slowed_vvd	yanked_vvd	housing_vvg	focuses_vvz
buses_nn2	specialists_nn2	starters_nn2	kilometers_nnu2
providers_nn2	obsession_nn1	neon_nn1	interactive_jj
productivity_nn1	starter_nn1	ramp_nn1	penis_nn1
businessman_nn1	headlines_nn2	output_nn1	unpredictable_jj
interactions_nn2	lethal_jj	interviewed_vvn	technician_nn1
cosmic_jj	zoo_nn1	overweight_jj	touchdown_nn1
overly_rr	raiders_nn2	plasma_nn1	seasonal_jj
vibrant_jj	motors_nn2	shortage_nn1	shack_nn1
ironic_jj	peanut_nn1	microphone_nn1	capitalism_nn1
livestock_nn	innings_nn	format_nn1	hmm_uh
yep_uh	broccoli_nn1	developers_nn2	arthritis_nn1
heartbeat_nn1	awesome_jj	investor_nn1	sweetie_nn1
armored_jj	fictional_jj	outcomes_nn2	entrepreneur_nn1
detectives_nn2	aftermath_nn1	touch_ii32	abs_jj
toddler_nn1	canned_jj	finals_nn2	scoring_nn1
erie_jj	evolutionary_jj	developer_nn1	backdrop_nn1
inning_nn1	vulnerability_nn1	biologist_nn1	comeback_nn1

1890s	1900s	1910s	1920s
movie_nn1	computer_nn1	gon_vvgk	okay_rr
na_to	global_jj	electronic_jj	video_nn1
television_nn1	shit_nn1	aids_nn1	iraqi_jj
environmental_jj	movies_nn2	servings_nn2	airport_nn1
nuclear_jj	weekend_nnt1	sox_nn2	boyfriend_nn1
basketball_nn1	soccer_nn1	pickup_nn1	sexy_jj
garage_nn1	calories_nnu2	agenda_nn1	robot_nn1
ta_to	jazz_nn1	helicopter_nn1	cholesterol_nn1
overall_jj	coverage_nn1	fuck_nn1	workout_nn1
therapy_nn1	genes_nn2	kidding_vvg	bikes_nn2
terrorist_jj	muslim_jj	featuring_vvg	airlines_nn2
clinic_nn1	aircraft_nn	lipstick_nn1	antibiotics_nn2
mommy_nn1	someday_rt	goddamn_jj	and/or_cc
wireless_nn1	quarterback_nn1	teammates_nn2	nonprofit_jj
gym_nn1	technologies_nn2	windshield_nn1	vitamin_nn1
islamic_jj	pm_ra	someplace_rl	hometown_nn1
phones_nn2	rookie_nn1	lineup_nn1	activist_nn1
basically_rr	emissions_nn2	parked_vvd	airline_nn1
scheduled_vvn	suitcase_nn1	nationwide_rl	briefcase_nn1
awareness_nn1	buddy_nn1	part-time_jj	playoffs_nn2
initially_rr	activists_nn2	cafeteria_nn1	wheelchair_nn1
analysts_nn2	hispanic_jj	backseat_nn1	yoga_nn1
tablespoon_nn1	muslims_nn2	prestigious_jj	allegedly_rr
sector_nn1	regulatory_jj	helicopters_nn2	coordinator_nn1
sweater_nn1	expertise_nn1	dick_nn1	footage_nn1
coastal_jj	airplane_nn1	motivated_vvn	insulin_nn1
full-time_jj	jeep_nn1	vitamins_nn2	columnist_nn1
locals_nn2	routinely_rr	sponsored_vvn	creativity_nn1
homework_nn1	diesel_nn1	podium_nn1	feedback_nn1
flashlight_nn1	skiing_nn1	limo_nn1	processing_nn1
weekends_nnt2	grid_nn1	overseas_jj	c'm_vv0
homeland_nn1	researcher_nn1	recordings_nn2	ecological_jj
ok_jj	minimal_jj	airplanes_nn2	hormone_nn1
behaviors_nn2	feminist_jj	consultants_nn2	onstage_jj
hiking_vvg	robots_nn2	postwar_jj	firefighters_nn2
sexually_rr	syndrome_nn1	planners_nn2	pipeline_nn1
bombing_nn1	carbohydrate_nn1	viruses_nn2	implementation_nn1
deadline_nn1	priorities_nn2	lightweight_jj	highlights_nn2
motel_nn1	all-star_jj	protesters_nn2	nazi_jj
motivation_nn1	coconut_nn1	touchdowns_nn2	audio_jj
unemployment_nn1	prostate_nn1	flips_vvz	airports_nn2
hike_nn1	nutrients_nn2	artwork_nn1	operational_jj
trauma_nn1	buddies_nn2	campuses_nn2	slacks_nn2
catalog_nn1	containers_nn2	cleanup_nn1	stairwell_nn1
motorcycle_nn1	entrepreneurs_nn2	paranoid_jj	nylon_nn1
kinda_rr	trillion_nno	highlight_nn1	input_nn1
payroll_nn1	hormones_nn2	breathhtaking_jj	aerobic_jj
spotlight_nn1	artifacts_nn2	aspirin_nn	teammate_nn1
vodka_nn1	featured_vvd	small-town_jj	receptionist_nn1
short-term_jj	psychiatrist_nn1	comics_nn2	workouts_nn2

1930s	1940s	1950s
okay_jj	sidebar_nn1	backpack_nn1
computers_nn2	girlfriend_nn1	infrastructure_nn1
pizza_nn1	t-shirt_nn1	fuckin_rr
fucking_jj	online_rr	ncaa_nn1
israeli_jj	pc_nn1	backup_nn1
wildlife_nn1	teenage_jj	freeway_nn1
fuck_vv0	teenager_nn1	t-shirts_nn2
nba_nn1	teenagers_nn2	ponytail_nn1
sunglasses_nn2	mainstream_jj	girlfriends_nn2
guidelines_nn2	radar_nn1	salsa_nn1
siblings_nn2	upcoming_jj	linebacker_nn1
reportedly_rr	supermarket_nn1	spokeswoman_nn1
laser_nn1	innovative_jj	sunni_nn1
playoff_nn1	asshole_nn1	steroids_nn2
electronics_nn1	workplace_nn1	excerpted_vvd
therapist_nn1	microwave_nn1	pakistani_jj
bullshit_nn1	palestinians_nn2	robotic_jj
racism_nn1	fridge_nn1	pesticides_nn2
mph_nnu	stereo_nn1	superstar_nn1
paperwork_nn1	supportive_jj	hosting_vvg
processor_nn1	desktop_nn1	interface_nn1
fucking_rr	postseason_nn1	parameters_nn2
programming_nn1	sensors_nn2	spacecraft_nn
labs_nn2	monitor_vvi	recycling_nn1
tourism_nn1	israelis_nn2	award-winning_jj
cds_nn2	iraqis_nn1	offseason_nn1
medications_nn2	basics_nn2	surfing_vvg
gop_nn1	breakthrough_nn1	venues_nn2
demographic_jj	fda_nn1	nightstand_nn1
seafood_nn1	irs_nn1	environmentally_rr
condo_nn1	estrogen_nn1	automated_jj
rearview_nn1	predators_nn2	high-risk_jj
saudi_jj	vietnamese_jj	antioxidants_nn2
integrator_nn1	monitoring_vvg	charismatic_jj
behavioral_jj	bikini_nn1	oregano_nn1
low-income_jj	monitoring_nn1	nascar_nn1
adrenaline_nn1	airborne_jj	life-threatening_jj
bam_vv0	cardiovascular_jj	on-site_jj
condom_nn1	antibiotic_nn1	geek_nn1
hp_nn1	graffiti_nn	surreal_jj
homeowners_nn2	staffers_nn2	upbeat_jj
commercials_nn2	health-care_nn1	module_nn1
vinyl_nn1	late-night_jj	weaponry_nn1
plywood_nn1	thai_jj	chiles_nn2
dj_nn1	viewer_nn1	pollutants_nn2
marijuana_nn1	burgers_nn2	addictive_jj
implemented_vvn	predator_nn1	aerospace_nn1
boutique_nn1	yogurt_nn1	clueless_jj
skiers_nn2	targeting_vvg	retro_jj
condoms_nn2	paperback_nn1	multimedia_nn

1960s	1970s	1980s
affordable_jj	online_jj	headnote_nn1
mets_nn2	lifestyle_nn1	high-tech_jj
sustainable_jj	african-american_jj	laptop_nn1
ecosystems_nn2	suv_nn1	videos_nn2
upscale_nn1	parenting_nn1	globalization_nn1
medium-high_jj	genome_nn1	database_nn1
activism_nn1	high-profile_jj	booker_nn1
arguably_rr	workforce_nn1	pcs_nn2
healthcare_nn1	preheat_vv0	cilantro_nn1
mantra_nn1	ceos_nn2	african-americans_nn2
ecosystem_nn1	nonstick_nn1	hip-hop_jj
rehab_nn1	sitcom_nn1	high-end_jj
trendy_jj	jihad_nn1	ppg_nnu
wetlands_nn2	tsp_nnu	sustainability_nn1
lasers_nn2	dumpster_nn1	phd_nna
filmmakers_nn2	recycled_jj	condos_nn2
filmmaker_nn1	networking_nn1	counterterrorism_nn1
autism_nn1	pricey_jj	rapper_nn1
hosted_vvd	carb_nn1	databases_nn2
videotape_nn1	handheld_jj	gdp_nn1
disco_nn1	world-class_jj	biotech_nn1
gays_nn2	tofu_nn1	state-of-the-art_jj
makeover_vv0	sunscreen_nn1	catwoman_nn1
multicultural_jj	deregulation_nn1	mid-level_jj
broadband_jj	fast-food_jj	same-sex_jj
marina_nn1	gurney_nn1	buyout_nn1
attendees_nn2	mid-1990s_mc2	algorithms_nn2
hosted_vvn	ethnicity_nn1	pesto_nn1
boutiques_nn2	updates_nn2	biotechnology_nn1
one-on-one_mc	cardio_nn1	preservice_nn1
entitlement_nn1	shiites_nn2	glynnis_nn1
benchmark_nn1	countertop_nn1	download_vvi
scheer_vv0	sushi_nn2	mentoring_vvg
hologram_nn1	magisterium_nn1	tugger_nn1
fucked_vvn	transnational_jj	laptops_nn2
debuted_vvd	cloning_nn1	stakeholders_nn2
real-time_jj	groundwater_nn1	sweatpants_nn2
dismissive_jj	spokesperson_nn1	wastewater_nn1
fucked_vvd	trailhead_nn1	starbucks_nn2
restructuring_nn1	antidepressants_nn2	wetland_nn1
on-line_jj	hispanics_nn2	minivan_nn1
postmodern_jj	meds_nn2	pager_nn1
real-world_jj	bikers_nn2	gawain_nn1
cuz_nn1	mid-1980s_nn2	basher_nn1
lasagna_nn1	biking_vvg	creationism_nn1
simulations_nn2	recycling_vvg	feel-good_jj
deco_nn122	veggies_nn2	firewall_nn1
cornerback_nn1	midlife_nn1	ipo_nn1
surfers_nn2	superhero_nn1	camcorder_nn1
batmobile_nn1	lifestyles_nn2	downtime_nnt1

APPENDIX B

Empath categories

In Section 4.3.3, we mention that we used lists of words provided by the tool Empath in order to classify the lemmatized words according to categories that represent different semantic fields. We also say that, due to the high number of categories predefined by Empath (194 in total), we selected eight that showed interesting results and are relevant for our discussion: *government*, *internet*, *journalism*, *leader*, *negative emotion*, *politics*, *social media* and *technology*. In this appendix, we list the words that comprise each category used in this dissertation.

Government (113 words): abolish, accordance, accounting, activist, administration, administrative, administrator, advisor, agency, allied, ambassador, amendment, applicable, association, authorities, campaign, capitalist, citizenship, civil, civilian, coalition, commission, communist, congress, consensus, conservation, constitution, contribution, cooperation, cultural, decree, democracy, democratic, dictatorship, economy, elect, election, embassy, empire,

employ, enforce, enforcement, ethical, facility, federal, finance, financial, fund, funding, global, govern, government, homeland, illegally, immigration, implementation, independence, international, involvement, jurisdiction, law, liability, mandate, monetary, nation, negotiate, negotiation, obligation, opposition, organization, organization, parliament, partnership, petition, policy, politician, politics, populace, poverty, program, propaganda, protocol, province, recruiting, regime, regional, regulation, representative, republic, republican, resource, revolution, rule, ruling, safeguard, sector, senate, senator, socialist, societal, society, tax, terrorism, treason, treasury, tyranny, unethical, unified, united, unjust, unlawful, utopian, welfare.

Internet (79 words): access, account, application, archive, article, broadcast, browser, celebrity, chat, chrome, click, coding, compute, computer, cursor, cyber, data, dating, document, download, edit, editing, editorial, edits, email, explorer, facebook, firewall, forum, gamer, gaming, glitch, google, hack, hacker, hacking, header, homepage, icon, info, information, interface, internet, keyword, laptop, link, mobile, multiplayer, networking, online, page, post, posted, profile, program, programming, publish, reporting, research, router, scam, screen, search, segment, server, site, spam, spreadsheet, subscriber, surf, tab, trend, updated, virtual, web, website, wireless, worldwide, www.

Journalism (69 words): administration, analysis, article, assignment, biography, booklet, bookstore, bulletin, campaign, cite, column, composition, copy, corporate, credential, documentary, edition, editor, editorial, essay, excerpt, freelance, gazette, guidebook, handbook, headline, history, homepage, informational, inked, jot, journal, journalism, journalist, leaflet, literary, literature, magazine, newsletter, newspaper, notebook, obituary, outdated, paper, paperwork, paragraph, photo, photographer, photography, print, printed,

printing, printout, publicist, publish, publishing, report, reporter, research, scribbling, showbiz, site, subscription, summarize, tabloid, thesis, typewriter, write, writing.

Leader (98 words): administrative, administrator, admiral, adviser, advisor, agent, ambassador, assist, association, authority, battalion, bishop, boss, briefing, campaign, candidate, captain, chairman, chancellor, chief, command, commandant, commander, commanding, committee, competent, corporate, corporation, council, crew, dictatorship, diplomat, diplomatic, directive, director, duty, elect, election, elite, emperor, employer, enforcer, executive, father, federation, founder, general, government, governor, headquarters, hotshot, in-command, informed, join, leader, leadership, leading, member, mission, negotiator, obey, operation, order, organization, organization, oversee, overthrow, parliament, peacemaker, politician, president, promote, qualified, rank, ranking, rebel, renowned, representative, responsibility, rule, ruler, ruling, senate, senator, sergeant, sir, specifically, strategist, strongest, successor, superior, supervise, supervisor, supreme, syndicate, treasurer, trusted, veteran.

Negative emotion (94 words): accident, afraid, alone, angered, angry, ashamed, bad, badly, beat, beaten, blame, break, care, confused, crazy, crushed, cry, crying, dead, death, depressed, die, dieing, disappointed, drunk, dying, either, fault, fight, fighting, freaked, frightened, fucking, furious, guilty, hard, hate, hated, heartbroken, hell, hit, horrible, hurt, hurting, hurts, insane, kill, killed, killing, last straw, lie, loose, lose, losing, lost, mad, mean, monster, pain, pissed, poor girl, poor guy, punch, raped, react, reason, sad, scare, scared, scary, seeing, shocked, sick, so much pain, stop, stupid, surprised, swear, terrible, terrified, thinking, threatened, tortured, trouble, unthinkable, upset, violent, wanted, weak, worried, worse, worst, worst part, wrong.

Politics (78 words): activist, advisor, advocate, ambassador, amendment, aristocracy, campaign, candidate, citizen, citizenship, committee, communist, community, congress, consensus, conservative, conspiracy, constitution, controversial, controversy, corruption, council, declaration, decree, delegate, democracy, democratic, dictatorship, diplomacy, diplomat, dispute, divided, division, doctrine, elect, election, extremist, federation, fundraising, govern, governor, ideology, influential, involvement, jurisdiction, leadership, liberal, liberation, loyalist, monarch, monarchy, nation, national, nationwide, negotiation, overthrow, parliament, philosophy, policy, politically, politician, politics, presidential, province, provincial, regime, representative, republican, revolutionary, ruling, senate, senator, socialist, society, sovereign, spokesperson, tyranny, unify.

Social media (51 words): bio, browser, camera, chat, comment, computer, contact, delete, email, facebook, fan, follow, followed, follower, following, follows, format, forum, friends, hashtag, homepage, icon, instagram, interaction, message, messaging, messenger, multimedia, network, notification, online, onscreen, password, post, posting, profile, site, snapshot, status, tablet, texts, trend, tweet, twitter, typed, update, video, vine, viral, web, www.

Technology (118 words): 3d, advanced, android, audio, automate, battery, binary, browser, cable, camcorder, cellphone, cellular, cloning, coding, communication, compute, computer, computerized, connector, console, cyber, data, database, desktop, developer, device, digital, download, electronics, engineer, engineering, experimental, firewall, format, futuristic, gadget, gaming, generator, glitch, grid, hack, hacker, hacking, handheld, innovative, install, integrate, interactive, interface, intergalactic, internet, invention, inventor, ipad, keyboard, laboratory, laptop, machinery, mainframe, malfunction, manufacture, messaging, microchip, mo-

bile, module, monitoring, multiplayer, navigation, network, nexus, nuclear, online, operational, optical, outdated, portable, powered, processor, program, programmer, programming, projector, prototype, quantum, radar, research, researcher, robot, router, satellite, scanner, scanning, scientific, scientist, screen, sensor, server, simulator, site, software, solar, spacecraft, spaceship, system, tablet, tech, technical, technician, technological, technology, transmitter, typing, upgrade, virtual, virus, web, website, wireless.

APPENDIX C

Papers and presentations

Papers included in this dissertation

The research presented in this dissertation comprises the following five papers – all published, accepted or submitted for publication at the time of the dissertation defense –, which are reproduced here with minor changes and with consent from co-authors.

Chapter 2: Building a diachronic corpus of comments extracted from news portals and websites

Paper I:

Cunha, E. L. T. P., Magno, G., and Almeida, V. (2017). A elaboração de um coletor e de um corpus de comentários extraídos de portais de notícias. In *Anais do X Congresso Internacional da ABRALIN*, pages 764–771. Associação Brasileira de Linguística (ABRALIN).

Paper II:

Cunha, E. L. T. P., Magno, G., and Almeida, V. (under review). *Xereta: A Brazilian corpus of online news comments*. Manuscript submitted for publication.

Chapter 3: Establishment and obsolescence of linguistic items in a diachronic corpus

Paper III:

Cunha, E. L. T. P. and Wichmann, S. (in press). An algorithm to identify periods of establishment and obsolescence of linguistic items in a diachronic corpus. *Corpora*.

Chapter 4: Diachronic corpora and quantitative approaches to the lexicon: the case of the term *fake news*

Paper IV:

Cunha, E. L. T. P., Magno, G., Caetano, J., Teixeira, D., and Almeida, V. (2018). Fake news as we feel it: Perception and conceptualization of the term “fake news” in the media. In Staab, S., Koltsova, O., and Ignatov, D., editors, *Social informatics*, volume 11185 of *Lecture Notes in Computer Science*, pages 151–166. Springer, Cham.

Paper V:

Cunha, E. L. T. P., Magno, G., Caetano, J., Teixeira, D., and Almeida, V. (under review). *Quantifying the conceptualization of the term “fake news” in Brazilian and English-speaking media sources*. Manuscript submitted for publication.

Papers not included in this dissertation

Besides the previously mentioned papers, the author of this dissertation contributed to the writing of the following articles published, accepted or submitted for publication during his PhD studies:

Cunha, E. L. T. P. and Cambraia, C. N. (in press). A linguagem de uma fraude: Análise das confissões forjadas em nome dos irmãos Naves. *Fórum Linguístico*.

Cunha, E. L. T. P. and Lourenço, L. “Letteratura di immigrati”: Composições poéticas publicadas na imprensa italiana belo-horizontina no início do século XX. In *Revista da Imigração Italiana em Minas Gerais*, 2019.

Caetano, J. A., Magno, G., **Cunha, E. L. T. P.**, Meira Jr., W., Marques-Neto, H. T., and Almeida, V. (2018). Characterizing the public perception of WhatsApp through the lens of media. In *Proceedings of the 2nd International Workshop on Rumours and Deception in Social Media (RDSM 2018)*.

Otoni, R., **Cunha, E. L. T. P.**, Magno, G., Bernardina, P., Meira Jr., W., and Almeida, V. (2018). Analyzing right-wing YouTube channels: Hate, violence and discrimination. In *Proceedings of the 10th ACM Conference on Web Science (WebSci’18)*, pages 323–332. Association for Computing Machinery (ACM). [Best student paper award]

Xue, M., Magno, G., **Cunha, E. L. T. P.**, Almeida, V., and Ross, K. W. (2016). The Right to be Forgotten in the media: A data-driven study. In *Proceedings on Privacy Enhancing Technologies (PoPETs)*, 2016(4):389–402. [Also presented at the 16th Privacy Enhancing Technologies Symposium (PETS 2016), Darmstadt, Germany, 2016]

Lourenço, L. and **Cunha, E. L. T. P.** (under review). *Toponímia de influência indígena nos bairros de Belo Horizonte*. Manuscript submitted for publication.

Oral and poster presentations in conferences (as presenting author only)

During his PhD studies, the author of this dissertation presented the following communications and posters in conferences:

- *Contribuições para a coleta e a compilação de um corpus de comentários de portais de notícias*. 11th International Congress of the Brazilian Linguistics Association (ABRALIN), Maceió, Brazil, 2019. [Oral communication]
- *Fake news as we feel it: Perception and conceptualization of the term “fake news” in the media*. 10th International Conference on Social Informatics (SocInfo 2018), Saint Petersburg, Russia, 2018. [Oral communication]
- *A elaboração de um coletor e de um corpus de comentários extraídos de portais de notícias*. 10th International Congress of the Brazilian Linguistics Association (ABRALIN), Niterói, Brazil, 2017. [Oral communication]
- *Toponímia de influência indígena nos bairros de Belo Horizonte* (with Lorenza Lourenço). 11th International Congress of the Brazilian Linguistics Association (ABRALIN), Maceió, Brazil, 2019. [Poster presentation]

Talks as an invited speaker

Finally, the author of this dissertation was invited to give the following academic talks during his PhD track:

- *Linguística computacional: Uma área entre dois mundos* [Computational linguistics: An area between two worlds]. Federal Institute Southeastern of Minas Gerais (IF Sudeste MG), Barbacena, MG, Brazil, 2018.

- *I dialetti sul Web: Possibilità di preservazione e rivitalizzazione linguistica?* [Dialects on the Web: Possibility of linguistic preservation and revitalization?]. Settimana della Lingua Italiana nel Mondo. Faculty of Letters of the Federal University of Minas Gerais (FALE/UFGM), Belo Horizonte, MG, Brazil, 2018.

- *Linguística forense: O linguista como investigador e perito* [Forensic linguistics: The linguist as investigator and expert witness]. Faculty of Letters of the Federal University of Minas Gerais (FALE/UFGM), Belo Horizonte, MG, Brazil, 2018.

- *Computadores e sociedade: Novos desafios na era digital* [Computers and society: New challenges in the digital era]. Federal Institute Southeastern of Minas Gerais (IF Sudeste MG), Barbacena, MG, Brazil, 2017.

- *Linguística computacional: Problemas, tendências e desafios* [Computational linguistics: Problems, trends and challenges]. Una University Center, Betim, MG, Brazil, 2015.

- *Generalizations regarding 200 years of lexical evolution of English* (with Søren Wichmann). Work in Progress Series, Max Planck Institute for Evolutionary Anthropology (MPI-EVA), Dept. of Linguistics, Leipzig, Germany, 2015.

