Evolution of Linguistic Properties of the UNGD Speeches from 1946 to 2022: An Analysis of Linguistic Similarities to Fake News

Julius Pfundstein | University of Leipzig | Julius.pfundstein@gmx.de

Abstract: An investigation of the evolution of linguistic properties across all speeches of the UNGD between 1946 and 2022 aims to measure the extent to which language has changed in its resemblance to fake news. To avoid the heuristic arbitrariness of labeling fake and factual, language properties such as average sentence length, linguistic complexity, readability, sentiment and fake news likelihood of a BERT-based fake-news classifier are measured in a time series analysis. The results reveal a complicated and contradictory linguistic topography. Words of crisis are increasing in proportion, but the share of emotionalizing modal adverbs, degree adverbs, negations and swear words are decreasing. The average sentence length decreases, readability increases, but lexical complexity increases, and language becomes more objective. The classifier shows a valley around the 1980s with an increased likelihood of fake news in the 1950s and since the early 2000s.

Keywords: United Nations General Debate, Fake News, Linguistic Evolution, Zero-Shot-Classification, Time-Series Analysis

1 where we are coming from

Today, the term *fake news* is an indispensable part in the academic debate on contemporary trends in political communication. Although recent research findings about the extent of distribution and consumption of fake news do not provide any reason for alarmism[1/2/3], the rapidly advancing capacities of large language models and generative adversarial networks attest to the growing relevance of the debate about authenticity in the digital public sphere.

The consensus on the significance of the scientific evaluation of fake news is contrasted by the uncertainty about the meaning and function of the term. From a critique of digital capitalism to a political style, definitions and applications of the floating signifier vary depending on the debate and discipline.[4] In a systematic literature review Engelberger and Lecheler classify definitions by distinguishing the following ingredients: Fake news are (1) low in facticity (2) presenting itself as genuine (3) intentionally deceiving. While there is consensus on (1), (2) and (3) are optional for some authors.[5] In the following we also claim this optionality for the third element, because the methods used are linguistic in nature and do not allow any statement to be made about the intention of the representatives of the UN. What is meant by the use of the term *fake news* in the following, however, is the linguistic similarity to fake news or, in other words, the hypothesis that fake news can be considered to be a stylistic device

[1] Moore, Ryan C.; Dahlke, Ross; Hancock, Jeffrey T. (2023): Exposure to untrustworthy websites in the 2020 US election.

[2] Altay, Sacha; Berriche, Manon; Acerbi, Alberto (2023): Fake news on Fake news: Conceptual and Methodological Challenges.

[3] Fletcher, Richard; Cornia, Alessio; Graves, Lucas; Nielsen, Rasmus Kleis (2018): Measuring the reach of 'fake news' and online disinformation in Europe.

[4] Farkas, Johan; Schou, Jannick (2018): Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood.

[5] Egelhofer, Jana Laura; Lecheler, Sophie (2019): Fake news as a two-dimensional phenomenon: a framework and research agenda.

that is identifiable by linguistic properties.

For much of the existing research, it is considered a methodological necessity to label the investigated corpora as either fake or factual. Since classification is expensive and requires technical expertise, it is often left to private sector providers who issue and update lists. These are usually determined via crowdsourced majority voting or journalistic quality criteria. In many methods, the resulting labels are attributed to a articles or sources of articles as a semantic unit, thus negating the complication of the space between the binary labels. Research based on these lists is limited in its robustness, as the addition or deletion of a single item can create or avoid significances.[6] In addition, not all fake news is falsifiable. Often, only a finer granularity allows details to be refuted, while generalizations, especially those that lack linguistic precision, are not identifiable as untrue even when compared with knowledge databases such as wolfram alpha. This is because even conspiracy theories are often accompanied by a body of recognized science, which can, however, differ considerably in its assessment of the impact, extent and underlying intentions.

However, to avoid the immeasurable bias of this methodological arbitrariness, there is a different approach to measure fake news. Instead of manually checking statements for accuracy, linguistic properties which are indicators of fake news are defined a priori. These indicators are taken from existing research literature, which has often also used manually labeled data for their identification. The measurement of these language properties is transparent and robust, but this quality is paid for by the impossibility of drawing direct conclusions about the actual presence or absence of fake news. Even the use of a classifier trained with labeled data can only be evaluated as an aggregation of all identified language properties, which ultimately only evaluate the extent of linguistic similarities with already previously classified fake news.

The goal is to analyze language in terms of its capacity to convey fake news. To this end, basic language properties that have been assessed as relevant for fake news in an existing body of research will be measured and compared in a time series.

The research is conducted on a corpus of the United Nations General Assembly. The corpus is intended to counteract the general research trend of deriving contemporary developments mostly from research corpuses originating in the US or continental Europe and to provide a standardized and officially translated format that is easy to compare. For this project, the following two questions are posed to the corpus:

RQ1: How did linguistic properties of the UNGD speeches change from 1946 to 2022?

RQ2: To what extent can the linguistic trends of the UNGD be seen as indicators of an increasing resemblance to fake news?

---

2 the UNGD and how to read it

There is a severe graphical bias in the quantitative research of political communication. The existing body of research causes a path dependency of technical affordances, which causes researchers to investigate the same platforms, states and people with the help of existing APIs and a set of methods fit for a geographically and politically exclusive spaces. To escape this focus on the American Presidency Project and the UKs parliament transcripts, an international body of research is to be put together. This poses several difficulties, as most states

[6] Fletcher, Richard; Robertson, Craig T.; Nielsen, Rasmus Kleis (2021): How Many People Live in Politically Partisan Online News Echo Chambers in Different Countries? This paper is not directly about fake news, but it shows the same problem in a particularly powerful way: Fox News just exceeds the arbitrarily defined value for bipartisan media in its political slant and thus provides that the now overestimated conservative echo chamber can be said to consume media outside the filter bubble particularly frequently.

only publish government statements in their domestic language. For many languages, there are still no reliable machine translations and, furthermore, the frequency and formats differ significantly. Direct protocols have fundamentally different linguistic properties than reports developed retrospectively and so a change in language could always be confused with a change in format. It is therefore evident that the research body must have originated from one and the same forum and that its framework has changed as little as possible over as long a period as possible. With these constraints, only one option remains for the desired text corpus - the United Nations General Debate (UNGD).

The General Debate is part of the UN General Assembly, which is held annually in September, and plays a prominent role in international relations. As of 2024 193 recognized states represent 99.46 % of the world's population and 88.50 % of its land mass - a figure that may be surpassed by organizations from sporting or commercial bodies, but the UN has monopolized the arena of international relations.[7]

Since 1946, every recognized state has been given the opportunity to deliver a speech at the annual general assembly in September. There are no thematic constraints, no vetoes by the permanent members of the Security Council and no opportunities for interruption. For this autonomy, the General Assembly pays with its inability to adopt binding decisions, let alone impose the sanctions regime provided for in Article 41. Resolutions, however, should not be underestimated. They build up public pressure and have a socializing effect on the participants.[8]

Since its foundation, the UN has grown into many fields of activity. Today 15 specialized agencies coordinate international cooperation, define standards and develop guidelines for critical infrastructure and technology. Participation in these organizations and the resulting intellectual property is a prerequisite for integration into a global economy. The substantive spillover has also elevated the status of the UN to a prominent role in diplomacy. The list of speakers is only partially annotated with the speaker's function, but random samples show that the share of speeches delivered by foreign ministers has steadily decreased in favor of heads of state. The UN also enjoys a high level of trust among the world's population. In the 2024 trust barometer, despite a sharp decline this year, the UN still score 58 points, two points above the average for the national governments of the 28 countries surveyed.[9]

The study is conducted on the transcripts of all speeches from 1946 to 2022. These can be found in Harvard's Dataverse and downloaded as TGZ-file.[10] Unzipped, the 206 MB dataset is structured into 77 folders for each session. Each session contains a TXT-file for each state. The TXT-files contain machine readable text by OCR processing of the original scans. The documents have been stripped of any boilerplate (footer, header, frontpage), sequential numbering of pages or paragraphs and links to resolutions using Python's Regular Expressions library. After processing we are left with an unformatted continuous text.

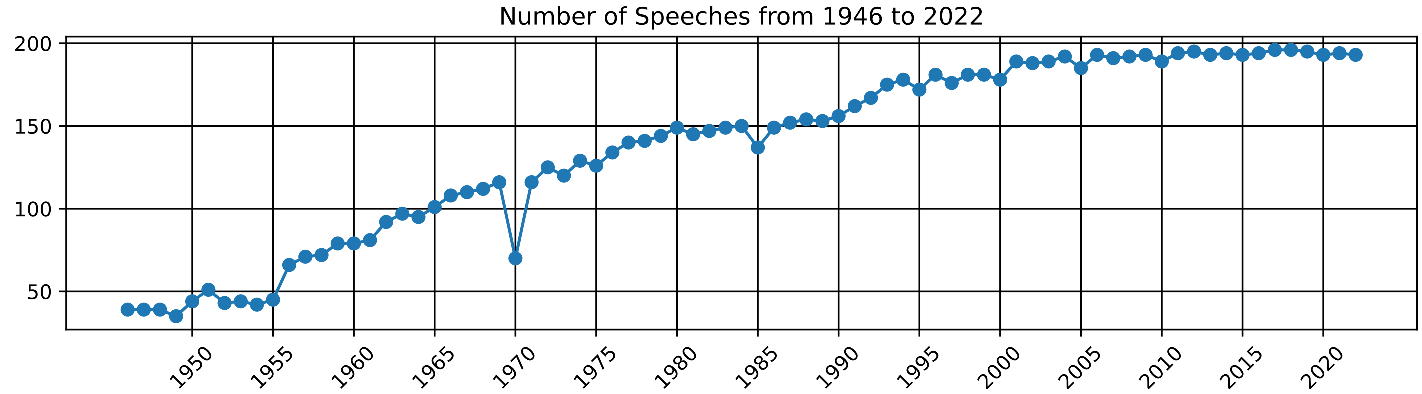
It is important to note that the number of states (and other entities) recognized by the UN and thus admitted to speak at the UNGD is not constant.

[7] WorldData (2024): Member States of the United Nations.

[8] van den Rul, Celine (2020): Why Have Resolutions of the UN General Assembly If They Are Not Legally Binding?

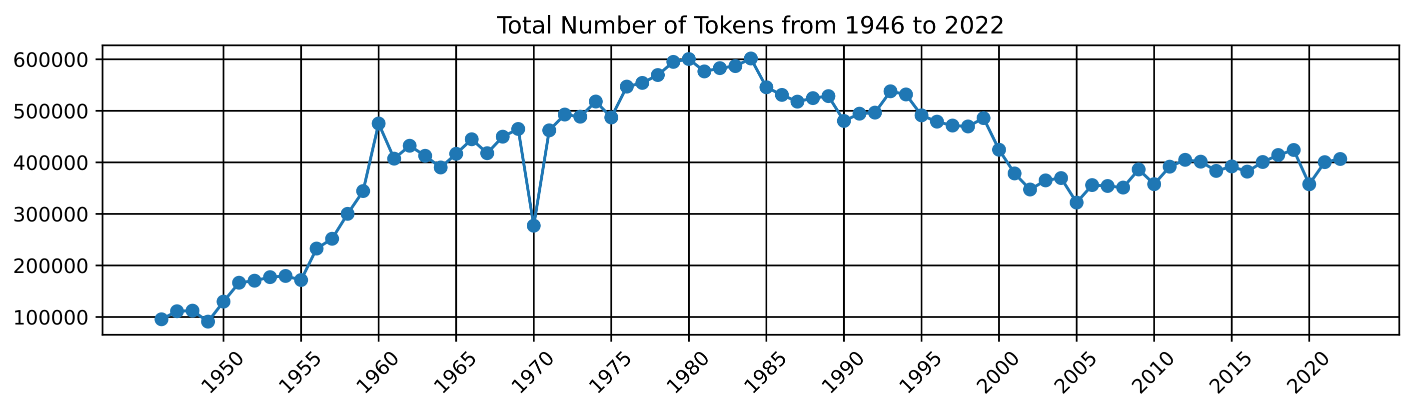
[9] Edelman (2024): 2024 Edelman Trust Barometer. Global Report.

[10] Jankin, Slava; Baturo, Alexander; Dasandi, Niheer (2017): United Nations General Debate Corpus 1946-2022.



Of the 51 founding members, only 39 spoke at the first session of the UNGD in 1946. In the course of decolonization and the disintegration of political entities, the number of UN member states rises to 193 in 2024. The fact that in some years there are more speeches than there are states is due to the participation of Palestine, the Holy See, the International Red Cross and other international organizations. The drop in 1970 can be explained by the unique format chosen for the 25th anniversary of the UNGD, which caused some states to hold their address earlier in the General Assembly and thus was not labeled as UNGD. Also, the number of states not exercising their right to deliver a speech has steadily decreased due to the increasing professionalization of international relations and the growing recognition of the UN as an agency.

As there is neither a minimum nor a maximum for the length of speeches, this size also varies considerably. Everything from brief acknowledgements to lengthy elaborations is permitted and practiced.



In the 1950s, the number of words increases along with the number of speeches. However, the peak of 600,000 tokens is reached in 1980, when only 149 speeches were delivered. The literature offers no plausible explanation for the decline in average speech length. However, when the relatively stable median length is considered, it can be argued that it is primarily the gradual decline in very long speeches, together with the increasing number of speeches overall, that forms the unimodal distribution over the years. Long speeches appear to be a particular property of the protagonists of the cold war. Of the hundred longest speeches, 44 were held in the 1960s. Since the beginning of the 1990s, only Ghaddafi's infamous speech in 2009 has made it into the top 100. In the same list, the Soviet Union appears 21 times, India 10 times and Fidel Castros Cuba 6 times.

This data already suggests an important limitation for research on the text corpus. This is because changes in the linguistic properties do not necessarily have to be attributed to the zeitgeist of the language or political communication as such. A change in the function of the UN would also be plausible. For example, geopolitical and ideological conflicts of the cold war were often carried out in the framework of the general debate, especially in the early history of the world organization. Today, the UNGD seems to invite fewer lengthy debates and thus also has a different set of language properties.

The summary of the dataset also shows that some metrics cannot be applied to the dataset without further ado, as the varying text lengths would cause distortions in the time series comparison. These must be addressed with normalizations and dynamic windows.

3 what features to look for

Now that we have a formatted corpus, we must determine which linguistic properties should be investigated to measure the languages resemblance to fake news. A vibrant body of research has been established around the question of whether fake news, can be identified using quantifiable features that represent linguistic traits, writing styles, format and sentiment. To identify linguistic properties that are significant identifiers for fake news, we will draw on three research papers.

1. In 2017 Rashkin et al. examined language properties from satire, propaganda, hoaxes and trustworthy news. They tested for correlations between the occurrence of a total of 18 language properties and the four types of text. 11 of the 18 properties were dictionaries with grammatical signal words, the remaining 7 tested for thematically sorted word groups of the LIWC dictionary. The result was a list of those properties sorted by the factor by which they occur more frequently in untrustworthy news than in trustworthy ones. Swear words (LIWC - Swear), direct address (second person singular), modal adverbs, degree adverbs and the first-person singular form the top five properties with the greatest overproportionality in fake news and are thus deemed suitable properties for the investigation of the UNGD corpus. The results also included properties that are markers for trustworthy news. This list was topped by numbers which will also be included in the list of investigated properties. [11]

2. Already before the widespread adoption of the term fake news and research of political communication that revolves entirely around the internet, computational linguistics were investigating markers for lies and deception. Based on the assumption that "liars can be reliably identified by their words - not by what they say but by how they say it", Newman et al. found that liars tend to tell stories that are less complex, less self-relevant, and more characterized by negativity.[12]

3. In 2023 Petrou et al. examine the evolution of linguistic features of fake news in a time series comparison. They find that over the period from 2009 to 2019, out of 534 linguistic features, it was mainly the thematic (here: psychological) dictionaries that were used "to demonstrate the intention of the writers to convince readers that the content is realistic by using certain mechanisms, for example emotionally influencing readers by overdramatizing certain events through persuasive language". As a proxy for this dramatization, the following features have been found to be most significant to differentiate fake from factual: A metric to quantify vocabulary richness, the readability and the negation rate. Other findings concern the capitalization of entire words, headlines and layout. These phenomena are not relevant for our corpus.[13]

4. Finally, there are ongoing attempts to identify fake news with the help of artificial intelligence. Berrondo-Otermin and Cabezuelo provide a review of these efforts.[14] Particularly fitting to our case is the approach of Tida et al. who use a fine-tuned BERT model to represent semantic and linguistic properties in a high dimensional

[11] Rashkin, Hannah; Choi, Eunsol; Jang, Jin Yea; Volkova, Svitlana; Choi, Yejin (2017): Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking, S. 2931–2937.

[12] Anwar, Imran Nazeer Shahneela Yousaf (2023): Analyzing Linguistic Shifts in Political Discourse: A Corpus-Based Study of Political Rhetoric in the Digital Age.

[13] Petrou, Nikolas; Christodoulou, Chrysovalantis; Anastasiou, Andreas; Pallis, George; Dikaiakos, Marios D. (2023): A Multiple change-point detection framework on linguistic characteristics of real versus fake news articles.

[14] Berrondo-Otermin, Maialen; Sarasa-Cabezuelo, Antonio (2023): Application of Artificial Intelligence Techniques to Detect Fake News: A Review.

vector space.[15] What is unique to this approach is that transformers such as BERT do not process the text sequentially, thus allowing for detection of dependencies across the entire text. Since the model of Tida et al. is not publicly available, a similar model by Romero is used in the following.[16]

This gives us a list of 13 linguistic properties and one zero-shot classifier:

ID	Property	Implementation	Example or Explanation
Thematic or Psychological Dictionaries			
T1	Swear Words	Wiktionary Category: English swear words	*damn*, *bloody*, *crap*
T2	Crisis Words	GPT 3.5 -constructed	*war, catastrophe, emergency*
T3	Numbers	Pythons Regular Expressions Library	123, thousand, twelve, 1.5
Grammatical Dictionaries			
G1	Self-References	First Person Singular	*I, me, my, myself*
G2	Direct Addresses	Second Person Singular	*you, your, yourself*
G3	Modal Adverbs	Wiktionary Dictonary	*clearly, perhaps*
G4	Degree Adverbs	Wiktionary Dictionary	*very, almost, barely, totally*
G5	Negations	Negation rate	*no, none, not, false, didn't*
Stylometric Properties			
S1	Sentence Length	Count token between sentence end markers.	Average sentence length
S2	Lexical Diversity	Moving Average Type-Token Ratio	Word set divided by word list
S3	Readability	Flesch and Kincaid's Readability Ease	See Chapter 4 – S3
S4	Sentiment - Polarity	Average polarity per token	*good* and *bad* cancel each other out
S5	Sentiment - Subjectivity	Average absolute polarity per token	*good* and *bad* add up
Zero Shot Fake News Classifier			
C1	Fake News Likelihood	BERT-based and Facebook trained ZS-Classifier	Labels based on feature-interrelationship

4 how to get the features

To answer the research questions the each property will be measured on each speech individually. Then the average of the values is calculated to prevent long speeches and thus representatives that speak long from contributing more than their colleagues to the value of one year.

The individual properties are evaluated in the following methods:

T1 - Swear Words: A publicly available dictionary of English-language swear words is used for this purpose. The dictionary can be found as a category on Wikimedia's Wiktionary.[17] Three terms that contain *jesus* as well as the term *spastic* were removed from the list as it requires context understanding to distinguish their function as swearword from other meanings. Then the proportion of words that occur in both the corpus and the swear dictionary is determined and divided by the total number of words in the corpus.

T2 – Crisis Words: No publicly available dictionary of words related to crisis has been found. Therefore, ChatGPT 3.5 was tasked with the following: "create a dictionary with 100 entries for words with semantic similarity to the word crisis". Since the addition of a random temperature prevents deterministic results in this large language model, the list can be found in the GitHub repository.[18] The calculation is the same as for **T1**.

T3 – Numbers: Numbers are identified using the Python

[15] Tida, Vijay Srinivas; Hsu, Sonya; Hei, Xiali (2022): A Unified Training Process for Fake News Detection based on Fine-Tuned BERT Model.

[16] Romero, Manuel (2021): bert-tiny-finetuned-fake-news-detection: Hugging Face.

[17] Wiktionary - The free dictionary (2024): Category: English Swear Words.

[18] Pfundstein, Julius (2024): UNGD-linguistic-patterns. Investigation of the lingusitic evolution of speech in the United Nations General Debate for 1946 to 2022.

package Regular expressions.[19] For this purpose, written numbers are recognized as well as numerical numbers. The calculation is the same as for **T1**.

G1 – Self-Reference: Any use of the first-person singular is understood as self-reference. The calculation is the same as for **T1**.

G2 – Direct Address: Any use of the second-person singular is understood as a direct address. The calculation is the same as for **T1**.

G3 – Modal Adverbs: Modal adverbs express the speaker's degree of certainty about what is being said. These adverbs often convey notions like possibility, necessity, obligation, or likelihood, functioning similarly to modal verbs (like "can," "must," "might," etc.) but in an adverbial form. A dictionary of English modal adverbs was found on Wiktionary. [20] The calculation is the same as for **T1**.

G4 – Degree Adverbs: Degree adverbs modify adjectives, other adverbs, or verbs to indicate the intensity, extent, or degree of the quality or action described. A dictionary of English degree adverbs was found on Wiktionary. [21] The calculation is the same as for **T1**.

G5 – Negation: Includes signal words of rejection and negation and their abbreviations. The calculation is the same as for **T1.**

S1 – Sentence Length: Calculated by the average number of words per sentence. Line breaks are explicitly not considered sentence endings, as the OCR scans also encode them for page breaks. Therefore, only dots and question marks are used to identify the end of the sentence.

S2 – Lexical Diversity: Lexical diversity calculates the ratio between types (unique words) and tokens (all words with duplicates). Since the text bodies are different in length and a text with increasing length also has an increasing probability of word repetition, it is necessary to have a fixed window size, to calculate the type-token ratio within and then calculate the averages over the entire corpus. This approach is called moving-average type-token ratio (MATTR).[22]

S3 – Readability: Readability is calculated by considering sentence length and word length with the readability test of Flesch and Kincaid. The result is a number between 0 and 100, where 100 is very easy to read and 0 is very difficult to read. The parameters of the formula were determined in an experimental design:

$$206.835 - (1.015 \times \frac{TotalWords}{TotalSentences}) - (84.6 \times \frac{TotalSyllables}{TotalWords})$$

S4 – Sentiment Polarity: Python's TextBlob is used for the sentiment polarity. For this purpose, a dictionary-based method is triangulated with a rule-based method (e.g. inversion of polarity in case of negation). The result is a value between -1 (very negative) and 1 (very positive).[23]

S5 – Sentiment Subjectivity: The sentiment subjectivity is also calculated with Python's TextBlob. It calculates the absolute value for all polarities and returns a value between 0 (very objective) and 1 (very subjective).

C1 – Fake News Likelihood: The classifier was trained with prelabelled news articles and employed a miniature version of Google's BERT model to map them into a high-dimensional feature space. When the classifier is applied to the UNGD speeches it seeks to infer classes by recognizing already known patterns. To avoid binary labels and get more comparable decimal numbers a SoftMax function is used to transform the output in a likelihood of getting a certain label. [24]

[19] Python (2024): re - Regular expression operations. Version 3.12.5.

[20] Wiktionary - The free dictionary (2024): Category: English modal adverbs.

[21] Wiktionary - The free dictionary (2024): Category: English degree adverbs.

[22] Covington, Michael A.; McFall, Joe D. (2010): Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR).

[23] Loria, Steven (2024): TextBlob Documentation. Release 0.18.0.post0.

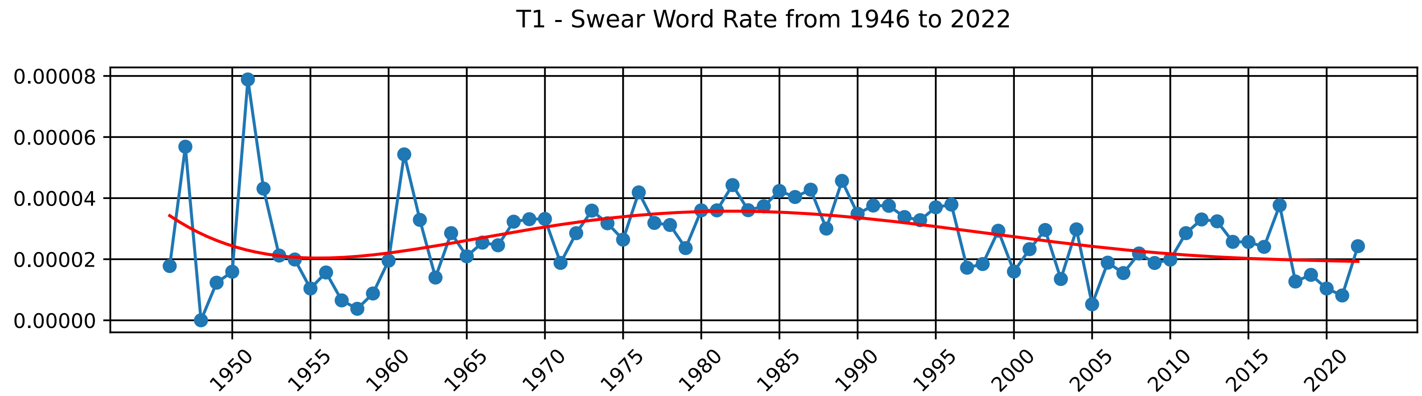
[24] Romero, Manuel (2021): bert-tiny-finetuned-fake-news-detection: Hugging Face.

5 what the features tell

In this section the time series comparisons for the 13 language properties and the classifier are examined in some detail. Particular attention is paid to patterns, trends, outliers and, where possible, a point of reference outside the UNGD dataset.

The displayed section of the y-axis is automatically measured as the space between minimum and maximum, so it is necessary to look at the axis ticks to estimate values.

The years 1947 and 1948 are special in that the protocols were not transcriptions of the original speeches but created retrospectively with the speakers in the third person. This not only introduces new grammatical properties, but also new vocabulary, as statements have been shortened to their formal consequences while "linguistic decorations" were neglected. To avoid distortion those years are omitted for all grammatical and stylometric analysis as well as the classifier.

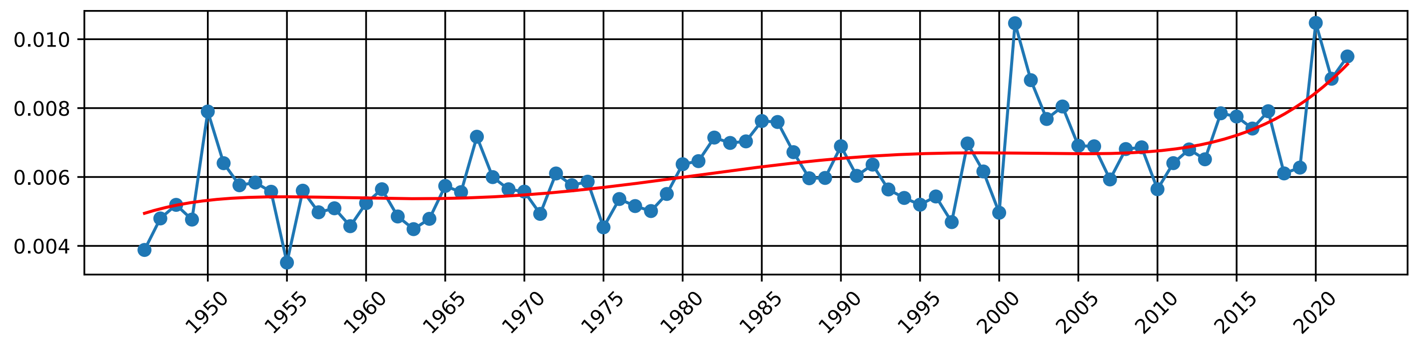


The use of swear words in the UNGD corpus fluctuates without a clear trend between 0.00001 and 0.00005. The polynomial regression (degree = 5) in red shows that the rate of swear words peaked in the 1990s and dropped again at the turn of the millennium. While the outliers on the left are due to the new constellations in an evolving set of member states, the year 1961 can be seen as the first significant outlier.

To put these results into perspective: A comparison with research literature on the social function of swear words suggests that even the highest value from 1951 falls short of the average curse rate of professional interaction by more than 50%.[25]

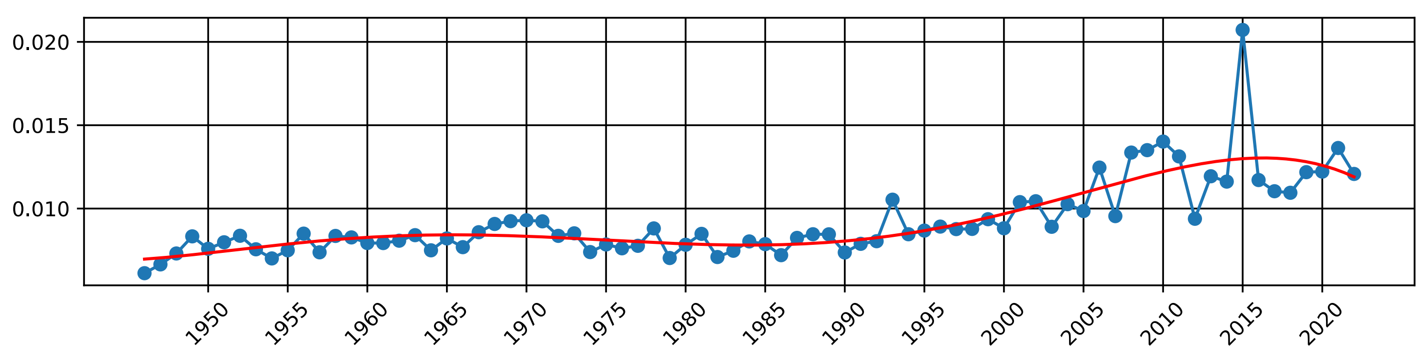
[25] Debray, Carolin (2023): Swearing, identity and power in professional interaction. In: *Journal of Pragmatics* 215, S. 145–158.

T2 - Crisis Word Rate from 1946 to 2022



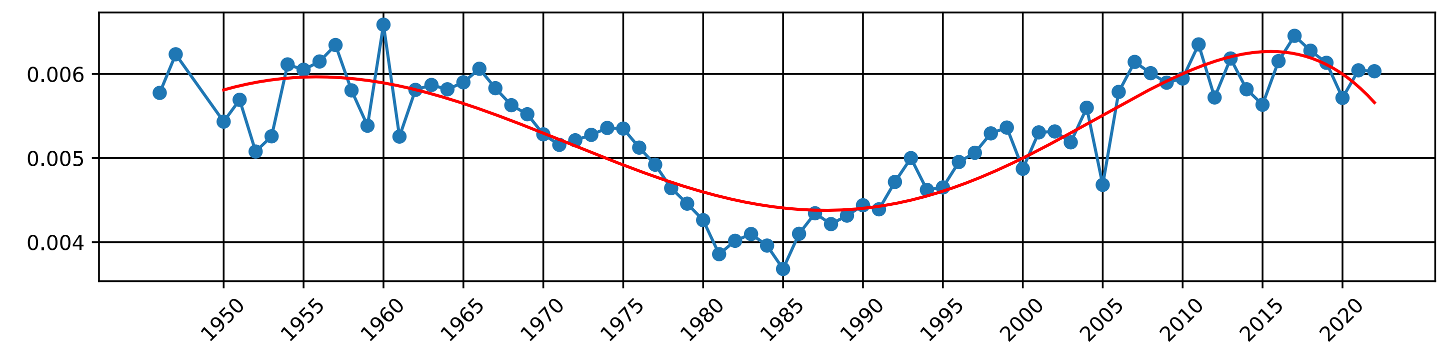
The occurrence trend for words semantically associated with crises shows an increase from 0.005 in the 1950s and 1960s to 0.007 in the 2010s. The years of the terrorist attacks of September 11 and the outbreak of the coronavirus mark the maximum values. It is also interesting to note that other historically significant crises, such as the Cuban missile crisis, the Vietnam War and the collapse of the Soviet Union, did not trigger such a high rate of crisis words. The high increase in recent years should not yet be overestimated, especially given the high explanatory power of Corona, the Russian-Ukrainian war and the deterioration of Sino-US relations. More significant is the gradual increase in the preceding decades, which, in combination with the assumption that it is not the intensity of the crises that has increased, suggests a dramatization of language use.

T3 - Number Rate from 1946 to 2022



The share of numbers in the overall spoken language of UNGD speeches has risen from 0.007 to 0.012 since 1950. 2015 is an isolated outlier with a relative frequency of more than 2 percent. Format errors cannot be conclusively ruled out for this extreme value and alternative explanations such as the Greek debt crisis, the refugee crisis, Chinese stock market turbulence, the 70th anniversary of the United Nations (243 mentions of "70 years" in 2015) and the adoption of the 2030 Agenda for Sustainable Development (270 mentions of "2030") would suggest similar outliers in other years (1995, 2008, 2018).

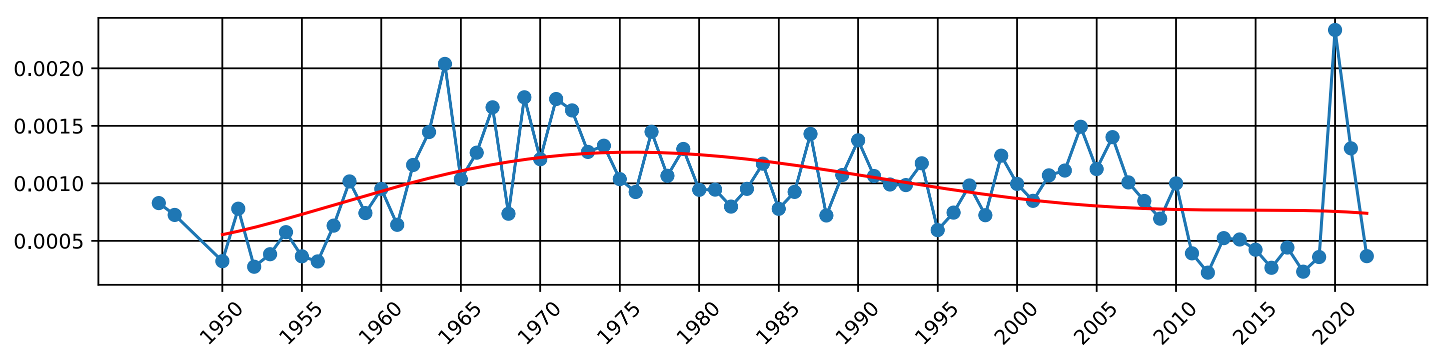
G1 - Self Reference Rate from 1946 to 2022



The graph on the occurrence of self-references shows a pronounced valley around the late 1980s with two local maxima in the 1950s and the 2010s. The range of values is narrower overall, with values between 0.004 and 0.006. Particularly striking is the rapid decline in the first-person singular from 1960 to 1980 and the rapid increase in the two subsequent decades. It is counter-intuitive that this is an inversion of the T1-Swear words curve, as both groups of words can be attributed to subjective and experience-based storytelling.

A comparison with Google's Books Ngram Viewer shows that the UNGD corpus is very close to the average English-language in that regard. The drastic increase of the "I" is even somewhat stronger in the contemporary literature, with values of 0.003 in 1980 and 0.007 in 2020.[26]

G2 - Direct Address Rate from 1946 to 2022



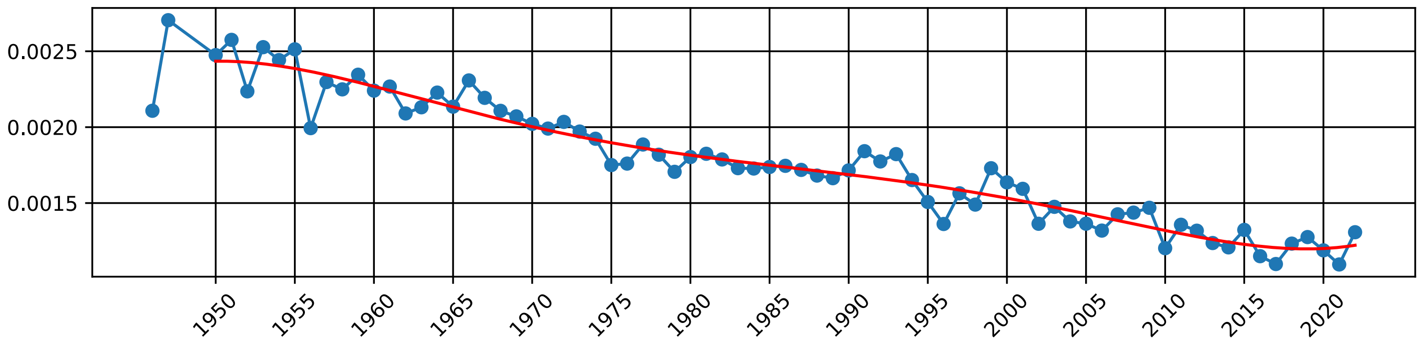
The graph of the occurrence of the second person singular seems to be related to the previously surveyed first-person singular. In consideration of the fact that sentences usually only bear one pronoun, this divergence makes sense. The outbursts of the Corona years, which seem to be characterized by direct address to the global public or the present representatives, are exciting here. However, the rapid increase only lasts for a short time and by 2022 the rate has already fallen back to the level before the outbreak.

The Ngram Viewer shows a clear deviation from the language pattern of the overall literature. Not only is the second person singular at 0.05 significantly above the value of the UNGD corpus, but since the turn of the millennium instead of slowly declining the rate has risen almost synchronously with the first-person singular. [27]

[26] Google LLC (2024): Google Books Ngram Viewer. (query-settings: I + me + mine + myself + my; 1945-2022; English; Case-Insensitive; Smoothing:3).

[27] Google LLC (2024): Google Books Ngram Viewer. (query-settings: you + you’re + yourself + yours; 1945-2022; English; Case-Insensitive; Smoothing:3).

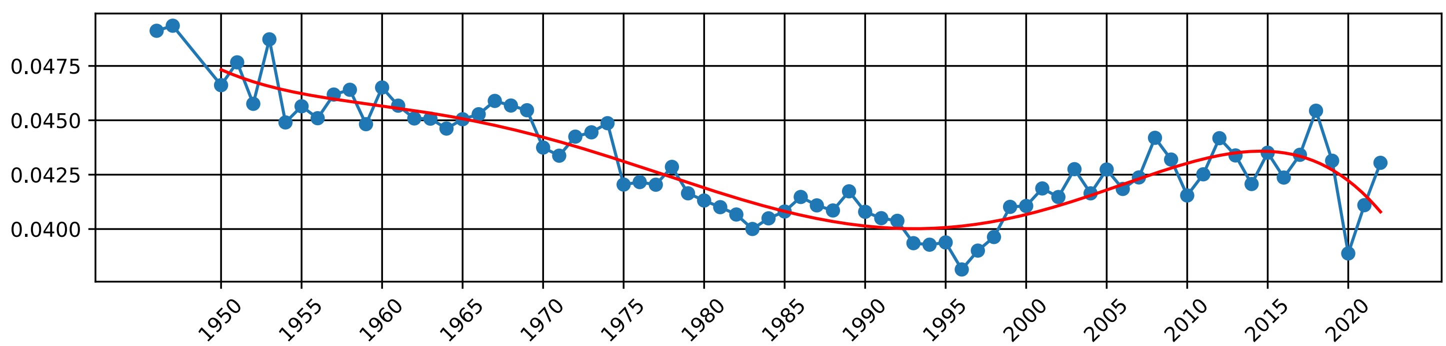
G3 - Modal Adverb Rate from 1946 to 2022



The graph of the share of modal adverbs in total text shows a clear and continuous downward trend. Between 1950 and 2020, the use of the word group halved from 0.0025 to 0.0125. Even in years of crisis this development is not halted.

A comparison with the Ngram Viewer suggests that the trend of the UN is once again contrary to the trend in general language patterns, which has either stagnated or increased since 1990.[28]

G4 - Degree Adverb Rate from 1946 to 2022



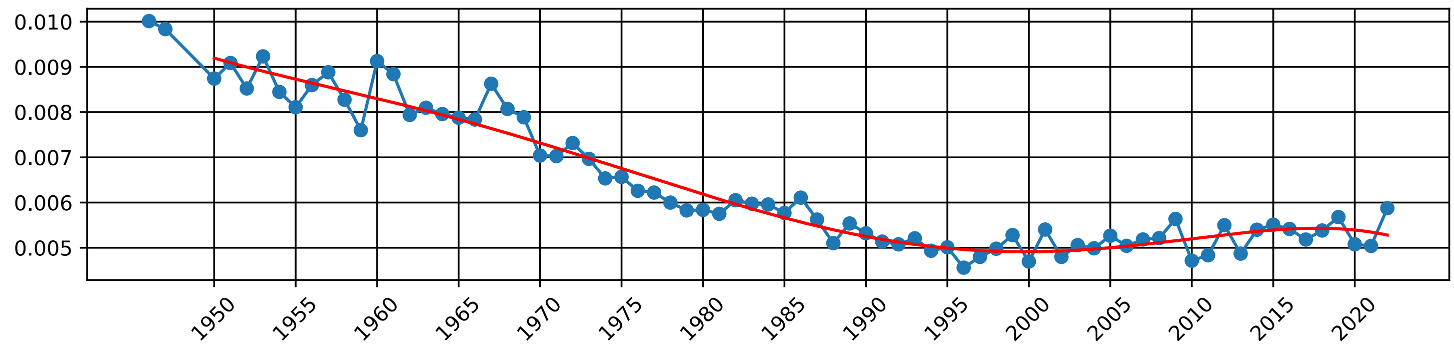
The graph of degree adverbs shows a similar decline over the entire period as its related modal adverbs. However, there has been a renewed increase in their use since the mid-1990s. Whether this increase has already reversed remains to be seen, as 2020 has already proven to be a year of exceptions in other metrics and one should not place much trust in polynomial regressions towards the edges.

The English-language literature for the same period shows largely stagnating values.[29]

[28] Google LLC (2024): Google Books Ngram Viewer. (query-settings: clearly, perhaps, obviously, actually, apparently, really, likely, maybe; 1945-2022; English; Case-Insensitive; Smoothing:3).

[29] Google LLC (2024): Google Books Ngram Viewer. (query-settings: very, hugh, big, small, tiny; 1945-2022; English; Case-Insensitive; Smoothing:3).

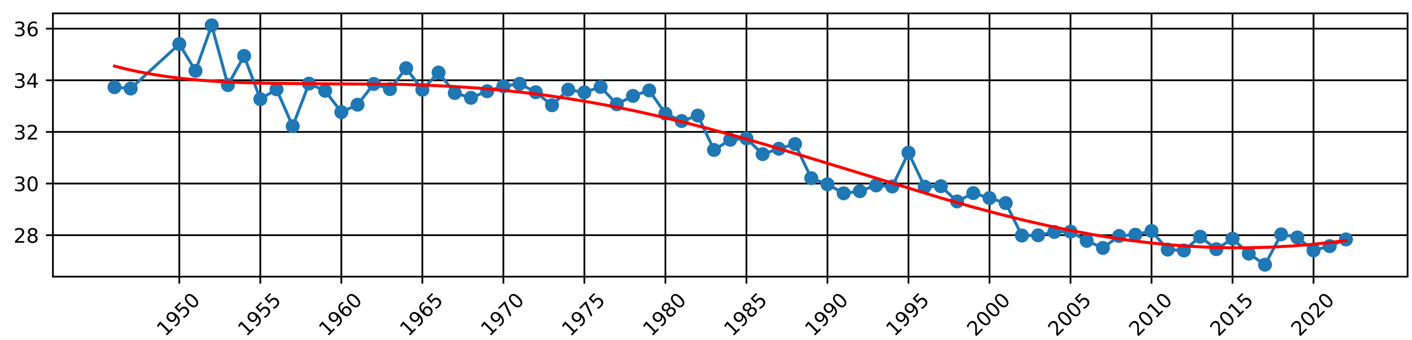
G5 - Negation Rate from 1946 to 2022



The rate of negation words also falls by almost 50% over the course of the 77 UNGD sessions. Once again, the low point was reached in the mid-1990s, with a slight increase in the following decade.

The proportion of negation in political communication has not yet been systematically measured, so no reference value can be provided here. Related research can, however, demonstrate a continuous increase in negations in the State of the Union speeches of US presidents.[30]

S1 - Average Sentence Length from 1946 to 2022

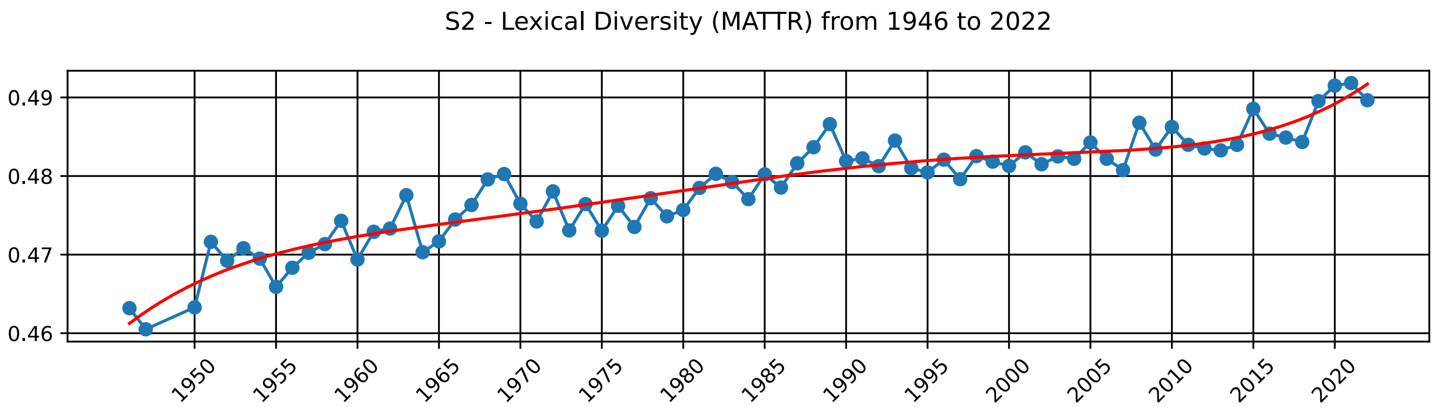


The average length of sentences has also decreased significantly since the first UNGD session in 1946. From 34 words per sentence at the beginning, 28 remain in 2020. The shrinkage occurs evenly between 1970 and 2010, with a plateau before and after that. It is also interesting to note that for a corpus in which all speeches from one year are merged, the same visualization shows a similar curve for the calculation of the international average, but with values of 30 in 1950 and 22 in 2020. This shows that there is a negative correlation between the overall length of the speech and the average sentence length, which can perhaps be explained by the fact that short speeches are prepared in advance and thus the linguistic standard may be particularly high.

An analysis of speeches by US presidents and UK prime ministers shows a similar trend over an even longer period, but with a lower number of words per sentence. The average sentence length in the two bodies of text is estimated to have fallen from 30 to 16 between 1900 and 2000. [31]

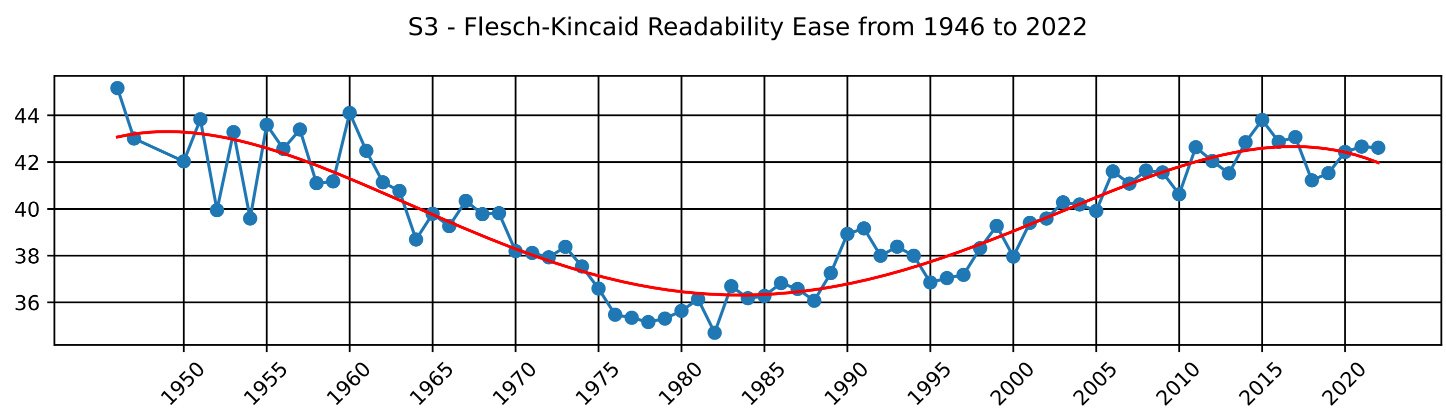
[30] Duran, Jose Manuel (2018): A corpus study of negation and their disruptive patterns in political discourse.

[31] Tsizhmovska, Natalia L.; Martyushev, Leonid M. (2021): Principle of Least Effort and Sentence Length in Public Speaking.



Linguistic diversity increased continuously but only slightly over the period under review. A possible explanation for this would be the multitude of issues representatives are expected to address by increasingly professionalized epistemic communities.

Overall, however, at 0.46 to 0.48, the MATTR value for a window size of 500 tokens is well below the average for the first five books of the Sherlock Holmes adventures, which were used as an example by the creators of the Moving Average Type Token Ratio. Here the value is between 0.46 and 0.58 while the average for English-language text is reported to be 0.6.[32]



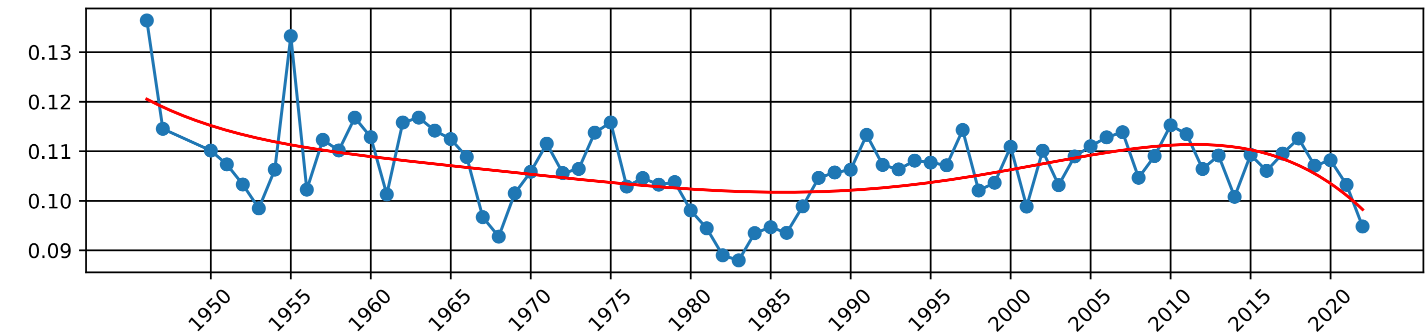
The readability ease shows the valley already known from the self-reference and intensity adverbs around the mid-1980s.

Readability fluctuates between 44 and 36, meaning that the entire period was rated as college (50-30) or "difficult to read" according to Flesch and Kincaid's grading system. For a rating of 40, an average of 25 words per sentence and 167 syllables per 100 words are assumed. Since we know that the number of words per sentence is significantly higher, it can be concluded that the number of syllables per word is somewhat lower. [33]

[32] Covington, Michael A.; McFall, Joe D. (2010): Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR).

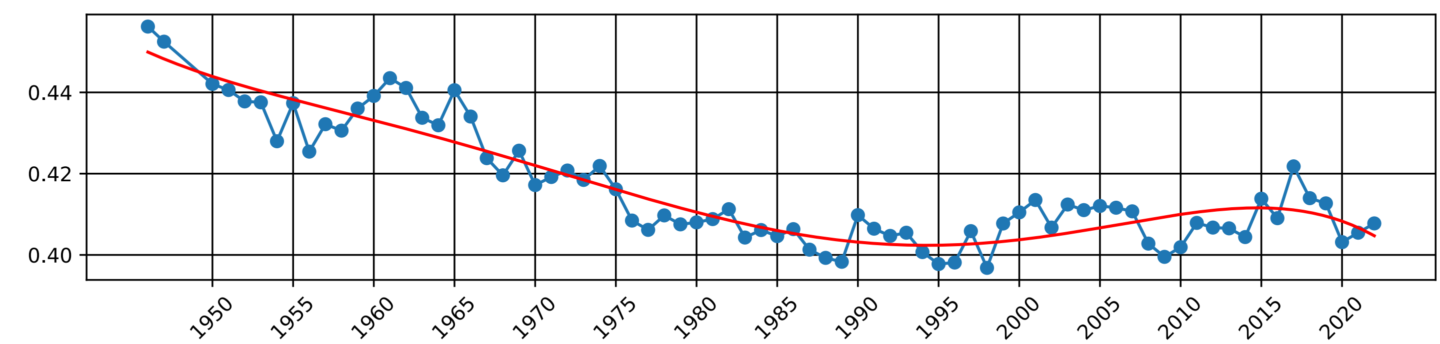
[33] Kincaid, J. P.; Fishburne, Jr.; Robert P., Rogers; Richard L., Chissom; S, Brad (1975): Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.

S4 - Sentiment Polarity from 1946 to 2022



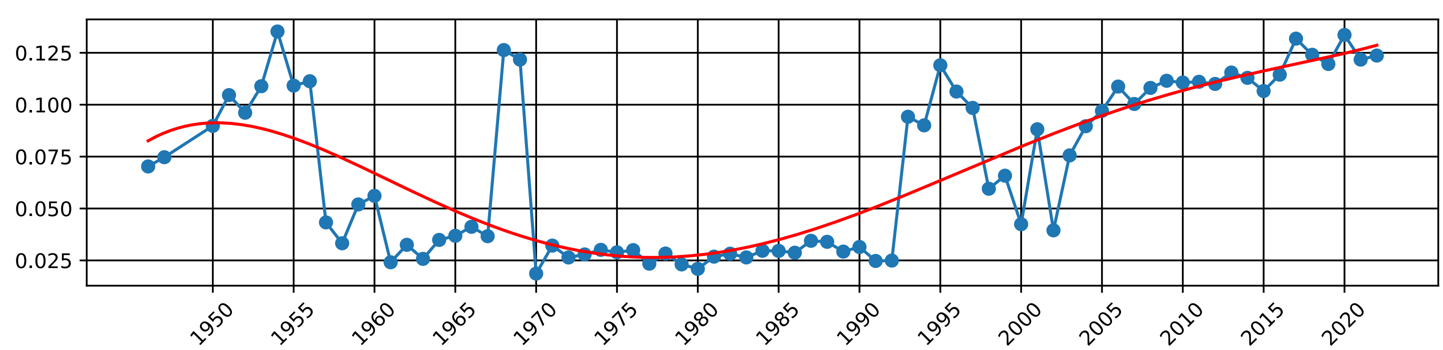
The evolution of sentiment polarity fluctuates strongly and shows no significant tendencies or clearly identifiable time intervals. There is only some movement at the edges, but it is not possible to speak of a generalized trend in the evolution of the sentiment polarity. The first session in 1946 and the year 1955, in which 15 states were admitted to the membership list, could be characterized by positive welcoming and thank-you speeches and the dip after the corona crisis could still prove to be a local minimum.

S5 - Sentiment Subjectivity from 1946 to 2022



Sentiment subjectivity shows a trend that is essentially like that of polarity, but is much more pronounced, so that we can speak of a continuous decline from 1950 to 1990 and an upswing in the early 2000s. Sentiment is highly dependent on the genre and the topic addressed, which is why it is particularly difficult to find reference values here.

C1 - Fake News Likelihood from 1946 to 2022



The development of fake news likelihood draws a graph that is already well known. The graph begins with an increased probability in the 1950s, then finds its minimum in the late 1970s before rising again to a new maximum in the early 2020s. Even at the end of the graph, the curve gives no indication of the value approaching a limit or even a point of reversal. However, high outliers in 1968 and 1969 as well as the drastic increase in 1994 cast doubt on the method. With the escalation of conflicts in Vietnam (1025 mentions across all speeches in 1968), Israel (738) and Czechoslovakia (253), the late 1960s offer several events that encourage the use of language that is also considered to convey fake news. 1994, on the other hand, with conflicts in Rwanda (74 mentions in 1994), Somalia (281) and Bosnia (313), received much less attention from the UN and yet shows a similar probability of fake news. Yet the consistency with which the individual texts have remained above the 10 percent mark since the turn of the millennium is remarkable and deserves more attention than can be paid to it in this report.

6 what to take away

Before we can answer our research questions, a few caveats need to be expressed.

Inconsistencies in language properties: Some language groups, such as modal adverbs, can contradict each other. The words clearly and maybe, for example, both belong to the word group of modal adverbs and yet have almost opposite semantic functions.

Comparative values: Each text body authorship format comes with its own set of interventions in the language. From conventions of greeting to transcription and translation. For this reason, the properties must be understood above all in their comparison with each other, i.e. in a time series comparison, and the classification based on external texts must be treated with caution.

Blackbox classifier: In 2020, the University of Tübingen conducted research into an image recognition system. The system was designed to label images and thus recognize different species of fish, among other things. The system performed well in classifying tenches, but when a bitmask was used to determine which pixels of the image were particularly decisive for the classification, it was found that it was the hands of the proud anglers and not the properties of the tenches they were looking for. Our zero-shot classifier could also have been affected by proud anglers and drawn the conclusion from its training data that the boilerplate of a website associated with fake news, or individual formulations are sufficient for the fake news label.

RQ1: How did linguistic properties of the UNGD speeches change from 1946 to 2022?

Each of the 13 language properties measured has shown an evolution that, firstly, can be clearly distinguished from a purely random fluctuation and, secondly, has an intensity that renders it relevant for research. However, the individual properties paint a picture that is rich in contradictions and complications.

In the area of vocabulary, for example, the language became more dramatic and emotionally charged measured by the increase in crisis words and the first-person singular according to the research literature. The increase in the occurrence of numerical quantifications and the decrease in modal adverbs, degree adverbs and negations, on the other hand, describes a more objective and scientific style.

Language complexity poses similar difficulties. While the continuous decline in average sentence length and the increase in readability since the 1980s indicate a simplification of language, lexical diversity peaked in 2020.

It is also interesting to note that the evolution of the researched language properties was often not linear, but that the years between 1975 and 1995 show the highest or lowest values in 7 of the 13 properties (T1, G1, G2, G4, G5, S3, S5).

Based on these results the following hypotheses emerge:

H1 - Professionalization: An increasing professionalization of political communication and international relations in accordance with an efficiency logic derived from marketing would be possible. In their standard work on political communication, Lehrman and Schnure make recommendations on language properties. They recommend limiting the number of words per sentence, a positive overall tone, the avoidance of negations and the extensive use of content-free filler words such as modal adverbs and degree adverbs. In fact, of 13 properties, only 2 clearly contradict the professional recommendations: The decrease in direct address (G2) and the increase in lexical diversity (S2). All other points either have no clear tendency or follow the suggested guidelines.[34]

H2 - Back to normal: Describing the current period as the climax or nadir of a linear development has a dramaturgical and scientific appeal, but the measured properties only partially support this theory. It would be equally possible to describe the years between 1975 and 1995 as a time of outliers, when there was a lot of swearing and little talk about oneself, with speeches that were difficult to understand and a highly objective. In contrast, the edges of the record are at a similar level in these metrics and so we could speak of a caesura that has already been overcome. The period is strikingly synchronized with the hot phase of the cold war fever-curve, but other theories cannot be ruled out. The valleys of the graphs can be explained not only by the peculiarity of the late 20th century, but also by the linguistic similarity between the political culture of the post-war period and the modern era.

RQ2: To what extent can the linguistic trends of the UNGD be seen as indicators of an increasing resemblance to fake news?

The statement that the language of UNGD talk has changed over time in a way that indicates an increase in political communications resemblance to fake news being proliferated cannot be confirmed. Of the nine properties that are positively correlated with the probability of fake news (T1, T2, G1, G2, G3, Gg4, G5, S3, S5), only the use of crisis words is significantly higher in 2020 than in 1950. Of the negatively correlated properties (T3, S1, S2, S4), only the average sentence length is significantly lower in the 2020s than at the beginning of the recordings.

On the other hand, with the decrease in modal adverbs, degree adverbs and negations, as well as the increase in lexical diversity and increasing subjectivity, five properties contradict the theory. The other six refer to the unique properties of the 1980s or do not allow any statement to be made about the trend of development.

Finally, there is the fake news likelihood classifier. It also tends to support H2, although the first increased probability in the decade after the founding of the United Nations is based on only ten data points. By contrast, the second hill, which begins in 2005, is more robust and stable. Like with the tench, one would now have to dig deeper into the network architecture to find out which features and patterns were particularly decisive for such a classification.

In summary, the language of political communication in international relations has not changed in a way that represents or contributes to the proliferation of fake news. Alarmist voices that claim that the evolution of language has become a vehicle for populism, the age of the post-factual and ultimately fake news, can be rebuked with complications and contradictions.

7 List of References

Altay, Sacha; Berriche, Manon; Acerbi, Alberto (2023): Misinformation on Misinformation: Conceptual and Methodological Challenges. In: Social Media + Society 9 (1), 205630512211504. DOI: 10.1177/20563051221150412.

Berrondo-Otermin, Maialen; Sarasa-Cabezuelo, Antonio (2023): Application of Artificial Intelligence Techniques to Detect Fake News: A Review. In: Electronics 12 (24), S. 5041. DOI: 10.3390/electronics12245041.

Covington, Michael A.; McFall, Joe D. (2010): Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). In: Journal of Quantitative Linguistics 17 (2), S. 94–100. DOI: 10.1080/09296171003643098.

[34] Lehrman, Robert (2010): The political speechwriter's companion. A guide for writers and speakers.

Debray, Carolin (2023): Swearing, identity and power in professional interaction. In: Journal of Pragmatics 215, S. 145–158. DOI: 10.1016/j.pragma.2023.07.002.

Duran, Jose Manuel (2018): A corpus study of negation and their disruptive patterns in political discourse. In: Letras (56), S. 15. DOI: 10.5902/2176148531175.

Edelman (2024): 2024 Edelman Trust Barometer. Global Report.

Egelhofer, Jana Laura; Lecheler, Sophie (2019): Fake news as a two-dimensional phenomenon: a framework and research agenda. In: Annals of the International Communication Association 43 (2), S. 97–116. DOI: 10.1080/23808985.2019.1602782.

Farkas, Johan; Schou, Jannick (2018): Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. In: Javnost - The Public 25 (3), S. 298–314. DOI: 10.1080/13183222.2018.1463047.

Fletcher, Richard; Cornia, Alessio; Graves, Lucas; Nielsen, Rasmus Kleis (2018): Measuring the reach of 'fake news' and online disinformation in Europe.

Fletcher, Richard; Robertson, Craig T.; Nielsen, Rasmus Kleis (2021): How Many People Live in Politically Partisan Online News Echo Chambers in Different Countries? In: JQD 1. DOI: 10.51685/jqd.2021.020.

Google LLC (2024): Google Books Ngram Viewer.

Jankin, Slava; Baturo, Alexander; Dasandi, Niheer (2024): United Nations General Debate Corpus 1946-2022. Harvard University.

Kincaid, J. P.; Fishburne, Jr.; Robert P., Rogers; Richard L., Chissom; S, Brad (1975): Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel.

Lehrman, Robert (2010): The political speechwriter's companion. A guide for writers and speakers.

Loria, Steven (2024): TextBlob Documentation. Release 0.18.0.post0.

Moore, Ryan C.; Dahlke, Ross; Hancock, Jeffrey T. (2023): Exposure to untrustworthy websites in the 2020 US election. In: Nature human behaviour 7 (7), S. 1096–1105. DOI: 10.1038/s41562-023-01564-2.

Nazeer, Imran; Yousaf, Shahneela; Anwar, Nazia (2023): Analyzing Linguistic Shifts in Political Discourse: A Corpus-Based Study of Political Rhetoric in the Digital Age. In: Pak. J. Humanit. Soc. Sci 11 (4). DOI: 10.52131/pjhss.2023.1104.0661.

Palmer, Martha; Hwa, Rebecca; Riedel, Sebastian (Hg.) (2017): Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.

Petrou, Nikolas; Christodoulou, Chrysovalantis; Anastasiou, Andreas; Pallis, George; Dikaiakos, Marios D. (2023): A Multiple change-point detection framework on linguistic properties of real versus fake news articles. In: Scientific reports 13 (1), S. 6086. DOI: 10.1038/s41598-023-32952-3.

Pfundstein, Julius (2024): UNGD-linguistic-patterns. Investigation of the lingusitic evolution of speech in the United Nations General Debate for 1946 to 2022. GitHub. https://github.com/Yakuzzo/UNGD-linguistic-patterns.

Python (2024): re - Regular expression operations. Version 3.12.5: Python.

Rashkin, Hannah; Choi, Eunsol; Jang, Jin Yea; Volkova, Svitlana; Choi, Yejin (2017): Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In: Martha Palmer, Rebecca Hwa und Sebastian Riedel: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, S. 2931–2937.

Romero, Manuel (2021): bert-tiny-finetuned-fake-news-detection: Hugging Face. https://huggingface.co/mrm8488/bert-tiny-finetuned-fake-news-detection/tree/main.

Tida, Vijay Srinivas; Hsu, Sonya; Hei, Xiali (2022): A Unified Training Process for Fake News Detection based on Fine-Tuned BERT Model.

Tsizhmovska, Natalia L.; Martyushev, Leonid M. (2021): Principle of Least Effort and Sentence Length in Public Speaking. In: Entropy (Basel, Switzerland) 23 (8). DOI: 10.3390/e23081023.

van den Rul, Celine (2020): Why Have Resolutions of the UN General Assembly If They Are Not Legally Binding?

Wiktionary - The free dictionary (2024): Category: English degree adverbs. https://en.wiktionary.org/wiki/Category:English_degree_adverbs.

Wiktionary - The free dictionary (2024): Category: English modal adverbs. https://en.wiktionary.org/wiki/Category:English_modal_adverbs.

Wiktionary - The free dictionary (2024): Category:English Swear Words. https://en.wiktionary.org/wiki/Category:English_swear_words.

WorldData (2024): Member States of the United Nations.