CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Language in the age of AI technology: From human to non-human authenticity, from public governance to privatised assemblages

Iker Erdocia[1] (ID), Britta Schneider[2], Bettina Migge[3]

[1]Dublin City University, Ireland; [2]Europa-Universität Viadrina Frankfurt (Oder), Germany and [3]University College Dublin, Ireland
Corresponding author: Iker Erdocia; Email: Iker.erdocia@dcu.ie

Abstract

Large language models based on machine-learning technologies are reshaping linguistic contexts and understandings of language. We explore these reconfigurations by investigating discursive positionings of traditional institutional guardians of power in language in response to these changes. Focusing on the discourse of the Real Academia Española (RAE), we show how RAE's social functions, ways of asserting authority, and the nature, function, and rightful ownership of RAE's standard language have been reimagined. Crucially, RAE presents itself as a professional soft power that protects the rights of Spanish speakers. Drawing on tropes of authenticity and endangerment, it conceptualises language generated by machine-learning technologies as inauthentic and as destroying the authentic Spanish of human Spanish speakers. We argue that these discourses are indexical of a power struggle where the role of traditional language norming institutions is reshaped in the face of sociotechnical innovations that are in the hands of global commercial companies. (Standard language, AI technology, language academies, authority in language, big tech, Real Academia Española)*

Introduction

Machine-learning technologies like large language models have led to changes in many areas of social life, leading to new practices and understandings. Sociolinguistics investigates the linguistic practices of technology users and their social indexicalities (Blommaert 2015; Androutsopoulos 2021), but there is little research on how linguistic contexts and understandings of language practices are being reshaped by algorithmic machine-learning tools. This is due to difficulties accessing underlying algorithmic processes and an absence of methods for studying increasingly intertwined human–machine-generated practices and the somewhat invisible processes of dissemination of these practices (Kelly-Holmes 2022). One way to explore these reconfigurations is to investigate the discursive positionings and negotiations of traditional language authorities in response to these changes.

The discourses of language institutions associated with standard languages are ideal objects for such an investigation. These institutions are ubiquitous in European national contexts. State and public institutions and lay people have traditionally recognised them as the rightful social authorities in matters of what constitutes 'correct' and prestigious varieties of language. However, their role in society and that of standard languages have been changing. Research has investigated standard language change (Vandenbussche 2022) including processes of destandardisation and restandardisation (Ayres-Bennett & Bellamy 2021), and language institutions' impact on the development of standard languages in different national contexts (McLelland 2021). However, we do not know how traditional guardians of authority in standard languages fare vis-à-vis language generating machine-learning technologies developed and owned by big tech companies (Google, Meta, Amazon, Apple),[1] nor what new conceptualisations and discourses about standard languages are emerging due to such technologies.

In this article, we inspect contemporary language ideological debates on language norming and standardisation and language-norming institutions' own understanding of their functions in the context of the establishment and distribution of large language models, which imply a rising dominance of big tech companies in national public spheres. Analysing the discourse of the Real Academia Española 'Royal Spanish Academy', the traditional language norming institution of Spain, using a critical discourse analysis (CDA) approach, this study investigates how European institutional guardians of power in language negotiate their position and that of 'their' standard language and critique new public language actors, namely commercial digital industries. We demonstrate that RAE has redefined its social function, its approach to asserting its authority, and its understanding of what it considers to be the nature and ownership of 'correct' language. We argue that these changes are indexical of the transformation of power relations in society, characterised by a progressive weakening of state institutions driven by the growing dominance of for-profit, private entities.

The article is organised into five sections. Following this introductory section, we present the background for this study, discussing the notion of standard language ideology, language models as sociotechnical power assemblages and their relationship to language and society, and RAE. We then present the research methods and analyse RAE's recent discourse on language and authority. The last section summarises the findings, exploring their implications for the place of traditional language authorities and standard languages and the relationship among (standard) languages, machine-generated language, and societies in the digital era.

Language authority: National publics, media technologies, and RAE

This section briefly discusses the factors that have traditionally contributed to the historical formation of linguistic authority in national public contexts and the role and function of media technologies and language norming institutions such as RAE in this context before considering how discourses are reshaped in the age of digital capitalism.

Standard language ideology

Standard language ideologies dominate social and educational discussions about language, entailing the normative idea that there is one form of language that is 'better' than others. Although an important tenet in Western countries, 'a cursory inspection of the facts will reveal that these standard varieties are nothing more than the social dialect of the dominant class' (Guy 2011:162). There are various strategies of legitimation and among them is the belief that this dominant variety represents modernist ideals of regularity, clarity, purity, and rationality (Bauman & Briggs 2003). Discourses of cultural nostalgia that construct older forms as inherently 'good' and refined due to their relationship with a mythical past (Durrell 2000) and ideologies of stability in which change is understood as 'bad' are also common (see Hickey 2012).

Discourses about 'good language' rely on an understanding that there is a social space beyond private face-to-face relationships in families or local communities for which a shared and ordered language is necessary. Sociological accounts (e.g. Habermas 1962/1989) refer to them as *public spaces* and define them as a 'communicative field in which rational actors can achieve agreement on grounds of rational interaction' (Heyd & Schneider 2019:5). But the assumption that public interaction is founded on orderliness and rationality has been strongly critiqued (Gardiner 2004). With changes in public spheres in the digital age, conflicting discourses on who is supposed to define public norms of linguistic conduct have emerged in many settings.

In sociolinguistics, standard language is not seen as linguistically 'better' and the attribution of prestige to some forms of speaking and writing is understood as an outcome of sociohistorical developments. In Europe, 'the idea of a single nation and a single form of language emanating from its centre is a predominant theme in eighteenth century [English] writings ... an integral part of the national language complex' (Hickey 2012:7).

In the history of European modernity, standard languages play an important role in the construction of legitimacy and authority of specific groups in nation states. Thus, the establishment of positions of public authority and the establishment of language norms mutually relate to each other. Gal & Woolard (1995) describe standard languages as the cultural construction of a (supposedly) neutral 'voice from nowhere' although the process of norming was sometimes based on the ideas of a single person (Hickey 2012:13).

In order to establish the image of such an apparently 'neutral' way of speaking, discourses of both ANONYMITY and AUTHENTICITY come into play (Gal & Woolard 1995:134). Elite ways of speaking are attributed with the authority of anonymity—they are authoritative because they are constructed as the language of everyone, of 'no-one-in-particular', based on an image of 'aperspectival objectivity' (Gal & Woolard 1995:134)—and are discursively constructed as a neutral code of communication that represents everyone within public space: 'They are positioned as universally open and available to all in a society, if only, as Michael Silverstein (1996) reminds us, we are good enough and smart enough to avail ourselves of them' (Woolard 2016:25). At the same time, the authority of anonymity exists simultaneously with the authority of authenticity. The standard language is constructed as legitimate because it is neutral and understood as 'authentically' expressing

membership and belonging to a national group (Woolard 2016:25). In some cases, standard languages are used by supranational groups, as in the case of English or Spanish, but these are still understood as divided according to the logic of nations (e.g. CANADIAN English, MEXICAN Spanish, etc.) and very often, the place of historical origin (e.g. England or Spain) still holds considerable influence over constructions of legitimate language elsewhere (Paffey 2021). Such nationalised (standard) languages coexist in a hegemonic relationship with 'deviating' (private non-standard) language practices. And only those who conform to the ideals of the standard language are able to participate in the powerful discourses that bring into being public spaces.

The role of media technology assemblages in framing and shaping public language

The construction and distribution of language policies rely on the existence of writing technologies and print industries—without writing and homogenised print, linguistic homogeneity, standard languages and the establishment of language policies would hardly be conceivable (Ong 1982). Dictionaries and grammars are tangible norm-givers that inform about the current norms and are typically understood as unquestioned linguistic authorities in national contexts, supported and often created by language academies like RAE. However, in recent years there have been heated debates about public language norms and language authorities, such as RAE, and 'there is, indeed, considerable evidence of the diminishing influence of traditional authorities who promote the ideology of the standard' (Ayres-Bennett & Bellamy 2021:18). Public and lay discourses often point to changing media ecologies such as privately owned social-media platforms to explain the diminishing influence of traditional language authorities.

Digital media have different affordances compared to more traditional media in terms of the ways in which individuals can participate and engage in publicly visible discourse. This may influence the discussions of traditional authorities when, for example, decisions by language-norming institutions are heatedly debated in social media contexts. Individual users can question authoritative stances, making it visible that traditional and formerly unmarked hegemonic institutions also hold a social position, rather than being understood as a 'voice from nowhere'. Paffey (2021:252–56) discusses an example where RAE's decisions on the definitions of meanings of words related to gender became the subject of a Twitter (X) debate. The Twitter account of the institution encourages interactive, participatory engagement and 'followers of the RAE Twitter feed are likely—and expected—to respond to the specific norms contained in the publications' (Paffey 2021:253). In some cases, RAE changed lexical entries in dictionaries, possibly as a reaction to such debates. The fact that decisions to update can involve public digital debate and are linked to social and political stances has recently become much more obvious; previously such deliberations took place behind closed doors and the decisions on language were presented as anonymous and objective 'truth'.

At the same time, digital media platforms are a hybrid of private and public. The general public can participate in discourses that are in principle available to everyone. Yet, the owners of the infrastructure are private companies. This means that

even though the global community of Spanish speakers can interact with author-itative institutions like RAE on Twitter and similar, it is predominantly US-based tech companies that provide the means for this interaction. Social media platforms are predefined spaces that, 'at least on the technical side, do not allow participants to develop shared rules of conduct and communication' (Heyd & Schneider 2019:8). Ultimately, it is the owners of platforms who decide what is seen and what is not seen in digital publics via processes of content moderation (Gorwa, Binns, & Katzenbach 2020). Platform infrastructures—encouraging users to attract attention in the form of shares, likes, views, and clicks (Maly 2021)—are built on an underlying belief that digital data represents reality. Aligned with quantitative economic ideals, they reflect a desire to collect more and more data and to track and monitor users (Bode & Goodlad 2023).

The data that speakers produce on these platforms and in online spaces serve as central input to machine-learning language models which are highly complex sociotechnical assemblages (Pennycook 2024). They are based on human language that is transformed (e.g. books) or appears (e.g. social media) as machine-readable data. The data is then curated by companies, and algorithmic design defines which patterns are reproduced (Schneider 2022). In this sense, the models are an outcome of collective intellectual work (Pasquinelli 2023) that has provided the data, which is exploited by those who use the data to build and sell machine-learning technologies to customers. In addition, companies use the data collected from individuals for (non-public) practices of marketing and surveillance (Zuboff 2019).

Such capitalist developments cannot be seen as isolated from political contexts, since states regulate the services of platform owners,[2] including how and which data can be collected and for which purposes. Additionally, the development of machine-learning technologies has historically been funded by the US military (Crawford 2021). Machine-learning technologies that generate language can there-fore be understood as an interactional assemblage, in which companies who have privatised data (for a critical discussion of data epistemologies, see Bode & Goodlad 2023) and who have the material means to build the models are in a hegemonic posi-tion. It is them who define what goes into the models and who has which access. Despite the entanglement of companies with states, we understand companies as private actors and states as public, given that state regulations (at least currently and in the Western world) are sanctioned in democratic political processes and company decisions are not. This distinction between private and public plays an important role in our later analysis.

Although these globally acting commercial platform providers are not per se interested in language, they continuously collect language data and partici-pate in discussions about language, and their widely used language-based tools impact on standard language ideologies. Companies' conceptualisations of lan-guage are based on the idea that language is machine-readable user data. They call this a 'usage-based' approach. Overall, historical user data defines the lin-guistic output of machines, and this output influences what speakers under-stand as 'correct' language. The precise working practices and data sets are in the hands of private companies and not available for public inspection or research. The technical and ideological approach of US-based tech companies to language contrasts with the discourses of traditional norming authorities

like RAE. The tensions between different constructions of authenticity and anonymity, but possibly also new lines of discursive legitimation, require empirical investigation.

Language-norming institutions: The case of RAE

In order to give national populations access to national language norms, institutions like national education systems, national mass media, or publishers are crucial. In some countries—such as Spain, France, and Italy—there are also language academies, such as the RAE. These have traditionally been exclusive institutions consisting of a select group of individuals who make authoritative decisions about language management behind closed doors. These decisions are often grounded in specific linguistic standards set out in dictionaries, grammars, and other reference works that serve as prescriptive norms for media and educational institutions. Their authority functions as a social gatekeeping mechanism that defines appropriate language use and reinforces social hierarchies and inequalities.

The Real Academia Española was established in 1713 with the mission of overseeing the Spanish language. RAE exemplifies a complex relationship between the public and private spheres. While it is a semi-public, state-sponsored institution operating under a charter granted by the Spanish Crown, RAE functions as an independent entity. It receives funding from both public and private sources, including the Spanish government and the Fundación pro RAE. This foundation, which is presided over by the Spanish king and the president of the Bank of Spain, includes members from some of the largest corporations in Spain, and carries out additional fundraising to support RAE's initiatives.

Academicians in RAE are distinguished individuals, predominantly male (the first female academician was appointed in 1978), who have traditionally come from the humanities disciplines such as philology, literature, philosophy, law, and journalism. However, more recently RAE has diversified its representation to include scholars from the sciences. This new context is reflected in our study, which particularly draws on the response of RAE's president, Santiago Muñoz Machado, a law scholar, to the inaugural speech of a newly appointed member of the academy, Asunción Gómez Pérez, a computational science scholar, in 2023.

RAE's position has transitioned from an ideology of linguistic nationalism, characterised by a standardisation approach that sought to provide a single, pure language to homogenise a multilingual nation within the boundaries of the state (Lodares 2002), to an ideology of a global language (Moreno-Fernández 2016). Standard language ideology still underpins RAE's course of action, but its current stance amounts to what Del Valle (2007:250) calls a 'moderate, almost inconspicuous, form of prescriptivism'. In its capacity as verbal hygienist (Cameron 1995), RAE exhibits an ambivalent discourse. It retains the rhetoric of linguistic conservatism, acting as a custodian of the language. By contrast, it shows a degree of broadmindedness by embracing a more inclusive and accountable approach to language management by, for example, prioritising correctness based on actual linguistic usage. Like other language management agencies (Edwards 2012), RAE has modified the way it perceives its role as a linguistic authority to accommodate the key developments of different historical periods. This adaptation involves a blend of continuity and change, consisting of two main developments, adopting a pan-Hispanic

view of language and embracing the digital world. Both of these are present in our analysis.

RAE's fixation with purism and linguistic nationalism began to change in the last decades of the twentieth century. This coincided with the consolidation of the Asociación de Academias de la Lengua Española (ASALE; 'Association of Spanish Language Academies'), which comprises twenty-three member academies, from Spain and Spanish-speaking countries in Hispanic America, as well as from the United States, the Philippines, and Equatorial Guinea. Against long-standing criticisms of Eurocentrism, in the 1990s RAE embraced a pan-Hispanic approach to both language and language policy (Del Valle 2007). This new approach entailed the recognition of Spanish as an internally variable language and the subsequent adoption of a pluricentric norm represented in the catchphrase 'unity in diversity' that encapsulates both the persistent concern about potential linguistic fragmentation and the enthusiastic embrace of diversity. Yet, despite the collaboration with its 'sister' academies in ASALE, RAE still holds a prominent, if not hegemonic, position within this association that reproduces colonial hierarchies (Del Valle 2013).

In the twenty-first century, RAE disseminates its authority with traditional printed publications and strives to communicate and reinforce their authority with online apps and, particularly, an active engagement with the public through its social media channels (Paffey 2021:246). Against accusations of elitism, these interactions in the public sphere have allowed RAE to establish a more democratic appearance, with its new approach formally complying 'with the protocols of a legitimate democracy grounded in open and rational debate' (Del Valle 2007:254). This democratic legitimacy stems from two sources: the linguistic truth claims based on the professional scrutiny of language experts (see Erdocia & Soler 2024), and the pulse-taking of the national public space by maintaining a permanent dialogue with speakers, representative institutions, and social and economic actors. In sum, RAE's interaction with the public is critical as it provides the basis for a popular legitimacy beyond the academic realm. The emergence of machine-learning technologies fundamentally altered both the dynamics of language use and the role of language academies in the human–machine era. It is, therefore, unsurprising that RAE's president describes this new period as 'challenging and exciting', remarking that RAE is 'entering a second era in its institutional life' (Muñoz Machado 2023:119).

In summary, standard languages in the age of European modernity are hegemonic constructions of authority that are intertwined with national concepts of the social and with the ability of some groups to establish power in national contexts by producing an image of their ways of talking as 'neutral' and 'professional' and appropriate for public uses. In an age of digital and machine-learning language technologies, a reconfiguration of publics is observable, and there are new discourses on the legitimacy of public language. Traditional language-norming institutions such as RAE provide particularly interesting insights into these debates and the newly emerging forms of language ideology.

Methodology and background

We take a critical discourse analysis approach in our study which analyses structural relations of domination, discrimination, and control as manifested through

language. We adopt a discourse-historical approach (Reisigl & Wodak 2016) as it allows us to explore RAE's discursive strategies and their adaptation in the context of the economic and ideological shifts brought about by developments in machine-learning language technology. Discourse is conceptualised as socially constitutive and socially conditioned: discourses shape and are shaped by the historical, sociopolitical, and ideological circumstances in which they are embedded. This approach enables us to link texts and their genres with a particular social activity (Fairclough 2015) to specific fields of action or areas of the social world 'defined by different functions of discursive practices' (Reisigl & Wodak 2016:28).

One such field of action involves the formation of public attitudes towards a matter of social interest such as the national standard language. Considering that discourses can influence and organise social practices within a field of action (Reisigl & Wodak 2016), one of the assumptions underlying our analysis is that RAE's language ideological frameworks seek to shape public opinion and the perspective of public institutions regarding the necessity of enforcing normative language. Ultimately, these ideological frameworks advocate for entrusting language academies with an instrumental role in the development of machine-learning technologies.

We understand the emergence of big tech companies as new agents in language regulation vis-à-vis RAE's long-standing linguistic authority as a matter of structural relationships of control, power, and dominance (Reisigl & Wodak 2016). We interpret RAE's discursive and institutional repositioning as part of a struggle for hegemony in technology-mediated language policy, the result of which may impact public perceptions of standardised language and the role of language academies in the twenty-first century.

Our analysis follows a two-step process (Wodak 2015) to examine RAE's discursive practices: we begin by mapping out the thematic content of texts and then move on to an in-depth analysis that examines the discursive strategies, representations, and argumentation situated within RAE's historical trajectory and the broader ideological context of digital capitalism. This prompts us to focus on a set of oppositions found in the data in the second step of our analysis. One key tension in RAE's discourse lies in problematising a distinction between the public and private spheres, given that language technologies are corporate assets beyond public institutional control. The priorities of the corporate sector are opposed to those of RAE, conceptualised as a public entity. Other binary structures include distinctions between European and US traditions in the (de)regulation of the market, anthropocentric versus machine-oriented views of language authenticity, and centralised versus decentralised approaches to linguistic standardisation.

Our methodological approach involves utilising the combination of different types of texts. We collected our material in two phases. First, we ran searches on RAE's website and in general search engines on the topic of AI (this being the most popular public term to refer to these technologies), with keyword searches for terms in Spanish such as 'RAE + *inteligencia artificial*' 'artificial intelligence', 'RAE + *sistemas digitales*' 'digital systems', and 'RAE + *procesamiento del lenguaje natural*' 'natural language processing'. This first search phase resulted in many documents with dates ranging from 2019 to 2024. After thorough scrutiny based on their relevance to our research objectives, we selected our initial set of data: one report, one press release, the content of RAE web pages, and two news pieces featuring events with public

statements from RAE's president, Santiago Muñoz Machado. This material provided us with a general, yet incomplete, overview of RAE's approach to machine-generated language. Second, to gain a more systematic understanding of the academy's discursive stance, we decided to include in our data a speech on AI delivered by RAE's president. The speech was RAE's institutional response to the inaugural address of Asunción Gómez Pérez when she was appointed to the RAE in May 2023.

To give some contextual information, this newly appointed member is the first and only 'AI expert' among the RAE academicians. In her speech, entitled *Inteligencia artificial y lengua española* 'artificial intelligence and Spanish language', Asunción Gómez Pérez presents her aim of putting machine-learning technologies at the service of the Spanish language. To do so, she explains the need to ensure reliable linguistic materials in Spanish in formats appropriate for such technologies and the crucial role that RAE can play in training these technologies in the use of normative Spanish. She notes that achieving such goals requires close collaboration among public administration, big tech companies, small and medium-sized enterprises, universities, research centres, and educational institutions. The focus of her speech is less on questioning the distinct nature of the language produced by machine-learning technologies and more on the role that the academy could play in ensuring that linguistic outputs conform to traditional norms. To mark the importance of appointing an academician with such an uncommon expertise, the RAE president himself presented the institution's traditional response to her inaugural address, a highly symbolic act as such a response from the president had not happened for the last ninety years of the academy's history. Public figures were among the audience, including the Spanish government's First Vice-President and Minister for Economic Affairs and Digital Transformation, who presided over the event.

Our analysis focusses on the president's twenty-four-page-long discourse because it provides an authoritative and systematic account of RAE's institutional stance on the challenges posed by machine-learning technologies. In fact, following his appointment as president in 2019, Santiago Muñoz Machado personally pushed for a strategic plan that led RAE to engage with such technologies. We analysed the original texts in Spanish.

Findings: RAE's discourse about language and authority in the digital age

In what follows, we situate our discussion within RAE's historical trajectory and the ideological frame of digital capitalism, prioritising extracts from the speech of RAE's president in our analysis. The presentation is organised thematically and the examples were translated by the first author after the analysis.

Old concerns in new times

As expected, RAE's position regarding the specific impact of machine-learning technologies on language usage, and more broadly the future of Spanish, draws on past 'anxieties' (Del Valle 2007) that are closely related to discourses of language endangerment (Duchêne & Heller 2007) and values such as the beauty, quality, and unity of the language. These concerns persist in contemporary discourses, now tailored to address the emerging challenges within the technological landscape. Example (1) illustrates this, drawing on the intrinsic value placed on linguistic unity

within Spanish, a foundational principle upheld by RAE that continues to under-pin many of the academy's assertions today. Going beyond the surface-level impact of machine-learning technologies on human language, RAE's president notes that there are

> (1) the purely linguistic repercussions of this use, that is, the quality and acces-sibility of the language spoken by the machines and the risk of it damaging its unity, maintained until today as one of the greatest conquests of the orderly expansion of Spanish in the world … Language is the main value of a people's culture, and Spanish is this for a community that includes almost six hundred million people. A deterioration in the quality, expressive capacity, beauty or unity of Spanish due to the developments of artificial intelligence would be a cultural injury of the first order. (Muñoz Machado 2023:123–24)

Of course, the uncritical use of expressions like 'greatest conquests' and 'orderly expansion', along with references to the vast number of Spanish speakers, to refer to a supposedly natural expansion of the language worldwide clearly overlooks Spain's colonial history.

RAE provides concrete examples of the influence of machine-learning technolo-gies on technologically mediated language usage and advances the risks they may pose for language users and professionals. For instance, RAE notes that 'there are keyboards with automatic correction systems that ignore almost twenty per cent of the words that appear in our dictionary' (Muñoz Machado 2023:128) and 'our dictio-nary has 94,400 entries, but most automatic correctors use fewer than 80,000 words from foreign dictionaries' (Muñoz Machado in Iglesias Fraga 2021), thus implying a simplification of the vocabulary available to everyday users of Spanish. Moreover, RAE expresses its concern lamentingly, noting that 'machines do not use the pan-Hispanic canon and follow the canon of Silicon Valley, which may be respectable but is different from the standardised language' (Muñoz Machado, in Lorenci 2023). Following the moderate linguistic imperialism that led to the universalisation of English due to the global trade network in the twentieth century, RAE's critique can be interpreted as accusing Silicon Valley companies of a new form of linguis-tic imperialism—this time, for not adhering to the standard form of languages other than English. This stage in the progression of language technologies has implications for standard Spanish, as the linguistic output generated by the algorithmic design of language technologies may shape speakers' perceptions of what constitutes 'correct' language, in the sense of standard language cultures (see the section on LANGUAGE AUTHORITY above).

The disruption that machine-learning technologies may cause in the way people communicate through language is not simply a matter of insufficient technological development or the inadequacy of existing corpora in Spanish. RAE warns about the possibility of a disproportionate presence in large language models of linguistic features specific to certain Spanish-speaking regions or programmers rather than others. This, argues RAE, may result in so-called digital dialects, a new type of risk to the future of the language, as they 'strain unity and lay the foundations for a fragmentation of language use that academic norms have managed to avoid for more than three hundred years' (Muñoz Machado 2023:127).

Against this backdrop of the destabilisation of traditional norms through technological means, RAE's mandate of preserving the unity of the language is sustained. This time, however, the challenges of the digital age, particularly those arising from the asymmetrical balance between the public and private sectors, compel the academy to evolve its strategies and explore new avenues to assert its traditional hegemonic role as the authoritative arbiter in linguistic affairs. As with many standardisation efforts, there are also ideological driving factors in this case.

Reclaiming authority in the era of machine-learning technology

Traditional normative institutions like RAE feel threatened due to the social changes brought about by language technologies. To maintain their status as dominant language policy actors in the twenty-first century, they must broaden their sphere of influence and activity into the digital realm, which is outside of their national constituencies. Before focusing on their repositioning vis-à-vis technological companies, let us examine how RAE perceives the manners in which its authoritative work has traditionally impacted society. This is illustrated in (2).

(2) The works of RAE have always been accepted and considered as obligatory rules throughout the three centuries of the institution. The Academy has no sanctioning power at its disposal with which to repress offenders, but its authority and prestige determine that its rules constitute a singular "soft law" whose observance is essential for anyone who wishes to be a member of a Spanish-speaking community as a literate person. It is society itself that repudiates the barbaric, inappropriate, or incorrect use of the common language. (Muñoz Machado 2023:124)

With a slight tone of cultural nostalgia, in this discourse we see that notions of language are closely linked with notions of society and social hierarchy. Most important for understanding how linguistic authority is enacted in current times, RAE explains that its regulatory power is a sort of invisible disciplinary mechanism through which it imposes its supposedly unquestioned authority—its 'voice from nowhere'. Following this view, authority is not the result of a top-down process that is continuously enacted upon speakers but instead reflects the general, anonymous speakers' beliefs about and adherence to linguistic norms. When applied to the digital realm, RAE's argument in (2) implies that Spanish-speaking technology users would be reluctant to embrace conventionalised speech behaviours that do not authentically mirror their own, particularly standard Spanish. This assertion, however, is still to be proven empirically.

In a discourse often structured around a public-private dichotomy, RAE presents itself as the natural representative and enforcer of people's will and popular sentiment, a public-spirited institution widely legitimated by society (Del Valle 2007). It thereby establishes a democratic-looking foundation for linguistic prescriptivism, one that transcends a simplistic focus on normative prestige alone. As expected, RAE adheres to the conventional ideal of the standard language as constituting a common good but also elevates its regulatory role to a similar status, characterising it as an object of general public interest.

That said, this kind of soft power, which is partly state-sponsored through public funding and symbolic representation (the Spanish king is RAE's highest representative), may no longer be influential enough in the tech era. This perception of threat to the traditional role of normative institutions is illustrated in (3).

(3) There are many millions of agents who own the language, whose mutations and variants this institution meticulously monitors throughout the universal geography of the Spanish language. In recent years, and with increasing intensity, the changes are not the exclusive work of the Spanish-speaking community, because new agents have been introduced into the language system: technology companies that use artificial intelligence, which are potential regulators or, at least, prescribers of the language that belongs to us, with the capacity to impose variants that may not coincide with the common uses of humans. (Muñoz Machado 2023:125–26)

Detached from the specific normative culture prevailing in Spanish-speaking territories, big tech companies often fail to comprehend or prioritise accommodating this notion of linguistic authority in their language models. Hence, considering RAE's previous pleas for economic and institutional support from state bodies to fight against the changing language policy power dynamics prompted by machine-learning technologies, it can be inferred that government and state representatives (some of whom were present at the highly formal event where Muñoz Machado delivered this speech) are among the intended recipients of RAE's messages.

In this new landscape where tech companies have emerged as influential players in language policy, example (3) illustrates some of the ideas that RAE traditionally uses to support their normative mandate. These include an allegedly judgement-free notarial role (Paffey 2021) or systematic descriptive approach based on actual linguistic usage and a celebratory appraisal of the internal diversity of the language in dictionaries, grammars, and other reference work. Of course, contrary to reductionist claims that standards are created by speakers rather than academicians (see Paffey 2021), linguistic authority allows for the transition from describing language to prescribing its use, which in RAE's case has been defined as 'moderate prescriptivism' (Del Valle 2007:249). Accordingly, RAE's principal function is the professional monitoring of the authentic linguistic practices of real human speakers, which at heart implies the construction of authenticity itself (Bucholtz 2003:403).

Against this backdrop, large language models embody the ideology of anonymity, that is, the disposal of any social and cultural situatedness in language, potentially erasing vernacular forms and imposing instead voices from nowhere. Rather than accepting this language of no-one-in-particular, available to everyone in the deterritorialised digital space, RAE's claim to authority-as-authenticity views the language generated by technology as inauthentic and the language produced by humans as genuine, particular, and localised, that is an AUTHENTIC, HUMAN-PRODUCED STANDARD LANGUAGE for an IMAGINED COMMUNITY OF AUTHENTIC SPEAKERS. We refer to it as an IMAGINED community because RAE presents its pan-Hispanic pluricentric approach as neutral Spanish—a disembedded global language not owned by

Spain or any particular nation state (Moreno-Fernández 2016; Paffey 2021:251). In fact, RAE's promotion of a general, politically neutral, and geographically ubiquitous standard variety somewhat resembles the anonymous language produced by machine-learning technologies, which RAE criticises for being an outside influence.

Yet, RAE's repertoire of arguments is not confined to the ideological concepts of linguistic authority and authenticity. As illustrated in (3), RAE's characterisation of Hispanic speakers and their communities, the real owners of the language and inducers of language change, as defenceless against big tech companies suggests an unfair distribution of power over standardised language. As part of its strategy in the battle for control over language in the digital realm, RAE connects its mission with a new discourse of 'rights', as we comment on in more detail in the following section.

Fresh approaches for new challenges

Let us focus now on the broader framing of the regulating role of language academies in the era of machine-learning technologies. In a context of underregulated global markets where single states no longer have the capacity to regulate both continuous technological advancements and the companies behind them (Castells 2010), RAE posits the control of language-related technology as being embedded in broader discussions around the legal and moral imperative to regulate technological advancements. In other words, for RAE, the challenges posed by machine-learning technologies go hand in hand with those posed by language-specific technologies.

As noted above, RAE underwent a process of opening towards Hispanic America—whether this was on its own initiative or out of political necessity (Del Valle 2007)—to recognise and integrate the linguistic authority emanating from American academies. This resulted in a pluricentric view of the Spanish language that proudly expresses the catchy message of 'unity in diversity'. However, in this new technological phase, RAE aims to find new potential allies in a supra-statal association other than Hispanic America. RAE's regulatory efforts, particularly in language technologies, align naturally with the European regulatory tradition and, more concretely, with the EU's principle of legal security, which contrasts with deregulatory models prevalent in the US. This tension between the two traditions, including the contrasting ideological principles underpinning them, is illustrated in the following example.

(4) Given that the challenge of defending culture and rights from the risks posed by artificial intelligence is an important one, there is no doubt that the best option for states and the European Union to pursue, and the latter rather than individual states given the scale of the problem, is to regulate it as soon as possible. In Anglo-American economic circles, they are more inclined to favour self-regulation. European culture has always preferred the regulation of new inventions ... The use of natural language by artificial intelligence is a goal that has already been achieved. It is clear that legal and ethical limits will have to be set for the protection of values and rights, either through self-regulation or regulation. (Muñoz Machado 2023:123)

Despite referring to politico-economic frameworks in this specific example, this regulatory–deregulatory binary in the market economy may be applied to the language domain and, particularly, to the different traditions in which standard languages are institutionally managed. For instance, while English has developed in a rather decentralised manner without a single clear state-sponsored authority over the language, in the case of many European languages, public bodies of the corresponding states, often language academies, enjoy the privileged role of regulating language. It follows that machine-learning technologies may pose a more significant challenge to languages with a widely accepted body of linguistic authority, such as Spanish or French, than to those without a professional authority body or the traditional binding to a nation state. In fact, given that big tech companies are based in the Anglosphere and their large language models are mostly fed with corpora in English, it is unsurprising that voices critical of the inattentive management of language by machine-learning technologies emerge in countries with national standard languages other than English. This sense that something needs to be done about the impact that such technologies would appear to have overshadowed the traditional concerns of RAE, such as the corrupting effects of Anglicisms on the purity of Spanish.

In sum, RAE contends that the traditional continental European regulatory culture should be upheld and enforced. In addition to framing language management from a legal perspective, as previously noted, RAE uses the discourse of rights and values as the new basis for justifying their authoritative mandate. More specifically, RAE includes democratic moral values such as equity and accessibility to claim that language is a public good, thus advancing the pressure for linguistic validation in language technologies. This is exemplified in the next example.

(5) The simplifications, jargon, and dialects that have been introduced by social networks and which may be generalised by artificial intelligence require special attention. The duty to use clear/plain language is largely related to the preservation of individual rights, which cannot be adequately exercised in the face of obscure or almost encrypted communications for those who lack minimum digital skills. For this reason, the idea of accessible language must be added to that of clear/plain language. The language of artificial intelligence must be adapted to people's natural abilities. (Muñoz Machado 2023:135–36)

This example contains a tacit criticism that non-standard language has been incorporated into large language models via social network data. Notably, however, RAE introduces critical nuances other than their view of the deficient suitability of those models for Spanish-speaking communities. In apparent opposition to techno-solutionist approaches (Morozov 2013) and views of technological developments as universally applicable, such as Meta's 'No Language Left Behind' universal translation project which is designed to 'learn' new languages with less training data, RAE emphasises its public-service orientation. By framing the debate as 'human vs. machines', it adopts an anthropocentric stance in favour of human-centred abilities over freely available, yet often inadequate, artificially mediated communicative tools.

This focus on plain language and, more generally, the moral principle of the accessibility of communication tools (RAE 2024) adds a layer of complexity to the prevailing, but rather uncritical, belief that big tech companies have democratised the use of machine-learning systems. Ultimately, as detailed in the following section, RAE considers that professional, human linguistic supervision is necessary to ensure that machine-learning technologies effectively enable the use of language-mediated digital communication. Such supervision resembles in some ways the existing human-centred approaches that tech companies use for content moderation, including in detecting abusive language in the training data.

Striving to exert authority over tech industries

Linguistic authority bodies have traditionally relied on the backing of the modern nation state-supported infrastructure such as print media, public broadcasting, and an education system. While RAE has made efforts to modernise the standardisation process, for example, by engaging with speakers on social media (Paffey 2021), there is a perception that the relevance of its normative work is shrinking (Ayres-Bennett & Bellamy 2021), especially with the widespread adoption of language technology in everyday tasks. RAE is explicit about potential remedies for this situation, which involve tech companies, as exemplified in (6).

> (6) We have convinced humans how they should use Spanish based on our reputation. But we cannot make machines follow the same rules without talking to the manufacturers … we are obliged to position ourselves in the digital world so that artificial intelligence begins to speak in Spanish, to impose reason in this new universe and to prevent the language from getting out of hand and favouring the big multinational technology companies. (Muñoz Machado in Iglesias Fraga 2021)

In this context of non-standard deviations, RAE aims to extend its normative mandate to cyberspace, which can be interpreted as reproducing the monopoly of linguistic prescriptivism in the context of privately owned language technologies. This entails that tech companies, which typically favour deregulation, have a commercial focus, and are based in non-Spanish-speaking jurisdictions, must be linguistically disciplined. This includes, for example, rigorous curation of the data sets used for training machine-learning algorithms, which are held away from public scrutiny, and that language models conform to standardised linguistic norms. Yet, it must be noted that tech companies view speakers as consumers and are primarily concerned with their online behaviour rather than their linguistic varieties or national identities. Machine-learning technologies aim to replicate recognisable genres and, more broadly, to generate language that users perceive as meaningful and correct, without necessarily prioritising the traditional linguistic normativity. Put differently, these technologies represent a different type of normativity: one focused on vast amounts of data and profit generation, rather than on meaningful communication potential. Therefore, such an ambitious goal of exerting the power of language authority over the corporate sector demands a concerted effort from both public and private stakeholders. This complex interaction between

different actors, sometimes with opposing agendas, illustrates the sociotechnical assemblage that defines language policy in the era of machine-learning technologies.

To achieve this goal, RAE has embarked, along with one of Spain's State Secretariats, on Lengua Española e Inteligencia Artificial (LEIA; 'Spanish Language and Artificial Intelligence'), a public–private partnership that aims to ensure the use of standard Spanish in technological products by, for example, making RAE's dictionary data sets and linguistic corpora available to tech firms (RAE 2020). In recent years, RAE has established agreements with Google, Amazon, Microsoft, Twitter, and Facebook to ensure that their voice assistants, word processors, search engines, chatbots, instant messaging systems, and social networks comply with RAE's approved standards for good Spanish usage (see Muñoz-Basols, Palomares Marín, & Moreno-Fernández 2024). Furthermore, RAE plans to implement a certification system to assess the quality of Spanish used in digital systems.

LEIA is included in the Strategic Project for Economic Recovery and Transformation on the New Economy of Language (Gobierno de España 2022), an overarching project financed through the European Union's Next Generation EU programme. Within this overarching project, RAE's LEIA is generally framed as making AI 'think in Spanish' or facilitating tech companies in the deployment of their services in a COMMON LANGUAGE' (emphasis added). However, the overall strategic project has manifestly market-oriented goals, as its main aim is to enhance the potential of Spanish (and Spain's co-official languages) as a driver of economic growth and international competitiveness (Gobierno de España 2022). This approach aligns with the principles of digital capitalism and comes as no surprise because it matches its funding source: the EU's recovery plan, designed to support member states' economic recovery following the Covid-19 pandemic. Despite its economic-led approach, this project opens a window of opportunity for RAE to renew its influential status in society, particularly during a period of an acute shortage of public financial resources for the academy.

To conclude, it is worth noting that by taking the lead in these initiatives, RAE solidifies its position as the pre-eminent linguistic authority over those of the other Spanish-speaking countries and reinforces Madrid as the continuing centre of linguistic authority in the Hispanic world. It does so both symbolically, by assuming the role of leading voice and driving force in standardisation matters for Spanish in the digital realm, and practically, by securing and managing European funding for this purpose and acting as the main interlocutor with big tech companies. This not only evokes past colonial practices but also reminds us that those who profited most from colonialism are also the ones reaping the greatest benefits from machine-learning technologies (Mejias & Couldry 2024).

Discussion and concluding remarks

We have discussed how traditional institutions of linguistic authority negotiate their position in the face of new language actors that have arisen in the sociotechnical assemblages shaped by commercial digital industries. Our investigation has focused on RAE's discourse about its position and its critique of big tech companies from Silicon Valley. After summarising our observations, we discuss what the tensions

between commercial entities and (semi)public institutions like RAE suggest about the redistribution of power structures.

RAE's discourse draws on traditional tropes of authenticity and endangerment, now presenting commercial technologies as threats to the unity and expressive and aesthetic qualities of Spanish as a global language. This threat is presented as emanating from language produced by machines, which is understood as artificial and inauthentic. The humans behind artificially generated language, that is, the Anglophone companies in Silicon Valley, are mentioned but the language that is produced by their companies is treated as supposedly different from human language, namely as 'non-human' language. Machine-learning technologies, operating on the basis of the idea that language equals machine-readable written data, and that language should not be regulated, are understood as having the potential to disrupt and destroy authentic Spanish, defined as Spanish produced by humans, and consequently the world's community of Spanish speakers.

Comparing RAE's discourse with previous notions of standard language, we see continuities and changes. In traditional conceptualisations of national standards as constructed and supported by an institution like RAE, language norms were perceived as unmarked authority, functioning as a powerful gatekeeping mechanism. An institution like RAE did not, in the past, have to justify its decisions, operating instead within (supposedly) accepted national or supranational traditions. When faced with multinational digital corporations, however, traditional language academies need to justify their existence and assert their authority. In the discourse analysed above, RAE invokes aspects of the traditional endangerment discourse and emphasises its long-recognised professionality in matters of the Spanish language, demanding that private technology companies recognise this special knowledge as linguistic authority. It now bases its authority on the supranational unity of Spanish-speaking countries and on European values that privilege the regulation of technological innovation. Language norms are no longer presented as symbols of refinement and education but as the representation of a democratic defence of the people against external and 'non-human' entities. While linguistic authority in national contexts is traditionally understood as anonymous and the language of 'everyone', newer discourses of legitimation declare a standard language a common good. It is defined as owned by the members of its (imagined) speaker community and must be based on the language commonly used by this community and represent and respect its linguistic diversity (as also documented by RAE); it also functions as a vital force that unites a global community of speakers. It is this common good that is being threatened by a force that is constructed as foreign, unnatural, and non-human.

What do these discourses suggest about the redistribution of power? There appears to be tension between global commercial actors and quasi-traditional national actors, both of whom purport to represent the public but have different understandings of what that entails. RAE presents this as a tension between the human community and a non-human asocial voice. It presents an image of machine-learning language technologies championed by globally acting commercial companies from Silicon Valley as threatening traditional culture, authenticity, democratic values, and unity. And yet, what is at stake here can also be interpreted as a power struggle over who is able to assert a voice 'from nowhere'.

Underlying this discussion is the question of who controls public space—traditional national institutions or global multinational and commercial companies? The latter have access to massive proprietary data sets, which they use to create and disseminate language, effectively driving the privatisation of language. Powerful private interests dominate the principle of social ordering (Zuboff 2019:192); there is an emerging global public space that is governed by commercial companies rather than national political institutions. The formerly unquestioned hegemonic linguistic sovereigns must try and find a place in this reconfigured social economy.

The observation that non-public, commercial entities have become vital actors in global power arrangements has been discussed and problematised in sociological theory. Bauman's writings on the structures of late modernity seem to be confirmed here; he argues that 'in the fluid stage of modernity, the settled majority is ruled by the nomadic and extraterritorial elite' (Bauman 2012:13), which is neither democratically elected nor visible but is shaped after 'the old-style "absentee landlords"'. It rules without 'welfare concerns' or 'the mission of "bringing light", "reforming the ways", morally uplifting, "civilizing" and cultural crusades' (Bauman 2012:13). Accordingly, the state and its institutions are no longer the 'plenipotentiary of reason or the master-builder of the rational society' (Bauman 2012:48) but are 'replaced by a 'shopping mall' in which freed individuals shop around among the offerings from the commercial providers for the best fit to satisfy their (individualised) needs rather than pursue a common goal' (Bauman 2012:20).

While sociolinguistics tends to discuss national linguistic norming as a practice of oppression, institutions such as RAE that are responsible for the creation and maintenance of shared signs arguably also contribute to social community and shared identity formation. In this new context then, like RAE, sociolinguistics continues to be 'dominated by theoretical and methodological preferences for offline, spoken discourse in fixed and clearly definable time-space, socio-cultural and interpersonal contexts and identities' (Blommaert 2019:486) and conceptualises face-to-face (or user-to-user) language use as the locus of language. In the context of big tech realities, sociolinguistics must urgently find a new place and rethink much of what it assumes 'to be natural, primordial and commonsense about language' (Blommaert 2019:486; see also Erdocia, Migge, & Schneider 2024). In the post-digital context, current understandings of language and the emergence, distribution, and sustaining of language (Migge et al. 2025) will be increasingly restructured by the algorithmic practices of big tech companies (Kelly-Holmes 2022) and language in the offline world will be inseparably intertwined with that in the online world. It is currently unclear what effect commercial language technology provision will have on the future of shared language and on the future of social communities.

What is clear is that the profit-oriented desires of big tech companies and the norming desires of national and supranational political institutions will continue to function as important centrifugal and centripetal forces in this ever-amplifying process. Normative ideas about language will continue to have an important social function, but we will likely observe the emergence of reconfigured constructions of anonymous voices and authentic language.

Notes

* This work was supported by the Faculty of Humanities and Social Sciences, Dublin City University.

[1] By 'machine-learning technologies', we mean tools that are based on algorithms that detect and reproduce patterns in data sets. We refrain from calling these tools 'artificially intelligent', since it is a marketing term that mystifies the functioning of algorithmic matrix multiplications and thus enforces the power of a handful of companies that have the financial and material resources to build large language models from scratch (see Katz 2020; see Bender, McMillan-Major, Gebru, & Shmitchell 2021 for a critical assessment of the hegemony of 'Big Tech').

[2] Note, for example, that in the US, Sec. 230 of the US Telecommunications Act allows social media companies to be treated as a 'common carrier' (like a phone or courier company), which means that they are not made responsible for what appears on their platforms (thanks to a reviewer for bringing this to our attention).

References

Androutsopoulos, Jannis (2021). Polymedia in interaction. *Pragmatics and Society* 12:707–24.

Ayres-Bennett, Wendy, & John Bellamy (2021). Introduction. In Wendy Ayres-Bennett & John Bellamy (eds.), *The Cambridge handbook of language standardization*, 1–24. Cambridge: Cambridge University Press.

Bauman, Zygmunt (2012). *Liquid modernity*. Cambridge: Polity Press.

Bender, Emily; Angelina McMillan-Major; Timnit Gebru; & Shmargaret Shmitchell (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT'21: Proceedings of the 2021 ACM conference of fairness, accountability, and transparency*, 610–23: https://doi.org/10.1145/3442188.3445922.

Blommaert, Jan (2015). Commentary: Superdiversity old and new. *Language & Communication* 44:82–88.

Blommaert, Jan (2019). From groups to actions and back in online-offline sociolinguistics. *Multilingua* 38(4):485–93. https://doi.org/10.1515/multi-2018-0114.

Bode, Katherine, & Lauren M. E. Goodlad (2023). Data worlds: An introduction. *Critical AI* 1. https://doi.org/10.1215/2834703X-10734026.

Bucholtz, Mary (2003). Sociolinguistic nostalgia and the authentication of identity. *Journal of Sociolinguistics* 7:398–416.

Cameron, Deborah (1995). *Verbal hygiene*. London: Routledge.

Castells, Manuel (2010). *The rise of the network society*. 2nd edn. Oxford: Wiley.

Crawford, Kate (2021). *Atlas of AI*. New Haven, CT: Yale University Press.

Del Valle, José (2007). Embracing diversity for the sake of unity: Linguistic hegemony and the pursuit of total Spanish. In Duchêne & Heller (eds.), 242–67.

Del Valle, José (2013). Linguistic emancipation and the academies of the Spanish language in the twentieth century: The 1951 turning point. In José Del Valle (ed.), *A political history of Spanish: The making of a language*, 229–45. Cambridge: Cambridge University Press.

Duchêne, Alexandre, & Monica Heller (eds.) (2007). *Discourses of endangerment: Ideology and interest in the defence of languages*. London: Continuum.

Durrell, Martin (2000). Standard language and the creation of national myths in nineteenth-century Germany. In Martin Durrell (ed.), *Das schwierige 19. Jahrhundert*, 15–26. Berlin: de Gruyter.

Edwards, John (2012). Language management agencies. In Bernard Spolsky (ed.), *The Cambridge handbook of language policy*, 418–36. Cambridge: Cambridge University Press.

Erdocia, Iker; Bettina Migge; & Britta Schneider (2024). Language is not a data set: Why overcoming ideologies of dataism is more important than ever in the age of AI. *Journal of Sociolinguistics* 28(5):20–25. Online: https://doi.org/10.1111/josl.12680.

Erdocia, Iker, & Josep Soler (2024). In pursuit of epistemic authority in public intellectual engagement: The case of language and gender. *Multilingua* 43:1–28.

Fairclough, Norman (2015). *Language and power*. London: Longman.

Gal, Susan, & Kathryn A. Woolard (1995). Constructing languages and publics: Authority and representation. *Pragmatics* 5:129–38.

Gardiner, Michael (2004). Wild publics and grotesque symposiums: Habermas and Bakhtin on dialogue, everyday life and the public sphere. *Sociological Review* 52:28–48.

Gobierno de España (2022). PERTE Nueva economia de la lengua. Approved by the Consejo de Ministros, 1 March 2022. Online: https://planderecuperacion.gob.es/como-acceder-a-los-fondos/pertes/perte-nueva-economia-de-la-lengua.

Gorwa, Robert; Reuben Binns; & Christian Katzenbach (2020) Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7(1). https://doi.org/10.1177/20539517198979.

Guy, Gregory (2011). Language, social class, and status. In Rajend Mesthrie (ed.), *The Cambridge handbook of sociolinguistics*, 159–85. Cambridge: Cambridge University Press.

Habermas, Jürgen (1962/1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. Cambridge, MA: MIT Press.

Heyd, Theresa, & Britta Schneider (2019). The sociolinguistics of late modern publics. *Journal of Sociolinguistics* 23:435–49.

Hickey, Raimund (2012). Standard English and standards of English. In Raimund Hickey (ed.), *Standards of English: Codified varieties around the world*, 1–31. Cambridge: Cambridge University Press.

Iglesias Fraga, Alberto (2021). La RAE quiere 'imponer' sus normas del español a la inteligencia artificial y espera fondos del Gobierno para ello. *El Español*, 1 December 2021. Online: https://www.elespanol.com/invertia/disruptores-innovadores/innovadores/tecnologicas/20211201/rae-quiere-imponer-espanol-inteligencia-artificial-gobierno/631187754_0.html.

Katz, Yarden (2020). *Artificial whiteness: Politics and ideology in artificial intelligence*. New York: Columbia University Press.

Kelly-Holmes, Helen (2022). Sociolinguistics in an increasingly technologized reality. *Sociolinguistica* 36:99–110. Online: https://doi.org/10.1515/soci-2022-0005.

Lodares, Juan Ramón (2002). *Lengua y patria: Sobre el nacionalismo lingüístico en España*. Madrid: Taurus.

Lorenci, Miguel (2023). El español debe estar en el centro de la inteligencia artificial, el Metaverso y la diplomacia. *El Correo*, 23 March 2023. Online: https://www.elcorreo.com/culturas/espanol-debe-estar-centro-inteligencia-artificial-metaverso-20230323143055-ntrc.html?ref=https%3A%2F%2Fwww.elcorreo.com%2Fculturas%2Fespanol-debe-estar-centro-inteligencia-artificial-metaverso-20230323143055-ntrc.html.

Maly, Ico (2021). Ideology and algorithms. *Ideology, Theory, Practice*. Online: https://www.ideology-theory-practice.org/blog/ideology-and-algorithms.

McLelland, Nicola (2021). Language standards, standardisation and standard ideologies in multilingual contexts: Introduction. *Journal of Multilingual and Multicultural Development* 42(2):109–24.

Mejias, Ulises A., & Nick Couldry (2024). *Data grab: The new colonialism of Big Tech (and how to fight back)*. London: Penguin.

Migge, Bettina; Britta Schneider; Didem Leblebici; Iker Erdocia; Mandy Lau; Auli Viidalepp; Beatrice Savoldi; Martina Podboj; Philipp Meer; Mohammad Alenezi; & Agnese Sampietro (2025). Conceptualising language in the human-machine era: Language ideologies and language as data. In Dave Sayers (ed.), *Language in the human-machine era*. Cambridge, MA: MIT Press, to appear.

Moreno-Fernández, Francisco (2016). La búsqueda de un 'español global'. VII Congreso Internacional de la Lengua Española. Instituto Cervantes – RAE. Online: https://congresosdelalengua.es/puerto-rico/paneles-ponencias/espanol-mundo/moreno-fancisco.htm.

Morozov, Evgeny (2013). *To save everything, click here: Technology, solutionism, and the urge to fix problems that don't exist*. New York: Allen Lane.

Muñoz-Basols, Javier; María del Mar Palomares Marín; & Francisco Moreno-Fernández (2024). El Sesgo Lingüístico Digital (SLD) en la inteligencia artificial: Implicaciones para los modelos de lenguaje masivos en español. *Lengua y Sociedad* 23(2):623–47.

Muñoz Machado, Santiago (2023). Contestación. In Asunción Gómez-Pérez, *Inteligencia artificial y lengua española*, 117–43. Madrid: Real Academia Española. Online: https://www.rae.es/sites/default/files/2023-05/Discurso%20Ingreso%20Asuncion%20Gomez-Perez_0.pdf.

Ong, Walter J. (1982). *Orality and literacy: The technologizing of the word*. London: Routledge.

Paffey, Darren (2021). State-appointed institutions: Authority and legitimacy in the Spanish-speaking world. In Wendy Ayres-Bennett & John Bellamy (eds.), *The Cambridge handbook of language standardization*, 234–62. Cambridge: Cambridge University Press.

Pasquinelli, Matteo (2023). *The eye of the master: A social history of artificial intelligence*. London: Verso.

Pennycook, Alastair (2024). *Language assemblages*. Cambridge: Cambridge University Press.

RAE (2020). Lengua Española e Inteligencia Artificial (LEIA). Online: https://www.rae.es/leia-lengua-espanola-e-inteligencia-artificial.

RAE (2024). S. M. el Rey preside la clausura de la I Convención de la Red Panhispánica de Lenguaje Claro [Press release], 21 May 2024. Online: https://www.rae.es/sites/default/files/2024-05/NdP_RAE_S.%20M.%20preside%20la%20clausura%20de%20la%20I%20Convencio%CC%81n%20de%20la%20Red%20Panhispa%CC%81nica%20de%20Lenguaje%20Claro.pdf.

Reisigl, Michael, & Ruth Wodak (2016). The discourse-historical approach (DHA). In Ruth Wodak & Michael Meyer (eds.), *Methods of critical discourse studies*, 23–61. London: SAGE.

Richard, Bauman, & Charles L. Briggs (2003). Making language and making it safe for science and society: From Francis Bacon to John Locke. In Richard Bauman & Charles L. Briggs (eds.), *Voices of modernity: Language ideologies and the politics of inequality*, 19–69. Cambridge: Cambridge University Press.

Schneider, Britta (2022). Multilingualism and AI: The regimentation of language in the age of digital capitalism. *Signs and Society* 10(3):362–87. https://www.journals.uchicago.edu/doi/10.1086/721757.

Silverstein, Michael (1996). Monoglot 'standard' in America: Standardization and metaphors of linguistic hegemony. In Donald Brenneis & Ronald Macaulay (eds.), *The matrix of language*, 284–306. Boulder, CO: Westview.

Vandenbussche, Wim (2022). The pursuit of language standardization research as a mission for true sociolinguists. *Sociolinguistica* 36(1–2):219–29. https://doi.org/10.1515/soci-2022-0026.

Wodak, Ruth (2015). *The politics of fear: What right-wing populist discourses mean*. London: SAGE.

Woolard, Kathryn A. (2016). *Singular and plural: Ideologies of linguistic authority in 21st century Catalonia*. New York: Oxford Academic.

Zuboff, Shoshana (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: Public Affairs.