



## **Text Mining Strategies applied on the annual reports of the International Monetary Fund. A look at the crisis<sup>1</sup>**

Feldman, Ana Nora \*

Universidad Nacional Arturo Jauretche – Provincia de Buenos Aires - Argentina  
– [anafeldman@gmail.com](mailto:anafeldman@gmail.com)

The annual reports of the International Monetary Fund issued annually under the name of “World Economic Outlook” for the years 2005 to 2012 are analyzed in this Paper by using the techniques of Statistical Analysis of Textual Data<sup>2</sup>(ATD). This text mining tool allows us to see the way the IMF describes the world crisis in their reports, highlighting their strengths and weaknesses in their role as the ultimate guarantor of the global economic balance.

In particular, this paper looks at the result of the analysis, having processed the textual corpus with Taltac<sup>3</sup> software. Sen and Kliksberg<sup>4</sup> consider that the Reports on Global Economic Prospects may be useful to understand the behavior of the IMF in the context of the financial crisis.

The three factor model proposed by Paul Nemiña<sup>5</sup> (2009) is used. This allows the data analysis and also highlights the logic related to internal interests of the institution and its typical bureaucracy, typical of all major multilateral institutions (Feldman, 1994).

These reports, meet the goal of preaching the hegemonic conception on markets and policies, seeking to satisfy related communication and marketing strategies in order to align public opinion, government officials and objectives behind this concept.

This lexicometric analysis has identified important differences over time. The text included in the 2008 Report is longer than the other ones (representing 16% of the total corpus), therefore indicating the presence of “conceptual dissonance”<sup>6</sup> (Feldman, 1994, 2005). The richness of the vocabulary<sup>7</sup> changes substantially in 2009 (9.56% of corpus). In both cases, the main difficulty is naming things by their name (such as when referring to “crisis”).

The distribution analysis of some statistically significant word allows the understanding of the speeches’ axis over the years, strongly related to the political objectives of the Institution.

In the paper to be submitted other lexicometric and corpus word distribution features will be analyzed, offering an IMF communicational study thanks to the use of ATD related statistical tools. The various strategies implemented by the IMF will also be demonstrated to a) initially, ignore the crisis, b) later, underestimate the crisis and c) finally, declare the end of the crisis.

**Keywords:** textual data analysis; text mining; economic language; crisis.

---

<sup>1</sup> A selection from a Doctoral Thesis in Economics from the National University of La Matanza (2014).

<sup>2</sup> This statistical analysis is based on “analyse des données” proposed by J.P. Benzécri (1973). Indexes and classical statistical measures in linguistics and in lexical statistics were developed by Ch. Müller (1973), M. Tourmier (1980) and P. Lafon (1984). In 1988, L. Lebart and A. Salem, in “Statistique Textuelle” (DUNOD), systematized and formalized the technique. Since 1990 JADT (International Seminar) are held.

<sup>3</sup> [www.taltac.it](http://www.taltac.it)

<sup>4</sup> Sen, A. y Kliksberg, B. 2007. *Primero la Gente*. Ed.Deusto. 9na edición Ed. Temas, 2011

<sup>5</sup> <http://rrii.flacso.org.ar/pablo-nemina-investigador-asociado/>

<sup>6</sup> Extensive explanations used derived from the fact that they have to explain something “unfamiliar”.

<sup>7</sup> V/N, where (V) is the amount of words and (N) occurrences



## 1. Introduction

The denial of the crisis, only recognized at the end of 2009, is consistent with the mission that the IMF considers to be met, that is to say, lecturing on how governments should run their economies (Weisbrot, Lamb and Sandoval, 2009). The main question regarding the IMF's position responding to the largest financial crisis after the Great Recession of 1930 is to understand how it foresaw, saw, described and analyzed the background of the world's economy during the last decade.

The aim of this paper is to understand, from the discourse analysis, which has been the position of the International Monetary Fund (IMF) regarding the global economic and financial crisis that erupted at the end of the year 2007. Much has been discussed about the foresight of the crisis and what was the position of the IMF on its consequences. This paper analyzes the Fund's reports and its contents, applying techniques of statistical analysis of textual data for an appropriate selection of words and contexts.

## 2. Objectives

In order to perform the aforementioned analysis, within a content analysis point of view, an exploration (Text Mining) of the "World Economic Outlook" annual reports was carried out using Statistical Analysis of Textual Data (ATD) tools.

Reports from 2005 until 2012 were selected to record and analyze the period that preceded the crisis. Both the contents and contexts will allow us to understand which was the way the IMF used to describe the crisis and which are the issues associated with it.

The hypothesis is that the World Economic Outlook reports may be useful to understand the behavior of the IMF in the context of the financial crisis. The reports would be fulfilling the purpose of preaching the hegemonic conception of markets and policies, seeking to satisfy objectives related to communication and marketing strategies, in order to align public opinion, officials and governments behind this concept.

## 3. Statistical Analysis of Textual Data

In recent decades the use of Statistical Analysis of Textual Data (ATD) techniques has shown a great development. For example, in connection with content analysis, discourse analysis, lexical statistics or lexicography, the application of these methods is easier now thanks to the use of a software which is available in the market. These instruments allow, among other things, to treat large volumes of data through the identification of words, segments or sentences of interest by using some quantitative parameters (frequency, statistical distribution, etc.).

The ATD allows the exploration and analysis of texts as diverse as literary ones, answers to open ended questions in questionnaires, interviews or focus groups, political speeches, newspaper articles and other material of significance for researchers.

Nowadays, this tool is more often used by official institutes of statistics, starting from the collection of data to the analysis of occupational categories used in surveys. A very interesting example of the application of these techniques can be found in the work on "distribution and use of time" developed by Prof. Sergio Bolasco and his team together with the Italian Institute of Statistics (ISTAT) (Bolasco S. et al., 2007).

The text exploration and analysis tools performed through multidimensional exploratory statistical resources (analyse des données) used for the analysis of large data matrixes is related to an ecological approach to treat qualitative data (Bolasco, 1993) in order to allow the identification of interesting data (concepts, words, phrases). This is an inductive method that considers language as a "system".

The following table depicts the potentiality of the method.

## POTENTIALITY OF THE STATISTICAL ANALYSIS OF TEXTUAL DATA

		By its generality	By its focus
		↓	↓
		Creates a synthetic vision	Focus on interesting details
Explores the relations of co-occurrence	Instrument of statistical analysis →	Maps of Core Issues (context typologies)	Words associations (sequences analysis)
Verifies the relations of similarity and difference between parts of the corpus	Instrument of statistical analysis →	Correspondence analysis (Cluster analysis)	Specificities

### 4. Main Results

The Reports of the years 2005, 2007 and 2008 are highlighted due to their extension (the last one concentrates the 16% of occurrences of the corpus).

#### DISTRIBUTION OF OCCURRENCES BY YEAR OF THE CORPUS

Year	Occurrences	%	$\bar{X}$ difference with
2005	133551	13	+
2006	121995	12	-
2007	147937	14	+
2008	168906	16	+
2009	114498	11	-
2010	124365	12	-
2011	120454	11	-
2012	124630	12	-
<b>Total</b>	<b>1056336</b>	<b>100</b>	

### FORMS, OCCURRENCES AND RICHNESS OF VOCABULARY IN THE CORPUS

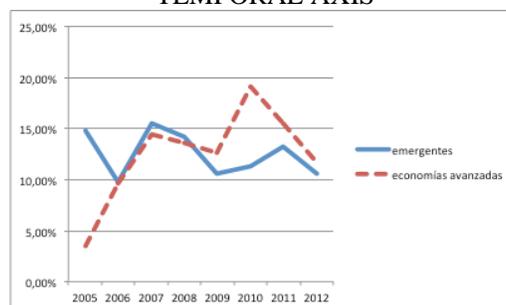
Year	# forms (V)	Occurrences (N)	Richness of Vocabulary (V/N)
2005	12301	133551	9,21%
2006	11203	121995	9,18%
2007	12454	147937	8,42%
2008	15283	168906	9,05%
2009	10951	114498	9,56%
2010	11332	124365	9,11%
2011	10656	120454	8,85%
2012	11112	124630	8,92%

It is worth to point out that during the year in which the IMF recognizes the occurrence of the financial crisis the richness and variety of the language reaches its peak.

Even if they use an economic language, some features related with the IMF's Reports have been brought out in this paper. Besides talking (and a lot) about itself, references are made mostly to those who wrote the reports, that is to say, to the technical staff of the IMF. The second institution in terms of frequency which is referred more often is the United States of America, the country that almost has a hegemonic power and the only one with veto within the IMF. Therefore, the main focus of attention related with the work of the IMF's staff is strictly linked to the interest of the most powerful countries. In the eyes of the IMF, the problems of the USA are also the problems of everybody else, followed by the European Union (and the Euro Zone), then Japan and other (advanced) Economies.

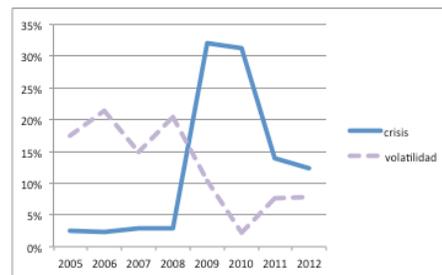
According to the level of relative economic development, countries are classified as advanced economies and emerging markets. The performance of advanced economies in the Reports grows as the crisis is recognized.

### DISTRIBUTION OF THE "EMERGENT" AND "ADVANCED ECONOMIES" IN THE TEMPORAL AXIS



The crisis has been untimely admitted in its Reports: this was in 2009, as the following graph shows. This is, mainly, a financial and/or banking crisis focused in Europe, specifically within the Euro zone. Although it has been already stated that the IMF openly mentions the occurrence of a crisis in 2009, it must be taken into account that this organization has not been unfamiliar with several signs that depicted the proximity of a crisis. Between 2005 and 2008, the IMF uses the term “volatility” as a signal that “corporate long term assets become short term securities for investors” and, therefore, the economy becomes “more dependent of the speculative actions over such securities, generating higher volatility and instability” (Rapoport & Brenta, 2010).

Corpus time distribution of the words “volatility” and “crisis”



Once the crisis was admitted and defined, a new concept substituted “volatility”. Thus, “debt” seems to be the new way of describing the crisis in a new scenario as of 2011/2012 onwards.

## 5. Conclusions

During the months of September and October of 2008, several Stock Exchanges and Banks were shaken, generating a variety of reactions worldwide. According to Ignacio Ramonet (2009), these events marked for capitalism the ending of an era, while George Soros (2012, p.31) stated that it was the beginning of a new paradigm that should consider the then preexisting one as “profoundly erroneous”. Nevertheless, the situation was far from being new. Back in 2002, Joseph Stiglitz (p. 18) pointed out that transparency related problems were the basis of the confidence crisis that had been affecting the IMF since the global financial crisis of 1997-1998.

In that context it was considered appropriate to analyze the documents published by the IMF to enlighten, with textual data statistical analysis tools, the topic interest and focus of every annual report taken into account. As it has already been said, the underlying idea is that language is considered as a “system” and that the study of its behavior and of its features allows identifying the values of the speakers (Feldman, 2008).

The analysis of documents issued within crisis contexts, both in connection with internal or institutional crisis (related to power and legitimacy) or external ones, as is the case of the still ongoing global financial crisis starting in 2007, allows understanding the strategies followed by the Fund in order to define its role in such context.

Messages ought to be analyzed, but also the messenger, that is to say, the IMF, which is a “public institution, established with funds coming from taxpayers spread all over the world” (Stiglitz, 2002, p.49) that work together with the Ministries of Finance and Central Banks assuming its control “through a complicated voting system mainly based on the economical power of the countries” which appeared at the end of the Second World War, and with only one country having the privilege of executing an “effective veto” (the United States of America). One message, one messenger, one goal and one mission have failed, because the IMF did not fulfill its role in helping to avoid the occurrence of another global economic crisis comparable to the Great Depression of the 30s.

The analyzed documents allow establishing some clues concerning the motivations that lie behind such failure. Due to the contradictions and the denying of a reality that was giving clear messages in



connection with the depth of the upcoming crisis, enables us to notice that within the IMF so as in other multilateral organizations exists a problem related with their governability.

But beyond who is in charge of the IMF, its speech is contradictory, ambiguous and focused to achieve communicational and propaganda-like objectives. The Introduction and the Executive Brief are written to be read by market players and its defendants, including the specialized hegemonic press. Reading the Reports from a paralinguistic dimension as proposed in this paper has enabled the stripping of intriguing aspects related to such behavior.

The distribution analysis of some statistically significant words allows the understanding of the speeches' axis over the years, strongly related to the political objectives of the Institution and it also enables us to perceive how the Fund tried first to hide the crisis, then to underestimate it and finally to declare the end of the crisis while it still persists.

## References

- BOLASCO, S. (2005). *Statística Testuale e Text Mining: alcuni paradigmi applicativi*. Quaderni di Statistica, Vol.7.
- BOLASCO S. (2013). *L'analisi automatica dei testi. Fare ricerca con il text mining*. Carocci Editore, Roma.hb
- BOLASCO S., D'Avino E., Pavone P. (2007) *Analisi dei diari giornalieri con strumenti di statistica testuale e text mining*, in *I tempi della vita quotidiana. Un approccio multidisciplinare all'analisi dell'uso del tempo*. Roma, ISTAT.
- FELDMAN, A. (1995), "Il concetto di sviluppo umano secondo le Nazioni Unite: analisi del contenuto" in Bolasco, S., Lebart, L., Salem, A. (eds.). *JADT 1995 - Analisi statistica dei dati testuali*, Roma, CISU, 2 voll.
- FELDMAN, A. (2000), "Introducción al Análisis de Datos Textuales", en *Cuadernos de Análisis Multidimensional de Datos n.1*, Universidad Nacional de Tres de Febrero, diciembre 2000
- FELDMAN, A. (2002), *La integración del Análisis Textual a los estudios de Calidad de Servicio y Satisfacción del Cliente*, en A.Morin & P. Sébillot, Ed., *6èmes Journées internationales d'Analyse statistique des Données Textuelles*, volumen 1, Saint Malo, 13-15 de marzo 2002.
- FELDMAN, A. (2003), *L'uso dell'Analisi Statistica dei Dati Testuali in una ricerca sistematica*, *Sociologia e Ricerca Sociale n.70*, 2003 nuova serie, Franco Angeli, Roma (Italia)
- FELDMAN, A. (2005), "Una Herramienta para la Escucha. Detección de Necesidades y Cambio de Actitudes a Través del Análisis de la "Voz Del Cliente", presentada en el I Congreso de la Sociedad Argentina de Investigadores de Marketing y Opinión Pública, Buenos Aires, 21 y 22 de abril de 2005
- FELDMAN, A. (2007) "Argentina – Italia. Mamá = Mamma? Mujeres, Vida Cotidiana y Participación Política", presentada en el II Congreso de la Sociedad Argentina de Investigadores de Marketing y Opinión Pública, Buenos Aires, 10 y 11 de abril de 2007.
- LEBART, L. (1985/86). *Qui pense quoi? Evolution et structure des opinions en Frances de 1978 à 1984*. *Consommation n. 4*, Paris.
- LEBART, L. et Salem, A. (1994). *Statistique Textuelle*, Paris: Dunod.
- NEMIÑA, Pablo. (2009) *Aportes para un esquema de análisis del comportamiento del FMI en crisis financieras a partir de su actuación durante la crisis argentina (2001-2002)*. DOCUMENTOS DE INVESTIGACIÓN SOCIAL NÚMERO 8. ISSN 1851-8788. IDAES. UNSAM.
- RAMONET, Ignacio (2009). *La Catástrofe Perfecta. Crisis del Siglo y Refundación del Porvenir*. Claves del Siglo XXI. Capital Intelectual. Buenos Aires
- SEN, Amartya y Kliksberg, Bernardo. 2007. *Primero la Gente*. Ediciones Deusto. 9na edición Editorial Temas, 2011.
- SOROS, George (2012). *Financial Turmoil in Europe and the United States*. Versión en francés *Le Chaos Financier Mondial. La crise de la dette, la crise de l'euro: comment s'en sortir*. Presses de la Cité. Paris.
- STIGLITZ, Joseph (2002). *Globalization and Its Discontents*, Stiglitz, J.E., W.W. Norton & Company, ISBN 978-0393051247
- WEISBROT, M., CORDERO, J. y SANDOVAL, L.. April 2009. *Empowering the IMF: Should Reform be a Requirement for Increasing the Fund's Resources?* Center for Economic and Policy Research. Washington, D.C. 2009. [www.cepr.net](http://www.cepr.net)