Special Issue: Natural Language Processing in Psychology

Original Article

Uncovering the Secret Life of We-Pronouns in the German Parliament

Computational Text Analysis in a Large-Scale Speech Dataset

Robin Bachmann and Ilka H. Gleibs

Department of Psychological and Behavioural Science, London School of Economics and Political Science, UK

Abstract: In social and political psychology, pronoun use, especially "we", plays an important role reflecting social identity and predicting leadership success in the political realm and beyond. Previous research focused on top-level political leaders and word-counting methods. We examined We-language and reelection success based on 349,783 speeches from 3,630 members of the German federal parliament from 1949 to 2021. We combined traditional We-counts with new Natural Language Processing (NLP) tools to explore syntactic and semantic contexts. We-language predicted next-term reelections and total number of reelections from the frontbench to the backbench and across many social and political factors. There were differential effects for We-language versus Our-language, We-specification, and We-categories. Our study revealed an upward trend of We-language in the parliament in recent decades. This research illustrates the use of new large-scale speech datasets and NLP-based tools for language and leadership research. We give valuable insights into the subtle language behind successful leadership.

Keywords: social identity, leadership, parliamentary debate, pronoun use, computational text analysis

In contemporary society, politicians face the complex task of representing diverse groups, acting on their behalf, and simultaneously unifying and shaping those to garner support for political initiatives and enact meaningful change. Language, their primary tool in a parliamentary democracy, plays a crucial role in this endeavor, making understanding how leaders leverage language for identity leadership vital. This paper illustrates how extensive political speech datasets and advancements in computational language analysis can contribute to a deeper understanding of the interplay between identity leadership, language, and political success.

Political Leadership as Identity Leadership

Leadership research has undergone a significant transformation, moving away from the traditional *great man* concept to embrace contextual, transformational, and social dimensions. The *Social Identity Approach to Leadership* (Haslam et al., 2020; Hogg et al., 2012) is a prime example of this evolution. In this approach, leadership is rooted in group processes and revolves around a shared social identity (Tajfel & Turner, 1979), enabling both leadership and followership (Haslam & Platow, 2001; Hogg, 2001). Haslam et al. (2020) synthesized *Identity Leadership (IL)* in four key components: prototypicality ("being one of us"), advancement ("doing it for us"), entrepreneurship ("crafting a sense of us"), and impresarioship ("making us matter"). Within this framework, leadership operates as "power through us" (Haslam et al., 2020, p. 57). Leaders must embody the group's identity and then lead through shaping the group's self-concept and actions. The robustness of the approach has been demonstrated in both meta-analytical work (Steffens et al., 2021) as well as large-scale, cross-cultural studies (Bracht et al., 2023; van Dick et al., 2018).

Political psychological research has explored social identity (Brewer, 2001; Huddy, 2001) and leadership through social identity (Mols et al., 2023; Monzani et al., 2024), examining various political phenomena such as national identity and international cooperation (Mols et al., 2009; Reicher & Hopkins, 2000), or political identity and partisanship (Huddy & Bankert, 2017; Steffens et al., 2018). It suggested that political leaders achieve effectiveness by aligning themselves and their political agenda with a shared social identity among leaders and followers (Haslam et al., 2020; Reicher & Hopkins, 2000).

Language of Political and Identity Leadership

Language analysis is a well-established research tool in social and political psychology (Dehghani & Boyd, 2022), typically focusing on political speeches as "natural language" and employing qualitative methods (Hopkins, 2023). Condor et al. (2013) identified two analytical perspectives: argument and speaker identity. While political speakers may strategically avoid identifying with an audience (Fløttum, 2010), Condor et al. highlighted their task to convey implicit identity alignment with the audience, emphasizing the significance of first-person-plural pronouns, specifically "we." Similarly, the Social Identity Approach to political leadership utilizes We-pronouns as reference to a shared social identity and IL language, drawing on extensive quotations from political speeches (Haslam et al., 2020), detailed case studies on political leaders (Haslam et al., 2023), and specific qualitative speech analysis methods (Gleibs et al., 2018).

Simultaneously, IL research utilizes quantitative language analysis. Computational word counting methods gained prominence in psychological research (Pennebaker et al., 2003; Tausczik & Pennebaker, 2010), offering a reliable tool for uncovering subtle language cues such as the "secret life of pronouns" (Pennebaker, 2011), and various group processes (Pennebaker, 2011; Van Swol & Kane, 2019). For example, Steffens and Haslam (2013) leveraged this quantitative and computational analysis predicting election results by quantifying the use of "we" and "I" in presidential candidates' speeches. Other research explored complex pronoun combinations in parliamentary debates (Chulvi et al., 2024).

Notably, previous quantitative studies often treated We-use as indicator for social identity construction and leadership without controlling for other factors influencing pronoun use (Chulvi et al., 2024; Steffens & Haslam, 2013; also Fladerer et al., 2021), such as political ideology (Jost & Krochik, 2014), political power (Houck et al., 2022), social status (Kacewicz et al., 2014; Shen, 2024), and other sociodemographics (Rouhizadeh et al., 2016).

Expanding Research With Large-Scale Parliamentary Speech Data

Leveraging an extensive dataset of German parliamentary speeches from 1949 to 2021, we investigated IL as We-language, contributing to IL research in the political realm on three key aspects:

First, IL is claimed to be not only an insightful analytical perspective but also an effective leadership practice (Haslam et al., 2017, 2020). An intriguing question arises whether there is a general trend in the adoption of IL by politicians. For example, Jordan et al. (2019) observed a shift toward less analytical and more assertive political leadership styles in English-speaking countries over the past century. Our data, spanning over 70 years, enabled us to investigate long-term trends in We-language in the German parliament, both in absolute terms and in comparison to I-language.

Research Question 1 (RQ1): Is there a long-term increase in IL/We-language in the German parliament from 1949 to 2021? Is there a relative increase of We-language compared to I-language?

Second, demonstrating the effectiveness of political IL poses challenges. Experimental IL studies, more prevalent in organizational settings (Steffens et al., 2021; van Knippenberg, 2011), are limited in the political domain (Gleibs & Haslam, 2016; Steffens et al., 2014). Steffens and Haslam (2013) proposed an approach focusing election results, showing that We-language in Australian presidential candidates' speeches predict their election. Our study examined whether We-language in parliamentary speech predict the reelection of parliamentarians. Our data enabled two novel perspectives: election sequences for studying the evolution of power in parliament and speaker diversity for assessing We-language controlling for other speaker characteristics.

Research Question 2 (RQ2): Does IL/We-language in parliamentary speeches predict next-term reelection and total reelections into the parliament?

Third, IL research on political leadership often examined prominent figures, including historical individuals (Haslam et al., 2020) or current state leaders such as Barack Obama (Augoustinos & De Garis, 2012; Steffens et al., 2014) or Donald Trump (Haslam et al., 2023). Some research extended the analysis, focusing on presidential spouses as "identity mediators" (Gleibs et al., 2018). However, more attention is needed for second-tier political leaders. Parliamentary debates (Ilie, 2017) provide a rich source of IL from both frontbenchers and backbenchers.

Research Question 3 (RQ3): Are there differences in the predictive value of IL/We-language for reelection between frontbenchers and backbenchers?

Uncovering the Meaning of "We" With New Computational Text Analysis Tools

We employed quantitative analysis, incorporating new computational text analysis tools. Traditional word-counting methods, treating all word instances equally,

are limited in capturing contextual meaning (Boyd & Schwartz, 2021), especially evident with pronouns such as "we." To whom "we" refers, and consequently, the related social identities, can only be understood within the context. Qualitative speech analysis excels at incorporating contextual richness (Gleibs et al., 2018). Even previous quantitative studies reanalyzed We-references qualitatively (Steffens & Haslam, 2013). However, this approach faces limitations with large-scale text data, such as our parliamentary dataset. Advancements in computational text analysis, especially *Natural Language Processing* (*NLP*), are revolutionizing our capacity for understanding words in contexts (Berger & Packard, 2022; Jackson et al., 2022). In our study, we adopted a twofold approach, combining traditional We-counts with advanced NLP-based tools to explore the syntactic and semantic context of We-words, contributing to a more nuanced understanding of We-language and its effectiveness on three aspects:

We employed word-counting to address an unanswered question in previous research. While established dictionary tools such as the *Linguistic Inquiry and Word Count* (*LIWC*, Tausczik & Pennebaker, 2010) combine personal ("we") and possessive pronouns ("our"), IL research has inconsistently excluded possessive pronouns without exploring potential differential effects (Fladerer et al., 2021; Steffens & Haslam, 2013). However, researchers advocated for distinguishing between collective identity ("we") and collective ownership ("our;" Verkuyten & Martinovic, 2017), illustrating how Our-language can hinder or enhance IL through implicit exclusion (Nijs et al., 2021, see also Haslam et al., 2023). To investigate this dynamic, we examined We-language ("we Germans") and Our-language ("our country") as separated predictors for reelection as well as their interaction.

Research Question 4 (RQ4): Do both We-language and Our-language predict reelection? Do they interact?

Furthermore, we utilized NLP-based tools to address more nuanced aspects of We-language. We-use was characterized as implicit identity reference (Condor et al., 2013), and even with contextual information, determining the referent of "we" can be challenging. IL research posited, however, that the effectiveness of IL relies on making identity salient and explicit, allowing leaders to portray themselves as prototypical and shape group identity (Haslam et al., 2020; Hogg et al., 2012). NLP-based syntactic analysis enabled us to identify instances where "we" is specified by a dependent syntactic construction, for example, a predicative noun ("We are Germans"). Our study assessed whether these We-specifications influence the predictiveness of We-language for reelection.

Research Question 5 (RQ5): Do both We-language and We-specification predict reelection? Do they interact?

Lastly, our research delved into the semantic meaning of We-use. Both political leadership (Mols et al., 2009; Reicher & Hopkins, 2000) and IL research (Gleibs et al., 2018; Haslam et al., 2020) have linked We-use to specific social identities. Steffens and Haslam (2013), for example, rated how often presidential candidates specified "we" as nation, government, or party. Our NLP-based analysis allowed us to identify syntactically dependent words, facilitating a large-scale semantic analysis of We-specifications. We assessed the relative frequencies of these semantic specifications to explore the predictiveness of specific We-categories for reelection.

Research Question 6 (RQ6): Which We-categories predict reelection?

Method

Samples

We used a newly available corpus of speeches held in the German federal parliament ("Deutscher Bundestag") from its establishment in 1949 to 2021 (Richter et al., 2023, opendiscouse.de). From digitalized parliamentary protocols, the corpus extracted individual speeches and their speakers. We included only complete parliamentary terms, excluding the current term, and only members of parliament, excluding president, deputy president, government members, and guest speakers. Furthermore, we included speeches between 500 and 100,000 words and speakers with this word count per term. Three samples addressed our research questions: The first sample (N_0 = 349,783) included all speeches meeting our criteria. The second sample (N_1 = 9,243) aggregated variables for each speaker within each term, resulting in a longitudinal sample with terms as repeated observations for reelected speakers. The third sample (N_2 = 3,620) aggregated variables for each speaker across all terms, resulting in a sample with each speaker represented only once. Thus, we used data from 3,620 distinct speakers with on average 2.6 reelections and 97 speeches per speaker over 19 terms.

Measures

As language measures, we used the *We-LIWC* and *I-LIWC* from the German version of the *Linguistic Inquiry and*

Table 1. We-categories with specifying words

We-category	Specifying words (German in brackets)
Europe	Europe (Europa), European (Europäer)
Nation	Germans (Deutsche), country (Land), Germany (Deutschland), human (Mensch), people (Volk), society (Gesellschaft)
State	Democrat (Demokrat), federal republic (Bundesrepublik), politician (Politiker), federal government (Bundesregierung), federation (Bund), state (Staat), citizian (Bürger), politics (Politik), government (Regierung)
Parliament	Delegate (Abgeordneter), federal parliament (Bundestag), colleague (Kollege), parliament (Parlament), parliamentarian (Parlamentarier), legislator (Gesetzgeber), member of parliament (Bundestagsabgeordneter)
Party	Socialdemocrat (Sozialdemokrat), greens (Grüne), left (Linke), union (Union), CDU, FDP, liberals (Liberale), SPD, coalition (Koalition), parliamentary group (Fraktion), CDU/CSU, party (Partei), opposition (opposition), AfD, CSU

Note. All lemmatized nouns in syntactic dependency to a first-person plural personal pronoun with at least 10 occurrences sorted by descending frequency.

Word Count (*LIWC-G*, Meier et al., 2018), counting all first-person plural and first-person singular pronouns, respectively. Own We- and Our-measures differentiated between personal and possessive pronoun counts. We utilized *Natural Language Processing (NLP),* employing *spaCy* (spacy.io) and a German language model (*de_core_news_lg-3.7.0)*, analyzing the syntax of all speeches, to develop two measures: *We-specifications* counted all first-person plural personal pronouns specified by one of the following depending syntactic constructions: noun phrases ("we Germans"), appositions ("we, the Germans"), prepositions ("we as Germans"), and predicate nouns ("we are Germans"), also in relative clauses ("we who are Germans;" see Electronic Supplementary Materials, ESM 1, Appendix D for the code). *We-categories* counted all first-person plural personal pronouns semantically specified by words related to one of the following categories: Europe, nation, state, parliament, and party. We categorized all lemmatized nouns with at least 10 occurrences in the identified dependent syntactic constructions (see Table 1). All language indicators were defined as counts per 1,000 words relative to the total word count. We applied logarithmization due to their left-skewed distributions.

Among the variables measuring success of political leaders in parliament, *next-term reelection* (1 = *yes*, 0 = *no*) indicated for each parliamentary term if a speaker spoke again in parliament in a following term, *previous reelections* indicated for each term the sum of all previous terms a speaker spoke in parliament, and *total reelections* indicated the sum of all reelections across all terms. *Speech time*, the total word count in each term (also left-skewed and logarithmized), differentiated frontbenchers and backbenchers, assuming that frontbenchers have more speech time compared to backbenchers. In addition, based on the speech corpus (Richter et al., 2023), we added the variables *party* (levels: "Linke" ~ left, "SPD" ~ socialdemocratic, "Grüne" ~ green, "FDP" ~ liberal, "CDU" ~ conservative, "AfD" ~ far-right), *power* (levels: government party, opposition party), *age*, *gender* (1 = *female*, 0 = *male*), and *academic title* (1 = *Dr title*, 0 = *none*; see ESM 1, Appendix C, Table C1).

Analysis

We-language trends (RQ1) were analyzed with timeline graphs, the link between We-language and reelection (RQ2–6) with group comparisons and multiple linear regression models. We used random-effects models for predicting next-term reelection and between-effects models for predicting total reelections. Despite next-term reelection being dichotomous, we applied linear regression models for more robust estimations, particularly given the inclusion of interaction and multilevel terms (Gomila, 2021). All numeric predictors were standardized. Random-effects model controlled for number of previous terms in parliament as factor. Both random-effects and between-effects models controlled for parliamentary term as factor. All regression models were calculated a second time, adding the variables age, gender, academic title, party, and power.

Results

Trend Analysis

Figure 1 depicts the evolution of We-language and I-language in the German parliament from the first parliamentary term starting in 1949 to the last complete term ending in 2021 (RQ1). I-language (LIWC) started in the 1950s high, declined over two decades, rose in the 1970s again for two decades, and declined since the 90s again ($M_{49} = 27$, $SD_{49} = 30$, $M_{72} = 18$, $SD_{72} = 27$, $M_{90} = 25$, $SD_{90} = 37$, $M_{17} = 20$, $SD_{17} = 31$, $t_{49\sim72}(13{,}570) = -24.2$, $p < .001$, $d = -.33$, CI = [−.35, −.30], $t_{72\sim90}(40{,}744) = 22.2$, $p < .001$, $d = .21$, CI = [.20, .23], $t_{90\sim17}(42{,}788) = -18.5$, $p < .001$, $d = -.17$, CI = [−.18, −.15]).

Interestingly, We-language (LIWC) followed a similar development until the 90s, being around 1.5 times less frequent than I-language. However, while I-language decreased since the 90s, We-language continued increasing, surpassing I-language in the last term for the first time in frequency ($M_{49} = 18$, $SD_{49} = 17$, $M_{72} = 8$,

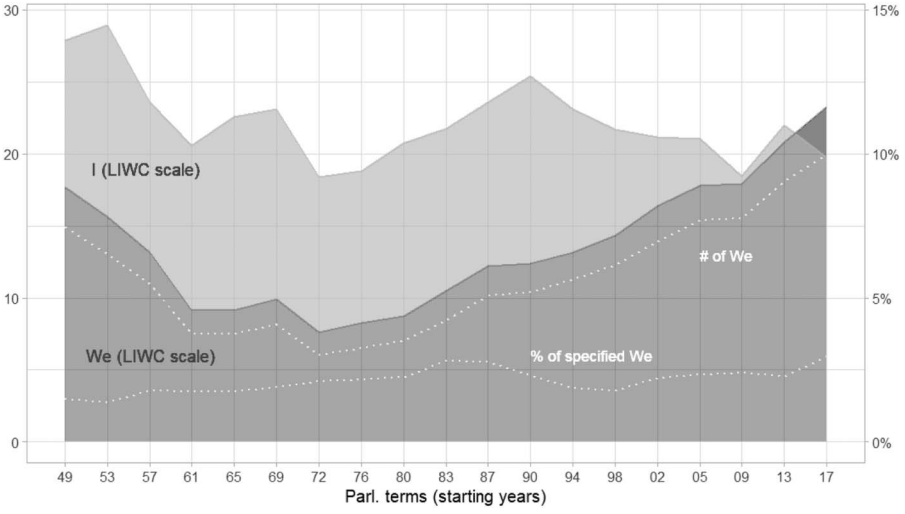

Figure 1. Frequency of We- and I-language across parliamentary terms. We- and I-LIWC-scales include both personal and possessive pronouns, while "we" counts include only personal pronouns. The first three measures are counts per 1,000 words (left *y*-axis), while the fourth measures specified "we" as percentage of its counts (right *y*-axis).

$SD_{72} = 14$, $M_{17} = 23$, $SD_{17} = 22$, $t_{49\sim72}(12{,}257) = -46.4$, $p < .001$, $d = -.64$, CI = [−.67, −.61], $t_{72\sim17}(52{,}132) = 64.7$, $p < .001$, $d = .86$, CI = [.84, .88]). Differentiating We- and Our-pronouns, We-pronouns were consistently dominant, driving the rise of We-language. The percentage of specified We-pronouns also increased from the 50s to the last term ($M_{49} = 1.5$, $SD_{49} = 7.4$, $M_{17} = 3.0$, $SD_{17} = 9.5$, $t_{49\sim17}(12{,}227) = 13.3$, $p < .001$, $d = .17$, CI = [.14, .21]).

Due to the large sample, all changes are significant. While the effect size was moderate to large for We-language, it was small for I-language and specified We-pronouns.

Regression Models

Group comparisons (see ESM 1, Appendix A) and correlation matrices (see ESM 1, Appendix C, Tables C2-3) already suggested a link between We-language and reelection. To examine RQ2–6, our regression analysis comprised three steps: Model 1 examined predictors individually; Model 2 combined We-language with an additional predictor; and Model 3 explored their interaction (see summary Table 2).

First, We-language (personal pronouns) was a significant predictor for reelection, with one *SD* increase boosting next-term reelection chances by 5% and total reelections by almost half a term (RQ2).

Second, when including speech time in the regression, We-language retained its predictive value, with comparable size to speech time for total reelections (RQ3). For both next-term and total reelections, a significant interaction indicated that We-language were more predictive the more speech time a speaker had, yet remained significant for speakers with 1.5–2 *SD*s below the average

Table 2. Regressions on reelection: standardized estimates and standard errors

		Model 2		Model 3		
Predictor	Model 1	We (1)	Pred (2–4)	We (1)	Pred (2–4)	Int
Next-term reelection (random-effects model)						
1. We[a]	.05*** (.01)					
2. Speech time[b]	.11*** (.00)	.02*** (.01)	.10*** (.01)	.03*** (.01)	.10*** (.01)	.01*** (.00)
3. Our[c]	.03*** (.00)	.04*** (.01)	.01* (.01)	.04*** (.01)	.01† (.01)	−.01* (.00)
4. We specified[d]	.01* (.00)	.05*** (.01)	−.00 (.01)	.05*** (.01)	−.01 (.01)	.00 (.00)
Total reelections (between-effects model)						
1. We[a]	.42*** (.03)					
2. Speech time[b]	.49*** (.03)	.30*** (.04)	.41*** (.03)	.43*** (.04)	.44*** (.03)	.16*** (.03)
3. Our[c]	.16*** (.03)	.42*** (.04)	−.00 (.04)	.43*** (.04)	.00 (.04)	.02 (.03)
4. We specified[d]	.06† (.04)	.44*** (.04)	−.08* (.04)	.47*** (.04)	−.10* (.04)	.05 (.03)

Note. N1 = 9,243 (random-effects model). *N2* = 3,620 (between-effects model). Models control for parl. term (factor) and the random-effects model also for number of previous terms in parliament (factor). See ESM 1, Appendix C, Table C4 for models with additional control variables. [a]# of first-person plural personal pronouns per 1,000 words. [b]# of words. [c]# of first-person plural possessive pronouns per 1,000 words. [d]# of first-person plural personal pronouns with syntactic specification per 1,000 words (see measurement part).
†$p < .1$. *$p < .05$. **$p < .01$. ***$p < .001$.

speech time, indicating its value for both front- and backbenchers (see ESM 1, Appendix C, Figure C1).

Third, while Our-language (possessive pronouns) was a significant predictor individually, it lost its predictive value entirely for total reelections and nearly entirely for next-term reelection when We-language was included in the regression (RQ4). Additionally, for next-term reelection, the negative interaction effect indicated even decreased predictive power. Thus, We-language rather than Our-language predicts reelection.

Fourth, We-specifications were individually a significant predictor for reelection (RQ5). However, when combined with We-language, We-specifications lost its predictive value for next-term reelection and even had a negative predictive value for total reelections. The interactions were not significant. Thus, across We-language levels, We-specifications have rather a negative than a positive influence.

Fifth, We-categories differed in being significant predictors for reelection (RQ6). For next-term reelection, We-as-state ($\beta = .01$, $SE = .00$, $p = .05$) and We-as-party ($\beta = .01$, $SE = .00$, $p < .01$) were significant predictors, while We-as-Europe ($\beta = .00$, $SE = .00$, $p = .90$), We-as-nation ($\beta = -.00$, $SE = .00$, $p = .58$), and We-as-parliament ($\beta = .00$, $SE = .00$, $p = .99$) were not. For total reelections, the pattern reversed. We-as-Europe ($\beta = .12$, $SE = .04$, $p < .01$), We-as-nation ($\beta = .07$, $SE = .03$, $p = .03$), and We-as-parliament ($\beta = .09$, $SE = .04$, $p = .02$) were significant predictors, while We-as-state ($\beta = .00$, $SE = .03$, $p = .89$) and We-as-party ($\beta = -.03$, $SE = .03$, $p = .32$) were not. However, the overlapping confidence intervals suggest nonsignificant differences between We-categories, making it difficult to interpret them (see ESM 1, Appendix C, Figure C3).

Importantly, our findings remained robust, even when including the variables age, gender, academic title, party, and power (see ESM 1, Appendix C, Table C4, and Figure C2 and C4). Notably, We-language retained a significant predictor for both next-term and total reelections, with only slight reductions in estimate sizes. This highlights its predictive value across various social and political factors that also in our sample were associated with both language use and reelection outcomes.

Discussion

Our study aimed to enhance our understanding of the interplay between real-life IL, language, and political success, capitalizing recent advancements in political speech datasets and computational language analysis. Findings, considering around 350,000 speeches, revealed a long-term increase in IL language in the German parliament, with We-language surpassing I-language in recent years for the first time since 1949 (RQ1). IL language, specifically We-pronouns, robustly predicted both next-term reelection and total reelections into parliament (RQ2). The analysis, including over 3,600 political leaders from different leader levels, indicated that IL language was more predictive for frontbenchers but remained relevant for backbenchers with far less speech time (RQ3). We-language (personal) compared to Our-language (possessive pronouns) were not only the dominant driver behind IL increase in parliament but also as the stronger predictor for reelections, with Our-language in some cases reducing predictive power (RQ4). NLP-based analysis revealed more nuanced findings on We-language. We-specifications combined with We-language lost their predictiveness for next-term reelection and even negatively predicted total reelections (RQ5). Higher-level We-categories like We-as-nation or We-as-parliament predicted total reelections, while lower-level We-categories like We-as-party predicted next-term reelection (RQ6). All findings held robustly beyond various social and political factors associated both with language use and reelection outcomes.

Research Implications

These findings substantially expand empirical evidence for the prevalence and effectiveness of IL in the political realm. Adding to previous language trend analysis in English-speaking countries (Jordan et al., 2019), our study is the first to indicate a long-term increase in We-language in the German political context. Our study is also the first to illustrate IL effectiveness across a broad range of political leaders beyond top level figures (Haslam et al., 2023; Steffens & Haslam, 2013; Steffens et al., 2014) and their mediators (Gleibs et al., 2018), speaking to the fact that IL is not only an analytical lens (Haslam et al., 2020) and teachable skill (Haslam et al., 2017) but increasingly adopted as real-life political leadership praxis.

This study advances our understanding of different components of IL language and their impact on IL effectiveness. While previous studies incorporated pronouns in broader leadership concepts (Jordan et al., 2019; Pennebaker, 2011), investigated We-use compared to other pronouns (Chulvi et al., 2024), inconsistently considering possessive pronouns (Fladerer et al., 2021; Steffens & Haslam, 2013), our study showed that We-(personal pronouns) and not Our-language (possessive pronouns) drives IL trends and more strongly predicts political success, supporting the differentiation between collective identity ("we") and collective ownership ("our," Verkuyten & Martinovic, 2017). Furthermore, our negative finding on We-specifications supports that implicit or vague collective identities might be advantageous for

leaders (Condor et al., 2013) and challenges to some extent emphasis on salient and explicit IL ("crafting a sense of us," Haslam et al., 2020; see also Hogg et al., 2012). While previous research found in speeches of elected presidential candidates more higher-level We-categories (nation and government) but not lower-level We-categories (party; Steffens & Haslam, 2013), our study includes some evidence that higher-level We-categories (nation or parliament) might predict long-term and lower-level We-categories (party) short-term success, adding to research on super- and subordinate identities, intergroup leadership, and partisanship (Huddy & Bankert, 2017; Pittinsky, 2009; Reicher & Hopkins, 2000).

Finally, this study demonstrates the utility of new NLP-based tools to investigate nuanced language and leadership dynamics in extensive datasets. We developed a linguistic indicator for We-specifications based on syntactic dependency parsing. This also allowed for large-scale semantic analysis, offering valuable insights into the contextual meaning of "we" and enabling the examinations of specific We-categories. These methodological advancements represent a significant step beyond traditional word-counting (Boyd & Schwartz, 2021; Tausczik & Pennebaker, 2010).

Limitations and Future Directions

The study estimated the association between IL language and reelections, though causality was not established. It introduced innovative measures for We-specification and We-categories but these measures need validation through association with other IL indicators and replication across diverse contexts for robustness. Despite moving beyond word counting methods, the study still reduces the linguistic complexity of IL to predefined syntactic constructions and semantic categories. Future research could further leverage NLP tools to investigate IL language and its complexity. While our We-specification measure might serve as a rudiment "identity entrepreneurship" indicator, similar indicators could be developed for the remaining three IL dimensions, identity prototypicality, advancement and impresarioship (Haslam et al., 2020). While we opted for a transparent indicator, future studies might consider training algorithms or employing existing NLP language models to identify IL language (Banks et al., 2023; Rathje et al., 2023). Finally, future research could tap into the extensive parliamentary dataset to investigate heterogeneity in IL language success across historical periods, political ideologies, power dynamics, and inequality-related variables (Houck et al., 2022; Jost & Krochik, 2014; Kacewicz et al., 2014; Rouhizadeh et al., 2016; Shen, 2024).

Practical Implications

The study's implications extend to politicians and practitioners in politics. The linguistic indicators can be valuable tools for their own analyses of political speeches, leadership skills, and the alignment between leaders, groups, and political agendas from a social identity lens. The strategic use of We-language emerges as a concrete means for politicians to enhance political effectiveness and improve reelection prospects. The study highlights that the effectiveness of IL might rely on We-language rather than Our-language and a careful We-specification and We-category selection. Politicians should balance concreteness and identity levels in their communication strategies for success in both the short and long term.

Conclusion

In summary, this paper contributes to our understanding of the relationship between social identity leadership, language, and political success. By leveraging an extensive political speech dataset and new computational text analytical methods, we have explored the evolution of We-language in the German parliament and its predictive capacity for political success for a wide range of leaders. This research not only differentiated between different IL language components and their effectiveness but also introduced innovative indicators for investigating IL language with automated text analysis in large-scale datasets. Ultimately, this study significantly expands the evidence for the increasing importance and potential of IL language in the political realm.

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/2151-2604/a000564

ESM 1. Appendix A: Group comparisons (additional analysis). B: Psychological ownership and reelection (additional analysis). C: Additional tables and figures. D: spaCy DependencyMatcher code.

References

Augoustinos, M., & De Garis, S. (2012). 'Too black or not black enough': Social identity complexity in the political rhetoric of Barack Obama. *European Journal of Social Psychology*, *42*(5), 564–577. https://doi.org/10.1002/ejsp.1868

Bachmann, R., & Gleibs, I. H. (2024, April 24). *Uncovering the secret life of We-pronouns in the German parliament: Computational text analysis in a large-scale speech dataset* [Pre-registration]. https://osf.io/nhae8

Banks, G. C., Ross, R., Toth, A. A., Tonidandel, S., Goloujeh, A. M., Dou, W., & Wesslen, R. (2023). The triangulation of ethical leader signals using qualitative, experimental, and data science methods. *The Leadership Quarterly*, *34*(3), Article 101658. https://doi.org/10.1016/j.leaqua.2022.101658

Berger, J., & Packard, G. (2022). Using natural language processing to understand people and culture. *American Psychologist*, *77*(4), 525–537. https://doi.org/10.1037/amp0000882

Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, *40*(1), 21–41. https://doi.org/10.1177/0261927X20967028

Bracht, E. M., Monzani, L., Boer, D., Haslam, S. A., Kerschreiter, R., Lemoine, J. E., Steffens, N. K., Akfirat, S. A., Avanzi, L., Barghi, B., Dumont, K., Edelmann, C. M., Epitropaki, O., Fransen, K., Giessner, S., Gleibs, I. H., González, R., Laguía González, A., Lipponen, J., … van Dick, R. (2023). Innovation across cultures: Connecting leadership, identification, and creative behavior in organizations. *Applied Psychology*, *72*(1), 348–388. https://doi.org/10.1111/apps.12381

Brewer, M. B. (2001). The many faces of social identity: Implications for political psychology. *Political Psychology*, *22*(1), 115–125. https://doi.org/10.1111/0162-895X.00229

Chulvi, B., Molpeceres, M., Rodrigo, M. F., Toselli, A. H., & Rosso, P. (2024). Politicization of immigration and language use in political elites: A study of Spanish parliamentary speeches. *Journal of Language and Social Psychology*, *43*(2), 164–194. https://doi.org/10.1177/0261927X231175856

Condor, S., Tileagă, C., & Billig, M. (2013). Political rhetoric. In L. Huddy, J. S. Levy, & D. O. Sears (Eds.), *The Oxford handbook of political psychology* (2nd ed., pp. 262–297). Oxford University Press.

Dehghani, M., & Boyd, R. L. (Eds.). (2022). *Handbook of language analysis in psychology*. Guilford Publications.

Fløttum, K. (2010). EU discourse: Polyphony and unclearness. *Journal of Pragmatics*, *42*(4), 990–999. https://doi.org/10.1016/j.pragma.2009.08.014

Fladerer, M. P., Haslam, S. A., Steffens, N. K., & Frey, D. (2021). The value of speaking for "us": The relationship between CEOs' use of I-and we-referencing language and subsequent organizational performance. *Journal of Business and Psychology*, *36*(2), 299–313. https://doi.org/10.1007/s10869-019-09677-0

Gleibs, I. H., & Haslam, S. A. (2016). Do we want a fighter? The influence of group status and the stability of intergroup relations on leader prototypicality and endorsement. *The Leadership Quarterly*, *27*(4), 557–573. https://doi.org/10.1016/j.leaqua.2015.12.001

Gleibs, I. H., Hendricks, K., & Kurz, T. (2018). Identity mediators: Leadership and identity construction in campaign speeches of American presidential candidates' spouses. *Political Psychology*, *39*(4), 939–956. https://doi.org/10.1111/pops.12448

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, *150*(4), 700–709. https://doi.org/10.1037/xge0000920

Haslam, S. A., & Platow, M. J. (2001). The link between leadership and followership: How affirming social identity translates vision into action. *Personality and Social Psychology Bulletin*, *27*(11), 1469–1479. https://doi.org/10.1177/01461672012711008

Haslam, S. A., Reicher, S. D., & Platow, M. J. (2020). *The new psychology of leadership: Identity, influence and power* (2nd ed.). Routledge.

Haslam, S. A., Reicher, S. D., Selvanathan, H. P., Gaffney, A. M., Steffens, N. K., Packer, D., Van Bavel, J. J., Ntontis, E., Neville, F., Vestergren, S., Jurstakova, K., & Platow, M. J. (2023). Examining the role of Donald Trump and his supporters in the 2021 assault on the US Capitol: A dual-agency model of identity leadership and engaged followership. *The Leadership Quarterly*, *34*(2), Article 101622. https://doi.org/10.1016/j.leaqua.2022.101622

Haslam, S. A., Steffens, N. K., Peters, K., Boyce, R. A., Mallett, C. J., & Fransen, K. (2017). A social identity approach to leadership development: The 5R Program. *Journal of Personnel Psychology*, *16*(3), 113–124. https://doi.org/10.1027/1866-5888/a000176

Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review*, *5*(3), 184–200. https://doi.org/10.1207/S15327957PSPR0503_1

Hogg, M. A., van Knippenberg, D., & Rast I. I. I, D. E. (2012). The social identity theory of leadership: Theoretical origins, research findings, and conceptual developments. *European Review of Social Psychology*, *23*(1), 258–304. https://doi.org/10.1080/10463283.2012.741134

Hopkins, N. (2023). Political Language. In L. Huddy, D. O. Sears, J. S. Levy, & J. Jerit (Eds.), *The Oxford handbook of political psychology* (3rd ed., pp. 310–346). Oxford University Press.

Houck, S. C., Conway, L. G., III, & Zubrod, A. (2022). Automated integrative complexity: Language analysis tools for psychological research. In M. Dehghani & R. L. Boyd (Eds.), *Handbook of language analysis in psychology* (pp. 173–188). Guilford Press.

Huddy, L. (2001). From social to political identity: A critical examination of social identity theory. *Political Psychology*, *22*(1), 127–156. https://doi.org/10.1111/0162-895X.00230

Huddy, L., & Bankert, A. (2017). Political partisanship as a social identity. In W. R. Thompson (Ed.), *Oxford research encyclopedia of politics* (pp. 1–30). Oxford University Press.

Ilie, C. (2017). Parliamentary debates. In R. Wodak & B. Forchtner (Eds.), *The Routledge handbook of language and politics* (pp. 306–325). Routledge.

Jackson, J. C., Watts, J., List, J. M., Puryear, C., Drabble, R., & Lindquist, K. A. (2022). From text to thought: How analyzing language can advance psychological science. *Perspectives on Psychological Science*, *17*(3), 805–826. https://doi.org/10.1177/17456916211004899

Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends in politics and culture through language of political leaders and cultural institutions. *Proceedings of the National Academy of Sciences*, *116*(9), 3476–3481. https://doi.org/10.1073/pnas.1811987116

Jost, J. T., & Krochik, M. (2014). Ideological differences in epistemic motivation: Implications for attitude structure, depth of information processing, susceptibility to persuasion, and stereotyping. In A. J. Elliot (Ed.), *Advances in motivation science* (pp. 181–231). Elsevier Academic Press.

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, *33*(2), 125–143. https://doi.org/10.1177/0261927X13502654

Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2018). *"LIWC auf Deutsch": The development, psychometrics, and introduction of DE-LIWC2015*. PsyArXiv. https://doi.org/10.31234/osf.io/uq8zt

Mols, F., Haslam, S. A., Platow, M. J., Reicher, S. D., & Steffens, N. K. (2023). The social identity approach to political leadership. In L. Huddy, D. O. Sears, J. S. Levy, & J. Jerit (Eds.), *The Oxford handbook of political psychology* (3rd ed., pp. 804–442). Oxford University Press.

Mols, F., Jetten, J., & Haslam, S. A. (2009). EU identification and endorsement in context: The importance of regional identity salience. *Journal of Common Market Studies*, *47*(3), 601–623. https://doi.org/10.1111/j.1468-5965.2009.01817.x

Monzani, L., Bibic, K., Haslam, S. A., Kerschreiter, R., Wilson-Lemoine, J. E., Steffens, N. K., Akfirat, S. A., Ballada, C. J. A., Bazarov, T., Aruta, J. J. B. R., Avanzi, L., Bunjak, A., Černe, M., Edelmann, C. M., Epitropaki, O., Fransen, K., García-Ael, C.,

Giessner, S., Gleibs, I., … & van Dick, R. (2024). Political leaders' identity leadership and civic citizenship behavior: The mediating role of trust in fellow citizens and the moderating role of economic inequality. *Political Psychology*. Advance online publication. https://doi.org/10.1111/pops.12952

Nijs, T., Martinovic, B., Verkuyten, M., & Sedikides, C. (2021). 'This country is OURS': The exclusionary potential of collective psychological ownership. *British Journal of Social Psychology*, *60*(1), 171–195. https://doi.org/10.1111/bjso.12386

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Press.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pittinsky, T. L. (2009). *Crossing the divide: Intergroup leadership in a world of difference*. Harvard Business Press.

Rathje, S., Mirea, D. M., Sucholutsky, I., Marjieh, R., Robertson, C., & Van Bavel, J. J. (2023). *GPT is an effective tool for multilingual psychological text analysis*. PsyArXiv. https://doi.org/10.31234/osf.io/sekf5

Reicher, S., & Hopkins, N. (2000). *Self and nation*. Sage.

Richter, F., Koch, P., Franke, O., Kraus, J., Warode, L., & Heine, S. (2023). *Open Discourse: Towards the first fully Comprehensive and Annotated Corpus of the Parliamentary Protocols of the German Bundestag*. SocArXiv. https://doi.org/10.31235/osf.io/dx87u

Rouhizadeh, M., Ungar, L., Buffone, A., & Schwartz, H. A. (2016, November). Using syntactic and semantic context to explore psychodemographic differences in self-reference. In J. Su, K. Duh, X. Carreras (Eds.), *Proceedings of the 2016 Conference on empirical Methods in Natural Language Processing* (pp. 2054–2059). Association for Computational Linguistics. https://doi.org/10.18653/v1/D16-1219

Shen, L. (2024). LIWC indicators of cross-country income inequality at the United Nations General Debate (1987–2020): An elastic net analysis. *Journal of Language and Social Psychology, 43*(1), 118–131. https://doi.org/10.1177/0261927X231190861

Steffens, N. K., & Haslam, S. A. (2013). Power through 'us': Leaders' use of we-referencing language predicts election victory. *PLoS ONE*, *8*(10), Article e77952. https://doi.org/10.1371/journal.pone.0077952

Steffens, N. K., Haslam, S. A., Jetten, J., & Mols, F. (2018). Our followers are lions, theirs are sheep: How social identity shapes theories about followership and social influence. *Political Psychology*, *39*(1), 23–42. https://doi.org/10.1111/pops.12387

Steffens, N. K., Haslam, S. A., & Reicher, S. D. (2014). Up close and personal: Evidence that shared social identity is a basis for the 'special' relationship that binds followers to leaders. *The Leadership Quarterly*, *25*(2), 296–313. https://doi.org/10.1016/j.leaqua.2013.08.008

Steffens, N. K., Munt, K. A., van Knippenberg, D., Platow, M. J., & Haslam, S. A. (2021). Advancing the social identity theory of leadership: A meta-analytic review of leader group prototypicality. *Organizational Psychology Review*, *11*(1), 35–72. https://doi.org/10.1177/2041386620962569

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–37). Brooks/Cole.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

van Dick, R., Lemoine, J. E., Steffens, N. K., Kerschreiter, R., Akfirat, S. A., Avanzi, L., Avanzi, L., Dumont, K., Epitropaki, O., Fransen, K., Giessner, S., González, R., Kark, R., Lipponen, J., Markovits, Y., Monzani, L., Orosz, G., Pandey, D., Roland-Lévy, C., … Haslam, S. A. (2018). Identity leadership going global: Validation of the Identity Leadership Inventory across 20 countries. *Journal of Occupational and Organizational Psychology*, *91*(4), 697–728. https://doi.org/10.1111/joop.12223

van Knippenberg, D. (2011). Embodying who we are: Leader group prototypicality and leadership effectiveness. *The Leadership Quarterly*, *22*(6), 1078–1091. https://doi.org/10.1016/j.leaqua.2011.09.004

Van Swol, L. M., & Kane, A. A. (2019). Language and group processes: An integrative, interdisciplinary review. *Small Group Research*, *50*(1), 3–38. https://doi.org/10.1177/1046496418785019

Verkuyten, M., & Martinovic, B. (2017). Collective psychological ownership and intergroup relations. *Perspectives on Psychological Science*, *12*(6), 1021–1039. https://doi.org/10.1177/1745691617706514

History
Received November 15, 2023
Revision received February 17, 2024
Accepted March 22, 2024
Published online July 12, 2024

Acknowledgments
We would like to thank the two anonymous reviewers and the editors of this special issue for valuable feedback to our manuscript.

Conflict of Interest
We have no conflicts of interest to disclose. No generative models were used for this study.

Open Science
We preregistered the project on the Open Science Framework page (https://osf.io/nhae8/; Bachmann & Gleibs, 2024).

Funding
This study was not funded. Open access publication enabled by London School of Economics and Political Science.

ORCID
Robin Bachmann
https://orcid.org/0000-0003-1673-2620
Ilka H. Gleibs
https://orcid.org/0000-0002-9913-250X

Robin Bachmann
Department of Psychological and Behavioural Science
London School of Economics and Political Science
Houghton St
London WC2A 2AE
United Kingdom
r.e.bachmann@posteo.de