

Understanding the Linguistic Complexity of World Bank Annual Reports

Thea Madsen

t.madsen@exchange.hertie-school.org

Xiaoyan Hu

x.hu@mpp.hertie-school.org

Dang Ngoc Huy

h.dang@mpp.hertie-school.org

1. Research paper summary (2 pages)

Bankspeak: The Language of World Bank Reports, 1946–2012, by by Moretti Franco, and Dominique Pestre *New Left Review*, vol. 92, 2015, pp. 75–99 (2015) is the key research paper that inspires our selection of research topic. It investigates the linguistic development of the notoriously incomprehensible World Bank Annual Reports from 1946 until 2012. The paper provides two main investigations into the semantic transformations and grammatical patterns of the Bank.

Investigating the semantic transformations of the first 20 years of the annual reports, Moretti and Pestre found that nouns were central and frequent (Moretti and Pestre 2015: 2). Adjectives were rare and only used descriptively. Verbs concretized action e.g. to encourage, support or improve. The nouns could be grouped in two main clusters of primarily the economic activities - loans and investments to sponsor infrastructural development projects; and the much smaller cluster describing the operation - demands, analysis and evaluation of programs. The last twenty years were characterized by new semantic clusters. The first cluster of finance indicated the move from agriculture and industry to financial activities e.g. growth, interest and debt. The second cluster of management indicated the change toward monitoring, controlling, rating, goals, agendas, opportunities, challenges and strategies rather than previously suggesting and assisting solutions and publishing descriptive reports.

Another strand of investigation of the paper into the grammatical patterns of the reports uncovered a bureaucratization of the Bank's discourse generating a syntax so opaque that it could be difficult for others to understand due to a lack of concreteness for instance words like key, global, innovative and enlightened (ibid.: 13). The Bank also used two to three times higher frequency of nominalizations - words typically ending in -tion, -sion and -ment (e.g. implementation, extension, development) - than comparable academic prose. This has the effect of transforming actions and processes of the Bank into abstract objects e.g. "South-

South cooperation" rather than plainly writing that the Bank is supporting countries cooperating with each other (ibid.: 14).

In sum, Moretti and Pestre found that the World Bank language semantically and grammatically developed into essentially another language (ibid.: 2). In the first decades after 1946, the language was simple, descriptive and straightforward; in the last decades before 2012, the language developed into so-called *Bankspeak*, which is a self-referential, complex and codified organizational syntax removed from everyday language difficult for people outside the World Bank to understand. It is this Bankspeak that we wish to continue the investigation of from 2012 until the last report published in 2018. Moretti and Pestre criticized the World Bank for writing the reports in a bankspeak too unclear for its readers to understand how the bank is handling its money and who is responsible for the consequences of its politics. Diverse actors within - attesting to that the Bankspeak is also an acknowledged problem for some World Bank employees - and outside the World Bank have criticized the institution for and warned against continuing down the distancing path followed for decades [4].

Although subject-specific language can be efficient when communicating with peers, the complexification of the language challenges the understanding of the Annual Reports of general readers, which is problematic because these are the people that the World Bank is intended to serve. According to the World Bank, in 2019, it is a global partnership institution encompassing 189 member countries with staff from more than 170 countries and offices in over 130 locations [5]. The goal of the World Bank is to aide in implementing sustainable solutions to reduce poverty and build shared prosperity in developing countries. The World Bank Annual Reports are intended to provide insight into how the World Bank cooperates with its partnering countries to support global sustainable development, promote shared prosperity and end extreme poverty by 2030. Failing to communicate in clear and plain language is a transparency issue and a failure to holding itself accountable to stakeholders, ensure successful policy recommendations. If people can-

not understand their financial institutions, they are unable to challenge institutional workings, which can lead to distrust, lack of representation and ultimately inefficient policies unable to fulfil their objectives. If the World Bank does not alter its course, fulfilling its goals will be fundamentally challenged because its language is avoidably difficult, which can impair thoroughly learning from past evaluations to make the necessary future implementations.

After receiving inspiration from Moretti and Pestre on the content of our research, we also look into other publications to seek guidance on the methodology of complexity measurement. Already in the Bankspeak report, Moretti and Pestre looked into some aspects that can be an inspiration to the assessment of language complexity. The language of the World Bank evolves throughout the 66 years. Despite that the two authors pointed out the norminalization and singularization of the Bank language, they did not directly conclude on the complexity of the text as a whole.

This gap was filled with ample research in political messaging and language sophistication. Some political scientists adopt measures from education and communication studies, such as the research by Bischof and Senninger (2018) [3]. Their research tried to re-establish the link between the language applied during election campaigns and citizens' ability to position parties in the ideological space. To do this, they sourced election manifestos from all major political parties in Austria and Germany in the period 1945 – 2013. In the first half of their research on language complexity, Bischof and Senninger applied the LIX formula to generate a score on the readability of each manifesto. Then the authors compared those scores with that of other types of text ranging from Cinderella to Habermas to locate the complexity level in the human conception.

They also put forward a new text sophistication measurement focusing on political communication. To determine the complexity, they combined human judgment as the “golden standards”, Bradley-Terry model as well as random forest algorithm to “predict” the easiness parameter for a given new text. Comparing to previous literature, the research by Benoit et al. [3] could have pushed the precision of political text analysis to another level by involving efficiently human coders and also use a prediction model to incorporate a broader range of textual complexity indicators. However, there is also a question of scalability as the need for human coders and judgment in itself could lead to bias in the model as well as a difficulty to replicate without the necessary resources.

2. Project Description (1 page)

2.1. Motivation

Every year, the World Bank publishes an annual report that explores the challenges encountered by developing na-

tions and what it is doing to help people help themselves. From its inception in 1944 until present day, its mission statement has shifted from the rebuilding of Europe towards fostering international development. In a way, these annual reports are mirrors that reflect the reality of their time and shine a light on the advancement of international development and the progress of developing nations. Yet, it is questionable if these reports could offer any real insights at all. Criticisms have been drawn from both within and outside of the organization on the unnecessarily complex and impenetrable language that is utilized in these reports, making them inaccessible to the very audience they are intended to serve.

As Jonathan Lancaster (New Yorker, 2014) eloquently expounded: “The language of money is a powerful tool, and it is also a tool of power. Incomprehension is a form of consent. If we allow ourselves not to understand this language, we are signing off on the way the world works today.” The World Bank Annual Reports are important, but it is even more important that these reports can be read, understood and processed by those they could bring the most value.

This paper, therefore, seeks to comprehend the linguistic sophistication of the World Bank Annual Reports. It will build upon the works done by Franco Moretti et al (2015) which has established that an overtly sophisticated and codified Bankspeak semantic permeates these reports. The research will focus on exploring the complexity of language used in these publications from 2013 up until now and investigate whether such systematic complexity still exists within their writing and communication of ideas.

2.2. Task

The main task of our research will be to analyze the corpus of the World Bank Annual Reports from 2013 to 2018 to evaluate their linguistic complexity and examine whether this is a trend that is consistent over the years. The input of our research will be the main text of these reports, and the output will be a comparison that can measure their readability through metrics such as the Gunning Fog index and the Flesh Reading Ease formula.

2.3. Data

Fortunately, all of World Bank Annual Reports from 1946 to present days are preserved in the organization's data portal. The recent reports average at about 100 pages in PDF format with accompanying graphs, tables and photographs. Rough OCR versions of these reports in plain text format are also available, which will be the main data source that forms the corpus in our analysis. In the initial phase, we will process the data as is, without any further pre-processing or transformation. The bulk of our research will focus on the reports from 2013 to 2018 to understand the linguistic sophistication of these texts.

2.4. Method

Whereas Franco et al. (2015) [1] focused more on how clusters or words, specific terminologies and the Bank's vocabulary developed over the years in the Annual Reports, our research will take a more quantitative approach by investigating the textual and readability complexity of these reports in the aggregate level. We will explore lexical diversity and sophistication of these texts by endeavoring to answer the two main research questions:

1. Is there, in fact, a trend of increasing linguistic complexity in the World Bank Annual Reports from 1947 to 2012?
2. Is the same level of complexity maintained in recent reports from 2013 to 2018?

To answer these two questions, our research will focus on assessing the readability of these publication by using the Gunning-Fog Index, Flesch-Kincaid Grade Test and the Flesch Reading Ease Formula.

- **Readability Score:** we will use, as the main metrics, the **Flesch Reading Ease (FRE) Formula**, which is one of the classic readability formulae. This index is distinct from others in that it adopts a score range of decreasing difficulty from 0 (hard to understand) to 100 (easy to understand). The score will operate as the evaluation of how easy it would be for a piece of text, and in this case, a specific annual report, to be understood and engaged with.
- **Readability Grade Levels:** our research will utilize the **Gunning-Fog Index**, and **Flesch-Kincaid Grade Test**, both have witnessed increasing usage in the literature on accounting and financial text understanding. These two metrics capture the number of years of education a person would need to undergo in order to read a text once and understand it.

With these indexes, we will be able to measure the linguistic complexity of the economist language of these reports and understand whether they might cause readability problems. The textual sophistication measurement model, illustrated in Benoit et al. (2019) [2], though being a fascinating approach to understanding complexity, is difficult to replicate in this context, as it adds an extra layer of human coders and judgement. In addition, the conclusion that the Benoit et al. (2019) [2] reached was that complexity analysis using FRE score is relatively similar to the best performing model of predicted probability in the paper, particularly in trend and general direction. Therefore, we conclude that the FRE is still a good measure with which to gauge an understanding of sophistication in the World Bank Annual Reports.

In the final phase, if time permits, we will compare the complexity of the World Bank Annual Reports with those of another financial institution to inspect whether such increasing complexity is perhaps a trend among these organizations.

2.5. Baseline

Based on the interpretation of these different metrics, our baseline will differ for each of them:

- **Flesch Reading Ease Formula:** we will set the baseline at 60 as it corresponds to Plain English and easy to be understood by anyone with an 8th grade education;
- **Flesch-Kincaid Grade Test:** we will consider 8 to be the standard (with 80 percent of American being able to understand the content with ease).
- **Gunning-Fog Index:** similar to the other metrics, we will set the baseline of this index at 8 (the range of 8-10 is often recommended as best for ease of understanding).

3. Evaluation

Achievement for this project will be a successful implementation of the methodology for measuring textual sophistication with the corpus of World Bank Annual Reports. We expect that there will be a trend of increasing difficulty in readability of the World Bank Annual Reports over the years, particularly from the late 80s and early 90s onwards, and that this difficulty level will continue until the most recent publications. As there exists no other quantitative research on the linguistic complexity of the World Bank Annual Reports, except for Moretti et al. (2015) [1], it is difficult to anchor down a specific or previously discovered state-of-the-art performance measure. However, we hope that this paper will provide a basis for continual research into the readability of these organizations' reports and research.

4. Contributions

Although all coding and writing process will be implemented as a team together, we have decided to delegate areas of specific responsibility to make the project management more efficient: Xiaoyan Hu will lead the collection and preprocessing of the data from the World Bank portal. She will work in tandem with Thea Madsen on the literature review. Thea Madsen will be in charge of exploratory analysis of the corpus and work with Dang Ngoc Huy on the inspection of the language. She will also contribute to the literature review and analysis of the results. Dang Ngoc Huy will be responsible for implementing the assessment on the language complexity and reports on the methodology, results and analysis of this appraisal process.

5. Project Page

The GitHub page for the project can be accessed **here**.

References

- [1] Franco Moretti and Dominique Pestre. *Bankspeaks: the Language of the World Bank Report*. New Left Review, vol. 92, 2015, pp. 75–99, 2015
- [2] Kenneth Benoit, Kevin Munger, and Arthur Spirling. *Measuring and Explaining Political Sophistication Through Textual Complexity*. American Journal of Political Science, 2019.
- [3] Daniel Bischof and Roman Senninger, *Simple Politics for the People? Complexity in Campaign Messages and Political Knowledge*. European Journal of Political Research 57 (2): 473–495, 2018.
- [4] Corinne Purtill, *The World Bank's Misguided Effort to Improve Its Writing Signal an Economics Communication Crises*. Quartz, 11 June 2017.
- [5] World Bank Official Website - Mission Statement, <https://www.worldbank.org/en/who-we-are>.