

Scientific
Journal of
**Applied
Social and
Clinical
Science**

**CONTRIBUTIONS OF
LEXICOMETRY TO
PERCEPTIONS OF THE
ENERGY TRANSITION**

Marcela F. Albornoz

Juan M. Paris

All content in this magazine is licensed under a Creative Commons Attribution License. Attribution-Non-Commercial-Non-Derivatives 4.0 International (CC BY-NC-ND 4.0).



Abstract: Within the framework of a research project, belonging to a University, which addresses problems related to the hydrocarbons industry, a survey is carried out during the month of September 2021 to investigate the perception of citizens regarding different issues related to the industry. The energy transition emerges as a solution to climate change that is commonly associated with the burning of fossil fuels. This paper aims to investigate through vocabulary how these ideas are installed in people. The answers to the open questions are analyzed: “Explain in your own words what you understand by energy transition”, “What is your opinion about the energy transition?” and “Who must lead this change that leads to the energy transition?”, using statistical analysis of textual data. These questions are part of a larger survey. The use of textual statistics in the analysis of answers to open questions provides information about what the investigated subjects think since the ideas are manifested through vocabulary. Additionally, this type of analysis makes it possible to differentiate and group subjects, establishing categories that do without the subjectivity of the researcher. The R language is used, complementing it with the SPSS Statistics software. 188 surveys are analyzed in order to identify lexical differences and similarities between age groups. Using Lexicometry, the words are contextualized by building contingency tables of the different groups. These characteristics identified in the lexicon are studied by Correspondence Factor Analysis applied to the added lexical table. During this process, responses are not modified or categorized. This work makes it possible to determine which words are installed in people’s minds regarding the energy transition and corroborates that there is a lexical difference in perceptions between different age groups.

Keywords: Lexicometry; Energy Transition;

Open Questions.

INTRODUCTION AND GOAL

Within the framework of a research project that addresses problems related to the hydrocarbons industry, belonging to the National University of Comahue, a survey is carried out during the month of September 2021 to explore the perceptions of citizens regarding different related topics. with the industry.

One of the issues addressed is the energy transition. Since the energy transition emerges as a solution or mitigation to climate change, which is commonly associated with the burning of fossil fuels, the issue is directly related to the hydrocarbons industry.

To investigate the ideas, judgments and perceptions of the population, there are various possibilities. They can be through the implementation of surveys with closed questions, of simple interpretation, or through open questions. The most appropriate method of inquiry are surveys with open questions or semi-structured interviews, since what the individuals surveyed express with respect to a certain topic is a reflection of the ideas or conceptions underlying them.

The objective of this study is to investigate through vocabulary how these ideas are installed in people with respect to the energy transition. It is worth mentioning that it is not about analyzing the veracity of the answers, but about visualizing the different perceptions and points of view. It is not the objective of this study to analyze the energy transition “per se”, for which reason technical issues related to it will not be delved into.

For this, multivariate descriptive statistics techniques are used, more specifically those developed for the treatment of qualitative or categorical variables that constitute a basic tool for the analysis of textual data or lexicometry.

The statistical analysis of textual data or

lexicometry (Lebart and Salem, 1988,1994) is an area of statistics that is developed due to the need to have a tool to analyze the open questions made in surveys. Lexicometry tries to remove the subjective gaze of the researcher (Becue, et al., 1995) and analyze the textual data after various codings in order to obtain information on the frequency of words, the context in which they are found, the frequency of said contexts, richness of vocabulary, etc. These frequencies are subsequently analyzed using multivariate statistical techniques.

The development of these techniques has made the statistical analysis of texts become an interdisciplinary tool, integrated by statistics, discourse analysis, linguistics, computer science, survey processing, documentary research and is increasingly used in various fields of the social sciences: history, politics, economics, sociology, psychology, among others (Albornoz, 2018).

The use of Textual Statistics or Lexicometry (Bécue, 1991; Lebart and Salem, 1994) in these investigations, can facilitate the knowledge of the ideas that the investigated subjects possess; since they are manifested in vocabulary management, specifically in the predominant use of certain words and in the frequencies of their use (Baccalá and de la Cruz, 1995, 2000).

Textual Statistics works with contingency tables that can be responses (individuals) for different words, which is called Lexical Table, or groupings of responses (texts) for different words, which is called Aggregate Lexical Table.

The analysis of this information is performed using simple correspondence factor analysis (SFA).

The AFC applied to the lexical tables gives a visualization of the proximities between individuals and between forms, allowing to observe which forms and/or expressions differentiate the individuals. It proceeds by comparing lexical profiles. With the use of this method, it is not a matter of knowing

what individuals say, but whether they say the same thing (Bécue, 1991).

In summary, the CFA allows to highlight the major structural features related to both sets (graphic forms and individuals or groups of individuals) analyzed, through projections on subspaces of reduced dimension, but maintaining the maximum dispersion of the original data.

METHODOLOGY

A survey is carried out during the month of September 2021 to explore the perceptions of citizens regarding different issues related to the energy transition. The answers to three open questions are analyzed in this study:

1. Explain in your own words what you understand by energy transition.
2. What is your opinion on the energy transition?
3. Who must lead this change that leads to the energy transition?

188 surveys are analyzed, using the R language version 4.1.1 (2021-08-10) and complemented with the SPSS Statistics version 25 software, for information processing.

The population is segmented into five age groups, according to the following ranks: 25 or less, 26 to 35, 36 to 45, 46 to 55, 56 or more years.

In the analysis of the corpus, the techniques of textual statistics or lexicometry are applied. The connecting words (de, y, que, etc.) and the articles (una, un, la, el) are eliminated, selecting for the study the graphic forms that by themselves have meaning, such as nouns, verbs, adverbs and adjectives. During this process, responses are not modified or categorized, thus honoring what each individual has expressed. For the words to have a significant contribution, those with a frequency greater than or equal to 7, with a length greater than or equal to 4 characters

or letters, and that are present in at least 4 documents, are considered.

To assess whether there is a lexical difference between the age groups, contingency tables are constructed, in which the different age groups occupy the rows and the words occupy the columns.

ANALYSIS OF THE RESULTS

Taking into account the objective proposed for this research, the thresholds are set: minimum number of documents, minimum number of characters and minimum frequency.

Certain words that present a natural association in the language are detected in the lexicon, that is, chains of words that are repeated in the same order. Therefore, structures such as “fossil fuels”, “energy matrix” and “I don’t know” are considered as words.

It is worth mentioning that in all the questions analyzed, the percentage of individuals who stated that they had no knowledge on the subject (“I don’t know” responses) is less than 10%. This indicates that individuals have, for the most part, knowledge about the topics covered.

For a better interpretation, the answers to the questions are analyzed individually. The results of each of them are presented below.

QUESTION 1: EXPLAIN IN YOUR OWN WORDS WHAT YOU UNDERSTAND BY ENERGETIC TRANSITION

Figure 1 shows the first factor plane of the Simple Correspondence Factor analysis applied to the lexical table. It shows that the first axis explains 42.98% of the variability and the second 25.10%.

In it, it is observed that there is no

1 For this reason, the age groups are not included in the factorial plane of figure 2.

2 Italic typography is used to indicate the words analyzed (which appear in the factorial plane).

3 Source: https://es.wikipedia.org/wiki/Transici%C3%B3n_energ%C3%A9tica

association of words with the different age groups in particular¹, which means that the lexicon is mostly shared by all groups. In Figure 1 only the words are displayed.

The words used to describe what they understand by energy transition (figure 1), are generally structured according to variants of the following discourse:

“The energy transition is to change² of an energy matrix based on the burning of fossil fuels (hydrocarbons such as oil or coal), which produce polluting emissions for the environment, through renewable, clean and sustainable energies such as wind and solar.”

It is worth mentioning that within sustainable energies only wind and solar energy are installed in the language; The remaining clean energies (hydraulic, geothermal, tidal, biomass, etc.) are not mentioned.

Although a definition has been structured using all the words analyzed, in general, the individuals surveyed use only some of the words mentioned, providing partial definitions on the subject in some cases.

Another aspect to highlight is that, if we compare the structured definition with the definition that appears in Wikipedia, both present a high degree of similarity:

“The current energy transition arises from the need for climate actions to mitigate global warming.⁵ To stay within the 1.5°C proposed by the Paris Agreement, it is necessary to stop greenhouse gas emissions by 2040 or 2050.⁶ This implies decarbonize energy systems, that is, abandon fossil fuels such as oil, natural gas, lignite and coal, and replace them with sources that do not generate or generate low emissions such as nuclear fuel (uranium) and renewable energy sources such as wind, hydroelectric, solar, geothermal, marine, wave, among others”³.

Although it is not the objective of this study

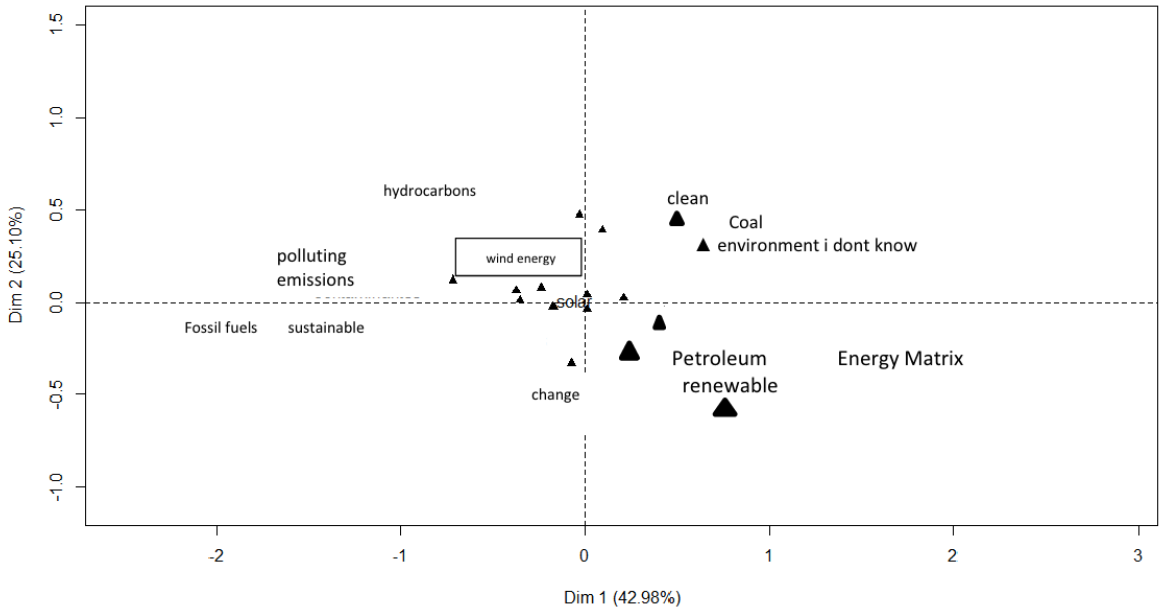


Figure 1. Factorial plane corresponding to question 1. Only the words are graphed.

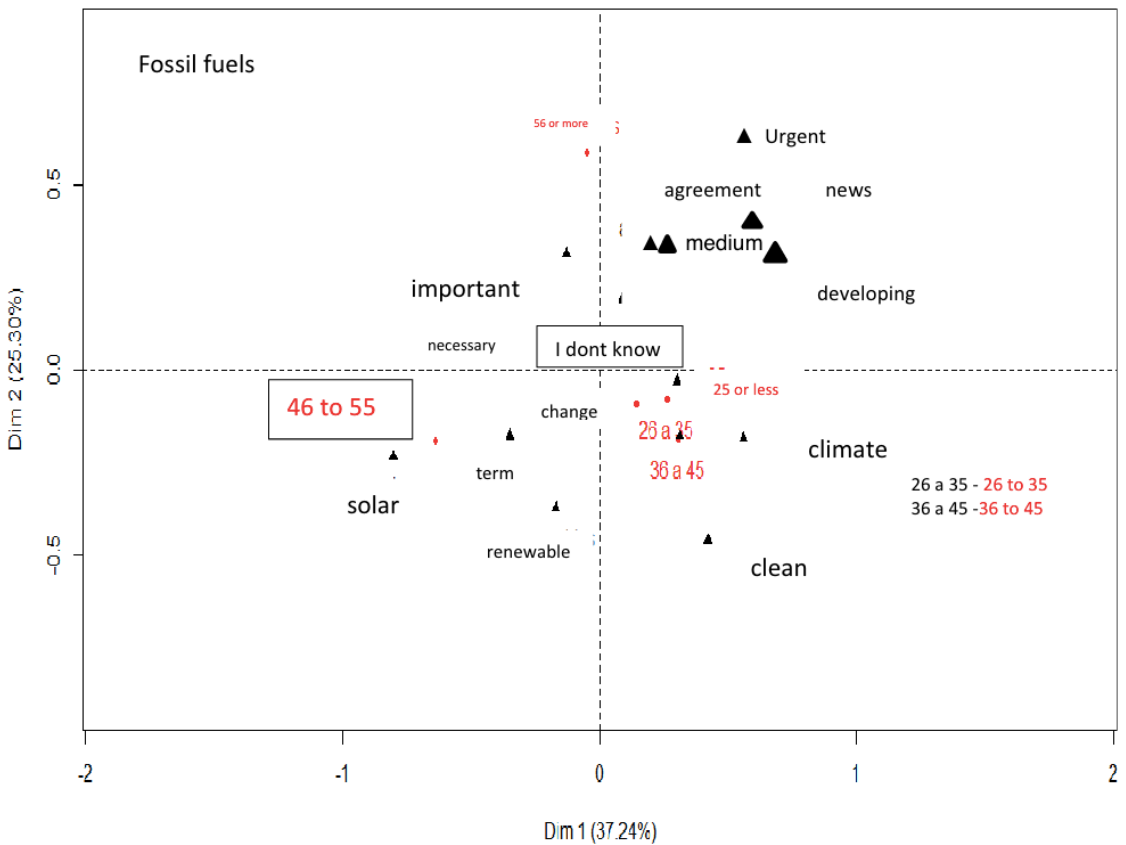


Figure 2. Factorial plane corresponding to question 2. The words are graphed in black and the age groups in red.

to evaluate the veracity of the statements of the individuals, but rather to try to visualize their perceptions, it is evident that the individuals surveyed have a perception of the meaning of energy transition that is mostly correct.

QUESTION 2: WHAT IS YOUR OPINION ON THE ENERGY TRANSITION?

Figure 2 shows the first factor plane of the Simple Correspondence Factor analysis applied to the lexical table. It shows that the first axis explains 37.24% of the variability and the second 25.30%. The first axis makes it possible to differentiate the groups of individuals under 45 years of age, from those between 46 and 55 years of age; while the second axis differentiates individuals over 56 years of age with respect to the rest of the groups.

When analyzing the factorial planes, the distances between words, the distances between age groups and the distances between words and age groups must be observed with special care.

The closeness between three age groups under 45 indicates that they share much of the lexical content, while the distance between these groups and those over 56 and those between 46 and 55 indicates that these groups use different words such as answer to the question posed. There is a vocabulary shared by the groups (words generally located in the central sector of the graph and equidistant from the age groups) and characteristic words of age groups in particular.

In the following list the words are expressed within the context in which they have been used.

Vocabulary shared by all groups:

Change from fossil energies to renewable energies / agree / don't know / implementation of necessary measures and policies / important.

Groups with individuals under 45 years of

age emphasize the development of new clean energies to stop or mitigate climate change.

Additionally, individuals between 46 and 55 years old mention solar energy as an alternative and those over 56 years old mention the decrease in the burning of fossil fuels.

One of the words most used by all the groups has been to change. Those over 56 show a tendency to consider it urgent, while the rest of the groups consider that the change towards renewable energies will take place in the long term.

QUESTION 3: WHO MUST LEAD THIS CHANGE THAT LEADS TO THE ENERGY TRANSITION?

Figure 3 shows the first factor plane of the Simple Correspondence Factor analysis applied to the lexical table. It shows that the first axis explains 50.78% of the variability and the second 33.21%.

For the youngest (25 years or less), the responsibility of leading the energy transition falls on the people, society, people and in industries.

On the other hand, for the older age group (56 years or older), the energy transition must be led by: international organizations or organizations, political leaders, public policies, and governments. These last three words mentioned (political, political and governments) are also lexicons shared by with the rest of the groups.

It is also part of the shared vocabulary: The implementation of renewable energies and companies.

The developed countries. This last vocabulary being shared by the age groups mainly of intermediate ages, but not representing an option for those aged 25 or less (note that they are far apart in the factorial plane).

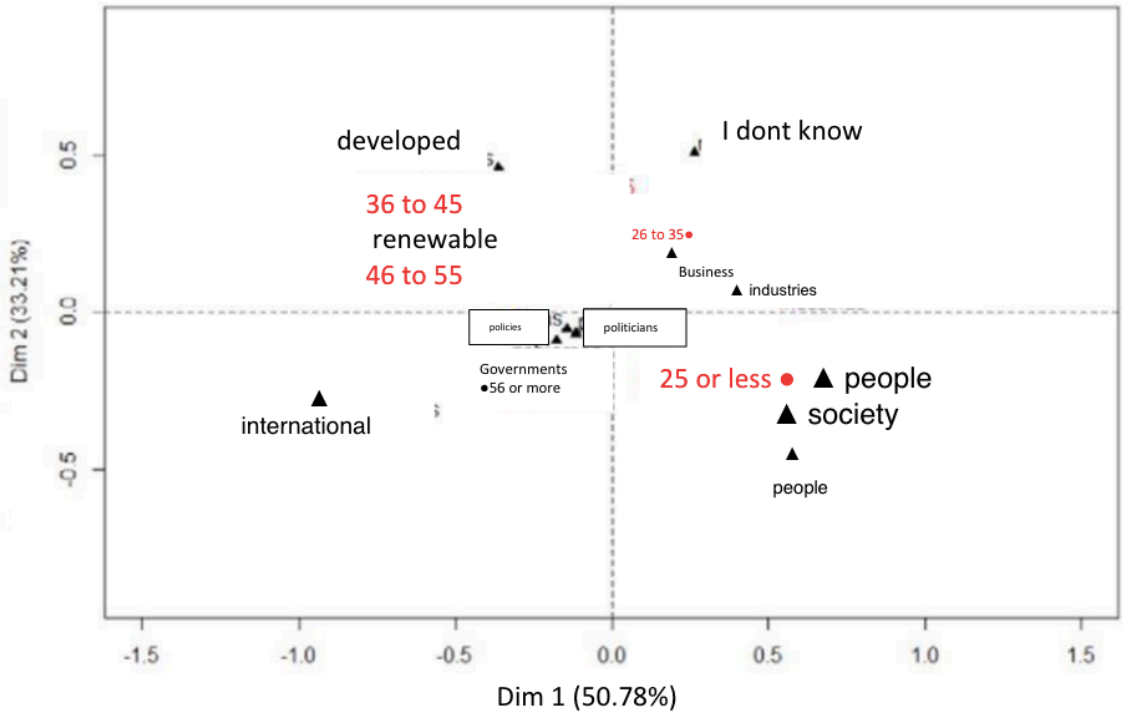


Figure 3. Factorial plane corresponding to question 3. The words are graphed in black and the age groups in red.

CONCLUSIONS

The analysis of the characteristics of the lexicon allows us to approximate what ideas individuals associate with respect to the energy transition through different age groups. The results that emerge from the analyzes show that the different groups handle a specific vocabulary, since it is possible to distinguish words that are specific to each one.

In all the questions analyzed, the percentage of individuals who stated that they did not have knowledge on the subject is less than 10%. This indicates that individuals have, for the most part, knowledge about the topics covered.

Regarding the definition that individuals provide about the energy transition, although it is not the purpose of this work to evaluate

the degree of veracity of the information, it conforms to the formal definitions that exist about it; which indicates that the common of the individuals handles correct information on the subject.

Regarding personal opinions on the energy transition, all age groups consider it important. Those over 56 show a tendency to consider it urgent, while the rest of the groups consider that the change towards renewable energies will take place in the long term.

Finally, the youngest (25 years or less) consider that the responsibility of leading the energy transition must fall on individuals, society, people and industries, while for the older group (56 years or more) the energy transition must be led by international agencies or organizations.

REFERENCES

- Albornoz, M (2018). *Aportes de la Lexicometría en Indagaciones sobre la Enseñanza de la Matemática*. Publicia, ISBN 978-620-2-43054-8. Dusseldorf, Alemania.
- Baccalá, N.; de la Cruz, M., (2000). La importancia de la estadística textual aplicada al estudio de las concepciones de enseñanza. In M. Rajman and J-C. Chappelier, editors, *Proc. Of 5es Journées internationales d'Analyse statistique des Données Textuelles*, Volume II. Copy Service Pillet, Martigny, pag. 519-524. Lausana , Suiza
- Bécue Bertaut, M. (2018). *Textual Data Science with R*. Chapman & Hall/CRC Computer Science & Data Analysis) 1st Edición, ISBN-13: 978-1138626911. Florida, EEUU.
- Bécue-Bertaut, M. (1991). *Análisis de datos textuales, Métodos estadísticos y Algoritmos*, CISIA, París
- Bécue, M.; Pagès, J. (1999). Intra-Set Multiple Factor Análisis. Application to textual data. In: *proa. of the 9th International Symposium on Applied Stochastic models and Data Analysis*, J. Jansen et al.(Eds.). Universidad de Lisboa. Lisboa, pp.51.60.
- Bécue, M.; Pagès, J. (2000). *Analyse Factorielle Multiple intra-tableux. Application à l' analyse simultanée de plusieurs questions ouvertes*. In *JADT 2000, 5ème Journées Internationales d' Analyse statistique de Donnée Textuelles*, Rajman M. et Chappelier J.C.(Eds.)EPFL, Lausanne, pp. 425-432.
- Bécue, M.; Pagès, J. (2003). *Análisis factorial múltiple para tablas de contingencia: Estudio de mortalidad en las comunidades autónomas de España*. 27 Congreso Nacional de Estadística e Investigación Operativa, Lleida, 8-11 de abril de 2003
- Kiers, H. A. L. (1991). Hierarchical relations among three-way methods. *Psychometrika*, 56: 449-470. Lébart, L.; Salem, A. (1994). *Statistique textuelles*, Dunod, Paris.