"It's not the what but (also) the how": characterizing left-wing populism in political texts

María Paz Raveau[1]*, Claudio Fuentes-Bravo[2], Miguel Ángel Fernández[1], Juan Pablo Couyoumdjian[3,4] and María José del Solar[1]

[1]Faro, Universidad del Desarrollo, Santiago, Chile, [2]Facultad de Derecho, Universidad de Chile, Santiago, Chile, [3]Facultad de Gobierno, Universidad del Desarrollo, Santiago, Chile, [4]Facultad de Economía y Negocios, Universidad del Desarrollo, Santiago, Chile

Despite all the elasticity and even ambiguity surrounding the concept of populism, the existing paradigms converge in the recognition of a populist rhetoric. By using Natural Language Processing (NLP) tools we propose a set of linguistic and discursive markers to identify populist markers in Presidential speeches. The performance of these markers is subsequently tested against the Global Populism Database (GPD). We set-up a multinomial regression model to study the predictive power of these markers on the GPD populist score, focusing on left-wing populist leaders in Spanish-speaking countries in Latin America. We are thus able to characterize (left-wing) populism as a style of communication, as well as to understand what is behind this rhetoric. Our results show that ingroup and emotional content are more present in populist speeches. We also find a positive relation between populism and the use future tense and conditional connectors, which suggest an intention to manipulate the audience. These results have implications both for the current understanding of (left-wing) populist rhetoric and for the conceptualization of populism itself.

KEYWORDS

populism, natural language processing, linguistics, speech analysis, Latin America

1 Introduction

Within contemporary discourse in political science the definition of populism remains a debated frontier. Scholars have dedicated a vast amount of effort to isolate its defining elements (Weyland, 2001). However, as Mudde and Rovira Kaltwasser (2018) have noted, a growing number of academics and commentators are utilizing the term as a trendy catchphrase rather than as a rigorous conceptual construct.

Presently, the academic construct recognizes two dominant paradigms. The ideational approach, that under the view of Hawkins and Rovira Kaltwasser (2018), envisions populism as a stark divide between the virtuous "people," with an unwavering commitment to uphold popular sovereignty at any cost, and the morally corrupt "elites." In contrast, the political-strategic approach suggests that populism emerges from power-seeking opportunistic leaders, characterized by fluctuating ideological inclinations, utilizing policies primarily as instruments for their objectives (Weyland, 2003). Can these seemingly disconnected theoretical conceptions coexist within the same phenomenon? In this article, we assert that they indeed can. Whether viewed as a thin ideology (Mudde, 2004), a distinctive worldview dichotomizing the virtuous masses and the corrupt elite (Hawkins and Rovira Kaltwasser, 2018), or as a strategic political maneuver (Weyland, 2001; Betz, 2002), these elements converge in one pivotal arena: the construction of populist rhetoric.

Our argument leverages on natural language processing (NLP) tools to illuminate how language offers an alternative avenue to analyze the tools of populism.

In the context of comparative politics, we are interested in populism as a theoretical conception that can travel through time and space. From a historical perspective, the trajectory of Latin American politics has been tightly entwined with the concept of populist leadership, showing waves of emergence and decline of this phenomenon (De la Torre and Arnson, 2013). Moreover, its significance is not tied to any historical trajectory (De la Torre, 2007); instead, it travels through the region's political history, manifesting multiple ideological and contextual transformations. As posited by De la Torre (2017), distinct waves of populism across the region exhibit both continuities and variations that shed light on the trajectory and endurance of democratic regimes. Notably, Latin American politics has witnessed several critical populist epochs: the onset of classical populism in the 1930s and 1940s, the resurgence of populist discourses intertwined with neoliberal policies in the 1990s, and the emergence of left-wing populism in the new millennium (Conniff, 2020). More recently, we have seen the emergence of a new populist radical right (Rydgren, 2018; Mudde, 2019) in the region.

Empirical studies on populism have grown in importance. Canovan (2004) points out that most studies on contemporary populism approach it as a discourse. Such studies aim to determine, for example, the type of populism or the degree of populism present in the language of a given political agent (Bonikowski and Gidron, 2016; Oliver and Rahn, 2016; Kestler, 2022). This perspective was led by Laclau (2005) and Mouffe (2005) and since then, several pieces of literature have approached populism via discourse theory (Miscoiu et al., 2008; Stavrakakis, 2017). In our article we use automatic text processing techniques to analyze populism in speeches, and we do not ascribe to any particular school of discourse analysis. The integration between automatic text processing techniques and discursive theories is a pending task for language sciences.

The operationalization of populism through populist discourse assumes a distinct stylistic approach (Jagers and Walgrave, 2007). Our argument rests on the premise that language is the cornerstone element for studying populism across time and space. Whether viewed as a confrontation between elites and citizens, a dominance narrative, or a calculated power-grab by politicians, specific linguistic traits and features persist across diverse waves of populism. This has to do with the systematic and intensive use of discursive strategies aimed at creating realities that consolidate a predetermined vision. Noteworthy traits linked to the populist style encompass straightforward, emotionally charged language, a confrontational rhetoric pitting the "people" against elites, and a prevalence of the first-person perspective, among others (Charaudeau, 2009; Drămnescu, 2014; Oliver and Rahn, 2016; Rabotnikof, 2018). While contextual factors and institutional quality influence populism, this investigation focuses on features extrapolated from speeches. Our intention is to characterize populist discourse rather than encapsulate populism, specific leaders, or policies.

In this work, we aim to identify populist markers in presidential speeches, using different Text-Mining and Natural Language Processing techniques. Automatic identification of populism in political text is not new. Previous works have proposed some metrics based on dictionaries (Bonikowski and Gidron, 2016; Oliver and Rahn, 2016). However, as Hawkins et al. (2019) noticed, dictionary based methods work better when applied to a single country, where the context is held constant. Hence, here we propose some non-content markers - such as complexity, discursive connectors and syntactic elements - to identify populism regardless of the speech's content. In order to keep language constant in our analysis we focus on Spanish-speaking countries in Latin America, using data from the Global Populism Database (GPD). The GPD assesses the level of populism in speeches of 241 chief executives (324 government terms) from 74 countries globally (Hawkins et al., 2019). This gives us a populism score for each speech. Given data constraints we focus on left-wing populists. Setting up a multinomial regression model we are able to test different linguistic markers on populism. Our results show a positive effect of ingroup and emotional content on left-wing populism score, in accordance with previous, more general, studies. The positive effects of the future tense and conditional connectors suggest an effort to manipulate the audience by increasing the force of the argument. The negative or null effect of the self reference (use of I or We), which we also find, remains to be explained.

This article contributes to populism and comparative politics in three dimensions. Firstly, by focusing on the main features of (left-wing) populist rhetoric we offer a way to analyze populism in a manner that unifies the main existing conceptions; in doing so, we aim to characterize populist speech, not populism itself, or a populist leader or policy. In this sense, and in second place, this work represents a pioneering methodological effort to characterize the discursive features of populism utilizing NLP methods. This involves the adaptation of linguistic markers into Spanish and the operationalization of NLP methods. Thirdly, our study has important implications: if the foundation of populism lies within leaders' linguistic attributes it becomes feasible to assess its impact on other forms of political discourse and ascertain whether a contagion of populist style extends to mass media and even individuals' articulation of the political sphere.

2 On populism: multifaceted origins and conceptual clarity

When examining the causes of populism, of any inclination, we find that they are intimately related with contextual dynamics. In this regard, Mudde and Rovira Kaltwasser (2017) highlight the centrality of the divide between the elite and the populace, a regularity that explodes in turbulent societal moments. A divergent viewpoint emerges through the perspective of Aguilar and Carlin (2017) who see populism as a result of profound structural inequalities within democratic systems. On the other hand, Albertazzi and McDonnell (2015) focus on the erosion of political and social norms as a catalyst for populism. Furthermore, Mudde and Rovira Kaltwasser (2018) contend that populism emerges as a response to the crisis of representation within democratic frameworks, thus shedding light on the constructive potential of populist movements in bolstering democratic engagement and

accountability. Additionally, Mudde (2004) asserts that populism often finds its original causes in societies pivotal moments, typically during crises. These considerations underscore the necessity of integrating country-specific effects and contextual factors when studying variations in populism across time and nations.

When comparing several definitions, the core of the populist construct resides in what Hawkins (2018) has eloquently termed "a cosmic struggle between a reified will of the people and a conspiring elite". This is the fundamental link connecting populist approaches, irrespective of temporal or geographical contexts. It represents a political praxis rather than a standalone ideology, distinguished by the ethical dichotomy between "the people" and the elite (Bonikowski, 2016). It is important to note that populism endeavors to unite the nation's interests through this "cosmic" conflict, invariably involving the creation of an adversary (the elites) for the people to contest, often galvanized by a charismatic leader. These ideational facets are intrinsically linked to the linguistic construction of populist discourses, encompassing aspects like ingroup virtues or vices, distinctive grammatical features, and a rhetoric envisioning a shared future for "the people." On the other hand, the political-strategic perspective conceives populism as a strategic pursuit of power by political actors, emphasizing mobilization and persuasion of individuals (Olivas Osuna, 2021). Like any other politician, populist leaders are power-seeking rational actors who act on a value-based appeal with no single ideological driving force. In this sense they are viewed as engaging in calculated political manipulation.

But as Rueda (2021) points out, populism presents several characteristics beyond disputes between ideational, strategic, or rhetorical approaches. Moreover, charisma and telegenic appeals are part of its defining terms. As Urbinati (2019) discusses, populism relies on the power of language and rhetoric to construct a direct representation and communication tool that appeals to specific public sections. In this regard, Jäger (2023) characterizes the phenomenon as a distinctive mode of political communication, marked by closeness to the masses and detachment from the "establishment," orchestrated to maximize popular support (Mudde, 2017). Touraine and Armiño (1989) extend this narrative, asserting that populism derives potency from its endeavor to nurture political and social cohesion mechanisms, championing national culture and identity against external dominance.

In this paper we study populism following the framework advocated by Hawkins et al. (2019), seeking to unravel the linguistic dimension of populism. In devising a language-based framework to gauge populism, we undertake a theoretical exploration of the linguistic expressions of this multifaceted concept. The problem entails political manipulation and the use of language and discursive strategies in this context. De Cleen and Glynos (2021) reinforce the linkage between populism and language, spotlighting language's role as an outcome of the phenomenon. We adopt the two seemingly disparate perspectives on populism we have discussed, interweaving them through the pivotal role of language as its quintessential expression. On the one hand, we leverage Meijers and Zaslove (2021)'s contention concerning the ideational dimensions of populism. Alongside, we embrace the political-strategic approach pioneered by Weyland (2001). While subject to recent critique, the premise that populism emanates from rational

decisions made by leaders to sway voters carries profound linguistic implications. This notion underscores the necessity for simplicity and directness in the linguistic structure of populist rhetoric.

The attempt to identify populism through the prism of language is not novel. Preceding works have advanced metrics grounded in dictionaries. For instance, Bonikowski and Gidron (2016) dissected populism in presidential campaigns via a dedicated populism dictionary, comprising terms indicative of populism's presence. Likewise, Oliver and Rahn (2016) curated dictionaries for both political and economic elites. Nonetheless, the contextual variance across countries renders dictionary-based methods more apt for single-country analyses. Thus, as will be discussed in further detail below, we posit non-content markers such as syntactic complexity, discourse connectors, and linguistic elements as potent tools for detecting populism irrespective of specific content. Our proposition rests on utilizing natural language processing, harnessing computational advancements to fathom populism as both discourse and phenomenon, profoundly intertwined with leaders' articulation of the world surrounding them.

3 Left-wing populism in Latin America

Populism has been a recurrent phenomenon across Latin America over a century (Conniff, 2020). Nevertheless, the main characteristics and contexts in which it has emerged are as diverse as the region's history. The account of the phenomenon by Knight (1998) highlights the role of crisis and mobilization, transforming the populist style into an influential tool that leaders use to establish a direct and emotional connection with the masses. Furthermore, Cammack (2000) unravels the roots of populism in the significance of political institutions and the idea of structural crisis.

A historical account of the stages of populism in Latin America traces its origins back to the early 20th century, coinciding with the rapid modernization and societal metamorphosis in countries like Argentina, Mexico, and Brazil, which can be traced to conditions both endogenous (attitudes, popular organization, and capacity of the leaders) and exogenous (economic conditions, technological revolutions, and institutional arrangements) (Ruth and Hawkins, 2017). As these nations grappled with transformative shifts, populist movements arose as a reaction, aiming to galvanize the working class and rural sectors against established elites (De la Torre, 2007). The first wave of populism in the 1930s and 1940s witnessed the emergence of notable figures such as Juan Domingo Perón in Argentina, Getulio Vargas in Brazil, and Lázaro Cárdenas in Mexico. Characterized by charismatic appeals and welfare-oriented policies, these leaders cultivated wide constituencies. Subsequently, a new wave of populism took root in the late 1990s and early 2000s, featuring luminaries like Hugo Chávez in Venezuela, Evo Morales in Bolivia, and Rafael Correa in Ecuador. This wave diverged by vehemently rejecting neoliberal tenets and championing socialist and nationalist agendas. This period witnessed the enactment of policies encompassing industrial nationalization, social program expansion, and constitutional reforms, all aimed at redistributing power and resources to marginalized segments of society (De la Torre, 2017). More recently, we have seen the emergence of a new populist radical

right in the region; this includes Jair Bolsonaro in Brazil and Nayib Bukele in El Salvador. However, language and data constraints limit a systematic linguistic analysis of these leaders' rhetoric, and we will thus focus on the political discourse of left-wing populist leaders.

In Latin America, left-wing populist leaders exhibit recurring traits, including charismatic allure, an emphasis on the concerns of the people, and a pervasive anti-elitist narrative. Often positioning themselves as advocates for the underprivileged and the labor force, these leaders pledge to combat corruption, inequality, and oppression (De la Torre, 2017). Altogether, frequently relying on personalized leadership styles, occasionally they undermine democratic institutions in the process.

Populism, a multifaceted and intricate phenomenon, has profoundly shaped Latin America's political panorama (Houle and Kenny, 2018). Thus, the roots of populism are deeply embedded in the region's socioeconomic transformations, while globalization, neoliberalism, and the ascension of socialist and nationalist ideologies have molded its contemporary expressions. With good reason, Latin America has been considered the land of populism (De la Torre, 2017). And prominent theorists and the most complex theories have emerged in the region (Basset and Launay, 2013). Is there a way to characterize left-wing populism across the region? Linguistics and natural language processing may have the answer.

4 Data and methods

4.1 Data

The Global Populism Database evaluates, via human codification, the level of populism in speeches of 241 chief executives[1] (324 government terms) from 74 countries globally (Hawkins et al., 2019). The dataset comprises 1,240 speeches, mainly spanning from 2000 to 2022. Employing a textual analysis technique termed "holistic grading", the GPD constructs a rubric featuring divergent categories and anchor texts to train coders. This method yields a populism score for each speech, thus facilitating the development of a regression model to explain the populism score.

Our analysis is confined to speeches from Spanish-speaking Latin American countries due to the nature of the markers. In order to increase our sample we also take Spain into account. Given data availability in the GPD database, this means we only include left-wing populists. After excluding 21 speeches due to poor text quality, our dataset reduces to a total of 290 speeches from 18 countries and 67 leaders. For each term in office, the database provides a populism score on four types of speeches: campaign speech, ribbon-cutting speech, international speech, and a famous speech which presumably represents the leader at his or her best. Table 1 shows the speeches. Here we include the different types of speeches and different terms by each leader (this average score was calculated only for demonstrative purposes). The number of speeches of each type and country are shown in Table 2. While the speeches from GPD date mainly from the year 2000 onward, the authors of this database also included some earlier speeches and leaders when they recognized cases of populism. While this leaves us with an

unbalanced dataset, we have included a dummy variable to account for different time periods, and run an additional model excluding the earlier speeches.

4.2 Populism score

In order to identify the speech markers of populist discourse, we build a multinomial regression model on the populism score assigned by the GPD. As each speech is graded by more than one coder, we used the average score per speech as the dependent variable. The coding is based on a grading technique in which the coders apply an integer grade scale and a rubric to identify rough attributes of texts at each grade, for which they are trained through repeated exposure through anchor texts (Hawkins et al., 2019). The texts are assigned a decimal grade scale from 0 through to 2, where a score of 0.5 rounds up to 1 and 1.5 rounds up to 2.[2] These three categories can be understood as 0 being "not at all or slightly populist," 1 being "moderately populist" and 2 being "extremely populist." The detailed description of each category can be found in Hawkins et al. (2019). In our work, we use the average score per speech as a categorical variable, using the same criteria as the GPD for rounding up the decimal scores.

4.3 Explanatory variables

A number of features have been identified with a populist style: (i) the use of language must be simple, direct and emotional, to convey the voters that the leader, as well as his party, belong to "the people" (Drămnescu, 2014; Oliver and Rahn, 2016). It may also include metaphors, indelicate language and insults toward the adversary (Charaudeau, 2009); (ii) Use of "I" (the leader, who represents the people) and "We" (the party, the people) (Charaudeau, 2009; Oliver and Rahn, 2016); (iii) increase of "ingroup" feelings, by emphasizing common interests and values, such as nationality, religion (Drămnescu, 2014; Oliver and Rahn, 2016), or the notion of a "heartland," which evokes an idealized vision of the past (Taggart, 2004); (iv) seduction by fear: related to the "ingroup" feelings, there is also the fear of the "enemies of the people," which can be internal or external (Drămnescu, 2014; Oliver and Rahn, 2016). A common interpretation of the last two points is that of the "people vs. elite" problem. Here, the enemy is internal, and may be embodied by political elites (parties, government, and ministries), but also by the media or the intellectuals or economic elites (Jagers and Walgrave, 2007; Bonikowski and Gidron, 2016; Mudde and Rovira Kaltwasser, 2017). Finally, (v) the populist speech shows a moralizing language, which reduces reality to a conflict between the good (the people) and the evil (the elites) (Oliver and Rahn, 2016; Rabotnikof, 2018).

The preceding traits have been previously proposed by different authors, and here we aim to operationalize some of them, and test them in our model. But we can go further, and use Critical Discourse Analysis to delve deeper into the

---

1   The term "chief executive" includes presidents and prime ministers.

2   In previous versions of the GPD, coders applied an integer grade scale with values 0, 1, and 2.

TABLE 1 Average populism score by country and leader, from Global Populism Database.

Country	Leader	Average score	Country	Leader	Average score
Argentina	Peron	2.000	Bolivia	Sanchez	0.300
	Menem	1.000		Mesa	0.167
	de la Rua	0.175		Morales	1.5, 1, 1.35
	Duhalde	0.533	Chile	Lagos	0.083
	Kirchner	0.250		Bachelet	0.000, 0.213
	Fernandez	0, 0.188		Piñera	0, 0.15
	Macri	0.038	Uruguay	Batlle	0.217
Colombia	Pastrana	0.025		Vazquez	0.25, 0.1
	Uribe	0.000		Mujica	0.125
	Santos	0, 0.0625	Ecuador	Noboa	0.233
Costa Rica	Rodriguez	0.083		Gutierrez	0.967
	Pacheco	0.222		Palacio	0.389
	Arias	0.000		Correa	1.25, 1.73, 1.52
	Chinchilla	0.000		Moreno	0.188
	Solis	0.533	Panama	Torrijos	0.222
Dominican Republic	Mejia	0.075		Martinelli	0.500
	Fernandez	0.333, 0.25		Varela	0.250
	Medina	0.175, 1.375	Guatemala	Berger	0.000
El Salvador	Flores	0.000		Colom Caballero	0.125
	Saca	0.583		Molina	0.467
	Funes	0.500		Morales	0.150
	Sanchez Ceren	0.625	Paraguay	Duarte	0.500
Honduras	Flores	0.000		Lugo	0.000
	Zelaya	0.500		Franco	0.000
	Lobo Sosa	0.333		Cartes	0.117
	Hernandez	0.175	Peru	Toledo	0.333
Mexico	Cardenas	1.000		Garcia	1.000
	Fox	0.250		Humala	0.500
	Calderon	0.125		Kuczynski	0.033
	Peña Nieto	0.000	Spain	Aznar	0.050, 0.050
	Lopez Obrador	0.963		Zapatero	0, 0.362
Nicaragua	Bolaños	0.000		Rajoy	0.0125, 0.225
	Ortega	1.25, 1.1, 1			
Venezuela	Chavez	1.83, 1.83, 1.67			
	Maduro	1.567			

The speeches in the GPD date from 1990, with three exceptions (Cárdenas, 1935; Peron, 1946; Menem, 1989), but most speeches are from 2000. For leaders with more than one term, average populist scores are listed in chronological order.

rhetorical phenomenon. In accordance with Weyland (2001) and the political-strategic approach, we hypothesize that populism can be viewed as a form of manipulation. The cognitive pragmatic model of manipulation focuses on the operational linguistic and pragmatic aspects of information processing that are involved in the success of the manipulation (Maillat and Oswald, 2009, 2011; Oswald and Maillat, 2013; Oswald, 2014). In this pragmatic

approach to human communication, the theory of relevance (Sperber and Wilson, 1995; Wilson and Sperber, 2012) offers a model of information processing that can be a useful framework for developing a model of discursive influence or manipulation. Based on this, we can characterize manipulation as a double constraint: it tries to hide or remove critical content that weakens the message, and at the same time, it tries to highlight the "favorable"

TABLE 2 Number of campaign speeches, famous speeches, international speeches, and ribbon-cutting speeches, per country, included in this analysis.

Country	Campaign	Famous	International	Ribbon	Total
Argentina	8	8	5	6	27
Bolivia	3	4	4	3	14
Chile	5	5	5	5	20
Colombia	4	4	4	4	16
Costa Rica	2	5	4	5	16
Dominican Republic	5	5	4	5	19
Ecuador	4	7	7	6	24
El Salvador	3	4	4	4	15
Guatemala	3	4	3	4	14
Honduras	1	3	4	1	9
Mexico	4	5	4	4	17
Nicaragua	4	4	3	3	14
Panama	2	3	3	3	11
Paraguay	1	4	4	3	12
Peru	2	4	4	4	14
Spain	6	6	6	6	24
Uruguay	2	4	4	3	13
Venezuela	3	4	4	0	11
Total	62	83	76	69	290

information that strengthens the message (Maillat and Oswald, 2009, 2011).

Based on this perspective, and in order to complement our analysis, we also incorporated an additional set of pragmatic features. Thus, we have classified the independent variables into four classes: (i) content variables, (ii) complexity, (iii) pragmatic features, and (iv) control variables. Each variable, whether binary or continuous, is estimated at the sentence-level, and then a speech average is calculated.

4.3.1 Content variables

Linguistic Inquiry and Word Count Dictionary (LIWC): LIWC dictionary is composed of around 6,400 words and word stems (Pennebaker et al., 2015). It has also been translated into Spanish (Ramírez-Esparza et al., 2007) and includes general descriptor categories, standard linguistic dimensions, punctuation categories, personal concern categories, and 41 word categories tapping psychological constructs (e.g., affect, cognition, biological processes, drives).

Moral Foundations Dictionary (MFD): MFD is a collection of words representing the positive and negative aspects of five moral foundations (Haidt et al., 2009): Care/harm (concern for the suffering of others), Fairness/reciprocity (concerns about unfair treatment and inequality), Ingroup/loyalty (concerns related to obligations of group membership), Authority/subversion (endorsing social hierarchy), and Purity/degradation (concerns about physical and spiritual contagion). MFD also includes an

eleventh category of "general morality." Although the MFD is only available in English, we used the semi-automated method developed by Matsuo et al. (2019) to translate the dictionary into Spanish.

Here we utilize the MFD to evaluate the use of terms related to both positive (Ingroup Virtue) and negative (Ingroup Vice) aspects of ingroup identification. As for the emotional and moralizing language, we have also included the Morality Ingroup and Morality Vice categories. From LIWC we incorporated the negative and positive emotion categories (EmoNeg and EmoPos).

4.3.2 Complexity

Complexity in language originally derives from research in education, and has been subsequently used in other areas. Within the field of political science, Spirling (2016) uses the Flesch score, a metric that considers the number of syllables relative to the number of words found in documents, to show a linguistic simplification of parliamentary speeches after the Second Reform Act in Britain. The Flesch score has also been used to analyze differences in language complexity between liberals and conservatives (Schoonvelde et al., 2019), and to study how the members of Parliament adapt the complexity of their speeches to their constituents features (Lin and Osnabrügge, 2018). Despite the extended use of the Flesch score and other indicators of lexical diversity—such as the *Type-Token Ratio* (TTR) (Malvern et al., 2004)—these metrics rely solely on lexical features. Here we go beyond lexical complexity, and

incorporate syntactic measures of complexity. Our work builds upon the research of Wolfe-Quintero et al. (1998), Ortega (2003), Lu (2010) on second language acquisition.[3]

Syntactic complexity measures the length of "production units"—such as sentence and clauses—and the amount of subordination, coordination and sophistication of particular syntactic structures (Ortega, 2003). The first attempts to develop measures of language complexity that go beyond lexical aspects date back to the late 1970s, with the works on proficiency in a second language by Larsen-Freeman and Strom (1977) and Larsen-Freeman (1978). Later on, Wolfe-Quintero et al. (1998) and Ortega (2003) studied the cumulative evidence on second language development and proficiency. Based on these works, Lu (2010) proposed a computational approach for automatic analysis for some of these syntactic complexity measures. From the fourteen measures selected, six were covered in both Wolfe-Quintero et al. (1998) and Ortega (2003), and another three were recommended by the former for further work.

Lu's computational approach was possible thanks to the Stanford parser, a novel generative model for natural language tree structures, which allowed to model lexical dependency and syntactic structure (Klein and Manning, 2002). Since then, the original model has been permanently refined by the Stanford NLP Group, so in this work we use Stanza, a Python NLP package for linguistic analysis that includes Spanish models (Qi et al., 2020). Because the parser differs from one language to another, our operationalization of syntactic complexity measures is inspired in Lu's work, but it has been specially derived—using Stanza—for the Spanish language. Following Lu (2010), in Table 3 we define the syntactic structures and complexity metrics.

4.3.3 Pragmatic features

In every act of making a statement there is a speaker and a receiver, and the former attempts to influence the latter. Therefore, in addition to analyzing the statement's content and syntax, we must take care of the relation between the text and the users. In semiotic analysis, the study of the relation between the signs and the interpreter has been called "pragmatics," or the "pragmatical dimension of semiosis" (Morris, 1938).[4]

Acknowledging speech as a type of communicative activity, we intend to capture the enunciator's intention through different elements, such as discursive connectors. We must also study verbs, as they are elements especially marked by the presence of the enunciator (Gutiérrez Araus, 2015). Since there can be multiple verbs in one sentence, our operationalization considers only the verb that acts as the root of the sentences. In sentences containing at least one verb, the root will be a verb. Sentences without verbs are removed from our model. In periphrastic constructions we use the verb that is conjugated.

---

3   Linguistic analysis can also be studied from different perspectives, such as Functional Grammar (Halliday and Matthiessen, 2004), Critical Discourse Analysis (Fairclough, 2013) and other theoretical approaches (Gidron and Bonikowski, 2013).

4   As Morris (1938) clarifies, "pragmatics" must be distinguished from "pragmatism," just as "pragmatical" from "pragmatic."

Discursive connectors. A speech is not only a sequence of sentences or clauses. Through a series of operations, any piece of speech constitutes a semantic-pragmatic unit. Connective elements (also called connectors) establish logical-semantic relations between two sentences, clauses or paragraphs (Calsamiglia and Tusón, 2012). Connectors guide the reader to interpret the text in the way the writer intended (Montolío, 2001). There are different types of connectors:

- Conditional: introduces a restriction between two elements, so one of them should be understood as a requisite for the other one to be true (if, after, as long as, once).
- Additive: aims to link two pieces of information that are similar, or point to the same direction (in addition, furthermore, besides).
- Adversative: signals a contrast between the information present in both elements (but, yet, however, instead).
- Consecutive: signals the cause-effect relation between two elements, so the element in which the connector appears should be understood as a consequence of the previous one (as a result, otherwise, therefore, then).
- Causal: emphasizes the cause in a cause-effect relation (for, because).

For each type of semantic relation, a list of discursive connectors in Spanish was extracted from Calsamiglia and Tusón (2012) and Quintero Ramírez (2015). The list of connectors consists of single words (for example "but", "however"), as well as in expressions such as "in the same way." After tokenizing each sentence, we counted how many connectors we identified of each type. The list of connectors we use can be found in Supplementary Table 1. Let us note that connectors may also be implicit, in which case we are unable to account for them, adding uncertainty to this variable. However, implicit connectors are more common in expressive or colloquial language. By contrast, a neutral language style, as we should expect to find in presidential speeches, should use explicit connectors to avoid misunderstandings (Calsamiglia and Tusón, 2012).

Verb - Person. In the discursive approach, the speaker constitutes an essential component of the communicative situation. This vision acknowledges that the enunciation is a bipartite act, that is built between two people. Here we intended to represent the speaker mark with the grammatical person of the main verb (Calsamiglia and Tusón, 2012). We distinguished three alternatives, that cover 93% of our data:

- First person: (i) the singular first person (the "I") suggests that the speaker takes responsibility in the content, and imposes himself on to others; (ii) in the plural first person (the "We") the speaker is part of a collective, and thus the responsibility is diluted. Other uses of the "We" are the majestic plural first person[5] and the inclusive one, which intend to bring the speaker closer to its audience (Calsamiglia and Tusón, 2012).
- Third person: according to Benveniste (1971) and Ricoeur (1990) the grammatical third person is a "non-person,"

---

5   Symbolic and traditional sign of distinction, use of the Pope and the King.

TABLE 3 Measure definitions, based on Lu (2010).

Syntactic structures	
Sentence	A sentence is a group of words delimited with a period, question mark, exclamation mark, quotation mark, or ellipsis.
Clause (C)	A clause is defined as a subject and a finite verb (includes independent clauses, adjective clauses, adverbial clauses, and nominal clauses).
Dependent clause (DC)	Subset of previous categories, and includes finite adjective, adverbial, or nominal clause.
Coordinate phrases (CP)	Comprise (non clausal) adjective phrases, adverb phrases, noun phrases, and verb phrases, connected by a coordinate conjunction
Complex nominals (CN)	Comprise (i) nouns with modifiers (adjective, possessive, prepositional phrase, relative clause, participle, or appositive), (ii) nominal clauses, and (iii) gerunds and infinitives in subject position.
Metrics	
Sentence complexity ratio	Number of C/number of S
Dependent clause ratio	Number of DC/number of C
Coordinate phrases per clause	Number of CP/number of C
Complex nominals per clause	Number of CN/number of C

because it may refer to anything; object, animal, or human. As there is no reference to the protagonist of the enunciation, he is foreign to the referred world. This may imply a more "neutral" worldview, although this neutrality may not correspond to a real objectivity (Calsamiglia and Tusón, 2012).

- Passive voice: use of passive voice or passive grammatical constructions. Sentences in this category are focused on the action, and do not necessarily specify who or what is performing the action. Therefore, it has been used as an indicator of a low sense of responsibility or agency by the speaker (Goñi et al., 2023).

The aforementioned analysis was automatized using the Stanza package for Python, provided by the Stanford Natural Language Processing Group (Toutanova et al., 2003; Chen and Manning, 2014; Manning et al., 2014). As the literature suggests that the use of first person is indicative of populist discourses (Charaudeau, 2009; Oliver and Rahn, 2016), we created an additional variable that recognizes any reference to the first person in the sentence (and not only when the root verb is conjugated in first person). This variable is built by identifying any verb conjugated in the first person and also the "I" and "We" categories of LIWC dictionary, which incorporates relative, possessive and personal pronouns.

Verb - Tense. Tense is a relationship between the "chronological time"—i.e., the time being talked about—and the "linguistic time" or time of speaking (Real Academia Española et al., 1999). A chronological time before, during or after the linguistic time will correspond to past, present and future, respectively. Among simple verb tense we find present, past, imperfect past and future. Compound verb tense include periphrastic future (going to + infinitive) and present perfect (have + participle). Our operationalization distinguishes the values "present", "past" (including regular and imperfect past), "future," "periphrastic future," and "past participle."

4.3.4 Control variables

Average words per sentence. As most of our explanatory variables count elements within a sentence, we have included the average words per sentence as a control.

Type of GPD speech. To control for speech context, we added a categorical variable to account for the four types of speeches (campaign speech, ribbon-cutting speech, international speech, and famous speech).

Country. To control for unobserved effects at the country level, we added a dummy variable per country.

Period. To control for temporal populist waves, we also added a categorical variable to account for time period. Following Conniff (2020), we distinguish three periods: classic populism (until 1990), neopopulism (from 1991 to 2000) and contemporary populism (from 2001 to present). As the GPD covers mainly from 2000 onwards, only 5 speeches of our subset fall into the classic populism category. Therefore, we have replicated our results excluding these 5 speeches, thus using only two categories (see Supplementary Table 7).

Institutional controls. An alternative to control for unobserved effects at a country/time level is to assume that those variables will have a direct effect on the quality of institutions. Thus, we extracted the *Liberal democracy index* (*v2x¡ibdem*) from the Varieties of Democracy (V-Dem) project (Coppedge et al., 2019).

4.4 Model

Due to the three-level categorical nature of the populism score, we tested a multinomial logistic regression and an ordered logistic regression. The proportional odds assumption of parallel regression holds that the relationship between each pair of outcome group is the same. Although a Brant-Wald test was applied in order to assess the proportional odds assumption, and the results showed favorable enough for an ordinal logistic regression or probit methodology, our theoretical interest lies in assessing the differences between the levels of populism themselves. Therefore, we also applied a multinomial logistic regression in order to identify any differences worth noting that may have been overlooked by the ordinal logistic regression. In order to ensure the viability of a multinomial logistic model, a Hausman-McFadden test of IIA assumption was performed. The results show to be favorable for the use of this model. As both statistical approaches gave the same

TABLE 4 Operationalization—Variables incorporated in final models.

Variable type	Dimension	Subdimension	Variable
Dependent	Level of populism	Populism score	GPD average populism score
Control	Lexical Speech Contextual		Av. words per sentence Type of GPD speech Country Period Liberal democracy
Independent	Content Complexity Pragmatic features	Emotion Ingroup Sentence complexity Coordination Particular structures Discursive connectors Verbal features	Positive emotions Negative emotions Ingroup Virtue Sentence complexity ratio Coordinate phrases per clause Complex nominals per clause Conditional connectors Causal connectors Exclusive connectors Future tense Third person

overall results, we opted for the multinomial logistic regression in order to better explore the existing differences between the levels of populism (absent, somewhat present, and completely present). See the Supplementary material section for details of the result for both the Brand-Wald and the Hausman-McFadden tests. Supplementary Tables 8, 9 show alternative formulation using ordinary linear square regression, logistic regression, and ordinal logistic regression.

The independent variables for the model were selected to cover each explanatory dimension, and based on their correlation levels (see the Supplementary material section for details of the result of the correlations between variables). Table 4 shows the variables included in the baseline model.

5 Results

According to Table 5, the Ingroup Virtue variable has a consistent and positive effect on the populism score, even when using only content and control variables (see Supplementary Table 4). This is consistent with the widely accepted view that the populist speech refers to "the people" or, as stated by Jagers and Walgrave (2007): "I listen to you because I talk about you." A good example of this can be seen in the following sentence, extracted from a 2019 speech by Nicolás Maduro[6]:

---

6 The quotes in this section are from speeches in the GPD and were selected by looking for sentences that scored high on the variable of interest. There is not necessarily a link from the sentences themselves to the populism score, because this score is associated to the speech, not to particular sentences.

Confiamos en el sistema electoral, lo dije hoy cuando fui a votar, confío en el pueblo de Venezuela y valió la pena decir esa hermosa expresión, ha valido la pena confiar en el pueblo de Venezuela, seguiremos confiando en el pueblo de Venezuela, sólo con el pueblo podremos construir patria, podremos consolidar la independencia, podremos avanzar en la ofensiva, en la superación de la pobreza. (*We trust the electoral system, I said it today when I went to vote, I trust the people of Venezuela and it was worth it to say that beautiful expression, it has been worth it to trust the people of Venezuela, we will continue to trust the people of Venezuela, only with the people we will be able to build a country, we will be able to consolidate independence, we will be able to advance in overcoming poverty.*)

Along with Ingroup Virtue, the negative emotion variable (Emoneg) consistently shows a positive and significant effect on populism, supporting the hypothesis of the emotional language. The next sentence scores high in negative emotion, and serves as an example:

Yo creo que hay que tener paciencia y llegará el momento, en que ellos mismos, se van a dar cuenta que es un gran error caer en ese tipo de campaña; de estar sembrando mentiras, infamias, calumnias, odio, resentimiento, violencia... así no se construye un país! (*I believe that you have to be patient and the time will come when they themselves will realize that it is a big mistake to fall into that type of campaign; of sowing lies, infamy, slander, hatred, resentment, violence... that's not how a country is built!*) (Ortega, Nicaragua, 2007)

Besides the emotional nature of language, the populist style is supposed to be simple. Given the high levels of correlation among our linguistic complexity variables, we chose three of them. These variables represent different types of complexity. The Syntactic Complexity ratio suggests the use of short and direct sentences, with less clausal constructions. The Complex Nominals accounts for the use of prepositional phrases in the nominal, that is, of a richer construction not at the verb, but at the noun level. Only the Complex Nominals has a negative and significant effect in the high populism category (Model A); that is, the less complex, the more populist is the speech. This effect is not significant in Model B, although it is consistently significant in some alternative versions of Model A (see the Supplementary material). The following quote shows an example of a sentence with low complexity, containing only one clause and no complex nominal:

Ya están construidos los hospitales de Salamanca, Puerto Aysén, Puerto Natales, Porvenir y Puerto Williams. (*The hospitals of Salamanca, Puerto Aysén, Puerto Natales, Porvenir and Puerto Williams have already been built.*) (Bachelet, Chile, 2017)

Although this sentence is relatively short, a longer sentence will not be necessarily more complex: the models control by the average words by sentence, so the count of clauses or complex nominals will be relative to the speech length. On the other hand, a longer sentence may contain more elements, as in the previous example, but not a more complex syntactic structure.

TABLE 5 Multinomial logistic regression.

| | Dependent variable: populism score | | | |
| | Model A | | Model B | |
	Mid	High	Mid	High
Content variables				
Ingroup Virtue	3.398**	7.747***	3.763**	6.468**
	(1.133)	(2.039)	(1.282)	(2.346)
Negative emotion	3.728*	13.024***	4.138*	13.733***
	(1.652)	(3.484)	(1.863)	(3.670)
Positive emotion	1.384	2.228	1.133	0.122
	(0.960)	(2.404)	(1.079)	(2.898)
Complexity				
Sentence complexity ratio	−0.108	−2.971	−0.112	−3.177
	(0.754)	(1.957)	(0.798)	(2.110)
Coordinate phrases per clause	−0.137	7.478	0.764	10.519
	(2.144)	(5.327)	(2.309)	(6.096)
Complex nominals per clause	−1.730	−11.360*	−0.600	−10.147
	(1.719)	(4.905)	(1.903)	(5.354)
Verb variables				
Future tense (total)	7.207**	16.902**	8.332**	17.307*
	(2.615)	(6.033)	(3.191)	(7.252)
Third person	5.634*	14.548**	6.836**	13.284*
	(2.448)	(4.879)	(2.641)	(5.509)
Connectors				
Conditional	7.832	20.190*	9.865*	24.596**
	(4.181)	(8.215)	(4.651)	(9.498)
Causal	−0.678	10.387	−0.693	9.061
	(3.008)	(8.627)	(3.135)	(9.383)
Exclusive	−7.567	−23.171	−9.955	−29.485
	(6.788)	(15.314)	(7.532)	(18.093)
Control variables				
Av. words per sentence	−0.095	−0.104	−0.095	−0.048
	(0.076)	(0.189)	(0.078)	(0.218)
Period (neopopulism)	−12.680***	−13.987***	−11.497***	−13.464***
	(1.757)	(2.264)	(2.187)	(2.811)
Period (contemporary)	−8.333***	−9.278***	−6.259**	−8.745**
	(1.378)	(2.052)	(1.999)	(2.902)
Type of speech - Famous			−1.782**	−2.904**
			(0.597)	(1.101)
Type of speech - International			−1.583*	−2.682
			(0.740)	(1.420)
Type of speech - Ribboncutting			−1.871**	−2.562*
			(0.679)	(1.276)

(Continued)

TABLE 5 (Continued)

| | Dependent variable: populism score | | | |
| | Model A | | Model B | |
	Mid	High	Mid	High
Liberal democracy (V-Dem)			−11.819*	−15.022*
			(4.735)	(6.083)
Country	Yes	Yes	Yes	Yes
Constant	5.580**	4.335	10.337***	12.822**
	(1.916)	(3.643)	(2.756)	(4.849)
Akaike Inf. Crit.	378.302	378.302	372.796	372.796

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

In the initial data exploration, highly populist speeches showed the highest use of the "I." Particularly those by Hugo Chávez, as seen in the following example:

> Yo no voy a tomar el sueldo del presidente como referencia, porque no puedo, no debo, no debo, porque yo no tengo prácticamente gastos, yo no debería tener sueldo más bien. (*I'm not going to take the president's salary as a reference, because I can't, I shouldn't, I shouldn't, because I have practically no expenses, I shouldn't even have a salary.*) (Hugo Chávez, Venezuela, 2007[7])

Yet, although highly populist speeches score high in the use of singular first-person, there are many non-populist speeches that do so as well, as seen in this quote:

> Yo creo que Chile está listo para enfrentar transformaciones que permitirán tener el país moderno que todos queremos. (*I believe that Chile is ready to face transformations that will allow us to have the modern country that we all want.*) (Bachelet, Chile, 2014)

According to our model, the use of singular first-person does not predict populism in speeches. In Table 5, this is shown in the positive and significant effect of the Third Person variable, which represent the person of the root verb and is the alternative to the first person. Given the enormous variability of subjects in sentences, it is difficult to specify who the third person refers to. However, some words emerge repeatedly—such as the subject "people" or the name of the country as a subject, the nation, or elements like inequality, employment, children, etc. An example of this can be seen in the following sentence:

> Esa es la vocación democrática del pueblo boliviano. (*That is the democratic vocation of the Bolivian people.*) (Morales, Bolivia, 2006)

To verify the third person effect, we replace this variable with singular and plural first person variables, and with additional first person variables that include pronouns (see Supplementary Table 6). Our result are robust to different person specification and to alternative models. On the other hand, the use of passive voice was not tested in the model due to data imbalance.

As for other pragmatic variables, there is a positive and significant effect of the future tense on the populist score. This may aim to get the future close to the present scenario, increasing the reality and importance of future events (see Supplementary Table 5 for alternative models on verb tense).

> Pronto comenzaremos a ver los resultados y esos resultados serán mucho más elocuentes que cualquier discurso porque son los que nos ponen en el camino de una Argentina de pie y en paz. (*Soon we will begin to see the results and those results will be much more eloquent than any speech because they are what put us on the path to an Argentina standing and at peace.*) (Duhalde, Argentina, 2002)

Regarding the discursive connectors, only the conditional connectors show a significant effect in the high populism category (Model A). When controlling by all contextual factors (Model B), the conditional connectors increase its significance in all populism categories. The conditional connectors reinforce the effect of future events being more real, as conditioned realities are perceived as more credible.

> Por otra parte, ya hemos hecho del conocimiento de la nación que el gobierno está preparado para limitar sus presupuestos cuando las condiciones económicas así lo reclamen. (*On the other hand, we have already made the nation aware that the government is prepared to limit its budgets when economic conditions demand it.*) (Cárdenas, México, 1935)

---

7 http://www.todochavez.gob.ve/todochavez/2705-intervencion-del-comandante-presidente-hugo-chavez-durante-acto-de-juramentacion-como-presidente-de-la-republica-bolivariana-de-venezuela-para-el-periodo-2007-2013, accessed on June 2024.

6 Conclusions

In this paper we have studied the linguistic-discursive dimension of populism, focusing on left-wing populist leaders in

Latin America. Our results show a positive effect of ingroup and emotional content on the degree of populism in political speech. We also find a positive effect of the future tense and conditional connectors, which suggest an intention to manipulate the audience, and a negative or null effect of the use of first person.

These results have implications both for the current understanding of (left-wing) populist rhetoric and for the conceptualization of populism itself. In the first aspect, our work intersects with critical discourse theory, which helps us to interpret the results. For example, Van Dijk (1994), Wodak and van Dijk (2000), and KhosraviNik (2010) describe the phenomenon of positive-self and negative-other representations, to show how manipulation operates. For the discursive manipulation to succeed, the negative representation of the "other" and/or the positive self-representation of the "ego" must be perceived as epistemically strong, while the critical reasons necessary to question them remain absent. On the other hand, the positive and significant effect of the Ingroup Virtue variable on populism relates to this phenomenon, and reminds us about the double-constraint nature of manipulation (hide critical content and highlight the favorable information). In this argumentative line, the use of the future tense could be understood, in turn, as another strategy (among several others) to increase the epistemic force of the content, since it establishes the future as the present scenario, bringing the time of the events closer, figuratively speaking. Conditional connectors also reinforce this effect, as conditioned realities are perceived as more credible. Future tense could also contain an element of prediction or modality (Lyons, 1997).

A popular hypothesis is that self-reference is a distinctive characteristic of populist leaders. Our results on left-wing populism tentatively challenge this deep-rooted belief. The field of language psychology can help us understand this situation. It has been documented that, contrary to a widely established belief, the linguistic behavior of the narcissistic personality does not necessarily correlate with the overt use of the first person (Ireland and Mehl, 2014; Holtzman et al., 2019). In fact, narcissistic linguistic behavior is more related to the avoidance of tentative words (maybe, probably), and the tendency toward dominance (Cheng et al., 2010) and exploitation (Raskin and Terry, 1988). If charismatic populist leaders would exhibit narcissistic personality traits, their speeches should then reflect other main characteristics, beyond the discarded positive correlation with the use of the first person, such as the certainty in future events, the tendency to manifest superiority and the manipulation of the audience's emotions. Though the relationship between populism and narcissism posits an interesting question, establishing this relationship requires instruments and methods specific to the field of personality psychology.

The positive effect of the use of the third person on populism remains to be explained. At this point, we can only formulate hypotheses and suggest future studies. A random reading of 30 sentences in the third person, extracted from the speeches we study, allows us to venture a hypothesis related to speech acts.[8] A classification of speech acts distinguishes between assertive, compromising, directive, declarative and expressive acts,

depending on their intention (Searle, 1969). Of the 30 statements analyzed, only 5 assertive statements are intended to describe realities and can be subject to verification. The remaining 25 are statements in which their truth depends exclusively on the person who utters it. That is, they depend on whether or not the sender of the speech has the intention of doing what he says, whether or not he believes in what he is stating, whether or not he intends to fulfill his commitments, etc. Following the definition of a macro speech act by Van Dijk (2013), we believe that the dominant macro speech act of these political discourses is declarative. That is, these are discourses that seek to create realities, decree or define the course of events.[9] Thus, the shift from the first to the third person may be another strategy of the speaker to hide the agency of the actions and the responsibility of the commitments.

In sum, our linguistic results suggests that the mechanism at the base of populist discourse, specifically, left-wing populist discourse, is manipulation. By generating a strong identity effect on a limited set of ideas, it weakens the critical surveillance of the audience (Sperber and Wilson, 1995). On the other hand, the systematic and intensive use of discursive strategies aim to create realities that consolidate a predetermined vision (coinciding with the leader's objectives) as well as to configure a predisposition or propensity to action that coincides with that predetermined vision.

Beyond the linguistic analysis, our results help shed light on the conceptualization of populism itself. Although they are conceived as two different approaches, the ideational and strategic approaches converge in our model. Recall that the dependent variable is extracted from GPD, whose coding strategy responds to the ideational approach. However, and together the content variables that go along with the ideational approach, we have incorporated pragmatic variables, which point to the speaker's intention. Relying on critical discourse theory, our results suggest an intention of manipulation by the populist leader. This is consistent with the politic-strategic approach of populism, which conceives it as an instrument for opportunistic leaders to reach and remain in power. In this sense, our results show that both approaches, in fact, coexist in the texts we have studied. Our work has delved into the characterization of populism within political speeches, discerning linguistic attributes across contexts. This examination does not negate the significance of context; in fact, it underscores its importance. As we study the "how" of populism we are better able to shed light on a topic that retains relevance in the broader sphere of comparative politics.

In closing it is important to note that we believe that the framework we have developed can also be extended to other discourses. We have proposed a novel methodological way to characterize the discursive features of populism utilizing NLP methods. If the foundation of populism lies within leaders'

8　Available upon request.

9　In other contexts, the frequency of assertive speech acts is much higher (see for example, Vaezi et al., 2014; Pagmar, 2016; Asayesh et al., 2020), since the speaker's intention is to collaborate in order to communicate, trying to provide the receiver with true, relevant and clear information. Instead, populist discourse appears to be fundamentally persuasive. The sender's objective is to capture the support of his interlocutor through strategies such as manipulation.

linguistic attributes, it becomes feasible to assess its impact on other types of political texts. Moreover, this would allow researchers to examine other effects at play, such as the contagion of populist style and how it extends to mass media.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: the Global Populism Database, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/LFTQEZ.

Author contributions

MR: Conceptualization, Formal analysis, Methodology, Writing – original draft. CF-B: Methodology, Writing – review & editing. MF: Conceptualization, Writing – original draft. JC: Conceptualization, Writing – original draft. MS: Formal analysis, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpos.2024.1435712/full#supplementary-material

References

Aguilar, R., and Carlin, R. E. (2017). Ideational populism in Chile? A case study. *Swiss Polit. Sci. Rev.* 23, 404–422. doi: 10.1111/spsr.12268

Albertazzi, D., and McDonnell, D. (2015). *Populists in Power*. London: Routledge. doi: 10.4324/9781315725789

Asayesh, Z., Tabatabaee Lotfi, S. A., and Sarkeshikian, S. A. H. (2020). Evaluating speech acts in English language teaching software: the case of Tell me More. *J. Mod. Res. English Lang. Stud.* 7, 29–50. doi: 10.30479/jmrels.2020.11660.1451

Basset, Y., and Launay, S. (2013). "A latin-american polemic: populism," in *Contemporary Populism: A Controversial Concept and Its Diverse Forms*, eds. S. Gherghina, S. Mişcoiu, S. Soare (Cambridge: Cambridge Scholars Publishing), 140–166.

Benveniste, E. (1971). *Problems in General Linguistics*. Oxford: University of Miami Press.

Betz, H. G. (2002). "Conditions favouring the success and failure of radical right-wing populist parties in contemporary democracies," in *Democracies and the populist challenge*, eds. Y. Mny, and Y. Surel (London: Palgrave Macmillan), 197–213. doi: 10.1057/9781403920072_11

Bonikowski, B., and Gidron, N. (2016). The populist style in American politics: presidential campaign discourse, 1952-1996. *Soc. Forces* 94, 1593–1621. doi: 10.1093/sf/sov120

Bonikowski, B. (2016). Three lessons of contemporary populism in Europe and the United States. *Brown J. World Affairs* 23, 9–24.

Calsamiglia, H., and Tusón, A. (2012). *Las cosas del decir*. Barcelona: Ariel.

Cammack, P. (2000). The resurgence of populism in Latin America. *Bull. Lat. Am. Res.* 19, 149–161. doi: 10.1111/j.1470-9856.2000.tb00096.x

Canovan, M. (2004). Populism for political theorists? *J. Polit. Ideol.* 9, 241–252. doi: 10.1080/1356931042000263500

Charaudeau, P. (2009). *Reflexiones para el análisis del discurso populista*. Revue Discurso y Sociedad. Available online at: http://www.patrick-charaudeau.com/Reflexiones-para-el-analisis-del.html (accessed January 3, 2024).

Chen, D., and Manning, C. D. (2014). "A fast and accurate dependency parser using neural networks," in *Proceedings of the 2014 Conference on empirical methods in natural language processing (EMNLP)*, eds. A. Moschitti, B. Pang, W. Daelemans (Kerrville: Association for Computational Linguistics), 740–750. doi: 10.3115/v1/D14-1082

Cheng, J. T., Tracy, J. L., and Henrich, J. (2010). Pride, personality, and the evolutionary foundations of human social status. *Evol. Hum. Behav.* 31, 334–347. doi: 10.1016/j.evolhumbehav.2010.02.004

Conniff, M. L. (2020). A historiography of populism and neopopulism in Latin America. *History Compass* 18:e12621. doi: 10.1111/hic3.12621

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Marquardt, K. L., et al. (2019). *V-dem methodology v9*. Varieties of Democracy (V-Dem) Project. doi: 10.2139/ssrn.3441063

De Cleen, B., and Glynos, J. (2021). Beyond populism studies. *J. Lang. Polit.* 20, 178–195. doi: 10.1075/jlp.20044.dec

De la Torre, C. (2007). The resurgence of radical populism in Latin America. *Constellations* 14, 384–397. doi: 10.1111/j.1467-8675.2007.00453.x

De la Torre, C. (2017). *Populism in Latin America*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780198803560.013.8

De la Torre, C., and Arnson, C. (2013). "Introduction: The evolution of Latin American populism and the debates over its meaning," in *Latin American populism in the twenty-first century*, eds. C. de la Torre and C. Arnson (Washington, DC: Woodrow Wilson Center Press), 1–36.

Drămnescu, M. (2014). "Communication. Political speech. Populist speech," in *Topical Communication Uncertainties*, eds. S. Vlăduescu, F. Smarandache, D. Gifu, and A. Ěšenescu (ZIP Publishing.), 242–256.

Fairclough, N. (2013). *Critical Discourse Analysis: The Critical Study of Language*. London: Routledge. doi: 10.4324/9781315834368

Gidron, N., and Bonikowski, B. (2013). *Varieties of populism: Literature review and research agenda*. Working Paper Series, Weatherhead Center for International Affairs, Harvard University. doi: 10.2139/ssrn.2459387

Goi, I., Raveau, M. P., and Fuentes-Bravo, C. (2023). Analytical categories to describe imaginations about the collective futures: From theory to linguistics to computational analysis. *Futures* 156:103324. doi: 10.1016/j.futures.2024.103324

Gutiérrez Araus, M. L. (2015). "El enunciador en el discurso oral y el verbo español," in *Aspectos de la subjetividad en el lenguaje: Curso sobre "Lo subjetivo en el lenguaje", celebrado en Zaragoza, en el marco de la Cátedra "María Moliner" de la Institución Fernando el Católico (IFC), en noviembre de 2011* (Institución Fernando el Católico), 133–152.

Haidt, J., Graham, J., and Joseph, C. (2009). Above and below left-right: Ideological narratives and moral foundations. *Psychol. Inq.* 20, 110–119. doi: 10.1080/10478400903028573

Halliday, M., and Matthiessen, C. (2004). *An Introduction to Functional Grammar*. London: Hodder Arnold.

Hawkins, K. A. (2018). "The ideational approach," in *Routledge Handbook of Global Populism*, ed. C. De la Torre (London: Routledge), 57–71. doi: 10.4324/9781315196923

Hawkins, K. A., Aguilar, R., Silva, B. C., Jenne, E. K., Kocijan, B., and Rovira-Kaltwasser, C. (2019). "Measuring populist discourse: the global populism database," in *EPSA Annual Conference in Belfast, UK, June*, 20–22.

Hawkins, K. A., and Rovira Kaltwasser, C. (2018). Measuring populist discourse in the United States and beyond. *Nat. Hum. Behav.* 2, 241–242. doi: 10.1038/s41562-018-0339-y

Holtzman, N. S., Tackman, A. M., Carey, A. L., Brucks, M. S., Küfner, A. C., Deters, F. G., et al. (2019). Linguistic markers of grandiose narcissism: A LIWC analysis of 15 samples. *J. Lang. Soc. Psychol.* 38, 773–786. doi: 10.1177/0261927X19871084

Houle, C., and Kenny, P. D. (2018). The political and economic consequences of populist rule in Latin America. *Govern. Opposite.* 53, 256–287. doi: 10.1017/gov.2016.25

Ireland, M. E., and Mehl, M. R. (2014). "Natural language use as a marker," in *The Oxford Handbook of Language and Social Psychology*, ed. T. M. Holtgraves (Oxford: Oxford University Press), 201–237.

Jäger, A. (2023). Populism and the historians: Richard Hofstadter and the birth of a global populism debate. *Hist. Polit. Thought* 44, 153–194.

Jagers, J., and Walgrave, S. (2007). Populism as political communication style: an empirical study of political parties' discourse in Belgium. *Eur. J. Polit. Res.* 46, 319–345. doi: 10.1111/j.1475-6765.2006.00690.x

Kestler, T. (2022). Radical, nativist, authoritarian or all of these? Assessing recent cases of right-wing populism in Latin America. *J. Polit. Latin Am.* 14, 289–310. doi: 10.1177/1866802X221117565

KhosraviNik, M. (2010). The representation of refugees, asylum seekers and immigrants in British newspapers: a critical discourse analysis. *J. Lang. Polit.* 9, 1–28. doi: 10.1075/jlp.9.1.01kho

Klein, D., and Manning, C. D. (2002). "Fast exact inference with a factored model for natural language parsing," in *Advances in Neural Information Processing Systems*, 15.

Knight, A. (1998). Populism and neo-populism in Latin America, especially Mexico. *J. Lat. Am. Stud.* 30, 223–248. doi: 10.1017/S0022216X98005033

Laclau, E. (2005). *On Populist Reason*. New York: Verso.

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quart.* 12, 439–448. doi: 10.2307/3586142

Larsen-Freeman, D., and Strom, V. (1977). The construction of a second language acquisition index of development. *Lang. Learn.* 27, 123–134. doi: 10.1111/j.1467-1770.1977.tb00296.x

Lin, N., and Osnabrügge, M. (2018). Making comprehensible speeches when your constituents need it. *Res. Polit.* 5:2053168018795598. doi: 10.1177/2053168018795598

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *Int. J. Corpus Linguist.* 15, 474–496. doi: 10.1075/ijcl.15.4.02lu

Lyons, J. (1997). *Semántica lingüística: una introducción*. Paoli: Paidos.

Maillat, D., and Oswald, S. (2009). Defining manipulative discourse: the pragmatics of cognitive illusions. *Int. Rev. Pragm.* 1, 348–370. doi: 10.1163/187730909X12535267111651

Maillat, D., and Oswald, S. (2011). "Constraining context: a pragmatic account of cognitive manipulation," in *Critical discourse studies in context and cognition*, ed. C. Hart (Amsterdam: John Benjamins Publishing Company). doi: 10.1075/dapsac.43.04mai

Malvern, D., Richards, B., Chipere, N., and Durán, P. (2004). *Lexical Diversity and Language Development*. Cham: Springer. doi: 10.1057/9780230511804

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., and McClosky, D. (2014). "The stanford CoreNLP natural language processing toolkit," in

Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, eds. K. Bontcheva, and J. Zhu (Kerrville: Association for Computational Linguistics), 55–60. doi: 10.3115/v1/P14-5010

Matsuo, A., Sasahara, K., Taguchi, Y., and Karasawa, M. (2019). Development and validation of the Japanese moral foundations dictionary. *PLoS ONE* 14:e0213343. doi: 10.1371/journal.pone.0213343

Meijers, M. J., and Zaslove, A. (2021). Measuring populism in political parties: appraisal of a new approach. *Comp. Polit. Stud.* 54, 372–407. doi: 10.1177/0010414020938081

Miscoiu, S., Craciun, O., and Colopelnic, N. (2008). *Radicalism, Populism, Interventionism. Three Approaches Based on Discourse Theory*. Cluj-Napoca: EFES.

Montolío, E. (2001). *Conectores de la lengua escrita*. Barcelona: Ariel.

Morris, C. W. (1938). "Foundations of the theory of signs," in *International Encyclopedia of Unified Science*, eds. O. Neurath, R. Carnap, and C. F. W. Morris (Chicago: Chicago University Press), 1–59.

Mouffe, C. (2005). "The "end of politics" and the challenge of right-wing populism," in *Populism and the mirror of democracy London*, ed. F. Panizza (New York: Verso), 72–98.

Mudde, C. (2004). The populist zeitgeist. *Govern. Opposit.* 39, 541–563. doi: 10.1111/j.1477-7053.2004.00135.x

Mudde, C. (2017). "Populism: an ideational approach," in *The Oxford Handbook of Populism*, eds. C. R. Kaltwasser, P. Taggart, P. O. Espejo, and P. Ostiguy (Oxford: Oxford University Press), 27–47. doi: 10.1093/oxfordhb/9780198803560.013.1

Mudde, C. (2019). *The Far Right Today*. New York: John Wiley Sons.

Mudde, C., and Rovira Kaltwasser, C. (2017). *Populism: A Very Short Introduction*. Oxford: Oxford University Press. doi: 10.1093/actrade/9780190234874.001.0001

Mudde, C., and Rovira Kaltwasser, C. (2018). Studying populism in comparative perspective: reflections on the contemporary and future research agenda. *Comp. Polit. Stud.* 51, 1667–1693. doi: 10.1177/0010414018789490

Olivas Osuna, J. J. (2021). From chasing populists to deconstructing populism: a new multidimensional approach to understanding and comparing populism. *Eur. J. Polit. Res.* 60, 829–853. doi: 10.1111/1475-6765.12428

Oliver, J. E., and Rahn, W. M. (2016). Rise of the Trumpenvolk: populism in the 2016 election. *Ann. Am. Acad. Pol. Soc. Sci.* 667, 189–206. doi: 10.1177/0002716216662639

Ortega, L. (2003). Syntactic complexity measures and their relationship to l2 proficiency: a research synthesis of college-level l2 writing. *Appl. Ling.* 24, 492–518. doi: 10.1093/applin/24.4.492

Oswald, S. (2014). "It is easy to miss something you are not looking for a pragmatic account of covert communicative influence for (critical) discourse analysis," in *Contemporary Critical Discourse Studies*, eds. C. Hart, and P. Cap (London: Bloomsbury Publishing), 97–120.

Oswald, S., and Maillat, D. (2013). Biases and constraints in communication: argumentation, persuasion and manipulation. *J. Pragmat.* 59:137–140. doi: 10.1016/j.pragma.2013.08.014

Pagmar, D. (2016). *Frequency, form, and distribution of illocutionary speech acts in Swedish parent-child interaction (dissertation)*. Master's thesis, Stockholm University, Faculty of Humanities, Department of Linguistics.

Pennebaker, J. W., Boyd, R. L., Jordan, K., and Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Technical report, University of Texas at Austin.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., and Manning, C. D. (2020). "Stanza: a Python Natural Language Processing toolkit for many human languages," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, eds. A. Celikyilmaz, and T.-H. Wen (Kerrville: Association for Computational Linguistics). doi: 10.18653/v1/2020.acl-demos.14

Quintero Ramírez, S. (2015). Identificación de los conectores discursivos de más alta frecuencia en notas periodísticas deportivas. *RLA* 53, 47–71. doi: 10.4067/S0718-48832015000200003

Rabotnikof, N. (2018). Populismo: conceptos, vocabularios y experiencias. *Diánoia* 63, 3–22. doi: 10.22201/iifs.18704913e.2018.81.1571

Ramírez-Esparza, N., Pennebaker, J. W., García, F. A., and Suriá, R. (2007). La psicología del uso de las palabras: un programa de computadora que analiza textos en español. *Rev Mexicana Psicol.* 24, 85–99.

Raskin, R., and Terry, H. (1988). A principal-components analysis of the Narcissistic Personality Inventory and further evidence of its construct validity. *J. Pers. Soc. Psychol.* 54:890. doi: 10.1037//0022-3514.54.5.890

Real Academia Española, Bosque, I., Demonte, V., Carreter, F. L., and Lucero, M. V. P. (1999). *Gramática Descriptiva de la lengua española, volume 1*. Madrid: Espasa.

Ricoeur, P. (1990). *Sí mismo como otro*. Ciudad de México: Siglo Veintiuno Editores.

Rueda, D. (2021). Is populism a political strategy? A critique of an enduring approach. *Polit. Stud.* 69, 167–184. doi: 10.1177/0032321720 962355

Ruth, S. P., and Hawkins, K. A. (2017). "Populism and democratic representation in Latin America," in *Political Populism*, eds. O. Mazzoleni, C. Holtz-Bacha, and R. C. Heinisch, 255–274. doi: 10.5771/9783845271491-255

Rydgren, J. (2018). *The Oxford Handbook of the Radical Right*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780190274559.001.0001

Schoonvelde, M., Brosius, A., Schumacher, G., and Bakker, B. N. (2019). Liberals lecture, conservatives communicate: Analyzing complexity and ideology in 381,609 political speeches. *PLoS ONE* 14:e0208450. doi: 10.1371/journal.pone.0208450

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139173438

Sperber, D., and Wilson, D. (1995). *Relevance: Communication and Cognition*. Hoboken: Blackwell Publishing.

Spirling, A. (2016). Democratization and linguistic complexity: the effect of franchise extension on parliamentary discourse, 1832-1915. *J. Polit.* 78, 120–136. doi: 10.1086/683612

Stavrakakis, Y. (2017). Discourse theory in populism research: three challenges and a dilemma. *J. Lang. Polit.* 16, 523–534. doi: 10.1075/jlp.17025.sta

Taggart, P. (2004). Populism and representative politics in contemporary Europe. *J. Polit. Ideol.* 9, 269–288. doi: 10.1080/1356931042000263528

Touraine, A., and Armiño, M. (1989). *América Latina: política y sociedad*. Madrid: Espasa-Calpe Madrid.

Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 conference of the North American chapter of the association for computational linguistics on human language technology-volume 1* (Kerrville: Association for Computational Linguistics), 173–180. doi: 10.3115/1073445.1073478

Urbinati, N. (2019). Political theory of populism. *Ann. Rev. Polit. Sci.* 22, 111–127. doi: 10.1146/annurev-polisci-050317-070753

Vaezi, R., Tabatabaei, S., and Bakhtiarvand, M. (2014). A comparative study of speech-acts in the textbooks by native and non-native speakers: a pragmatic analysis of New Interchange Series vs. locally-made EFL textbooks. *Theory Pract. Lang. Stud.* 4:167. doi: 10.4304/tpls.4.1.167-180

Van Dijk, T. A. (1994). Discourse and inequality. *Lenguas Modernas* 10, 19–37.

Van Dijk, T. A. (2013). "Ideology and discourse analysis," in *The Meaning of Ideology* (London: Routledge), 110–135. doi: 10.1093/oxfordhb/9780199585977.013.007

Weyland, K. (2001). Clarifying a contested concept: populism in the study of Latin American politics. *Comp. Polit.* 34, 1–22. doi: 10.2307/422412

Weyland, K. (2003). Neopopulism and neoliberalism in Latin America: how much affinity? *Third World Q.* 24, 1095–1115. doi: 10.1080/01436590310001630080

Wilson, D., and Sperber, D. (2012). *Meaning and Relevance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139028370

Wodak, R., and van Dijk, T. (2000). *Racism at the Top: Parliamentary Discourses on Ethnic Issues in Six European States*. Klagenfurt: Drava Verlag.

Wolfe-Quintero, K., Inagaki, S., and Kim, H.-Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexiyt*. Honolulu: University of Hawai'i Press.