From the media to Wikipedia: the relationship between Chilean media news and malicious edits made in the virtual encyclopedia during the Social Outbreak of Chile in 2019.

Introduction

Several studies, including Geiß et al. 2016, show the relationship between mentions of a topic in the news and visits to Wikipedia. Based on different methods, these studies affirm that if an event or topic is repeatedly mentioned in the media, visits to its Wikipedia article and similar articles will increase.[1] The interesting thing is that in some cases these visits end up being transformed into editions that can become vandalism.

During the Social Outbreak in Chile in 2019, there was a contradiction between the reality thought or imagined collectively with the reality reported by the hegemonic media (Elias Valenzuela, 2020). You can also see the different approaches of the media regarding the news that covered the event. Cabrera (2020) finds differences such as the fact that the hegemonic media concentrated on narrating the events from the "high command", while the independent media narrated the events from interviews with civilians. The study analyzes, for example, the case of *Televisión Nacional* (TVN) which had a speech pointing to civilians as those responsible for most cases of violence in the streets, but media such as *Piensa Prensa* offered a very different version in which they pointed to the police as the cause of the violence. This position is not new, since the appearance of the mass media that there are authors who consider them as means of manipulation, where this control comes from the homogenization of taste and the reproduction of a version of reality, which can be shaped by who is in charge of the media. Nowadays, this is reflected in the fact that the main media in Chile continues to belong to elite group companies that influence the public agenda and the informative treatment of what is transmitted. Therefore, the role played by the media in the 2019 Social Outbreak is key to understanding the impact it had on different platforms, such as Wikipedia.

Taking the Social Outbreak as a temporary event, it is possible to locate the news around the events that occurred, furthermore, the fact that they have varied approaches raises the question of what topics were the most influential and made people seek information in other places. As Wikipedia is one of the most visited sites in Chile, this platform can be used as an object of study and the visits can be analyzed together with the history of editions of those articles that, as a result of the news, were the most visited during the social crisis. With the detailed history of the articles, it is possible to locate those editions that were reverted and cataloged as vandalism, in order to classify them and know which issues are the ones that most generate this type of edition. With this, it would be possible to know if what happens in Wikipedia is related to the media and if the vandalisms have any relation with the information that appears in the news. All of this is in order to be able to anticipate this type of edition and understand how Wikipedia can contribute to the verification of the information.

---

[1] This can be clearly seen with the Pandemic article on Spanish Wikipedia, where its number of daily visits was not so relevant until March 11, 2020, when a drastic peak in its visits occurred. On that day, the WHO declared the Covid-19 disease a Pandemic, causing all the media to focus on this news and, therefore, affecting the number of visits to the article. Link to graphic.

Objectives and specific questions

The main objective is to know the relationship of the news from the media in Chile, such as TVN, T13, Piensa Prensa and El Mostrador, with the vandalisms that were detected in the Spanish Wikipedia during the social outbreak of Chile in 2019.

Specific research questions:

1) What news content from the media is the one that produces the most malicious editions on Spanish Wikipedia during the Social Outbreak in Chile in 2019?
2) What are the types of vandalism that occur most on Spanish Wikipedia based on the topics analyzed in the media during the Social Outbreak in Chile in 2019?
3) What is the relationship between the news in the media and the vandalism detected during the Social Outbreak in Chile in 2019?

We know that the news of the Chilean media have a relationship that could be direct or indirect with the visits of an article in the Spanish Wikipedia; this relationship would be linked to the amount of information delivered about a specific event, in this case the Social Outbreak in Chile in 2019. Therefore, a hypothesis of this work may be that the greater the amount of news with pejorative content about the authorities in charge, the greater the amount of vandalism in their articles; taking as reference other articles related to the topic.

Relationship of this work with the current state of knowledge in the field

As previous works on this subject, very few can be found. Starting from the fact that around only 37% of the works on Wikipedia come from the social sciences and these have been declining over time (Bar-Ilan & Aharony, 2014). Another important element is the fact that each Wikipedia is different, both in its content and in its community of editors; making the most studied Wikipedia the one in English. Therefore, when thinking about antecedents, the search must be broadened toward the same concepts that will be used in the study and not on works that contain the same research topic (because there aren't).

If we look only at the social studies that focus on Spanish Wikipedia, we find that they are mostly focused on community behavior (Claes & Deltell, 2019) or content editing in a collaborative environment (Zazo et al., 2015). Unfortunately, this platform has not yet been fully developed in the area of social sciences, mainly because the potential behind it is not well known; this is in contrast to the natural sciences where Wikipedia is used as a large database both for training machines and for use in the area of natural language processing. Knowledge of the "behind the scenes" of the platform opens up a vast area of research in the social sciences since Wikipedia has its community of editors who have managed to organize themselves, have their own voting system and who have achieved set standards despite cultural differences and the constant changing of society.

In the case of the media and its relationship with Wikipedia,[2] two approaches are repeated the most. The first answers the question of how much time passes between an event appearing in the news and the creation of its Wikipedia article (Fetahu et al., 2015). The second approach goes towards the

---

[2] There are studies about it on Wikipedia in English, German and Japanese.

influence that the media have on the platform, either using the agenda-setting (Maurer & Holbach, 2016) or the dynamic transactional model (DTM) (Geiß et al., 2016). Both approaches have their limitations, but they still conclude that collectively the media do have some bearing on what happens on Wikipedia.

With the Covid-19 pandemic, several people have been monitoring the use of Wikipedia as a main source of information. This is because each country has different management of the pandemic and in some countries, the information provided by the media is doubtful (Wikipedia Contributors 2021 a). This gives us a precedent to focus the investigation since it shows us how it is that in times of crisis, people turn to Wikipedia to be able to centralize information (Ribeiro et al., 2020).

This research is framed in a specific context that is the Social Outbreak in Chile in 2019. To contextualize this process, there are several works, most of which are framed in understanding the social crisis and what led to this. Some researchers also focus on those media that emerged as a result of the Social Outbreak (Flores & González, 2020), where they show the importance of social networks as a means of information at that time. Although there is no information from studies that relate the social crisis with Wikipedia, these works help to understand the context and also serve to understand that the media used will be closely related to the results of the study, since there was a difference in the information delivered to the public. They also enlighten us on the different violations of freedom of expressions that occurred, such as the obstruction of access to public information, secret government meetings with businessmen and major TV channels and written media, conferences without the right to question by the journalists, professionals, and audiences questioning the informative treatment of national channels, monitoring communicators by intelligence services, among others (Fundación Datos Protegidos & Observatorio del Derecho a la Comunicación, 2020, p. 1). In short, when choosing the media to be used, the position taken in the face of the crisis must be considered.

Following this last line, we have the case of vandalism on Wikipedia. In this area, three-quarters of the studies have been from the area of computing about thinking of methods to automatically detect these edits (Tramullas et al., 2016). This same study mentions the lack of research into the reason for the vandalism and who makes these editions. The latter is extremely interesting, since there is a tendency to think that those who vandalize Wikipedia are only trolls, but as mentioned (Shachaf & Hara, 2010) there are different types of vandalism that are carried out by people with different motivations.

Overall, there are no previous approaches to this in the Spanish Wikipedia, only on English and German Wikipedia. The methods used there are insufficient because they only looked for vandalisms in one article, and here we're not considering only the article about the social outbreak, we're considering all of the articles related to Chile in a frame of time. Knowing how the media affects Wikipedia vandalism leads us to understand how vandalism works socially and this can be of great benefit when creating new encyclopedia policies.

In conclusion, Wikipedia has its own community that in turn has different cultures interacting with each other, therefore the question of what leads people to generate vandalism will be very different between each platform (Hara et al., 2010). Putting this together with the Social Outbreak and the role played by the media at that time leads us to ask ourselves if these vandalisms detected in the Spanish Wikipedia are related to the information provided by the media and how in times of a crisis, where people use Wikipedia as their main source, we can face these types of edits and get ahead of them.

Methodology

The main objective of this research is to be able to know the relationship between the news of the media in Chile and the malicious editions that were detected in the Spanish Wikipedia during the Social Outbreak of Chile in 2019. In this way, it is necessary to describe the methodological design that will be carried out to meet the objectives of the study; thus, the approach of this research has been of a mixed nature, both qualitative and quantitative, since in this way the data on malicious editions in the encyclopedia can be crossed together with the analysis of the content of each news that will be within the corpus to use.

First of all, and by way of defining the type of research, it is established that the analysis of the data will be both descriptive and correlational, as well as explanatory. The choice of these is directly related to the research question since descriptive studies "serve to analyze what a phenomenon and its components are like and how they manifest themselves. They allow to detail the phenomenon studied basically through the measurement of one or more of its attributes" (Vásquez, 2008, p.1), so then the facts that make up the research problem can be delimited with respect to the news of certain media of communication in Chile and the vandalisms detected in Spanish Wikipedia. On the other hand, the study is correlational, understood as one where "it is intended to visualize how various phenomena are related or linked to each other, or if, on the contrary, there is no relationship between them. The main thing is to know how a variable can behave knowing the behavior of another related variable" (Vásquez, 2008, p.3), in this way it will be possible to analyze if those vandalisms detected in the Spanish Wikipedia, during the Social Outbreak in Chile of 2019, have some relationship concerning the news of certain media outlets that came out during those days. Finally, this work is of an explanatory type, since it seeks to find the reasons or causes that cause this phenomenon (Ander Egg, 2011); with the objective of being able to explain the conditions in which the research problem occurs.

Universe and sample

As a first approximation, the universe of the sample is defined as all those Chilean people who have access to the Internet, because editions made in the digital environment of Wikipedia will be analyzed.

An important part of the investigation is to be able to define the criteria for selecting the samples. In this sense, there will be sampling criteria for a) the news of the Chilean media, b) the pernicious editions database in Spanish Wikipedia and c) the time in which the study will focus.

In the first instance, the formation of a news corpus is needed to analyze, understanding this as "a finite collection of materials, previously determined by the analyst, with a certain (inevitable) arbitrariness, on which he will work" (Barthes, 1993, p.80). The same author highlights two key elements for the construction of the corpus, the first is that the sample is large enough to achieve a framework of similarities and differences between the elements, and the second is that it be as homogeneous as possible with the sample. To meet these criteria, we place ourselves within the framework of the Chilean media, encountering the problem mentioned in previous chapters where the main media in Chile continue to belong to elite group companies that influence the public agenda and the informative treatment of what is transmitted. Taking the latter into account together with the various studies that detail the difference in the information delivered to the citizens, the secret meetings of the government with the main TV channels and written media, and the contradiction between the reality thought or imagined collectively with the reality reported by hegemonic media (Elias Valenzuela, 2020; Cabrera, 2020; Fundación Datos Protegidos & Observatorio del Derecho a la

Comunicación, 2020) it is extremely necessary to expand the search criteria to media that are characterized by being independent and pluralistic. Thus, the choice of media to be used will be the news from Televisión Nacional de Chile (TVN) and channel 13, for the traditional media, while for the independent media the news from the digital newspaper El Mostrador and the digital media Piensa Prensa. As a result, there will be four Chilean media, two traditional and two independent, thus fulfilling the homogeneity of the sample.

In the same way, it is necessary to limit the sample of the database that contains the malicious editions in the Spanish Wikipedia. First of all, you should know that malicious edits can be found in two ways, the first is by viewing the history of each article and identifying the reverted edits, and the second is by using the edits made by a bot specialized in reverting malicious edits. The latter is much friendlier, since it allows you to centralize the information and it is not necessary to search article by article, but from the reversals of the same bot one can see which articles are the most vandalized and thus find patterns of themes. In this case, SeroBOT[3]reversals will be used, since it is currently the most reliable tool in these cases and it also has easy access to the data. We must bear in mind that not all vandalism will serve us, since these editions must be crossed with issues related to the context in which the investigation is carried out, that is, the Social Outbreak in Chile in 2019. So from this great base, only those editions that are related to Chile or the topics found in the news sample of the media will be used. To meet these criteria, an algorithm will be used. This algorithm will search the history of SeroBOT and will identify the edits reverted that have some relevance to the article "Chile". This will be done by using the "what links here" for every edit made by SeroBot in a frame of time. Then, with this corpus, the algorithm will do this process again but with different articles, for example "Chilean government", "2019–2022 Chilean protests" and "Santiago Metro" (because the Social Outbreak started with the rise in the subway fare).

It should be noted that once you have the topics, along with the vandalism, it will be necessary to see the history of those articles to check if there is any malicious edition reverted manually, in other words, some users may have been pending of an article in order to monitor that there are no harmful editions and thus find manual reversals that can be used in the investigation.

Finally, we have to decide the temporal criterion of the investigation. Currently, the Social Outbreak in Chile of 2019 does not have exact dates. Some sources consider its beginning on October 18, while others take its beginning on October 14, the day on which high school and university students organized themselves to massively evade the subway passage. Along these same lines, some consider that the process ended on November 15 with the Peace Agreement,[4] others say that it ended on March 18 with the arrival in Chile of the Covid-19 pandemic, there are even some who consider that the process will end when the new Constitution is approved. However, for this research, the delimitation will be defined by the work of Watt et.al (1993) and the simplification of their model made by Ardévol-Abreu et.al (2020) shown in figure 1. Although the objective of this research is not the agenda-setting theory, this scheme helps us to delimit the sample time, that is, based on this model we can determine that from point D onwards the data, both news, and vandalism, do not have much relevance because its appearance in the media is declining. In this way, it is convenient for us to use

[3] Registered on April 19, 2018, it has more than a million editions and reverts around 1,100 vandalisms per day. To date, its error rate is 1%.
[4] Agreement for Social Peace and the New Constitution, signed by different political actors.

dates that are between point A and point D, which would be from October 18 to November 15, with

the Peace Agreement as the decay point; when the Government responds to requests from citizens.
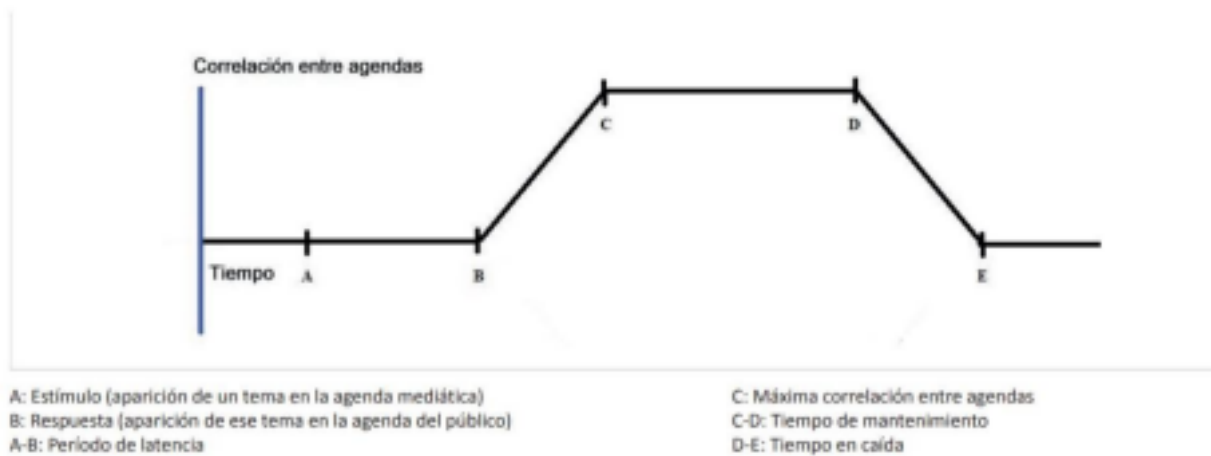

A: Estímulo (aparición de un tema en la agenda mediática)
B: Respuesta (aparición de ese tema en la agenda del público)
A-B: Período de latencia

C: Máxima correlación entre agendas
C-D: Tiempo de mantenimiento
D-E: Tiempo en caída

Figure 1. Time frames of the agenda setting. Watt et.al (1993) model simplified by Ardévol-Abreu et.al (2020)

In summary, the sample of this study will be the news between October 18 and November 15, which are related to the Social Outbreak of Chile in 2019, from the traditional media TVN and C13 and the digital media El Mostrador and Piensa Prensa. Along with this, it will be used, in the same period, the reverse editions in Spanish Wikipedia of the bot called SeroBOT.

Information collection techniques

To collect the information, a documentary analysis technique will be used, defined as "the search for a specific answer based on the investigation of documents" (Baena, 2017, p. 68). Therefore, the first step will be the collection of news concerning the subject to be studied and it will be done from two points of view: the general and the particular. In this way, it will be possible to categorize the topics of interest so that they can be analyzed later. In addition, according to Arias (2012), these investigations can be of three types: 1) monographic, 2) Measurement studies of independent variables from secondary data, and 3) Correlational from secondary data. Thus, this research would be focused on the correlational aspect, since in this way it is possible to identify the variables and determine the relationship between them.

Information analysis techniques

For this research, the information analysis technique that will be used will be the thematic content analysis, which aims to discover the meaning of a message. Specifically, it consists of classifying and/or coding the various elements of a message into categories to make its meaning appear in an appropriate way (Mayer & Quelle, 1991). Thus, those contents that deal with issues regarding the Social Outbreak in Chile in 2019 will be analyzed, which will allow knowing quantitatively and qualitatively the key issues that the media of the sample emphasized around the social revolt.

Impact and outcomes

One of the most important impacts of studying how news affects different platforms is being able to understand how the media shapes our society. The difference, in this case, is that although the news shapes our algorithmic spheres of social networks, it cannot affect "what Wikipedia shows us", since this is an information platform without algorithms involved.

Although there are studies on how in times of crisis news helps to generate the respective articles on Wikipedia, most of these studies are done on English Wikipedia, ignoring the socio-political context of the different regions. In Latin America, we are experiencing a series of social processes that have led to crisis and misinformation in various countries,[5] and the role of the press has been crucial in reporting these events; even though in many countries the traditional press ends up on the side of those who oppress communities and falls in the face of fake news. In this sense, Wikipedia has been key for communities to be informed objectively and that is why the question of "how the news affects Wikipedia?" is crucial, especially in the Latin American context.

On the other hand, merely knowing how the news affects us does not leave us much to be able to act later, since the subject is very broad. This is why focusing on malicious edits is an interesting approach, as understanding how news are related to vandalism can help the Wikipedia community take concrete action in the short term.

One of the expected results is to be able to recognize some pattern within the media, perhaps review which medium is the one that has the most impact on Wikipedia; traditional media are probably the ones that have the most impact since they have greater visibility in the territory. Understanding the vandalism that happens during a sociopolitical crisis helps us a lot to prepare for the next crisis. Having concrete results on how the news impacts vandalism can help to review article protection policies, perhaps helping to identify which ones can be protected before and not wait for the article to be vandalized. In addition, we cannot rely only on "the news tells the truth", we have seen --especially in Latin America-- that the media opt for likes and do not always verify the information. Therefore this investigation will also encounter misinforming news that is very likely to affect Wikipedia directly.

Also, and in line with movement strategies, this research is totally related to identifying impact issues and understanding what role misinformation plays in times of crisis in the Wikipedia environment. That way we can know how our projects (not only Spanish Wikipedia) can be misused or manipulated by detecting threats with significant potential for harm, such as misinformation.

Dissemination

The first step is publishing the paper in an Open Access Journal. This journal can be the [International Journal of Information and Communication Technology](#) or the [Information, Communication & Society,](#) both accepts Open Access articles. Even though these Journals have great H-index, they're in english. That's why the Spanish version of this study will be available online with no cost. The results

---

[5] Protests in Chile, Colombia, Perú, Argentina, and more.

will be adapted to all levels of understanding* through a series of infographics and open talks about it. It will also be considered working with a WM chapter to generate a series of indications about WP in times of socio political crisis for the region. It will also be great to present this in conferences,

classrooms and even in a wiki research talk at the WMF.

*Wikimedia community and people outside the wikimedia community.

Bibliography

Arias, F. (2012). *El proyecto de investigación, introducción a la metodología científica*. Venezuela: Editorial Episteme.

Ardévol-Abreu, Alberto; Gil de Zúñiga, Homero; McCombs, Maxwell E. (2020). *Orígenes y*

desarrollo de la teoría de la agenda setting en Comunicación. Tendencias en España (2014-2019). Profesional de la información, v. 29, n.4, e290414

Ander Egg, E. (2011). *Aprender a investigar : nociones básicas para la investigación social.* Editorial Brujas. ISBN: 978-987-591-271-7

Bar-Ilan, J., & Aharony, N. (2014). *Twelve years of Wikipedia research.* Proceedings of the 2014 ACM conference on Web science - WebSci '14. Published. https://doi.org/10.1145/2615569.2615643

Barthes, R. (1993). *La aventura semiológica*; Ediciones Paidós; Barcelona, España.

Baena, P. G. M. E. (2017). *Metodología de la investigación* (3a. ed.). Retrieved from http://ebookcentral.proquest.com Created from bibliotecacijsp on 2018-07-30 15:50:55

Cabrera Cares, Diego Ignacio. (2020). *El trato mediático y el uso de Twitter en el estallido social chileno.* Repositorio institucional de la Universidad de La Laguna

Claes, F., & Deltell, L. (2019). *Wikipedia en español. Comportamiento de la comunidad hispanohablante en el trabajo colaborativo en Internet.* Estudios sobre el Mensaje Periodístico, 25(3). https://doi.org/10.5209/esmp.66992

Elias Valenzuela, Arturo Alfonso. *Medios de Comunicación e imaginario social en la rebelión del 18 de octubre en Chile: Una relación contradictoria. 2020.* 84 páginas. Trabajo de Conclusión de Curso (Graduación en Ciencia Política y Sociología–Sociedad, Estado y Política en América Latina) Universidade Federal da Integração Latino-Americana, Foz do Iguaçu, 2020.

Fetahu, B., Anand, A., & Anand, A. (2015). *How much is Wikipedia Lagging Behind News?* Proceedings of the ACM Web Science Conference. Published. https://doi.org/10.1145/2786451.2786460

Flores Herrera, V. y González Pozo, D. (2020). *Estallido social en Chile 2019: comparación del tratamiento sobre el conflicto de la prensa tradicional y los medios nacidos en redes sociales.* (Trabajo Fin de Grado Inédito). Universidad de Sevilla, Sevilla.

Fundación Datos Protegidos & Observatorio del Derecho a la Comunicación. (2020, enero). *Libertad de expresión en el contexto de las protestas y movilizaciones sociales en Chile entre el 18 de octubre y el 22 de noviembre de 2019.* https://datosprotegidos.org/wp-content/uploads/2020/01/Informe-LibExpChile_CIDH_18oct_22novFDP-ODC_Chile1-1.pdf

Geiß, S., Leidecker, M., & Roessing, T. (2016). *The interplay between media-for-monitoring*

and media-for-searching: How news media trigger searches and edits in Wikipedia. New Media & Society, 18(11), 2740–2759. https://doi.org/10.1177/1461444815600281

Hara, N., Shachaf, P., & Hew, K. F. (2010). *Cross-cultural analysis of the Wikipedia community.* Journal of the American Society for Information Science and Technology, *61*(10), 2097–2108. https://doi.org/10.1002/asi.21373

Mayer, R.; Ouellet, F. (1991). *Metodología de investigación para trabajadores sociales*. Boucherville, Gaëtan Morin Éditeur.

Maurer, M., & Holbach, T. (2016). *Taking Online Search Queries as an Indicator of the Public Agenda.* Journalism & Mass Communication Quarterly, *93*(3), 572–586. https://doi.org/10.1177/1077699015610072

Ribeiro, M.H., Gligoric, K., Peyrard, M., Lemmerich, F., Strohmaier, M., & West, R. (2020). *Sudden Attention Shifts on Wikipedia Following COVID-19 Mobility Restrictions.* ArXiv, abs/2005.08505.

Shachaf, P., & Hara, N. (2010). *Beyond vandalism: Wikipedia trolls*. Journal of Information Science, 36(3), 357–370. https://doi.org/10.1177/0165551510365390

Tramullas, J., Garrido-Picazo, P., & Sánchez-Casabón, A. I. (2016). *Research on Wikipedia Vandalism.* Proceedings of the 4th Spanish Conference on Information Retrieval. Published. https://doi.org/10.1145/2934732.2934748

Vásquez Hidalgo Isabel. (2008). *Tipos de estudio y métodos de investigación*.

Watt, james H,; Mazza, Mary; Snyder, Leslie (1993), *Agenda setting effects on television news coverage and the effects decay curve.* Communication research, v.20, n.3, pp. 408-435.

Wikipedia Contributors 2021 a, *COVID-19 misinformation by governments*. (2021, 30 junio). En *Wikipedia*. https://en.wikipedia.org/wiki/COVID-19_misinformation_by_governments -Versión 1030266518

Zazo Rodríguez, Ángel F.; Figuerola, Carlos G.; Alonso Berrocal, José Luis. *Edición de contenidos en un entorno colaborativo: el caso de la Wikipedia en español.*// Scire. 21:2 (jul.-dic. 2015)57-67. ISSN 1135-3716.