

The Bible of AI™ OpenScience • Authors' publications

State of the art of data science in Spanish language and its application in the field of AI

Celia Medina Lloret

Published on: Mar 11, 2021

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Abstract

This study analyses the state of data science and AI in Spanish language. The motivation is driven by our previous project: <https://doi.org/10.21428/39829d0b.d871fa5c> - “The art of Artificial Intelligence from a lexical perspective”. The search for information about AI in Spanish language was a failure, so it must rely on English language. This study is divided into two interlinked parts: the first one exposes what data science is, the benefits produced, and the education that a data scientist should have. Therefore, a relation of data science and AI with Spanish language. Questions are raised about how the training of data scientist in Spain is, and the analysis of diagrams, which shows the percentage of Spanish and English academic articles. Moreover, there are many contributions of members and representatives of Latin American Languages Academies. They comment on the lack of AI glossary written in Spanish language. The study of art provides results that indicate the absence of involvement of Spanish with AI and all the subareas, which consequently adversely affect to the education of future professionals.

Key words: AI, data science, Spanish language and education.

1. Introduction

1.1 What is known?

“The idea that all change should be sweet, slow and stable does not sprout from rocks. It represented a common cultural bias, that is, an answer from the liberalism of nineteenth century for a world in revolution [...]” Those words were used by the palaeontologist Stephen J. Gould to describe the term “gradualism”. This concept directly addresses the revolution of planet because of the development of TICs and the progress of new educational areas as artificial intelligence (Manuel Castell, 1999, p.1). Generally, when we refer to AI, a significant percentage of population has a wrong idea of what it is because AI is not refer to machines that do the tedious work which human being does not carry out. Artificial intelligence is much more (IT Digital Media Group, 2017). There are many branches of knowledge in AI that help us daily, for instance, data science enables us to identify the hidden relationships between variables with the aim of stablishing and creating different versions and predictive classifications (R.A. Salas Rueda, 2019) (R.D. Salas Rueda, 2019). For more information, see also the next example. The acronym “MENA” is used in Spain to denominate unaccompanied foreign minors (BBC News, 2019). However, thanks to data mining technique (an area of data

science), it has been found that: “This abbreviature evokes a concerned, criminalized and moralist discourse from an adult centric and nationalist perspective (Scientific journal of Communication and Education: Comunicar, 2021).

1.2 A bit of history

In last years, the term data science has achieved a considerable prestige in world. In 1962, people was already talking of it, when the American statistical, John W. Tukey, started to awake his interest in the future of mathematic statistic as experimental science. His contributions were focused on studying different techniques of data analyses. Some time later, his proposal rose to such heights that he published the book: “Exploratory Data Analyses” (EDA, *Exploratory Data Analyses*) (Turkey et al., 1977).

1.3 What is the purpose?

This project tends to photograph instantly state of the art of data science in artificial intelligence. The analysis is applied in several areas of study, but without losing sight of our objective: education. Nowadays, the diffusion of the English language has become in a globalised process. Thus, English language is considered a language of international communication or lingua franca (Graddol, 1997 y 2006). Globalisation is the responsible for the fragility of other languages and, specifically, Spanish language (Joaquín Guerrero, 2010). Spanish is an orphan language in terminology, use, and vocabulary to be used in AI. This adversely affects on the exploration of new techniques applied to different fields of study. That also means a precariousness in notable areas as education and, concretely, in universities.

2. Development

2.1 Data Science and Artificial Intelligence

For some time now, classic disciplines such as statistics and distributed systems have been pushed into the background in the presence of propulsion of a new and important doctrine: data science (Wil van der Aalst, 2016). Before continuing, the extension of definition of data science and its application in AI included in the introduction of this project, it should be highlighted. In 2015, the doctor Alex Liu, a pioneer data scientist, and, most importantly, very experimented in this area, he redacted a definition of data science in a publication in *IBM Corporation*: “Data science is an interdisciplinary field about processes and systems to gain knowledge or perspectives of large data volumes in many ways. They could be structured or not structured, what implies the

continuation of some of data analyses fields such as data mining and predictive analyses, as well as the discover of knowledge and data mining.”

In 2018, José Luis Marín, head of corporative *Technology Strategy*, published an article about data science, the branches that composed it and the function developed by certain specialities. Machine Learning and Deep Learning, they constitute two fields of knowledge of artificial intelligence which are of fundamental importance for data science, bearing in mind that they combine with other technologies such as: Natural Language Processing (NLP), visualization of data and experimental design. In addition, the description that he includes about the labour that *Machine Learning* and *Deep Learning* execute in this doctrine: “Both look for the composition of systems of orthographical prediction or automatic translation, autonomic cars or applied and artificial vision systems in uses cases as spectacular as the shops of AmazonGo.”

2.2 Data science: a practical approach

To be certain about the application of data science in real world, this part of the project is oriented to do a basic exploration of state of the art of data science. To carry out this undertaking, it is evaluated the sector of tourism, an area related to society, economy, and culture. As with any discipline, technology has also revolutionized tourism with Internet. This interconnected network has changed the way of registration in a hotel, mode of transportation and even, it has let us virtually move to place what we would like to visit. Definitely, “when modes of transportation evolve, technology and living conditions all over the world, tourism changes” (Jorge Bonilla, 2013, p.35).

Recently, the volume of data has experimented a considerable growth because of its application in generating knowledge and determination in making decisions. Mainly, the idea is that a large majority of enterprises and organizations design their own successful strategy to augur a brighter future (Octavio Lerena, 2019). To get it, it is necessary to foresee an expert system, that is, an intelligent information system composed by two modules: a base of knowledge and a rules interpreter. The base of knowledge contains the information requested of a concrete problem, meanwhile, the rules interpreter overcomes this problem by offering answers and orienting user to solution (Bohanec, M. et al., 1983). For further details, it is better to see how theory is applied to practice.

The project is realised by members of University Community of Colima (México). It consists of evaluating tourist destinations via data science technology. Based on those

feeling provoked by previous tourists, its main objective is proportioning information to new tourist when they choose a touristic destine (Amaya Molinar C. M et al., 2017). The first step will be selecting the factors which contribute to offer an outcome. Subsequently, through the analyses of text and social media and from the implementation of processes of learning machine as Deep Learning, needed information will be compiled to create a structure or prediction model (Itelligent, 2017). It is right here when two essential terms related to data science appear: Data Mining and Big Data. Data Mining is referring to: “By means of using various tools and algorithmics techniques, data mining looks for patterns of hidden interest with the objective of predicting future situations with a certain degree of probability (Ana M. Polo, 2016, P.3). The definition of Big Data says that: “The technology Big Data is capable of capturing, storing, operating, and processing quickly and easily large volumes of data by gaining from them. Fundamentally, it is focused on predictive analyses and identifying trends by using different techniques, for example data mining. With the definition of models and the use of diver technologies, the study promotes converting data in a valuable asset” (Thais Balagueró, 2017, párr. 10). Both procedures complement each other to register trends and find “golden data”. After gathering all possible information, the study tries to analyse valuations, attitudes, and feelings of tourists to classify the recollected data by means of a process denominated as feelings mining. Its objective is extracting subjective data by applying NLP (Ana M. Polo, 2016, 17). After following this methodology, the academic research offers positive outcomes because data science helps to identify services, establishments, and agents what, respectively, harm and favour the imagen of a tourist destine to be chosen. In general terms, the management of mass data and the development of decision support systems have provided the possibility of benefitting big quantities of data from Internet to bring a raft of benefits. As well as touristic sector, other areas such as developers outlined in data science. For instance, the system Statihouse is used to qualify, in a statistic way, property offers and, also to prognostic the price. As a result, we can underline another successful case in the application of data science technology (J.I. Pérez Rave, 2018).

2.3 Natural Language Processing in Spanish language

Fifties can be considered the point of departure for natural processing language. The stage after was marked by Alan Turing, who questioned the intelligibility of machine. As well as Noam Chomsky, who developed its generative grammatic to formalize grammatical rules (Spanish Society for Natural Language Processing, 2020). From that moment, different associations and societies were consolidated. They were focused on exploring the evolution and growth of NLP as Mexican Association for

Natural Language Processing. This entity is composed by twelve investigation groups that coordinate and share views. Some of them are the Investigation in Computation Centre (CIC-IPN) and the Institute of Investigation in Applied Mathematics and Systems (IIMAS - UNAM) (AMPLN, 2009). As with the Spanish Society for Natural Processing Language, which was created with the aim of promoting activities or giving information related to NLP: information about congresses, publications, job offers and investigation groups (SEPLN, 2021). Apart from institutions, there are some academic works that associate NLP and Spanish language. Looking for scientific databases, there are some academic works oriented to NLP. The first one is entitled "Applications of Natural Language Processing", developed by M. Beatriz, professor at National Polytechnic School of Quito (Equator) and J. M. Gómez, research professor at University of Alicante. The objective is contributing to understand texts in which NLP techniques are applied (M.B., & Gómez, J.M., 2013). Jesús Vilares, a member of the department of Computer Science and Information Technologies, highlights the presence and domine of other languages as English. Furthermore, in 2005, he writes an article entitled "Applications of Natural Language Processing", in which it is mentioned the background position of Spanish language in the study of the viability of NLP application in information retrieval systems of Spanish documents. Other studies are focused on improving present techniques such as automatic translation of texts and documents classification. This is possible due to the implementation of new tools, particularly, a syntactic and morphological analyser. This information has been gathered by the article "NLPT - Suite: Suite for Natural Language Processing in Spanish". It was prepared by many constituents of Cuban corporations, in particular Santiago of Cuba: Pattern Recognition Studies and Data Mining Centre, University of Orient. The Enterprise of Applications Development, Technologies and Systems, and Applied Linguistics Centre (Ramírez-Cruz et al., 2010).

2.4 What is the role of data scientist?

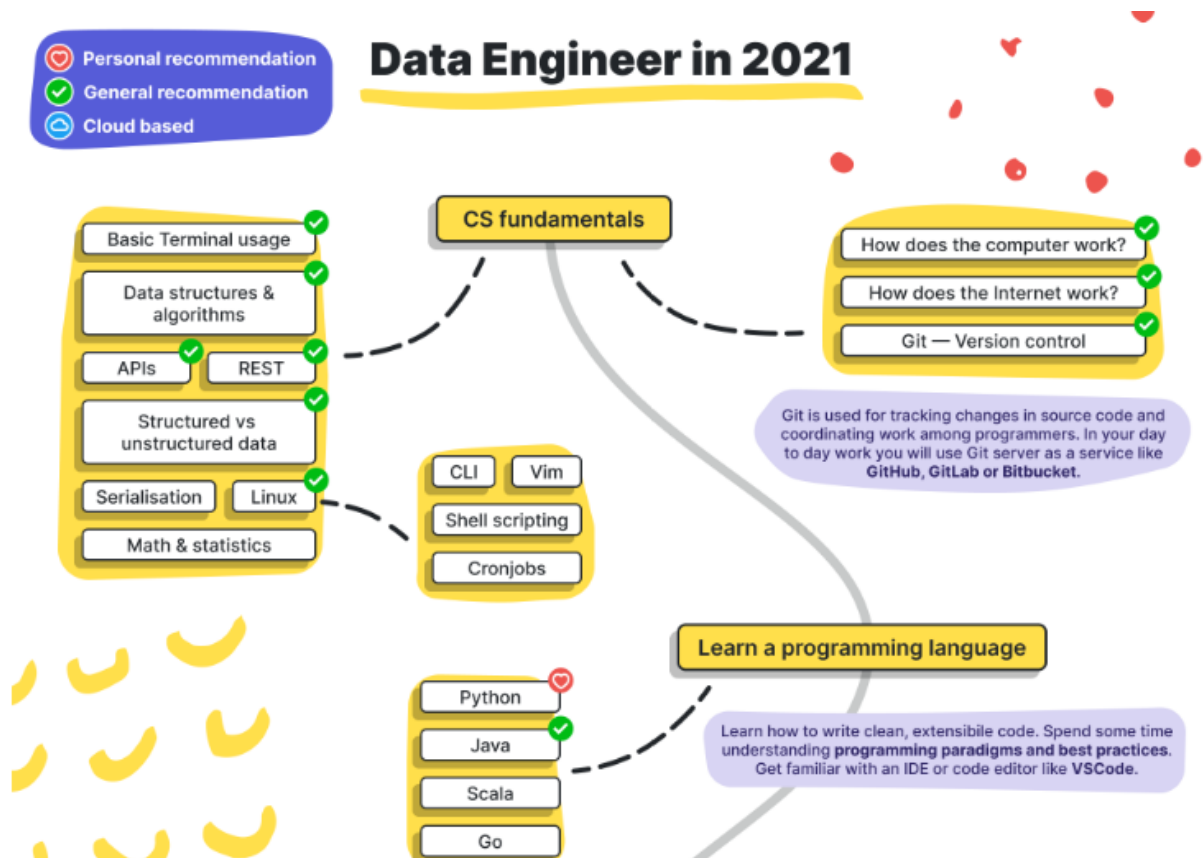
From the beginning of 2020, 44 zettabytes of data are managed in the world, that is the same thing as saying that the number of bytes in the universe was forty time greater that the number of observable stars. The problem comes with data management. That is the reason why we start talking about data scientist, who conventional wisdom identifies as: "A statistic who works in San Francisco" (Xataka, 2020).

In 2006, Jonathan Goldman began working in the social network LinkedIn. He started creating theories, new patterns, and models, and he also explored the connections

between people. Hence, J. Goldman speculated that there will be new possibilities on the horizon. One of his ideas was the suggestion between profiles, what means to incentivize people meets other people and with whom it shared knowledges, education, formation centres or abilities. In that moment, Goldman became a reference of data scientist. However, to talk about data scientist is unviable without knowing one of the most powerful tools of data science, dataset. A dataset is defined as: “group information, published or selected by an only one reference and available to access and download in several formats” (W3C Data Catalog Vocabulary, 2020). On the list that Allen Institute for AI makes public about recent datasets, we would like to underscore the following: ATOMIC 2020, an atlas of reasoning of daily common sense organized by textual descriptions. Scruples, a corpus and reference point to predict ethical trials of communities about real life anecdotes. RuleTaker, a dataset used to teach transformers to reason. And GenericsKB, a repository with a big base of knowledge about generics sentences. As we mentioned before, datasets should be always revised by a data scientist. At times, the worse enemy of data scientists is not having an appropriate dataset. Others, data are sorely lacking to get conclusions. And, on other occasions, dataset is of poor quality. To this should be added the deficiency of languages variety in datasets design because in almost all cases datasets are redacted in English. This interminable list, where newest datasets appear, is another proof of the marginalisation of Spanish language. For this reason, we need to rethink of: ¿Why do they admit the subjugation of science and universal use of data? And ¿Is manipulated, remote-controlled and condemned a data scientist?

According to Open Group industry, a data scientist cooperates with business leaders to resolve problems by means of comprehension, preparation, and data analyses. In this way, a data scientist could predict emerging trends and offer recommendation with the objective of optimizing commercial outcomes. Therefore, Open Group details certain essential abilities that a data scientist should have to augur well for the future. Some of these qualities are business acumen to understand a problem and seeking a solution. Apart from these attributes, a person dedicated to data science should deal with learning machine to manage predictions (Álvarez y Coll-Serrano, 2018).

The picture below is a roadmap, at which all of the knowledges and abilities of data scientist or engineer of data are represented. The graphic has been used with the permission of the author, Alexandra Abbas: GitHub url – <https://github.com/alexandraabbas>.



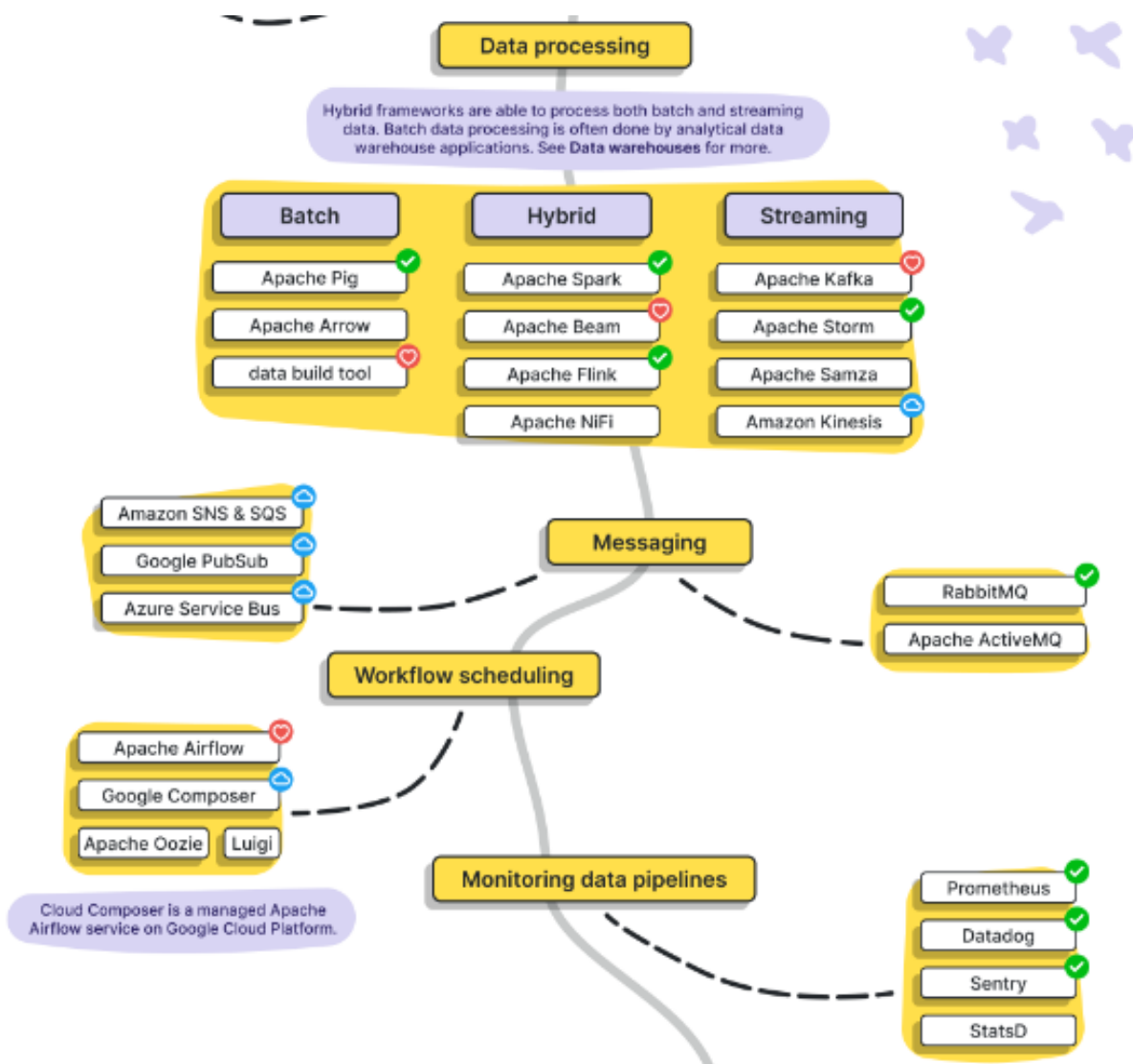




Image 1.1

Evaluating the training, education, and preparation that a data scientist should have, we have a question: What would be expected of a data scientist for his/her education in Spain?

2.5 What is the link between AI and English language?

“Globalisation is a politic phenomenon characterized by the decline of mediate instructions and direct confrontation between human being and global forces (J. Guéhenno, 2010, párr. 1). This is a historical event which is gradually marking phases because it impacts the world: politics, economy, mode, publicity, modes of transport, technologies, and language. This study is focused on analysing the repercussion that globalisation has over language. Miquel Siguán, an old member of Emeritus Free College and European Academy, pointed out that one of the consequences of globalisation is the contact between people who speaks different languages. Consequently, we have gone from monolingual society (only one language) to multilingual (more than one language).

English is one of the languages which has had more repercussion in globalisation. During the middle of the twentieth century and coinciding with the end of Second

World War (1945), English language had turned into the first language of international communication (Miguel Siguán, 2008). Last years, the development of English language has been accelerated by the growth of countries such as United States. At the same time, it has provoked the use of the Saxon language in business, films, songs, television programmes and publicities. To be sure, English language has also headed the domine of Internet, programming languages and email messages between people who vernacular language is not English (Guk Cook, 2003). In contemporary world of languages, this transformation is called “English as a Lingua Franca (ELF)” (Guy Cook, 2003). Around 1.500 million of people speak English in the world. Approximately, 575 million are native speakers. For instance, United States has not an official language, but American English language is used in regulations, legislations, and other official pronouncements (Rosa Fernández, 2020). This implies that English language is indispensable in corporate world (M. Inés, Teixeira, 2021). That is to say it is the vehicle in the development of new faculties as artificial intelligence.

2.6 Spanish language and artificial intelligence

“AI speaks English language, and we have to procure that Spanish language gains a prominent position in AI world, as well as in technological world”. These are the words that Santiago Muñoz Machado, director of RAE (Spanish Royal Academy) and president of ASALE (Spanish Royal Academies Association), pronounced in November 2019 at the 25th congress of ASALE. In addition, Spanish Royal Academy presented the project “Spanish Language and Artificial Intelligence (LEIA)”, which should accomplish two objectives: making a good use of Spanish Language in machines and operating with artificial intelligence to impulse an appropriate use of Spanish language. In this project, many agreements with technological enterprises (Telefónica, Facebook and Google) are included. These organizations promise to add Spanish language in the development of chatbots, voice assistants, and word processors to take advantage of the benefits from artificial intelligence in Spanish language (Spanish Royal Academy, 2019). As Pilar López, president of Microsoft Spain, says, these agreements are an incentive to Spanish developers, engineers, and investigators because it is an important phase for technological industry and language. At that time, Spain started getting involved with artificial intelligence by devising an investigation strategy for driving economic and social benefits of the country (Ministry of Science and Innovation, 2020).

Numerous investigation articles that circulate on Internet about AI, they are redacted in English. The following study offers data from searches, repositories, and academic

databases with the aim of comparing the numbers of publications in English and Spanish. The methodology consists of choosing those portals of Internet used for writing this paper. Furthermore, we have into account other searchers focused on academic web resources such as BASE (Bielefeld Academic Search Engine) and ERIC, a search patrocinated by the Institute of Educational Sciences of the United States Department of Education (Lenis Querales, 2017). Firstly, we tend to analyse the data offered by Google Scholar, where displays that the number of registered publications in Spanish about artificial intelligence adds up to 17.000 in a period of four years (2017-2021). However, at the same time of period, the number of English documents is 454.000 (Google Scholar, 2004). Now, we continue the study of art in Dialnet, a scientific and hispanic portal in Internet developed by the University of La Rioja (Spain). This database provides 5.632 documents in Spanish versus 2.224 in English (Dialnet, 2001). Moreover, Scielo, a scientific and electronic library, has an existence of 480 articles in Spanish and 569 in English language (Scielo, 1997). In one of the most recognised scientific Internet portals in world, WorldWideScience, there exists a difference of 898 documents between English and Spanish papers. The Germanic language leads with a repertory of 2.7491 publications (WorldWideScience, 1997). Another academic reference included in the study is Academia.edu. This database was created for academics and it has format of social network. According to Academia.edu, English language hosts 5.000 writings about AI in comparison with Spanish language, with just 60 publications (Academia.edu, 2008). BASE is an academic searcher that test the connection between English and Spanish language with AI. This database hosts 620.724 articles in English and 22.704 in Spanish language (BASE, 2009). To conclude with data collecting, we show the registered data in the bibliographic database ERIC. ERIC recognises more than 2.700 articles written in English language about artificial intelligence and anyone in Spanish language (ERIC, 1993). According to this information, we show these diagrams:

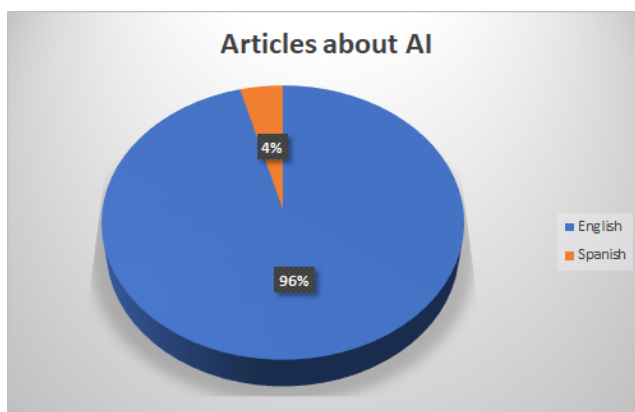


Image 1.2

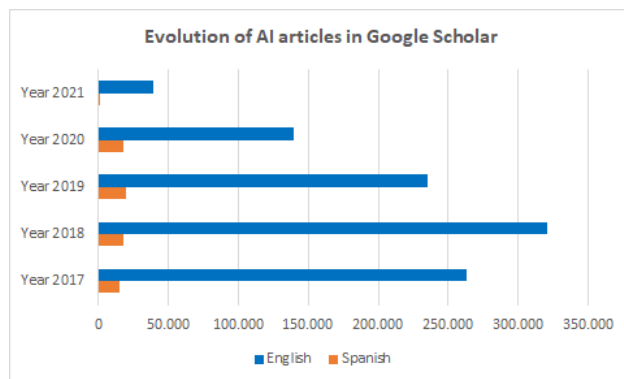


Imagen 1.3

The image 1.2 is a pie chart that represents the percentage of scientific articles about artificial intelligence in Spanish and English language. In the case of Spanish language, publications account for only 4 per cent of the graphic. In contrast to the Saxon language, which accounts 96 per cent of the totality. The bar diagram of the graphics 1.3 shows the progression and decrement of artificial intelligence articles written in Spanish and English language for four years. As can be seen, the predominancy of English language is extraordinary about the decadency of Spanish language, which accounts below 50.000 articles. What demonstrates that there is this abysmal difference between two languages in AI? Clearly, this issue has to be with globalisation of English language. This was stated by two licenses and professors of English language, A. Estrada and V. García in their essay: “Language and Globalisation: ¿A new term for an advantage phenomenon?” When the expression “rector language” started to use in order to refer to those language that heads several disciplines because of globalization. These results underscore the fact that major deficiencies that English language has in artificial intelligence area.

With the acquired experience with the study “The art of Artificial Intelligence from a lexical perspective” and reinforced by other searchers, we have some doubts about the existence of glossary about artificial intelligence in Spanish or Latin American linguistic works. To resolve this question mark, we consult external resources, mainly in Latin American Language Academies. Mr. Gonzalo Ortiz Crespo, a member of the Ecuadorian Academy of Language responded to this question by saying that: “This issue should not surprise because artificial intelligence is an area relatively new in computation science. As well as the inventions associated with the use of AI are under the swain of developing countries. Moreover, the language which describe the inventions is a matter of beginners because it is a technic language not commonly used, so it is difficult that this type of language can transcend to linguistic area.” Maia Sherwood, representative of Puerto Rican Academy of Language, defends that

technological innovation emerges from English speakers countries and therefore the first name of artificial intelligence terms has to be written in English language. In some ways, she blames the Spanish language variations, and she raises the use of the Saxon language as lingua franca. Moreover, Sherwood gives a solution by creating a terminological bank as a reference for Spanish speakers. By this way, population consults it, and it can create its own proposals. From Colombian Academy of Language, they explain that Spanish Language Dictionary is the only one linguistic work in which expressions about artificial intelligence could be included. Although being a written composition that compiles general vocabulary of all Spanish speaking countries, incorporating technicisms of various sciences is impossible. Furthermore, it is important to add the lack of resources that universities have to AI education, as well as other disciplines which are related to this area. This topic is also examined in detail later by focusing on attention to Spanish universities.

2.7 How does the lack of resources affect to Spanish universities?

On 17 February 2020, in Spain it was celebrated an artificial intelligence journey on universities. “Those countries, that have not a positioning in relation to AI, they are set to expand themselves much less, lose competitiveness and destroy jobs. I am sure that universities technological culture can be strengthened to potentiate training in the stem careers – science, technology, engineer, and mathematics. This lets the development of AI and other disruptive technologies, as well as its application in all scientific and disciplinary ambits.” These were the declarations of the president of 1millionBot, Andrés Pedreño. To this can be added, declarations of more professionals as Senén Barro, professor of computing sciences and AI in the University of Santiago of Compostela: “We are not educating an enough number of professionals in intelligent technologies in Spain, we are not doing it, demands are superior to what we are addressing. If we do not educate professors in a different ways and making a reliable transformation in education, it is impossible to get it. The main problem is teachers training, who will continue to train more students”. As Pedro Miguel Ruiz, Vice rector of strategy and digital university of the University of Murcia: “Technology of AI should be used to create a competitive university”. Experience tells us the stagnated growth of Spanish universities and AI. Even worst, the instability of AI and Spanish language.

3. Conclusions

Data science is a dense and extensive area in AI field. It brings advantages in terms of precision. The education of a data scientist should be based on statistic knowledge, data bases, visualization of data and programming languages. A data scientist must

acquire equal competences in an English speaking country that in a Spanish one. In this project, one of the problems is the scantiness of information in Spanish language, not only in data science, but also in AI. In fact, the diagrams reflect an insufficiency growth of articles written in Spanish about AI. According to database Google Scholar and comparing 2017 year with 2018, it was observed an increasement of 15,38% the documents found in Spanish. Meanwhile, English documents experimented a growth of 22,05% in the same period. To this should be added the absence of competitiveness in technology that exists in Spanish universities, as well as the scarcity of Spanish linguistics work of an AI glossary.

The evidence of this study demonstrates that Spanish is not a completely language developed in the field of AI. As a consequence, some professions as a data scientist and any other in this area are destined to disappear due to a low growth. The main cause is not to act to get a decent education, plentiful of future and innovation.

From this research, we motivate to future academic projects to continue giving voice and thinking about this issue. One of the reason is because Spanish language is a Romanic dialect language that is spoken by million of people in the world. It is also a way of promoting the educational system of Spanish-speaking countries. Additionally, the students who want to dedicate themselves professionally to data science, Big Data or AI, they can do it in their own language without English as the only linguistic support.

4. Acknowledgments to collaborators

We thank organizations and personalities that make possible this project by collaborating on updated information and testimonies which enhance the investigation value.

D. ^ª Alexandra Abbas

[linkedin.com/in/alexandraabbas](https://www.linkedin.com/in/alexandraabbas)

<https://datastack.tv/>

D. Gonzalo Ortiz Crespo

[linkedin.com/in/gonzalo-ortiz-crespo-b495a81b](https://www.linkedin.com/in/gonzalo-ortiz-crespo-b495a81b)

<http://www.academiaecuatorianadelalengua.org/#>

D. ^ª Maia Sherwood

[linkedin.com/in/maia-sherwood-droz-6687061](https://www.linkedin.com/in/maia-sherwood-droz-6687061)

<https://www.academiapr.org/>

Miembros de la Academia de la Lengua de Colombia

<https://www.academiacolombianadelalengua.co/>

5. Bibliography

Abbas, A. (2021, 15 de enero). Roadmap to becoming a data engineer in 2021.

Recuperado de: <https://github.com/datastacktv/data-engineer-roadmap>

Ahumada Polo, A.M. Minería de datos, de textos y sentimientos, Instituto Tecnológico de Orizaba.

Alex, A. (2019). *Data Science and Data Scientist*. IBM Analytics, IBM Corporation.

Allen Institute for AI. *Datasets*. <https://allenai.org/data>

Asociación Mexicana para el Procesamiento del Lenguaje Natural. (2021). *Cicling*. <https://www.cicling.org/ampln/>

Balagueró, T. (01 de noviembre de 2017). Qué es la minería de datos en big data. *Blog de Empresa y Nuevas Tecnologías*. <https://www.deustoformacion.com/blog/gestion-empresas/que-es-mineria-datos-big-data>

Bohanec, M., Bratko, I. y Rajkovic, V. (1990). DEX: An expert system shell for decision support. *Sistemica* 1.1, 145-157.

Bonilla, J. (2013). Nuevas tendencias del turismo y las tecnologías de información y las comunicaciones. *Turismo y Sociedad*, 14, 33-45. <http://www.redalyc.org/articulo.oa?id=576261184003>

Buzai, G., Baxendale, C. Análisis Exploratorio de Datos Espaciales. *Geografía y Sistemas de Información Geográfica (GEOSIG)*. <http://ri.unlu.edu.ar/xmlui/handle/rediunlu/702>

Castells, M. (1999). La revolución de la tecnología de la información. La era de la revolución: economía, sociedad y cultura, 1.

Coll-Serrano, V. y Álvarez-Jareño J.A. (2018). “Científico de datos”, la profesión el presente. *Métodos de Información*, 9 (16), 113-129. <http://dx.doi.org/10.5557/IIMEI9->

[N16-113129](#)

Davenport T.H. y Patil D.J. (2012). Data Scientist: The Sexiest Job of the 21s Century. *Harvard Business Review*, 1-5. [hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/pr 2/](http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/pr2/)

Emmert-Streib, F., Moutari, S., y Dehmer, M. (2016). The process of Analyzing Data is the Emergent Feature of Data Science. *Frontiers in Genetics*. 7:12. <https://www.frontiersin.org/articles/10.3389/fgene.2016.00012/full>

Fernández, R. (2020). Los idiomas más hablados en el mundo en 2020. *Statista*. <https://es.statista.com/estadisticas/635631/los-idiommas-mas-hablados-en-el-mundo/>

García, J., Molina, J.M, Berlanga, A., Patricio, M.A, Bustamante, A.L, y Padilla, W.R. (2018). *Ciencia de Datos. Técnicas Analíticas y Aprendizaje Automático en un Enfoque Práctico*. Publicaciones Altaria.

Garrido, J. (2010). Lengua y globalización: inglés global y español pluricéntrico *Historia y Comunicación Social*, 15, 63-95. <https://dx.doi.org/10.5209/HICS>

Gelbukh, A. (2010). Procesamiento del Lenguaje Natural y sus Aplicaciones. *Komputer Sapiens*, 1, 6-32.

Gómez-Quintero, J.D, Aguerri, J.C, Gimeno-Monterde, C. (2001). Representaciones mediáticas de los menores que migran solos: Los MENAS en la prensa española. *Revista Científica de Educomunicación*, 66, 95-105. <https://doi.org/10.3916/C66-2021-08>

Grupo de trabajo de la Estrategia del Procesamiento del Lenguaje Natural 2020. (2021, 03 de febrero). Estrategia Procesamiento del Lenguaje Natural 2020. <http://www.sepln.org/actualidad/noticias/publicacion-de-la-estrategia-de-procesamiento-del-lenguaje-natural>

Guéhenno. J. (2010). The impact of globalisation on strategy. *Survival - Global Politics and Strategy*, 40, 5-19. <https://doi.org/10.1080/713660009>

Hernández, M.B., & Gómez, J.M. (2013). Aplicaciones de Procesamiento de Lenguaje Natural. *Revista Politécnica*, 32. Recuperado a partir de https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/view/32

Hoaglin, D. (2003). John W. Tukey and Data Analysis. *Statistical Science*, 18(3), 311. <http://www.jstor.org/stable/3182748>

Itelligent. (2018, 31 de mayo). Glosario de términos sobre Inteligencia Artificial, Big Data & Data Science. Big Data e Inteligencia Artificial. <https://itelligent.es/es/tag/analisis-de-sentimiento/>

Lerena, O. [Octavio Lerena – ResearchGate]. (2020, 30 de Julio). Métodos y aplicaciones de la ciencia de datos para las políticas de CTI, vol. 1 – Redes sociales, minería de textos y clustering. <https://www.researchgate.net/project/Metodos-y-aplicaciones-de-la-ciencia-de-datos-para-las-politicas-de-CTI>

Marín, J.L. (2018, 05 de abril). Ciencia de datos, machine learning y deep learning. *Innovación*. <https://datos.gob.es/es/blog/ciencia-de-datos-machine-learning-y-deep-learning>

Melara, J.R, Gómez, M.A, Asenjo, A. y Madariaga, B. (2017). ¿Dónde lleva la Inteligencia Artificial a las TIC? *It user Teach and Business*, 29, 1-3.

Ministerio de Ciencia, Innovación y Universidades y Grupo de Trabajo de Inteligencia Artificial. (2019). Estrategia Española de I + D + I en Inteligencia Artificial.

Molne Estrada, A.T. y García Benítez V. ResearchGate. (2001, Abril). Idioma y Globalización: ¿Un nuevo término para un viejo fenómeno? https://www.researchgate.net/publication/262752232_Idioma_y_globalizacion_Un_nuevo_termino_para_un_viejo_fenomeno

Pedreño, A., Oliver, N., Martín Garijo, E., Barro, S., Pascual, C., Ruiz, P.M., Piriz, S., Rouhiainen, L. y Sánchez, C. (2020, 19 de febrero). Jornada <<La Inteligencia Artificial en las universidades. Jornada exclusiva Universidades, Torrejuna OST, Alicante.

Pérez-Rave, J.I. (2018). Statihouse: desarrollo tecnológico basado en ciencia de datos para explorar estadísticamente el sector inmobiliario. *Revista chilena de ingeniería*, 27, 133-130. <http://dx.doi.org/10.4067/S0718-33052019000100113>

Ramírez-Cruz, Y., Viant Morán, R., Ríos García y J., Fernández Cairó, C [ResearchGate]. (2010, Enero). NLPT-Suite: Suite para el Procesamiento del Lenguaje Natural en español. https://www.researchgate.net/publication/337448731_NLPT-Suite_Suite_para_el_Procesamiento_del_Lenguaje_Natural_en_espanol

Real Academia Española de la Lengua. (2019, 8 de noviembre). La RAE presenta el proyecto Lengua Española e Inteligencia Artificial (LEIA) en el XVI Congreso de la ASALE. <https://www.rae.es/noticia/la-rae-presenta-el-proyecto-lengua-espanola-e-inteligencia-artificial-leia-en-el-xvi>

Siguan, M. (2017) Las lenguas y la globalización. http://www.euskara.euskadi.eus/contenidos/informacion/artik26_1_siguan_08_07/es_siguan/adjuntos/Miquel-Siguan-cas.Pdf

Teixeira, M.I. (11 enero de 2021). Las lenguas más habladas en el mundo en 2020. <https://blog.lingoda.com/es/lenguas-mas-habladas-en-el-mundo/>

The Open Group Professional Certification Program. (2018). Conformance Requirements for the Data Scientist Profession (Open CDS). https://www.academia.edu/38263720/Data_Scientist

Van der Aalst W. (2016). *Data Science in Action*. In: Process Mining. Springer, Berlin Heidelberg. https://doi.org/10.1007/978-3-662-49851-4_1

Vilares, J. (2005). Aplicaciones del procesamiento del lenguaje natural en la recuperación de información en español [Tesis doctoral, Universidad de Coruña – Departamento de Computación]. <http://hdl.handle.net/2183/5682>

W3C Recommendation. (2020, 4 de febrero). Data Catalog Vocabulary (DCAT), V.2. <https://www.w3.org/TR/vocab-dcat/#class--dataset>