Quality & Quantity
https://doi.org/10.1007/s11135-026-03038-9

How creative are communication firms? An empirical test of Zipf's Law in Spanish advertising and public relations sectors

Jessica Fernández-Vázquez[1] · Brenda Vázquez-La Hoz[1] · Aida Galiano[1] · Juan Manuel Martín-Álvarez[1]

Received: 6 April 2026 / Accepted: 30 July 2026

Abstract
Company names constitute one of the earliest and most enduring forms of corporate communication. Before organisations engage in formal messaging or stakeholder interaction, their legal names function as symbolic signals of identity and differentiation. Prior research has treated naming mainly as an individual strategic or branding decision, paying limited attention to system-level communicative patterns across industries. This study examines corporate naming as a system-level communicative phenomenon and examines symbolic differentiation in Spain's advertising, public relations, and communication sectors. The analysis covers 26,432 firms founded between 1975 and 2023. Names are analysed using Zipf's Law, with lexical concentration indicating communicative saturation or differentiation. Temporal regimes are identified through structural break analysis, and Zipfian coefficients are estimated across periods and regions. Results show consistently low lexical concentration, consistent with a relatively differentiated naming repertoire. Regional variation is observed in the estimated coefficients. Equal-sized Monte Carlo subsampling shows that unequal sample sizes account for part of the original regional dispersion, but pooled interaction models and pairwise comparisons indicate that residual regional heterogeneity remains after controlling for sample size. Lexical concentration also follows a non-linear temporal trajectory, increasing during the expansion years preceding the 2008 financial crisis and declining thereafter. The regional evidence is therefore interpreted as statistically supported but moderate heterogeneity rather than as an invariant ranking of regional naming systems, whereas the temporal patterns remain descriptive.

Keywords Zipf's law · Corporate naming · Lexical differentiation · Advertising · Public relations · Communication firms

1 Introduction

The practice of naming in commercial and organizational contexts has long been recognized as a fundamental act of communication. A firm's name constitutes one of its most visible symbolic assets, playing a central role in shaping brand identity, conveying intended associ-

Extended author information available on the last page of the article

Published online: 12 August 2026

ations, and facilitating stakeholder recognition (Chan and Huang 1997; Lowrey et al. 2003). As a communicative device, a corporate name serves not only to identify an organization but also to signal its values, positioning, and intended market image. For many stakeholders, it represents the first point of contact with an organization and therefore contributes to the formation of perceptions regarding legitimacy, professionalism, and distinctiveness.

From a marketing perspective, naming has traditionally been viewed as a strategic decision aimed at influencing consumer perceptions and supporting brand-building processes (Ahn and La Ferle 2008). Previous research has examined how names shape evaluations of products and organizations through their phonetic characteristics, semantic associations, memorability, and cultural connotations (Leong et al. 1996; Baker et al. 2004; Yan et al. 2010; Johnson 1980; Alserhan and Alserhan 2012). This body of work has consistently demonstrated that naming decisions can affect recognition, recall, credibility, and overall brand effectiveness.

In sectors such as advertising, public relations, communication, and market research, however, the distinction between corporate identity and brand identity is often less clear-cut than in other industries. In these contexts, the legal name of the company may itself function as a quasi-brand, operating simultaneously as a formal organizational identifier and as a symbolic representation of the services, values, and creative capabilities associated with the firm (Usunier and Shaner 2002; Huang and Chan 1997; Chung and Eoh 2019). Consequently, naming becomes more than a simple administrative requirement; it becomes a strategic communicative resource through which organizations seek visibility, differentiation, and symbolic positioning within increasingly competitive markets.

This communicative dimension is particularly relevant in industries whose activities revolve around the production and management of symbolic content. Firms operating in advertising, public relations, and communication services compete not only through tangible resources but also through reputation, symbolic differentiation, and other intangible resources. As a result, company names frequently serve as signals of distinctiveness, cultural embeddedness, and professional identity (Tolochko et al. 2024; Schmeltz and and Kjeldsen 2016; McDevitt 2014). Understanding how such organizations name themselves therefore provides valuable insights into broader processes of organizational communication and symbolic differentiation.

In recent years, researchers have increasingly examined naming practices not only as isolated managerial choices but also as products of broader sociocultural, institutional, and regional forces (Shipley et al. 1988; Buitrago et al. 2019; Martínez and Muñoz 2014; González-del Río et al. 2011). Particularly in communication-intensive and symbolic-service sectors, names may carry symbolic capital and reflect regional innovation patterns or localized market identities (Fox 2011; Heredero Díaz and Chaves Martín, 2015; Gutiérrez-Chico et al. 2024). This shift in focus opens the door to understanding naming behaviours through aggregate patterns across geographies, industries, and historical periods rather than as isolated branding decisions.

This broader perspective also connects naming practices to fundamental organizational tensions between legitimacy and differentiation. From an institutional perspective, organizational names are not merely labels but symbolic representations of organizational identity that communicate conformity with, or deviation from, prevailing expectations within a field. Glynn and Abzug (2002) argue that organizational naming reflects processes of symbolic isomorphism through which organizations seek legitimacy while simultaneously expressing distinctive identities. Corporate names therefore constitute visible signals through which firms position themselves within institutional environments.

This tension has been further developed in the literature on optimal distinctiveness. Organizations benefit from conforming to established categories because similarity enhances legitimacy, recognition, and stakeholder acceptance. At the same time, excessive similarity may reduce visibility and limit differentiation from competitors. Consequently, organizations face the challenge of balancing pressures to fit in with pressures to stand out (Deephouse 1999; Zhao et al. 2017; Yang and Li 2025). Recent evidence suggests that company names can function as signals of conformity or distinctiveness, with organizational outcomes depending on how such naming choices position firms relative to prevailing norms and peer identities (Amore et al. 2024). Naming practices therefore represent one of the most immediate and visible mechanisms through which this balance can be negotiated.

From this perspective, the lexical structure of corporate naming systems provides an indirect indicator of how organizations collectively navigate the trade-off between legitimacy and distinctiveness. Highly concentrated naming repertoires may suggest stronger tendencies toward symbolic convergence, whereas more differentiated repertoires may reflect greater efforts to establish unique identities. Examining naming patterns at the population level therefore contributes not only to communication and branding research but also to broader debates concerning institutional conformity, symbolic differentiation, and optimal distinctiveness.

Furthermore, recent contributions have expanded the interdisciplinary scope of naming studies by connecting them to broader cultural, institutional, and communicative frameworks. Scholars such as Costa (2012) and Ferrari et al. (2020) have emphasized the symbolic and performative dimensions of corporate discourse, demonstrating how names participate in the constitution of professional identities and market imaginaries. Meanwhile, Gürtler and Miller (2022) examined naming within the creative industries, foregrounding the role of affect, aesthetic judgment, and symbolic differentiation. Matus-Mendoza (2025) contributed a sociolinguistic lens by tracing the indexicality of names across transnational contexts, while Maza-Maza et al. (2020) explored the intersection of brand strategies and regional embeddedness in organizational communication.

In the domain of institutional legitimacy and identity, Olivares-Delgado et al. (2016) showed how naming reflects tensions between market logic and local identity, a perspective echoed by Pinillo-Laffón (2015) in his analysis of symbolic resources in entrepreneurial ecosystems. Complementary to this, Pinillos-Laffón et al. (2016) investigated how narrative structures and linguistic choices in company names affect legitimacy and recognition, especially in culturally saturated or competitive fields. Salazar (2020) added further nuance by linking naming strategies to processes of digital entrepreneurship and brand storytelling, while Shamsollahi et al. (2014) provided an empirical framework for measuring name distinctiveness through semantic clustering and lexical distance metrics. Collectively, these studies highlight the relevance of naming for communication, branding, and organizational research while also supporting the need for empirical approaches capable of capturing variation in naming practices across contexts.

Despite these advances, existing studies have predominantly examined corporate names from three perspectives. First, many studies analyze how specific naming characteristics—such as phonetics, semantics, memorability, or linguistic structure—influence consumer perceptions, brand recall, or organizational legitimacy (e.g., Leong et al. 1996; Baker et al.

2004; Lowrey et al. 2003; Shamsollahi et al. 2014). Second, other contributions focus on naming strategies adopted by particular firms, sectors, or family businesses (e.g., Fox 2011; Pinillos-Laffón 2015; Pinillos-Laffón et al. 2016). Third, recent studies have explored the symbolic, cultural, and institutional dimensions of naming practices through qualitative or case-based approaches (e.g., Costa 2012; Ferrari et al. 2020; Schmeltz and Kjeldsen 2016).

However, these studies rarely examine naming systems at the level of the entire organizational population. Consequently, we still know little about how lexical resources are distributed across all firms within a sector, whether naming systems exhibit systematic patterns of concentration or differentiation, or how such patterns evolve across territories and historical periods. To our knowledge, no previous study has combined a population-level dataset with Zipfian analysis in the context of a national communication sector.

This study addresses that gap by examining the full population of 26,432 firms founded in Spain's advertising, public relations, and communication sectors between 1975 and 2023. In this study, population-level refers to the analysis of the complete population of firms founded in a given sector and country during the observation period. Rather than examining individual company names, selected brands, or limited samples of organizations, we analyze the entire population of firms in the sector. This approach allows us to investigate the aggregate structure of the naming system itself, identifying patterns of lexical concentration and differentiation across regions and historical periods.

In this article, we conceptualise company names not merely as linguistic artefacts or branding devices, but as foundational acts of corporate communication. A firm's legal name is often the first communicative signal it sends to external publics, preceding marketing campaigns, strategic messaging, or formal stakeholder engagement. As such, naming practices contribute to shaping the symbolic environment in which organisations seek legitimacy, recognition, and differentiation. Examining corporate names at an aggregated level therefore allows us to move beyond individual naming decisions and analyse how communicative repertoires are structured across organisations, regions, and historical periods.

To address this gap, we draw on Zipf's Law, one of the most widely documented statistical regularities in language (Zipf 1945). Zipfian distributions describe the relationship between the frequency of a word and its rank, typically revealing that a small number of words account for a large proportion of all occurrences, while most words appear only rarely. When applied to collections of company names, Zipfian analysis provides a systematic way of assessing lexical concentration and symbolic differentiation. Highly concentrated distributions indicate reliance on a limited set of recurrent lexical elements, whereas flatter distributions suggest a broader and more differentiated symbolic repertoire. In this study, symbolic differentiation refers to the degree of variation in the lexical resources used by organizations when constructing their corporate names. Higher symbolic differentiation indicates a broader and less concentrated naming repertoire, whereas lower symbolic differentiation reflects greater reliance on a limited set of recurrent lexical elements.

This approach has been validated in multiple linguistic contexts (Sicilia-García et al. 2002; Sigurd et al. 2004) and has subsequently been extended to explain communicative efficiency and system-level optimization (Piantadosi 2014; Thurner et al. 2015). Corral et al. (2015) further highlight its applicability beyond natural language, including urban and organizational systems. These findings support the use of Zipfian metrics as a tool for quantifying lexical concentration and symbolic differentiation in naming systems across time and space.

1.1 Lexical differentiation in communication industries

Previous research suggests that firms operating in communication-intensive sectors rely heavily on symbolic resources, reputation, and differentiation (Schmeltz and Kjeldsen 2016; Gürtler and Miller 2022; McDevitt 2014). Because naming constitutes an important communicative resource in these industries, firms are expected to avoid excessive lexical concentration and to develop relatively differentiated naming repertoires.

H1 *Firm names in the communication and advertising sectors display relatively low lexical concentration, reflecting a relatively differentiated naming repertoire.*

1.2 Regional variation in naming systems

Previous studies have shown that naming practices are influenced by sociocultural, institutional, and regional contexts (Shipley et al. 1988; Buitrago et al. 2019; Martínez and Muñoz 2014). Research has also suggested that names may reflect regional identities and local patterns of innovation (Fox 2011; Heredero Díaz and Chaves Martín, 2015; Gutiérrez-Chico et al. 2024). Consequently, naming repertoires are unlikely to be homogeneous across territories.

H2 *Lexical differentiation in firm names exhibits regional variation.*

1.3 Temporal evolution of naming repertoires

Naming systems evolve alongside broader economic, institutional, and technological transformations. Changes in market conditions, digitalization processes, and sectoral development may alter the symbolic resources available to firms and influence naming practices over time. Therefore, naming repertoires are expected to display temporal variation rather than remain stable across historical periods.

H3 *Naming repertoires evolve over time, with evidence of increasing lexical differentiation in recent years.*

This study makes three contributions to the literature. First, it extends the study of corporate naming from individual organizations to the population level by examining naming as a collective communicative system. Second, it introduces Zipfian analysis as a quantitative tool for measuring lexical concentration and symbolic differentiation in organizational naming. Third, it provides new evidence on how naming practices vary across regions and historical periods within one of the most symbolically intensive sectors of the economy. For the regional analysis, equal-sized subsampling and pooled interaction models are used to distinguish sample-size effects from residual territorial heterogeneity.

2 Data and methods

To address the research questions, we compiled a dataset of all firms registered in Spain under CNAE 2009 codes 7021, 73, 731, and 7311 between 1975 and 2023. These classifications encompass activities in public relations, advertising, and market research, sectors in which naming constitutes a particularly visible communicative and symbolic resource. Using this corpus, we estimate Zipfian coefficients for company names across temporal regimes and geographical regions. Less negative coefficients indicate lower lexical concentration and a more even distribution of lexical elements across names. Regional and temporal comparisons are then used to assess variation in naming patterns across contexts.

Methodologically, the analysis combines structural break estimation and rank-frequency modelling. First, endogenous temporal regimes are identified through the Bai and Perron (2003) multiple breakpoint procedure applied to the annual series of firm foundations. This approach allows the historical evolution of the sector to be segmented into statistically distinct periods. Second, within each temporal regime and administrative region, Zipfian coefficients are estimated for the distribution of words used in company names. These coefficients provide a measure of lexical concentration and differentiation within the naming system.

This two-stage strategy enables the examination of whether naming patterns vary across historical periods and territories. Because the regional samples differ substantially in size, the main regional estimates are complemented by within-region repeated subsampling, equal-sized Monte Carlo subsampling, pooled region-by-log-rank interaction models, and pairwise coefficient comparisons. These additional procedures distinguish the internal stability of each regional estimate from the statistical comparability of estimates across regions.

This study draws on administrative data extracted from the SABI (Sistema de Análisis de Balances Ibéricos) database, one of the most comprehensive sources of firm-level information in Spain and Portugal. SABI is maintained by Informa D&B and regularly updated to include the population of legally registered firms. For this study, we used version 310 of the database (June 13, 2025), which provides information on company names, geographic location, sector classification, employment, revenues, and foundation dates.

The dataset comprises 26,432 firms founded in Spain between 1975 and 2023. These firms belong to CNAE 2009 categories 7021 (Public Relations and Communication), 73 (Advertising and Market Research), 731 (Advertising), and 7311 (Advertising Agencies). These activities were selected because they operate in sectors where communication, branding, and symbolic differentiation constitute important dimensions of organizational activity.

Because CNAE classifications are hierarchically nested, firms were included only once in the final database according to their registered primary activity in SABI. Consequently, the use of categories 73, 731, and 7311 does not generate duplicate observations. The selected classifications should therefore be understood as defining the empirical scope of communication-, advertising-, and public-relations-related activities rather than as separate analytical populations.

Importantly, the analysis is based on the full population of firms registered in the selected sectors during the study period rather than on a sample. This population-level approach minimizes sampling bias and allows the identification of structural naming patterns across time and space.

For each firm, the database provides the official registered name, year of incorporation, autonomous community, province, legal status, and additional organizational information. Company names were cleaned and tokenized for lexical analysis, while foundation years were used to assign firms to the temporal regimes identified through the structural break procedure.

2.1 Estimating endogenous temporal regimes with Bai–Perron method

To assess the temporal dynamics of firm creation in the communication and advertising sectors in Spain, we implemented a structural break analysis on the time series of firm foundations per year (1975–2023). This procedure enables the identification of discrete shifts in the data-generating process, such as those associated with macroeconomic cycles, institutional changes, or industry evolution.

We used the breakpoints function from the *strucchange* (Zeileis 2006) package in R, based on the method developed by Bai and Perron (1998, 2003). This approach estimates multiple structural changes in linear regression models by minimizing the residual sum of squares (RSS) subject to a penalty term for complexity. The methodology allows for consistent estimation of both the number and location of breaks without requiring a priori knowledge of break dates.

Specifically, Bai and Perron's procedure fits a series of models with increasing numbers of breakpoints (from m = 0 m = 0 up to a user-defined maximum) and selects the most parsimonious model using information criteria such as the Bayesian Information Criterion (BIC). The algorithm uses dynamic programming to compute globally optimal partitions and provides confidence intervals for the estimated break dates.

We modeled the annual number of firm foundations (n_sociedades) as a function of calendar year (Año constitución), evaluating break models from 0 to 5 segments. The estimation was performed using ordinary least squares and allowed for heterogeneous variances across segments.

2.2 Zipf's law estimation framework

From a corporate communication perspective, Zipfian distributions can be interpreted as indicators of symbolic saturation versus communicative differentiation within an organisational field. Conversely, flatter Zipfian distributions reflect greater lexical diversity and a more differentiated symbolic repertoire. In this sense, Zipf's Law provides a quantitative tool for analysing how naming systems reflect broader structural dynamics of communication, identity, and differentiation at the industry level.

To analyze the degree of lexical concentration in the naming of firms, we test the empirical validity of Zipf's Law across temporal regimes and geographic regions. Zipf's Law (Zipf 1945) posits that the frequency of an element—in this case, a word—follows a power-law distribution in relation to its rank. More formally, the frequency *f* of a word is inversely proportional to its rank $r : f \sim r^{\alpha}$, where the parameter α (typically close to -1 in natural language) captures the slope of the log-log distribution (Esteban-Rojo et al. 2026).

To operationalize this, we construct word frequency distributions for company names across six temporal regimes (1975–1983, 1984–1990, 1991–1999, 2000–2007, 2008–2016, 2017–2023) and for each of Spain's 17 administrative regions. Within each subset, names are tokenized and standardized using lowercasing and a list of stopword exclusions. We then aggregate word frequencies and compute the log-rank and log-frequency of each word. Linear regression models are estimated using ordinary least squares (OLS) on the log-log transformed data:

$$log(Rank_i) = \delta + \alpha log(Size_i) + \varepsilon_i$$

The estimated slope α serves as an index of lexical concentration: steeper (more negative) slopes indicate dominance of a few repeated terms (higher concentration), whereas flatter (less negative) slopes imply greater lexical diversity (higher lexical differentiation).

All regressions were implemented in R (Team 2021) using the *lm()* function. The estimation process was fully automated through iterative pipelines written in functional programming style (*purrr*, *dplyr*) (Wickham and Henry 2025; Wickham et al. 2025), enabling the systematic and reproducible estimation of Zipfian slopes for each region and regime. For each model, we extract the estimated value of α, its 95% confidence interval, and the R-squared statistic to assess model fit.

In line with theoretical benchmarks from prior literature (Zipf 1945; Piantadosi 2014; Thurner et al. 2015), an αα close to −1 supports the hypothesis that naming behavior resembles natural language distributions. Deviations from this value are interpreted as follows: values of $\alpha < -1$ denote higher concentration (i.e., increased repetition of a few dominant words), while $\alpha > -1$ suggests flatter distributions with more even lexical use, typically associated with greater lexical differentiation.

Importantly, Zipfian estimation in this context does not assume any specific semantic interpretation of words but rather treats naming as a symbolic system where structural regularities can be detected and compared. This allows for a statistical assessment of naming lexical differentiation across time and space without imposing ex-ante normative definitions.

To further ensure robustness, models are validated by visually inspecting the linearity of log-log plots and by evaluating the distribution of residuals. Where necessary, high-frequency outliers (e.g., generic suffixes or firm-type indicators) are removed to avoid biasing slope estimates.

This framework provides a replicable method for assessing naming concentration in large collections of company names.

2.3 Robustness and formal comparison of regional coefficients

To evaluate the robustness and comparability of the regional Zipf coefficients, we implemented two complementary resampling procedures. The first evaluates within-region stability. For each autonomous community, we generated 500 random subsamples containing 80% of the firms in that region and re-estimated the Zipf coefficient using the same specification as in the main analysis. This procedure assesses whether individual regional estimates are sensitive to a small number of firms or influential lexical observations.

The second procedure evaluates between-region comparability under identical sample sizes. Because the smallest regional sample contains 115 firms, we randomly selected 115 firms without replacement from each of the 17 autonomous communities and re-estimated all regional coefficients. This procedure was repeated over 1000 Monte Carlo replications.

We additionally repeated the analysis using common sample sizes of 25, 50 and 100 firms per region to assess the sensitivity and convergence of the estimates.

Regional heterogeneity was evaluated in two additional ways. First, for every equal-sized replication, we estimated a pooled model containing region-by-log-rank interaction terms and tested the joint null hypothesis that the regional slopes were homogeneous. Second, we constructed Monte Carlo intervals for all 136 pairwise differences between regional coefficients. These procedures permit formal assessment of whether regional differences persist once unequal sample sizes have been removed. Full results are reported in Annex B.

The temporal coefficients continue to be estimated separately across the structural regimes identified through the Bai–Perron procedure. Because the equal-sized resampling and interaction analyses reported here concern the regional comparison, differences across temporal regimes are interpreted descriptively rather than as formal pairwise differences.

3 Results

3.1 Descriptive naming patterns

Figure 1 visualizes the evolution of firm creation between 1975 and 2023, overlaid with structural breakpoints estimated using the Bai and Perron (2003) procedure (Table 2). The figure shows clear temporal regimes in entrepreneurial intensity, with major inflection points around 1983, 1990, 1999, 2007, and 2016. These breaks define six statistically distinct periods that serve as the basis for our Zipfian analysis. Figure 1 complements this by displaying the top five most frequent words in company names across each temporal regime.

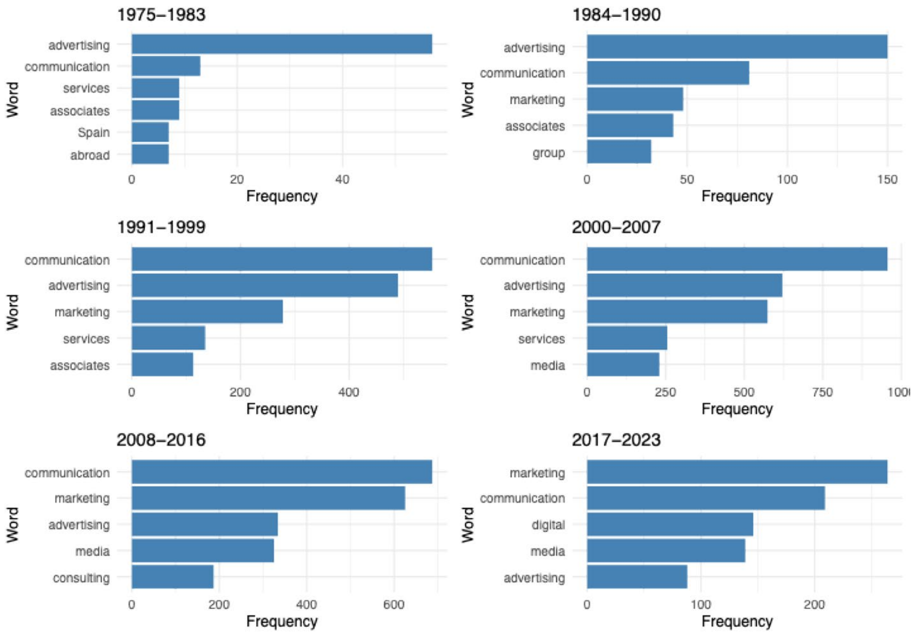

Fig. 1 Top five most frequent words in company names by temporal regime

It reveals both consistencies (e.g., the dominance of "communication" and "advertising") and notable shifts—for instance, the rise of "digital" and "media" in the 2008–2016 and 2017–2023 periods, reflecting broader technological transformations and branding strategies. This dual approach—integrating structural time segmentation with lexical frequency modeling—allows us to investigate whether naming lexical differentiation evolves in tandem with institutional and market changes.

Figure 2 disaggregates these patterns by region, highlighting notable geographic variation in naming conventions. Together, these figures raise important questions about the balance between creativity and repetition in naming—and how that balance may shift over time and space.

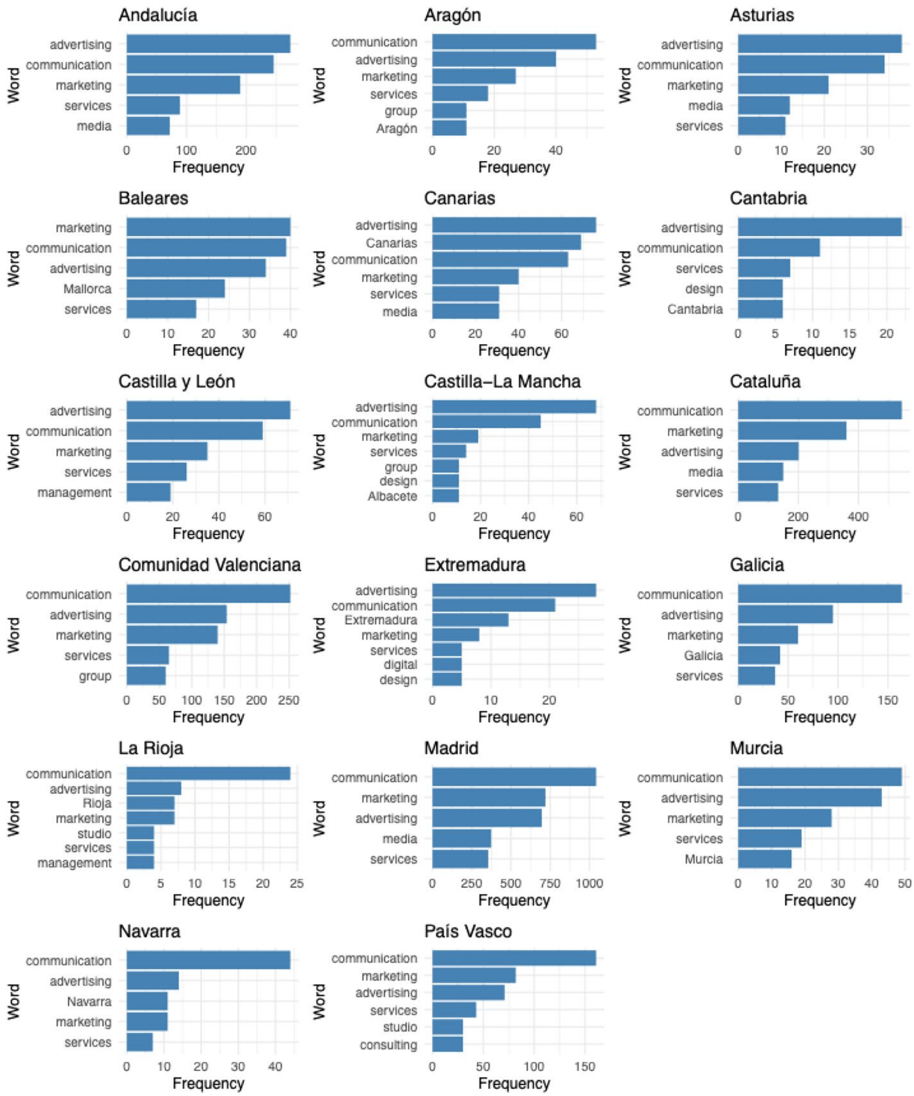

Fig. 2 Top five most frequent words in company names by region

3.2 Testing H1: lexical differentiation

The results indicate that the Zipf coefficients (α) estimated through log-log rank-frequency regressions are consistently and significantly above −1 across all specifications (Tables 1 and 3). This deviation from the theoretical benchmark of Zipf's Law suggests a distribution that is more balanced and lexically diverse than would be expected under a strict power-law regime. In practical terms, this reflects a relatively low level of lexical concentration and a broad distribution of lexical elements across company names. Overall, the evidence is consistent with H1.

Table 1 Zipf´s law regression results by region

Region	δ	p value	α	p value	R^2	p value Zipf´s law
Andalucía	4.049*** (0.033)	0.000	−0.540*** (0.005)	0.000	0.807	0.000e+00 (strongly rejected)
Aragón	2.545*** (0.048)	0.000	−0.419*** (0.008)	0.000	0.766	0.000e+00 (strongly rejected)
Asturias	2.355*** (0.060)	0.000	−0.423*** (0.011)	0.000	0.742	6.188e−193 (strongly rejected)
Baleares	2.648*** (0.046)	0.000	−0.423*** (0.008)	0.000	0.763	0.000e+00 (strongly rejected)
Canarias	2.962*** (0.045)	0.,000	−0.455*** (0.007)	0.000	0.767	0.000e+00 (strongly rejected)
Cantabria	1.884*** (0.059)	0.000	−0.373*** (0.012)	0.000	0.766	3.658e−140 (strongly rejected)
Castilla y León	2.762*** (0.051)	0.000	−0.446*** (0.009)	0.000	0.748	0.000e+00 (strongly rejected)
Castilla-La Mancha	2.550*** (0.056)	0.000	−0.437*** (0.010)	0.000	0.753	2.929e−244 (strongly rejected)
Cataluña	4.532*** (0.025)	0.000	−0.551*** (0.003)	0.000	0.816	0.000e+00 (strongly rejected)
Comunidad Valenciana	3.917*** (0.033)	0.000	−0.528*** (0.005)	0.000	0.808	0.000e+00 (strongly rejected)
Extremadura	1.950*** (0.071)	0.000	−0.384*** (0.015)	0.000	0.699	5.466e−126 (strongly rejected)
Galicia	3.188*** (0.040)	0.000	−0.470*** (0.006)	0.000	0.783	0.000e+00 (strongly rejected)
La Rioja	1.658*** (0.075)	0.000	−0.350*** (0.017)	0.000	0.675	2.999e−97 (strongly rejected)
Madrid	5.088*** (0.023)	0.000	−0.591*** (0.003)	0.000	0.827	0.000e+00 (strongly rejected)
Murcia	2.479*** (0.055)	0.000	−0.425*** (0.010)	0.000	0.746	4.859e−250 (strongly rejected)
Navarra	1.913*** (0.073)	0.000	−0.378*** (0.015)	0.000	0.676	3.939e−123 (strongly rejected)
País Vasco	3.145*** (0.041)	0.000	−0.463*** (0.006)	0.000	0.767	0.000e+00 (strongly rejected)

Standard errors are reported in parentheses. Statistical significance of the estimated coefficients is indicated using asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. The interpretation of the p-values follows these guidelines: if the p-value is less than 0.01, the null hypothesis is considered strongly rejected; if the p-value lies between 0.01 and 0.05, it is rejected; if it falls between 0.05 and 0.10, the rejection is considered marginal; and if it exceeds 0.10, the null hypothesis is not rejected. In the context of Zipf's Law, the null hypothesis (H_0:$\alpha = -1$)

These findings are consistent with the communicative nature of the sectors under study. Advertising, public relations, and communication firms compete in markets where visibility, distinctiveness, and symbolic positioning constitute important organizational resources. Under these conditions, firms may face incentives to avoid highly repetitive naming practices and to adopt more differentiated lexical repertoires. The relatively low levels of lexical concentration observed in the data therefore suggest that naming functions not merely as an administrative requirement but also as a mechanism of symbolic differentiation within highly competitive communication environments.

3.3 Testing H2: regional variation

The coefficients estimated from the complete regional samples display substantial variation (Table 1 and Fig. 3). In the unadjusted estimates, less negative coefficients are observed in regions such as La Rioja, Cantabria, Navarra and Extremadura, whereas Madrid, Cataluña, Andalucía and Comunidad Valenciana exhibit comparatively steeper slopes. However, because the regional samples differ considerably in size, these complete-sample estimates should not be interpreted as an invariant ranking of regional naming systems.

The robustness analyses reported in Annex B clarify the extent to which these differences reflect sample size. The within-region resampling exercise shows that each regional coefficient is internally stable: the mean absolute difference between the original and resampled estimates is 0.017, Pearson's correlation is 0.999 and Spearman's rank correlation is 0.988. Nevertheless, the equal-sized Monte Carlo analysis shows that controlling for sample size modifies the magnitude and ordering of several regional coefficients. The mean absolute difference between the complete-sample coefficients and the equal-sized estimates is 0.133,

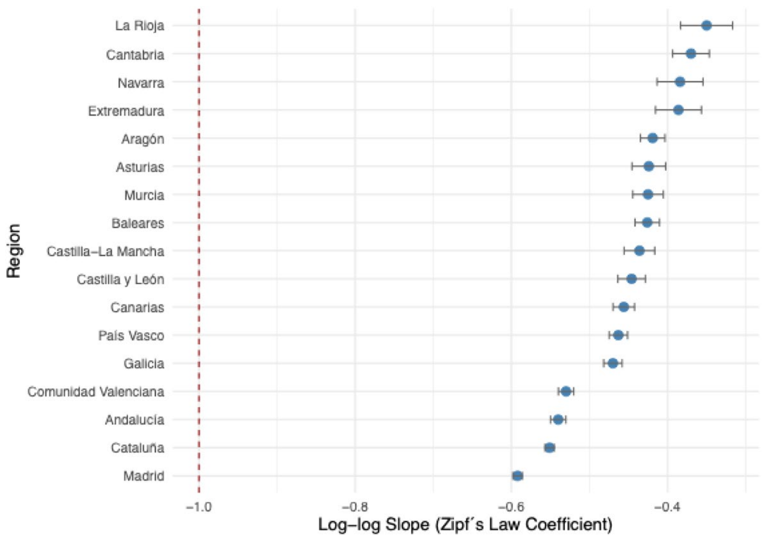

Fig. 3 Zipf coefficients by region. *Note* The figure reports coefficients estimated from the complete regional samples. Because regional sample sizes differ, the relative ordering shown here should not be interpreted as invariant. Equal-sized Monte Carlo estimates and formal regional comparisons are reported in Annex B

and the mean between-region range decreases from 0.245 at a common sample of 25 firms to 0.137 at 115 firms per region.

Importantly, equalising sample sizes reduces but does not eliminate regional heterogeneity. In all 1000 equal-sized replications, a pooled model with region-by-log-rank interactions rejects the null hypothesis of homogeneous regional slopes. In addition, Monte Carlo intervals exclude zero for 17 of the 136 pairwise regional comparisons. These results support H2 in a qualified sense: regional variation persists after controlling for sample size, but it is more moderate and concentrated in a subset of regional comparisons rather than representing a stable ordering across all regions.

The residual differences may be associated with regional market conditions, institutional environments or locally embedded naming practices. However, the present analysis does not identify the causal mechanisms responsible for those differences. Interpretations based on market density, territorial identity or nominal saturation should therefore be understood as plausible explanations for future investigation rather than as conclusions established by the present evidence.

3.4 Testing H3: temporal evolution

The structural break analysis applied (Table 2) to the time series of newly founded firms identified five statistically significant breakpoints in the annual number of company formations. These breaks delineate six distinct temporal regimes that segment the evolution of the sector between 1975 and 2023. The estimated break years are 1983, 1990, 1999, 2007 and 2016. These were obtained by minimizing the Bayesian Information Criterion (BIC), which balances model fit and parsimony. Because the number of firms also differs across temporal regimes, the magnitude of the period-specific coefficients and their relative ordering should be interpreted descriptively. The formal equal-size and pooled interaction procedures applied to the regional analysis are not used to establish pairwise differences between temporal regimes.

The first regime spans from 1975 to 1983 and represents a foundational phase for the sector, marked by relatively low but gradually increasing activity. The second regime, from 1984 to 1990, corresponds to a period of expansion, likely driven by Spain's economic modernization and integration into the European Community. The third period, covering the years 1991 to 1999, reflects continued sectoral development under conditions of economic liberalization and the digital transition.

Table 2 Structural breakpoints in the time series of firm foundations (1975–2023)

Number of breaks (m)	Break years	R^2	RSS	BIC
0	–	–	4,829,960.8	714.2
1	2007	–	703,598.3	631.4
2	2006, 2016	0.8571	332,898.4	606.4
3	1999, 2008, 2016	0.8367	123,604.4	569.6
4	1999, 2007, 2011, 2016	0.8571	77,324.5	558.3
5	1983, 1990, 1999, 2007, 2016	0.8571	58,916.7	556.6

The first regime spans from 1975 to 1983 and represents a foundational phase for the sector, marked by relatively low but gradually increasing activity. The second regime, from 1984 to 1990, corresponds to a period of expansion, likely driven by Spain's economic modernization and integration into the European Community. The third period, covering the years 1991 to 1999, reflects continued sectoral development under conditions of economic liberalization and the digital transition.

Between 2000 and 2007, the fourth regime captures the boom years leading up to the global financial crisis, during which firm creation reached particularly high levels. The fifth regime, spanning from 2008 to 2016, reflects the post-crisis adjustment period, characterized by a marked slowdown and reconfiguration of the business landscape. Finally, the sixth regime begins in 2017 and continues through 2023, representing a recovery and transformation phase likely shaped by the rise of digital platforms and the lingering effects of the COVID-19 pandemic.

These regimes provide a data-driven structure for interpreting changes in entrepreneurial dynamics over time. Importantly, they serve as the temporal framework for analyzing the evolution of naming concentration through the application of Zipf's Law in the subsequent sections.

Figure 4 visually represents the temporal evolution of firm creation in the Spanish communication and advertising sector between 1975 and 2023, measured by the annual number of newly registered companies. The solid blue line depicts the trajectory of firm foundations over time, revealing both cyclical dynamics and long-term trends. The red dashed vertical lines indicate the five structural breakpoints identified through the Bai and Perron methodology, corresponding to the years 1983, 1990, 1999, 2007, and 2016.

The figure highlights a steady increase in firm creation from the mid-1970s through the early 2000s, peaking around the years immediately preceding the 2008 financial crisis. After this peak, the number of new firms exhibits a clear decline, particularly sharp in the final years of the series, potentially reflecting the effects of market saturation, technological disruption, and broader economic challenges.

Each vertical break denotes a statistically significant change in the underlying trend, marking transitions between the six regimes used in the subsequent Zipfian analysis. These

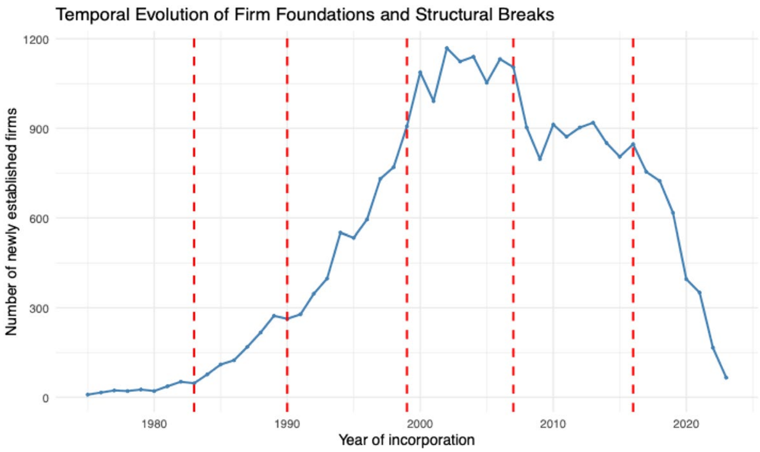

Fig. 4 Temporal evolution of firm foundations and structural breaks

Table 3 Zipf's law regression results by regime

Regime	δ	p value	α	p value	R^2	p value Zipf's law
1975–1983	2.059*** (0.065)	0.000	−0.384*** (0.013)	0.000	0.694	2.929e−167 (strongly rejected)
1984–1990	3.282*** (0.040)	0.000	−0.484*** (0.006)	0.000	0.792	0.000e+00 (strongly rejected)
1991–1999	4.598*** (0.027)	0.000	−0.569*** (0.003)	0.000	0.824	0.000e+00 (strongly rejected)
2000–2007	4.997*** (0.023)	0.000	−0.582*** (0.003)	0.000	0.822	0.000e+00 (strongly rejected)
2008–2016	4.709*** (0.024)	0.000	−0.554*** (0.003)	0.000	0.799	0.000e+00 (strongly rejected)
2017–2023	3.785*** (0.031)	0.000	−0.492*** (0.004)	0.000	0.768	0.000e+00 (strongly rejected)

Standard errors are reported in parentheses. Statistical significance of the estimated coefficients is indicated using asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. The interpretation of the p-values follows these guidelines: if the p-value is less than 0.01, the null hypothesis is considered strongly rejected; if the p-value lies between 0.01 and 0.05, it is rejected; if it falls between 0.05 and 0.10, the rejection is considered marginal; and if it exceeds 0.10, the null hypothesis is not rejected. In the context of Zipf's Law, the null hypothesis (H_0:$\alpha = -1$)

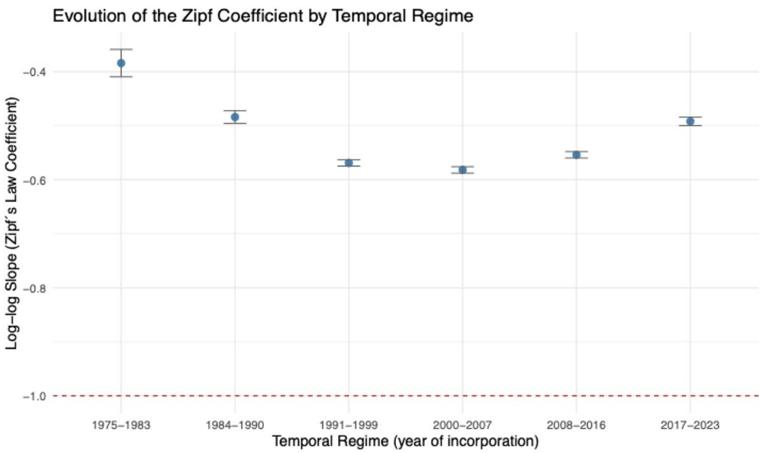

Fig. 5 Evolution of the Zipf coefficient by temporal regime

turning points provide empirical justification for analyzing naming dynamics across temporally distinct economic and institutional contexts, reinforcing the importance of considering structural change in longitudinal studies of corporate behavior.

The Zipf regressions estimated for each temporal regime (Table 3 and Fig. 5) reveal a non-linear evolution in naming patterns over time. Lexical concentration increases between the first regime (1975–1983) and the fourth regime (2000–2007), reaching its highest level during the expansion years preceding the financial crisis. Thereafter, the trend reverses, with coefficients becoming progressively less negative in the 2008–2016 and 2017–2023 periods. The most recent regime exhibits substantially lower concentration than the middle stages of sector development, suggesting a renewed tendency toward lexical differentiation in contemporary naming practices.

This pattern coincides with broader transformations in the communication sector, although the present analysis does not allow causal attribution. During the expansion years of the 1990s and early 2000s, rapid firm creation and market growth may have encouraged the repeated use of a relatively limited set of sector-specific terms, increasing lexical concentration. By contrast, the post-2008 period coincides with digitalization, the emergence of new communication services, and the fragmentation of professional niches. These developments may be associated with a broader range of symbolic resources available to firms. However, the evidence presented here should be interpreted as descriptive rather than as evidence of a causal relationship. The decline in lexical concentration observed after 2008 is therefore consistent with a more heterogeneous and specialized communication landscape.

3.5 Regional and temporal interactions

Finally, when Zipf coefficients are cross-tabulated by region and temporal regime (Table 4 and Fig. 6), a layered descriptive pattern emerges. Several regional trajectories point toward lower lexical concentration in recent periods, although the magnitude and timing of the changes vary across territories. Because the region-by-period cells differ markedly in the number of firms, these trajectories should not be interpreted as formal evidence that particular regions experienced statistically distinct temporal changes. They are reported as exploratory patterns that may guide future analyses based on pooled region-by-period interaction models and comparable cell sizes.

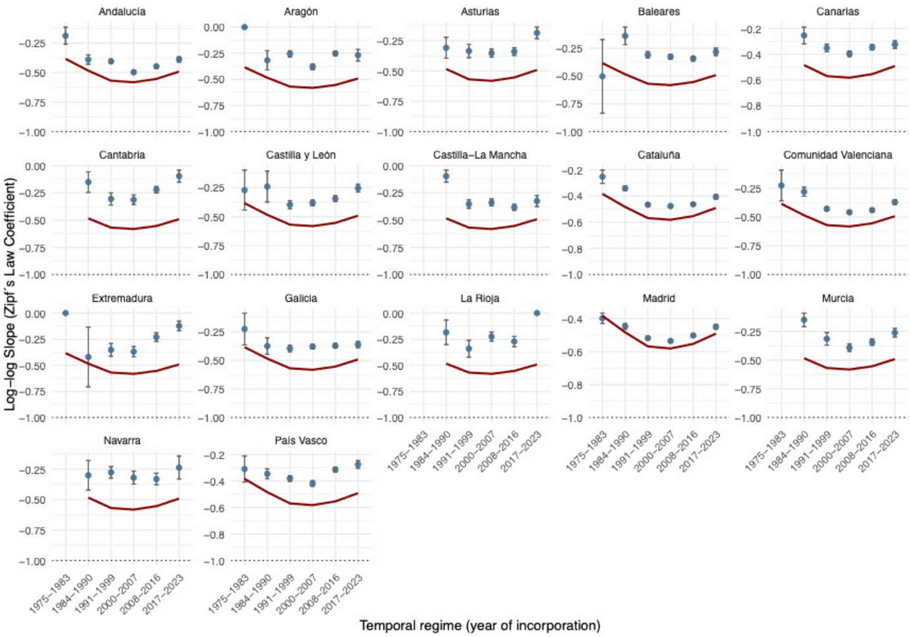

Fig. 6 Evolution of the Zipf coefficient by temporal regime and reg

4 Conclusions

Rather than treating company names as isolated branding decisions, this study conceptualizes corporate naming as a population-level communication system. Using a comprehensive dataset of 26,432 firms operating in Spain's advertising, public relations, and communication sectors between 1975 and 2023, we examined the degree of lexical concentration and symbolic differentiation in corporate naming practices through the application of Zipf's Law.

The findings reveal that naming practices in these sectors are characterized by relatively low levels of lexical concentration, suggesting a broad and differentiated symbolic repertoire. Naming patterns are not completely homogeneous across space or time. In the regional analysis, equal-sized Monte Carlo subsampling shows that unequal sample sizes explain a substantial part of the dispersion observed in the complete samples. Nevertheless, pooled interaction tests and pairwise comparisons indicate that residual regional heterogeneity remains after controlling for sample size. The temporal analysis, interpreted descriptively, indicates that the most recent period is characterized by lower lexical concentration than the middle stages of sector development.

The findings contribute to existing research in several respects. First, they extend the study of corporate naming beyond the level of individual firms by examining naming as a collective communicative phenomenon. Second, they introduce Zipfian analysis as a replicable quantitative framework for measuring lexical concentration and symbolic differentiation in organizational populations. Third, they provide new evidence on territorial and historical variation in naming practices while showing the importance of separating substantive heterogeneity from variation generated by unequal sample sizes.

From a corporate communication perspective, the findings suggest that naming does not occur in a neutral environment. Residual regional differences remain after controlling for sample size, indicating that naming repertoires may be associated with territorial or institutional contexts. However, the robustness analysis also shows that part of the variation initially attributed to regional context reflects differences in the number of firms observed. Market density, territorial dynamics and historical saturation therefore remain possible explanatory mechanisms rather than effects directly established by the present analysis.

The study also has practical implications for branding professionals, entrepreneurs, and policy makers. Awareness of broad regional and temporal patterns in naming may contribute to more context-sensitive naming strategies and provide additional insights into the symbolic dimensions of entrepreneurial ecosystems. However, practitioners should not interpret the unadjusted regional coefficients as a fixed hierarchy of more or less differentiated territories.

The interpretation differs between the regional and temporal analyses. For regions, equal-sized Monte Carlo subsampling, pooled interaction models and pairwise coefficient comparisons provide formal evidence of residual heterogeneity after controlling for sample size. For temporal regimes, the estimates remain descriptive because formal equal-sized and pooled comparisons across periods are not undertaken.

Several limitations should be acknowledged. The analysis focuses exclusively on legal company names within a single national context and industry group. Furthermore, the study examines lexical concentration rather than semantic meaning, phonetic characteristics or stakeholder perceptions. Consequently, lexical differentiation should not be interpreted as a direct measure of creativity, but rather as an indicator of symbolic diversity within naming systems.

The robustness analyses also identify an important methodological qualification. Within-region repeated subsampling confirms that the regional estimates are internally stable, but equal-sized Monte Carlo subsampling shows that unequal sample sizes account for a substantial proportion of the dispersion observed in the complete regional samples. Although pooled interaction tests and pairwise comparisons confirm that residual regional heterogeneity remains, only 17 of the 136 regional pairs are statistically distinguishable under the Monte Carlo procedure. The findings therefore support the existence of regional heterogeneity without supporting a definitive ranking of all regional naming systems.

The temporal analysis remains subject to a separate limitation. The temporal regimes contain different numbers of firms, and the study does not implement an equivalent equal-sized resampling or pooled interaction analysis across periods. Period-specific differences should therefore be interpreted as descriptive trajectories rather than formal pairwise differences. Future research could extend the equal-sample framework to temporal and region-by-period comparisons, incorporate semantic and phonetic dimensions of naming, and examine relationships between naming differentiation and organizational outcomes such as survival, growth, reputation or brand performance.

Annex A. Complete regression results and log–log plots

See annex Table 4 and Figs. 7, 8.

Table 4 Zipf´s law regression results by regime and region

Regime | Region	δ	pvalue	α	pvalue	R^2	p value Zipf´s law
1975–1983 | Andalucía	0.562*** (0.105)	0.000005822	−0.188*** (0.037)	0.00001426	0.42	6.775e−22 (strongly rejected)
1975–1983 | Baleares	0.920** (0.282)	0.01139	−0.503** (0.169)	0.01791	0.524	1.880e−02 (rejected)
1975–1983 | Castilla y León	0.555*** (0.166)	0.006481	−0.271** (0.088)	0.0104	0.464	4.632e−06 (strongly rejected)
1975–1983 | Cataluña	0.972*** (0.094)	1.103E−16	−0.254*** (0.026)	1.408E−15	0.525	7.326e−46 (strongly rejected)
1975–1983 | Comunidad Valenciana	0.550*** (0.156)	0.002207	−0.224*** (0.068)	0.003622	0.367	5.400e−10 (strongly rejected)
1975–1983 | Galicia	0.501*** (0.146)	0.004105	−0.225*** (0.071)	0.006661	0.42	3.070e−08 (strongly rejected)
1975–1983 | Madrid	1.910*** (0.072)	1.102E−69	−0.399*** (0.016)	3.827E−65	0.745	4.979e−96 (strongly rejected)
1975–1983 | País Vasco	0.836*** (0.122)	5.862E−07	−0.310*** (0.050)	0.000002325	0.628	1.191e−12 (strongly rejected)
1984–1990 | Andalucía	1.718*** (0.084)	1.75E−44	−0.390*** (0.020)	2.126E−41	0.718	1.246e−63 (strongly rejected)
1984–1990 | Aragón	0.961*** (0.131)	1.463E−08	−0.318*** (0.047)	6.576E−08	0.57	1.688e−16 (strongly rejected)
1984–1990 | Asturias	0.928*** (0.123)	9.716E−09	−0.309*** (0.044)	4.61E−08	0.59	4.161e−17 (strongly rejected)
1984–1990 | Baleares	0.375*** (0.102)	0.001104	−0.138*** (0.040)	0.001826	0.317	3.846e−18 (strongly rejected)

Table 4 Zipf´s law regression results by regime and region

Regime | Region	δ	pvalue	α	pvalue	R^2	*p* value Zipf´s law
1984–1990 | Canarias	0.832*** (0.103)	1.849E−10	−0.254*** (0.033)	1.026E−09	0.551	1.095e−26 (strongly rejected)
1984–1990 | Cantabria	0.326*** (0.096)	0.004737	−0.151*** (0.048)	0.007663	0.433	1.679e−10 (strongly rejected)
1984–1990 | Castilla y León	0.671*** (0.181)	0.0008547	−0.240*** (0.068)	0.001418	0.3	4.964e−12 (strongly rejected)
1984–1990 | Castilla-La Mancha	0.251*** (0.069)	0.001457	−0.096*** (0.028)	0.002405	0.336	1.416e−20 (strongly rejected)
1984–1990 | Cataluña	1.906*** (0.052)	2.961E−141	−0.342*** (0.010)	5.943E−133	0.716	5.707e−245 (strongly rejected)
1984–1990 | Comunidad Valenciana	1.195*** (0.082)	1.562E−29	−0.279*** (0.020)	1.384E−27	0.583	6.238e−71 (strongly rejected)
1984–1990 | Extremadura	0.695** (0.216)	0.01831	−0.420** (0.146)	0.02815	0.58	7.403e−03 (strongly rejected)
1984–1990 | Galicia	1.321*** (0.122)	1.082E−15	−0.375*** (0.037)	1.545E−14	0.635	3.135e−24 (strongly rejected)
1984–1990 | La Rioja	0.364*** (0.109)	0.007706	−0.184** (0.060)	0.01229	0.482	9.298e−08 (strongly rejected)
1984–1990 | Madrid	2.679*** (0.052)	1.079E−240	−0.448*** (0.009)	2.183E−227	0.768	3.624e−279 (strongly rejected)
1984–1990 | Murcia	0.412*** (0.078)	0.00001524	−0.150*** (0.030)	0.00004052	0.483	5.959e−21 (strongly rejected)
1984–1990 | Navarra	0.699*** (0.131)	0.00008171	−0.301*** (0.062)	0.000202	0.613	9.619e−09 (strongly rejected)
1984–1990 | País Vasco	1.487*** (0.075)	1.089E−40	−0.346*** (0.019)	9.185E−38	0.726	4.251e−67 (strongly rejected)
1991–1999 | Andalucía	2.402*** (0.050)	1.124E−220	−0.405*** (0.009)	2.715E−208	0.757	1.477e−299 (strongly rejected)
1991–1999 | Aragón	1.196*** (0.068)	8.176E−43	−0.256*** (0.015)	3.282E−40	0.579	2.505e−114 (strongly rejected)
1991–1999 | Asturias	1.336*** (0.105)	2.195E−22	−0.336*** (0.028)	7.213E−21	0.,594	2.671e−42 (strongly rejected)
1991–1999 | Baleares	1.496*** (0.068)	2.796E−58	−0.310*** (0.015)	1.295E−54	0.652	2.656e−118 (strongly rejected)
1991–1999 | Canarias	1.760*** (0.071)	6.602E−73	−0.351*** (0.015)	1.736E−68	0.664	4.378e−127 (strongly rejected)
1991–1999 | Cantabria	1.125*** (0.094)	2.481E−18	−0.306*** (0.028)	8.41E−17	0.647	7.416e−36 (strongly rejected)
1991–1999 | Castilla y León	1.867*** (0.079)	1.237E−59	−0.397*** (0.018)	8.85E−56	0.713	2.738e−84 (strongly rejected)
1991–1999 | Castilla-La Mancha	1.539*** (0.084)	2.293E−39	−0.353*** (0.020)	1.04E−36	0.68	4.334e−66 (strongly rejected)
1991–1999 | Cataluña	3.198*** (0.037)	0	−0.467*** (0.006)	0	0.803	0.000e+00 (strongly rejected)
1991–1999 | Comunidad Valenciana	2.552*** (0.051)	4.184E−234	−0.428*** (0.009)	4.759E−221	0.765	3.236e−292 (strongly rejected)
1991–1999 | Extremadura	1.274*** (0.104)	1.78E−18	−0.352*** (0.031)	4.023E−17	0.666	1.417e−30 (strongly rejected)
1991–1999 | Galicia	2.033*** (0.071)	3.087E−89	−0.395*** (0.015)	7.301E−84	0.699	6.650e−129 (strongly rejected)

Table 4 Zipf´s law regression results by regime and region

Regime | Region	δ	pvalue	α	pvalue	R^2	*p* value Zipf´s law
1991–1999 | La Rioja	1.080*** (0.121)	4.324E−11	−0.341*** (0.041)	3.059E−10	0.633	5.353e−19 (strongly rejected)
1991–1999 | Madrid	3.716*** (0.037)	0	−0.519*** (0.005)	0	0.801	0.000e+00 (strongly rejected)
1991–1999 | Murcia	1.273*** (0.106)	1.326E−21	−0.315*** (0.028)	3.398E−20	0.549	3.092e−46 (strongly rejected)
1991–1999 | Navarra	1.031*** (0.084)	1.78E−19	−0.277*** (0.024)	4.72E−18	0.644	8.402e−43 (strongly rejected)
1991–1999 | País Vasco	2.113*** (0.057)	6.262E−140	−0.381*** (0.011)	9.178E−132	0.722	9.153e−211 (strongly rejected)
2000–2007 | Andalucía	3.301*** (0.043)	0	−0.497*** (0.007)	0	0.795	0.000e+00 (strongly rejected)
2000–2007 | Aragón	1.935*** (0.065)	3.583E−91	−0.380*** (0.013)	1.483E−85	0.729	1.500e−136 (strongly rejected)
2000–2007 | Asturias	1.618*** (0.074)	5.544E−52	−0.353*** (0.017)	1.852E−48	0.707	7.034e−86 (strongly rejected)
2000–2007 | Baleares	1.712*** (0.059)	2.685E−95	−0.326*** (0.012)	1.35E−89	0.684	1.005e−178 (strongly rejected)
2000–2007 | Canarias	2.208*** (0.059)	3.754E−145	−0.396*** (0.011)	9.718E−137	0.722	7.696e−208 (strongly rejected)
2000–2007 | Cantabria	1.248*** (0.080)	1.067E−27	−0.313*** (0.022)	1.251E−25	0.69	4.869e−52 (strongly rejected)
2000–2007 | Castilla y León	2.024*** (0.063)	2.512E−107	−0.382*** (0.013)	6.479E−101	0.715	2.786e−161 (strongly rejected)
2000–2007 | Castilla-La Mancha	1.649*** (0.073)	6.618E−62	−0.338*** (0.016)	4.91E−58	0.656	3.267e−113 (strongly rejected)
2000–2007 | Cataluña	3.474*** (0.034)	0	−0.477*** (0.005)	0	0.783	0.000e+00 (strongly rejected)
2000–2007 | Comunidad Valenciana	2.991*** (0.043)	0	−0.458*** (0.007)	0	0.779	0.000e+00 (strongly rejected)
2000–2007 | Extremadura	1.511*** (0.097)	1.964E−29	−0.369*** (0.025)	2.261E−27	0.668	1.389e−46 (strongly rejected)
2000–2007 | Galicia	2.210*** (0.052)	4.894E−183	−0.380*** (0.010)	9.272E−173	0.723	1.880e−277 (strongly rejected)
2000–2007 | La Rioja	0.871*** (0.080)	5.066E−18	−0.225*** (0.022)	8.216E−17	0.547	7.299e−54 (strongly rejected)
2000–2007 | Madrid	4.124*** (0.031)	0	−0.536*** (0.004)	0	0.802	0.000e+00 (strongly rejected)
2000–2007 | Murcia	1.867*** (0.078)	6.243E−62	−0.392*** (0.017)	5.764E−58	0.706	1.840e−89 (strongly rejected)
2000–2007 | Navarra	1.288*** (0.101)	6.575E−23	−0.320*** (0.027)	2.416E−21	0.584	2.077e−46 (strongly rejected)
2000–2007 | País Vasco	2.426*** (0.058)	1.213E−179	−0.419*** (0.011)	1.019E−169	0.722	1.036e−236 (strongly rejected)
2008–2016 | Andalucía	2.908*** (0.045)	0	−0.448*** (0.007)	0	0.761	0.000e+00 (strongly rejected)
2008–2016 | Aragón	1.269*** (0.058)	2.56E−63	−0.252*** (0.012)	1.427E−59	0.6	8.486e−169 (strongly rejected)
2008–2016 | Asturias	1.560*** (0.069)	1.887E−53	−0.341*** (0.016)	7.811E−50	0.725	2.162e−90 (strongly rejected)

Table 4 Zipf´s law regression results by regime and region

Regime | Region	δ	pvalue	α	pvalue	R^2	*p* value Zipf´s law
2008–2016 | Baleares	1.788*** (0.054)	1.096E−107	−0.342*** (0.011)	8.274E−101	0.748	3.252e−180 (strongly rejected)
2008–2016 | Canarias	1.869*** (0.058)	5.26E−114	−0.345*** (0.011)	2.421E−107	0.689	9.011e−200 (strongly rejected)
2008–2016 | Cantabria	0.913*** (0.058)	1.637E−29	−0.221*** (0.015)	3.79E−27	0.668	9.149e−77 (strongly rejected)
2008–2016 | Castilla y León	1.774*** (0.065)	2.739E−85	−0.345*** (0.013)	3.937E−80	0.68	4.510e−150 (strongly rejected)
2008–2016 | Castilla-La Mancha	1.871*** (0.064)	3.04E−79	−0.383*** (0.014)	4.786E−74	0.762	5.647e−114 (strongly rejected)
2008–2016 | Cataluña	3.329*** (0.036)	0	−0.464*** (0.005)	0	0.768	0.000e+00 (strongly rejected)
2008–2016 | Comunidad Valenciana	2.802*** (0.045)	0	−0.439*** (0.007)	0	0.767	0.000e+00 (strongly rejected)
2008–2016 | Extremadura	0.897*** (0.078)	1.467E−19	−0.230*** (0.021)	3.253E−18	0.567	3.118e−56 (strongly rejected)
2008–2016 | Galicia	2.152*** (0.055)	1.113E−168	−0.372*** (0.010)	2.53E−159	0.701	2.658e−267 (strongly rejected)
2008–2016 | La Rioja	1.029*** (0.090)	2.02E−18	−0.272*** (0.025)	4.047E−17	0.594	1.437e−43 (strongly rejected)
2008–2016 | Madrid	3.818*** (0.034)	0	−0.503*** (0.005)	0	0.765	0.000e+00 (strongly rejected)
2008–2016 | Murcia	1.725*** (0.066)	3.048E−77	−0.344*** (0.014)	1.659E−72	0.69	5.065e−135 (strongly rejected)
2008–2016 | Navarra	1.361*** (0.092)	6.206E−28	-0.331*** (0.024)	5.138E−26	0.635	2.940e−52 (strongly rejected)
2008–2016 | País Vasco	1.772*** (0.056)	6.414E−125	−0.314*** (0.010)	4.528E−118	0.628	1.341e−260 (strongly rejected)
2017–2023 | Andalucía	2.217*** (0.060)	7.83E−152	−0.389*** (0.011)	2.01E−143	0.688	4.670e−228 (strongly rejected)
2017–2023 | Aragón	1.020*** (0.102)	7.55E−16	−0.269*** (0.028)	8.397E−15	0.522	5.478e−41 (strongly rejected)
2017–2023 | Asturias	0.677*** (0.083)	8.157E−12	−0.186*** (0.024)	4.594E−11	0.459	1.355e−45 (strongly rejected)
2017–2023 | Baleares	1.250*** (0.069)	9.052E−40	−0.282*** (0.017)	4.104E−37	0.652	4.353e−88 (strongly rejected)
2017–2023 | Canarias	1.534*** (0.068)	3.048E−57	−0.325*** (0.015)	1.501E−53	0.687	1.649e−106 (strongly rejected)
2017–2023 | Cantabria	0.251*** (0.069)	0.001457	-0.096*** (0.028)	0.002405	0.336	1.416e−20 (strongly rejected)
2017–2023 | Castilla y León	1.087*** (0.067)	3.226E−33	−0.254*** (0.017)	6.435E−31	0.638	7.949e−82 (strongly rejected)
2017–2023 | Castilla-La Mancha	1.252*** (0.094)	2.661E−22	−0.325*** (0.026)	1.102E−20	0.651	2.098e−41 (strongly rejected)
2017–2023 | Cataluña	2.583*** (0.047)	0	−0.407*** (0.008)	2.279E−293	0.72	0.000e+00 (strongly rejected)
2017–2023 | Comunidad Valenciana	2.065*** (0.055)	1,249E−147	−0.369*** (0.010)	4.248E−139	0.723	2.448e−232 (strongly rejected)
2017–2023 | Extremadura	0.425*** (0.076)	5.784E−07	−0.123*** (0.023)	0.000001432	0.323	9.329e−44 (strongly rejected)

Table 4 Zipf´s law regression results by regime and region

Regime | Region	δ	pvalue	α	pvalue	R^2	*p* value Zipf´s law
2017–2023 | Galicia	1.834*** (0.065)	7.795E−85	−0.361*** (0.014)	1.193E−79	0.706	8.755e−138 (strongly rejected)
2017–2023 | Madrid	3.051*** (0.043)	0	−0.451*** (0.007)	0	0.745	0.000e+00 (strongly rejected)
2017–2023 | Murcia	1.036*** (0.069)	3.337E−26	−0.262*** (0.019)	3.24E−24	0.68	8.996e−59 (strongly rejected)
2017–2023 | Navarra	0.649*** (0.124)	0.00001774	−0.237*** (0.048)	0.00004304	0.481	7.552e−15 (strongly rejected)
2017–2023 | País Vasco	1.270*** (0.066)	2.78E−46	−0.275*** (0.015)	2.582E−43	0.64	4.178e−107 (strongly rejected)

Standard errors are reported in parentheses. Statistical significance of the estimated coefficients is indicated using asterisks, where *** denotes significance at the 1% level, ** at the 5% level, and * at the 10% level. The interpretation of the *p*-values follows these guidelines: if the p-value is less than 0.01, the null hypothesis is considered strongly rejected; if the p-value lies between 0.01 and 0.05, it is rejected; if it falls between 0.05 and 0.10, the rejection is considered marginal; and if it exceeds 0.10, the null hypothesis is not rejected. In the context of Zipf's Law, the null hypothesis (H_0:$\alpha = -1$)

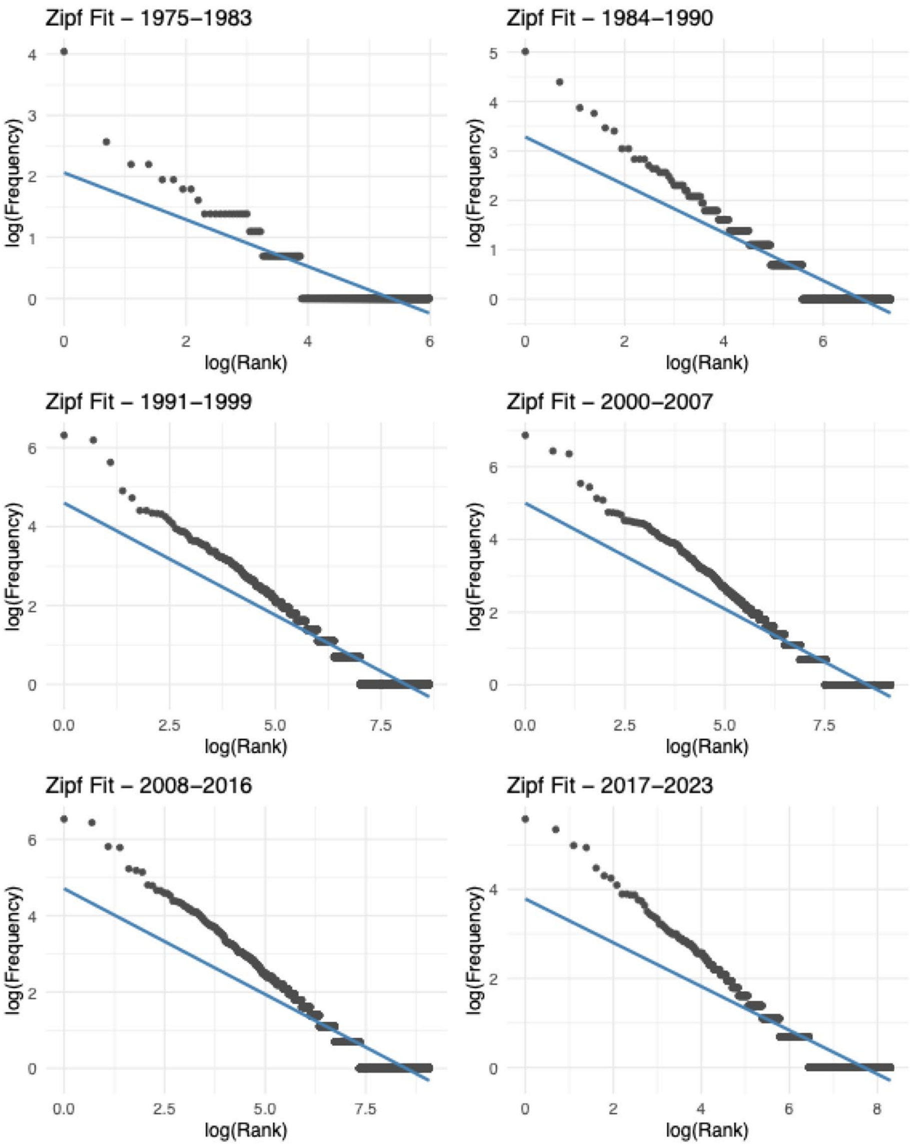

Fig. 7 Log-log plots by regime

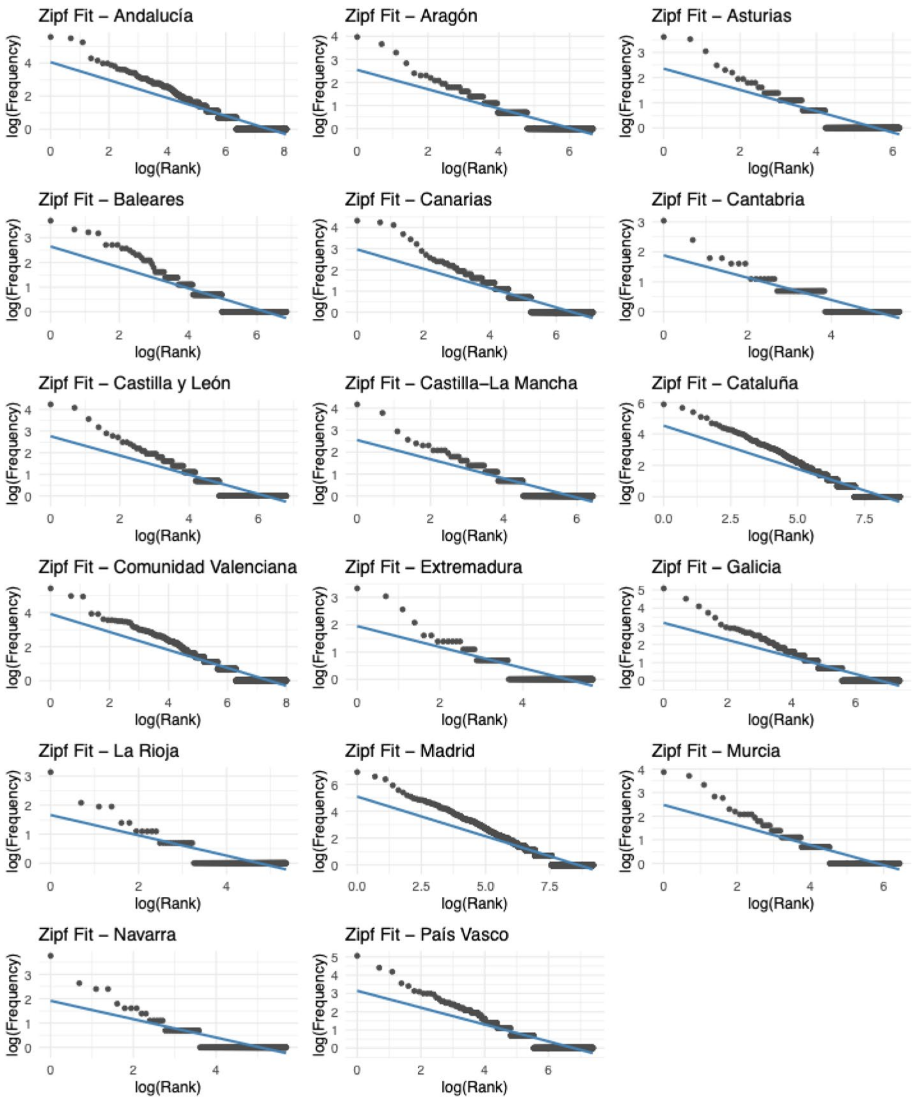

Fig. 8 Log-log plots by region

Annex B. Robustness analyses of regional Zipf coefficients

Within-region robustness

The first robustness exercise evaluates the internal stability of the estimated Zipf coefficients within each autonomous community. Its objective is to determine whether the regional estimates are driven by a small number of firms or by highly influential lexical observations.

For each of the 17 autonomous communities, we generated 500 random subsamples containing 80% of the firms in the corresponding regional sample. The Zipf coefficient was re-estimated for every subsample using exactly the same estimation procedure as in the main analysis. The resulting bootstrap distribution allows us to evaluate the stability of each regional coefficient under repeated random resampling while preserving the original characteristics of each regional dataset.

The results indicate a very high degree of internal robustness. Across the 17 regions, the mean absolute difference between the original coefficient and the average bootstrap estimate is only 0.017, while the maximum observed difference is 0.025. Likewise, the agreement between the original coefficients and the bootstrap estimates is almost perfect (Pearson's $r = 0.999$, $p < 0.001$), and the regional ordering remains essentially unchanged (Spearman's $\rho = 0.988$, $p < 0.001$).
These findings indicate that the estimated regional Zipf coefficients are highly stable and are not driven by a small number of influential firms or lexical observations. However, because this procedure preserves the original sample size of each region, it does not address whether comparisons across regions may be affected by differences in regional sample size (Figs. 9, 10 and Table 5).

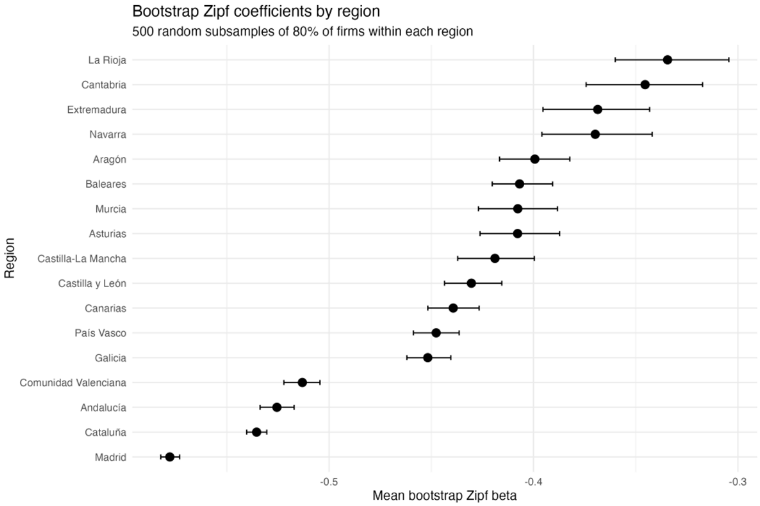

Fig. 9 Bootstrap confidence intervals for each region

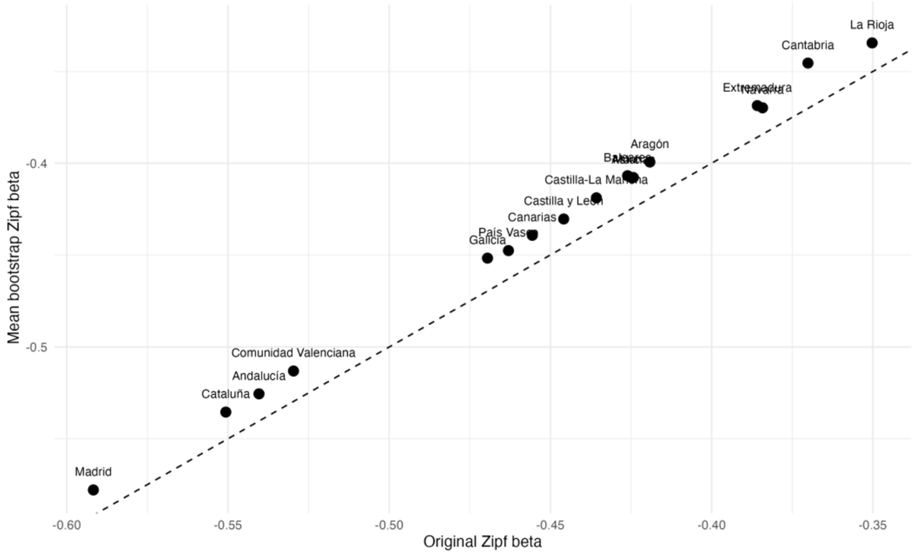

Fig. 10 Original and bootstrap Zip coefficients

Table 5 Summary of bootstrap robustness analysis

Statistic	Value
Number of regions	17
Bootstrap replications per region	500
Subsample proportion	80%
Mean absolute difference between original and bootstrap coefficients	0.017
Maximum absolute difference	0.025
Pearson correlation (original vs. bootstrap coefficients)	0.999***
Spearman rank correlation	0.988***

***$p < 0.001$

Between-region robustness under equal sample sizes

Although the bootstrap analysis demonstrates the internal stability of each regional estimate, it does not evaluate whether differences between regions are influenced by unequal sample sizes. Because the number of firms varies substantially across autonomous communities, we conducted an additional Monte Carlo robustness analysis in which all regions were compared using identical sample sizes.

The smallest regional sample in the database contains 115 firms. Accordingly, for each Monte Carlo replication we randomly selected 115 firms from every region without replacement and re-estimated all regional Zipf coefficients. This procedure was repeated 1000 times. To examine the sensitivity of the results to sample size, the same exercise was also performed using equal sample sizes of 25, 50 and 100 firms per region.

The first objective of the equal-sample analysis is to determine whether unequal regional sample sizes artificially inflate the apparent degree of regional heterogeneity.

Figure 11 shows that the dispersion of regional Zipf coefficients decreases steadily as the common sample size increases. Correspondingly, Table 6 reports that the mean between-region range falls from 0.245 when only 25 firms are sampled per region to 0.137 when all regions are compared using 115 firms, representing a reduction of approximately 44%. Similar reductions are observed for the standard deviation and the interquartile range.

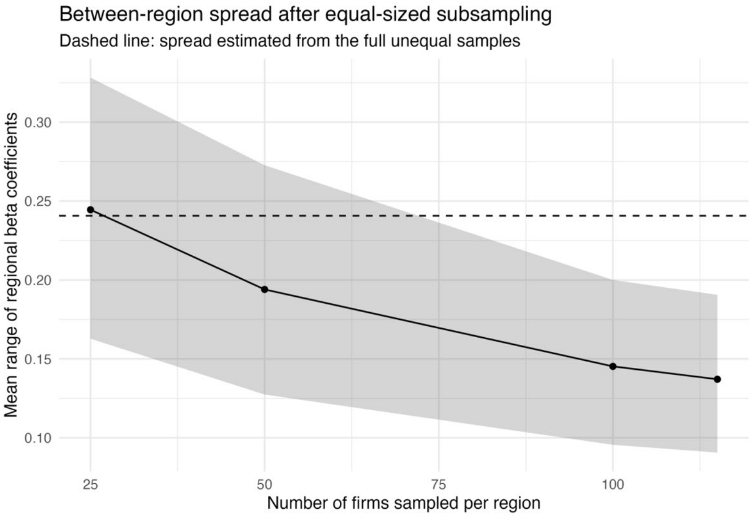

Fig. 11 Between-region spread after equal-sized subsampling. The dashed line reports the spread obtained from the full unequal samples

Table 6 Effect of common sample size on between-region dispersion

Firms per region	Mean SD	Mean range	95% MC interval for range	Mean IQR
25	0.069	0.245	[0.163, 0.328]	0.089
50	0.054	0.194	[0.127, 0.273]	0.068
100	0.041	0.145	[0.096, 0.200]	0.054
115	0.039	0.137	[0.091, 0.191]	0.053

These results indicate that unequal or small regional samples exaggerate the apparent degree of regional separation. Nevertheless, regional dispersion does not disappear after sample sizes are equalised.

Equalising regional sample sizes also modifies the estimated coefficients themselves.

Figure 12 compares the coefficients obtained from the original unequal regional samples with the Monte Carlo means estimated from equal-sized samples. Several regions exhibit noticeable shifts after controlling for sample size, indicating that part of the original regional ordering was influenced by differences in the number of firms available in each autonomous community.

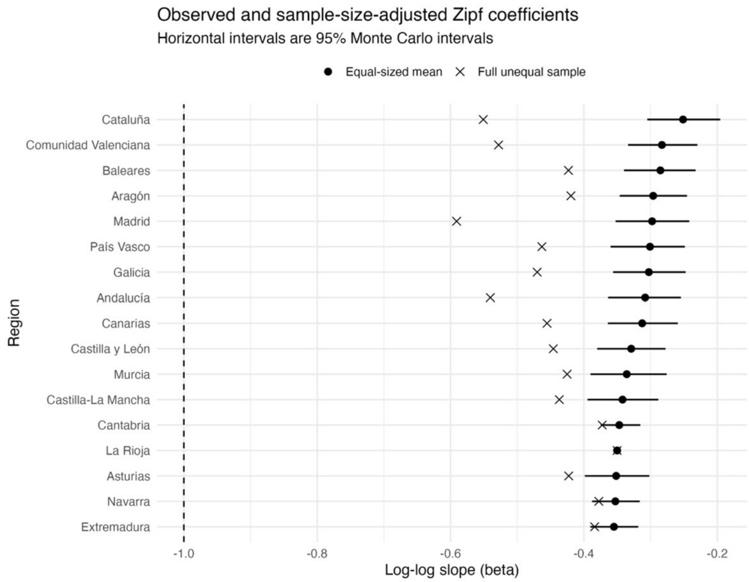

Fig. 12 Observed coefficients versus equal-sized Monte Carlo means and 95% intervals. The shifts show why the full-sample regional ranking should not be treated as invariant

Table 7 summarises these changes quantitatively. The mean absolute difference between the observed coefficients and the equal-sized Monte Carlo estimates is 0.133, with a maximum absolute difference of 0.300. Likewise, the correspondence between the original and equal-sized estimates is substantially weaker than that observed in the within-region bootstrap analysis, confirming that unequal sample sizes affect the precise ordering of regional coefficients.

Table 7 Summary of the Monte Carlo equal-sample robustness analysis

Statistic	Result
Regions	17
Monte Carlo replications per specification	1.000
Common sample size in main comparison	115 firms per region
Mean absolute change: observed vs equal-sized beta	0.133
Maximum absolute change	0.300
Pearson correlation: observed vs equal-sized beta	-0.736 ($p=0.0008$)
Spearman rank correlation	-0.694 ($p=0.0020$)
Regional coefficients rejecting beta $=-1$	100% of valid replications
Pooled interaction tests with $p<0.05$	100% of 1.000 replications
Median pooled-interaction p-value	1.66e$-$13
Pairwise intervals excluding zero	17 of 136 (12.5%)

Consequently, the regional ranking reported in the main text should be interpreted cautiously and not as an invariant ordering of regional naming systems.

To evaluate the influence of the common sample size on the estimated coefficients, we repeated the Monte Carlo analysis using equal samples of 25, 50, 100 and 115 firms per region.

Figure 13 shows that the regional coefficients converge progressively as the number of firms increases. Once approximately 100 firms per region are available, additional observations produce relatively small changes in the estimated coefficients, indicating that the estimates become increasingly stable.

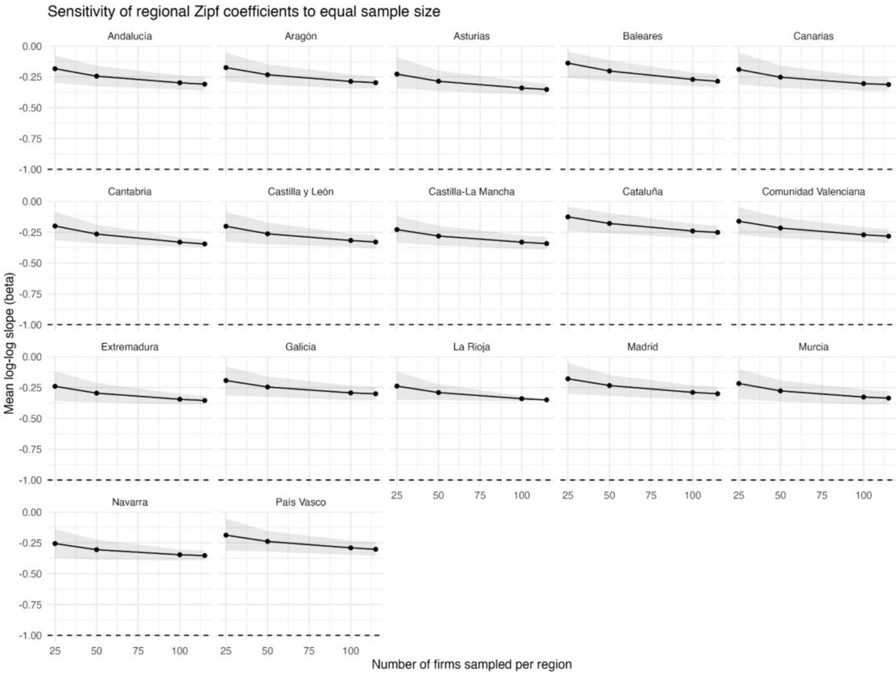

Fig. 13 Sensitivity of regional coefficients to the common number of firms sampled per region

Finally, we evaluate whether regional heterogeneity persists after controlling for sample size.

Figure 14 presents the Monte Carlo distributions obtained from 1000 equal-sized samples of 115 firms per region. Although the regional distributions overlap more than in the original analysis, several remain clearly separated, suggesting that regional differences are not entirely explained by unequal sample sizes.

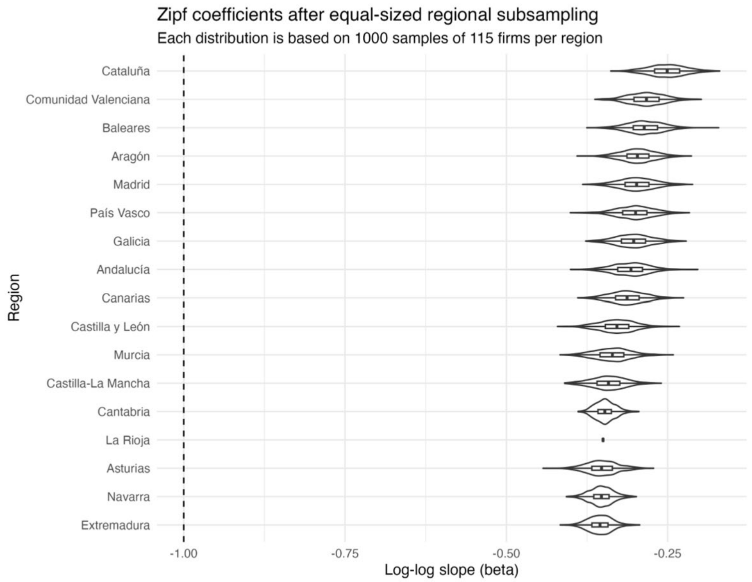

Fig. 14 Monte Carlo distributions obtained from 1000 equal-sized samples of 115 firms per region

This conclusion is supported by formal comparisons of the regional slopes. Across all 1000 Monte Carlo replications, pooled regressions containing region-by-log-rank interaction terms rejected the null hypothesis of homogeneous regional slopes (median $p = 1.66 \times 10^{-13}$; 97.5th-percentile $p = 1.3 \times 10^{-5}$). In addition, Monte Carlo intervals excluded zero for 17 of the 136 pairwise regional differences, showing that residual heterogeneity is concentrated in a subset of regional comparisons rather than being universal. As a separate result, all regional coefficients also remained significantly different from the theoretical benchmark of -1 across the valid replications.

Overall interpretation

The two robustness analyses reported in this appendix address complementary methodological questions. The bootstrap analysis demonstrates that the estimated regional Zipf coefficients are internally stable and are not driven by a small number of influential observations within each regional sample. The equal-sample Monte Carlo analysis shows that unequal regional sample sizes account for a substantial proportion of the originally observed regional dispersion and influence the ordering of several regional coefficients.

Importantly, however, controlling for sample size does not eliminate regional heterogeneity. Although equalising regional sample sizes substantially reduces the magnitude of the observed dispersion, statistically significant regional differences remain. Accordingly, the regional results reported in the main text should be interpreted as evidence of residual regional heterogeneity after controlling for sample size, rather than as a definitive ranking of regional naming systems. This combined robustness assessment therefore strengthens the substantive conclusion that regional variation exists, while providing a more conservative and statistically defensible interpretation of its magnitude.

Author contributions J.F.-V., B.V.-L.H., A.G., and J.M.M.-Á. contributed equally to the conception, writing, and revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The Authors declare that there is no conflict of interest.

References

Ahn, J., La Ferle, C.L.: Enhancing recall and recognition for brand names and body copy: a mixed-language approach. J. Advert. **37**(3), 107–117 (2008). https://doi.org/10.2753/JOA0091-3367370308

Alserhan, B.A., Alserhan, Z.A.: Naming businesses: names as drivers of brand value. Compet. Rev. **22**(4), 329–342 (2012). https://doi.org/10.1108/10595421211247169

Amore, M.D., Epure, M., Garofalo, O.: Organizational identity and performance: an inquiry into nonconforming company names. Long Range Plann. **57**(1), 102396 (2024). https://doi.org/10.1016/j.lrp.2023.102396

Bai, J., Perron, P.: Estimating and testing linear models with multiple structural changes. Econometrica (1998). https://doi.org/10.2307/2998540

Bai, J., Perron, P.: Computation and analysis of multiple structural change models. J. Appl. Econom. **18**(1), 1–22 (2003). https://doi.org/10.1002/jae.659

Baker, W.E., Honea, H., Russell, C.A.: Do not wait to reveal the brand name: the effect of brand-name placement on television advertising effectiveness. J. Advert. **33**(3), 77–85 (2004). https://doi.org/10.1080/00913367.2004.10639170

Buitrago, S., Duque, P.L., Robledo, S.: Branding corporativo: una revisión bibliográfica. Económicas CUC **41**(1), 143–162 (2019). https://doi.org/10.17981/econcuc.41.1.2020.Org.1

Chan, A.K., Huang, Y.Y.: Brand naming in China: a linguistic approach. Mark. Intell. Plan. **15**(5), 227–234 (1997). https://doi.org/10.1108/02634509710177297

Chung, J., Eoh, J.: Naming strategies as a tool for communication: application to movie titles. Int. J. Advert. **38**(8), 1139–1152 (2019). https://doi.org/10.1080/02650487.2019.1593719

Corral, Á., Boleda, G., Ferrer-i-Cancho, R.: Zipf's law for word frequencies: word forms versus lemmas in long texts. PLoS ONE **10**(7), e0129031 (2015). https://doi.org/10.1371/journal.pone.0129031

Costa, J.: Construcción y gestión estratégica de la marca: Modelo MasterBrand. Luciérnaga Comun. **4**(8), 20–25 (2012)

Deephouse, D.L.: To be different, or to be the same? It's a question (and theory) of strategic balance. Strateg. Manag. J. **20**(2), 147–166 (1999). https://doi.org/10.1002/(SICI)1097-0266(199902)20:2%3c147::AID-SMJ11%3e3.0.CO;2-Q

Esteban-Rojo, F., Galiano, A., Martín-Álvarez, J.M., Vázquez-La Hoz, B.: Testing Zipf's and Gibrat's laws in the Spanish university system: evidence from a decade of institutional transformation. Stud. High. Educ. (2026). https://doi.org/10.1080/03075079.2026.2680550

Ferrari, M., Pesantez-Coronel, P., Ugalde, C.: Proceso de naming: teoría vs. práctica. Pensar Public. Revista Int. Invest. Public. **14**(1), 13–27 (2020). https://doi.org/10.5209/pepu.67142

Fox, R.: Naming an organisation: a (socio)linguistic perspective. Corp. Commun. **16**(1), 65–80 (2011). https://doi.org/10.1108/13563281111100980

Glynn, M.A., Abzug, R.: Institutionalizing identity: symbolic isomorphism and organizational names. Acad. Manag. J. **45**(1), 267–280 (2002). https://doi.org/10.5465/3069296

Gürtler, S., Miller, B.: Branded for survival: naming effects on the life expectancy of new companies. Mark. ZFP J. Res. Manag. **44**(2), 44–58 (2022). https://doi.org/10.15358/0344-1369-2022-2-44

Gutiérrez-Chico, F., González-Fuente, I., Pulleiro-Méndez, C.: La diplomacia del naming en el fútbol: El caso de ESPAÑA-Kosovo (2020–2024). Hist. Comun. Soc. **29**(2), 325–334 (2024). https://doi.org/10.5209/hics.98675

Heredero Díaz, O., Chaves Martín, M.Á.: Las asociaciones “marca producto” y “marca ciudad” como estrategia de “city branding”: una aproximación a los casos de Nueva York, París y Londres. Rev_arab **15**(2), 63 (2015). https://doi.org/10.5209/rev_ARAB.2015.v15.n2.47857

Huang, Y.Y., Chan, A.K.K.: Chinese brand naming: from general principles to specific rules. Int. J. Advert. **16**(4), 320–335 (1997). https://doi.org/10.1080/02650487.1997.11104699

Johnson, R.C.: Blacks and women: naming American hostages released in Iran. J. Commun. **30**(3), 58–63 (1980). https://doi.org/10.1111/j.1460-2466.1980.tb01992.x

Leong, S.M., Ang, S.H., Tham, L.L.: Increasing brand name recall in print advertising among Asian consumers. J. Advert. **25**(2), 65–81 (1996). https://doi.org/10.1080/00913367.1996.10673500

Lowrey, T.M., Shrum, L.J., Dubitsky, T.M.: The relation between brand-name linguistic characteristics and brand-name memory. J. Advert. **32**(3), 7–17 (2003). https://doi.org/10.1080/00913367.2003.10639137

Martínez, J.A., Muñoz, J.: Análisis y evolución histórica de los nombres de marcas de zapatillas deportivas. RICYDE. Rev. Int. Cienc. Deporte **37**(10), 235–263 (2014). https://doi.org/10.5232/ricyde2014.03705

Matus-Mendoza, M.: The interface of Latinos, language and space in restaurant naming in Philadelphia. Latino Stud. **23**, 105–136 (2025). https://doi.org/10.1057/s41276-024-00509-8

Maza-Maza, R.L., Guaman-Guaman, B.D., Benítez-Chávez, A.M., Solis-Mairongo, G.: Importancia del branding para consolidar el posicionamiento de una Marca corporativa. Killkana Soc. **4**(2), 9–18 (2020). https://doi.org/10.26871/killkanasocial.v4i2.459

McDevitt, R.C.: “A” business by any other name: firm name choice as a signal of firm quality. J. Polit. Econ. **122**(4), 909–944 (2014). https://doi.org/10.1086/676333

Olivares-Delgado, F., Pinillos-Laffón, A., Benlloch-Osuna, M.T.: An approach to patronymic names as a resource for familiness and as a variable for family business identification. Eur. J. Fam. Bus. (2016). https://doi.org/10.1016/j.ejfb.2016.06.001

Piantadosi, S.T.: Zipf’s word frequency law in natural language: a critical review and future directions. Psychon. Bull. Rev. **21**, 1112–1130 (2014). https://doi.org/10.3758/s13423-014-0585-6

Pinillos-Laffón, A.: Nombres corporativos en empresas de carácter familiar en Alicante (España). Criterios denominativos. Rev. Mediterr. Comun. **6**(2), 149–163 (2015). https://doi.org/10.14198/MEDCOM2015.6.2.07

Pinillos-Laffón, A., Olivares Delgado, F., Rodríguez Valero, D.: El nombre de la marca corporativa. Una taxonomía de los nombres de empresa familiar en España. Rev. Lat. Comun. Soc. **71**, 750–774 (2016). https://doi.org/10.4185/RLCS-2016-1119

Team, R.C.: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2021)

Río, J.G., Canellas, O.A., Albiñana, B.J., Royo, T.M.: El nombre de marca: interrelación de factores lingüísticos y corporativos. Rev. Lingüíst. Leng. Apl. **6**, 181–193 (2011). https://doi.org/10.4995/rlyla.2011.902

Salazar, A.: El naming de las marcas. Una breve reflexión. Prop. Intelect. **22**, 345–348 (2020). https://doi.org/10.53766/PI/2021.22.09

Schmeltz, L., Kjeldsen, A.K.: Naming as strategic communication: understanding corporate name change through an integrative framework encompassing branding, identity and institutional theory. Int. J. Strateg. Commun. **10**(4), 309–331 (2016). https://doi.org/10.1080/1553118X.2016.1179194

Shamsollahi, A., Amirshahi, M., Ghaffari, F.: Brand name recall: a study of the effects of word types, processing, and involvement levels. J. Mark. Commun. **23**(3), 240–259 (2014). https://doi.org/10.1080/13527266.2014.930068

Shipley, D., Hooky, G.J., Wallace, S.: The brand name development process. Int. J. Advert. **7**(3), 253–266 (1988). https://doi.org/10.1080/02650487.1988.11107064

Sicilia-Garcia, E.I., Ming, J., Smith, F.J.: Extension of Zipf’s law to words and phrases. In: COLING 2002: The 19th International Conference on Computational Linguistics. (2002)

Sigurd, B., Eeg-Olofsson, M., Van Weijer, J.: Word length, sentence length and frequency–Zipf revisited. Stud. Linguist. **58**(1), 37–52 (2004). https://doi.org/10.1111/j.0039-3193.2004.00109.x

Thurner, S., Hanel, R., Liu, B., Corominas-Murtra, B.: Understanding Zipf’s law of word frequencies through sample-space collapse in sentence formation. J. r. Soc. Interface. **12**(108), 20150330 (2015). https://doi.org/10.1098/rsif.2015.0330

Tolochko, P., Balluff, P., Bernhard, J., Galyga, S., Lebernegg, N.S., Boomgaarden, H.G.: What's in a name? The effect of named entities on topic modelling interpretability. Commun. Methods Meas. **18**(4), 349–370 (2024). https://doi.org/10.1080/19312458.2024.2302120

Usunier, J.C., Shaner, J.: Using linguistics for creating better international brand names. J. Mark. Commun. **8**(4), 211–228 (2002). https://doi.org/10.1080/13527260210146000

Wickham, H., Henry, L.: purrr: functional programming tools. R package version 1.1.0, https://purrr.tidyverse.org. (2025)

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D.: dplyr: a grammar of data manipulation. R package version 1.1.4, https://dplyr.tidyverse.org. (2025)

Yan, R.N., Hyllegard, K.H., Blaesi, L.F.: Marketing eco-fashion: the influence of brand name and message explicitness. J. Mark. Commun. **18**(2), 151–168 (2010). https://doi.org/10.1080/13527266.2010.490420

Yang, J., Li, Y.: An integrated review on optimal distinctiveness: multidimensionality and contingency. Manag. Rev. Q. (2025). https://doi.org/10.1007/s11301-025-00503-x

Zeileis, A.: Implementing a class of structural change tests: an econometric computing approach. Comput. Stat. Data Anal. **50**(11), 2987–3008 (2006). https://doi.org/10.1016/j.csda.2005.07.001

Zhao, E.Y., Fisher, G., Lounsbury, M., Miller, D.: Optimal distinctiveness: broadening the interface between institutional theory and strategic management. Strateg. Manag. J. **38**(1), 93–113 (2017). https://doi.org/10.1108/JKM-07-2017-0260

Zipf, G.K.: The meaning-frequency relationship of words. J. Gen. Psychol. **33**(2), 251–256 (1945). https://doi.org/10.1080/00221309.1945.10544509

Authors and Affiliations

Jessica Fernández-Vázquez[1] · **Brenda Vázquez-La Hoz**[1] · **Aida Galiano**[1] · **Juan Manuel Martín-Álvarez**[1]

✉ Juan Manuel Martín-Álvarez
juanmanuel.martin@unir.net

Jessica Fernández-Vázquez
jessica.fernandez@unir.net

Brenda Vázquez-La Hoz
brenda.vazquez@unir.net

Aida Galiano
aida.galiano@unir.net

1 Universidad Internacional de La Rioja, Logroño, Spain