

Qualitative Analysis with ATLAS.ti 22 in Social Sciences: New Tools and Specific Applications

El análisis cualitativo con ATLAS.ti 22 en Ciencias Sociales: Nuevas herramientas y aplicaciones concretas

Ramiro N. Pérez Ripossio

pramiro907@gmail.com

<https://orcid.org/0000-0001-9816-0499>

Ph.D. in Social Sciences (UBA); Master's Degree in Social Research (UBA); Bachelor's Degree in Sociology (UBA); Lecturer (UBA).

Received: November 30, 2022

Accepted: April 24, 2023

Abstract

This article analyzes the main functions of the ATLAS.ti 22 software in reference to a doctoral research already concluded. This research qualitatively addressed the migratory projects of South American transvestites/trans residing in the AMBA. Forty-four individual interviews were carried out with the migrants using the Grounded Theory as the main method. The article provides the use of the main functions of this version of the software in order to make other researchers aware of its developments. Thus, the following resources are developed: concept clouds, sentiment analysis, folders, opinion extraction and co-occurrence tables.

Keywords: ATLAS.ti - qualitative research - methodology - Grounded Theory

Resumen

Este artículo analiza las principales funciones del software ATLAS.ti 22 en referencia a una investigación doctoral ya concluida. Esta investigación abordó de manera cualitativa los proyectos migratorios de las travestis/trans sudamericanas que residen en el AMBA. Se realizaron 44 entrevistas individuales a las migrantes empleando la Teoría Fundamentada como principal método. El artículo aporta el uso de las principales funciones de esta versión del software con el fin de que otros investigadores estén al tanto de sus novedades. Así, se desarrollan los siguientes recursos: nubes de conceptos, análisis de sentimientos, carpetas, extracción de opiniones y tablas de co-ocurrencias.

Palabras clave: ATLAS.ti - investigación cualitativa - metodología - Teoría Fundamentada

Introduction

This study aims to link the software functions of ATLAS.ti with Grounded Theory, thus providing elements for researchers to develop their investigations more efficiently. To characterize this experience, a completed and approved doctoral thesis is taken as a reference, in which the migratory projects of South American transvestite/trans women residing in the Metropolitan Area of Buenos Aires (AMBA, for its Spanish acronym) were addressed.

When analyzing the latest version of ATLAS.ti (version 22), a relevant article is proposed in order to provide other researchers with an update about the resources offered by this software. It is known that

quantitative research programs such as SPSS or R are crucial to accomplishing the workflow. However, given that qualitative research generally involves a small number of cases or situations, computer resources may not be used, resorting to manual analysis instead.

In the doctoral thesis developed by the author of this article between the years 2017-2019, ATLAS.ti 8.3 and 22.2 versions were used to analyze the migration of South American transvestite/trans women residing in AMBA. This research showed that South American transvestites and trans women leave their places of origin due to vulnerabilities, discrimination and violence received from their family members, acquaintances and other members of the civil societies involved. Likewise, migration represents a core investment strategy that improves the position in the social space by increasing resources of various kinds.

The AMBA area appeals to them as they can have access to rights, experience freedom and increase their financial resources. Nevertheless, it was also shown that migration to this area has several pitfalls, since they are subject to discrimination as a result of the articulation between their gender identities, ethnic status and social class. Therefore, some of them decide to migrate to Western European countries to find more thriving environments.

The valuable empirical data stemming from the 44 interviews and participant observations led to an efficient organization of the information, and, therefore, it was necessary to incorporate the ATLAS.ti software. This software allowed for greater transparency, systematization and rigor to the analysis. Although version 8.3 was used during the doctoral thesis development, this new version 22 was used to deepen the analysis by testing new resources. The new software version offers several essential new features for a better analysis. Some of them include word and concept clouds, sentiment analysis, folders and Sankey diagrams when applying co-occurrence tables.

The academic literature at the local level has briefly addressed the importance of ATLAS.ti for social research. However, at an international level, several studies have characterized the functions of this software for social research (Varguillas, 2006; Woods et al, 2016, among others). This shows that the automation of qualitative data analysis has been crucial for decades as it represents one of the methodological concerns of current relevance.

The different sections shown below first address how data were processed and analyzed in the aforementioned doctoral thesis in order to acquaint the reader with the crucial methodological phases when applying the software functions. Then, the main functions of the software and its history are briefly presented. Finally, several functions, such as word/concept clouds, sentiment analysis, opinion extraction, folders, co-occurrence tables, coefficients and diagrams are introduced in different sections.

Data processing and analysis

Data processing is often one of the aspects that receive the least attention in qualitative research. According to Valles (2005), qualitative researchers have been concerned about the analysis and presentation of their results, but have underestimated the way data are processed and handled. As a consequence, the limited attention given to these procedures may affect the quality and rigor of the research as a whole. Although the use of a specific software is not necessary to perform these procedures, since the processing can be done manually or using a program from the Office package (Excel or Word), data processing is an essential methodological procedure to reduce, systematize and analyze the data. Therefore, CAQDAS (Computer-Assisted/Aided Qualitative Data Analysis Software) may not be used for their elaboration, but should always be present, since the methodological procedure as such represents an essential condition in the research process. In this process, the researcher transforms the information into data, assigning meaning to them, that is to say, a theoretical framework that allows them to approach the data according to the research goals. Therefore, these procedures must be applied rigorously, considering that the applied techniques provide variety and complexity to the information that is later used to structure the analysis. When there is no processing, or when the analysis lacks rigor, it is presupposed that the data are self-explanatory, assuming the reality as something given.

In this research, Excel 2016 was first used by entering the initial categories of analysis as well as the interview extracts whose contents expressed their meaning in their different variations. The first eight interviews were processed using this grid; however, the growth of the empirical material warranted the need to incorporate a software that would facilitate the analysis and provide the necessary resources to achieve the greatest systematization and rigor. In this way, ATLAS.ti versions 8.3 and 2 were

incorporated. It is relevant to note that, at this stage, preliminary analyses were conducted insofar that the grid created in Excel represented an instance of coding and analysis of the main categories, based on the identification of words and sentences that were observed in the interviews and that were relevant from the theoretical point of view (see Seid, 2016).

Given that, at the local level, qualitative research often lacks the use of software to process the data produced, the resources are often limited to specific programs in the Office package. When using quantitative methods, the manual approach becomes impossible due to the volume of empirical material involved, which usually becomes unmanageable. Qualitative data processing packages make it possible to integrate different types of information and conduct analysis procedures such as code co-occurrence, which illustrates the link and codependency of the categories of analysis, also allowing to verify the consistency of the substantive theory that is being constructed. In other words, these resources represent vectors to provide quality and systematization to the analysis.

In the case of this research, the main resources applied were: code assignment, code families, code hierarchies coding, notes, word clouds, memos, diagrams, the design of code co-occurrence tables and code-document, some of which are found in the appendix. Memos are analytical devices in which interpretations of the empirical material are constructed. Codes represent labels that name the categories and reduce data; the notes are first-hand written elaborations that allow for initial interpretations and code families enable to sort and hierarchize the categories to provide structure to the emergent theory. In regard to memos, according to Vivar et al. (2010), descriptive, analytical, theoretical and methodological memos have been created which allowed users to incorporate the analysis procedures based on the data provided by the interviews and participant observation.

Then, the variations constructed in accordance with the properties and dimensions of the respective categories were entered into a matrix elaborated in the program Word 2016, making it possible to illustrate how the categories of analysis were generated. ATLAS.ti software is useful when working with Grounded Theory (Strauss and Corbin, 2002), as most of its functions have been articulated consistently with this method (see San Martín Cantero, 2004).

Given that there is often an overlap of phases in qualitative research, the distinction between processing and analysis is related to clarity in presentation, since "...in qualitative research, it is not possible to separate the processing stage from the analysis stage: the development of the former entails the latter" (Cohen and Gómez Rojas, 2014:13). In other words, creating a sample, selecting the cases that provide relevant information and establishing coding and analysis procedures, through the comparison of empirical material, represent instances involved in the same process.

One of the tools proposed by Grounded Theory, and which leads to data analysis, entails different coding types: open, axial and selective, which give rise to the construction of categories, sub-categories, properties and dimensions. In order to achieve this goal, it is necessary to apply the analysis method called constant comparison, in which incidents must be detected and compared according to groups or individuals using maximum and minimum variations so that the categories are constructed robustly (see Seid, 2017). This procedure consists, first, in reducing the differences to design the categories and then highlighting the differences to find variations regarding the properties and dimensions. This instance implies being able to delimit the main categories relevant for the analysis and later to strengthen them through the incorporation of properties that show significant variation. It was through this procedure that it was decided to focus on constructing robust and dense categories, showing the greatest variability and richness possible.

As pointed out on other occasions (Perez Ripossio and Cirone, 2019), the initial microanalysis or "line-by-line" analysis, includes open and axial coding and makes it possible, at first, to discover categories and begin to unveil the relationships between concepts. The analytical process that enables researchers to identify concepts and discover their properties and dimensions in data is known as open coding. Starting from the field records, comparisons that seek to identify similarities in the contents are made. These incidents, which can be objects, actions, events, ideas or interactions and that share a common denominator according to the researcher's theoretical perspective, are identified with a common code. In this way, they are placed in existing or emerging categories at the time of their discovery, and constant comparison allows to dynamically generate properties of the categories. The names of the categories can be established by the researcher according to either the evocations that they produce from a theoretical context, or they may be named using *in vivo* codes, that is, using the words from the interviewees' own language.

Axial coding implies that the researcher has already established the main categories of analysis and that the initial analysis procedure was finished; therefore, the categories begin to be incorporated, refined and improved. This coding makes it possible to link the categories to their sub-categories in order to achieve more thorough explanations of the phenomena. Thus, axial coding refers to the procedure that, taking a category as an axis, links its properties and dimensions and, thus, gives it depth and structuring. At this point, the theory is formed. Axial coding aims to systematically construct dense categories linked to each other, which is certainly a complex task. The comparison, in this instance, is focused on observing the categories with the data in order to refine and strengthen them.

Thus, some of the main categories were: the reasons for migration, the age when assuming their gender identity, the type of social representations of Argentina as a migratory destination, the migration mode, the type of discrimination, the type of relationship with their long-distance families, the type of transit migration and the type of temporary return experiences. There are also some secondary categories, related to the main ones, namely: the mode of practicing prostitution/sex work, the type of sexual practices, the type of emotional bonds, drug use, the type of body interventions, genital reassignment, obtaining ID cards and type of bond between transvestites/trans women, among others.

On the other hand, it is necessary to clarify that the coding proposed by Grounded Theory does not consist of temporal stages and a specific linearity. In other words, open recoding may be necessary to the extent that the generation of new interviews elucidates relevant categories that either had not been observed or taken into account at the beginning of the research (see Cohen and Seid, 2019). This aspect is relevant, in this research, it was necessary to go back and recode interviews when emerging categories of analysis were not initially considered.

Selective coding implies the discovery of one or two categories that have analytical power. For this research, the migratory project and gender identity were the core categories, as they permeate the set of constructed categories. They have analytical power and the definitions of the theoretical framework account for its theoretical density. This does not mean that other analysis categories are not central, but some of them include the rest, containing them and establishing relationships with them.

On the other hand, with regard to participant observations, the field diary was handwritten and, then, a specific memo was created in ATLAS.ti, in which some notes of greater relevance were elaborated. In this memo, some condensed, expanded, methodological notes as well as notes of analysis and interpretation were detailed (Valles, 2000), conferring greater rigor to the work done through interviews. The memos created using the software fell under the categories of analysis developed through tables presented in the following chapters and which, as already mentioned, complemented the information provided in the interviews from the initial stages. In addition, the program generated reports with interpretations of the empirical material at various levels, which facilitated the writing process.

Finally, analytical networks or diagrams were also created using the program. These resources represent visual analytical devices that allow users to approach the creation of substantive theory. The networks link the codes referring to the categories and allow the analysis to be incorporated through the hierarchical connection of constructed codes. As quoted by Strauss and Corbin, "In the end, it is important to have a clear and graphic version of the theory that synthesizes the major concepts and their connections" (2002, p. 258).

ATLAS.ti: its main functions and brief history

This section neither exhaustively presents the history of the software nor attempts to highlight its main milestones. Only a concise contextualization is made so that the readers can situate themselves and learn a little more about the origin of the program.

ATLAS.ti is one of the most widely used software tools in qualitative research, although it may also be applied to quantitative approaches (Valles, 2001). While it is not the only one, given that there are others such as NVivo, Etnograph and AQUAD, it leads the international market. Its development began towards the end of the 1980s. By the 1990s, the acronym CAQDAS (Computer-Assisted/Aided Qualitative Data Analysis Software) was used to determine the set of programs for qualitative data analysis. In short, the expansion of computers during those years consolidated the use of computer systems to analyze qualitative data. However, computerization has not only brought positive results, as it has sometimes been suggested that computer functions are somewhat automated and that there is a lack of holistic view when observing the data.

The functions provided by ATLAS.ti are difficult to address because of their diversity and complexity. Some of them are basic but useful for starting to organize, process and analyze data. The creation of quotations, codes, memos, networks and co-occurrence tables are the most used resources provided by the initial versions of the software. Several of these functions were already mentioned in the previous section. Therefore, the following sections will include the more innovative functions, bearing in mind the recent update provided by version 22 of the program.

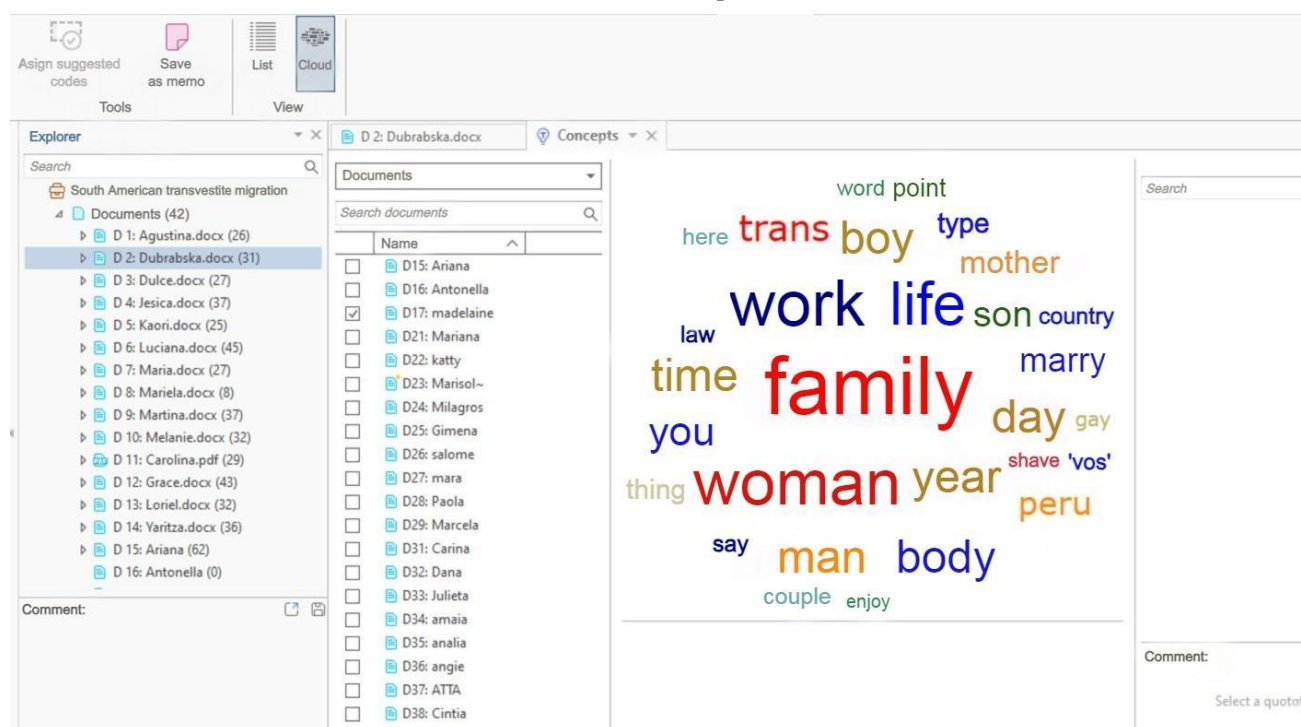
Word and concept clouds with more options

The new word clouds offered by the software in its version 22 make it possible to differentiate a set of terms present in the documents we worked on. This analytical resource allows to start an exploratory analysis, by identifying those terms which are most frequently present and may be relevant for the analysis.

Both the use of adjectives and adverbs serves as a way of approaching the microanalysis and line-by-line analysis. As mentioned above, this exploratory analysis procedure enables to discover frequent words, so as to then qualitatively think about why they are present in the interviewees' discourse. Strauss and Corbin (2002) have highlighted the importance of the words for the Grounded Theory when an initial microanalysis is intended. This resource, as well as the word list, can also be useful for performing content analysis, which is another qualitative inquiry strategy. Identifying words that occur repeatedly may identify incidents in empirical material. It is a starting point for observing and understanding the meaning of these words and delimiting their theoretical relevance. The empirical material may even surprise researchers with words that had not been considered by the theoretical framework and that may be coded as in vivo codes given their theoretical relevance.

Furthermore, this version enables to search concepts from which greater density codes and even categories of analysis could be developed. Concepts are ways of approaching the analysis that may be saved as memos, providing a clear view of each one of them and their frequency of occurrence, as shown in the following figure.

Table No. 1: *Concept clouds.*



Source: own elaboration.

The above table is an extract from the mentioned doctoral research. This concept cloud made it possible to start elaborating significant assertions for the analysis. For example, the relevance of recurrent concepts among the interviewees, such as “family”, “woman”, “body” and “work” can be observed.

Sentiment analysis

This new version 22 classifies people’s feelings in interviews or documents as neutral, positive and negative. Although it is necessary for the researcher to accompany the classifications produced by the software, in general terms, this allows to evaluate the value judgments made by the interviewees. In addition, feelings are related to people’s subjective experiences, usually enabling to identify the most relevant passages from the interviews or other types of documents.

Version 22 may also identify whether the predominant feelings expressed are positive, neutral or negative, which makes it easier to identify the emotions conveyed by the interviewees and to code the incidents more efficiently. People’s emotions and feelings are usually a crucial aspect in qualitative analysis; therefore, having analysis tools to visualize these elements is a positive factor.

This can be especially useful for certain field studies such as the sociology of emotions. However, the limitation of this resource is that it does not delve deeply into the meaning of the feelings by itself and, sometimes, it misunderstands the actual meanings conveyed by the interviewees during the interview. This fact was observed in the research about migratory projects of South American transvestites/trans women. All the same, it could also be observed that most interviewees expressed negative feelings towards their lives in their places of origin.

In short, this is an initial automated resource to approach people’s emotions when it is relevant to address them. However, the final interpretive decisions will be left to the discretion of the researchers.

Opinion extraction

This resource resembles the sentiment analysis, but identifies verbs or nouns that affect feelings. The interesting fact is that it provides a bar chart showing the frequency of occurrence of these elements and how they are affected. It conducts the data analysis not only for those who focus their work on the sociology of emotions but also for those who deal with aspects related to people’s feelings. It may also be useful to know the interviewees better.

Once again, the main limitation of this tool lies in the fact that the meaning of the associations made by the software is not guided by a specific theoretical framework. In addition, some inconsistencies were observed, as, for example, when the transvestite/trans women reported situations of violence or suicide attempts, the software suggested coding them as positive feelings.

Despite some inaccuracies, it is necessary to clarify that the software does not analyze the data. On the contrary, it provides tools for the researchers to interpret the empirical material and suggestions generated by the functions. This interpretive task can be facilitated by these functionalities.

Folders

Folders may have different functions but, mostly, I think they may be useful for creating categories. Categories are concepts that have a certain level of abstraction and that make it possible to advance in the analysis by identifying their properties and dimensions.

The creation of categories makes it possible to code the empirical material and integrate the different codes into specific folders. This goes beyond the code family as it not only links the empirical material in its different aspects, but also hierarchizes them, being able to assign different roles to each of the codes. The creation of folders is a resource that researchers can use once they have already established code fusions and families, thus creating denser categories of analysis.

In addition, the folders may bring together diverse or scattered codes that the researcher decided not to merge due to their different nature. It is a way of grouping varied information and not losing the richness of qualitative analysis. The folders represent one of the new features of the software, as previous versions did not offer this option.

In the research taken as an example, the folders facilitated the grouping of specific codes that were presented in a scattered manner. Furthermore, when the analysis was advanced, the folders gathered and

included codes of core categories labeled as “discrimination”, “violence”, “representations” and “bonds”, among others.

Co-occurrence tables, “C” coefficient and diagrams

Co-occurrence tables in ATLAS.ti make it possible to observe the relationship between the different codes, by exploring their links in order to favor the development of the emergent theory. These types of tables favor conceptualizations and analysis because, by observing the fusion of codes, the empirical material can be intervened through relationships while connections can be also discarded. This resource has been provided by the software for a long time; however, Sankey diagrams are a recent resource that allows codes and their relationships to be visualized creatively.

Sankey diagrams are not the only graphs that can be visualized, as it is also possible to visualize them through bar charts. Furthermore, in this same tool, the quotations referring to those codes can also be visualized in a pop-up window located on the right side of the screen. This allows for a three-part visualization that favors the analytical experience, since the researcher, in addition to observing the co-occurrence table and the diagrams, can also have access to the quotations, interpreting the relevance of the links. In addition, by observing the co-occurrence table, the code frequencies can also be visualized, making it possible to understand the scope of the research and whether theoretical saturation has been achieved.

Table No. 2: *Co-occurrence tables and Sankey diagram.*



Source: own elaboration.

On the other hand, it is important to point out that the “C” coefficient measures the strength of the association between the category codes used to analyze the empirical material. This coefficient is the result of dividing the co-occurrences of codes and the occurrences or rootings. The relevant fact is that, similar to other coefficients, it can vary between 0 and 1, showing that the values that are closest to one determine a stronger association between category codes. Beyond this, the result of the coefficients is usually close to zero.

In the research taken as an example, the category code “police violence” co-occurred with cases of arrest on ten occasions, giving a C coefficient of 0.26. This proved that, when transvestites/trans women are arrested or detained, the main situations of violence from law enforcement take place. Ultimately, it

allowed to confirm that the arrest and subsequent deprivation of liberty that followed were a form of violence.

Conclusions

Qualitative analysis is a fundamental phase of the research process because it contributes to knowledge by generating new findings or conclusions that differ from those considered by other research studies. In this article, some tools provided by ATLAS.ti which are useful to facilitate and improve this process, were briefly described. Word clouds, sentiment analysis and the identification of concepts allow to identify incidents by performing an exploratory analysis of the information. Folders allow information to be better organized in terms of categories and to integrate codes into concepts of greater theoretical density. Finally, co-occurrence tables are more complex resources for analysis that make it possible to observe the links between codes in tables and graphs to determine whether theoretical saturation has been achieved.

Indeed, not all the functions offered by the software that have been integrated into other versions, have been described here. Moreover, the functions to edit documents and import comments made on social networks in a simpler way are also useful. This latter resource facilitates the processing and analysis of qualitative data that come from virtual spaces.

ATLAS.ti represents a unique qualitative analysis experience, serving as a bridge between researchers and data. This program promotes transparency, communication, reduction and interpretation of empirical material, making the qualitative research process simpler, while also ensuring a more rigorous and systematic approach. Despite these advantages, it has the limitation of a certain level of automation in the analysis procedures, which researchers must tackle according to their theoretical perspectives.

When working with a significant number of interviews, documents and observations, as in the research addressed in this article, it is crucial to have software tools that facilitate data processing and analysis. This article outlines various functions and examples of the implementation of ATLAS.ti, aiming to provide researchers with information on its usage so that they can apply it to their own research.

Bibliography

- Cohen, N., & Seid, G. (2019). Producción y análisis de datos cualitativos. En N. Cohen & G. Gómez Rojas (Comps.), *Metodología de la investigación, ¿para qué? La producción de los datos y los diseños* (pp. 203–228). Teseo; CLACSO.
- Pérez Ripossio, R. N., & Cirone, M. (2019). Travestis y trans sudamericanas que migran hacia el Área Metropolitana de Buenos Aires: Una aplicación de la teoría fundamentada. En H. Lewin (Coord.), *XIII Jornadas de Sociología de la Universidad de Buenos Aires*. Universidad de Buenos Aires.
- San Martín Cantero, D. (2014). Teoría fundamentada y Atlas.ti: Recursos metodológicos para la investigación educativa. *Revista Electrónica de Investigación Educativa*, 16(1), 104–122.
- Seid, G. (2016). Procedimientos para el análisis cualitativo de entrevistas: Una propuesta didáctica. En *V Encuentro Latinoamericano de Metodología de las Ciencias Sociales: Métodos, metodologías y nuevas epistemologías en las ciencias sociales: desafíos para el conocimiento profundo de Nuestra América* (pp. 1–26). Universidad Nacional de La Plata. Facultad de Humanidades y Ciencias de la Educación. Centro Interdisciplinario de Metodología de las Ciencias Sociales. http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.8585/ev.8585.pdf
- Seid, G. (2017). La pluralidad de procedimientos para alcanzar validez en las investigaciones cualitativas. *Revista Latinoamericana de Metodología de la Investigación Social (ReLMIS)*, 12, 41–55. <http://www.relmis.com.ar/ojs/index.php/relmis/article/view/165>
- Strauss, A., & Corbin, J. (2002). *Bases de la investigación cualitativa: Técnicas y procedimientos para desarrollar la teoría fundamentada*. Universidad de Antioquia.
- Valles, M. S. (2000). *Técnicas cualitativas de investigación social*. Síntesis.
- Valles, M. S. (2001). Ventajas y desafíos del uso de programas informáticos (p. ej., ATLAS.ti y MAXQDA) en el análisis cualitativo: Una reflexión metodológica desde la *grounded theory* y el contexto de la investigación social española. En *Seminario sobre Investigación Avanzada Cualitativa Asistida por Ordenador* (pp. 1–26). Fundación de Centros de Estudios Andaluces.

Varguillas, C. (2006). El uso de ATLAS.ti y la creatividad del investigador en el análisis cualitativo de contenido. *Laurus*, 12(Número extraordinario), 73–87.

Woods, M., Paulus, T., Atkins, D. P., & Macklin, R. (2016). Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS.ti and NVivo, 1994–2013. *Social Science Computer Review*, 34(5), 597–617.
<https://doi.org/10.1177/0894439315596311>

Para citar: Pérez Ripossio, R. N. (2023). Qualitative Analysis with ATLAS.ti 22 in Social Sciences: New Tools and Specific Applications. *Perspectivas Metodológicas*, 23(27). <https://doi.org/10.18294/pm.2023.4324>



Tanto la revista *Perspectivas Metodológicas* como todos sus contenidos se encuentran publicados bajo la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0). <https://creativecommons.org/licenses/by/4.0/>. Esta licencia permite copiar, redistribuir, remezclar, transformar y construir a partir del material en cualquier medio o formato, incluso con fines comerciales. El ejercicio de estos derechos está condicionado al cumplimiento de ciertos requisitos: se debe otorgar el debido reconocimiento a la autoría original, incluir un enlace a la licencia correspondiente e indicar si se han realizado modificaciones al contenido. Asimismo, no pueden imponerse restricciones legales ni aplicarse medidas tecnológicas que limiten los usos autorizados por la licencia.

Translated by Rocío Andersen and Andrea Martin, students of the Undergraduate Degree in Translation Studies at Universidad Nacional de Lanús. Reviewed by professor María Victoria Illas.