Towards a Phenomenographic Framework for Exploratory Visual Analysis of Bibliographic Data

Martin Ruskov, Sara Sullam

Department of Languages, Literatures, Cultures and Mediations, University of Milan, Italy

Abstract

A recurring challenge when studying history of translation is interpreting catalogue metadata. On one hand such interpretation is limited by the fact that data present in catalogue records is tabular and nominative, and not quantitative. On the other hand, such research is guided by tacit knowledge of scholars in the humanities, and thus it could be challenging to reproduce its results. We take inspiration from phenomenography, a discipline within educational research that examines how students perceive the phenomena being learned. We adopt the view that scientific inquiry is a collective form of learning. By doing this, we turn to the phenomenographic theory that variation is necessary to understand the phenomena being studied, and is achieved through three distinct patterns of variance: contrast, generalisation and fusion. We propose an approach to visualise the combination of nominal data and tacit knowledge by subjecting it to these three patterns. We illustrate our approach with two case studies from literary translations between Italy and the UK in the post-war 20th century. Our claim is that on one hand this guides scholars on how to analytically approach their research questions, on the other it drives them to externalise and validate hidden assumptions. Our approach offers a way of doing reproducible science not only when conducting literature research with bibliographic data. It is also applicable in the wider cases within the humanities when tabular data are available.

Keywords

Literary Transfer, Phenomenography, Bibliographic Data, Tacit Knowledge

1. Introduction

A recurring challenge in translation history is the interpretation of bibliographic metadata, a task readily available for analysis within the digital humanities [17, 22]. Yet, most of the data in bibliography are nominal, which strongly reduces the applicability of commonly used techniques from comparative analysis, semantic language modelling or quantitative visualisations. Moreover, complex historical and critical analysis typical for research in the humanities remains out of the scope of numerical approaches [16]. This is partly due to the fact that in research projects with a humanistic focus scholars tend to formulate hypotheses within their discipline relying heavily on previously acquired – and therefore tacit – knowledge shared by other members of their scientific community. This has posed a challenge to scholars' collaboration in the digital humanities, most relevantly in the creation of a shared research environment where researchers coming from different backgrounds can elaborate research questions relevant at a real interdisciplinary level, i.e. not relying on tacit knowledge. John Biggs places

tacit knowledge within an extensive taxonomy of knowledge types. Beyond other commonly discussed forms such as theoretical and procedural knowledge, he also includes tacit, intuitive, metatheoretical and conditional knowledge. Some of these categories include forms of understanding that are provisional or even counterfactual, and scientific rigour would require the validation of such knowledge through critical reasoning or triangulation with other knowledge or evidence [4]. In this perspective, understanding tacit knowledge in research, and externalising as much of it as possible, helps expanding the boundaries of science by allowing for the definition, discussion, operationalisation, and validation of both what is being studied and how it is studied. This is especially so considering the need to have quantitative analysis go hand in hand with qualitative inquiry, recently brought to evidence by Franco Moretti [16].

To respond to this challenge, we have turned to phenomenography - i.e. the discipline that investigates how learners perceive the phenomena they study [14, 13]. This discipline draws a parallel between learning at the individual level and scientific research, i.e. learning at the collective level [13]. The three patterns of variance identified within phenomenography – contrast, generalisation and fusion – may prove relevant for collaborative interdisciplinary research. Our short paper thus purports to test the viability of the transfer of these notions from phenomenography to bridge humanistic and statistical interpretations in an interdisciplinary research environment. We aim to inquire whether and how these patterns may provide a systematic and reproducible approach towards a visual analysis of bibliographic metadata applied to translation history. Specifically, we define our research question as follows:

RQ: How the phenomenographic three patterns of variance could shape a reproducible method to perform exploratory visual analysis of bibliographic data?

2. Case Study: Translations of Fiction between Italy and the UK

We focus on translations of Italian 20th century fiction into English (UK market) and of British 20th century fiction into Italian in the period 1945-1970[1]. We created two symmetrical datasets of bibliographic data: one collecting Italian titles translated in the UK, the other containing titles in English translated into Italian. For the first dataset, we sourced partial data from Healey [11]. In the second case, no comprehensive bibliography was available, so we built our dataset sourcing data from publishers' trade historical catalogues [9, 6, 10, 21, 3]. We considered these to contain cleaner and more reliable data than the one available through the Italian ICCU[2]. For our purposes we had to enrich these datasets with information that was not provided neither in Healey nor in the trade catalogues, which included the title in the original

---

[1]For the present paper we decided to consider exclusively Italian publishers who (i) also imported titles to the UK and (ii) for whom a complete historical trade catalogue was available. These included major Italian publishers Mondadori, Feltrinelli, Einaudi, Garzanti, Bompiani, whose export activity covers 152 titles out of a total of 215 Italian titles published in the UK. While our choice allowed us to build a representative enough dataset and to focus on exchanges between publishers who both imported and exported titles, While these provide representative evidence, because trade catalogues are available for the major Italian publishers, who were the ones who traded titles with the UK, we are aware that we still need to fully operationalise the retrieval of all titles of British fiction translated by Italian publishers from ICCU.

[2]Central Institute for the Union Catalogue of Italian Libraries and for Bibliographic Information, accessible from the Online Public Access Catalogue of SBN - National Library Service of Italy: https://opac.sbn.it

language, publisher and year of first publication of the source text and was sourced from the ICCU and the British Library Catalogue[3].

We focus on a subset of the collected data, consisting of the bibliographic information of the first edition in the source language, and the first translated edition in the target language. For each pair original-translation, we consider author and title. Furthermore, on both ends we consider publisher and year of publication[4]. A clear limitation of this approach emerges for translations published several years after the original publication, which may imply a change of publisher. Whether the title was acquired from the first or the second publisher is a piece of information that our datasets do not contain. This is only one of the arguments for the need to treat findings of our computational analysis as a provisional result that needs to be verified and explained, as opposed to definitive findings. In this sense we propose to combine quantitative and qualitative analysis: since the latter implies a good deal of tacit knowledge for a hypothesis to be formulated and tested, the challenge is to find a method to obtain traceable provisional results by automatising most operations that do not rely on tacit knowledge. As a consequence, this also supports the discernment of where tacit knowledge might be involved.

3. Analysis of Bibliographic Data

When working with metadata and without a relevant text corpus, recent research in digital humanities has turned towards network analysis [17, 1, 7]. Yet, what still remains important is the ability to interpret in a reproducible manner the atomic elements of such a network: a particular relationship between two nodes (authors, publishers or other entities), or even an individual node in context. Bibliographic data is mostly nominal information and captures only a very limited set of relationships. In the rare cases when relationships between entities exist (e.g. communication between publishers), archives are partial. As a consequence, such relationships are not explicitly represented and often remain as tacit knowledge of humanities researchers. This is particularly true for entities of the same type, the differences among which might expose important patterns to be learned. Thus, our proposed approach is not an alternative, but a complement to network analysis. It could help inform relevant decisions for network modelling.

Educational research has traditionally had strong focus in the communication of knowledge and different possible perceptions. In the 1980s the discipline of phenomenography emerged as the study of the different ways in which phenomena are perceived. Within this discipline, a theory of variation was elaborated, for which successful learning followed three distinct patterns of variance: contrast, generalisation and fusion [13]. Contrast posits that a dimension of the phenomenon under consideration cannot be discerned unless its different values are experienced. Akin to the approach of scientific controlled experiments, the focus on contrast enables the understanding of what changes with the variance of the dimension of interest. When it comes to generalisation, changes in other dimensions of the phenomenon need to be perceived while keeping the dimension of interest invariant. This enables both further understanding of

---

[3]British National Bibliography, accessible from the British Library's main catalogue: https://explore.bl.uk

[4]For each dataset, this results into six-columns tables: two columns are numerical, providing years of publication, and four nominal, containing author, title and the two publishers.

the impact of each value of the dimension of interest and the understanding of the surrounding context of this dimension. Finally, fusion of different dimensions is crucial to understand the interplay between them. This means having a way to experience the independent combinations between the different dimensions under scrutiny. Visualisation plays an important role in communication of knowledge and in fact, these patterns of variance are also familiar in data visualisation. For example, Bach and colleagues present the patterns of comparison and repetition [2] that might be seen as related to contrast and generalisation. However, to understand the dimensions of interest, the phenomenographic patterns need to be experienced incrementally in the above order. Noteworthy, the same aspect of the phenomenon could be a dimension in one context, and a value in another [13].

Bibliographic records are represented as a combination of names (title, author, publisher, etc.), dates (e.g. publication date), identifiers (e.g. ISBN, URL) and possibly other data. In computer science terms, these data are the attributes of each bibliographic record and all attributes for the same data across all records form a dimension. From an analytical perspective, these dimensions are mostly nominal. There are only few formal operations that can be performed on nominal attributes. Without additional taxonomic knowledge of the underlying domain of the attribute (e.g. publishers), all that can be said about two different instances is whether they are equal or not [8]. Such an equality operation allows for the possibility to group and count equal instances. However, further operations typical for quantitative analysis, like ranking any two values or measuring the distance between them, are not well defined [8]. This limitation precludes further analytical operations like ordering or clustering the data.

Partially due to the above reasons, exploratory visualisations have been traditionally used for quantitative [18, 19] and only rarely for nominative data [5]. Moreover, research in data visualisation has explored the dimensionality of visualisations, in which projections of data along a single line is considered to be a visualisation of one dimension, mapping of two data dimensions one against the other is considered to be of two dimensions and so on [19]. Examples of one-dimensional visualisations that could be used for the visualisation of nominal data and its counts are bar charts, line charts and pie charts, of two-dimensional - heatmaps (Figure 2), and of two or more dimensions - diagrams that represent dimension parallelly like alluvial diagrams (Figures 3 and 5) and slope chart diagrams (Figure 6).

4. Proposed Approach

We operationalise variation by applying the three patterns to the defining characteristics we have identified for our research domain: the combination of nominal data and tacit knowledge of contributing scholars. In phenomenographic terms the definition of dimensions and values may be flexible – one aspect could be a dimension in one case and a value in another [13], e.g. a specific publisher can be seen as both one of many, and the dimension in which a specific book series are defined. However, the consideration of such flexibility remains beyond the scope of this paper and subject to future research. Rather, we take a naive approach by considering the dimensions formed by attributes of bibliographic records as phenomenographic dimensions and the corresponding nominal data as the phenomenographic values. Here is how the three patterns of variance could be represented in visualisations:

Contrast. A first step is focusing on a single dimension of interest. This could be performed not only as considering the possible values of this dimension, but also aggregating the other dimensions by counting the distinct values there that correspond to each of the values of the dimension of interest. Even though multiple other dimensions could be aggregated, any resulting charts representing counts are unidimensional.

Generalisation. The second way to experience variation is to fix a dimension to a particular value and visualise a chart juxtaposing other dimensions under this condition. In this case it is beneficial to take advantage of multidimensional representations, visualising two or more complementary dimensions.

Fusion. The final step in experiencing variation is to create a visualisation of the big picture, possibly a type of map that includes all possible data, risking to obscure the possible details of interest. To examine the interplay of several dimensions, they need to be plotted on a graph with high dimensionality.

When it comes to externalising tacit knowledge, we propose that be done by explaining the why and how the patterns of variation were applied. Clearly, this is not exclusive to tacit knowledge, but also could lead to eliciting intuitive, metatheoretical and conditional knowledge.

5. Illustrative Case Study Results

We have tested our approach by visualising[5] two case studies from our datasets of literary transfer. The first one focuses on Heinemann as a sort of wholesale importer to Italy. The second on Vasco Pratolini as the only Italian author in the period who had two distinct publishers translating more than one literary work of his. Heinemann's case was chosen because of the high number of partnerships with Italian publishers, and especially with Mondadori. Besides being Italy's major publisher, Mondadori is the one with most accessible data, which provided an apt test-case for our study. Pratolini's case was chosen because of the author's presence across several publishers' catalogues, which in merely visual terms, is outstanding and therefore elicits questions as to possible interpretations on the qualitative level.

### 5.1. The role of Heinemann in translations to Italy from English

One important research question within literary transfer studies concerns relationships between publishers, a crucial factor for the establishment of translated authors. For the sake of simplicity, we consider a case study contained only in the UK to Italian dataset.

---

[5]We illustrate our case studies with the RAWGraphs visualisation app [15] for two reasons. First, beyond the widerspread chart types for quantitative data, it also features charts that support dimensions containing nominal data. The second useful feature of RAWGraphs is the export of the charts in vector graphics, which allows for easier post-processing.
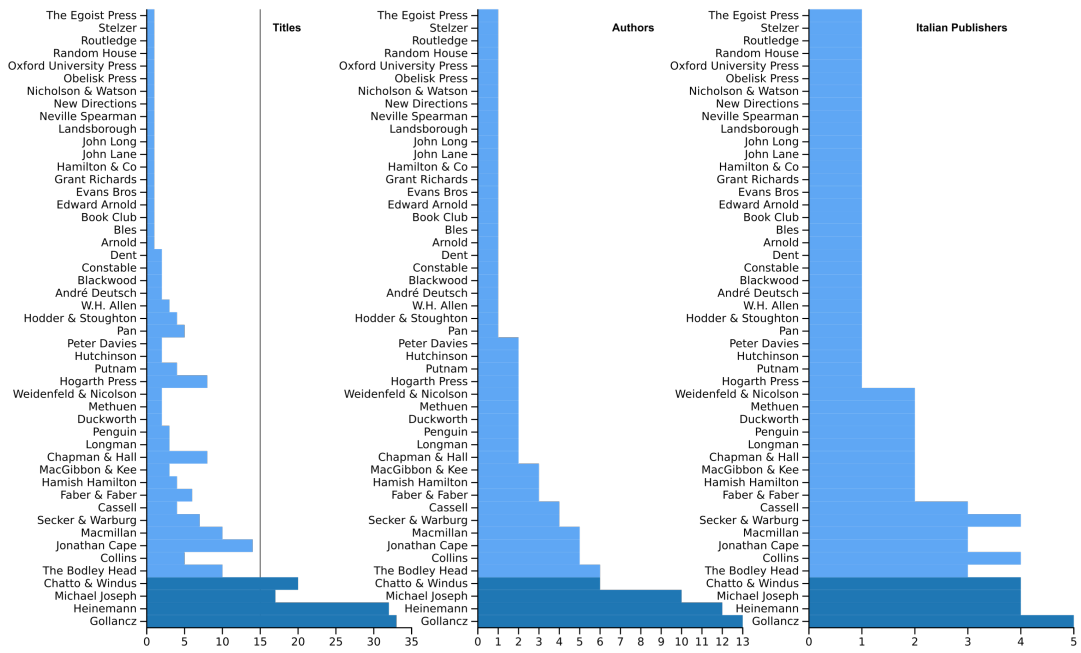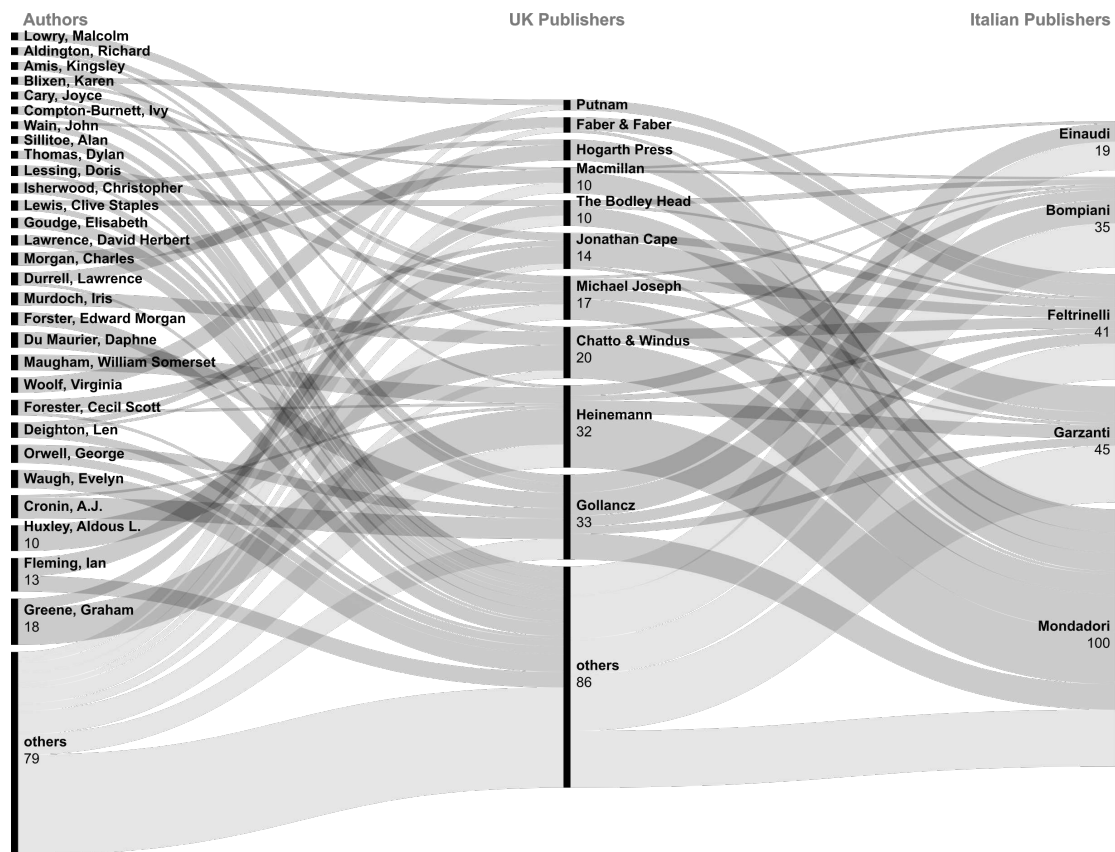
Figure 1: Viewing UK Publishers through the lens of Contrast, respectively aggregating Titles, Authors and Italian Publishers. Also highlighted Publishers surpassing threshold line for Titles count.
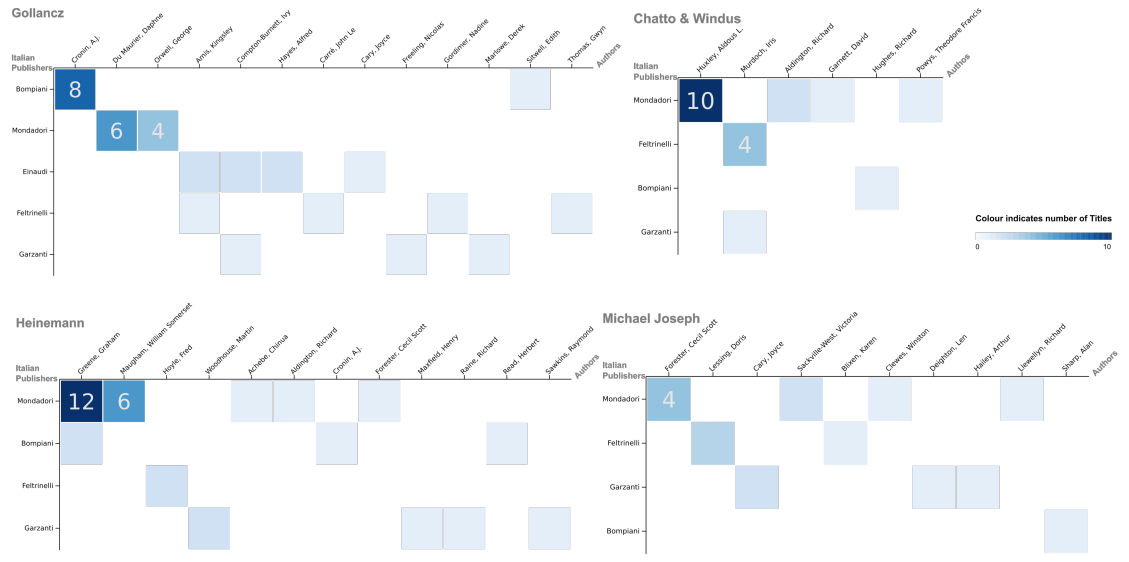
Figure 2: Viewing key UK Publishers through the lens of Generalisation. Counts shown only for values above 3.

A first step is identifying the (quantitatively) big players in the process. To employ contrast, we look at all UK publishers involved in literary transfer to Italy. This view shows us that the market of exports to Italy is very segmented with only 4 out of 49 publishers having translated more than 15 books each (Figure 1). Yet these 8% of publishers are cumulatively responsible for 43% of studied translations. Focusing on these four key UK publishers, in a further generalisation step we look at authors and Italian publishers of translations by this UK publisher (Figure 2). In line with contextual knowledge that Heinemann is close to exclusive publisher of G. Greene, W. Somerset Maugham, and others, here it is visible that these authors form an important part of Heinemann's exports and correspondingly that there is a strong partnership between Heinemann and Mondadori. Finally, fusion through an alluvial diagram allows us to visualize the role of Heinemann in the big picture (Figure 3) - as one of the two strongest exporters to Italy thanks to its important role in the UK and strong relationship with one of the big players in Italy. Whereas Gollancz outperforms Heinemann in numbers of translated titles, it appears less persistent with its partnerships, both with authors and Italian publishers.

Figure 3: Viewing most translated UK Authors, and their UK and Italian Publishers through the lens of Fusion. Shown only Authors of at least 3 Titles are shown, Counts shown only for values of 10 or more.

A further step in this investigation would be to engage in a study of archives to understand whether Heinemann lost its role on the Italian market in 1961 when it was acquired by Tilling and was left by Greene and others [20], but this is beyond the scope of this paper.

5.2. UK Publishers' Reception of Italian Authors

Another question that helps understand the role of publishers is what authors were successful when not working with the same publisher.

This leads us to consider Italian authors that have been translated by more than one publisher (Figure 4), a case of applying contrast on the authors. We consider all five authors that have at least four works translated and at least two publishers, for each of the Italian and UK markets. This leads us to generalise over these authors and compare how their translations were published (Figure 5). This reveals that Pratolini is the only author that was published twice by each of two UK publishers. This suggests that Italian publishers were not the ones that defined these consecutive relationships. Rather a possible interpretation could be that Vallecchi had a role in opening the door to Pratolini to diverse UK publishers. Finally, to visualise the potential impact of this limitation alongside with the changes of UK publishers, we employ fusion to display the Italian authors that have four or more translated titles showing the publication years by country and the corresponding UK publishers (Figure 6). Here we see that there was a particularly big gap in some translations of Pratolini.

Archival research into the communication between editors reveals that it was not the first
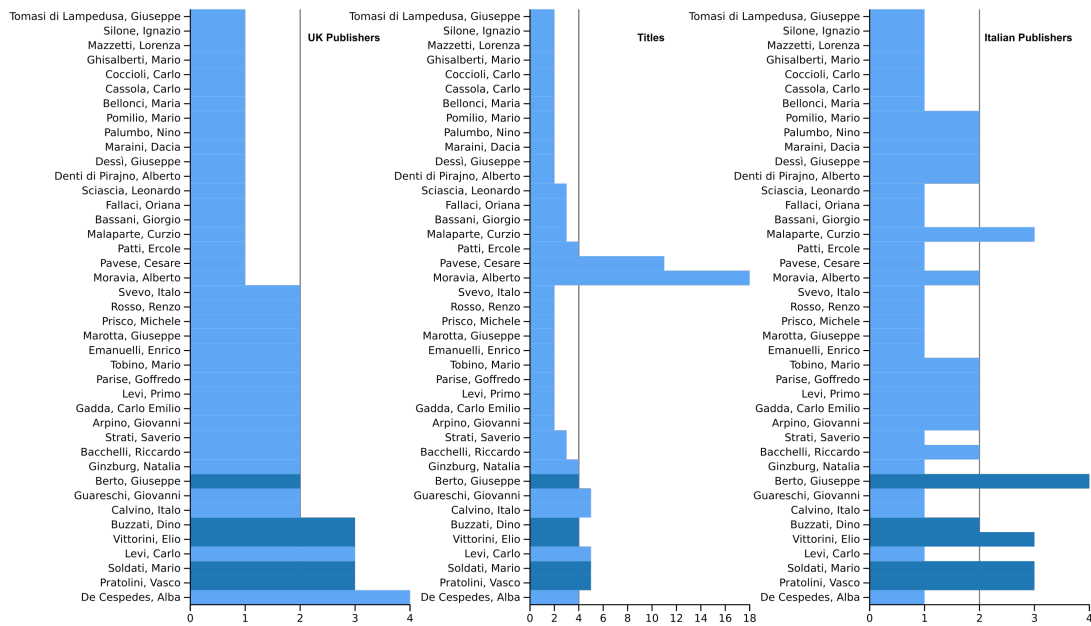

Figure 4: Viewing Italian Authors with at least two translated Titles through the lens of Contrast. Also visible threshold lines and highlighted Authors reaching all of them.

publisher (Vallecchi) that managed the translation on the Italian side. This illuminates an issue that still needs to be addressed. Whereas for short periods between first publication in original language and the first translation, it might be realistic to assume that it was the first publisher that managed the translation, when more time passes this becomes more complicated, and we need to also include subsequent Italian reprints and editions in our analysis.

6. Conclusion

In this paper we proposed a new approach towards visually analysing bibliographic data, addressing two challenges that we identify to be critical: the nominal character of bibliographic data and the tacit nature of complementary knowledge of scholars in translation studies and in the humanities in general. One could argue that the steps of contrast, generalisation and fusion are intuitive and thus already widely applied. Yet, it is the systematic use of the combination of these that makes the approach conform with phenomenography. To validate our proposed approach, we apply it to two case studies from literary translation between Italy and the UK in the post-war period of the 20th century. We approached simplistically the mapping between the conceptual terms of phenomenography and the constructed objects of tabular data representation. This interpretation is overly limiting, e.g. research in cognitive load [12] suggests that people are capable of mentally constructing domain-specific schemas that can potentially span across multiple descriptive dimensions. This opens to the possibility that for example a phenomenographic value could include a combination of values across bibliographic fields, possibly also including aspects of tacit knowledge, e.g. considering the partnership between UK and Italian publishers as a dimension, incorporating corresponding researcher knowledge.

The creation of a research environment where data is sourced and analysed in collaboration between scholars in the humanities and in computer sciences right from the start is for us a priority. To do this, it is crucial to negotiate a set of questions that are relevant on both sides
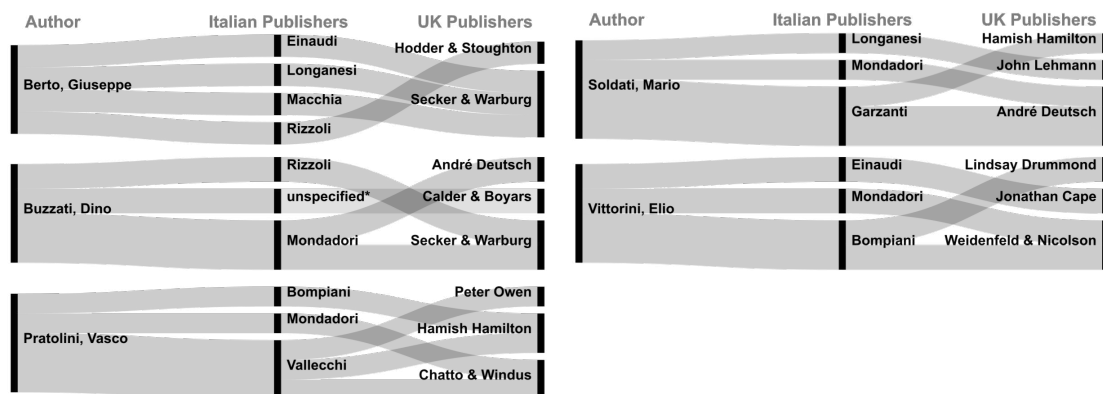
Figure 5: Viewing selected Italian Authors and their Italian and UK Publishers through the lens of Generalisation.
* one of the translations of Buzzati is a collection, so no single Italian Publisher can be specified.
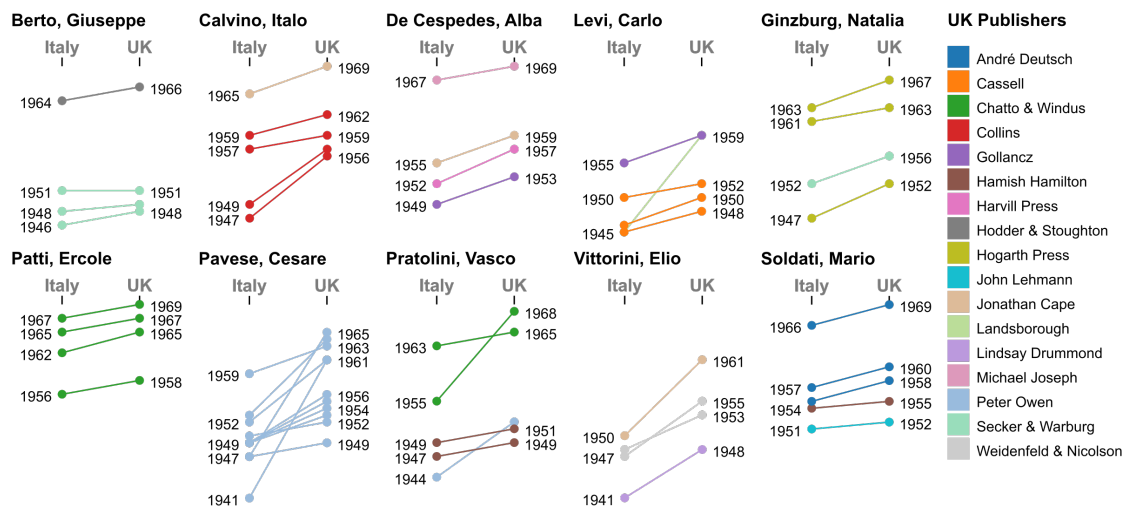
Figure 6: For presentation efficiency, viewing only a curated selection of highly translated Italian Authors, their Italian and UK publishers and the dates of the respective first publications through the lens of Fusion. The colours indicate UK publishers.

and likely to contribute to both quantitative and qualitative analysis, since neither of these, alone, can impact the way of writing translation history.

Acknowledgments

References

[1] R. Ahnert, S. E. Ahnert, C. N. Coleman, and S. B. Weingart. "The Network Turn: Changing Perspectives in the Humanities". In: *Elements in Publishing and Book Culture* (2020). DOI: 10.1017/9781108866804.

[2] B. Bach, M. Stefaner, J. Boy, S. Drucker, L. Bartram, J. Wood, P. Ciuccarelli, Y. Engelhardt, U. Köppen, and B. Tversky. "Narrative Design Patterns for Data-Driven Storytelling". In: *Data-Driven Storytelling*. Boca Raton: A K Peters/CRC Press, 2018, pp. 107–133.

[3] L. Bianco and A. Borsani. *Le edizioni Einaudi: 1933-2018*. I millenni. Torino: Giulio Einaudi, 2018.

[4] J. B. Biggs. "Modes of learning, forms of knowing, and ways of schooling". In: *Neo-Piagetian theories of cognitive development: implications and applications for education*. Ed. by M. S. Andreas Demetriou. London: Routledge, 1994, pp. 30–51.

[5] J. Blasius and M. J. Greenacre. *Visualization of categorical data*. San Diego: Academic Press, 1998.

[6] Bompiani, ed. *Catalogo generale Bompiani, 1929-2009*. Milano: Bompiani, 2009.

[7] L. Bowker. "Digital humanities and translation studies". In: pp. 46–53.

[8] M. Chen, M. Feixas, I. Viola, A. Bardera, H.-W. Shen, and M. Sbert. *Information theory tools for visualization*. AK Peters visualization series. Boca Raton: CRC Press, 2016.

[9] Feltrinelli, ed. *Feltrinelli 60, 1955-2015: catalogo storico*. Milano: Feltrinelli, 2015.

[10] Garzanti, ed. *1975, Catalogo generale Garzanti*. Milano: Garzanti, 1976.

[11] R. Healey. *Twentieth-century Italian literature in English translation: an annotated bibliography, 1929-1997*. Toronto: University of Toronto Press, 1998.

[12] S. Kalyuga and A.-M. Singh. "Rethinking the Boundaries of Cognitive Load Theory in Complex Learning". In: *Educational Psychology Review* 28.4 (2016), pp. 831–852. DOI: 10.1007/s10648-015-9352-0.

[13] F. Marton. *Necessary Conditions of Learning*. Hoboken: Taylor and Francis, 2014.

[14] F. Marton. "Phenomenography–A Research Approach to Investigating Different Understandings of Reality". In: *Journal of Thought* 21.3 (1986), pp. 28–49. URL: https://www.jstor.org/stable/42589189.

[15] M. Mauri, T. Elli, G. Caviglia, G. Uboldi, and M. Azzi. "RAWGraphs: A Visualisation Platform to Create Open Outputs". In: *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*. Cagliari: Acm, 2017, pp. 1–5. DOI: 10.1145/3125571.3125585.

[16] F. Moretti. *Falso movimento: la svolta quantitativa nello studio della letteratura*. Extrema ratio. Milano: Nottetempo, 2022.

[17] D. Roig-Sanz. "Global translation history: Some theoretical and methodological insights". In: *Translation in Society* 1.2 (2022), pp. 131–156. DOI: 10.1075/tris.22010.roi.

[18] J. W. Tukey. *Exploratory Data Analysis*. Reading: Pearson, 1977. URL: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20%5C&path=ASIN/0201076160.

[19] M. Ward, G. G. Grinstein, and D. Keim. *Interactive data visualization: foundations, techniques, and applications*. Boca Raton: CRC Press, 2011.

[20] W. West. *The Quest for Graham Greene*. New York: St. Martin's Press, 2002.

[21] M. Zerbini and P. Moggi Rebulla. *Catalogo storico Arnoldo Mondadori Editore 1912-1983*. Milano: Mondadori, 1985. URL: https://www.fondazionemondadori.it/pubblicazione/catalogoi/.

[22] X. Zhou and S. Sun. "Bibliography-based quantitative translation history". In: *Perspectives* 25.1 (2017), pp. 98–119. DOI: 10.1080/0907676x.2016.1177100.