

Categories: A Decision – Not a Paradigm

Teresa Herzgsell

In order to form a basis for quantitative approaches to the magazines we are analysing, we have created a grid for data-collection¹ that has enabled us to split the contents of the 42 magazines of our corpus into a set of categories. That allows us to quantitatively sort and analyse the material we are working on, with regard to the research questions that are at the heart of our project.² This practice was born out of the necessity to gather information that can be computed; and reflects the fact that the “act of classification is of its nature infrastructural, which means to say that it is both organizational and informational, always embedded in practice” (Bowker / Star: 320). Such an approach is clearly opposed to the hermeneutic practice that is usually taught and used in the humanities. With the number of magazines our project is working on, the necessity arose to implement computational means of information processing that would be adept at handling such large quantities of information. In order to use computational tools unique data had to be established, and classification as a practice was introduced in the project. In order to classify, however, the ambiguities, that are usually addressed with the interpretational methods typically used in literary studies, had to be reduced.

¹ Hanno Ehrlicher and Jörg Lehmann, “Datenerhebung als epistemologisches Labor – Überlegungen am Beispiel der virtuellen Forschungsumgebung *Revistas culturales 2.0*”, in: Martin Huber, Sybille Krämer and Claus Pias (Eds.), *Forschungsinfrastrukturen in den digitalen Geisteswissenschaften. Wie verändern digitale Infrastrukturen die Praxis der Geisteswissenschaften?* Frankfurt am Main: CompaRe 2019, S. 40–57. Available online at <http://d-nb.info/1201549302/34>.

² For the project description see: <https://gepris.dfg.de/gepris/projekt/327964298>.

The following paragraphs will describe the categories we established, and address the ways we handled the most prevalent of these ambiguities.

Our grid came to consist of 19 columns: Magazine; Number of Contribution in Magazine; Contributor: Last name(s)/First name(s); Pseudonym / Abbreviation Used; Source of Contributor Data; Sex; Country of Origin; Year of Birth; Year of Death; Age at Contribution; Title of Contribution; Issue Number; Original Issue Date; Calculable Issue Date; Type of Contribution; Language; Dedication; Translator; Original Language.

We created the column 'Number of Contribution in Magazine' so as to keep the order in which the contributions appear in the magazines and to facilitate the location of a single contribution, particularly important when some issues could be quite large and did not always use the convenience of page numbers. We split multiple authorships into single lines, and indicated this in the relevant line by adding alphabetized lower-case letters to the numbering. In February 1925, in issue number 7 of *Proa* (Buenos Aires), the article "La moderna poesía en Cuba" co-authored by Cuban Félix Lizaso and Argentinian José Antonio Fernández de Castro was published. As this was the 176th contribution in the magazine, we formed two lines, one for Lizaso and one for Fernández de Castro, and indicated this by the numbering 176a and 176b. The contribution would obviously be counted as one while keeping the information on both authors intact and connectable to their other contributions as individuals.³

³ We have made one exception from this rule in the Spanish modernist magazine *Germinal*. In this publication we found a set of manifestos signed by the prisoners incarcerated in Barcelona in the Cárceles Nacionales and in the Castillo de Montjuich. With those contributions we kept the collective authorship intact, and forwent to split them up in the a/b/c structure. In this particular case, it would not have been productive to do so, as it is most unlikely that all 50 or so signatories had a part in creating the manifestos, and even more importantly the function of authorship in this case is decidedly a collective one that negates the exhibition of the (legal) responsibility function of authorship.

The data for the 'Contributor' column was recorded in the format of 'Last name / First name', as we encountered the entries in the secondary sources found under the column 'Source of Contributor Data'. We normalized the spelling of the name according to the form found on VIAF, since this enabled us to find the same contributors across the data, despite different spellings, typographical errors, use of pseudonyms, etc. In order to make the resolutions of pseudonyms and abbreviations fully comprehensible, the original contributor name that was given in magazines was kept in a separate column (Pseudonym / Abbreviation Used), in which we also noted unresolvable cases to facilitate later revisions. However, we refrained from noting typos or different spellings in this column.

Most sources for contributor data are online sources. As a standard, we decided on VIAF as our primary source for contributor data. This decision is grounded in a positioning paper by the Association for Digital Romance Studies (AG Digitale Romanistik).⁴ The paper explains the specific difficulties in Romance Studies, such as the dissemination of research across many language areas, geographies and research cultures. The problems arise around the issue of authorship description in research, as there are no conclusive and standardized authority files provided. Instead, scholars are faced with a rag rug of unsatisfactory, small, and unsystematic, national solutions. For those reasons, the association has recommended the use of VIAF as a primary source for authority data regarding authors. For our comparative approach as well as the subsequent provision of the data collection for the international research community, VIAF was the only sensible source, despite its deficiencies. From VIAF we used the first

⁴ See http://deutscher-romanistenverband.de/wp-content/uploads/sites/14/Open-Access-und-Forschungsdaten_Mrz-2017.pdf [last accessed 11 August 2020].

name that is given as the name of the contributor for our data. In some cases, this can lead to a problem, and that is that there are some double entries in VIAF. In such cases, we decided on the most accurate set of data, and provided the permalink to only this set. We are fully aware of the problems a source like VIAF causes, as it is solely based on grabbing mass information from the web and sorting it by frequency. Double entries are only the tip of the iceberg, and the grabbing mechanism inevitably leads to more mistakes that could be a whole research question in itself regarding swarm intelligence, namely, of the ramifications of a mistake becoming so prevalent and widespread that it threatens to replace the correct version. Nevertheless, this is a completely different subject, and is only germane to our project inasmuch as we are conscious of these problems, and make them transparent together with our choice to use VIAF whenever possible. Where there was no VIAF source we deferred to other online sources, and the following are a few of the non-online sources we have used, when we could not find any information on the internet:

In the corpus for the avant-garde, some information has been taken from the magazines directly, but also from Juan Manuel Bonet's *Diccionario de las Vanguardias en España 1907-1936* (Madrid: Alianza, 2007) (abbrev. DDLVE), and Guillermo Sheridans *Índices de Contemporáneos: Revista Mexicana de Cultura (1928-1931)* (Mexico D.F.: UNAM, 1988) (abbrev. INDEX *Contemporáneos*).

For modernist magazines, digital resources such as the *Biographical Dictionary of the Royal Academy of Spanish History* (<http://dbe.rah.es/>), the bibliographic data portal of the *National Library of Spain* (<http://datos.bne.es/>), and the *Encyclopaedia of Literature in Mexico* (<http://www.elem.mx/>) were consulted. Additionally, usage was made of printed catalogues (e.g. *Ensayo de un catálogo de periodistas españoles del siglo xix*, Manuel

Ossorio y Bernard, Madrid: Imprenta y Litografía de J. Palacios, 1903) and other specialised sources (e.g. *La mujer de letras o la letraherida: discurso y representaciones sobre la mujer escritora en el siglo XIX*, Pura Fernández and Marie Linda Ortega, España: Consejo Superior de Investigaciones Científicas, CSIC, 2008), especially to gather data about the less prominent contributors.

The information on 'Sex'⁵ was completed in accordance with these sources, where this was available, and with regard to the name of the contributor, where corroborative information was otherwise inconclusive.

The 'Country of Origin' is a more problematic category than meets the eye. We decided to use the countries as they are defined on today's political map and not the historical territories. When in doubt the contributors were pinpointed to these areas via their place of birth (wherever this could be established).⁶ The reasons for this are manifold, and clearly open to contestation. They are mostly pragmatic, and have to do with our quantitative approach, as much as with the visualization and mapping of the information. As our research is concerned with comparisons of magazines from the 1890s to the 1930s, the constantly changing political situation and shifting borders of the time become problematic. If historically accurate 'countries of origins' were to be

⁵ After the debates on sex and gender in the 20th century, it might seem reductive to use sex as a category. We do this, because we are working with sources on biographical data which do not differentiate between sex and gender and because in the historical period we examined, questions of gender-identity were still in the distant future; the sex of a contributor can therefore most often be deduced from the name. Opening discussions on these questions would be to distort the historical reality, where it was not unimportant what sex the author of a text had, and the systematic exclusion of most women from political and societal debates was still in full force.

⁶ We could not resort to using the place name (e.g. the town or city) of birth instead, as for many contributors it was difficult enough to find even a country of origin, and to attempt to trace contributors so exactly would have been unfeasible in view of their quantity.

used, this would lead to a fragmentation and distortion especially on maps, but also in statistical explorations. As we are operating with material from a turbulent time regarding state-building, some of the contributors would be born in the Viceroyalty of New Spain (1521-1821), others in the Federal Republic of Central America (1823-1841), while later-born contributors would have to be allocated to countries like Nicaragua (1938-today) or Cuba (1898-today) for instance. This doesn't even touch on the developments taking place in the Europe of the 19th century, when most of the contributors were born, and the building of nation states had only just begun.⁷ While states like France and Spain were relatively stable, even at that time, other areas, like the land current Germany is located on, were permanently changing. In cognizance of these problems and to facilitate comparative visualizations, statistics, and mapping, we came to the decision of using current country distinctions. We consciously decided to use the category 'Country of Origin' instead of 'Nationality', as there are several cases of magazine contributors where it would be difficult to ascribe the nationality because of their migration histories, or their ancestry. An especially complicated example of this would be the case of Max Aub, who has German ancestors, was born in Paris, and grew up in Valencia. Later he held a post as a cultural attaché in the Paris-based Spanish embassy, before anonymously being denounced as a political traitor and having to leave Europe for exile to Mexico, where he worked in many different capacities, including university teacher, screenwriter, editor, and translator, and died in 1972. In a biography like this, one single nationality ascription is negated as much as

⁷ These videos illustrate the changes in the world and specifically in Europe, and show quite vividly the problematic nature of 19th-century cartography: https://www.youtube.com/watch?v=b0zTNZ1n_VA and <https://www.youtube.com/watch?v=-6Wu0Q7x5D0> [last accessed 11 August 2020].

any form of cultural affiliation. As we were on the lookout for one clear localization, we decided that the 'Country of Origin', as in the 'Country of Birth', in the case of Max Aub would be France. Detailed differentiations, we leave to the respective scholarship, as questions of cultural or national belonging of single contributors are not at the heart of our research. Pinpointing Aub to France fulfils the requirements of our approach, as it still shows the transnational connections inherent in his persona.

We also collected the 'Year of Birth and Death'. We decided on the year instead of the exact date, as for many contributors finding the exact date would have been impossible, or would have posed a laborious expenditure. In the cases in which there was doubt, we verified the dates by checking specialized secondary sources. Where this was not possible, we used the first entry in VIAF to determine the birth and death year, in the very same way we normalised the spelling of the names of the contributor names.

The 'Title of Contribution' features the title as written in the magazine, including any overlines or subheadings. The datasheets can be searched for topics appearing in these titles. This provides a good initial overview, and when used creatively can provide a wide array of texts for one topic. If one were to search for contributions concerning Rubén Darío for example, not only searching for his name, but also checking entries containing 'Nicaragua' (as well as its inflections and conjugations), 'Poesía', 'Literatura', etc. would be a good way to find not all but a significant number of contributions concerning Darío.

Next to 'Issue Number' and its 'Original Issue Date' the datasheets also contain a 'Calculable Issue Date' in the format Year-Month-Day, in order to facilitate

computational processing of the data. When an exact date was not provided by the publications e.g. only the month or a period of months was given, we used the first day of the given time span as the 'Calculable Issue Date'. For example, issue 7 of *Amauta* states 'Marzo de 1927' as the date. This information we kept in the 'Original Issue Date' while the 'Calculable Issue Date' states '1927-01-01'.

The category that involves most of the interpretative work in our datasheets is the 'Type of Contribution' which is elaborated on in a second paper (see file "05_Teresa-Herzgsell_Categorization-as-Theory-and-Practice"), which also contains information on the theoretical framework concerning categorization in the project. In this category, we ascribe each contribution to one of seven transhistorical categories – Non-Fictional Prose, Review, Magazine Review, Fictional Prose, Drama, Lyricism, and Image – in order to make visible the formal markup of the publications, as well as allow for comparisons of the markup. Moreover, this category helps the systematic search within the data for specific texts and phenomena.

We also noted the 'Language' of the contributions, as not all foreign contributions were translated. One extreme example for this would be Vicente Huidobro's magazine *Creación/Creation* which involves five different languages in its 39 textual contributions (five of the contributions being images) – Spanish (8), French (26), German (3), Italian (1) and English (1). Where there were translations indicated we input the information (if given or obtainable) in the 'Original Language' as well as the 'Translator' columns. We used the same format for both the 'Contributor' and the 'Translator' columns, so that any cross-connections could be easily made. The same thinking was applied to the

'Dedication' column, which is why only the names of the dedicatees were recorded, rather than whole dedicatory sentences.

Bibliography:

Bowker, Geoffrey C. and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press, 1999.