Determining Political Text Complexity:
Conceptualizations, Measurements, and Application

PETRO TOLOCHKO
HAJO G. BOOMGAARDEN
University of Vienna, Austria

Language is of major importance in communicating politics to the people. Political texts, however, vary in their use of language beyond their particular contents. Some texts are structurally and linguistically more complex than others. Text complexity, in turn, has an impact on audiences' abilities to process and acquire information. Yet, the measures used to operationalize textual complexity are vastly different across studies. This study analyzes whether various complexity metrics measure similar underlying constructs and develops an approach to determine the complexity of texts. The current article provides evidence that text complexity is a multidimensional construct and thus should be studied more carefully. It subsequently validates this approach by applying it to a sample of political newspaper articles. The approach should inform future studies on the structure and effects of linguistic complexity on political communication and beyond.

Keywords: linguistic complexity, political communication, automated text analysis

Political action is the most evident manifestation of politics, and language plays a crucial role in motivating it. Language is the medium of politics, and—in addition to the actual content—the structure and characteristics of the medium matter. Political texts such as political news, speeches, social media utterances, and so forth, are important data that allow quantifying politics, and they vary widely in the way they are structured. Political information varies depending on the type of outlet, even when the subject matter is the same (e.g., Fowler, 2013), or involves added complexity in its presentation, for example, in the form of a hyperlink structure (e.g., Eveland & Cortese, 2004). Investigations into how information or language structure influences citizens and politics have been, and continue to be, of interest to the social sciences.

Political texts convey nontrivial information and are, therefore, by definition, complex. Evidence suggests that there are important differences in the level of complexity and in the structure of language used to produce such differences, across various political texts. Politicians vary in their language usage (Van Der Velden, Schumacher, & Vis, 2018), and tabloid newspapers may use a simpler style of language than quality newspapers (Tolochko & Boomgaarden, 2018). Differences in style and linguistic structure among

Petro Tolochko: petro.tolochko@univie.ac.at
Hajo G. Boomgaarden: hajo.boomgaarden@univie.ac.at

various media outlets have been apparent for a long time (e.g., McLuhan, 1964) and are still being investigated today (e.g., Baker, 2010). Political texts assume that the reader possesses at least some knowledge to understand them (i.e., basic political literacy); some texts require just the basics, whereas others demand an extensive repertoire of acquired knowledge. Given that citizens' understanding of political information, conveyed by journalists or politicians regarding political actors and policies, is a key component of a democratic society (Strömbäck, 2005), the complexity with which political information is presented is arguably among the key factors affecting citizens' involvement with and comprehension of political processes. But how complex political texts actually are or should be to foster comprehension and participation and what text complexity essentially is are still subjects of debate. Addressing such conceptual and measurement concerns advances the theoretical corpus of political communication as a field and facilitates a better understanding of political communication processes.

Most previous studies on text complexity have focused on its role in the construction of language and human development and hence relate to developmental psychology and linguistics (Solé, Corominas-Murtra, Valverde, & Steels, 2010). In psychology, complexity is regarded as a core attribute of communication (e.g., *integrative complexity*; Suedfeld & Tetlock, 1977). Research in political communication also points to considerable effects of the structure of a message on some aspects of citizens' political involvement, whether it be political socialization, acquisition of political information, or political information retainment (Eveland & Dunwoody, 2001; Kleinnijenhuis, 1991). These studies show differential effects of political information structures. People differ in their way of consuming differentially complex information (Prior, 2013), and different complexities influence learning processes (Eveland & Cortese, 2004; Eveland & Dunwoody, 2001). Although quite similar research exists linking text complexity to the processes of political socialization or political learning, the methodological approaches and theoretical concepts used in these studies are rather heterogeneous.

The vast majority of studies in the social sciences conceptualize complexity as *readability*, essentially measuring a single—syntactic—dimension of text complexity, inherently overlooking its equally important semantic counterpart. A distinction between syntactic and semantic text complexity is commonplace in linguistics (Ortega, 2003), but has not been acknowledged in the social sciences. Complexity is commonly operationalized based on simple ratios of word and syllable counts (Bischof & Senninger, 2018; Flesch, 1948; Spirling, 2016), which captures only a small slice of how textual complexity could manifest itself. The methodological and conceptual gaps in research on text complexity in political and communication science are surprising, given strong indications that such structures have nontrivial effects on political processes. To answer questions about text complexity effects, a more comprehensive conceptualization and corresponding measurement are needed. Conceptual clarification of text complexity would contribute to political communication insights, but potentially also in other domains of communication.

The goal of this study was threefold: (a) to show the dual nature of text complexity as a two-dimensional concept, (b) to provide a relatively holistic approach to determining the linguistic complexity of political texts from measures applied in various fields, and (c) to test and validate measurement instruments. The measures used here stem from various disciplines, yet, to the best of our knowledge, have never been employed concurrently and related to each other. We first determined whether the previous measures of text complexity form latent constructs relating to the main facets of linguistic structure in

political texts: semantic and syntactic complexity. Importantly, this study does not purport to "uncover" that there is more than one dimension of text given that it is a well-established fact in linguistics, and it would be futile to try to convince the reader about this fact again. We do not argue that the existence and applicability of various measures are inherently methodological flaws; conversely, this is an investigation into how (political) text complexity is operationalized and seen in the social sciences, the ways to measure it, and the consequences to the interpretation based on a single selected measure of text complexity. We tested our approach on a large corpus of political news articles from different types of newspapers across two different languages and validated it on additional sources. The study finally offers a brief supplementary validation of the measurements by considering audience perceptions of linguistic complexity. Establishing measurement of multiple indices of text complexity and analyzing them as various dimensions of this construct would provide a more precise and robust estimate of text complexity than using any one of them separately. Doing so, this study allows future research to systematically and comprehensively study the contents and consequences of (political) text complexity and to position themselves in the field of text complexity research.

Text Complexity in Communication Science

Two types of linguistic complexity are usually identified: objective complexity and agent-related complexity (Dahl, 2004). *Objective complexity* relates to the notion that complexity is an objective property of a system (e.g., a language). Although there is no universally agreed-on definition of objective complexity, the complexity of a system may be measured by the amount of information needed to re-create it (Dahl, 2004). Objective complexity, in turn, is divided into *system complexity* and *structural complexity*, in which system complexity deals with language as a system and encompasses the totality of all of its possible elements. Structural complexity deals with the structure of utterances, expressions, and their permutations. Sentences may be constructed according to the same grammatical rules, but consist of different structural elements and have various degrees of complexity. Finally, the notion of *agent-related complexity* deals with the difficulty of understanding language and the cost associated with doing so from an agent's perspective. In this study, the focus is on objective structural complexity, the formal characteristics of a language (Dahl, 2004), defined as "a formal property of texts and linguistic systems having to do with the number of their elements and their relational patterns" (Pallotti, 2015, p. 3).

A range of measures of structural text complexity has been used in the field of communication science. Kleinnijenhuis (1991) used readability—based on the Flesch Reading Ease Test (Flesch, 1948)—to measure text complexity, and information entropy (Shannon, Weaver, & Burks, 1963) to measure news saturation (the amount of information included in newspapers), while focusing on the relationship between individuals' education levels and complexity of political text they consume to explain the knowledge gap hypothesis (Tichenor, Donohue, & Olien, 1970). Notable recent examples applying text complexity in political communication include the analysis of British political elite speeches in the context of democratization of politics (Spirling, 2016) and the analysis of Austrian and German populist parties' rhetoric (Bischof & Senninger, 2018).

Most studies have found that different types of complexity affect people differently. Educated people benefit from more complex news, but added complexity proves to be detrimental to people with lower education (Kleinnijenhuis, 1991). Familiarity with the medium elicits positive effects on learning, whereas unfamiliarity

leads to detrimental effects of complexity (Eveland & Cortese, 2004). Complexity of political texts is likely to be an important factor in how people acquire, retain, and process political information, a prerequisite for participating in the political process. It may also explain heterogeneous learning patterns from different types of media or media genres. This strongly suggests that the subject of complexity deserves more careful attention to conceptualization and operationalization than it is currently receiving in the communication literature.

Conceptualizing Text Complexity and Its Dimensions

There are different approaches and multiple methods used to determine complexity in the social sciences and other fields. The reason for such methodological and theoretical diversity is mainly because there is no single established definition of text complexity. Although the ad hoc usage of the term is more prevalent in the social sciences than in linguistics, the terms *textual* or *linguistic complexity* are not exactly defined in the latter field as well (Kortmann & Szmrecsanyi, 2012). Another issue is that linguistic operationalizations are too technical or too specific and are difficult to apply in empirical social science research (to incorporate system complexity, for example, one would need a method to "measure" language in its entirety).

When referring to the complexity of language, scholars in the social sciences usually refer to the *readability of text* (e.g., Bischof & Senninger, 2018). One of the most prominent readability metrics is the Flesch Reading Ease score (Flesch, 1948), although many other indices exist. Essentially, this metric comprises ratios of words to syllables and words to sentences. There are two major problems with using the Flesch score to gauge the linguistic complexity of a political text. First, it was not designed to measure linguistic "complexity"; the Flesch Reading Ease Test was developed to measure how understandable and readable texts are for children (even more specifically, English texts from the first half of the 20th century). The second problem arises because of the multifaceted nature of natural language.

The study of language comprises various branches, with phonetics, phonology, morphology, pragmatics, syntax, and semantics being the overarching ones (Manning & Schütze, 1999, p. 112). The former four are likely less interesting from a social science perspective, but *syntax* (variations in linguistic structure) and *semantics* (variations in meanings) are two fundamental dimensions in operationalizations of political text complexity.[1] The readability index, dealing with sentence length, syllables, and words, is (a) inherently a syntactic measure, and (b) one that only superficially captures this aspect (hence, a crude proxy of the concept). Thus, when using readability to operationalize linguistic complexity, one is inevitably missing a large portion of variance. Linguistic complexity is known to be multidimensional in the field of applied linguistics (Ortega, 2003) and deserves that such multidimensionality is acknowledged in political and communication science.

The conceptual distinction between syntax and semantics is important. Research in linguistics, psychology, and cognitive science has shown that these dimensions have a different effect on information

---

[1] Using computational tools to measure true semantic characteristics of language most often requires handwritten rules (Jurafsky & Martin, 2014), and are almost intractable when dealing with large corpora. Thus, when referring to the "semantic" dimension, a *lexical* approximation is implied. Adjectives *semantic* and *lexical* are used interchangeably throughout this article.

processing and information acquisition. Semantic and syntactic disambiguation, a phenomenon that refers to understanding what exactly is meant by a linguistic unit, is a major area of research in linguistics and natural language processing (Jurafsky & Martin, 2014; Manning & Schütze, 1999) and is included in the complexity discourse. Semantic and syntactic complexity have varying effects on both short-term memory and recall in children (Daneman & Case, 1981), as well as reaction times and working memory in adults (Tan, Martin, & Van Dyke, 2017). Cognitive science studies have shown that manipulated errors in speech elicit a different brain reaction depending on the type of error—morphological, semantic, or syntactic (Friederici, Steinhauer, & Frisch, 1999)—or that different parts of the brain are responsible for processing semantic and syntactic cues (Rogalsky & Hickok, 2009). Communication scholars should benefit from a conceptual refinement and open a possibility for much more concrete questions tied to text complexity.

Moreover, semantics may relate to either *intensional* or *extensional* semantics (e.g., Von Fintel & Heim, 2011, pp. 102–103); therefore, semantic complexity may also be defined as either intensional or extensional (Frajzyngier, Hodges, & Rood, 2005, pp. 147–150). The fact that these two categories are inversely proportional makes the operationalization of semantic complexity somewhat problematic. Intensional semantic complexity assumes that the complexity of a lexical unit grows with the number of defining features associated with it; that is, a more abstract lexical unit, which required fewer features to describe, it is less complex than a more concrete one, which invariably demands more description. Conversely, the extensional approach assumes the reverse relationship: The more abstract a lexical unit is, the more complex it is because it allows for a broader range of interpretation and contextual cues. The approach taken in the current article is that of extensional semantic complexity. There are several reasons for this choice: First, our intention is to cover those indices that have been already prominently used to measure the linguistic complexity of the texts, and the vast majority of text complexity applications in the social sciences adhere to the extensional approach. In addition, quantitative approaches in linguistics also assume the extensional approach to semantics (Frajzyngier et al., 2005, pp. 143–157). Finally, having massive amounts of data necessarily restricts one choice of measures for reasons of computational tractability.

In what follows, we review and concurrently apply various measures of complexity used in previous research that are expected to reflect these two broad dimensions, and subsequently investigate their relative performances and sensitivities. In doing so, we focus on whether various manifested measures are as expected underlying either of the two dimensions of text complexity: semantics and syntax. Using various metrics that broadly capture semantic and syntactic dimensions, we are essentially applying multi-item scales to determine the text complexity, thus, providing a more accurate and robust measurement of the concept. We include a two-step validation procedure to provide face validity as well as predictive validity of the method.

Measuring Text Complexity

Text Complexity as Semantic Complexity

Several measures used to measure text complexity in the social sciences are outlined below. These are not the only indices employed for this task, but the section provides a broad overview of methods used to determine the complexity of a given text that are common across social science disciplines.

Lexical Complexity

Although there exists no measure to quantify the real semantic complexity of a text, a number of statistical methods have been developed to serve as proxy measures. *Lexical diversity* is the ratio of all unique words (*types*) encountered in the analyzed text to the total number of words (*tokens*). For example, the sentence "I am what I am" has three types: "I," "am" and "what," and five tokens and thus a lexical diversity of 3/5, or 0.6. *Lexical richness* is also used to measure the semantic complexity of a text; it is a probabilistic measure fitting a token frequency distribution onto a Poisson distribution (Yule, 2014).

Information Entropy

A notion of information entropy from information theory (Shannon et al., 1963) is also used to measure semantic complexity. Simply put, information entropy is a measure of uncertainty or unpredictability of information. Higher entropy of information requires more effort to decode. A semantic entropy (Dale, Moisl, & Somers, 2000) metric, which measures semantic ambiguity, is based on this notion. This metric determines how "difficult" it is to guess the next word in the corpus.

Text Complexity as Syntactic Complexity

Sentence Structure

Syntactic complexity overall relates to the formal structure of the text. Lu (2010), for example, proposes a method to determine syntactic complexity based on a set of measures chosen from meta-analysis research covering linguistic complexity studies (Ortega, 2003). These measures include mean length of a word, mean length of a sentence, and verb phrases per sentence, among others. Other, more complicated measures rely on sentence parsing and use the sentence tree structure to determine complexity.

Textual Readability

Readability is perhaps the most straightforward and the most popular method of determining text complexity (see, e.g., Baumann, 2005), the most prominent of which is the Flesch Reading Ease Test (Flesch, 1948). Readability indices use (some combination of) the ratio of the total number of words in the text to the total number of sentences, and the ratio of the total number of syllables to the total number of words. This standardized test, which is obviously unable to encompass the entire complexity of language, provides a basic framework to identify complexity. A text with a higher readability score is assumed to be less complex than a text with a lower readability score. Textual readability—as defined by the readability indices—is inherently a measure of the syntactic structure of text because it exclusively deals with the lengths of grammatical units (words and sentences) and ignores their permutations, repetitions, and so forth, which could at least be thought of as a proxy measure of lexical content.

Empirical Assessment of Text Complexity in Political News

Combining these linguistic dimensions will provide a more holistic approach to determining the complexity of a political text. Drawing on these measures, we examined a large corpus of political news, performing an analysis including the different types of metrics discussed above. An exploratory factor analysis established underlying text complexity dimensions from the manifest measures discussed above. The extracted components were then used as measures of complexity dimensions. To illustrate and provide an empirical application of these measures, we compared complexity scores between tabloid and quality newspapers to show the structural difference between these two press types.

For validation purposes, we used two supplementary text data sources to provide face validity for our approach: legal cases from the European Court of Human Rights and children's books. Finally, a crowdsourced experiment was designed to provide the predictive validity of the approach.

Drawing on linguistic theories and the nature of the complexity metrics discussed above, we expected that the structure of linguistic complexity would comprise two distinct subdimensions: semantic and syntactic.

H1:   *The measures commonly used to measure text complexity constitute two separate dimensions*: *semantic and syntactic.*

Furthermore, even if the data actually support the aforementioned hypothesis, the value of differentiating among the complexity dimensions will depend on whether such differentiation brings not only additional conceptual precision, but also practical benefit. An example that can be straightforwardly applied to communication science is the difference in linguistic complexity of tabloid and quality media. If the data show that (a) linguistic complexity comprises two separate (semantic and syntactic) dimensions, and (b) this difference is observable in applications that are important to communication science, one has a very strong argument to structurally differentiate linguistic complexity based on its dimensions.

Prior research on differences among newspapers provides essential information allowing us to consider the face validity of the dimensionality of complexity measures. Comparative studies analyzing quality and tabloid media usually note that not only do these two media have a distinct readership, but also the language they use is explicitly different even when they are covering the same issue (Fowler, 2013). Tabloid newspapers often present an oversimplification in lieu of objective coverage (Zelizer et al., 2000). Sometimes, the linguistic difference is so pronounced that what quality newspapers consider a good style might be considered bad by tabloids (and vice versa; Bagnall, 1993). A study that compared British quality and tabloid newspapers' noun phrase complexity (Jucker, 1992) suggests that quality newspapers tend to be more syntactically complex than tabloids. A modal and lexical analysis of British broadsheet and tabloid newspapers showed that tabloids use a more diverse vocabulary and their language is more emotional and biased (Timuçin, 2010). We refrained from formulating a directional hypothesis based on the lack of empirical evidence; however, we anticipated a difference in the structure of tabloid and quality newspapers:

RQ1:   Do tabloids and quality newspapers differ on the syntactic dimension?

RQ2:   Do tabloids and quality newspapers differ on the semantic dimension?

These latter two research questions serve to showcase an application and provide initial empirical results of our approach of measuring text complexity as a two-dimensional concept.

Method

To address our expectations and test the applicability of the method, we performed a series of analyses on a large corpus of English- and German-language newspapers (*N* = 70,320 newspaper articles). A series of Python programming scripts was used for the data processing and text data extraction, including natural language processing, based on Python's spaCy (Honnibal & Johnson, 2015) module. The preprocessing part included lemmatization and stopwords removal; thus, inflection was taken into account on the lexeme level and plural terms were transformed into a singular form.

Data

The sample included political news coverage during the entire year 2014 from quality and tabloid newspapers of the United States (*The New York Times* vs. *New York Daily News*), the United Kingdom (*The Guardian* vs. *The Mirror*), and Germany (*Frankfurter Allgemeine Zeitung* vs. *Bild*). Because the focus of the present investigation was on the textual complexity of political news articles, articles that were not related to political topics were filtered out during the initial data-gathering phase. Individual newspaper articles were chosen as the basic unit of observation.

Two additional data sources were used for validation purposes of the proposed method. First, a comparison was made with nonpolitical texts using a sample of 968 legal cases (from the European Court of Human Rights) and 24 children's books (a random sample from Project Gutenberg's "Children's Literature" section). We expected that, for the initial face validity, legal cases would show higher levels of complexity and children's books would be less complex on both dimensions.

To further examine the predictive validity of our approach, we conducted a crowdsourcing experiment to determine whether articles identified as complex were perceived as complex by the audience. Participants were recruited from the Crowdflower crowdsourcing platform (*N* = 427). In addition, three texts with high complexity and three texts with low complexity scores on all dimensions were randomly selected from the newspaper corpus. In total, 90 participants were filtered out from further analysis.[2] Thus, the final data set comprised 337 cases. The questions to gauge perceived complexity were "How difficult is it to comprehend the article?" "How easy is it to understand the grammatical structure of the article?" "How easy

---

[2] Each participant saw only one of the experimental texts. Every text included an attention check question: "This is a test sentence, please choose answer 1 (*very easy*) when answering this question." Participants who chose a different answer were subsequently excluded.

is it to understand the vocabulary of the article?" and "How much information is conveyed by the article?" All items were combined into a single scale with Cronbach's α = .83.

Measures of Text Complexity

Syntactic Complexity

Simple metrics such as average word length and average sentence length were the most basic syntactic complexity measures. Syntactic depth was calculated by modeling the sentence as a syntactic tree and counting the number of links from the root node to the terminal node. Syntactic dependency is a measure similar to syntactic depth; however, it comes from a different theoretical approach to syntax (Percival, 1990). The verb of a sentence is taken to be the central unit and all other syntactic units are directly or indirectly connected to it. The length of dependency was calculated by counting the longest dependency chain in the sentence, starting from the central verb of the sentence. The automated readability index (ARI) was chosen to gauge the readability of political texts, in which higher values of ARI indicate more complex texts. ARI was chosen because it is calculated by manifest linguistic variables (such as word length, sentence length, etc.) without using language-specific measures such as, for example, syllables. This feature makes ARI more or less language independent, which was an important factor for this research design.

Semantic Complexity

A number of measures were used to determine the semantic complexity of the political texts. Lexical diversity is the simplest measure of textual lexical complexity. It is the ratio of all unique words (types) encountered in the analyzed text to the total number of words (tokens). The Measure of Textual Lexical Diversity (MTLD) is a more complex measure of lexical diversity (McCarthy & Jarvis, 2010). The algorithm cumulatively calculates the type/token ratio for a text sequence. When it reaches a specific type/token ratio (0.72 is often taken as a cutoff point), it adds 1 to the number of total factors and resets starting from the word it stopped at, with a type/token ratio of 1.00; thus, the type/token ratio decreases as the text is parsed. This algorithm is applied to the whole length of the text, and in the end the total number of words is divided by the total factor count. For example, if the text length is 500 words, and there are six factors, then MTLD = 500/6 = 83.34. Although the type/token ratio and even the more sophisticated MTLD do not directly measure the complexity of utterances, they do provide information about the heterogeneity of text (in terms of lexical content) and thus, assuming the extensional definition of semantics, by proxy measure their complexity.

Lexical richness was measured using Yule's *I* metric (Yule, 2014, pp. 57–60). The measure is also based on the word frequencies in a text. Yule's measure of lexical richness was initially developed for author identification, and essentially, it measures the probability of two randomly selected words from a text being the same. Although Yule's *I* metric seems similar to lexical diversity measures, it captures a broader concept of word distributions over the text (rather than a simple ratio of types to tokens) and is related to the so-called "constancy" measures (Tanaka-Ishii & Aihara, 2015).

Finally, the semantic entropy measure is based on information entropy (Dale et al., 2000, p. 551). It is a logarithmic measure of uncertainty. Higher scores indicate a higher lexical variation in a sample, meaning higher semantic complexity. Semantic entropy estimates the uncertainty of encountering a certain lexical unit. In essence, it tries to quantify the same effect as lexical richness, although from a different mathematical perspective. The combination of these two measures should provide a robust estimation of lexical variability in text.

All of the "semantic complexity" indices presented above are proxies for semantic complexity. Because true measures of semantic complexity often require handwritten parsing rules and ad hoc assumptions about the representation of knowledge (e.g., Jurafsky & Martin, 2014), it would be counterproductive to try to implement them in a large-scale text analysis setting.

Results

Complexity Dimensions

A factor analysis was performed with Oblimin rotation to allow for cross-factor correlations. A base model with no rotation extracted two factors with Eigenvalues higher than 1.00 (KMO = 0.74), and the Bartlett test of sphericity was significant ($p < .001$). A two-factor model was then estimated, as can be seen in Table 1 (the zero-order correlation matrix is presented in Table A1 in the Appendix). The unstandardized descriptives are reported in Table A2 in the Appendix.[3]

---

[3] As can be seen from Table A2, the mean scores for the English and German samples are not uniform in some cases. This may be a function of the language difference (e.g., longer German mean word length [$M = 6.00$, $SD = 0.43$] opposed to the mean English word length [$M = 4.86$, $SD = 0.27$]). It may be unreasonable to compare German and English texts directly on their respective factor scores; however, given that the pattern is virtually the same across the languages, one could easily make directional comparisons (e.g., quality newspapers are more syntactically complex than tabloids for both English and German cases).

Table 1. Factor Analysis Models.

Variable	Model I Combined		Model II German		Model III English	
	Semantic	Syntactic	Semantic	Syntactic	Semantic	Syntactic
Lexical diversity	.93		.83		.95	
Yule's I	.82		.87		.83	
Semantic entropy	.93		.85		.95	
Measure of Textual Lexical Diversity	.89		.93		.82	
Sentence length		.87		.91		.95
Automated readability index		.94		.87		.94
Syntactic depth		.86		.92		.97
Syntactic dependency		.87		.92		.97
Cross-factor ρ	−.42***		−.42***		−.14***	
Eigenvalues	5.01	1.86	4.62	1.96	4.03	3.00
Variance	0.45	0.42	0.44	0.39	0.46	0.42
Cumulative variance	0.45	0.87	0.44	0.83	0.46	0.88
χ^2	13113.57***		4886.58***		8120.15 ***	

Note. Model statistics from factor analysis for the combined language model, as well as English and German languages. Factor analysis scores below .5 are omitted.
*p < .05. **p < .01. ***p < .001.*

The combined data reduction model and separate models for each language-specific sample overall strongly confirmed our expectations. The variables that were thought to measure the semantic complexity of the text—lexical diversity, semantic entropy, lexical richness, and MTLD—loaded on a single factor, further referred to as the Semantic factor. The second component incorporated the syntactic structure measures—ARI, average sentence length, syntactic depth, and syntactic dependency length. All of the variables loaded highly on one of the extracted factors (> .8).

The results revealed that the cross-factor correlation was negative for the combined data set, ρ = −.43 (*p* < .001); the German-only data set (i.e., *Frankfurter Allgemeine Zeitung* and *Bild*), ρ = −.42 (*p* < .001); and the English-only data set (i.e., *The New York Times*, *New York Daily News*, *The Guardian*, and *The Mirror*), ρ = −.14 (*p* < .001). This suggests that complexity is not only a multidimensional construct, but that its dimensions also are negatively correlated with each other. Texts or corpora thus can be more complex on one and less complex on another dimension. Here, it is important to note that the negative correlation coefficient was not a result of some measurement artifact (e.g., the MTLD not only measures semantic diversity, but also the opposite of syntactic complexity) because the operationalizations and mathematical formulas of aforementioned measures, while obviously measuring words, incorporated

different aspects of text that would not create negative correlations simply as an artifact (several formulas used in the current study are reported in Table A3 in the Appendix). Overall, the factor model confirmed our hypothesis: These metrics indeed are composed of the two linguistic dimensions of text complexity: semantic and syntactic.

Complexity Across Newspaper Types

Having established the dimensionality of the existing measures of text complexity, we further examined the differences among newspaper types (i.e., quality vs. tabloid newspapers), as reported in Figure 1. For the purpose of this comparison, we created composite "factor score" variables, one for each complexity dimension, by taking extracted scores from previous results (hereafter "factor scores"). All factor scores were then normalized to facilitate further analyses.
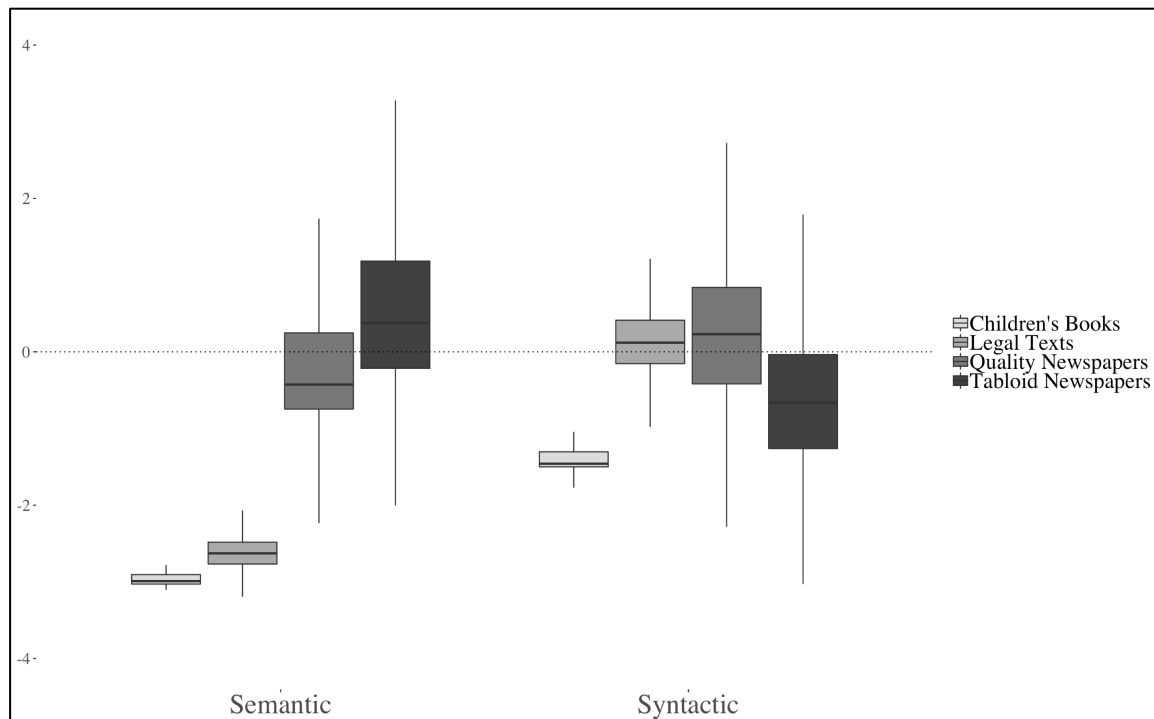

Figure 1. Complexity dimensions by outlet type.

A series of one-way analyses of variance was conducted to determine whether the selected newspapers differed in their mean scores on the two complexity factors. All models yielded significant differences for semantic complexity scores, $F(3, 71308) = 6121$, $p < .001$, and for syntactic complexity scores, $F(3, 71308) = 4033$, $p < .001$. A post hoc comparison (using Tukey's honestly significant difference test at a 95% confidence interval) yielded a significant difference across all means of all individual text types and individual newspapers. Therefore, all reported models and mean differences are significant at $p < .001$.

Semantic Complexity Model

Quality newspapers ($M = -0.17$, $SD = 0.84$) scored significantly lower than tabloid newspapers ($M = 0.56$, $SD = 1.04$) on semantic complexity. Interestingly, this pattern also held when comparing newspapers by country. *Frankfurter Allgemeine Zeitung* ($M = 0.87$, $SD = 0.80$) scored significantly lower than *Bild* ($M = 1.41$, $SD = 0.93$). *The Guardian* ($M = -0.42$. $SD = 0.55$) scored lower than the *Daily Mirror* ($M = 0.03$, $SD = 0.70$). Finally, *The New York Times* ($M = -0.55$, $SD = 0.48$) scored lower than the *Daily News* ($M = -0.09$, $SD = 0.53$).

Syntactic Complexity Model

The results were reversed for the syntactic complexity scores. Quality newspapers ($M = 0.24$, $SD = 0.90$) scored significantly higher than tabloid newspapers ($M = -0.61$, $SD = 0.99$). Once again, the same pattern held for the newspapers from individual countries. The German quality newspaper ($M = -0.62$, $SD = 0.51$) was much more syntactically complex than its tabloid counterpart ($M = -1.27$, $SD = 0.51$). *The Guardian* ($M = 0.34$, $SD = 0.88$) was significantly more complex when it came to formal structure than the *Daily Mirror* ($M = -0.14$, $SD = 0.68$). U.S. newspapers followed a similar trend: *The New York Times* ($M = 0.59$, $SD = 0.79$) scored higher on the syntactic complexity scale than the *Daily News* ($M = -0.17$, $SD = 1.28$).

In sum, both research questions were answered in the affirmative. Not only are quality newspapers syntactically more complex than tabloids, but also tabloids are semantically more complex than quality newspapers; this pattern also holds for newspapers from individual countries, as well as for both English and German languages. The latter points provide robust evidence for these differences.

Validation

To provide steps toward further validation of the approach, we compared nonpolitical texts (25 children's books and 1,000 legal texts) with the main corpus of newspaper texts. At face value, these text types should differ significantly in terms of their complexity, and findings should clearly display such differences.

Unsurprisingly, children's books proved to be the least syntactically complex texts in the sample ($M = -1.40$, $SD = 0.17$); legal texts ($M = 0.14$, $SD = 0.43$), however, turned out to be more syntactically complex even than tabloid newspapers, approaching the complexity of quality newspapers; the difference, however, was still significant at $p < .05$. Semantically, children's books were the least complex of all text types ($M = -2.96$, $SD = 0.01$), followed closely by legal texts ($M = -2.63$. $SD = 0.24$). In sum, this shows that the algorithm classified different text types according to face value expectations. Legal texts are expected to have a complex syntactical structure but employ a limited set of language cues (hence, the low score on semantic complexity). The language in children's books, as expected, is simpler than any other text type on both complexity dimensions because it is specifically designed to be understandable. Again, these findings also demonstrate the importance of distinguishing different types of complexity rather than relying on just one approach to measurement.

The small crowdsourcing experiment showed that texts that were classified as complex on all dimensions were perceived as more complex by human participants than less complex texts. A Welch's *t* test was performed to compare the means of two groups: high complexity texts versus low complexity texts. There was a significant difference in perceived complexity between the group that received the complex texts (*M* = 4.17, *SD* = 1.26) and the group that received the easy texts (*M* = 3.55, *SD* = 1.40), *t*(334.33) = 4.28, *p* < .01. Although this provides some broad validity to our complexity measures also in terms of peoples' perceptions, the question remains as to how different dimensions play into these perceptions, as further discussed below.

Discussion

This study deals with the structure of political texts, in particular with applying and analyzing various measures that are related to text complexity. Prior research on text complexity usually relies on very limited conceptual and measurement approaches. The focal interest here was to test whether different ways to determine text complexity can be used in conjunction with each other, that is, whether they measure a single construct of complexity or different latent subdimensions. To measure the complexity of political texts, we chose a range of measures from the fields of linguistics and social sciences based on their implementation in previous research. To test the measures and exemplify the approach, we looked at political news coverage in English- and German-language quality and tabloid newspapers across the year 2014. Although we are fully aware of potentially different complexities of other types of political texts (e.g., party manifestos, speeches, debates, or social media posts), we believe that our analysis of newspapers is a fruitful starting point for a methodological contribution to political communication research and potentially beyond. The ubiquitous measurement of textual readability is not necessarily enough to measure the multidimensional nature of the text complexity construct.

The empirical findings support what has been established already in linguistics. The two factors extracted from the data correspond to the semantic complexity and syntactic complexity dimensions. It is important to note that complexity is a more complex phenomenon than previously handled in the literature, and outcomes of political communication studies relating to text complexity may very well hinge on the measures that were selected (i.e., studies operationalizing complexity as readability). The factor analysis model clearly shows that each of these variables individually measures only one aspect of linguistic complexity. Text complexity is a multifaceted construct that includes at least two dimensions directly related to the structure of language—syntactic and semantic—and it should be treated as such.

To showcase the added utility of incorporating a multidimensional view of text complexity, we looked at comparisons of two complexity score measures—one for syntactic and the other for semantic— across different types of newspapers. Quality newspapers, in general, were statistically significantly different from tabloid newspapers on both complexity dimensions, with quality newspapers being syntactically more complex than tabloid media and tabloid newspapers being semantically more complex than quality newspapers. Interestingly, this pattern was the same for individual newspapers: All quality newspapers were more syntactically complex, but less semantically diverse than their tabloid counterparts. These findings show that differentiating between these two dimensions not only makes sense from a theoretical point of view, but also practically.

Researchers operationalizing complexity by using a single measure necessarily lose information about the second dimension that this particular measure does not cover. In addition, because these dimensions are negatively correlated, one might inadvertently come to different conclusions depending on the measure one chooses to operationalize complexity. This leads to one of the key points of the current study: The existence of multiple indices measuring text complexity is not detrimental to the measurement of this concept per se; nevertheless, we believe that using a single one of these measures without consideration as to which aspect of complexity it is related to inevitably leads to misinterpretations of results. If one is interested in the concept of complexity and in its multidimensional nature, one should incorporate scales, or at least a pair of indices, that measure both dimensions. Even if multidimensionality is not of interest to the researchers, reflection about what aspect of the construct they are measuring would provide more clarity to the design. Overall, our results show that semantically complex texts do not necessarily need to be also syntactically complex; thus, one needs to think carefully when choosing text complexity metrics and interpreting results.

Having empirically established that tabloid newspapers, for example, are syntactically relatively simple but semantically relatively complex, the implications of these findings quickly become very salient for political communication. Quality and tabloid newspapers differ in language even when covering the same topic (Bagnall, 1993), and this variation is not simply a byproduct of different editorial routines or standards, but an inherent distinction between the two media (Fowler, 2013). Tabloids and quality press have very distinct audiences and a different agenda (Trilling & Schoenbach, 2013). Kleinnijenhuis (1991) confirmed that simple news is beneficial for the lower educated strata, whereas complex news is beneficial for the higher educated; combining these facts with different dimensions of text complexity would potentially open doors for wide range of questions in political communication. We can now go one step further and ask whether the lower educated readers of political newspapers benefit from highly complex information (semantics) presented in a very simple way (syntax)? Is there any form of an interaction effect of different complexity dimensions on the effects of complexity on political processes? Is one of these dimensions more beneficial than the other for political learning? These questions could only be answered if we start to differentiate among the various aspects of "text complexity" and stop treating it like it is one-dimensional.

The study included two validation procedures: a sample of different text types and a crowdsourced experiment on perceived text complexity. The two text types different from political texts—legal texts and children's books—varied significantly from both the quality and tabloid newspapers in a theoretically expected way (although still statistically significant, legal texts and quality newspapers were quite similar on the syntactic dimension). According to the results, children's books were less complex on both dimensions, whereas legal texts were as complex as quality newspapers regarding their syntactic structure but did not have the lexical variation of the newspapers. A crowdsourcing experiment showed that the texts classified as complex by the algorithm were also perceived as being more complex by humans. It is important to note that the stimuli combined the complexity dimensions, while it is reasonable to assume that different dimensions (and combinations of these) would lead to different perceptions.

Based on these findings, we believe two aspects are important for future research in political communication dealing with the complexity of political texts. First, content analytical procedures need to do better in terms of motivating their choice of complexity measurements and acknowledging the limitations

of a restricted type of measurement in case they employ only a single dimension of text complexity. Ideally, such studies should relate to different aspects of complexity. Second, political communication research needs to acknowledge the possibility of differential consequences of different dimensions of text complexity for, for instance, political engagement and learning. It is likely that variation in semantic versus syntactic complexity would produce different knowledge gains, but our understanding of such differences is thus far rather limited. In addition, this method may also be used to uncover historic trends in textual data or effects of textual characteristics on audiences over time. For example, Spirling (2016) found that with time the speeches made by British cabinet ministers decreased in linguistic complexity, and Bischof and Senninger (2018) investigated how populist language is perceived by the voters; yet, both of the aforementioned studies used the Flesch measure as the metric for determining overall text complexity. Extending this question further, one could analyze whether such historical and political trends become simpler semantically, syntactically, in their content, or on every dimension simultaneously, and how these changes affect the voter.

The present investigation was concerned only with objective structural complexity of language and thus was confined to the manifest characteristics describing language generation (newspapers as the producers of language). Nevertheless, this is only one of the two parts of how text complexity affects various political processes, the second being language comprehension. Thus, an analysis of agent-related complexity would be the logical next step for future studies. Understanding this interplay between language generation and information acquisition would allow future research to better understand both the construct of text complexity and its role in the media environment. Furthermore, the current study assumed an extensional (as opposed to intensional) approach to semantic complexity; that is, the more ambiguity is associated with a lexical unit, the more semantically complex it is deemed to be. Although such an approach seems very reasonable, especially in the context of quantitative social science, future research may be interested in investigating the role that intensional semantic complexity plays in both its relationship to syntactic complexity and to how people perceive and acquire information from texts with high intensional semantic complexity.

The study is not without its limitations. As mentioned above, a logical next step for future research is the inclusion of even more text types into the analysis. Even though the current study applied this method to a heterogeneous sample of texts, extending the scope of our analysis to assessment of and comparison among other types of communication may help construct a broader spectrum of textual complexity and its variations among political texts. Another limitation of the current study is that the stimulus material for human validation was either complex or simple on all complexity dimensions. It is necessary to discover how these complexity dimensions affect various learning processes or information retention; yet, an intricate experimental design such as this would require a study of its own.

Another necessary limitation of this study refers to the concept of semantic complexity. Semantic complexity is a complicated concept to define, and unlike syntactic complexity, difficult to operationalize in a quantitative setting. All of the measures used in the current study were in essence lexical proxies for semantic complexity; these measures do not directly gauge the semantic content of any lexical unit (an automated algorithm would not know how "ambiguous" a word or a sentence is in the strictest term). The measurement of text complexity may be improved by incorporating more sophisticated text processing

techniques such as, for example, thesauri and semantic treebanks to account for phenomena such as synonyms and polysemy, which undoubtedly influence the general complexity of the text.

To sum up, even with the aforementioned limitations, this study provides important insight into how text complexity is defined, how it should be measured in the social sciences, and how such measurements differ from each other. The method used here should in no way be thought of as normative, and these results are neither an exhaustive description of the full range of measures nor the text complexity phenomenon as a whole. However, our approach does provide a firm guideline to properly measuring political text complexity on a large scale for future researchers. The main goal of the current study was to show that using a single measure to operationalize text complexity is insufficient and that researchers need to either acknowledge that they are measuring a single facet of complexity only (if relying on one index) or create more nuanced instruments to gauge political text complexity. We believe that this study makes an important step toward a more thorough investigation of political text complexity and its effects in political communication. Certainly, the theoretical notions and the methodological approach advanced here should also be of relevance to communication studies more broadly, providing a proper tool for further engagement with the actual structure of the language that is at the heart of communication.

References

Bagnall, N. (1993). *Newspaper language*. London, UK: Routledge.

Baker, P. (2010). Representations of Islam in British broadsheet and tabloid newspapers 1999–2005. *Journal of Language and Politics*, *9*, 310–338.

Baumann, J. F. (2005). Vocabulary–comprehension relationships. In B. Maloch (Ed.), *YEARBOOK—National Reading Conference* (Vol. 54, pp. 117–119). Oak Creek, WI: National Reading Conference.

Bischof, D., & Senninger, R. (2018). Simple politics for the people? Complexity in campaign messages and political knowledge. *European Journal of Political Research, 57*(2), 473–495.

Dahl, Ö. (2004). *The growth and maintenance of linguistic complexity* (Vol. 71). Amsterdam, Netherlands: John Benjamins.

Dale, R., Moisl, H., & Somers, H. (2000). *Handbook of natural language processing*. New York, NY: CRC Press.

Daneman, M., & Case, R. (1981). Syntactic form, semantic complexity, and short-term memory: Influences on children's acquisition of new linguistic structures. *Developmental Psychology*, *17*(4), 367–378.

Eveland, W. P., & Cortese, J. (2004). How web site organization influences free recall, factual knowledge, and knowledge structure density. *Human Communication Research*, *30*(2), 208–233.

Eveland, W. P., & Dunwoody, S. (2001). Applying research on the uses and cognitive effects of hypermedia to the study of the World Wide Web. *Annals of the International Communication Association*, *25*(1), 79–113.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*(3), 221–233.

Fowler, R. (2013). *Language in the news: Discourse and ideology in the press*. London, UK: Routledge.

Frajzyngier, Z., Hodges, A., & Rood, D. S. (Eds.). (2005). *Linguistic diversity and language theories* (Vol. 72). Amsterdam, Netherlands: John Benjamins.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: Sequential effects of syntactic and semantic information. *Memory & Cognition*, *27*(3), 438–453.

Honnibal, M., & Johnson, M. (2015). An improved non-monotonic transition system for dependency parsing. In L. Màrquez & C. Callison-Burch (Eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1373–1378). Lisbon, Portugal: Association for Computational Linguistics.

Jucker, A. H. (1992). *Social stylistics: Syntactic variation in British newspapers*. New York, NY: de Gruyter.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing*. London, UK: Pearson.

Kleinnijenhuis, J. (1991). Newspaper complexity and the knowledge gap. *European Journal of Communication*, *6*(4), 499–522.

Kortmann, B., & Szmrecsanyi, B. (2012). *Linguistic complexity: Second language acquisition, indigenization, contact*. New York, NY: de Gruyter.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

McCarthy, P. M., & Jarvis, S. (2010). MTLD, VOCD-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392.

McLuhan, M. (1964). *Understanding media: The extensions of man*. New York, NY: McGraw-Hill.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492–518.

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117–134.

Percival, W. K. (1990). Reflections on the history of dependency notions in linguistics. *Historiographia Linguistica, 17*(1), 29–47.

Prior, M. (2013). Visual political knowledge: A different road to competence? *The Journal of Politics, 76*(1), 41–57.

Rogalsky, C., & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, *19*(4), 786–796.

Shannon, C. E., Weaver, W., & Burks, A. W. (1963). *The mathematical theory of communication*. Chicago, IL: University of Illinois Press.

Solé, R. V., Corominas-Murtra, B., Valverde, S., & Steels, L. (2010). Language networks: Their structure, function, and evolution. *Complexity, 15*(6), 20–26.

Spirling, A. (2016). Democratization and linguistic complexity: The effect of franchise extension on parliamentary discourse, 1832–1915. *The Journal of Politics, 78*(1), 120–136.

Strömbäck, J. (2005). In search of a standard: Four models of democracy and their normative implications for journalism. *Journalism Studies, 6*(3), 331–345.

Suedfeld, P., & Tetlock, P. (1977). Integrative complexity of communications in international crises. *Journal of Conflict Resolution, 21*(1), 169–184.

Tan, Y., Martin, R. C., & Van Dyke, J. A. (2017). Semantic and syntactic interference in sentence comprehension: A comparison of working memory models. *Frontiers in Psychology*, *8*, 198.

Tanaka-Ishii, K., & Aihara, S. (2015). Computational constancy measures of texts—Yule's *K* and Rényi's entropy. *Computational Linguistics*, *41*(3), 481–502.

Tichenor, P. J., Donohue, G. A., & Olien, C. N. (1970). Mass media flow and differential growth in knowledge. *Public Opinion Quarterly, 34*(2), 159–170.

Timuçin, M. (2010). Different language styles in newspapers: An investigative framework. *Journal of Language and Linguistic Studies, 6*(2), 104–126.

Tolochko, P., & Boomgaarden, H. G. (2018). Analysis of linguistic complexity in professional and citizen media. *Journalism Studies*, *19*(12), 1786–1803.

Trilling, D., & Schoenbach, K. (2013). Patterns of news consumption in Austria: How fragmented are they? *International Journal of Communication, 7*(25), 929–953.

Van Der Velden, M., Schumacher, G., & Vis, B. (2018). Living in the past or living in the future? Analyzing parties' platform change in between elections, the Netherlands 1997–2014. *Political Communication, 35*(3), 393–412.

Von Fintel, K., & Heim, I. (2011). *Lecture notes on intensional semantics*. Archives of Massachusetts Institute of Technology, Massachusetts Institute of Technology. Retrieved from http://web.mit.edu/fintel/fintel-heim-intensional.pdf

Yule, C. U. (2014). *The statistical study of literary vocabulary.* Cambridge, UK: Cambridge University Press.

Zelizer, B., Bird, S. E., Brookes, R., Calabrese, A., Golding, P., Gripsrud, J., . . . & Macdonald, M. (2000). *Tabloid tales: Global debates over media standards*. Lanham, MD: Rowman & Littlefield.

Appendix

Table A1. Zero-Order Correlation Matrix.

Variable	1	2	3	4	5	6	7	8
1. Lexical diversity	1							
2. Yule's *I*	.71	1						
3. Semantic entropy	.98	.69	1					
4. Measure of Textual Lexical Diversity	.78	.71	.80	1				
5. Sentence length	−.55	−.51	−.56	−.53	1			
6. Automated readability index	−.14	−.09	−.15	−.07	.64	1		
7. Syntactic dependency	−.53	−.52	−.54	−.53	.94	.56	1	
8. Syntactic depth	−.52	−.52	−.52	−.52	.94	.57	.99	1

Table A2. Descriptive Statistics.

Variable	Combined sample	English sample	German sample
Lexical diversity	0.58 (0.15)	0.52 (0.11)	0.71 (0.13)
Yule's I	36.00 (41.09)	20.34 (15.13)	74.79 (58.50)
Semantic entropy	83.9 (6.09)	81.58 (5.01)	89.49 (4.67)
Measure of Textual Lexical Diversity	0.78 (0.06)	0.76 (0.04)	0.84 (0.05)
Sentence length	18.90 (7.15)	22.12 (5.54)	10.84 (3.39)
Automated readability index	12.52 (3.21)	12.51 (3.36)	12.33 (2.78)
Syntactic dependency	5.35 (1.46)	6.02 (1.11)	3.70 (0.75)
Syntactic depth	5.39 (1.44)	6.05 (1.09)	3.76 (0.78)
n	71,312	50,799	20,513

Note. Means and standard deviations (in parentheses) are reported.

Table A3. Formulas.

Lexical diversity	$\dfrac{N_{types}}{N_{tokens}}$	N_{types} = number of unique words N_{tokens} = total number of words
Yule's I	$\dfrac{M_1 \times M_1}{M_2 - M_1}$	M_1 = all words M_2 = sum of the product of observed frequency of each type to the second power
Semantic entropy	$-100\dfrac{\sum p_i \times \log p_i}{\log N}$	p_i = probability of encountering a certain word type N = total number of words
Automated readability index	$4.71\left(\dfrac{N_{char}}{N_{tokens}}\right) + 0.5\left(\dfrac{N_{tokens}}{N_{sentences}}\right) - 21.43$	N_{char} = number of characters N_{tokens} = total number of words $N_{sentences}$ = number of sentences