



Automated sports journalism. The AnaFut case study, the bot developed by El Confidencial for writing football match reports

Periodismo deportivo automatizado. Estudio de caso de AnaFut, el bot desarrollado por El Confidencial para la escritura de crónicas de fútbol



José Luis Rojas Torrijos. PhD in Journalism from the University of Sevilla, he holds a position as professor in the Journalism II Department at the same university. Professor Rojas Torrijos is a member of the research group entitled Communication & Social Sciences (COM&SOC) as well as a lecturer in the following Master's Degree courses: Innovation in Journalism at Miguel Hernández University, Sports Journalism at Pompeu Fabra University, Communication and Sports Journalism at the European University of Madrid, Sports Journalism at University CEU San Pablo-Marca and Sports Journalism at UCAM. He is also coordinator of the collective work entitled *Periodismo deportivo de manual*, and co-author of *En Antena*, the stylebook of *Cadena SER*.
University of Sevilla, Spain
jlrojas@us.es
ORCID: 0000-0002-7390-9843



Carlos Toural Bran. PhD in Communication Sciences from the University of Santiago de Compostela (USC). Professor Toural Bran also teaches Multimedia, New Formats and Cyber Culture in the Department of Communication Sciences at the same university. He has been Vice Dean of the Faculty of Communication Sciences at USC as well as a journalist since 2006, and has participated in the research group known as *Novos Medios*, of which he is secretary of the group. He is also editor of the journal of the Spanish Association of Communication Research, and president of the Galician Association of Communication Researchers (AGACOM).
University of Santiago de Compostela, Spain
carlos.toural@usc.es
ORCID: 0000-0002-0961-3925

Received: 31/07/2019 - Accepted: 30/10/2019

Recibido: 31/07/2019 - Aceptado: 30/10/2019

Abstract:

Sports coverage has become an expanding field for automated journalism due to the peculiarities of the competitions. Sporting events have an important statistical base that fosters the use of structured datasets and helps media outlets schedule information routines because of the cyclic and repetitive nature of matches and tournaments, and consequently, of news coverage. This article is a case study of AnaFut, the first bot developed by the digital native *El Confidencial*. This is the first Spanish media outlet to use this technology for the automated writing of football match reports.

Resumen:

El periodismo automatizado ha hallado en las coberturas deportivas un ámbito propicio para su expansión, debido a la propia naturaleza de las competiciones. Estas tienen una importante sustancia estadística que favorece el manejo de datos ordenados y permite la programación de rutinas informativas, dado el carácter cíclico y repetitivo de la celebración de partidos y torneos y, por extensión, del seguimiento que hacen los medios. Este artículo se detiene en el estudio de caso de AnaFut, el bot desarrollado por el nativo digital El Confidencial, primer medio español en usar esta tecnología para la

How to cite this article:

Rojas Torrijos, J. L. and Toural Bran, C. (2019). Automated sports journalism. The AnaFut case study, the bot developed by El Confidencial for writing football match reports. *Doxa Comunicación*, 29, pp. 235-254.

<https://doi.org/10.31921/doxacom.n29a12>

This study comprises the content analysis of eighty texts published on the web, semi-structured interviews of journalists from *El Confidencial*, as well as a questionnaire carried out with a panel of five experts. Results show that the development and application of Artificial Intelligence in journalism, and particularly in the sports field, is still in the initial stages, and that news organizations from the most technologically advanced countries are leading the way in terms of innovation.

Keywords:

Automated journalism, robot journalism, data journalism, sports journalism, journalistic match report.

escritura automática de crónicas deportivas. A través del análisis de contenido de ochenta crónicas publicadas, entrevistas semiestructuradas a periodistas del medio y cuestionarios a un panel de cinco expertos, los resultados indican que el desarrollo y la aplicación de la Inteligencia Artificial al periodismo, y particularmente al periodismo deportivo, se encuentra en una fase inicial donde las organizaciones mediáticas de los países más desarrollados tecnológicamente lideran la prospección en términos de innovación.

Palabras clave:

Periodismo automatizado, periodismo robot, periodismo de datos, periodismo deportivo, crónica periodística.

1. Introduction: automated journalism and newsroom automation

The relationship between information, journalists, newsrooms, media, algorithms, bots and Artificial Intelligence has been defined in various ways by many authors in recent years. They all have one common denominator, which is the undeniable relationship between the worlds of “computing, social science and communication” (Flew et al., 2012). For Carlson (2014), these are “algorithmic processes that convert data into informative, narrative texts with little or no human intervention other than initial programming”. Graefe (2016), on the other hand, defines the phenomenon as a “process of using software or algorithms to generate news automatically without human intervention after the initial programming of the algorithm”.

Thus, on the basis of these definitions, we also encounter different nomenclatures, ranging from automated journalism (Graefe, 2016), to algorithmic journalism (Dörr, 2016), to robotic journalism (Oremus, 2015). All of these are the result of a technological process that began at the end of the 1980s of the twentieth century (Túñez López, Toural Bran, Cacheiro Requeijo, 2018), which focused on the production of journalistic content, analysis and visualisation (Carlson, 2014; Gao, Hullman, Adar, Hecht and Diakopoulos, 2014; Diakopoulos, 2014; Young and Hermida, 2014; Broussard, 2015; Cervera, 2017).

In addition to the tasks resulting from the definition itself, for some authors (Lokot and Diakopoulos, 2016), the main tasks of bots have to do with participating in dissemination on social platforms, supporting and managing broadcasts, and adding web content (Starbird, Leysia, Hughes and Vieweg, 2010; Mittal and Kumaraguru, 2014), while at the same time being capable of identifying facts and events of journalistic relevance for later diffusion (Steiner, 2014).

1.1. State of the issue in the academic field

In the academic field, studies on the application of Artificial Intelligence for the automated creation of news have increased in the last 15 years both in the quantity and quality of the research, and in the number of deliberations on the phenomenon (Slater y Rouner, 2002; Powers, 2012; Levy, 2012; Karlsen y Stavelin, 2013; Matsumoto, Nakayama, Harada y Kuniyoshi, 2007; Napoli, 2012; Van Dalen, 2012; Clerwall, 2014; Edge, 2014; Latar, 2014; Carlson, 2014; Oremus, 2015;

Lecompte, 2015; Dörr, 2016; Graefe et. al, 2016; Fanta, 2017; Hansen, Roca-Sales, Keegan y King, 2017; Lindén, 2017; Marconi y Siegman, 2017; Usher, 2017; Renó y Renó 2017; Salazar, 2018; Wölkeer y Powell, 2018). This quantitative and qualitative increase in research has helped place the focus on automated content generation in the journalistic field as an essential phenomenon in terms of information production and consumption by audiences.

Graefe (2016a) and Dörr (2016) were among the first to develop a list of media that were already using automated methods for content production. On the other hand, it was Fanta (2017) who focused on news agencies in order to identify which of them produced content using algorithms and tools designed to produce information in an automated way.

Other researchers focused on the behaviour of the media when confronted with content automation (Lindén, 2017), as well as on the benefits of customizing local information through the application of data automation tools (Lecompte, 2015), the audience's perception of the informative content produced by applying Artificial Intelligence methods (Graefe, 2016b), and on the response of different audience segments (divided by age group) to texts produced by humans and robots (Slater and Rouner, 2002).

The question of the journalistic profession and how it relates to the phenomenon of automation has also been addressed (Túñez López, Toural Bran and Cacheiro Requeijo; 2018). For their part, Thurman, Dörr and Kunert (2017) have conducted interviews with journalists from the BBC, CNN, Thomson Reuters and others with the aim of finding out their opinion on some journalistic articles produced using automation techniques. Meanwhile, one of the first to address the issue of the relationship between journalists and automation was Van Dalen (2012), who analysed how several journalists reacted to the launch of StatSheet, a network of information websites specialising in sports topics written entirely by machines. Within this new context, the journalistic profession would reach, for some authors, a state in which the collaborative aspect would become much more important due to the liberation of journalists from certain assignments thanks to the application of automation methods and the use of work synergies (Clerwall, 2014).

Artificial intelligence (AI) is now an indisputable reality in newsrooms. Within the growing application of this technology in journalistic production, the technological development and application of bots and algorithms to the automatic writing of informative texts stand out.

2. Automated sports journalism

Apart from the world of finance, the so-called “robot journalism”, or “automated journalism”, has found the area of sports coverage to be a more suitable field for its development. This has been the case ever since major news agencies began producing teletypes from generated data and algorithms to broaden their coverage of different competitions. Some of these agencies have included AP and AFP in 2014, and Reuters in 2015.

Thus, news agencies paved the way for the rest of the important media organizations, which soon began to apply automated tasks to their coverage in order to reach more sites and audiences, and to respond to breaking news as dynamically as possible”.

To this day, more than twenty major media outlets in several countries use bots to expand, streamline and diversify their coverage of sporting events (Rojas Torrijos, 2019). Some apply them for continuous and last-minute tracking of results on Twitter (The Washington Post), others for the generation of graphics to enrich their live events (The Telegraph), while an increasing number of media organizations decide to use them to improve their local coverage (the MittMedia group in Sweden, or the Press Association agency for local press groups in the UK).

The spread of this technology in the field of sports is directly related to the nature of the competitions themselves. These have an important statistical base that favours the management of organised data and allows for the programming of informative routines given the cyclical and repetitive nature of matches and competitions and their monitoring in the media.

This increase in the use of automated information-writing bots has generated a debate within the profession in recent years. The debate has to do with setting boundaries between the functions of machines and people, as well as the consequences of their use in a journalistic world that were more agile, diverse, and of particularly higher quality.

So far, many of the experiments developed by benchmark international media organizations have shown that bots can be useful to the industry in covering more topics, expanding their audience, and enriching coverage with data that is more abundant and faster, and if possible, in real time through new digital and mobile platforms.

3. Case study: AnaFut, the new editor of *El Confidencial*

Bearing this context in mind, the present article addresses the case study of the first technology of this type developed by a Spanish media company for the automatic writing of sports reports. The case involves a bot nicknamed AnaFut, developed by the digital native *El Confidencial* through its journalistic innovation laboratory (*El Confidencial Lab*) in September 2017, and this media began using it in early 2018 to produce football match reports from the Second Division B and Third Division leagues on a national level.

3.1. Objectives and hypothesis

On one hand, this study analyses the structure and content of these automatically produced texts in order to extract quantitative and qualitative measurements on the type of language that is naturally-produced, and on the other hand, its consequences on product quality within the context of the Sports section of this media company.

The research is based on the hypothesis that the routine and statistical nature of football narratives favours the application of artificial intelligence in this area of information over others, and that the automation of sports reports is the result of the refinement in the programming of a series of pre-established editorial linguistic models that are generated from data and repeated without the need for human intervention once the texts have been published.

Therefore, the nature of the reports automatically generated by this bot causes the resulting text to have the tendency to reflect a series of linguistic repetitions, especially in the use of verbs and expressions when it comes to situations of victory, defeat or a tie in the course of a match, but which are amendable and avoidable from the moment the human editor prepares the machine.

Thus, the first sub-hypothesis should point out that these types of reports, in spite of working as a kind of template that predisposes the structure, order and length of the headlines and paragraphs in the body of the text, are improved by programming through the introduction of synonyms and second references, as well as statistical context data, in order for the reports to be more informative and attractive.

Moreover, as a second sub-hypothesis, these automated reports are supported by data and supposedly unquestionable facts, which makes it difficult to find subjective elements in them. When these data appear, they reflect the interpretation that the journalist-editor who has entered the data into the programming may have construed of situations that usually occur in a football match. Moreover, these particular 'readings' of the game, due to linguistic habits acquired among sports journalists, at times may not correspond exactly to what happened on the field.

Taking into account the starting point of the hypothesis and the sub-hypotheses of this case study, the objectives of this research are as follows:

- Analyse the structure and content of the sports reports generated automatically by the virtual editor AnaFut in *El Confidencial*.
- Obtain quantitative and qualitative measurements on the type of naturally-generated language produced, paying special attention to the degree of repetition existing in formulas referring to situations of victory, tie or defeat in the course of the matches reported.
- Examine the use of synonyms and second references in texts, as well as the input of statistical data that may lead to an improvement in the quality of the reports.
- Evaluate the extent to which the use of the technology can benefit this type of journalism, based on content analysis and the gathering of opinions through interviews with journalists and questionnaires given to experts, as this technology offers greater topic diversification, a wider scope of news coverage, and the removal of automatic tasks from editors so they can devote more time to reporting and research.

3.2. Sample and methodology

For this purpose, based on a sample that includes the detailed analysis of 80 automatically-generated reports published on the *El Confidencial* website during the 2018/2019 season, variables have been established that include length of the articles, length of paragraphs, types of headlines, lexicon used, etc. Moreover, the overall aim is to evaluate the degree of repetition of these texts and to consider the extent to which this technology can be improved on the basis of programming and human editorial intervention or supervision.

Interviews and questionnaires

In order to complete this analysis and give more context to the study, two other methodological techniques have been used: on one hand, semi-structured interviews with those directly responsible for the *El Confidencial Lab*, the department that developed the technology, as well as with those in charge of the media's Sports newsroom, where its application began; on the other hand, questionnaires were given to a panel of five experts in journalistic innovation.

The interviews were conducted with Alejandro Laso, director of *El Confidencial Lab*, and Víctor García, editor-in-chief of the media's Sports section. With regard to the questionnaires, they were sent to academics who specialise in innovation and new technologies in journalism from several Spanish universities. These individuals are as follows, in alphabetical order:

- Lluís Codina is full professor in the Department of Communication of Pompeu Fabra University, Director of the Support Unit for Quality and Teaching Innovation, and Coordinator of the Research Group on Digital Documentation and Interactive Communication. He is one of the leading figures in the field of digital journalism in a general sense, and especially in topics related to the semantic web, databases, journalistic documentation, and documentary information systems.
- Miguel Carvajal Prieto is a PhD lecturer in the Department of Social and Human Sciences at Miguel Hernández University in Elche, Director of the Master's Degree programme in Innovation and Journalism, and one of the most renowned experts in the study of innovation in Spain.
- Miguel Túnuez López is full professor of Organisational Communication and Communication Strategies in the Department of Communication Sciences at the University of Santiago de Compostela, Director of the Postgraduate programme in Contemporary Communication and Information, and one of the most active researchers in the field of information automation processes.
- Ramón Salaverría Aliaga is Full Professor in the Department of Journalistic Projects at the University of Navarra, Vice Dean of Research on the Communication Faculty of the University of Navarra, Sub-Director of the Centre for Internet Studies and Digital Life, and one of the most outstanding academic references in the field of cyber journalism and digital media at the national and international level.
- Xosé López García, Journalism Chair of the Department of Communication Sciences at the University of Santiago de Compostela, Director of the Novos Medios Research Group, and among those researchers with the most distinguished investigative careers in the field of digital journalism in Spain.

The standard questionnaire prepared and sent to the panel of experts consisted of the following five questions:

- What is your perception of the application of artificial intelligence to journalism?
- Speaking of this technology, what is your opinion regarding the contribution of bots to the automatic writing of news?
- To what extent can bots contribute to the broadening of sports coverage, one of the areas in which they are most widely used?
- To which topics and sports competitions do you think the media could best apply this technology?
- To what extent do you consider this technology improvable? Could this improvement possibly be achieved by editorial intervention or through supervision by the editors to give the finishing touch to the text before its final publication?

Analysis sheet

For the content analysis of this case study, an analysis sheet was drawn up containing a total of eleven aspects in order to check the degree of repetition of the bot's writing, or in other words, the extent to which the programming of that language is reflected, which is generated automatically on the basis of human input.

In this way, measurement of the report text by means of this analysis has made it possible to compare the most formal aspects to those most related to the content of the information in order to establish the degree of variety of these texts, and with it, the quality of the programming of the data introduced.

With regard to the more formal aspects, the following have all been analysed: The preparation of headlines; subheadings and pre-headings; the number of paragraphs; the structure; the order of the elements in the text; and the length of the sentences. Meanwhile, the use of statistical data and sources, as well as the use of lexicon (with special emphasis on the most heavily used verbs and adjectives), and other aspects of style that measure repetitions within the text (use of synonyms and second references, or not) have been studied in terms of content.

In this way, terms used by the bot when referring to similar situations such as victory, draw, defeat, coaches and replacements, referees and warnings, have also been evaluated. Moreover, if subjective or interpretative elements had slipped into those textual references, these would have been due to previous human intervention as well.

Therefore, the eleven items on the analysis sheet are arranged in three sections: elements of the headline, body of the text, and aspects of writing and style. These are as follows:

Elements of the headline

- Headline (includes result / does not include result)
- Subheading (includes result / does not include result; enhances headline or statistical references / does not enhance headline or statistical references)
- Pre-heading (broadens headline information / repeats headline)

Body of the text

- Length (number of paragraphs)
- Duration (estimated reading time)
- Sentences per paragraph

Aspects of writing and style

(Use of synonyms or second references in verbs and adjectives, subjective /interpretative elements)

- Ways to refer to victory
- Ways to refer to a tie
- Ways to talk about a defeat
- Ways to refer to coaches and replacements
- Ways to refer to referees and warnings

4. Results

The analysis of El Confidencial's automated football reports used a sample of eighty published texts related to all matches of the National Championship of the Second Division B League played on the 9th and 10th of February and the 2nd and 3rd of March, 2019.

This sample is considered adequate given the fact that it uses texts produced in the same way from the same source, which is a bot in this case. Furthermore, a selection of texts has been gathered on the *El Confidencial* website that were published throughout the same season involving the same protagonists, and were spread out over time for nearly a month in order to see if there had been an evolution or improvement in the technology used.

The Third Division in Spanish football consists of four groups of twenty teams each, and in the analysed reports, all of the clubs appear referenced on two occasions. In this way, when all of the teams are present, it is possible to obtain a better in-depth view of the texts and to observe if there is any type of difference or variation in the writings depending on the clubs and protagonists covered in the reports.

The first report written by *AnaFut* was published in *El Confidencial* on November 14, 2017, and as highlighted in Image 1, its debut was announced in the Sports section of this media on Twitter. In this tweet, the reader was made aware that a robot would bring all of the Second Division B and Third Division sports reports to the web

Image 1



Screenshot of the tweet from *El Confidencial Sports*. 14-11-2017.

ISSN: 1696-019X / e-ISSN: 2386-3978

4.1. Quantitative and Qualitative analysis

In a first approach within this study, it should be noted that *AnaFut's* automated reports have an established, almost unchanging structure. Taking the form of a template or mould, this structure is repeated continuously and consists of four distinct parts: pre-heading, headline, subheading, and body of the text.

The pre-heading, in white letters on a black grid, is concise (an average length of between four and eight words); the headline is somewhat longer (this occupies one line and a half on the computer screen display); and the subheading (always short, on one line). As far as the body of the text is concerned, the number of paragraphs is nearly standard, ranging from five (most often) to seven paragraphs.

Image 2



Screenshot. Report published on 09-02-2019.

The order of the paragraphs, such as the one shown in Image 2, is always the same:

- The first paragraph indicates the result, competing rivals, and location of the event, followed by background information (context), and consequences regarding the ranking of the contestants;
- The second paragraph summarizes what happened in the first half;
- The third paragraph summarizes what happened in the second half (these two paragraphs are shorter or longer depending on the number of goals scored, as the scoring player of each goal is included in the report);
- The fourth paragraph again reminds the reader of the position occupied by each team in the ranking at the end of the match (this has already been done in the first paragraph);
- The fifth paragraph informs the reader about the next day's matches and the teams involved.

Image 3



Screenshot. Report published on 03-03-2019

One or two additional paragraphs can be added to these reports. The publisher is the one who makes the decision. The initial assumption is that extending the length of the story depends on the importance of the match, but in reality the criterion of adding these paragraphs seems unclear and even random, as this occurred in only a few of the matches rather than all (12 of the 80 analysed).

Thus, paragraphs added are as follows:

- Another paragraph counts the changes made by the coaches of each team during the match, and also mentions the names of all players who participated in the match as well as the reserve players;
- One additional paragraph reports on the yellow and red cards shown by the referee and the players from each team who received the cards.
- These paragraphs are usually placed 4th and 5th in the order, respectively, so in these cases, the one referring to the classification moves to the sixth position and the one that deals with the next day's event closes the text in the seventh position.

As far as the headlines of the reports are concerned, they are informative and restrained, in many cases nearly always pointing out the contending teams, and in most cases the result. Inclusion of the result in the headline of these reports is justified because in this case, no data sheet is included, so all elements that usually appear in the data sheet must appear in the text.

Image 4



Screenshot of the headline of a report published on 03-03-2019.

As in the headline, there are hardly any subjective elements in the subheading (Table 1). The only two cases in which a certain subjectivity is apparent are those referring to a 1-0 result as a “worthy triumph”, and a 1-1 final as “a poor result” for the home team. As such, in both cases the subjective weight resides in using a descriptive adjective.

Regarding the subheadings and how they appear in combination with the headline of the reports, analysis of the texts produced by the bot reveals two models that appear most frequently: a subheading that merely repeats the headline data (the result if mentioned, team names, and even certain verbs and nouns) (Image 4); or one that points to an obvious fact (e.g., a tie is “a draw”, or “two points were ceded”, or phrases such as “they will fight for a win in the next match”).

Table 1

Combination headline-subheading		
Classification	Number of reports	% of total
Result in headline	51	63,75%
Subheading repeats result	13	16,25%
Subheading does not repeat/obviousness	31	38,75%
Subheading does not repeat/statistics	5	6,25%
Subheading does not repeat/subjectivity	2	2,5%
No result in headline	29	36,25%
Subheading with result	17	21,25%
Subheading without result/repeats headline	8	10%
Subheading without result/statistics	4	5%

Source: created by the authors

However, in 11% of the cases analysed, the subheading enhances and enriches the headline by introducing some statistical data in context (Image 5). This would therefore be an improved subheading model, which on other occasions would

also provide data referring to the name of the stadium in which the match was played or the coach of one of the teams (“the team coached by...”). On the other hand, this model would serve as a synonym or second reference to avoid further repetitions.

Image 5



Screenshot of the headline of a report published on 10-02-2019.

A defining element of the length of the reports written by *AnaFut* is the estimated reading time, which is indicated at the left margin of the screen under the signature and labels. Thus, the duration varies between 43 seconds for the shortest text and one minute for the longest. Only in three cases does the time exceed two minutes (Table 2). The standard duration is approximately one minute, which is the estimated reading time of half the reports studied.

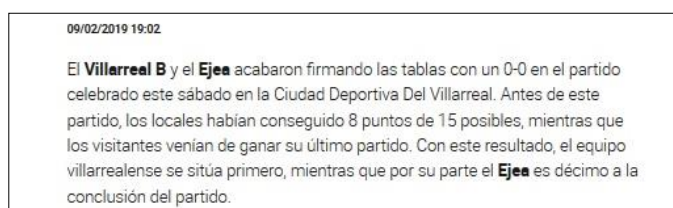
Table 2

Report length	Estimated reading time	Estimated reading time	% of total
43-50 seconds		12	15%
50-57 seconds		26	32,5%
58 seconds-1 minute		39	48,75%
More than 1 minute		3	3,75%

Source: created by the authors

As for the structure of the paragraphs, in the lead (Image 6), they usually consist of three sentences, or four at most, which are either simple or coordinated, and between one and three sentences in the rest of the paragraphs. There are hardly any subordinate sentences, and when they appear they are usually of the adverbial type introduced by “while”.

Image 6



First paragraph screenshot. Report published on 09-02-2019.

Even though they are brief, from an editing and style point of view these reports try to make reading easier for the user by highlighting in bold the names of the teams, the football players (goal scorers, replacement players, and goalkeepers when they have been the protagonists), and even the coaches.

The vocabulary used is predictably repetitive. Even so, there is an effort from the programming of the bot in the newsroom to use synonymous expressions that make the report seem less 'robotized'. In this way, in all of the eighty reports analysed, different expressions are combined in the text when referring to victory, draw and defeat, as well as to situations during the match, especially regarding the evolution and variation in the scoring.

When referring to victory, the verbs "to win", "to beat" or "to triumph" appear most often, while in draws the expressions "settle for sharing the points", "tie game", and "points shared" appear. In the case of a 0-0 result, the most common expressions are "no goals scored", "no team managed to score", or "no one was fortunate in scoring a goal". In this regard, obvious information and redundancies are sometimes reproduced, such as "neither of the teams scored a goal and therefore the scoreboard showed the same score", or "as no goals were scored, the result showed no alteration".

Moreover, there are references to the evolution of the results in the course of a match, such as: "they took the lead", "they took the lead on the scoreboard", "they managed to turn the scoreboard around", "they increased the distance on the scoreboard", "they widened the differences", "they shortened the distance", "they expanded their advantage", "they managed to come back" or "the game started in a positive way for...". Likewise, comments regarding scoring of the various goals usually begin with, "thanks to the success of..."

In addition, there are alternative ways of referring to the next day's events ("the next appointment on the calendar"), or to the replacement of players ("the coaches made all possible changes", or "both teams ran out of changes", or "both coaches used all the players on the bench"), and a phrase was even used to vary the way to say that half-time had arrived ("the players lined up at the entrance of the changing rooms").

Apart from the lexical alternatives that the editing department can introduce in order to improve the textual quality of these automated reports, another element that enriches these pieces is without doubt the incorporation of statistical data. These are becoming more and more frequent, not only in subheadings for the purpose of enhancing the headlines ("a win at home will mean a 6th league victory"), but also in the body of the text. The use of statistical data in these reports appears in the lead in order to contextualise the result in a cumulative series throughout the championship ("the locals are coming to the match with four consecutive victories", or "before the match, the locals had achieved 6 points out of 15 possible, while the visitors came to the competition after winning their last match").

Therefore, taking into account the objectives set for this research, both the quantitative and qualitative measurement of the naturally-generated language in these reports shows a significant advance in terms of the number of texts produced each day (forty), with each and every one of the teams in the category being included, and full coverage being ensured. In spite of this, all of the doubts about aspects related to the quality of the resulting information have still not been clarified.

Paying special attention to the degree of repetition in language referring to situations of victory, draw or defeat in the course of the matches covered, a clear effort can be observed in the use of synonyms and second references in the texts, as well as the progressive incorporation of statistical data that can improve the quality of the reports.

The truth is, these automated reports have allowed this digital media to expand its sports coverage by diversifying topics and extending them to other teams and categories, which otherwise would have been very difficult to accomplish through the use of human editors alone due to the fact that the priorities of information about football in everyday life are different (First Division, Champions League, and player signings).

In fact, to find out what impact this technology has had on the newsroom and to what extent it can be improved through the use of pre-programming and editing to increase the quality of information, the opinion of *El Confidencial* journalists was sought.

4.2. JFKL Interviews

AnaFut was launched and began writing reports in late 2017. The managers of *El Confidencial Lab* were responsible for explaining that this technology would deal with coverage of lower division football competitions that were not covered by journalists from the Sports section. The functions of the bot were clearly defined in relation to those of the professionals in order to facilitate their co-existence from the beginning.

This has been pointed out by Victor García, editor-in-chief of the Sports department¹, who refers to *AnaFut's* endeavour as “throw-away work”, as these are match reports, which usually have a shelf life of a couple of hours on the Internet. Therefore, these are very short texts aimed at fulfilling user interest in knowing the latest news, as close as possible to the updates provided by live match coverage that many of them watch on television and social networks. “From there, people look for other aspects of interest related to the game, and that’s where our imagination and creativity come into play to tell other kinds of stories”, says García, who talks of the importance of this technology in satisfying the need for fast information. “Here, whether it’s a bot or a journalist, it’s worth telling what happened, and later do other stories giving more detail, a conclusion, or something that can provide added value”, he adds.

In the opinion of Alejandro Laso, director of *El Confidencial Lab*, “the important thing is for editors to start understanding this technology”, because this is really about the bot “doing other work so that journalists can be journalists, and are able to produce other stories that generate organic traffic” to the web. The objective of this Sports bot was clear from the beginning, which was “to save editors time when doing their job, to cover niches and areas that were not possible due to a lack of professionals, to reach new audiences, and to make Sports people more capable of acting and disseminating”. In addition, says Laso, “robots do the dirty work that journalists usually have to do, but it’s the type of work that doesn’t let them shine”.

Although the football matches of the Second Division B have been used as a pilot experience to test the bot, the question is whether this technology can be applied to other sports. In this regard, Alejandro Laso believes “it can be applied to anything that has structured data as well as frequent, routine competitions”. He uses tennis as an example, where the

1 Telephone interview conducted on June 4th, 2019.

2 Telephone interview conducted on June 7th, 2019.

technology could be extended to the monitoring of Spanish competitors who are not covered by sports reports until they reach the semi-finals, despite having a very high number of supporters.

Meanwhile, Victor Garcia sees many options in the automation of event coverage of the Olympic Games due its status as a mega event where the general public seeks results until someone reaches the finals, or is on the verge of winning a medal. “If the result itself can be disclosed by a robot, the situation will be better for journalists, as they will have more freedom without the pressure of urgency, and thus be able to narrate with more elaborate detail”, he says.

Thus, despite the advantages of this technology, its limitations within journalism are also clear. This is why the debate also focuses on how artificial intelligence can be perfected by intervention or editorial supervision of the news writers so that higher quality texts can be produced through better preparation, or “training”, of the bot.

AnaFut is based on a programming meta-language developed in the Lab in which all editors are given the possibility of playing with all the variables. “We develop decision trees, a data structure, and we put this into different files that are used according to each situation”, Laso explains.

Therefore, depending on the amount of data entered through programming, a greater variety of lexical and statistical readings will be possible in the automated reports. Customisation of the news offering, real-time monitoring of the results, and the introduction of evolving graphics by means of data journalism are already the lines in which *El Confidencial* is working to expand the use of the bot.

Beyond the automated Second Division B reports, this technology is already being used by *El Confidencial* in other news areas such as politics (monitoring of electoral data), judicial news (monitoring of pardons, appointments to public positions - companies, politics), and stock market alerts, as well as other services such as lottery results and weather reports.

In sports, the bot is also starting to be used for match previews with information related to ‘Schedules, and where to see the...’³, which are texts that tend to always include the same data, and despite their repetitive nature they “are something that people are constantly looking for, so it generates a lot of traffic”, says Victor Garcia.

5. Discussion

The question of applying Artificial Intelligence to the production of news content is setting the development agenda of many media organizations, agencies and content producers, now more than ever. In spite of this, as pointed out by the experts consulted, we are at an early stage in the potential use that AI can bring to the sector.

5.1. *The perception of Artificial Intelligence and its application to journalism*

With regard to the perception of applying Artificial Intelligence to journalism, our experts point out that we are in a phase in which the implementation of tools and techniques of AI is taking place in countries where technological development

3 https://www.elconfidencial.com/deportes/motociclismo/2019-06-02/gran-premio-italia-moto-gp-horario-donde-ver_2046622/

is high, and also in journalistic organizations that excel in terms of innovation when searching for new approaches supported by technology and knowledge as a means of progress and growth.

The increase in the number of experiences, and above all in their quality and integration into the media structure, mean that content automation solutions will not only execute programmed commands, but will also be able to propose complex decisions based on their ability to learn.

5.2. *From the hand of bots to automatic news writing*

Bots write news automatically supported by processes and structures provided by Artificial Intelligence. Moreover, this is the most widespread⁴, use of AI in the field of journalism, and is nearly universal in this area. However, its extensive use is not positive *per se*. There is a possibility that journalistic organisations will fall into the trap of using this technology for the simplest applications, and that automated news generation will encourage vices and bad practices, examples of which could be the following:

The automated generation of content may become a simple aggregation of data.

As a result, the information might be repetitive since it comes from structures that are highly basic, very similar to Artificial Intelligence. The application of these procedures is costly and requires constant updating and process review.

It may result in newsrooms being emptied, as AI might replace journalists rather than being a tool for journalists to use, manage and integrate into the functioning of the media.

5.3. *The contribution of bots to the world of sports regarding coverage, topics, minor sports, and competitions*

The panel of experts who collaborated with this research have highlighted the capacity of Artificial Intelligence to provide media organisations with tools that will allow them to provide informative coverage of events and sports competitions characterised by its standardised structure (with norms or rules), repetition (scheduling of competitions), and the possibility of synthesising or translating these facts into data (massive statistical load of sports information).

The infinite content production capacity of a newsroom system based on algorithmic and AI solutions shows a strong influence on the considerations that the five experts have brought to the issue of the contribution of *bots* to traditional sports coverage. The classic model of conventional coverage is costly in terms of human resources and time investment, so the automated generation of content may imply the ability to reach an unlimited number of events and minor sports, or those not considered to be of general interest.

Supported by AI, the media would have the ability to increase exponentially the factual content and provide live coverage of a myriad of sports and types of competitions, as indicated by several of the experts consulted.

4 Some people believe that bots have a more direct impact in the transmission of content. However, the writing is determined to a greater extent by algorithms, which are the ones that replace the work of information professionals.

5.4. *Technology, Artificial Intelligence and the Journalistic Profession*

Some of the experts who participated on the panel have expressed a common concern about the possible indiscriminate use of bots, algorithms and other procedures related to Artificial Intelligence by media organisations. Media companies, agencies, and other content producers need to understand that the use of AI to automatically generate content might somehow eliminate workers from the business structure and reduce the number of journalists and media professionals, and this is a real possibility in a sector accustomed to being in permanent crisis.

After analysing the contributions made by the panel of experts, we are able to outline a discourse on how Artificial Intelligence should be integrated from a strategic perspective in organisations, using all of its potential to generate a culture of maximised synergy between the use of AI and the work of professional journalists. On one hand, the use of automated content generation will make it possible to offer more coverage in quantitative terms, while on the other hand, the liberation of journalists from time constraints could help them devote more hours in the day to working on real situations, issues and formats that require as much attention and human capacity as possible.

Artificial Intelligence will attempt to imitate, or in other words try to seem real, while journalists will have to differentiate their work, something that was generally agreed upon by the experts consulted. The human factor is essential in the production of content, and will continue as such, yet the editorialisation of AI processes (such as programming algorithms) will require the use of people as well. Journalists must have an important role in the organisation of tasks, supervision and process planning.

6. Conclusions

The content analysis of the sports reports, as well as the interviews with journalists and questionnaires sent to experts, show that the use of artificial intelligence for the automatic writing of informative texts offers more benefit than harm for this type of journalism. As can be seen from the results obtained in this paper, the use of the *AnaFut* bot for football match reports has meant faster production, a wider topic range, and greater diversification of *El Confidencial's* coverage of sports competitions, and at the same time has liberated journalists from automated and routine tasks.

According to its promoters, the purpose of automation is not to replace journalists with algorithms, but to allow professionals to devote more time to reporting, searching for stories, and investigating, so that the information product gains as a whole in quality and differentiation. In this regard, it remains to be seen how this quality might be enhanced in years to come by generating more natural language that is increasingly being attained and enriched from the point of view of lexical variety and the introduction of more statistical data, the purpose of which is to make the language less repetitive, more agile, more analytical, and in short, more informative.

7. Bibliographic references

- Broussard, M. (2015). "Artificial intelligence for investigative reporting: Using an expert system to enhance journalists' ability to discover original public affairs stories". *Digital Journalism* 3(6), 814-831.
- Carlson, M. (2014). "The robotic reporter: automated journalism and the redefinition of labor, compositional forms, and journalistic authority". *Digital Journalism* 3(3), 416-431.
- Cervera, J. (2017). "El futuro del periodismo es ciborg". *Cuadernos de periodistas: revista de la Asociación de la Prensa de Madrid* 34, 102-109.
- Clerwall, C. (2014). "Enter the robot journalist". *Journalism Practice* 8(5), 519-531.
- Diakopoulos, N. (2014). "Algorithmic accountability". *Digital Journalism* 3(3), 398-415.
- Dörr, K.-N. (2016). "Mapping the field of algorithmic journalism". *Digital Journalism* 4(6), 700-722.
- Edge, A. (2014, December 2). "Ophan: key metrics informing editorial at The Guardian". Retrieved from: <https://www.journalism.co.uk/news/how-ophan-offers-bespoke-data-to-inform-content-at-the-guardian/s2/a563349>.
- Fanta, A. (2017, September 24). "Putting Europe's robots on the map: automated journalism in news agencies". Retrieved from: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-09/Fanta%2C%20Putting%20Europe%E2%80%99s%20Robots%20on%20the%20Map.pdf>
- Flew, T., Christina Spurgeon, A.D. & Swift, A. (2012). "The promise of computational journalism". *Journalism Practice* 6(2), 157-171.
- Gao, T., Hullman, J. R., Adar, E., Hecht, B., & Diakopoulos, N. (2014). "Newsviews: an automated pipeline for creating custom geovisualizations for news". *SIGCHI conference on human factors in computing systems*, 3005-3014.
- Graefe, A. (2016). *Guide to automated journalism*. Retrieved from: https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php
- Graefe, A., Haim, M., Haarmann, B. & Brosius, H. (2016). "Readers' perception of computer-generated news: credibility, expertise, and readability". *Journalism* 19(5), 595-610.
- Hansen, M., Roca-Sales, M. Keegan, J. & King, G. (2017). *Intelligence: Practice and Implications for Journalism*. Brown Institute for Media Innovation and the Tow Center for Digital Journalism.
- Karlsen, J. & Stavelin, E. (2013). "Computational journalism in Norwegian newsrooms". *Journalism Practice* 8(1), 34-48.
- Latar, N.L. (2014, March 28). "Robot journalists: 'Quakebot' is just the beginning". Retrieved from: <http://knowledge.wharton.upenn.edu/article/will-robot-journalists-replace-humanl-ones>
- Lecompte, C. (2015). "Automation in the newsroom". *Nieman Reports* 69(3), 32-45.
- Levy, S. (2012). "The rise of the robot reporter". *Wired* 20(5), 132-139.
- Lindén, C. (2017). "Algorithms for journalism: the future of news work". *The Journal of Media Innovations* 4(1), 60-76.

- Lokot, T. & Diakopoulos, N. (2016). "News bots: automating news and information dissemination on Twitter". *Digital Journalism* 4(6), 682-699.
- Marconi, F. & Siegman, A. (2017, February 22). "The future of augmented journalism: a guide for newsrooms in the age of smart machines". Retrieved from: https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf.
- Matsumoto, R., Nakayama, H., Harada, T. & Kuniyoshi, Y. (2007). "Journalist robot: robot system making news articles from real world." *2007 IEEE International Conference on Robotics and Automation*, 1234-1241.
- Mittal, S. & Kumaraguru, P. (2014, June 17). "Broker bots: analyzing automated activity during high impact events on twitter". Retrieved from: <https://arxiv.org/abs/1406.4286>
- Napoli, P. (2012). "Audience evolution and the future of audience research". *International Journal on Media Management* 14(2), 79-97
- Oremus, W. (2015, October 9). "No more pencils, no more books". *Slate*. Retrieved from: <http://publicservicesalliance.org/wp-content/uploads/2015/10/Adaptive-learning-software-is-replacing-textbooks-and-upending-American-education.-Should-we-welcome-it.pdf>
- Powers, M. (2012). "In forms that are familiar and yet-to-be invented: american journalism and the discourse of technologically specific work". *Journal of Communication Inquiry* 36 (1), 24-43.
- Renó, D. & Renó, L. (2017). "Algoritmo y noticia de datos como el futuro del periodismo transmedia imagético". *Revista Latina de Comunicación Social* 72, 1.468-1.482.
- Rojas-Torrijos, J.L. (2019, June 3). "El creciente uso de bots y la expansión de las coberturas deportivas". *Mip.umh.es*. Retrieved from: <https://mip.umh.es/blog/2019/06/03/impacto-bots-informacion-deportiva/>
- Salazar, I. (2018). "Los robots y la inteligencia artificial. Nuevos retos del periodismo". *Doxa Comunicación* 27, 295-315.
- Slater, M. D. & Rouner, D. (2002). "Entertainment-education and elaboration likelihood: understanding the processing of narrative persuasion". *Communication Theory* 12(2), 173-191.
- Starbird, K., Leysia P., Hughes A. & Vieweg S. (2010). "Chatter on the red: what hazards threat reveals about the social life of microblogged information". *Computer Supported Cooperative Work*, 241-250.
- Steiner, T. (2014, March 17). "Telling breaking news stories from Wikipedia with social multimedia: a case study of the 2014 Winter Olympics". Retrieved from: <https://arxiv.org/abs/1403.4289>
- Thurman, N., Dörr, K. & Kunert, J. (2017). "When reporters get hands-on with robowriting: professionals consider automated journalism's capabilities and consequences". *Digital Journalism* 5(10), 1240-1259.
- Túñez López, J.M., Toural Bran, C. & Cacheiro Requeijo, S. (2018). "Uso de bots y algoritmos para automatizar la redacción de noticias: percepción y actitudes de los periodistas en España". *El Profesional de la Información* 27(4), 750-758.
- Usher, N. (2017). "Venture-backed news startups and the field of journalism". *Digital Journalism* 5(9), 1116-1133.

Automated sports journalism. The AnaFut case study, the bot developed by El Confidencial...

Van Dalen, A. (2012). "The algorithms behind the headlines". *Journalism Practice* 6(5–6), 648–658.

Wölker, A. & Powell, T. E. (2018). "Algorithms in the newsroom? News readers perceived credibility and selection of automated journalism". *Journalism* 00(0), 1-18. <https://doi.org/10.1177/1464884918757072>

Young, M. & Hermida, A. (2015). "From Mr. and Mrs. Outlier to central tendencies. Computational journalism and crime reporting at the Los Angeles Times". *Digital Journalism* 3(3), 381-397.

ISSN: 1696-019X / e-ISSN: 2386-3978