**Research Proposal: The Hebrew Book: From Manuscripts to Printing**

**Background**

The Hebrew book has charted a fascinating and circuitous path from manuscripts through printing to today’s digital books. The goal of this study is to consider different aspects of the development of the Hebrew book, focusing in particular on the transition from the manuscript period to the era of printing. The printing revolution occupies a special place the history of the Hebrew book, and printing has had a particular influence on Jewish culture. However, few studies have been devoted to the connections between the printed Hebrew book and Hebrew manuscripts. These include research by Abraham Berliner (1949), who studied the influence of the first printed books on the structure of religious life. Malachi Beit-Arié, a scholar of Hebrew paleography and codicology, dealt with the relationship between Hebrew incunabula and manuscripts (1976). Zeev Gris, also a scholar of the Hebrew book, dealt with the topic in his many publications, especially his article “Tradition and Change in the Transition from Manuscript to Print” (2015).

Traditional methods for the analysis of historical books and manuscripts usually focused on a qualitative examination of philology, including textual criticism and critical editions on the one hand, and paleography and codicology on the other. This study will integrate other research methods, such as the “livre et société” approach (Baruchson, 1993; Elyada, 1999), which views the book as an important source for understanding historical processes and social and cultural changes, and quantitative methods common in the study of the digital humanities.

***Livre et Société***

Generally, historical research deals with “macro” history — the “dramatic” events, such as wars, revolutions, and natural disasters, that impacted the course of history. However, the course of history also includes events on the “micro” level, events that touch on everyday life and which had a profound influence on cultural, social, political, and economic processes. “Livre et société,” (literally, “book and society”) is a branch of research was developed in France in the late 1950s by the historians Henri-Jean and Martin Lucien Febvre. This school emphasizes the connection between sources and scholarship in bibliography and cultural history and the social sciences. In other words, the book is not only studied as a subject in its own right, but also as an important source for understanding social and cultural processes and changes. The search for answers is aided by the sources, by scholarship and by interdisciplinary research methods (Baruchson, 1993). The “Livre et Société” approach deals with the “micro,” the everyday, and seeks to understand these processes through the history of the “book.” The multiple lenses through which the subject is studied raise a variety of questions: What is the role of the book in a certain society? To which degree is a certain public interested in written works in general, or a certain work in particular? These and other questions can reveal the social, economic, and cultural world of the society being investigated, and the changes it undergoes over time, alongside other variables: geography, customs, mentality, art, ideas, local and world historical events, and others.

In the study of the Hebrew book, too, there have been a handful of studies that have made use of the “livre et Société” method as a means for understanding cultural and historical processes. These include the historian Isaiah Sonne, who dealt with the connections between bibliographic phenomena and historical events (Sonne, 1950). Shifra Baruchson undertook a statistical study of the bibliographic lists of books included in the private libraries of the Jews of Manuta, Italy, in 1595, and found that the purchase of books in this period was, first and foremost, for utilitarian and scholarly purposes, and not only for cultural or social interest. It is also important to note the wide-ranging studies of Zeev Gris on Hasidic moral literature (Gris, 1989), on the history of the Hassidic book (Gris, 1991), on the world of Jewish books between the seventeenth and nineteenth centuries (Gris, 2002), and his recent comprehensive study on the history of the Hebrew book (Gris, 2015). Esther Kandelshein has also dealt with Hebrew printing in the physical sciences between the fourteenth and eighteenth centuries and with Hebrew printing in Palestine between 1577 and 1922 (Kandelshein, 2004). We should also note the studies by Hagit Cohen on Jewish bookstores in Eastern Europe in the second half of the nineteenth century (Cohen, 2005), and of Anat Guetta, who focused her research on printed books in the shin years as a source for the study of intellectual life in Jewish societies (Guetta, 2002).

***Digital Humanities***

Recent years have seen the development of new research avenues in the humanities. We live in a new and dynamic era that began with the invention of the computer in the first half of the twentieth century and continued with the invention of the internet, innovations that surpass even the printing revolution. This situation has also impacted scholarship. Humanities Computing, later known as Digital Humanities, is a new academic discipline, interdisciplinary by nature, that began in the first days of digital computing but developed especially in the 1990s with the World Wide Web, and which has continued to expand in recent years in Israel and abroad. The transition from Humanities Computing to Digital Humanities is not only semantic; it also indicates a substantial change. In the first stage, digital tools were developed to aid the scholar in his or her research, and to make it more efficient: textual corpuses and databases were built, which mostly served as the basis for quantitative studies, but research continued along established methods. The second stage saw the development of new research horizons that had not been possible before, and concepts such as “distance reading” (Moretti, 2005), “digital edition,” “big data,” and “visualization” were developed.

The term “digital humanities” refers to different computer applications that are used in research or teaching in the fields that are defined as part of the humanities. Many scholars have addressed the definition of the term. In *A Companion to Digital Humanities*, published in 2004, which coined the phrase “digital humanities” as part of the academic vocabulary, the goals of digital humanities were defined as follows:

“Using information technology to illuminate the human record, and bringing an understanding of the human record to bear on the development and use of information technology.”9

The discipline of digital humanities has become more and more popular and a focal point of academic activity. Today, there are hundreds of research centers devoted to the digital humanities around the world, the subject is studied at different levels at universities worldwide, and conferences and several academic journals are devoted to the topic.

***Digital Humanities and the Study of the Hebrew Book***

Important research in digital humanities was undertaken in Israel years before the field was recognized by its current name. In particular, one can point to the Responsa Project, which began in 1963 at the Weizmann Institute and continued under the auspices of Bar Ilan University. The first edition of the project was launched in 1967, and in 2007 the project was awarded and Israel Prize for Jewish Literature. Other projects include the historical Hebrew dictionary by the Academy for the Hebrew Language, the Friedberg Geniza Project, which has expanded to become a portal for projects on Jewish books and manuscripts, and the extensive activity of the Judaica Europeana project in Israel.

The National Library of Israel (NLI) occupies a central position, and is heading a number of large and important initiatives in the framework of the Library’s digitization activities. For example, the Ketubbot Collection, a unique website devoted to the Jewish *ketubbah*. The site includes more than 4,200 *ketubbot* from dozens of different collections around the world. The Ketubbot Collection of the NLI includes almost 1,800 original *ketubbot* from the entire Jewish disaspora. The concentration of the *ketubbot* on a single website allows for a deep, comprehensive, and broad view of the *ketubbah* as a Jewish document, as a Jewish work of art, and as an invaluable historical source. Similarly, the Historical Jewish Press project is a website that includes Jewish newspapers published in the past in different languages, countries, and periods. The electronic versions of the newspapers enable scholars to view the periodicals in their original appearance, as well as a powerful search tool that scours every word published in a periodical throughout the years of its publication. With Ktiv — the international collection of digitized Hebrew manuscripts — the National Library has updated its collection of photographed Hebrew manuscripts. The initiative aims to make Hebrew manuscripts accessible to people all over the world. The images of the manuscripts are preserved and presented with the most advanced technology and at the highest resolution, and will serve communities of scholars and readers.

Other initiatives include Dicta, a project that applies cutting-edge machine learning and natural language processing tools to the analysis of Hebrew texts (<http://dicta.org.il/)>, and the Digital Humanities department at Bar-Ilan University, which has sponsored projects such as the Multi-dimensional Ontology for Hebrew Proverb Search, and Towards a Cross-generation Social Network for Jewish Sages (<http://digitalhumanities.org.il/)>.

**Research Objectives and Expected Significance**

This project builds on our previous research. The project Towards the Ontopedia for Post-Medieval Hebrew Manuscripts, funded by the National Science Foundation (grant 342/15), developed an event-based ontology model for the analysis of manuscripts (Zhitomirsky-Geffet and Prebor, 2016). The ontological data model was constructed on the basis of the *livre et société* approach. The underlying philosophical perspective considers manuscripts and books as “living entities” and designs a data model of its narrative (fig. 1).

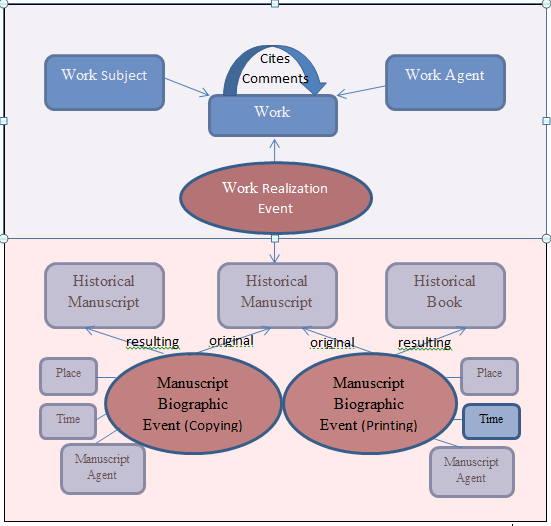


Figure 1: The ontology model of the work reproduction data.

The model includes the stages and milestones in the biography of a text, such as composition, copying, acquisition, and printing — one of the most important events in the history of a manuscript — and its influence on and interactions with other manuscripts, books, people, places, and historical and cultural events. This sequence of events and places constitutes a timeline of history on which manuscripts, books, individuals and their relationships can be placed. Using this data model, scholars can answer important research questions: How many works on certain subjects were copied or printed in a certain period in different countries? Who was involved in their creation and distribution, and how were these figures related to each other, if at all? What historical events influenced these individuals and their work?

This earlier study (grant 342/15), the results of which will be published in a series of articles now undergoing peer review, dealt primarily with questions of distance reading, related to big data and broad questions. In the proposed project we intend to continue our focus on a number of questions related to distance reading (Moretti, 20015). However, we will not deal with the entire corpus, but instead with selected portions of it, and will add close reading of popular works and authors. The current study will examine additional segmentation regarding authors whose works were widely distributed. These include:

Which works or portions of them were most copied? Where were these copies made? What information about the owners of the manuscripts have reached us? In this way, it is possible to learn where and when certain works or portions of them were most in demand, and to attempt thereby to reach conclusions about their use: Was it for theoretical study, for halakhic rulings, or for other purposes? And what was the fate of this author or work after the advent of printing? Did a certain work, which was popular in manuscript form, remain so after being printed? Are there regional and chronological differences?

We believe that the results of our proposed project will contribute to the study of Hebrew books and manuscripts and of Jewish cultural heritage

**Detailed Description of the Proposed Research**

***Working hypothesis***

The study of the history of the Hebrew book in Jewish society, and of its “republic of letters” across generations, can reveal cultural, literary, social, religious, and economic features of Jewish life over the course of generations, and adds an important layer to the understanding of the “People of the Book.”

***Experimental Design and Methods***

In stage one, we will define the corpus of the study and which works and figures will be included in the project. The information will be collected on the basis of the data that were included in the previous study (grant 342/15). In the first stage, the data on the twenty most popular authors will be collected (fig. 2). The manuscripts associated with each individual will be listed in the database. These associations include: author, editor, commentator, commentatee, etc. Other data regarding the manuscripts, gleaned from the existing database, will include such details as the date of the composition of the manuscript, the place of composition, and similar information. For this purpose, we will require the assistance of a programmer to retrieve the relevant entries and to build the new database. Despite the fact that the data was already compiled from the catalog, there is still a need for human labor to improve the entries as much of the information was written in the catalog in a non-uniform way (Zhitomirsky-Geffet and Prebor, 2016).

In the second stage, each figure’s works will be considered separately. For example, in examining the distribution of the number of manuscripts according to author, we found that the most popular author is Maimonides, with 2,279 manuscripts. In this study, we will examine other distributions of Maimonides’s manuscripts, as well as those of other widely-read authors: Which work or portion of a work was most widely copied? Where were the copies made? Our goal is to understand where and when certain works or portions of them were most in demand, and thereby to attempt to reach a conclusion as to whether a book was used for theoretical study, halakhic decision making, or other purposes.

In the next stage, we will add data on the relevant printed verions of the works discussed in the study, focusing on the first centuries of printing (until 1800). In order to locate the books, we rely on the database of the Bibliography of the Hebrew Book, which includes the national bibliography of the Jewish people, comprising detailed entries for all of Jewish literature printed in Hebrew letters, of any genre and in any language. The bibliography includes most of the books, pamphlets, and periodicals published in Hebrew, Yiddish, Ladino, Judeo-Arabic, and other Jewish languages found in Israeli and international collections, from the dawn of Hebrew printing (around 1460) to 1960 (<http://www.hebrew-bibliography.com/)>. The Treasury of the Hebrew Book, a commercial digital database of Hebrew bibliography, contains 90,000 entries of books and individual pages printed in Hebrew characters from the earliest printed books in 1469 to 1948. This database was edited by Yehoshua Winograd and Moshe Rosenfeld.

Comparing the data between manuscripts and printed books can teach us a great deal about the culture of reading and study of the Hebrew book in different Jewish communities. The transition from manuscripts to print was a true revolution in general, and in Jewish society in particular. The print revolution caused an increase in the quantity of books, and thus affected access to knowledge and its distribution among new populations, sparking new ideas and innovations, and expanding the circle of those who learned by reading (Einstein, 1979; Eliav-Feldon, 2000; Gris, 2002, 2015).

In the final stage of the project we will compare the first printed books — Hebrew incunabula — to manuscripts in order to better understand the transition between the periods. We will examine the introductions to the books and the people mentioned in them, and will compare the printed books to the manuscripts that have survived.

**Preliminary Results**

We found that there were 117,526 Codicological Units (CUs) in the catalogue of the National Library of Israel. Only 23,177 (19.72%) of the CUs were accurately dated. An additional 77,710 (66.12%) CUs had an estimated date or a date range. There were 44,338 unique individuals involved in the life cycle of the manuscripts’ corpus, each of whom was assigned one or more of 12 roles: author, scribe, owner, etc. (see fig. 4). 9,211 were authors of 55,513 manuscripts, and the average number of manuscripts per author was 6.16. The other 62,013 manuscripts were unauthored. These manuscripts were mostly copies of the fundamental Jewish texts such as the Bible, the prayer book, the Mishnah, and the Talmud. Table 1 summarizes the descriptive statistics of the manuscript entities extracted from the catalogue fields.

|  |  |
| --- | --- |
| **Entity data** | **Quantity** |
| Number of manuscript CU records | 117,526 |
| Number of accurately dated CUs (precision <= one year) | 23,177 |
| Number of inaccurately dated CUs (precision > one year) | 77,710 |
| Number of undated CUs | 16,639 |
| Number of authored manuscripts | 55,513 |
| Number of unauthored manuscripts | 62,013 |
| Average number of subjects per CU | 1.46 |
| Average number of authors per CU | 1.02 |
| Average number of persons involved per CU | 2.18 |
| Number of unique persons involved in CU lifecycles | 44,338 |
| Average number of CUs per author | 6.16 |
| Number of role types in the corpus | 12 |
| Number of identified autographs in the corpus | 999 |
| Number of distinct locations mentioned in the corpus | 3,605 |
| Average number of locations mentioned per CU | 0.35 |
| Number of CUs with a known place of copying | 16,032 |
| Number of persons with known life periods | 2,774 |
| Number of persons with known geodetic data | 3,009 |
| Number of events | 569,949 |
| Average number of events per CU | 2.18 |
| Average number of persons involved per event | 0.16 |
| Number of events with a known person | 91,632 |
| Number of events with a known location | 204,004 |
| Number of events with a known date range | 182,325 |

Table 1: Descriptive statistics of the manuscript corpus extracted from the catalogue fields

In addition, missing data (mostly dates and locations) was completed by inference from the constructed ontology and from external resources (Table 2). For example, there were 2,335 additional CUs whose place of copying was derived from known sites of scribal activity. More information about dates and locations was added from the Virtual International Authority File (www.viaf.org). 10,867 personal names in the catalogue were also found in VIAF.

|  |  |
| --- | --- |
| **Inferred entity data** | **Quantity** |
| Number of person life periods inferred from CUs | 6,750 |
| Number of persons with geodetic data derived from CUs | 3,009 |
| Number of person life periods extracted from VIAF | 1,494 |
| Number of people with geodetic data extracted from VIAF | 9,035 |
| Number of copying events whose place of copying  was derived from known sites of scribal activity | 2,335 |
| Number of composition events whose date of creation derived from author life periods | 36,277 |
| Number of composition events whose location of composition  derived from known sites of author activity | 27,056 |
| Number of archiving events whose location was derived from CUs current locations | 114,983 |

Table 2: Descriptive statistics of the manuscript corpus data completed by inference from the constructed ontology and from external sources.

In the following sections we present some the results obtained from diachronic and geospatial analysis of individual events as well as from cross-event relationships.

***Distribution of Authors***

9,211 individuals have been identified as authors of manuscripts in the corpus, and the average number of manuscripts per author is 6.16. The most popular authors are Maimonides (1138-1204) with 2,340 CUs, Hayyim ben Joseph Vital (1542-1620) with 1,696 CUs, and Abraham Ibn Ezra (1089-1164) with 1,405 CUs.

Each of these figures deserves a brief introduction. Maimonides was one of the most prolific, influential, and admired Jewish scholars of all generations and one of the most important philosophers of the Middle Ages, as well as being a physician and a communal leader. Hayyim ben Joseph Vital was one of the leading kabbalists of 16th century Safed. He began his study of Kabbalah at the age of 20, and was a student of the well-known kabbalist Rabbi Moshe Cordovero. Later, with the arrival of Rabbi Isaac Luria in Safed, Vital became his leading student. He was a central figure in the development of Kabbalah overall, and in the adoption of Isaac Luria's teaching in particular. Ibn Ezra was one of the most distinguished Jewish Bible commentators and philosophers of the Middle Ages.

Interestingly, we also find two non-Jewish authors among the top twenty figures: Averroes (1126-1198), a Muslim physician and philosopher who wrote commentaries and criticism of Aristotle and who influenced European philosophy in the Middle Ages, and Aristotle himself. The distribution of the 40 most popular authors is demonstrated in Figure 2.

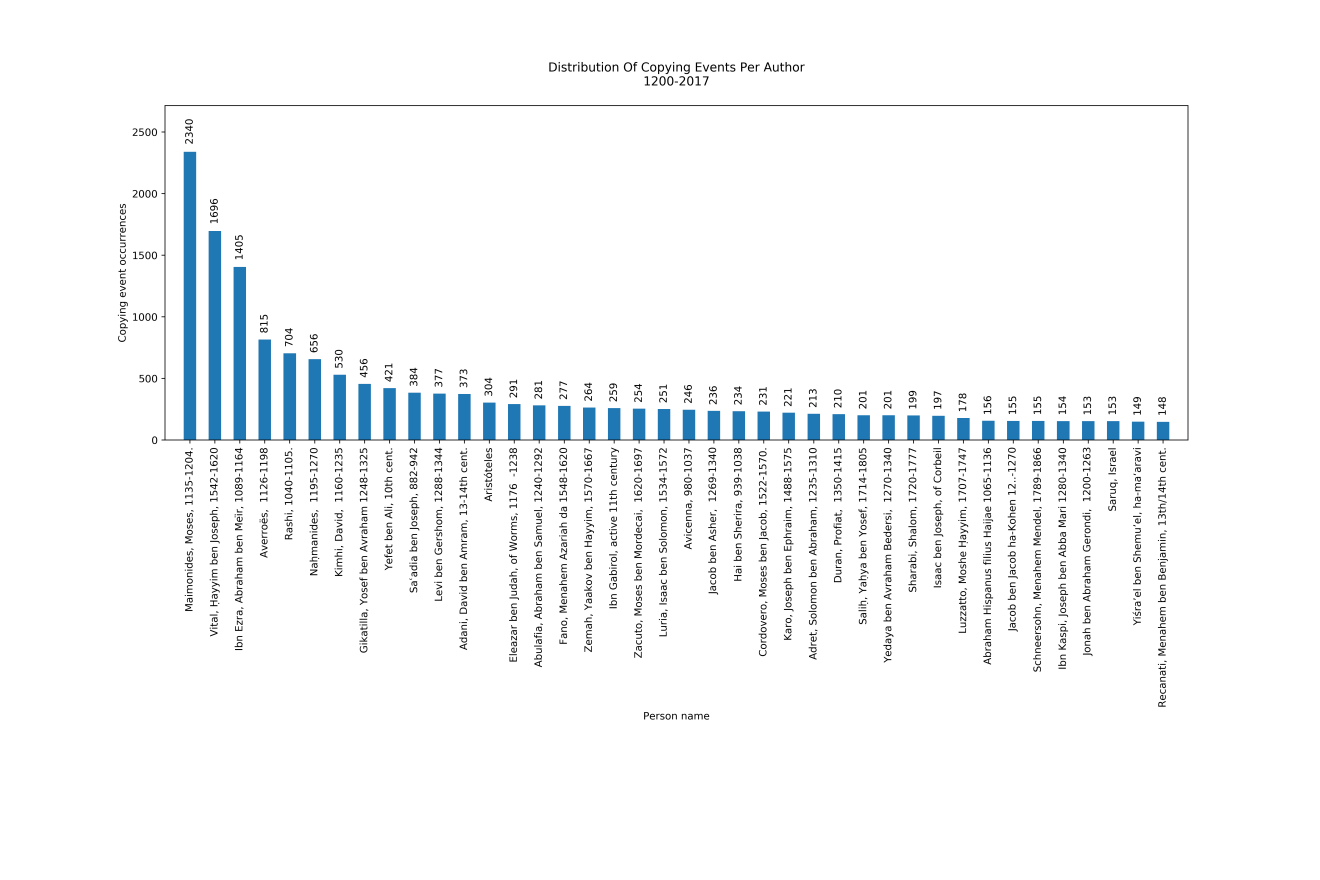


Figure 2: The number of manuscript copies for the 40 most popular authors in the corpus.

***Distribution of Subjects***

The Hebrew manuscript collection contains works that deal with a wide spectrum of topics, including almost all the main subjects of Hebrew literature, such as Jewish law and philosophy, as well as the sciences, Christianity, and other religious topics. Each manuscript includes, on average, 1.46 subjects. Figure 3 represents the distribution of subjects among the manuscripts.

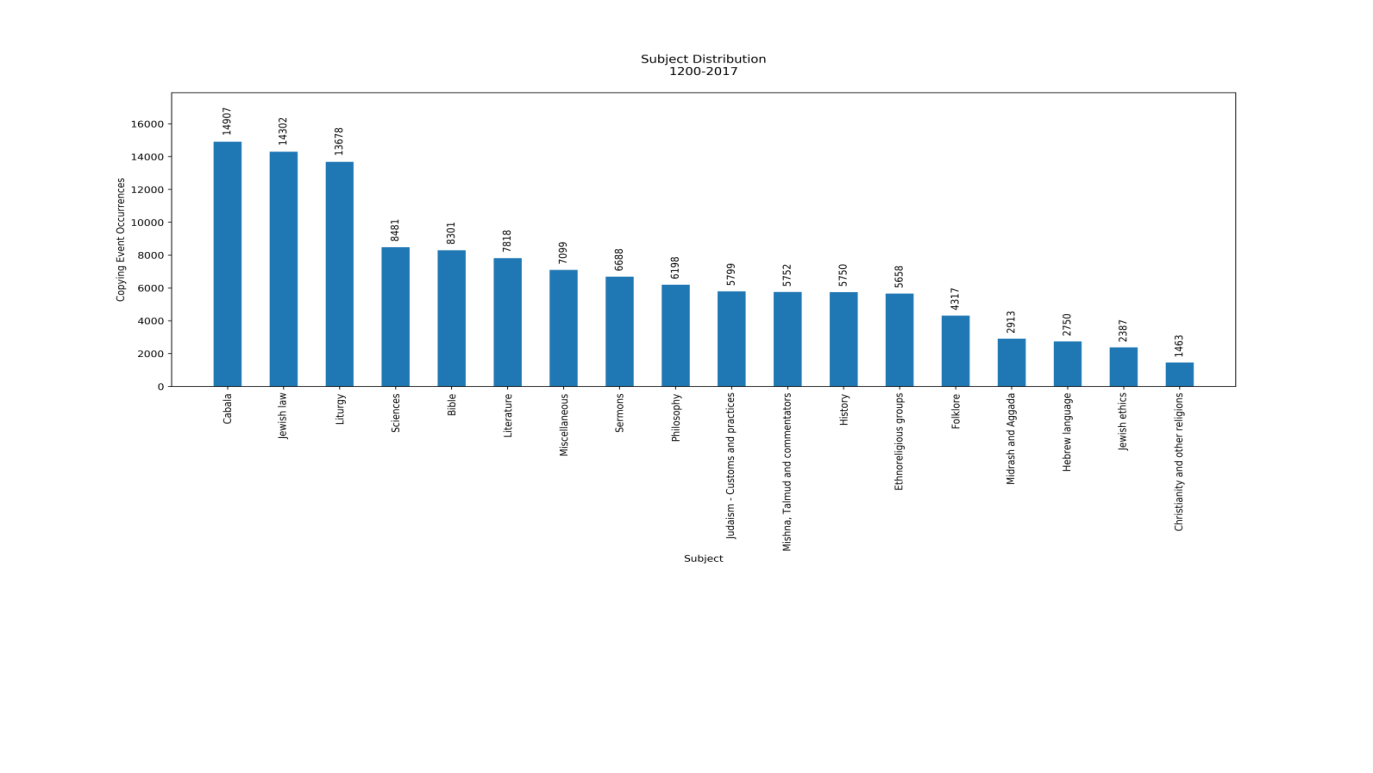
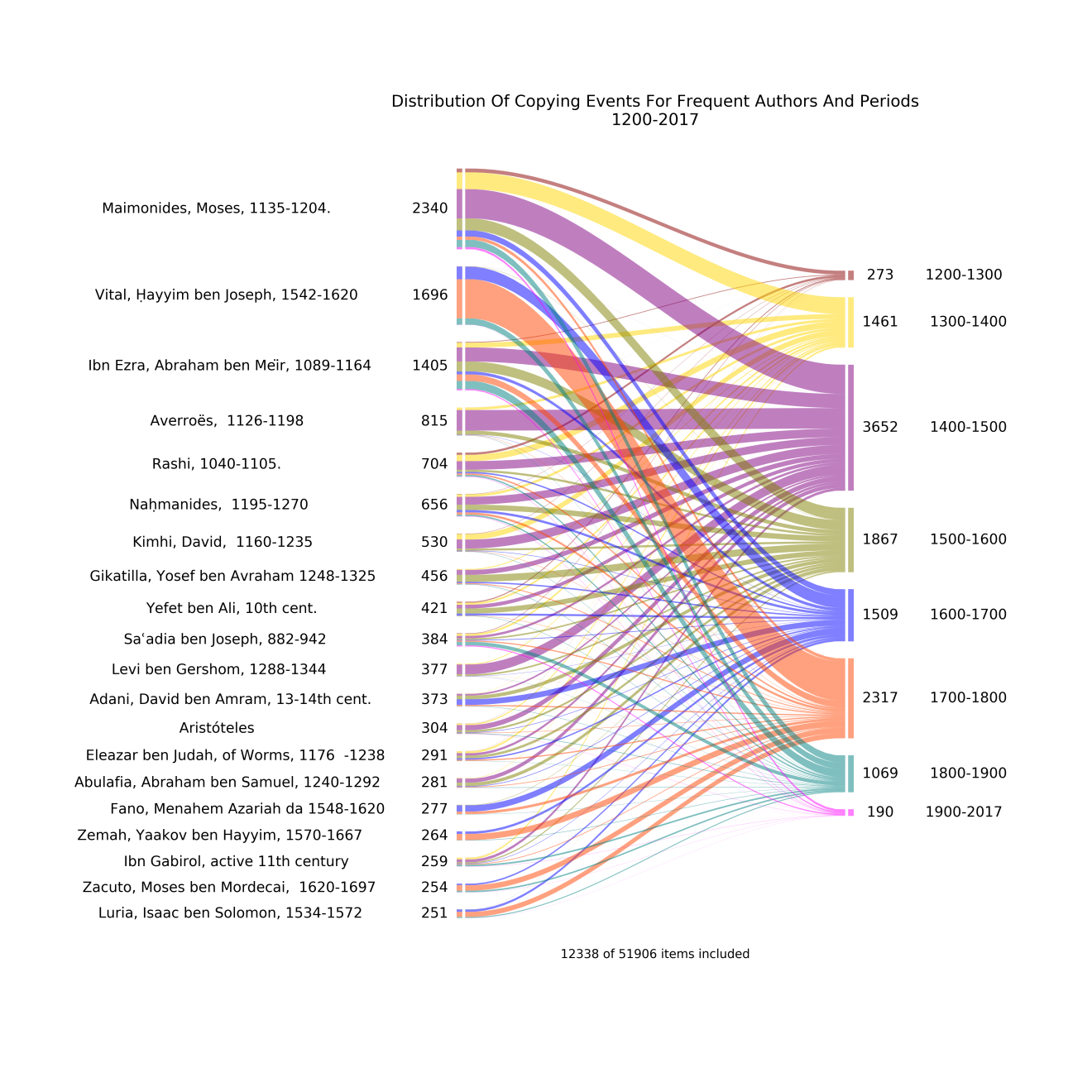
The most popular subjects in the manuscript corpus are Kabbalah, Jewish law, and liturgy, with 14,907, 14,302, and 13,378 CUs, respectively. At the bottom of the list we find Hebrew language with 2,750 CUs, Jewish ethics with 2,387 CUs, and Christianly and other religions with 1,463 CUs.

Figure 3:The full distribution of all subjects contained in the manuscript corpus.

**Composition event vs. copying event: The most copied authors in different time periods and countries**

Authors are directly linked with the composition events, while manuscript copying date is linked with the copying event in the ontology and both events are cross linked through their associated manuscript. Following the blue arrows in Figure 1 we could, therefore, retrieve the information required for this task. As can be observed from the bi-partite graph in Figure 4, the most copied author in the Hebrew manuscripts is Maimonides, with 2,340 copying events. But when we examined the distribution of authors according to copying event dates, we found that most authors experienced periods of greater and lesser popularity. These changes in popularity also depend on when they lived and their relative chronological proximity to the invention of printing. Most of the manuscripts of Maimonides’ writings were copied between 1300 to 1500, before the advent of Hebrew printing. But for the second most popular author, Hayyim ben Joseph Vital, the majority of the copying events were in the years 1700 to 1800, well after the invention of printing. There could be several explanations for this fact. Vital lived in the Holy Land, which lacked a printing press during this period (Kandelshein, 2004). Many of his writings also deal with Kabbalah, which is considered an esoteric teaching in Judaism that is passed down from teacher to student rather than widely distributed, and for this reason Kabbalistic works may have been printed less than writings on other subjects.

Figure 4: A bi-partite graph of the diachronic distribution of copying events for the twenty most popular authors.

The distribution of authors by copy event location is presented in Figure 5. Some authors were popular in many countries, such as Hayyim Vital and Ibn Ezra. Works by Maimonides were also copied in many different countries, foremost among them Italy and Yemen. This is not surprising since the connection between the Jews of Yemen and Maimonides is well known. Maimonides wrote his famous *Iggeret Teman* (Epistle to Yemen) (1173) in response to an inquiry by the head of the Jewish community in Yemen. The exchange of letters was occasioned by a crisis that affected the Jews of that country when many were forcibly converted to Islam; after their appeal, Maimonides helped abolish the decree by the Ayyubid dynasty, and the Jews of Yemen accepted the authority of Maimonides.

We also see authors whose copying events were concentrated in one country, seemingly the country where they lived. For example, works by Samuel Landau (1750-1834) in Bohemia; Shalom Sharabi (1720-1777) in the Land of Israel; and David ben Amram Adani (13-14th) in Yemen.‎

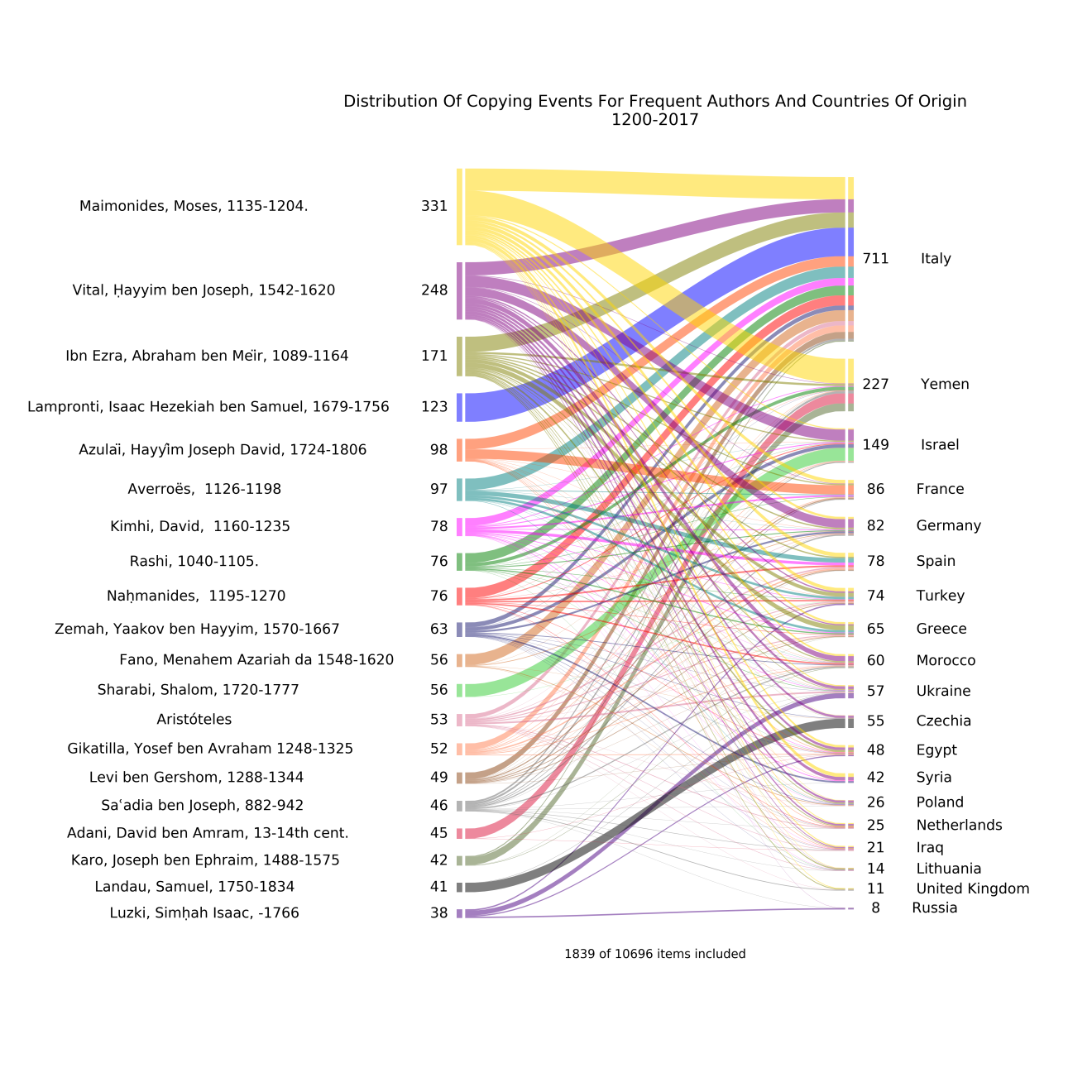


Figure 5:A bi-partite graph of the geographical distribution of copying events for the twenty most popular authors and countries.

**Composition event vs. copying event: The most copied subjects in different time periods and countries**

Figures 6 and 7 display the diachronic and geographical distribution of copying events by subject. Subjects are directly linked only to the composition events in the ontology, so the (indirect) connection between subjects and copying locations, required for this type of analysis, was inferred from other ontological relationships. Regarding the time of copying, it is interesting to note that works dealing with the sciences and philosophy were mostly copied in the fifteenth century, while works dealing with the three most popular subjects, Kabbalah, Jewish law, and liturgy were also copied often in the later period of 1700 to 1900, centuries after the invention of printing.

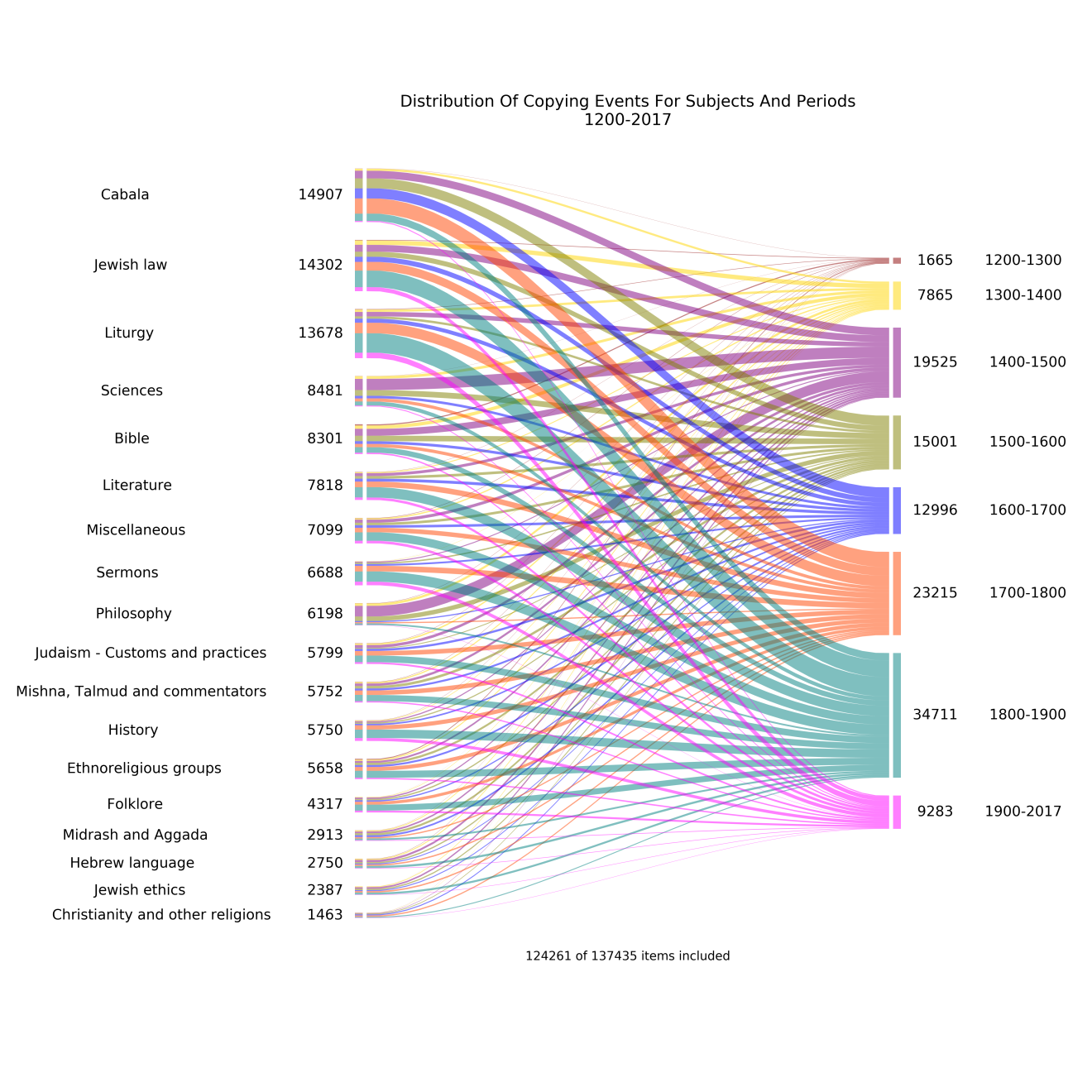


Figure 6: A diachronic distribution of subjects by copying event.

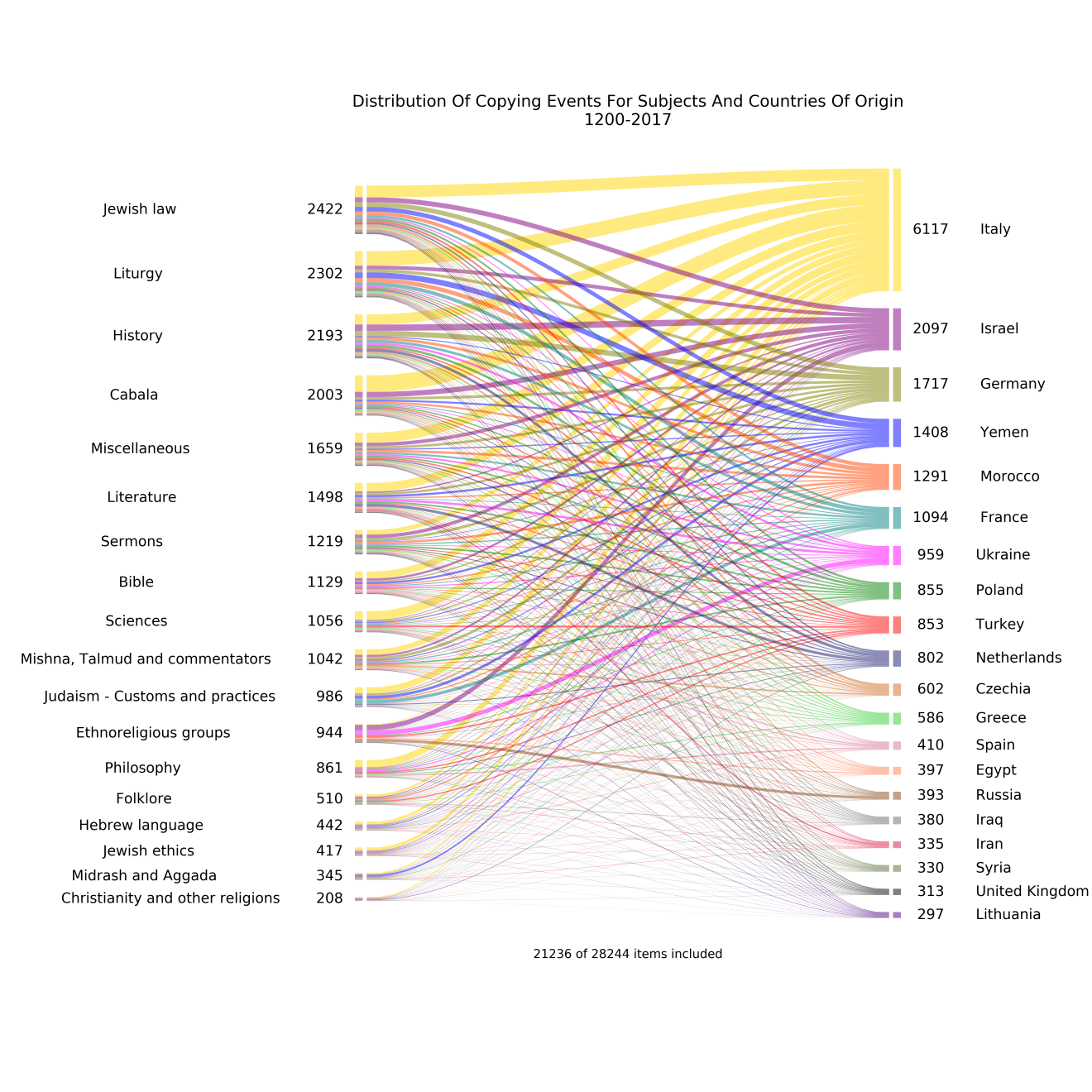


Figure 7: A geographic distribution of subjects by copying event for the twenty principal countries.

In conclusion, the above diagrams and their accompanying interpretations are only few examples of how the methods of distance reading and close reading can be combined based on the constructed ontology. In the proposed study these diagrams and others will be further analysed and interpreted in order to draw general conclusions concerning overall processes and influences. At the same time, we will use the data to pinpoint various phenomena related to individual authors and works and put them in their historical and cultural context.