Mapping German fiction in translation in the German National Library catalogue (1980-2020)

Lisa Maria Teichmann

German Studies
Department of Languages, Literatures, and Cultures
McGill University, Montreal

December 15, 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the
degree of
Doctor of Philosophy

TABLE OF CONTENTS

Abstract

This thesis project is dedicated to making translations visible and disentangling the roles of the author, place of publication, and language in carving the paths that German fiction travels in the world of literature. These channels of transfer are documented in archives of national literatures and this project presents an unprecedented study of translated German fiction by sourcing bibliographic data of translations from the German National Library. It is a first attempt at utilizing the wealth of data openly accessible through the library catalogue and at designing, exploring, and testing a quantitative model of analysis for German fiction in translation, drawing on social network analysis, descriptive statistics, and geomapping. By analyzing the role certain languages and authors play in the field of translation, this thesis investigates which works of fiction and authors are overrepresented and which ones are marginal in the translation network. This thesis builds on the notion of *translationalism*—that certain authors set themselves apart by connecting several linguistic communities in geographic space—as a driving force behind the literary transfer of translations. Translationalism, as argued here, invites the reader to re-evaluate what is categorized as world literature by drawing attention to the role of the national canon and the national collection in the global circulation of German fiction in translation.

Sommaire

Ce projet de thèse vise à rendre visibles les œuvres de traduction ainsi qu'à démêler les rôles de l'auteur·rice, du lieu de publication et de la langue dans les voies que peuvent emprunter les fictions allemandes pour intégrer le monde littéraire. De telles chaînes de transfert peuvent être trouvées dans les archives des littératures nationales et cette thèse présente une étude inédite des traductions de fictions allemandes en cherchant la source des données bibliographiques de traductions trouvées dans la Bibliothèque nationale allemande. Il s'agit d'une première tentative d'utilisation de cette richesse de données ouvertes à travers le catalogue de la bibliothèque, mais aussi de création, d'exploration et d'essai d'un modèle d'analyse quantitative usant de statistiques descriptives, de géolocalisation et d'analyse de réseaux sociaux pour étudier les traductions d'œuvres de fiction allemandes. En se concentrant sur le rôle que jouent les langues et les auteur·rice·s dans le domaine de la traduction, cette thèse cherche à comprendre quels ouvrages et artistes sont les plus représentés – et lesquels sont les plus négligés—dans les réseaux de traduction. Cette thèse s'appuie sur le concept de *translationalism*—rejeté par quelques auteur·rice·s qui décident plutôt de lier différentes communautés linguistiques entre elles dans un même espace géographique – en le considérant comme un moteur du transfert littéraire des traductions. Le *translationalism*, tel que nous l'envisageons ici, invite le lectorat à réévaluer ce que l'on perçoit comme littérature du monde en soulignant le pouvoir du canon national et de la collection nationale dans la circulation mondiale des traductions de fictions allemandes.

Acknowledgements

Introduction

"A focus on translation and contact among languages, however, implies new scales of literary history and new principles of literary belonging." (Walkowitz 2017, 44)

Why are certain authors or works of fiction representative of the national canon, yet do not cross their original language's boundaries through translation? This question was born out of an observation in my early study years when translating Turkish fiction into German. I would often encounter nationally famous authors, such as Ahmet Hamdi Tanpınar, Memduh Şevket Esendal,[1] or Ayşe Kulin,[2] whom we discussed in countless seminars at Istanbul and Boğaziçi University. However, I often discovered that few to none of their works had been translated into German or other major languages. The sharp national borders seemed to also extend to the realm of literature, and while Orhan Pamuk had reached worldwide fame with translations into various languages and claimed his Nobel Prize, those countless writers we discussed in Turkish literature seminars, lectures, conferences, and theater groups were unknown to readers outside of Turkey.

While reflecting on this observation in the German literary context, I started noticing similar cases. For instance, Ilse Aichinger has won more than 20 literary prizes; however, according to the German National Library catalogue, translations of her works are published in only 18 countries and 16 languages. Her novel *Die größere Hoffnung*, first published in 1948, was published in French translation in 1953 by Gallimard in Paris, in 1963 by Atheneum in New York in English translation, and the same year also in Italian by Garzanti in Milano. Only after 1995 were there seven more translations into Russian, Spanish, Slovenian, and finally, in 2020, Korean, which shows that only recently have her works reached more places in translation.[3] Aichinger stands out as an author whose position in the national canon is not reflected in the number of translated works, only 30 of which are listed in the catalogue of the German National library.

On the other hand, Herta Müller is translated in 42 countries and into 39 languages

---

[1] Only one novel translated into German, *Die Mieter des Herrn A.* (2007), besides a collection of short stories: Esendal, Memduh Şevket, Attila Babadostu, Dimitri Gül, Iskender Kantemur, Melanie Langsenlehner, Julia Margiol, Lisa Teichmann, and Verlag Literaturca. *Von Heuverkäufern, Glücksvögeln und Steinmörsern Kurzgeschichten in zwei Sprachen, Deutsch und Türkisch = Üfürükçü, otçuluk, taşhavan, kısmet*, 2015.

[2] Ayse Kulin has only two translations in German, *Der schmale Grad* (2010) and *Steinmauer und offene Fenster* (2007).

[3] There appears to be a big gap between the initial translations by prestigious publishing houses and further translations of almost 20 years.

with a total of 198 translations according to the German National library catalogue. The same year Müller won the Nobel Prize in literature in 2009, her novel *Atemschaukel* was first published in German and won the Deutscher Buchpreis.[4] At that point, *Atemschaukel* was only translated into Dutch and Swedish, but was quickly followed by 24 translations in 2010, including into Rumanian, Hungarian, Czech, Italian, and Norwegian. In 2011 eleven more translations of *Atemschaukel* came out in Russian, French, and Japanese among others.[5] By 2011 the novel had travelled to São Paulo, Tokyo, and Tehran with one last translation in 2014 into Thai. In comparison to Aichinger, Müller appears to have significantly more titles in translation, reaching more than twice as many countries and languages.

These two examples of translations of Müller and Aichinger illustrate an imbalance between the position of canonical authors in their national literature and the global circulation of their works in translation. Not only do they differ in the number of translated titles, but also in how widely distributed they are across languages and publishing places. Cases like these indicate that looking at translations reveals the linguistic and geographic dimensions beyond the national canon by gaining insight into the global circulation of specific authors. They also raise questions such as: what differentiates and/or connects these authors? What is the role of the languages they are translated into? Which places do their works travel to and which cultural literary communities do they reach?

A number of scholars in translation studies have made similar observations regarding the imbalance of translations according to their linguistic and geographic dimensions, their role in literary transfer (Bachleitner and Wolf 2004, Walkowitz 2017, Reynolds and Vitali 2021),[6] and networks of circulation in the formation of world literature. Johan Heilbron (1999) has

---

[4] This was 27 years after the publication of her first work *Niederungen* which took 17 years to get translated into English as *Nadir* with the University of Nebraska Press. See Flood, Alison. "Herta Müller Takes Nobel Prize for Literature," August 10, 2009. https://www.theguardian.com/books/2009/oct/08/herta-muller-nobel-prize-literature.

[5] The first translation in English only appeared in 2012. The latest translation had been into Lithuanian in 2020. While in 2010 all translations were mostly done by publishers based in Europe. *Atemschaukel* experienced a significant push with an increased number of translations happening in only two years.

[6] The term "transfer" in the context of translation is subject to an ongoing debate in translation studies. Some scholars frequently use the term "transfer", Heilbron tends more toward "flow" with a linguistic connotation. Even though Pym mentions that "transfer" was "understood as the simple moving of inscribed material from one place and time to another place and time" (Pym 2010, 13), he also refrains from using that term in the context of translation because of the numerous misunderstandings it caused in the field (see Pym 2004). Kershaw and Saldanha argue that "transfer" represents "the dominant conceptual metaphors in Western theory based on the TRANSLATION IS TRANSFER metaphor" (Kershaw and Saldanha 2013, 136). While in this thesis I only briefly address the arising challenges with this term regarding my model and analysis in the discussion chapter, I use the term "transfer" in the sense of the process of the network formation and its dynamics between authors, languages and publishing places by establishing connections with the objective to visualize these "routes of transfer."

argued that the global circulation of a literary work depends on which languages it is translated into and, more precisely, that translations in central languages (e.g., English) are also more likely to be distributed across peripheral languages. His argument also relies on the observation of an imbalance in that some languages have more translations than others, making them central in their position in the network of transfer. Both *Atemschaukel* and *Die größere Hoffnung* appear to have followed Heilbron's trajectory, being published in central languages such as English and French, as well as Dutch and Swedish, and then travelling to ever more peripheral languages and places. However, we can also see that *Atemschaukel* had reached several peripheral languages before being translated into French, which sets it apart from Heilbron's argument.

Comparing both authors reveals that while Aichinger has fewer translations into fewer languages overall than Müller, every second translated title of hers is published in a different language, making her a translationally successful author. Müller, on the other hand, has almost eight times as many titles, many of which appear to be translated into comparatively fewer languages. Each author appears to be distributed differently in translation, not only by means of reaching different languages and publishing places, but also because of how distributed their titles are overall. On the one hand, we can observe an imbalance of translated titles, but looking further at languages, we can also see how translational an author is. While scholars have used the term "translational" for literature that is written in various languages and incorporates translation as a genre-building element (Ivanovic and Seidl 2016, Hassan 2006),[7] I apply it as an analytical category to explore authors and works that follow Walkowitz's definition of "born translated," in that translational literature is widely distributed in linguistic and geographic space, connecting distant literary communities.[8] Accordingly, "translational" is a categorical term I use in this thesis following Rebecca Walkowitz who in her book *Born Translated* (2017) claims that specific authors set themselves apart by their translations in that they address various linguistic communities and geographic places simultaneously. She calls these works "born-translated" which thus form

---

[7] In their article "What is translational literature and how to classify it?: crowd-sourcing as a starting point for corpus building and type distinction in comparative literature" (2016), Ivanovic and Seidl propose a method of classification for translational literature that incorporates both, metadata and text-inherent features such as translations within the source text. Since my thesis project represents a bibliographic data and metadata analysis the proposed classification scheme cannot be applied to the catalogue data alone but would require a close reading of each work.

[8] She differentiates between the series denoting the collection of "many editions, each in a different language" and objects as "one edition, consisting of different languages" (Walkowitz 2017, 47). My definition of translational focuses on the first, series, while the object dimension is addressed in my analysis of bilingual and multilingual works and their role in the DNB catalogue.

their own category of analysis beside the national canon of representative authors. Walkowitz's concept of born-translated literature directly addresses the question of what differentiates certain authors by looking at the circulation of their translated works.

For my model, I focus my analysis on linguistic distribution as a measure for translationalism and therefore draw on Walkowitz's definition in order to understand the nature of translationalism. By using translational literature as a category of analysis to respond to Heilbron's argument on the role of languages in the imbalance of the circulation of translation, this thesis is dedicated to testing two hypotheses: first, that there is an imbalance of translated titles and their linguistic distribution which sets canonical and translational authors apart. And secondly, that specific authors bridge linguistic communities and publishing places despite this imbalance and therefore shape the network of literary transfer.

In order to test these hypotheses, this thesis addresses the following central research questions: whose, where, and in which languages are translations of German fiction published? Can we observe an imbalance in canonical versus translational authors by applying quantitative methods to a large number of translated works published originally in German? What is the role of canonical authors in the global circulation of German fiction in translation, and what is the role of non-canonical (translational) authors in connecting linguistic and literary communities in geographic space? Tied together, these questions converge in the role translations play in the national and global canonization of an author or work of fiction and how translations challenge our understandings of the national and the world literary canons of German fiction.[9]

They also open up methodological questions on how we can model translations for authors of German fiction by using bibliographic data from a linguistic and geographic viewpoint. How can we develop a model tailored to bibliographic translation data and the above questions by building on existing models of networks and geographic mapping? How can we visualize the networks of translational transfer? What is an accessible, available, and comprehensive resource for extracting and compiling a dataset of German fiction in translation?

The methodological and theoretical implications of these questions bring us to the central research problems this thesis is addressing, which can be summarized under the notion of the invisibility of translations. First, to date there are only a few quantitative studies

---

[9] I refer to German fiction as fiction published originally in German.

of German fiction in translation that are not restricted to a case study of a set of prestigious authors and/or languages, even though according to the translation database Index Translationum, German is the third most translated source language[10] and therefore occupies a key role in the network of translations.

Secondly, even though library catalogues have been identified as an important resource for translations data (Donahaye 2012), no quantitative study of translations of German fiction has been conducted on any of the national library catalogues of German speaking countries that is not limited to a genre, author, or period. Similarly, no dataset specializing in outgoing translations of German fiction is freely accessible to date. Consequently, until now quantitative models for bibliographic translation data have not yet been developed or adapted to these resources. Quantitative translation research is often market-driven, and studies mostly limited to descriptive statistics (percentages of translations in a given country, etc.). Further, despite the fact that, in translation studies and literary studies, models for translations have been formalized, translation studies, digital humanities (DH), and German studies still lack models for quantitatively analyzing translations and their transfer between languages and publishing places. There is a need for more models that allow for a comparative analysis of translation in library catalogues and national collections. Both of these challenges—the lack in a dataset as well as the lack of a quantitative model of analysis—call for developing a model that is tailored to translation data specifically, as well as the extraction of a dataset of German fiction in translation that includes the necessary variables.

Even though library catalogues are freely accessible and often follow an open access license for data use, there are a several challenges in locating and extracting bibliographic translation data. Due to its collection policies and data accessibility, the German National Library catalogue is an immensely rich open resource for bibliographic data, containing information on publication places, authors, and languages. However, due to its cataloguing practices, translations are not categorized as such in a standardized way. The information on whether the edition is a translation is only visible as part of the title or in the notes (e.g., "translated by", "trans. by", etc.). This makes translations hardly visible in that they cannot be easily differentiated from original editions and, hence, extracted from the catalogue. From

---

[10] Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Original Language." Accessed January 28, 2021.
https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=3&nTyp=min&topN=50&lg=0.
Additionally, Germany not only is home to one of the largest book markets in Europe, but also has a comparatively large share of translations globally.

their status in the German National Library catalogue, it is apparent that translations are generally not understood as part of a national or world literary canon. Instead, they appear to represent their own set of authors, following specific rules of cataloguing, with multilingual data depending on the target language. Additionally, translations appear as separate editions, not linked to the original publication in German. The marginal role of translations is therefore also visible in the lack of quantitative translation research on catalogue data, which requires a model that takes into account these specifications of the bibliographic data structure.

This thesis project is dedicated to making translations visible and disentangling the roles of authors, publishing places, and languages in carving the paths that German fiction travels on in the world of literature. These traces can be found documented in archives of national literatures, and this project represents an unprecedented study of translated German fiction by sourcing the bibliographic data from the German National Library. It is a first attempt at using the wealth of data freely accessible through the library catalogue by creating, exploring, and testing different models of analysis to investigate the observed imbalance—which works of fiction and authors are favoured and which ones are marginal in translation networks and global circulation—and to map the relationships in the global circulation and canonization of German fiction; that is, what is categorized as world literature. Following Mani's claim that "translations form the very foundations of world literature and global literary comparison and so do not always impede the idea of world literary access but do allow us to track the question of power and access" (Mani 2017, 41-42), this thesis also questions the classification frameworks for the canon of world literature within the German National Library catalogue and the role translationalism plays in transgressing its boundaries by applying theories and methods of sociological translation studies, social network analysis, and literary geography.

For this thesis project, a specific dataset of bibliographic translation data was extracted and curated from the German National Library catalogue, which will be published and therefore made accessible to future studies. The resulting dataset consists of over 35.000 translated titles by more than 6000 authors extracted from the German National Library, representing the first dataset compiled from the catalogue of German fiction of outgoing translations. By presenting the first quantitative study of German fiction in translation that visualizes translations in the German National Library catalogue, this thesis contributes toward making this rich resource of bibliographic data accessible to the field of digital humanities, German and translation studies, and comparative literature in the hopes of

opening the doors to quantitative translation studies within and beyond German literature.

In addition to these scholarly fields, the contribution of this thesis project further extends to authors, publishers, as well as educational institutions and lastly readers. For authors, my findings may raise the awareness of translation as an extended mode of production and of writing itself, a practice that as Walkowitz argues has been popular in recent years as more and more born-translated novels are being published. Additionally, publishers can draw on my analysis for their market analysis and therefore make use of my dataset and model to understand large translational trends across languages and publishing centres. In the educational context, this thesis can provide content as well as the digital tools to teach translation quantitatively and invite students to reflect on the translational dimension in literature, especially within world literature courses or courses where translations are on reading lists. Lastly, for readers, this thesis may inspire reflection on the genesis of the books we read in terms of their translations, and to ask ourselves which languages came first and how that connects or disconnects us to other reader communities. It also may change the way we view which authors we deem representative of a certain literature and why and which authors we have not yet discovered due to their translational trajectories. In terms of accessibility to those authors, this thesis may prompt us to have a closer look at our national libraries and collections as one of the gates through which a translation passes to arrive on our bookshelves.

Designing a model of analysis tailored to the specifications of the bibliographic data structure and the cataloguing practices of the German National Library for translations, which allows for testing the central hypotheses, is the second major objective of this thesis. Methodologically, this thesis project is an attempt to develop a model to analyze the dynamics in the translation network by mapping how literary fields are connected by people, publishing places, and languages. These three main variables of analysis correspond to the first central question of whose, where and into which languages translations are published. The latter is the main unit of analysis of the model, representing the overarching system of transactions between authors forming the network of circulation between linguistic communities and publishing centres. I argue that each language has a unique function in the network of translations and directs the movement of each work. Therefore, the network of translation is not restricted to national boundaries but is inscribed unto a map of linguistic communities.

Each chapter represents a building block of the model drawing on descriptive

statistics, social network analysis, and geomapping[11] to provide an explorative and multilayered picture of the literary history of translations. I follow three central notions throughout the analysis: concentration, connectivity, and communities. From measuring concentration, we can get a sense of the overall imbalance that the examples at the beginning of this introduction point toward. This allows me to test the hypothesis of an uneven distribution and Eurocentric focus of translations in specific languages and authors. From there, as I argue throughout this thesis, the space of the canon in the field of translation[12] and the position of specific authors that stand out in reaching various linguistic communities can be mapped. Following Heilbron's hypothesis of a centralized language network, I expect to find few languages and authors that have the most translations, which would support the argument that the library catalogue predominantly represents a core group of (canonical) authors and target languages.

Secondly, I measure an author's translationalism by applying social network analysis and geographic mapping to analyze their roles in connecting languages and publishing places. In line with Heilbron, my network model is constructed on the argument that the translational field's structure is a network in which authors constitute bridges (connections) between linguistic communities and geographic places while languages, which represent points of connection (nodes), occupy different positions (peripheral, semi-peripheral, central). I expect to find a distinct group of authors that are distributed across languages and publishing places: that is, authors who stand out in their translationalism. In looking further into authors that are located on the fringes of the network and who are only present in specific languages, I also expect to find a set of different authors and investigate their relationship to languages and publishing places.

Measuring the linguistic and geographic distributions of an author by their translations is especially crucial to see how factoring in translationalism as an analytical category challenges our understanding of the function of the library catalogue as both an archive and an agent in shaping German fiction in translation beyond its national boundaries.

---

[11] Unless stated otherwise, statistical testing, visualization, and mapping was done in R and RStudio using various libraries which are referenced throughout the chapters. The resulting R scripts will be published alongside the dataset.

[12] Drawing on Bourdieu's field theory that defines the artistic field as a "The space of literary or artistic position-takings, i.e., the structured set of the manifestations of the social agents involved in the field" (Bourdieu and Johnson 1993, 312), I use the term field of translation to describe the space in which authors, languages, and publishing places are positioned in and whose positioning condition the circulation of translations. In addition, I use the term translational field to describe the space within the field of translations, which is composed of authors that are linguistically widely distributed as a result of their translationalism.

In response to Venkat Mani's question of "how local and translocal, national and transnational, provincial and cosmopolitan actors and institutions work toward the creation of a world literary space" (Mani 2017, 41),[13] I argue that national library collections are representative of vernacular paradigms of a literary history and that translational literature, by placing itself in between applied categories, linguistic and geographical, challenges the underlying premises of the catalogue. Hence, library collections are representative archives of canonical and translational tendencies, and by making translations visible, tracing their network of authors, languages, and publishing places, we can see how they transgress the boundaries of the "national" beyond the national collection that the German National Library catalogue is representing.

For German Studies, this thesis aims at an extension of the German Studies paradigm beyond canonical, "nationally" identified and highly prestigious authors by introducing translations into the field as an invaluable resource to explore German literature at large. The proposed model may not only partially fill the gap of quantitative methods for bibliographic data in translation and literary studies, especially in German Studies, but also expand our understanding of representative authors and works of fiction in the field of translation.

Accordingly, I see this thesis situated within the translational turn in the humanities, expanding the concept of literary transfer by applying a language network model and visualizing translated authors as connections between linguistic communities. By curating and making available the dataset, as well as the code scripts for analysis, this thesis project further adds to the quantitative turn in translation studies, as well as multilingual digital humanities, by presenting a multilingual dataset and offering solutions in working with bibliographic translation data across different languages. As we will see, some of these solutions address multilingual place names, whose mapping represents a contribution to the geographic shift in translation studies[14] and the concept of publishing centres and their role in the circulation of translations. Within translation sociology, this thesis draws heavily on field theory by conceptualizing the translational field as a space where translationalism takes shape. Therefore, it not only addresses the roles of specific languages and authors, but also uses their connections and routes of transfer to visualize their positions in the field of

---

[13] Mani himself argues that "an entire network of social structures and institutions facilitates, and at times jeopardizes, the conditions under which a literary work can be identified as 'born translated'" (Mani 2017, 41). My thesis hence also aims at making visible these conditions that make it possible to identify works of translational authors within the bibliographic data of the German National Library catalogue.

[14] For a precursory project for mapping place names in Turkish which partially inspired this thesis project see: "A Literary Map of Turkey | Franz-Benjamin Mocnik, Lisa Maria Teichmann," 2015. http://projects.mocnik-science.net/literary-map-of-turkey/.

translation. In addition, this thesis also contributes to critical archival studies by making translations visible within national collections and catalogues, which hence opens the discussion on the role of these archives in the making of "world" literature. Also, in the interdisciplinary nature of this thesis project is another objective: developing a comparative model that can be adapted to other resources of bibliographic data for analyzing the literary transfer between literary cultures. Accordingly, the proposed model can be applied and hence may contribute to any sister fields of German literary studies that have been founded on the concept of a national representative canon of a specific literary culture and its language, such as Russian, Hispanic, Italian, or Turkish studies, and for which translation has previously been a marginal topic of analysis.

In accordance with the research aims, problems, and central hypotheses, as well as the nature of the catalogue data and the applied definition of translations to facilitate its extraction, the following guiding principles have set the scope of the study:

First, it is important to note that this study is about outgoing translations—works originally published in German which then have additional editions in other languages. There is a significant amount of existing market-related studies that address the translation publishing landscape in Germany and incoming[15] translations (see Wischenbart et al. 2010, 2018, and 2020) but only few specific case studies for outgoing translations. In focusing my project on translations of fiction originally published in German, I hope to shift the attention to outgoing translations, to add a counterweight to the market-driven models, and to provide a base of comparison for outgoing versus incoming translations on the global literary market.

Secondly, as mentioned above, this study focuses on four main variables: title, author, language, and publishing place. Again, since factoring in publishers as an additional variable moves the analysis into the sector of market-related research, and therefore goes into a different direction related to agency, principal component analysis (see Bourdieu and Johnson 1993), and revenue-based inquiries, my model is restricted to the network of languages, places, and authors.

Similarly, my two examples at the beginning of this introduction hint at the influence literary prizes can have on the likelihood of an author to get translated. Prestigious prizes such as the Nobel Prize, as well as the Deutscher Buchpreis, play a major role in attracting publishers to invest in a translation of an author and have a yet unmeasured influence on the

---

[15] By incoming, I mean translations of works originally published not in German but published in translation in Germany. E.g., a work by Jane Austen published in German translation by a publisher situated in Germany.

likelihood of a work to get translated. While this thesis project is limited to the library catalogue data, which does not give indication of prestige for an author such as prizes, revenues, or professional relationships, comparing my results to the lists of prize winners could yield interesting results. My dataset and results hence can provide a starting point for measuring the influence literary prizes have on translations in future studies.

Third, as a precursor to the data extraction I ask whether the German National Library is a comprehensive catalogue for translations from German compared to other catalogues, which is why this thesis also includes a comparison of the overall completeness of the DNB[16] (Deutsche Nationalbibliothek) and a data quality analysis. The data quality analysis represents novel information about the data completeness as well as missing and ambiguous bibliographic translation data and can also be applied to other catalogues to assess data quality. This means that, even though a manual verification of each of the 35000 entries has not been conducted, it can be evaluated whether the German National Library catalogue is representative of the translated titles by comparing the catalogued fields and titles of selected authors manually with other catalogues and databases[17] to get a general sense how comprehensive the catalogue is compared to other resources. Adapting the model to analyze the data quality of other catalogues, such as the Swiss and Austrian National libraries, is the main aim of making the dataset and developed code scripts available to the research community.

Lastly, my dataset is comprised of 40 years of published translations that have been catalogued by the DNB, and hence it represents a study going beyond most case studies that look at single authors, periods, or works. 1980 marks the year when translations increasingly appear in the catalogue and 2020 represents the final year of this thesis project. As we will see, cataloguing practices have an immense influence of the data that can be extracted and the results that can be obtained. Accordingly, my analysis is a "Momentaufnahme" (snapshot) of the ever-changing catalogue, to which editions are continuously added and data fields updated. Thus, this thesis is the starting point of taking snapshots of the DNB catalogue and its translations, which at a later point in time can allow scholars to trace the development of this national archive and the role of translations in it. This can serve as the basis for a longitudinal analysis to document cataloguing practices and the chronological analysis of German fiction in translation.

---

[16] I use the abbreviated form of Deutsche Nationalbibliothek (DNB) throughout this thesis.

[17] The main reason behind this was that the extraction of translation data is not the same for each catalogue.

My chapters are organized as follows: Chapter 1.1 and 1.2 are dedicated to reviewing the existing literature on the quantitative analysis of bibliographic translation data, outlining applied methods and models, as well as central arguments, to address the research problems and situate this thesis project in the scholarly field. In chapter 1.1, I first contextualize the exceptional role of German in the translational landscape and to what extent the field of literary studies lacks comprehensive, data-driven studies of German literature in translation. This chapter also discusses how studies have addressed the role of translation in the formation of the world literary canon with a focus on how translation studies have challenged this paradigm by drawing attention to its Eurocentric, Germanocentric biases and introducing translations as an analytic category which initiated the translational turn in literary studies. Lastly, I also discuss how these arguments have laid the foundation for the emerging methodologies for the quantitative analysis of world literature, canonization, and translationalism.

In chapter 1.2, I outline the methodological frameworks and related datasets within the field of the quantitative study of translation. I discuss the advantages and limitations of different translation datasets for market-based as opposed to literary analysis in order to highlight the methodological challenges and lack of models in quantitative translation research, focusing on the German context. In response to the lack of models for analyzing translation data of German fiction, I argue that the field needs to develop models for bibliographic data and datasets that correspond with the theoretical implications mentioned in chapter 1.1.

Chapter 2.0 is a response to the question of visibility of translation data in different library catalogues and proposes a definition of translation for extracting bibliographic translation data of German fiction from the German National Library catalogue. This chapter introduces the dataset and its properties, variables, and data quality, as well as the challenges of working with this data. It also compares the German National library catalogue with other catalogues and datasets, such as the Austrian National Library and the Index Translationum, to list the inconsistencies across existing resources and establish which resource is the most comprehensive in its cataloguing of translation data.

Chapter 2 is dedicated to the question of who and what is translated by analyzing titles and authors across languages in the German National library catalogue. I apply descriptive statistics like frequency distributions and correlation coefficients to measure the extent of the concentration of titles per language and language family to address the

arguments scholars of translation studies have raised regarding the Eurocentric tendencies and uneven distribution in translations. I expect to find a small number of central, concentrated languages that have most titles, as well as peripheral languages with few titles.

Next, I use title and language counts to situate authors in the canonical and translational fields to see if there is a distinguished group of titles or authors that stand out in their translationalism by their linguistic distribution in the German National Library catalogue. In identifying which authors are situated in and beyond the overlap of the canonical and translational fields, I address the question of the role of canonical authors in the circulation of the national canon through the library catalogue and test my hypothesis of an imbalance in translated titles and languages that sets specific authors apart. Based on my observation of Aichinger and Müller, I expect to see an imbalance of the number of titles and the number of languages authors are translated into, highlighting authors like Müller as compared to other canonical authors.

Chapter 3 aims at testing my second central hypothesis, that authors connect languages. Based on this hypothesis, my network consists of languages as nodes connected by authors as edges, meaning that languages are connected if they have translations by the same authors. By applying social network analysis to Heilbron's centre-periphery model and centrality measures, I test how connected the language network is and which languages have influential positions in the translation network by sharing translated authors. I also apply a community detection algorithm to discern sub-groups of languages based on the authors they share in order to classify peripheral and central languages and investigate which authors connect them. I identify which authors connect central to peripheral languages and whether those authors are also concentrated in their editions or linguistically distributed. My analysis explores the function of authors in the author-language network as bridges between languages and the role of canonical versus translational authors.

The last chapter maps the relationships between publishing places, languages, and authors in the translational network to track the channels of translation that connect different literary cultures in geographic space. As for language, I measure concentration of titles per publishing place and visualize the publishing centres and languages by geocoding the locations of the publishers extracted from the dataset. By also drawing connecting lines between places by shared authors, I investigate whether we can observe a similar centre-periphery structure as in the network graph from the previous chapter. I also look at sub-communities of authors and discuss how they challenge our understanding of a global distribution of German fiction in translation.

Keeping in mind the invisibility of translations, the imbalance of translations that varies across authors, and our underlying assumptions about world literature and the national canon, let us now delve into the DNB catalogue and explore the roles authors such as Aichinger and Müller, and languages such as English, French or Hungarian play in the global circulation and literary transfer of German fiction. Let us look at the map of translations to gain a deeper understanding of translations and trace the threads that connect one literary tradition or community to another.

1.1 Toward a quantitative framework for translational literature

The German publication market is one of the most active in Europe when it comes to publishing translations. With 208,240 translated titles with German as the original language, according to the Index Translationum, German is the third most translated language after English and French.[18] Nevertheless, even though databases like the Index Translationum and the German National Library provide access to translation data, a quantitative study of German fiction in translation has not yet been done. One reason for this lack in studies is that applying quantitative methods to German literature in translation is a young research subject from a computational standpoint. Even though access to bibliographic data increases and the applied methodologies diversify,[19] there is a persistent focus on smaller case-study specific research projects in this field. Hence, the field of German literature in translation is mostly limited to selected topics, languages, authors, or periods, for example "Georgische Literatur in deutschen Übersetzungen : Bibliographie – 1887 bis 2017" (Chotiwari-Jünger 2017), "Literarische Übersetzung als Universum der Differenz : mit einer analytischen Studie zu deutschen Übersetzungen des Romans Oblomov von I. A. Gončarov" (Tashinskiy 2018), "Der Verleger Johann Friedrich Cotta (1764-1832) als Kulturvermittler zwischen Deutschland und Frankreich : Frankreichbezüge, Koeditionen und Übersetzungen" (Hass 2015), or "Annäherung – Anverwandlung – Aneignung : Goethes Übersetzungen in poetologischer und interkultureller Perspektive" (May und Zemanek 2013).

Notwithstanding the case-study specific focus in translation studies and German studies, contextualizing the broader role of national literatures with their specific canonical authors in a global literary field and the movement of works, themes and topics has a long-lasting tradition in literary studies as the study of world literature. This concept—which in the majority of scholarly works is tied to Johann Wolfgang von Goethe's discussion of the term in 1827 with Eckermann (see Eckermann 1837)—has persistently been applied to look at why some authors are part of a global canon rather than remaining in their national literatures. Of course, the implication of a universal literature comprised of a selected number of authors or works has been critically challenged in its Eurocentrism and Germanocentrism

---

[18] Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Original Language." Accessed January 28, 2021.
https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=3&nTyp=min&topN=50&lg=0.
[19] Topic Modeling, cluster analysis, sentiment analysis, machine learning models, to name a few.

as well as its gender imbalance—most "world literature" is male and originally published in Europe. This prompted some literary and cultural scholars, such as Susan Bassnett, Doris Bachmann-Medick, Emily Apter and Rebecca Walkowitz, to widen the theoretical implications to focus on transfer, pluralism, and the distribution of works of literature by converging upon translation as an analytical category. This so-called "translational turn" opens the field to include post-colonial and gender theories and translation as an analytical category in the study of literature of the world. At the same time, a model for the analysis of translational literature has been taking shape through a sociological framework of translation (Bachleitner and Wolf 2004; Heilbron 1999; Sapiro 2016 and 2020). These recent studies have shed new light on how bibliographic translation data can be quantitatively modelled, which previous studies have not addressed. However, they have not developed the methodological framework to a point where it provides empirical evidence for their theoretical implications about the role of German in the translation market. Hence, I argue that making use of the wealth of available bibliographic data and modeling German fiction in translation spatially, as well as socially, is long overdue, especially since the possible findings have important implications for the broader domain of literary studies: canonization, world literature versus national literatures, and literary communities.

Consequently, this thesis is an attempt to fill the gap of modeling the social and spatial relations of authors, works, and languages in the field of German translated fiction at large and to further develop a methodological framework that takes translation as its main analytical category. Theoretically, it challenges the conceptual division of world literature versus national literature that scholars of German studies have been relying on in the past and builds on recent studies of translationalism, network theory, and literary geography. Methodologically, it is grounded in the sociological approach to translation proposed by scholars such as Heilbron, Bachleitner and Wolf, and Sapiro, while further building on their proposed models and applying them to a dataset of German fiction in translation (1990-2020) extracted from the German National Library, which is discussed in the following chapter (1.2.). To that effect, this chapter is dedicated to discussing why and how mapping German fiction in translation addresses the theoretical implications and concepts of world literature, canonization, and translationalism, as well as the role of German on the translational landscape, to illustrate what is at stake.

German literature in translation

According to the Index Translationum—one of the most extensive databases of translations[20]—Germany holds the top position as the country that published the most translations (269,724 in total), followed by Spain (232,853), France (198,574) and Japan (130,496).[21] According to the German National Library in 2019, out of 28733 published works of fiction, 995 were translations.

Moreover, the report "Publishing Translations in Europe Trends 1990-2005" (Büchler 2011), which is based on an analysis of the Index Translationum database and was published as part of the *Making Literature Travel* series of reports on literary exchange, translation and publishing, outlines the exceptional position Germany takes as a producer of translations in the European landscape of translations, labelling it as "one of the most productive countries as regards to translated books" (Büchler 2011, 67) according to the annual output in comparison to the European average.[22]

Even though this is common knowledge, studies on German translation mainly focus on case studies for a specific target language (Norrick-Rühl and Bold 2016; Gerber 2008, Tatlock 2012; Raven 2002), and a comprehensive study on the global spread of German language fiction in translation is still missing in the literary field. There are several reasons behind this. First, finding data on translations is a challenging task. Bestseller lists, annual reports of publishers, and platforms such as goodreads.com may give an overview on "the best," "the most popular," or the "most read" authors and works for a given year. These resources are directed at readers rather than at scholars who wish to gain insight into the overall distribution or numbers of translated works of fiction for each target language.

Secondly, access to statistics and data on translations is highly limited. Certain translation agencies and marketing companies collect the data but make it accessible only to paying members or publishing houses.[23] The Federation of European Publishers also collects data and publishes annual reports[24] on the translation industry in European countries,

---

[20] For the period between 1979 and 2009 with more than 2,000,000 entries in all disciplines: literature, social and human sciences, natural and exact sciences, art, history.

[21] Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Original Country." Accessed January 28, 2021.
https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=1&nTyp=min&topN=50&lg=0.

[22] Germany is one of the most productive countries as regards to translated books, the average annual output in the 1990-2005 period reaching nearly 10,000 titles, 52.5% of which is literature, somewhat higher than the European average of 50.5% (See Büchler 2011).

[23] An example for German literature is Rüdiger Wischenbart Conent & Consulting, which also publishes reports including publishing statistics regularly.

[24] "FEP Publications." Accessed January 28, 2021. https://fep-fee.eu/-Publications-.

including sale numbers. However, there is no distinction made between fiction and non-fiction and the raw data remains inaccessible to the public.

Nevertheless, in recent years valuable resources have been published in the form of datasets that provide an entry point to the study of German translated fiction at large. The Index Translationum[25] is one of those projects funded by UNESCO, which unfortunately did not continue to be extended past books published after 2009, but nevertheless provides open access to scholars of all fields to quantitatively study translation for roughly 400 languages. The Index Translationum is the only openly available and accessible resource to globally compare and analyze translation data. However, it is not documented how frequently and comprehensively the German translation data is updated. In 2019 a query on the online resource revealed that for the years after 2010, no entries with German as the original language are included. Additional queries and sources suggest that 2005 marks the last most comprehensive year of collected data in the Index Translationum (Büchler 2011, 4).

Besides the Index Translationum, the German National Library is one of the most comprehensive resources for bibliographic data of German fiction in translation. Looking at the annual report gives an idea of what position translated works hold in the inventory. The German National Library has a collection policy, according to which any printed or digital work originally published in Germany (for example translations) as well as any works on Germany (Germanica) are obliged to be submitted to the library.[26] According to the National Library's report from 2019, the number of translations (not specified if incoming or outgoing[27] and Germanica was 9,514 for 2019, 9,764 for 2018, and 7,773 for 2017 (Sälzer and Schmitz-Kuhl 2020, 46). Of 39,016,827 works, 15,962,386 are monographs in the current inventory of the German National Library as of 2019 (Sälzer and Schmitz-Kuhl 2020, 44). Translated works and Germanica therefore make up 24% of the complete catalogue as compiled until 2019 according to Sälzer and Schmitz-Kuhl.[28] It is not specified what the percentage of incoming or outgoing translations is, but we do see a slight decrease in the sum

---

[25] See "Index Translationum." Accessed March 20, 2019. http://www.unesco.org/xtrans/bsstatexp.aspx.

[26] See the section "Sammelpflichtige Veröffentlichungen aus dem Ausland" Deutsche Nationalbibliothek. "Unser Sammelauftrag." Accessed January 28, 2021.
https://www.dnb.de/DE/Professionell/Sammeln/sammeln_node.html.

[27] The authors of the 2019 report, Sälzer and Schmitz-Kuhl, combine translation and "foreign-language media works published abroad that relate to Germany and persons in the German-speaking countries" (Germanica) since it falls under the same collection mandate as translations and is therefore included in this statistic. See: Deutsche Nationalbibliothek. "Sammlung Körperlicher Medienwerke." Accessed January 28, 2021.
https://www.dnb.de/DE/Professionell/Sammeln/Koerperliche_Medienwerke/koerperliche_medienwerke_node.html.

[28] See chapter 2.0 for the percentages of translations in the DNB catalogue at the time of data extraction.

of titles from 2017 onwards.[29] Bibliographic data is freely accessible for 22,232,147 items (Sälzer and Schmitz-Kuhl 2020, 50), which makes the German National Library the most valuable resource for studying the corpus of translations produced in Germany. Both the Index Translationum and the dataset of the German National Library provide the resources for a comprehensive analysis of translated German fiction in a given period or language. Despite the role of German literature on the global market as well as the accessibility of data, it is even more surprising that even though Germany holds an extraordinary position both in importing and exporting translation and has a long-standing theoretical focus on world literature and German, to date, scholarship on German publishing lacks comprehensive, data-driven studies about literature and its role in the global translation market.

World literature and translation as an analytical category

Another point which further draws attention to this gap in scholarly literature is the fact that translation has been a pre-condition for the study of "world" literature. In literary studies, the question of who and what is translated is inseparably intertwined with the concept of world literature.[30] How have world literary studies addressed this relationship to translation? And how has this concept been challenged in the context of recent studies that also address colonialism, feminist theory, and Germanocentrism? In recent years, collections of essays have revisited the concept of world literature, critiquing its claims of universality, cultural appropriation, and gender. Two major publications with collected essays—*World Literature in Theory* (2014) and *The Routledge Companion to World Literature* (2012) both edited by David Damrosch—include earlier extensive critiques outlining the problematics of the world literature paradigm regarding these claims.

We can identify two major standpoints from which the world literature paradigm has been either endorsed or challenged: the generalization based on the universal application of

---

[29] See the dataset description for percentages per year.

[30] In 1827, Goethe famously noted in a letter to Eckermann that one must turn from "National Literature" to "world literature" as a new model of the coming epoch (Eckermann 1837, 325). According to the ideals of the Weimarer Klassik, Goethe refers to Greek literature as the archetype, while all other literary epochs should be looked at historically "appropriating to ourselves what is good" (Eckermann 1837, 325), while not focusing on the particular, but on world literatures as a model. He further draws attention to reading world literature as a "pattern". Interestingly, we can see the quantitative and structural nature of Goethe's words here, which push the reader to look beyond the confinements of national literatures, or even German literature. This very central statement that would turn out to be one of the most cited references to the concept of world literature to date, is embedded in references to numerous translations from Greek, Chinese, and Arabic from which Goethe's practice and method—if you will call it such—becomes apparent. Just looking at the West-östlicher Diwan exemplifies perfectly how far Goethe has been involved in applying translation to genre and metaphors to widen his literary scope.

world literature through Eurocentrism/Germanocentrism in literary canons and the persistent national and gender-related biases in the studies of world literature.

In his article "A View on the Unification of Literature" (originally published in 1922),[31] Zhen Zhenduo endorses the universal application of world literature as a theoretical framework, arguing that the study of literature must be universalized. For Zhenduo, world literature as such is a unity of all literature, "a unified study synthesizing the entire world's literatures and making literariness its main focus" (Zhenduo 2014, 67).[32] The alignment of Goethe's and Zhenduo's definitions of world literature is undeniable and deeply related to translation. Zhenduo sees the challenges of a universalized study of literature in the lack of a universal language, and he argues that inspiration and sentiment can be transferred if the texts are translated truthfully to the original.[33] For a scholar looking to synthesize "the entire world's literature," Goethe's and Zhenduo's discussion leaves much to be desired and poses many questions as to the universal applicability of this theory to all literatures. It is not surprising, therefore, that world literature continues to instigate scholarly discourse on the problematics of this concept.

Opposed to Zhenduo, David Damrosch, in his extended work on world literature, stresses that studies on literary history tend to claim a global universality while only analyzing one language, and therefore one national literature (mostly English and American).

> World literature courses, and the anthologies that served them, saw no incongruity in defining "the world" purely in terms of Western Europe and its classical and biblical antecedents, sometimes with a few Russian and American writers thrown in for good measure. (Damrosch 2008, 482)

While here Damrosch's critique on the Eurocentrism in courses and canons of world literature is clear, the author proposes to view world literature as linked to the movement of a work from a "national context to the spheres of world literature," pointing out that in recent years a literary globalization has been taking place. For him, a literary work is not born as world literature but becomes world literature, as in the case of Voltaire's *Candide* that "entered world literature when it crossed the English Channel to become *Candid* in English translation" (Damrosch 2008, 484).

---

[31] Reprinted in *World Literature in Theory* (Damrosch 2014).

[32] He is therefore in line with Edward Said's views—Said himself was a translator of Auerbach's essay "Philologie und Weltliteratur"—that world literature does not disempower national literatures but outlines their individual specifications (See Arac, 2011).

[33] The author makes an additional point here and also draws attention to the amount of translation in terms of market saturation, asking "Even if literature can be translated, can the amount of translations satisfy the need of scholars or even readers with a global interest?" (Zhenduo 2014, 63).

Damrosch further asks how one could write a history of world literature, looking at the processes and strategies a writer employs to move between regional, national, and global contexts, or what he calls "literary systems." He argues that literature has always been local and translocal, international and national in a global world,[34] ending his discussion on the purpose of a world literature by saying that it "will provide an invaluable map to locate our work in the wider world" (Damrosch 2008, 494). The author here draws a direct link between what constitutes the system of world literature: language and geography. In his critique of the claimed universal applicability of world literature as a theoretical framework but also as a category of literature itself, it is striking how Damrosch argues in terms of a geographic focus and exemplifies perfectly how unequally distributed literatures are in the field of literary history. According to his definition, world literature is both a system and a map through which processes of global distribution are made visible.

In a critique of world literature with the title "Should We Rethink the Notion of World Literature?" (originally published in 1972), the sinologist René Etiemble argues in line with Damrosch putting forward the inherent bias of the concept of world literature:

> Because the concept of Weltliteratur was coined in German (and by what a German!) it has always retained, at least for certain people, the taint of a germanocentrism. (Etiemble 2012, 87)

In other words, Etiemble criticizes the term world literature based on its Germanocentrism, arguing that all national literatures simply make literature without the adjective "world." Interestingly, in the beginning of the twentieth century, we can clearly observe how the concept of world literature has produced the first quantifiable bibliographies as a first step of canonization from within German literature.

As an example of this bias toward German, and more generally Western European, literature categorized as world literature, Etiemble mentions one of the first attempts to put together a master list of what should be translated and recognized as world literature: Hermann Hesse's essay "Eine Bibliothek der Weltliteratur" (originally published in 1929).[35] For this essay, Hesse asked writers to select their ideal library of world literature, arguing that world literature should be accessible to anyone. He also argues that some poems are untranslatable, which is why access to world literature is restricted by language. In an earlier

---

[34] As an example of this "simultaneous localization of the global and globalization of the local" (Damrosch 2008, 492), Damrosch mentions the development of the international script from the language of the Sumerians for a multi-lingual region.
[35] In Hesse, Hermann. *Schriften Zur Literatur.* Frankfurt am Main: Suhrkamp, 1972.

text titled "Übersetzungen," he also stresses the impossibility to translate poetry due to its complexity. For him, world literature is related to the idea of a masterpiece ("Meisterwerk"), and he criticizes all the less valuable books that flood the market.

First, according to Hesse, in order to build a library of world literature, one must look at the books that have not lost popularity throughout centuries, "the holy classics" of religious and mythical texts. Hence, he created a list of his "ideal" world literature library (that, according to Hesse, is only problematic because it appears to him as too perfect).[36] At the end of his list, Hesse asks how he could justify including mostly European literature and not Indian and Chinese classics such as the Bhagavad-Gita, or the "I Ging." He answers by saying that his collection is subjective and moody ("launenhaft") and that it would be hardly possible or even impossible to replace the world literature with an objective, fair list simply because it would be too long and only decades later would it become clear which works attain the status of a classic.[37]

In addition to Hesse's list, as an example of the canonization efforts of world literature works, Etiemble also mentions two volumes published in Germany and Austria after 1939 of *Die Weltliteratur and Vergleichende Zeittafel der Weltliteratur* by Adolf Spemann (1951), of which three-quarters of the texts are by French or German authors, with no mention of any of the Chinese poets and writers of the time. Just like Hesse's list, the selection of authors is, besides two to three exemptions of Japanese and Indian titles, mostly composed of German—Goethe appears with three titles—French, and English titles from which the Eurocentric focus of these lists is apparent.

These lists are not unique to the German context. Similar to Hesse, the French writer and poet Raymond Queneau also compiled a list based on books selected by writers from a list of 3,500 titles (see Queneau 1956).[38] All these efforts can be read as part of the canonization process of what we now call "classics" and are tied into the colonization dynamics of the time. Translation and the paradigm of world literature can also be read as a history of the colonization of language and the establishment of the lingua francas, as well as the "classics" canon and the educational canon which is widely read in schools today.

---

[36] See Hesse 1972.

[37] "Was in den Bestand gehört, der Generationen überdauert, darüber hat die eigene Zeit nicht zu urteilen" (Hesse 1972, 359).

[38] It is not surprising that the first ten titles are almost identical with the most translated authors of the time according to the Index Translationum. See: Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Author." Accessed January 28, 2021. https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=5&nTyp=min&topN=50.

Etiemble mentions this study to exemplify how homogeneously French the titles on the list are (considering that most writers were of French descent) and critiques the absence of major Chinese, Indian, and Arabic works. He points out the Eurocentrism of what was constructed as the first canons of world literature and draws attention to what is missing by comparing them to several compilations of world literature titles for educational purposes published in Japan and Egypt. The author argues that these lists portray a linguistically more diverse picture of world literature including a variety of Arabic, Persian, and Indian titles. Etiemble's point here is clearly to draw attention to the inherent bias within the world literature paradigm toward a set of specific languages and their representative authors.

From a quantitative standpoint, a look at Hesse's list as well as other canonized lists of world literature[39] will quickly reveal not only the enormous imbalance or bias of languages, but also of gender. Under these premises, what are we studying when looking at world literature in translation quantitatively? As an indicator that Hesse sees the heritage of world literature as entirely in the hands of men, he uses the exact phrase "our grandfathers and fathers" in the context of what has been forgotten from medieval times (Hesse 1972, 347) and about the books his "sons and grandsons" will read (Hesse 1972, 360). Even though some might argue that it is obvious that a text from 1929 will mainly contain male author names (it is still not obvious to me why this is or should be the case), I cannot stress it often enough how skewed the common idea of a world literature is in terms of (especially!) gender. In fact, Hesse mentions only two female authors: Bettina von Arnim and Droste Hülshoff.

However, these biases are not only present in lists by prominent writers and scholars from the former century, but they also almost identically match up with the content presented in one of the most accessed current resources to learn and read about literature, Wikipedia. The German Wikipedia only counts only twenty "classics" by female (international) writers,[40] more than half of which are authors mainly writing in English, followed by a majority of French and none in German or non-European languages.

One recent study on Wikipedia and world literature (Hube et al. 2017) ranks authors for their prominence on Wikipedia according to page length, page rank, in-and outgoing links, and number of page views. Across all variables, besides Shakespeare, German writers such as Goethe and Schiller as well as Heinrich Heine show up as the most highly ranked.

---

[39] *Bibliothek der Weltliteratur (BDW) (1962 - 1991)* See: Kritikatur - Das Kulturprotal. "Bibliothek Der Weltliteratur (BDW)." Accessed July 3, 2022. https://www.kritikatur.de/Bibliothek_der_Weltliteratur_(BDW), *Cotta'sche Bibliothek der Weltliteratur* (beginning 1882), and the *Manesse Bibliothek der Weltliteratur* series (beginning 1944).
[40] Wikipedia. "Frauenliteratur." Accessed February 26, 2020. https://de.wikipedia.org/wiki/Frauenliteratur.

Not a single female German author is present among the most highly ranked authors, even among different language versions, and the vast majority of writers are male. Except for non-European language editions such as Chinese, Arabic and Turkish, mostly English, French, or Russian writers dominate the chart (Hube et al. 2017, 20). Not only do we seem to anthologize, canonize, and study mainly European writers, meaning that the term world literature mostly encompasses works by German and French authors, but the majority of classics under the category world literature are also written by male authors. Huge et al.'s analysis of bias in Wikipedia is one of many studies that clearly critique the inherent bias in canonical lists of authors and works.

Interestingly, computational studies of world literature aim to raise awareness around the inherent bias of world literature and its canons. In their article titled "Gender and Vernaculars in Digital Humanities and World Literature," Bergenmar and Leppänen criticize that most quantitative studies of literature still lack a comprehensive discussion on gender and what they call "smaller languages in highly globalized scholarly practices" (Bergenmar and Leppänen 2017). They point out how the digital humanities studies of world literature conflict the aim to rewrite literary history in terms of gender and language:

> The large-scale methods of digital humanities and world literature, on the one hand, and the study of literary exchange, on the other, are to some extent conflicting: the first two tending towards general models and the second concerned with cultural specificity and (un)translatability. (Bergenmar and Leppänen 2017, 233)

The authors argue that in world literature, there is an absence of gender both theoretically and numerically (Bergenmar and Leppänen 2017, 237) and that gender hardly figures at all as an analytical category. Additionally, they criticize how the digital humanities' studies of world literature (e.g., Franco Moretti) focus on the novel and therefore are mainly Eurocentric due to the geographic origin of the novel as a genre.

> The comparisons of the rise of the novel in different parts of the world and in other languages rest on publication data made available by other scholars. In order to be able to study literature on a global scale, without the specialized knowledge of all the national literary traditions involved, data collected from previous research are used in the same manner—not read closely, but at a distance. (Bergenmar and Leppänen 2017, 236)

Here, they establish two important focal points: the place from which world literature is being studied and the risks of overgeneralization without context on the basis of one point of departure. This poses that limits to the study of world literature are still defined by a national context.

> Still, translation is always an adaptation to culture and language. If, for example, one wants to analyse the novel genre globally by distant reading a large corpus of texts translated into English, it is still in some sense the English novel that is being investigated. (Bergenmar and Leppänen 2017, 239)

They suggest that to make results of large-scale literary analysis meaningful, "the empirical data need to be contextualized and theoretically framed" (Bergenmar and Leppänen 2017, 235).

Spivak and Damrosch make a similar point, saying that "When it's done badly, the teaching of world literature is, in fact, methodologically naïve, culturally deracinated, philologically comprised, and ideologically suspect?" (Damrosch and Spivak 2011, 368). In line with Bergenmar and Lepännen, they also opt for more in-depth knowledge of a variety of languages, more collaborative scholarship, and pluralism instead of the Eurocentric study of world literature, both in terms of language and institutions doing research on this subject (Damrosch and Spivak 2011, 367).

Emily Apter makes a similar point of critique, arguing that "many recent efforts to revive World Literature rely on a translatability assumption" (Apter 2013, 3) and juxtaposing the universalized theories and disciplines that are based on the trope of world literature to untranslatables. The "translational humanities," she writes, as a paradigm is more and more recognized. Within this paradigm, both translation studies and the study of world literature align in their predicament of an underlying universality of language, literature and writing through "worldly criticism, politicized cosmopolitanism, comparability aesthetic, and a redrawn map of language geopolitics" (Apter 2013, 7). She openly criticizes the overtly pluralistic, ecumenical application of these practices in university programs, claiming that both, translation studies and world literature studies simply ignored their theoretical and mostly Eurocentric premises through the repeated study of "the classic," "the postmodern," etc.—which are all based on examples from the Western literary canon.[41]

To summarize all these different points of critique, we can draw the following conclusions. First, there is a persistent bias in the category of world literature, which is biased geographically and linguistically, as well as according to gender. Secondly, the way world literature has been studied up to now has been just as biased in similar ways. Third, in all these studies which voice a critique on the concept and category of world literature, one

---

[41] "Literary communities are gated: according to Western law and international statue, authors have texts, publishers have a universal right to translate (as long as they pay), and nations own literary patrimony, as cultural inheritance" (Apter 2013, 15).

connecting point of focus becomes apparent: methodology! None of the aforementioned scholars appear to completely discard world literature as an object of analysis, but instead seem to critique the way it has been studied and, more specifically, its limitations due to the bias in the selection of texts.

But what do they suggest as a counter methodology to answer questions such as "can world literature then be seen as a synonym for canonization of a 'world literature' in national terms?" and "where does translation stand in all of this?" I agree with Bergenmar and Leppänen, Damrosch, Etiemble, and Apter that a comprehensive, de-centralized study of world literature in terms of the flow of literary works needs to follow a pluralistic, diverse, and large-scale methodology;[42] however, I argue that first, such a framework needs to be centred around translation as the main object of analysis[43] and secondly, that it is necessary to develop a methodological model that can produce findings related to the questions above. In recent years, a redefining of world literature in terms of translation and translational has been taking shape that does not further perpetuate the canonicity and Eurocentrism of male classics.

The "translational turn"

Emerging methodologies in translation studies and the sociological study of translation have been contributing toward a bottom-up framework for the study of the flow and movements of literary works at large, beyond the categories of national and canonical works. Theoretically, this thesis draws heavily on the concept of "translational" literature as well as on sociological approaches to German fiction in translation for its methodology.

In her book "Born Translated" Rebecca Walkowitz claims that "the novel is the most international genre" and defines born-translated novels as novels that "have been written for translation from the start" (Walkowitz 2017, 2):

> Like born-digital literature, which is made on or for the computer, born-translated literature approaches translation as medium and origin rather than as afterthought. Translation is not secondary or incidental to these works. It is a condition of their production. (Walkowitz 2017, 3-4)

She approaches "world literature from the perspective of translation" (Walkowitz 2017, 44),

---

[42] René Etiemble, for instance, who wrote his reflection on world literature long before quantitative methods found their applications in the literary field, argued that world literature needs to be analyzed comparatively and at a large scale for unbiased results.

[43] Or as Apter puts it, "testing the hypothesis that translation and untranslatability are constitutive of world forms of literature" (Apter 2013, 16).

with each chapter analyzing a selected number of works to outline the different qualities of translational literature. She links born-translated literature to the development of a globalized market and the author's positioning within. Walkowitz employs a method she calls "close reading at a distance," her selection covers a wide range of authors (from Ishiguro's novels to Young-Hae Chang's born-digital multimedia work). Even though she uses a language of computational categories of analysis such as in her chapter "Sampling, Collating, and Counting," Walkowitz does not use any quantitative methods. In Walkowitz's discussion of "how contemporary novels have incorporated translation into production, and what this development does—what it needs to do—to our analytic categories and procedures" (Walkowitz 2017, 45), she outlines a number of factors that characterize born-translated novels: market-specific factors,[44] translation and geography[45], style, genre and topoi of translational literature, and the question of how "translation shapes the narrative structure of the contemporary novel" (Walkowitz 2017, 4).[46]

Walkowitz is not the first scholar to base her analysis on translation and its linguistic and geographic circulation. The focus on these categories also resonates with Johan Heilbron's sociological approach to the "translational field," which I discuss in detail in chapter 1.2. Additionally, several scholars of translation have referred to the study of translation across disciplines, languages, and theoretical literary traditions as the "translational turn" or the "cultural turn in translation" (Bachmann-Medick 2009; Bassnett 2014).

Susan Bassnett, an author Walkowitz surprisingly does not mention in her book, is a pioneering scholar in translation studies who—alongside André Lefevere—has propagated the study of the processes of translation since the mid-seventies. Coming from the manipulation school, founded in the 1970s, she has been interested in the "manipulatory

---

[44] She especially draws attention to the case of English being mostly a receiving language with an estimate of only 3% of translations in the book market of the U.S., meaning that books published in English do not need translation as a means of distribution on a global market.

[45] As mentioned above, Bassnett has repeatedly stressed the relationship of geography and translation. Walkowitz also draws this connection stating that born-translated works are "refusing to match language and geography" (Walkowitz 2017, 6). If, for example, the novel pretends to be written or take place in another language it is written in (as in the case of *Don Quixote*) (Walkowitz 2017, 4).

[46] "Walkowitz repeatedly argues that translational writers such as Orhan Pamuk display overgeneralization, which is reader-directed and fulfills the premise of world literature, the "Tendency to make guidebooks or cultural primers out of literary works from representative spaces" (Walkowitz 2017, 16-17). Born-translated poetry, according to Walkowitz, is "not so much a work for all readers as a work for other readers" (Walkowitz 2017, 17).

processes involved in the transfer of texts" (Bassnett 2011, 123).[47] Describing translation theory as a general theory of transaction, she propagates a turn to investigate factors beyond the aesthetical categorization of translation which was used to describe why some texts have entered one canon in translation while others remain unstranslatables. She sees this notion "as a crucial instrument of literary renewal" to "rethink of how we draw up literary histories, how we map out the shaping forces of past and present" (Bassnett 2011, 127-128).

Bassnett argues that, due to the move away from the analysis of certain authors and works claiming universality,[48] "the translational turn in cultural studies is now well underway" (Bassnett 2011, 136). Bassnett sees parallels in the developments of translational studies and cultural studies "towards a greater awareness of the international context and the need to balance local with global discourses" (Bassnett 2011, 133) through a methodology of coding and decoding and by "understanding the manipulatory processes that are involved in textual production" (Bassnett 2011, 136). For cultural studies, this entails looking at the set of power relations between source and target contexts, toward the analysis of text production across linguistic boundaries. In both cultural studies and translational studies she sees the need for a deeper exploration of how cultures construct their image of writers and texts, how texts become cultural capital across cultural boundaries, and the dynamics of translation politics.

The emphasis on geography and language Bassnett is voicing here also echoes Damrosch's proposed approach to translation. In her earlier text "From Cultural Turn to Translational Turn. A Transnational Journey" (2014), Bassnett suggests an investigation of the questions above through the combined study of literature, geography and translation.

> Any study of translation necessarily involves a geographical dimension, and the movement of literatures through translation requires an awareness of changing context of textual production. (Bassnett 2014, 235)

She here again stresses the importance of studying translations as text products and as text producers—the integration of theory and practice (Bassnett 2014, 236). Theory and practice to study translation in our age of "accelerating electronic communication" for Bassnett could emerge from the combined methodologies of translation, literature, and geography.

---

[47] "These constraints, or manipulatory processes involved in the transfer of texts have become the primary focus of work in translation studies, and in order to study those processes, translation studies have changed its course and has become both broader and deeper" (Bassnett 2011, 124).

[48] "As any translation studies scholar knows, a comparison of translations of the same text, particularly of a text that has been translated frequently, exposes the fallacy of universal greatness" (Bassnett 2011, 135).

Even though Bassnett sees a fluid evolution from the study of world literature to the study of translations, arguing that when we study translation we also study world literature,[49] Bassnett does not propose a clear methodology to model translation. As I illustrate in this section, this again emphasizes the reasons of a lacking comprehensive, quantitative study of translated fiction.

Doris Bachmann-Medick, whom Bassnett also refers to, sees the translational turn interlinked with the development of a methodological framework across disciplines. In the introduction to an edition of *Translation Studies* (2009), Bachmann-Medick points out that several studies from the early 2000s, including, among others (Papastergiadis 2000; Baker 2006; Renn 2006; Snell-Hornby 2009), Bassnett's study on translating terror, contribute to the development of the translation category within and beyond translation studies and therefore the birth of a "translational turn":

> But a 'translational turn' goes further because it is born specifically out of the translation category's migration from the translation studies into other disciplinary discursive fields in the humanities. (Bachmann-Medick 2009, 3)

She argues that only when the study of translation has been established as a methodologically grounded analytical category across disciplines is a translational turn put in motion. This entails the transcendence of the translational category from translational studies and from the analytical unit of the text to cultural translation and systems of transactions (see also Bassnett 2014). A methodology of translation is grounded in a close and distant analytical perspective, to study both: the translation as a text and the system of texts as a global network of transactions.

> The translation scholar always sets the micro and macro levels in a necessary interrelation: the smaller formats, textual and interactional analysis, are related to wider translational frameworks and vice versa. (Bachmann-Medick 2009, 3)

For Bachmann-Medick, translational frameworks also require that "translation as a differential concept must be thoroughly historicized" (Bachmann-Medick 2009, 10), and a translational turn also includes the theoretical exchange and translation of analytical categories and theories as well as their cross-cultural application. Bachmann-Medick claims that for a cross-cultural analysis, we must first problematize "cross-categorical translation" in terms of research categories that claim to be universally and cross-culturally applicable. The

---

[49] "What is significant today about the growth of world literature is that it offers a reappraisal of the significance of translation and proposes a shift of focus onto interconnectedness, on global literary and cultural flows on the one hand, and on question of agency on the other" (Bassnett 2011, 239).

study of migration and translation therefore plays a crucial role, since, in addition to works of fiction, their tools and units of analysis travel across linguistic and cultural borders. Bachmann-Medick further argues that for the humanities and Kulturwissenschaften or cultural studies, a translational turn would mean opening up to translation: to make the "contact zones" between disciplines bigger, to draw attention to the decentralized flux of theoretical schools of thought, transgressing boundaries of national literatures.[50]

These three different viewpoints on translation are all propagating an emerging practice in studying (world) literature at the crossroad of literary studies, comparative literature, and the comparatively more recent field of translation studies, which in turn includes historicization, the study of untranslatables, and the study of translation in terms of geography, language, and translation practice.

Emerging methodologies

Even though these recent notions of a translational turn suggest a methodological shift in the study of world literature, especially through mapping translations' movement and distribution, a quantitative study of translationalism for German fiction in translation that goes beyond specific case studies of single authors or genres and uses open access bibliographic data from library catalogues has not yet been conducted. Likewise, previous reports outline some of the factors, but do not analyze them in relationship with each other. Thus, I see this thesis project as a continuation of this conceptual shift, but with a methodological framework for the quantitative analysis of translated fiction, arguing that translational literature—as supported by previous studies on the field of translation—is the result of a broad spatial and social network composed of authors, languages, and publishing places.

In recent years, several empirical studies have employed the emerging methodological and theoretical framework of translational literature. Christine Ivanovic, who extensively studies translationalism in relation to the work of Yoko Tawada, proposes a classification model for translational literature for corpus analysis (Ivanovic and Seidl 2016), which defines translational as both the text itself and meta features. Similarly, Andrew Piper in his discussion on paraphrasis in Goethe's work (Piper 2010) takes the term translational literally, meaning translated by the author in the original text (as in Goethe's *West-Östlicher Divan*).

---

[50] "The 'translational' turn in the humanities finds its greatest scope at the points where disciplines make themselves pluralized and translatable within an emerging global knowledge society" (Bachmann-Medick 2009, 13).

Additionally, some studies map translations of one language comparatively, as in the case of Vaicenoniene's article on Anglo-American versus Lithuanian translation systems (2010) and Zhou and Sun in their article "Bibliography-based quantitative translation history" (2017), in which the authors outline current research models and challenges of a bibliographic data approach. In line with these initial studies that apply quantitative methods, this thesis takes translation as a categorical unit of analysis through the interplay of the global movement of German fiction in translation, spatial and geographic mapping, and social networks between author and target languages. Building on Bassnett's definition of translation as a combined practice of geography and language, Heilbron's outlined framework, and sociological studies of translation, this thesis builds on prior work in translation studies and bibliographic data analysis, hopefully providing a fruitful ground for emerging quantitative studies of translational literature.

1.2 A model for the quantitative study of translations

One would assume that for a language with one of the largest shares of translations on the market,[51] there would be a significant amount of empirical evidence, as well as numerous studies with detailed, accessible datasets of bibliographic data and statistical models, but as I will show in this chapter, this is not the case for German literature in translation. There are two main reasons for that. First, the methodological and theoretical frameworks to quantitatively model translation data are still developing, as is data availability and accessibility. Secondly, studies that have applied a quantitative framework to bibliographic data of translations limit their findings to market analyses and specific case studies and do not go further than suggesting to use descriptive statistics as the main methodology.

Nevertheless, in the field of translation sociology, a detailed model has been proposed by several scholars (Heilbron 1999, Sapiro 2016 and 2020, Bachleitner and Wolf 2004) which combines the analysis of social networks and spatial relations of translators, authors, and publishers in the field of translation. This framework has found partial application in studies of Norrick-Rühl and Bold (2016), Zhou and Sun (2017), and Susan Pickford (2016) who have all described the challenges and limitations of working with bibliographic data.[52]

In this chapter, I review how studies of literature have developed this framework to model translations in the field of sociological analysis, which is based on the presumption that translation is an artistic field and network for which language is the main unit of analysis. I also discuss studies that employ a quantitative method to translation and major reports on publishing statistics,[53] focusing on how they address data availability and inconsistency as primary challenges in developing a model for translations that transgresses national and lingual boundaries. Even though they employ different methods on different data—studies on translation use bibliographic data and reports consisting of mostly revenue-focussed survey

---

[51] See Chapter 1.1. and official numbers published by the Index Translationum. Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Original Language." Accessed January 28, 2021. https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=3&nTyp=min&topN=50&lg=0.

[52] A sociology of translation could be summarized under two major currents or sections of research: Research that looks at the factors driving translation (economical, cultural, social, etc.) on the one hand—Gisèle Sapiro applied a sociological method to the French and U.S. American context and Norbert Bachleitner and Michaela Wolf to the German context, focussing on specific factors that condition the positioning of literary works within the translational field—and scholars modeling translations as an artistic field with language as the central unit of analysis on the other (Heilbron). Scholars from both currents have pointed toward the usage of publishing statistics to describe and analysis the field of translation, even though their analysis is often merely providing an overview without a comprehensive data-driven method.

[53] Reports present a methodology to measure the output of translations across languages and have pointed out the central role of German on the translation market.

data from publishers—both the respective quantitative translation studies and the publishing reports have observed translation biases when it comes to period, author gender, publishers, and languages and have pointed out the central role of German in the global translation market.

I argue that even though frameworks for a quantitative, sociological analysis of translation have been proposed by numerous scholars (Heilbron 1999, Sapiro 2020, Bachleitner and Wolf 2004), the scholarly field is lacking application of these models and has not produced a significant amount of evidence. There are just a few published studies and reports that apply these frameworks (Norrick-Rühl and Bold 2016, Zhou and Sun 2017) to specific case studies, while large-scale projects focus on the economic factors (book market reports). In order to map and analyze all translations of German fiction in the German National Library catalogue, this thesis mainly draws on the field of literary sociology or sociology of translation (Heilbron 1999) as well as the above-mentioned precursory studies that apply a quantitative approach to modeling bibliographic data of translations. Additionally, I explore the concept of translationalism, building on the analysis of Bassnett, Walkowitz, and Reynolds and Vitali, by visualizing the canonical and translational networks within the national collection of the DNB. The methodological aim of this thesis is to propose a model for the analysis of bibliographic translation data extracted from the DNB and fill the gap in scholarly literature. To that effect, in the following section I give an overview of how quantitative studies of translation and literary sociology have been developing a methodology to analyze translations, with their advantages and caveats.

Translation sociology as a methodological framework: translation as a field and language as a unit of analysis

From the sociological sub-field of literary translation studies, some scholars have voiced the need to study the production and circulation of translated works of literature. Influenced by Pierre Bourdieu's theories regarding the field of cultural production, which look at the positions and power relations of agents within an artistic field (Bourdieu and Johnson 1993), Johan Heilbron has laid the theoretical groundwork for a sociological study of translation in 1999. In "A Sociology of Translation" (1999), Heilbron points out the shift from the linguistic and practice-oriented study of translation to analyzing translation as a "transnational cultural field" (Heilbron 1999, 432). He describes this as "the conceptual shift from source-text to target-context" (Heilbron 1999, 430). He defines translation as follows:

> From a sociological perspective, translations are a function of the social relations between language groups and their transformations over time. (Heilbron 1999, 430)

We can see a clear focus on language systems and their transfer through translation in his definition, as well as a focus on language groups and not nation states, a limitation which is so often the case in reports about the translation market, which I discuss in detail in the following section.

Heilbron also defines the sociology of translation as a "structural analysis of the international flow of translated books" (Heilbron 1999, 431). To explore this "international system, "he introduces a model for conducting a sociological analysis of literature with statistics, asking "how can one account for the uneven flow of book translations between various language groups? And how can one explain the varying role of translations within different language groups?" (Heilbron 1999, 431). According to his approach, to understand the world-system of translation, one must look at the position of a language group in the international map of translations. His proposition for a sociological study of this world-system of translations is a model that uses national and international data and statistics about language groups within the world-system of book transfer by applying the sociological methodology to translation, with language as the main unit of analysis. He also proposes to use book statistics as data, referencing the Index Translationum and pointing out that the available datasets lack coherence, making it impossible to conduct a comparative analysis between translation ratios. However, he only presents limited statistical evidence or findings that apply his proposed framework to the dataset he proposes. He instead describes the positions languages take in the translation field without providing statistics other than percentages of market shares per language. For example, in Heilbron's world-system, German has a central role (with English, French and Russian), while Spanish, Italian, Polish and Czech are languages with one to three percent of the world market in translations and hence have a semi-peripheral role. Languages with less than one percent of translations, such as Chinese, Japanese, and Arabic are peripheral, which leads Heilbron to the conclusion that "the size of language groups is clearly not decisive for their degree of centrality in the translation system" (Heilbron 1999, 434).

As we can see in the example, he employs a terminology similar to social network theory—such as centrality—for the study of hierarchical structures of language groups. He argues that "the dynamics of the international translation system is based on a core-periphery structure" or a hierarchical structure of central, semi-peripheral, and peripheral languages

(Heilbron 1999, 432).[54] The translational system of transfer is shaped by the language's position in the field as well as their relationships, which allows for the analysis of translation as a network. In regards to our example above, this means that works of fiction pass from German through a number of connections, which in return shifts the positioning of languages within the translation network. As I show in chapter 3, Heilbron's framework of language positions and networks lends itself to a social network analysis model of translations, and building on his central hypothesis regarding the language network contributes toward developing quantitative models for translation studies.

As I already mentioned, an application of his framework is still not fully developed in scholarly literature on translations. To further illustrate my argument regarding this lack in scholarly literature, I now turn to several other scholars that have subsequently built on a sociological framework for translations emphasizing their presented model and findings.

Giséle Sapiro, one of the prominent scholars in sociological translation studies, has repeatedly called for making use of available databases to map out the field of literary production. Her methodology, however, relies on qualitative data about the "social condition of circulation of literary works" (Sapiro 2016, 82), implying that an analysis of movements and positions in the field can be done by mapping out the factors that contribute to the movement of a literary work beyond its geographic borders. She proposes to study the relations between specific factors (social, economic, cultural) and their participation in the production of world literature[55] Sapiro further endorses this approach in a more recent article with the title "The Transnational Literary Field between (Inter)-nationalism and Cosmopolitanism" (2020). She categorizes the factors that "shape and condition the transnational literary field and that favour or hinder its unification" (Sapiro 2020, 482) and classifies them into: political (or more broadly ideological),[56] economic,[57] cultural, and

---

[54] For Heilbron, a methodological model of translation flow is hierarchical and longitudinal, including the following measures: centrality, monopolization, and levels of cultural importation. See chapters 2 and 3 for a detailed discussion.

[55] Referring to Bourdieu, Sapiro states: "To locate translations within the field of publishing, we need to understand its structure, which is polarized between large-scale and small-scale production . . . the pole of large-scale production is ruled by the law of profitability in the short term (embodied by best-sellers), whereas at the pole of small-scale production, the specific logics and values of fields of cultural production, in our case the literary field, prevail over commercial considerations" (Sapiro 2016, 87).

[56] For political factors, the author lists copyright, censorship, and other forms of control over production and circulation. As an example, she mentions that during the German occupation translations from French were suppressed, while French publishers such as Gallimard were obliged to publish as many German translations as possible.

[57] As for economic factors, Sapiro mentions that "the circulation of works in print depends on the book and press industries and on distribution networks" and that the publishing market has both cultural (linguistic) and political frontiers (Sapiro 2016, 85).

social,[58] with a particular focus on agents (authors, publishers, state representatives, literary agents, translators, and critics). In other words, by contextualizing each factor and mapping the dynamics spatially, we gain insight into why literary works move the way they do.

She argues that the position of a language, publisher, or author depends on the relationship and influence of several factors. For Sapiro, political and cultural factors are connected, and cultural diversity is often dependant on the relationship between the two. As an example, she points out that if an author gets published from a central position in the market (by a publisher such as Gallimard, for example) and if the source language is also occupying a central position, they will have an advantage over authors publishing in a peripheral language and position in the market. We can see that Sapiro applies Heilbron's social and spatial terminology here to describe the dynamics within the translation field. They both employ a network terminology of centre versus periphery and argue that the field is defined by the position of authors, publishers, and translators in the network of language groups. While Sapiro contextualizes the factors categorically and historically, Heilbron focuses on language as a unit of analysis. Even though, like Heilbron, she does not measure the extent of centrality and periphery of the listed authors and publishers, she further contextualizes how this terminology can describe the translational field as a network of social and spatial relations.[59]

While I agree with Sapiro's approach to closely look at the relation between agents and factors, and thereby contextualize the dynamics within the literary field by applying a social and spatial terminology, for me this only represents the groundwork for a quantitative study of translation and does not provide evidence that goes beyond very specific case studies. As I argued above, even though Sapiro endorses using quantitative methods and databases and supports that "quantitative data displays patterns and regularities that help us understand the social, political, and cultural factors that may hinder or trigger the circulation of symbolic goods" (Sapiro 2016, 82), her model only remains descriptive of the factors,

---

[58] According to Sapiro, cultural factors "play in the circulation of literary works and are embedded in power relations between social groups" (Sapiro 2020, 92), suggesting that "These factors are education, the book market, the nation state, political movements, international organizations (UNESCO), and specific authorities of the transnational field (such as the Nobel Prize)" (Sapiro 2020, 482). Here she draws attention to the underrepresentation of women on the market of translation in relation to their position in the field. She also points to the cultural factors that shape the development of the canon of world literature—which she notes has been limited to European writers—and the role of translation in the formation of literary and publishing fields: "For a long time, the canon of world literature was mainly composed of white, male, Western authors. As we have seen, it started to extend to non-Western cultures in the 1950s, thanks to a voluntary policy set up by UNESCO" (Sapiro 2016, 90).

[59] Additionally, Sapiro describes the role of prizes and programs funding translations in diversifying social factors, whose "centrality and peripherality are mediated by the means of production" (Sapiro 2016, 91).

providing examples and context for each of them and is not applied to quantitative data even though she argues in favour of a data-driven analysis of translations. Therefore, her articles are valuable as a reference for defining and classifying factors for specific case studies, but do not provide a clear methodology for quantitatively measuring the positions and relations within the translational field.

This has been the case for other scholars that have presented Heilbron's socio-spatial framework to analyze translation, such as a study by Norbert Bachleitner and Michaela Wolf (2004). In line with Sapiro and Heilbron, Bachleitner and Wolf discuss the possibilities of a sociological approach to the study of translation in the German context. In their essay, they define the macrostructure of the field of world literature as a framework to analyze the national translation markets (Bachleitner and Wolf 2004, 5). Just like Heilbron, the units of analysis for Bachleitner and Wolf are languages, organized in central versus peripheral, which need to be studied according to their position in the global field of production. They clearly point out that the translational field is not defined by states, but by language groups, meaning that German may also include any publications from Switzerland and Austria.

To describe the German field of translations, Bachleitner and Wolf apply Heilbron's categories of central, peripheral, and semi-peripheral languages to the percentages of translations per year and categorize German as a semi-peripheral language with a rich symbolic capital due to the translation tradition, market size, and prestigious book fairs. Just like English, German may therefore act as a gateway language to the global translation market. Hence, the position of the target language is decisive in the circulation of a given work to the global market and the canon of world literature. Here again, we can clearly see the application of a network terminology—languages as network points (or nodes) through which works of translation are passing—describing the relationship of one language to another. The focus on a language's position and its influence in the translation network (which Bachleitner and Wolf have used to describe the translational field), as well as the dynamics between centre and (semi-)periphery invoked by Heilbron, are core concepts of my model (see chapter 3) for exploring translationalism and canonicity of authors that connect language communities.

In addition to language, Bachleitner and Wolf also use a network terminology for literary centres (capitals) and argue that the global literary field is organized polycentrically (Bachleitner and Wolf 2004, 3) which, according to Bachleitner and Wolf, then visualizes the

hierarchical structures within the global market.[60] In other words, publishing centres are organized in a centre-periphery system, which contributes to the uneven distribution of translations in the linguistic space. German, as in their example mentioned above, would hence be overrepresented due to its multitude of publishing places across Germany, Switzerland, and Austria. By adding a spatial dimension to their model, they can contextualize the specific dynamics causing the uneven distribution of translations across languages as well as highlight sub-communities (as in the case of Austria), and their role in literary transfer. In addition to the language network, the uneven distribution of translations and the role of publishing centres in the DNB is a second core concept of the model proposed in this thesis.

In summary, from Heilbron, Sapiro, and Bachleitner and Wolf, we can identify the following two main pillars of a sociological model of translations: First, authors consistently use a network terminology of centre and periphery in addition to the geographic and spatial distribution of translations according to language groups. The fact that the sociological methodology presented by these authors is based on a mixed model of social network and spatial mapping of translations suggests that they are the main analytical methods for a model for the quantitative analysis of translation. Secondly, not only does the field of translation depend on the relation between social actors (translators, publishers, authors) and factors (social, cultural, political), but a weighted model, where each actor has accumulated what they call "symbolic capital" based on their connections. As I argue, this strongly corresponds to the social network theory of centrality, which not only takes into consideration the types of connections, but also the weight or number of connections. The studies discussed above, then, shed new light on a comprehensive, quantitative, sociologically grounded analysis of translation and suggest building the methodological framework in a combined model of spatial and social network theories. By building on these two dimensions of the model—the language and the spatial network—this thesis is not only an attempt to model and analyze bibliographic translation data spatially and sociologically, but also to document the arising challenges in working with this kind of data in the German literary context. Thereby, I hope

---

[60] In their analysis, they focus on literary and symbolic capital. Literary capital conglomerates in centres as well as in the form of prestige for certain authors. Financial support by associations and government institutions also play an important role for the export of literature in translation, which in the case of Germany's rather strong market is not significant. The accumulation of symbolic capital is dependent on the dynamics between the agents, which means a translator may gain symbolic capital through the prestige of an author and vice versa. However, since most translations are contract-based, publishers are also an important agent in the translation field, as well as lectors, the expectations of the readership, and censorship, which all have an influence on the decision of which texts get translated and indirectly influence the position of the work, author, and translator.

to provide both a model and best practices for scholars wishing to work with translation data.

Identifying challenges in modeling bibliographic translation data: accessibility and data quality

How has this preliminary framework been used to study translation and map the movement of literature through language groups, and what were the main challenges? Several recent studies have applied a quantitative methodology to the field of translation by exploring these categories, while also emphasizing additional categories such as literary prizes and gateway languages. Data availability and accessibility as well as consistency between datasets (especially national library bibliographies) are recurring challenges described by scholars who work on quantitative models of translation as well as reports on translation. In line with Norrick-Rühl and Bold (2016), Zhou and Sun (2017), and Pickford (2016), I argue that national library bibliographies, if carefully curated and annotated, can provide a comprehensive resource to study translation via a model based on the spatial and social network to analyze translation quantitatively.

In their article "Crossing the Channel: Publishing Translated German Fiction in the UK," Norrick-Rühl and Bold (2016) draw a parallel between the role of German and English as translation languages, focusing on the UK publishing market. Similar to Sapiro, they outline some factors that are driving publication numbers in both languages, but do not present a way to measure the extent to which each factor contributes to the translatability of a literary work. While they discuss percentages of so-called "bottleneck" languages, such as English, which enable translations to enter the global market, they do not in detail draw out the statistics on which they base their observations on and instead reference exemplary authors and titles.[61]

Based on their dataset, a "corpus of relevant titles has been identified and collected on the basis of the British Library's British National Bibliography (BNB) database," they identified four main categories of translated German fiction that make up the majority of translations on the British market: steady sellers, classics, and authors who already achieved

---

[61] "In our database, names such as Markus Heitz with his Dwarf novels (German publisher: Piper, Orbit; UK publisher: Little, Brown Book Group) or Sebastian Fitzek (German publishers: Lübbe and DroemerKnaur; UK publisher: Atlantic Books) and Frank Schätzing (German publisher: Kiepenheuer & Witsch; UK publisher: Quercus) with their highly readable but probably rather short-lived bestselling thrillers pop up regularly" (Norrick-Rühl and Bold 2016, 130). They also stress the importance of literary prizes as an influential factor for percentage of translation as "structures and mechanisms in place that support and promote translated (German) books" (Norrick-Rühl and Bold 2016, 135). Based on the numbers by the German Book Prize and the German National Library catalog and publisher's information, the authors claim that 60% of the prizewinners have been translated into English.

popularity to some extent; bestsellers, especially German fantasy and thriller authors; novels about the Nazi period or the Holocaust; and a selected number of contemporary German fiction, most of which are written by male authors. The authors fail to present how these findings are statistically distributed in their dataset and how this relates to the other factors mentioned (availability, visibility, funding, readership size, production costs, and translation rights). Again, this study is restricted by the geographic boundaries of one case study, and due to the curation of a specific dataset, the application of their model to a larger dataset such as the German National Library catalogue appears to be challenging. What then constitutes a suitable bibliographic dataset to analyze translated fiction quantitatively?

To address this question, Zhou and Sun (2017) offer a model to study the book history of translation and argue that bibliographic datasets are the ideal source for this purpose, based on their comparison between available bibliographic databases and the proposal of a statistical model to analyze translation frequencies. Compared to Norrick-Rühl and Bold, Zhou and Sun (2017) include detailed statistics and a review of available resources for a quantitative bibliography-based history of translation such as trade catalogs, national bibliographies, union catalogs, and OCLC WorldCat. The variables Zhou and Sun map out depend on the types of metadata available through the accessible bibliographic databases— titles published, publisher, date of publication, place of publication, author, number of book reviews, recommended retail price, sales figures, book format, subjects, genres, and language, suggesting these as measurable factors for analyzing the book history of translation. These metadata types are common across different library catalogues, including the German National Library, which makes it possible to partially adapt the proposed model to the context of German fiction in translation.

Zhou and Sun propose to measure the general output of translated books in percentages, the sum of all translated books for one original language annually, and the proportion of English in percentages for the sum of translated books worldwide. Additionally, they also compare different frequency measures, such as Lotka's Law and Price's Law for author frequency and the limitations of these measures. They also analyze the publisher distribution, with special attention to how selective the process is for books to be published in translation. If we compare this framework to Heilbron's proposed model, we can see that here the different measures are not put into relation with each other, disregarding the angle of language positions and how that influences the flow of translations. Thus, there are several challenges regarding the compatibility to apply one proposed model to another dataset. Zhou and Sun's proposition to also measure themes and subjects, when compared with the

available datasets for German fiction, illustrates that a model to measure translation is heavily dependent on the bibliographic data available. For instance, Zhou and Sun's dataset includes the annotated subject of each work (according to the Dublin Core Metadata Element Set or Library of Congress).[62] The German National Library catalogue, however, lacks a subject category and therefore this category cannot be included in a model.[63]

Nevertheless, Zhou and Sun present a clear framework for a model with for a quantitative translation history with bibliographic data. Their methods mainly focus on frequency counts reported in percentages and the top of the distribution, but do not establish a link between the factors or study their relation—and therefore, they omit the dynamics in the translational field. Even though this article is limited in its findings, it does draw attention to the main factors for translation when working with bibliographic data, while also showing the limitations of this approach. Zhou and Sun mention data quality and data accessibility as the central challenges and point out that most databases such as the Library of Congress do not include a translator field in their bibliographic data.[64] They make an excellent point on how findings based on a quantitative bibliography-based history of translation can be biased according to the dataset and its quality. Their findings are limited, but they raise important points about the inconsistencies between datasets. For my dataset I found similar inconsistencies, especially when comparing library catalogues.[65]

In line with Zhou and Sun, I argue that bibliographic translation data can be modelled, but must be contextualized in terms of data quality, availability, and accessibility.[66] In her study on the availability of Maghrebi Literature in English, Pickford (2016) presents a similar argument, noting that current resources like the Index Translationum are oftentimes incomplete when it comes to minor languages, while at the same time there is a bias toward translations from Maghrebi French compared to Arabic and Berber. She also stresses the necessity for the assembly of her own corpus with a variety of data sources, due to the fact that available resources such as the Index Translationum alone are incomplete.[67]

---

[62] This is not available through the German National Library catalogue.

[63] Similarly, for the geographic distribution of translated books, Zhou and Sun only list library holdings, which can be obtained from WorldCat (Zhou and Sun 2017, 215) as a resource.

[64] As an example, often data on the translator ends up in the comments section and is not uniform, which is why institutions work toward establishing metadata standards (BIBFRAME and MARC), which are still developing.

[65] Please refer to chapter 2.0. where I discuss dataset inconsistencies in detail.

[66] See chapter 2.0 for a detailed discussion of my dataset.

[67] Pickford assembled her corpus from the LIMAG database, the Three Percent series of databases (see University of Rochester. "Three Percent Translation Database." Accessed July 3, 2022. http://www.rochester.edu/College/translation/threepercent/translation-database/.), and the Index Translationum (dataset not available at the moment on their web portal). For documentation see Donahaye, Jasmine. "Three Percent? Publishing Data and Statistics on Translated Literature in the United Kingdom and Ireland." Making

Her analytical method is mostly based on frequency counts of language (Maghrebi French versus Arabic versus Berber), works translated over the period of 1955-1960, and a typology of publishers with translation counts, as well as translators by gender and language. She uses these counts to compare trends in a time series analysis for each language. Similar to Zhou and Sun, she also follows a descriptive statistical model, but from there identifies which factors are influencing the frequency distributions. For example, she correlates increased translation counts to literary prizes, particular publisher types—especially the independent press—and the gender of the translator and author.[68] In summary, Pickford's defined factors (language, translators, publishers, prizes, academic field) correspond to the factors outlined by Zhou and Sun, but Pickford directly correlates them to her statistical findings and problematizes how the dataset availability necessitates the curation of a comprehensive dataset.

In other words, a clear limitation in the study of the translational field is the dataset quality and accessibility, which can vary greatly and often require curating a specific dataset. When comparing each mentioned study's model with special attention to the statistical measures, we can conclude that the model and findings depend on the available bibliographic data. This might not come as a surprise since this is the case in most data-driven research (Mocnik et al. 2018). However, as Pickford has pointed out, translation data is often inconsistent, sparse, and the findings correspondingly limited, only including descriptive statistics. This thesis hence aims at addressing the main challenges of data accessibility and quality of bibliographic translation data identified by Pickford, as well as Zhou and Sun, in

---

Literature Travel. Mercator Institute for Media, Languages and Culture, Aberystwyth University, Wales, UK, December 2012. https://www.lit-across-frontiers.org/wp-content/uploads/2013/03/Publishing-Data-and-Statistics-on-Translated-Literature-in-the-United-Kingdom-and-Ireland-A-LAF-research-report-March-2013-final.pdf.), cross-checked with national library catalogues in France, the US, and the UK. She notes that "the method of compilation means that translations published in the author's country of origin rather than the target culture are likely to be somewhat under-represented" (Pickford 2016, 79), which raises an interesting point on the representability of a translation dataset for a certain linguistic community.

[68] For literary prizes she lists the influence of Naguib Mahfouz's Nobel Prize in 1988. She also mentions that specific themes oftentimes linked to the political context such as post-9/11 narratives and terrorism memoirs have a positive influence of translation counts. As another factor of translations Pickford also looks at the typology of publishers and stresses the influence of the independent press in the publication of translations from Maghrebi French and Arabic. She also finds a correlation between gender, publisher, and translator. Looking at gender distribution of author and translator, she finds a "pattern of distribution in cross-referencing gender of author and translator with publisher category, with women translators publishing books by women authors with university presses" (Pickford 2016, 89). Additionally, changes in the academic field are also correlated to a rise in translation—namely the rise of Francophone postcolonial studies in the 2000's, the postcolonial feminist theory's interest in female writers and the institutionalized establishment of world literature, that "as a field of specialism . . . emphasized not national traditions but processes of intercultural transfer and exchange; together with the growth in translation studies" (Pickford 2016, 88). She argues that there is a correlation between the need for world literature course material, where reading in the original language was not a condition anymore, but translation became essential.

regard to the DNB catalogue. As I show in chapter 2.0, while the catalogue does not necessitate the manual compilation of a dataset such as Pickford's since it is freely accessible, it does, however, require transparent documentation of how cataloguing practices have had an impact on the completeness, consistency, and data quality of the fields of interest, such as language, author, and publishing place. In line with Zhou and Sun, Norrick-Rühl and Bold, and Pickford, this thesis is also an attempt to make its bibliographic data accessible for future research alongside a detailed documentation of its challenges and utility for future research.

A market-based model for analysis: reports on the translation publishing market

While, as Pickford has pointed out, translation data often requires the careful curation of a case study-specific dataset, translation data is also collected and analyzed from several institutions, especially for market research, though this data is often not freely accessible and differs from the studies discussed above. It is therefore interesting to compare existing scholarly studies to publishing reports, which are based on data directly from publishers and institutions, by focusing on their statistical model. In this section, I will look at the different problematics linking publishing reports with the aforementioned studies—considering availability and accessibility, the lack of coherent methodology, the challenges of working in national versus linguistic spaces, and the observed translation bias—to establish how applicable their models are to a sociological framework, with the aim to distill the main building blocks of a methodological framework to model translation data.

Several annual reports provide an overview of the book market in Europe. Each institution has different data sources, but similar categories, such as titles, publishers, and turnover.[69] Since most reports are based on a unique, curated dataset, which is not available to the public, adapting and applying the same model to a different dataset is challenging. The institutions compiling reports on publishing data, therefore, act as gatekeepers to the available data. Hence, there appears to be a divide between institutions collecting publishing data and the scientific discourse in quantitative literary analysis, especially in literary sociology, which looks at similar analytical categories as the respective reports (publishers, titles, languages).

Methodologically, we can see coherence in the factors explored by the studies above

---

[69] *Buchhandel in Zahlen* (Börsenverein des Deutschen Buchhandels, 2020) is likely the most comprehensive and referenced report on the publishing industry in Germany. However, while it includes a section on translations into German, it does not include any numbers on translations from German—for the obvious reason that the data is limited to the market of Germany—therefore only giving insight into the incoming, not outgoing, literature. Since this thesis project focuses on outgoing literature, with German as the source language, it is not discussed in detail here.

and their statistical methods. Time series analysis and frequency distributions for a language-specific comparison is a common research practice to measure the translation flows. While most of the methods and findings fall under the category of descriptive statistics, such as reporting frequency distributions of titles, publishers, and revenues,[70] sociological approaches and data-driven methods provide a framework for quantitatively studying the translational field as a network of book transfer between languages to then gain insight into their positions and relations with each other. Publishing reports only provide insight into each category of analysis, and not on how these categories correlate.

Additionally, the focus on high-prestige authors and books poses a number of concerns for a literary analysis of the field of translation and does not address the language and author positions such as semi-periphery and periphery in the field of translation. Since they are centred around market data (revenues and turnovers as well as bestseller and prizewinner lists), they provide an overview which is mostly directed at an audience consisting of publishers.[71] Therefore, any findings are summarized according to market-specific outcomes and consequently national boundaries and only rarely linguistic communities, which poses clear limitations to the findings that can be obtained in the context of the translational field.

The report published under the title "The Book Sector in Europe, Facts and Figures" (Turrin 2017) by the Federation of European Publishers is based on collected data from their members through surveys among 28 national book publisher associations from EU member and candidate states. The dataset, survey instrument, and methodology are not freely accessible to the public, and each year the FEP releases a collection of book publishing statistics for the European market.[72]

This again supports the argument stated in the previous section, that data availability and accessibility are common challenges in this field, additional to the methodological inconsistencies between models across datasets. As long as the dataset is not accessible and only collected by a few institutions, a broad comparative quantitative analysis of fiction in translation proves to be challenging. This, as I argue in this thesis, often necessitates curating a custom dataset from one of the institutional resources, such as library catalogues, and requires an in-depth data quality assessment alongside a model tailored to the data

---

[70] For instance, the percentage of published titles per year in a given country, annual revenue per country and year, and publisher size according to annual revenue.
[71] See *Buchhandel in Zahlen* (2020).
[72] The latest extensive report by the FEP was published in 2017, and besides annual book sales, it includes frequency graphs and maps of new titles published in the EU + EEA, 2006-2015, the number of persons employed in book publishing, and number of enterprises.

consistency and specific questions, which go beyond descriptive parameters and revenue-based inquiries.

As for using the model presented in publishing reports, we can see that for a more specific inquiry about genre—especially fiction—publishing reports can only provide limited insight and applicable methods. It is apparent that the FEP clearly uses descriptive statistics focusing on the market, the turnover, the revenues, and number of publications, including e-commerce and expenditures. Fiction and other literary genres, for example, are only related to turnover and not to the artistic and literary implications of genre distributions. This makes sense considering that this report is directed at publishers, not literary scholars.

For the same reason, a market-based approach uses frequency counts related to the number of inhabitants within national borders, such as the size of the book market in each country (sale prices), without considering the linguistic diversity of the EU member states. Hence, a market-based approach also presupposes a nation state geography of the book market, disregarding linguistic communities. Austria, Germany, and Switzerland each have their own national market and book publishing industry, while they are one linguistic community united by the German language, which means their borders are not as sharply defined as for other nation states. If we take bilingual states, such as Belgium for example, summarizing the book publishing market not by language but by an enclosed geographic space is problematic for further analysis. According to a report of 2012 by the European Commission, 24% of survey respondents in member states read books, newspapers, or magazines in their first or second language besides the mother language (Special Eurobarometer 386, 2012, 45).[73] This suggests that people read in various languages and may form different linguistic readership communities within the boundaries of one nation state.[74]

The same goes for the geographic visualization of translations according to the distribution of titles, publishers, book shops, and revenues,[75] illustrating the differences in the different national book markets. This then is used to present "a testament to the contribution of the book sector to cultural diversity in Europe (Turrin 2017, 6). Each visualization is accompanied by a secondary map that scales the sums according to a fixed threshold of inhabitants (one million).[76] As we can see, linguistic communities are disregarded in this

---

[73] See "Special Eurobarometer 386. Europeans and Their Languages," June 2012.
[74] Interestingly, according to the report, the five most widely spoken foreign languages in Europe are English (38%), French (12%), German (11%), Spanish (7%) and Russian (5%) (19), which also coincides with the biggest translation markets in terms of frequency of published titles (See chapter 1.2).
[75] For example, in a map showing the number of bookstores per country in total and per 100,000 inhabitants or the number of new titles published annually in each country (based on an average of the last five years).
[76] The report also includes percentages of readers for each country, which are derived from EUROSTAT.

model, which in many ways goes against suggestions made by scholars of sociological literary studies, such as Bachleitner and Wolf, to choose linguistic communities over a scaled count according to national borders. Thus, this focus on national borders exemplifies the gap between the ongoing discourses on multilingualism, translational literature, and linguistic communities which have been striving over the past 20 years in the humanities[77] and the publishing institutions that have been issuing these reports.

It is challenging to consider Bachleitner and Wolf's and Heilbron's language model based on linguistic communities, solely relying on a market-based and revenue-driven model, for the sake of analyzing translations in a library catalogue. First, a library catalogue does not include the main variable used for market analysis, revenue, and secondly, translations present a multi-variable problem: one original can have many editions in different languages and based on that, have a different status in the national collection (e.g., canonical, classic, part of a sub-collection). This underlines the specific challenges of using bibliographic data of translations in the context of literary analysis, which drastically differ from those of a market-related analysis, calling for the development of its own models and methods of analysis.

Additionally, translation oftentimes appears as a marginal category in publishing reports. In the FEP report, only the percentage between home and export, which "for German publishers, exports represent more than 15% of turnover, the top destinations being German-speaking areas (Austria and Switzerland), a number of other EU countries and the US" (Turrin 2017, 11) and the number of sales of translation rights (titles) are of any interest. Even though translation rights and exports here are measured by title and not limited to turnover, one cannot get a detailed picture of the yearly fluctuations of the translation market by language group or country from the report.

Translation rights and licenses as an analytical category appear to be rare,[78] while frequency counts per year per country are a more common category for analyzing literature in translation, as a similar report published by Literature Across Frontiers (LAF) (Büchler 2011)—a platform for Literary Exchange, Translation and Policy Debate—using the Index Translationum shows. The report titled "Publishing Translations in Europe" (Büchler 2011) gives a global picture of translation between 1990 and 2005, shows results by language, including regional, dead, and non-European languages, and then presents selected EU country

---

[77] I discuss the use of maps as analytical tools for publishing and especially translation data in chapter 4.
[78] Partially because the data needs to be requested from publishers directly through conducting surveys (as in the case of the association of the German book trade).

profiles as case studies with a special focus on fiction.[79] The data visualized in frequency graphs shows the number of translated books over time globally, for Europe, and then for each country. Mainly visualizing data as trends over time, the LAF report shows that "about 80% of all translated books in the world are published in Europe" (Büchler 2011, 6) and stresses that 1990 marks a turning year in translation where most languages saw a steep rise in the source language.

Regarding the average share of German language in literature translation in 26 European countries (1990-2005), the report states:

> French kept its advantage over German by about 500 titles in an average year, except for a brief period between 1995-1998, when German was a serious contender for the second position. Although the two languages rose from the 1500-2000 range to 2500-3000 titles per year, their weight within the "next 25" group became weaker. (Büchler 2011, 10)

After this global view on German as a language group and its ranking compared to the output in other languages, the report summarizes percentages of literary translations from all translated titles in Germany, Austria, and Switzerland separately[80] and from the total number of translations of all countries while focusing on the incoming translations. In line with the other reports, the LAF shows that Germany is among the most productive countries in publishing translations, especially for literature (7.9% of all literature translations). English, French, and Italian are the most popular languages of published translations in Germany. The report focuses on German as a target language, therefore any information on the distribution of translated fiction with German as a source language must be extracted from each country's profile and the global overview.

Therefore, the LAF report provides a valuable base of comparison for the translation of incoming fiction per country, but only includes limited information on outgoing translation per language group. The distribution of German fiction on a global market is therefore not evident and besides the Index Translationum, which is limited in several ways (e.g., timeframe), no dataset specialized on outgoing translations of German fiction is freely accessible to date. Further, detailed analyses regarding translated fiction in a given target language are usually treated as small case studies supplementary to larger reports. One of the

---

[79] In comparison to more market-related reports, language and fiction as units of analysis are prioritized. Up to now, this is one of the few reports that emphasizes fiction in translation, which also further investigates why the trends are happening (e.g., the success of single authors and prizewinners).

[80] Including the yearly number of new literature titles translated from English, Italian and French in Germany (1990-2005), the percentage of English, Italian, and French literary originals translated in Germany (1990-2005), the yearly number of translated new literature titles by original languages in Germany (1990-2005), and the percentage share of the most important original languages of literary translation titles in Germany.

reasons for this is that the data accessibility and availability, as well as inconsistencies in bibliographic datasets, are often identified as challenging for an in-depth analysis of fiction in translation. Thus, it is all the more crucial to compile, render accessible, and document inconsistencies of a dataset on German fiction in translation from the most comprehensive resource available.[81]

Modeling fiction in translation

To avoid inconsistencies in existing translation databases, some projects curate their own database with the aim to present an analytical tool and framework to study translation. In 2012, LAF published another short analysis with the title "Three percent? Publishing data and statistics on translated literature in the United Kingdom and Ireland" (Donahaye 2012) about a case study of translations in the United Kingdom and Ireland, including data from the British National Library—according to the authors, the best resource for bibliographic translation data—, which focuses on the availability and accessibility of translation data. For their sample years (2000, 2005, 2008), the authors present a detailed dataset with a focus on fiction and genres according to the Dewey decimal classification system, which allows for a literary study of translation, especially genres, and provides a feasible framework for structuring translation data. By using the information from the Dewey system, the authors classified each bibliographic record according to poetry, fiction, and drama.[82] In summary, the findings include the percentage of all publications that are translations, the percentage of poetry, fiction and drama that is translation, and the percentage of all literary genres (the entire 800 Dewey range) that is translation, showing that translated fiction in the UK and Ireland is higher than the previously anticipated 3%.

In conclusion, this report clearly shows the benefits and caveats of working with translation data and points out that a detailed bibliographic dataset of translations is difficult to acquire. For instance, the authors agree with Pickford when arguing that resources such as the Index Translationum display inconsistencies in the data, due to the limits of the compatibility and standards of bibliographic cataloguing—which are addressed in detail in

---

[81] As I illustrate in chapter 2.0, even among national libraries that share the same official languages (such as Germany, Switzerland, and Austria), some appear to include more translations than others and each library catalogue therefore represents the overall distribution differently. This is especially apparent when comparing the different conditions of each national library and their laws regarding which works are collected. For instance, the law according to which each work that has its author's rights in Germany needs to be deposited at the national library. That way, translations are included in the library catalogue.

[82] As I argue in the following chapter, I have found the same to be true for data of German fiction in translation, influencing the decision to base this thesis project on a dataset extracted from the German National Library.

the following chapter. This is a common obstacle when working with translation data and limits the scope of analysis to a sample of years, genres, or authors.

Another series of reports that provides a framework focused on fiction in translation is the Diversity Report. Following previous reports published in 2008, 2010, 2016, and 2018, the Diversity Report[83] asks "who is translated across Europe?" and "what languages are more receptive for translations, or are better received by other territories, than others?"[84] by observing the relation between factors (such as author frequency and target languages, the role of bestsellers, and genre- and theme-related factors) and ranking authors and titles according to "impact points" which are based on the number of titles sold. The dataset is comprised of a list of 500 authors (bestselling, mid-list for book awards), 250 of which were selected and tracked in retail book catalogues bestseller charts for the respective years, across eight major markets. The report seeks to compare the authors' performances and to understand "non-traditional" authors and publishing models (e.g., self-publishing). From this list, Wischenbart et al. assembled a bibliographic dataset of 2000 entries for 12 countries according to "top sales ranks for titles in a half-dozen Western European markets of different sizes, namely Germany, France, Italy, the Netherlands, Spain, Sweden, and the United Kingdom, plus the United States" (Wischenbart et al. 2018, 19). In each report, selected case studies are presented by country, e.g., Austria and Slovenia in the 2018 report, or by specific authors to show larger trends, such as best-selling titles that were also featured in cinema and TV.[85]

Even though the Diversity Report represents one of the most comprehensive analyses of fiction in the German-speaking countries, their dataset focuses on high-prestige authors and titles. Suggesting that 500 authors are representative of the literary field in a given country contradicts what Pickford identifies as the importance of independent publishing mechanisms and the role of marginal authors for the literary field of a given country. Again, this model seems to align with a revenue-based approach as presented by the FEP report

---

[83] During the process of writing this thesis, a 2020 report came out, for which an author dataset has been curated of 102 authors (prizewinning, bestselling, newcomers) alongside a list of the 1,000 top selling titles in fiction to further zoom in on self-publishing and bestselling authors and their relation to other industries such as film, TV, and games. See Wischenbart et al. 2020.

[84] See: Kovač, Miha., Wischenbart, Ruediger Wischenbart., Jursitzky, Jennifer., Kaldonek, Sabine., Coufal, Julia.,. "Diversity Report 2010 : Literary Translation in Current European Book Markets : An Analysis of Authors, Languages, and Flows." [Wien?]: Wischenbart.com, 2010. http://www.wischenbart.com/diversity/report/Diversity-Report_2010.pdf.

[85] Additionally, Wischenbart et al. identify factors that the top selling authors had in common: "Writing a small series, like a trilogy (like Stieg Larsson or E. L. James), or creating a hero who appears in a dozen books (like Camilleri's Inspector Montalbano), is another helpful factor. But being well-branded as a writer, with an easily recognisable topic, can do the job, too. In addition, being published by an industrial corporate media house can certainly boost sales, but independents have some muscle to show as well" (Wischenbart et al. 2018, 24).

revealing a bias toward high-prestige translations; thus, the initial motivations to tackle diversity on the publishing market become problematic. Here we may argue that in addition to dataset availability and accessibility, one should consider bias of dataset curation and interpretation.

Bias is not only a factor inherent to the dataset, but also the presented findings, and one may ask if quantitative bibliography-based history of translation actually provides an idea about the field of translation or only the available datasets and their limitations. As an example, in the 2018 and 2020 Diversity Reports, another focus is funding schemes in support of translations (e.g. http://www.eurolitnetwork.com/european-literature-network/ or EUPL) and their impact on the translation market, which significantly pushes translations in certain target languages.[86] One of their main findings is the persisting bias on the translation market:

> […] only a remarkably small number of authors succeed in finding broader readerships across multiple languages, while most mid-list writers are well received in only a few other tongues and that even a successful translation of one work does not necessarily guarantee that subsequent books will be allowed to travel across linguistic borders. (Wischenbart et al. 2018, 10)

According to Wischenbart et al., "Western European bestseller markets are a regionally enclosed, globalised and male-dominated area" (Wischenbart et al. 2018, 27) with European bestseller lists are widely dominated by few western European authors, with women underrepresented. Here Wischenbart et al. overlook what I consider an important point about the limitations of a prestige-centred dataset curation. When only focusing on the best-selling or prize-winning authors, the findings presented do not come as a surprise, since gender-bias on mid- and final lists is not an uncommon fact.[87] The scholars therefore dismiss the majority of small-scale authors that play key roles in the transfer of literary works. To return to my previous question, between a quantitative analysis of bibliographic data leading to a comprehensive understanding of the translational field or being limited by data quality and accessibility, it becomes clear that the latter is often the case. This is especially problematic when drawing general conclusions about the translational landscape.

---

[86] The authors look specifically into German and stress the role German plays as a gateway language to the global translation market: "The observations of this report on Central Europe call for a hypothesis that a) there is a regional Central European translation area, and b) for the international success of Eastern or Central European authors, being published in German is a more natural first step than being published in English or French" (Wischenbart 2018, 74).

[87] See: Cima, Rosie. "Bias, She Wrote. The Gender Balance of The New York Times Best Seller List." Accessed August 23, 2022. https://pudding.cool/2017/06/best-sellers/.

Another problematic result Wischenbart et al. (2018) report is the gap between translated titles and demography with "hardly any titles from the overall output targeting inhabitants other than German non-migrant middle-class consumers" (Wischenbart et al. 2018, 17), while "the sheer number of languages is not in itself an evidence of diversity" (Wischenbart et al. 2018, 63). The simple assumption here is that a native Turkish speaker in Germany would read Turkish literature in German translation and not in Turkish, which is problematic considering studies showing Turkish readership mainly consumes literature in their maternal language.[88]

Additionally, in 2020 Wischenbart et al. also found that multicultural authors who write in more than one language also ranked higher than expected, especially if they also published in a strong target language such as German (see Wischenbart et al. 2020, 36). Similar to the FEP report, language is not the main unit of analysis, but instead the author and their position on the translation market is the focus. Thus, the main analytical framework presented in the diversity report is based on literary prestige and the factors that contribute to the ranking of certain authors and titles. For their data, they prioritize online book retail catalogues and not national bibliographies. It is not possible to say if the data is representative, especially since a complete list of authors is not enclosed. As mentioned in the 2020 report,

> The aim of this was not to build a complete bibliography but to look for publications that an interested reader and consumer could easily find and be able to purchase. (Wischenbart et al. 2020, 65)

As this is the case in most reports, the diversity report also focuses on market-driven data analysis aiming at publishers and readers rather than an academic audience. Both restricted accessibility to the curated dataset and non-formalized data curation practice make an evaluation of data quality according to other reports challenging, underlining the necessity that a comprehensive model applicable across datasets must address issues related to availability and accessibility, national versus linguistic spaces, and the observed translation bias.

Designing a model for a quantitative, sociological analysis of translation

In conclusion, in designing a methodological model for a quantitative, sociological analysis

---

[88] See: Teichmann, Lisa. "Migration and Literary Education: A Comparative Study of the Turkish American Community and the Turkish Austrian Community, San Francisco and Vienna." Marshallplan Foundation, 2016.

of translation, the proposed model in this thesis builds on and responds to the following points made in the publishing reports and studies discussed in this chapter:

First, reports and sociological studies both observe an uneven distribution in translation across languages, but the relation between different responsible factors is not clearly outlined. Therefore, a sociology of translation proposes to analyze the positions of languages in the field of translation, in addition to the role of publishing centres. This can provide a useful methodological framework based on bibliographic data to analyze translation and the language network. However, while there are some initial approaches and suggestions to use descriptive statistics and sociological approaches for which language is the main unit of analysis, there is no coherent methodology yet established to study translation quantitatively. Hence, the proposed model in this thesis builds on these preliminary approaches by introducing different measures for analyzing uneven linguistic and publishing centres by means of social network analysis and geographic mapping, as well as descriptive statistics.

Secondly, market-based analysis often overlooks the importance of linguistic groups versus national boundaries. Additionally, literary categories such as genre and period, as well as style and narrative, are broadly overlooked. From the publishing reports alone, for example, one cannot understand the role translation plays in the distribution of fiction on a global scale and how it transgresses national and linguistic boundaries. Sociology of translation suggests analyzing language groups as a unit of analysis and modeling the positions and interactions of those within a field, while the reports simply display a monolithic map of translation organized according to nation states and their respective markets. In literary studies, the field of literary sociology is devoted to analyzing the context of the book publishing market by looking at specific factors and the flows between language groups instead of nation states. Like Heilbron (1999) points out, the size of language groups does not correlate with their position in the translation system, which challenges the way the reports refer to population size and translations as causal variables. In line with Heilbron, my model also takes language as the unit of analysis in modeling translation networks and identifying sub-communities.

Third, data access and availability, as well as inconsistencies, are a major challenge for a quantitative study of translation. Previous studies in literary sociology all mention that existing databases are a valuable resource for studying the translational cultural field; however, a comprehensive application of their categories and factors have not yet been done. The main challenge for a sociological or any quantitative analysis of translation is that a

universal unit of analysis such as language would require formalized data curation practices and consistent categories. As I have outlined in the discussion of the aforementioned reports, this is not the case. However, as the studies by Norrick-Rühl and Bold (2016), Zhou and Sun (2017), and Pickford (2016) show, national library bibliographies, if carefully curated and annotated, can provide a comprehensive resource to study translation. In response to the lack of datasets for German fiction in translation, this thesis presents the first study of translations in the German National Library catalogue. The following chapter is dedicated to data accessibility and quality in regard to translations in the German National Library catalogue and the dataset that has been curated for this thesis project.

2.0 Extracting translations from the DNB

In line with the premise of this thesis—to make translations visible in the library catalogue—in this chapter I document and evaluate the data extraction process, cataloguing practices, and data quality, with special attention to the challenges and limitations of the applied approach. I want to show that translation data challenges the very structure of the catalogue because it poses challenges in extracting data due to translation not being a separate category, which in turn calls for a prior extraction filter and definition of translation. As mentioned in the introduction, this project focuses on outgoing translations (works originally published in German), since this thesis aims at mapping the geographic and linguistic traces of German fiction by means of translation. The extraction method and resulting dataset, comprised of over 35,000 titles in 86 languages, therefore correspond to this focus.

"Where is the translation data for outgoing German fiction translations and how can we extract it?" is the first question that arises when searching for translations of German fiction in the German National Library catalogue. Since most of the time translations are not their own category, they visually are not different from non-translated catalogue entries. Translation data therefore is often at first invisible in the library catalogue and can only be found based on a well-designed search query that includes some sort of definition of what a translation is. Poupaud et al. (2009) has described this central challenge of finding and extracting translation data by narrowing it down to a matter of definition and filtering. Poupaud asserts that "the term 'translation' needs to be defined explicitly' (Poupaud et al. 2009, 268) according to the operationalization of translation through the prior filter and the research filter that either do or don't give translation status to a given work of writing. A prior filter, for example, may be institutional, as in the case of a library catalogue where the selection and annotation process are handled on an institutional level, or commercial as in the case of bookseller databases. Additionally, a research filter is developed through the final selection of what is defined as translation by the researcher, which varies from project to project. It is therefore necessary to transparently discuss the pre-existing and applied filters as part of documenting how translations have been collected and curated into a dataset. Data extraction thus depends on how translation is defined and how it is operationalized as a filter.

How can we define outgoing translations in order to extract them from library catalogue? And how does this condition the data quality and representativeness? First, it is important to note that this thesis builds on Reynolds and Vitali's definition of translation as

an "act of translation" that includes "both the first publication of a new translation and its republication in a different place" (Reynolds and Vitali 2021, 3), meaning that each title in the dataset is seen as a unique linguistic event and can therefore include various editions which may be reprints of an existing translation. Secondly, for this project, a translation is defined as a work which has been published in a different language than what it was originally published in. As part of my research filter, the bibliographic entry therefore always includes at least two languages: one source and one target language.[89] However, this definition and the resulting dataset brings with it specific challenges, such as data representability, quality, and reliability, which are in turn conditioned by cataloguing practices.

In order to highlight the challenges and limitations of extracting translation data from the German National Library catalogue, the aim of this chapter, then, is to provide a field guide that includes various measures for data quality and representativeness to make visible the prior and research filters and contextualize them in the light of cataloguing practices. The following data quality assessment takes into account the various mentioned sources of bias to provide an overview of how representative the extracted dataset of German fiction in translation is, as well as to guide the reader through any data inaccuracies or inconsistencies that result from cataloguing practices. In order to document the various biases of the applied filter on the relevant variables and the resulting dataset, I discuss four major categories of data quality assessment methods: consistency, sampling bias, accuracy, and completeness.[90] After presenting an overview of how representative the dataset source (the DNB) is in comparison to other major translation databases—comparing the selected records with the Austrian National Library (ONB), the Index Translationum, and VIAF to review inconsistencies across datasets—this chapter moves on to document cataloguing practices and further evaluate the fitness of purpose for the dataset source. Additionally, a brief assessment of representativeness and bias is conducted by measuring falsely identified translations on a subset of data. Variable accuracy describes the errors and data heterogeneity within the DNB translation dataset by measuring the ratio of data to errors for all relevant variables. Completeness refers to the amount of missing data, which is important to assess before any

---

[89] For the compilation of the dataset, only the entry for the target language will be compiled, since tracing each translation to the original edition it is based on appears to be impossible within the framework of this project.
[90] See Olson 2003 chapter 2, 24-42, whose assessment of data quality and bias included the following: value representation consistency (varying orthography), changed induced inconsistencies (changes in the way the data has been recorded), valid values (value is accurate and consistently used), missing values, object-level accuracy (if objects are missing in database, appears to be complete but is not), and object-level inconsistencies (fluctuations in changes made to the dataset by removing or adding data).

manipulation or data modeling takes place.

Inconsistencies across databases: the DNB catalogue compared to ONB, VIAF, and IT

As a pre-step to data curation, the project had to evaluate how inconsistent or incomplete the data of the German National Library is compared to other library catalogues from German speaking countries. To see in what ways this database may be missing records compared to other translation databases and catalogues, I outline some specific records as they appear in the German National Library catalogue (DNB), the ONB, the VIAF, and the Index Translationum (IT) and calculate the percentage of values that match across records.

Resource	Years	Number of translated titles
DNB	1980-2020	35,972
VIAF	?-2017	11,767
IT	1979-2005	301,935 (including non-fiction)

Table 1. Sums of translated titles of German fiction in the DNB, VIAF, and IT databases.

The main challenge in assessing inconsistency lies in the timeframe and scope of data entries. While in terms of the number of translations the Index Translationum appears as the most comprehensive resource, it does not include any data after 2005.[91] The IT mentions German as number one of the top 10 target languages[92] with 301,935 translations documented between 1979 and 2005, while my dataset consists of 35,972 translations from German (fiction and non-fiction) in total between 1980 and 2020. VIAF has a total of 11,767 entries of works for German fiction in translation for a timeframe until 2017. Not only do we see significant differences in the size of the databases—the IT appears to be a very comprehensive resource comparatively, while VIAF seems to not even include half of the translations extracted from the German National Library—both VIAF and IT's timeframes do not reach that of the German National Library. Due to a lack of a comprehensive report and documentation, it is only possible to roughly compare the datasets, which is why a closer

---

[91] At the time of data curation, the Index Translationum web portal was under construction and no assessment could be made on the current state of the dataset. An initial extraction in 2017 showed that the dataset did not include work titles or publication year and publisher information, but only raw counts of translation numbers into each language.

[92] The latest extraction of March 2019 revealed close to 30,000 entries for German fiction. The database was under construction from 2019 until December 2021. See Index Translationum. "Contributions from Countries." Accessed January 28, 2021. https://www.unesco.org/xtrans/bscontrib.aspx?lg=0.

look at a subset of data to highlight the caveats is necessary at this point.

When comparing the DNB with ONB, VIAF, and IT for some prominent authors who have been originally published in German, we can get a sense of the dataset consistencies across different catalogues and databases. The Austrian author Ingeborg Bachmann's novel *Malina,* for instance, appears with 34 entries of translations in the ONB catalogue and 51 entries of translations in the German National Library dataset, which suggests that the latter is a more comprehensive resource. From the Austriacae dataset, only 18 entries of *Malina* translations are included, most likely because the novel was first published in Germany, and therefore most translations were registered at the German National Library. VIAF only lists eight entries for *Malina*: the Russian, Polish, Hebrew, French, and English translations are all part of one expression alongside the German original.[93] In comparison, the Index Translationum cites a total of 11 translations of Ingeborg Bachmann.[94] In summary, this suggests that the German National Library includes the most records on Bachmann's *Malina*, while other databases have less information, both in terms of catalogue entries and data categories available.

Another example that illustrates some of the dataset inconsistencies is Patrick Süskind's novel *Das Parfüm* (1985), which was first published with Diogenes in Zürich. The novel has 49 translations and received the PEN translation prize in 1987, as well as bestseller list status. ONB only lists the Bulgarian translation from 2007,[95] while the DNB lists 106 translated editions among 20 publications in German and many schoolbook adaptations. VIAF lists nine translations (Arabic, three editions in English, Hungarian, Croatian, Korean, Polish, and Russian).[96] Again, similar to the previous example, the German National Library appears to include the most records for this prominent novel.[97]

Kafka's *Der Prozess* is a second example of one of the most well-known German novels whose translations further highlight some inconsistencies across catalogues and datasets. *Der Prozess*, first published in 1925 by the Berlin publisher Die Schmiede, is also the example with the most bibliographic entries across all databases. While VIAF lists 92 expressions (10 of which are Chinese editions dating back to 1930),[98] the ONB lists 632, 201

[93] VIAF. "Bachmann, Ingeborg, 1926-1973. | Malina." Accessed August 23, 2022. http://viaf.org/viaf/182465068.

[94] The IT only includes author name and translation frequencies per language, which complicates tracing single works.

[95] Österreichische Nationalbibliothek. "Parfjumăt : Istorijata Na Edin Ubiec . 3. Izd." Accessed August 23, 2022. http://data.onb.ac.at/rec/AC08832393.

[96] VIAF. "Süskind, Patrick 1949- Das Parfum." Accessed August 23, 2022. http://viaf.org/viaf/220436158.

[97] Due to access and availability limitations, the Swiss Library Catalogue is not included here.

[98] VIAF. "Kafka, Franz, 1883-1924. | Prozess." Accessed August 23, 2022. http://viaf.org/viaf/183949498.

of which are translations, including multiple editions of translations into the same languages. The DNB lists 459, of which 416 are annotated as not German. Compared to the previous examples, this seems to be by far the most consistent data across all library catalogues, while the ONB and DNB appear to be comparably consistent in their translations for *Der Prozess*.

The above examples show how inconsistent databases are and the related challenges of working across national bibliographies. These examples zoom in on some underlying challenges when working with bibliographic data of German fiction. The biggest challenge in assessing inconsistency across databases and catalogues is that translation is not annotated in most bibliographies, and even translation databases that combine data from national bibliographies such as VIAF and Index Translationum are not following a standardized format. These examples nonetheless suggest that the German National Library is the most comprehensive catalogue for translation.

Data collection

The translation dataset analyzed for the purposes of this dissertation has been extracted from the German National Library, which has been collecting all translations by authors based in Germany.[99] The Datenshop of the Nationalbibliothek offers a free, open-source resource[100] for all bibliographic data of the library's catalogue. According to the legal deposit regulation (PflAV), "all publications issued in Germany, irrespective of their language" and "media works published abroad for which a publisher or a person who has a legal domicile, business premises or their principal residence in Germany has sold (licensed) the right to publish the work abroad"[101] need to be deposited at the German National Library. The latter specifically includes translations published outside of Germany.

An expert search for each year of publication was conducted in order to filter out translations based on my definition: any work classifies as a translation that has been originally published

---

[99] One set of bibliographic records may not exceed 10,000 entries and a user can conduct 200 queries.
[100] See "Creative Commons Public Domain." Accessed August 23, 2022.
https://creativecommons.org/publicdomain/zero/1.0/deed.
[101] Gesetz über die Deutsche Nationalbibliothek (DNBG)
§ 17 Auskunftspflicht: "Die Ablieferungspflichtigen haben der Bibliothek bei Ablieferung der Medienwerke unentgeltlich die zu ihrer Aufgabenerfüllung notwendigen Auskünfte auf Verlangen zu erteilen. Kommen sie dieser Pflicht nicht nach, ist die Bibliothek nach Ablauf eines Monats seit Beginn der Verbreitung oder öffentlichen Zugänglichmachung berechtigt, die Informationen auf Kosten der Auskunftspflichtigen anderweitig zu beschaffen." See Deutsche Nationalbibliothek. "Sammlung Körperlicher Medienwerke." Accessed January 28, 2021.
https://www.dnb.de/DE/Professionell/Sammeln/Koerperliche_Medienwerke/koerperliche_medienwerke_node.html.

in German while having another target language. The following line of Boolean search code extracts all works of fiction that originally appeared in German and then were published in other languages. In other words, it says: search for any work that has German as the original language (spo=ger) and limit all results to German literature (sgt=59) and fiction (sgt=B) for the year 2004 (jhr=2004).[102]

spo=ger and (sgt=59 or sgt=B) and (jhr=2004)[103]

For each year, the data has been extracted in CSV (comma-separated values) format from the Datenshop.[104] First extractions were done in 2017, while the final dataset was compiled in April 2021. As stated on the library's Datenshop page, metadata is constantly being updated, which may lead to varying numbers of titles and therefore requires running any statistical testing on the most recent dataset. A set of data has been extracted for each year between 1990 and 2020.[105] With this query, the following catalogue entry can be extracted:

Link zu diesem Datensatz	https://d-nb.info/970660944
Titel	1979 / Christian Kracht. [Trad. de Carmen Gauger]
Person(en)	Kracht, Christian (Verfasser)
Werk(e)	1979 (span.)
Verlag	Madrid : Alfaguara
Zeitliche Einordnung	Erscheinungsdatum: 2004
Umfang/Format	161 S. ; 23 cm
ISBN/Einband/Preis	84-204-6526-7 kart.
Sprache(n)	Spanisch (spa), Originalsprache(n): Deutsch (ger)
Sachgruppe(n)	830 Deutsche Literatur ; B Belletristik
Leipzig	Signatur: 2005 A 26906 Bereitstellung in Leipzig

Figure 1. Catalogue entry for a translation as shown in the database.

With the above query, a total of 35,972 outgoing translations were extracted. When

---

[102] In order to collect incoming translations to study the role of translation in the German publishing landscape, a different method would need to be developed that includes conditions as to the work being originally published in another language as German, but published in Germany, Austria, or Switzerland. This can be achieved by simply designing a similar search query that follows the same logic as the one used for this dataset, e.g., "geo=de and spo=eng and spr=ger and sgt=B" where "geo" selects works published in Germany, "spo" selects the original language and "spr" again selects publications in German.

[103] "spo" stands for original language, "sgt=59" for German literature, and "sgt=B" for Belletristik (fiction).

[104] Additional options for formats include MARC, which requires transforming the data into tables by using a Python script. Hence, CSV was the most accessible format to use in RStudio for the proposed model.

[105] In 2020, numbers for that year were still very low (64 book titles in May) relative to the average amount of titles per year, which is why the final dataset for 2020 was extracted in 2021.

looking at the total number of titles of fiction in the catalogue by using the same query omitting spo=ger, we can see that translations occupy 3.5% of fiction titles in the catalogue (n= 1,002,420 titles of fiction in all languages)[106].

Translations in the DNB catalogue

In this section, I unpack what can be expected within the library catalogue in terms of translations. This is important because it is necessary to establish how representative the extracted dataset from the German National Library catalogue is compared to other translation databases and to explore how cataloguing practices may pose specific challenges to modeling the extracted data.

Sampling bias

In order to establish how confident we can be that the search query indeed extracted translations based on my definition, I measured the false positives for the year 2020 in a subset of the data by manually annotating false positives.[107] Based on the research filter and definition of translated work, I counted all entries (n=552 from 35,972 total records in the catalogue for the year 2020) that have the language marked as German as translations into German, accounting for 28 of 552, which corresponds to 5%. However, identified outliers that can still be included are works in more than one language of translation, multilingual works,[108] and a small number of translations from English, which were published in Germany (3) and therefore selected. Additionally, translations from a German dialect (such as Plattdeutsch) are also included. Anthologies of selected texts by non-German authors also appear among the extracted catalogued items. Only three false positives[109] were included in the dataset for all entries marked as German. Additionally, 65 (11%) of all extracted translations have a publication place in Germany. This is an especially interesting category since after close examination, most of these records are self-published translations and

---

[106] When applying the query mentioned above for incoming translations, we can see that they are significantly more present in the catalogue with 114,969 titles of fiction corresponding to 11.4% of all fiction titles. These numbers correspond to the state of the catalogue in August 2022.

[107] A comprehensive analysis of precision and recall as applied by Erlin et al. (2021) would require having a comparable dataset with complete, consistent bibliographic data in the same categories as the DNB. As discussed in this chapter, due to data inconsistency in translation datasets, this lies beyond the scope of this project. For the same reason, an assessment of false negatives is not included here.

[108] Such as *Deutsche Gedichte zweisprachig: Kurmanjî-Kurdisch/Deutsch / herausgegeben und übersetzt von Abdullah İncekan* which was marked as German.

[109] False positives stand for titles that have been extracted as translations from German, but are actually translations from another language.

smaller publishers focused on translations. Overall, only the previously mentioned three translated works from English into German published by a German publisher can be identified as false positives. Based on this assessment of a smaller subset, it can be deduced that the definition and designed search query enable a comprehensive extraction of translations without producing a lot of noise in the data by also extracting non-translations.[110]

Summary statistics illustrating cataloguing practices

Secondly, besides defining a filter for extraction and establishing a definition of translation that reveals a reliable dataset, documenting cataloguing practices is another important challenge when working with translation data from the German National Library catalogue.

First, it is important to note that works of translations are not consistently catalogued[111]. As identified by Mäkelä et al. (2020), cataloguing practices often are represented by gaps in bibliographic datasets, which, according to Poupaud's argument, constitute the prior filter based on which translations are catalogued. I was informed by the Datenshop of the German National Library that the codes for language were not assigned before 1992. From 1992 onwards, they were assigned consistently to all Germanica and translated German works (cod=ru), and only from 2010 onwards to all publications; however, as part of their ongoing cataloguing work, they retroactively assign language codes to older entries. Each extracted dataset therefore may vary in completeness of variables.

Secondly, like most library catalogues, the German National Library also does not include a separate field to mark translations. The common indication that a work is a translation may appear as plain text (trans./Übersetzt von/übers./trad./çeviren/etc.) in another field and does not follow a standardized format. Irregularities like these, and the fact that translations are not categorized as such in the catalogue in a standardized way, are a drawback of current cataloguing practices and need to be investigated prior to data extraction.

Inconsistency in cataloguing processes pose additional challenges and limitations for extracting and modeling the translation data of the German National Library catalogue. Annual fluctuations in title data persist and visualizing the frequency distribution of title

---

[110] To assess false negatives, a classification of all records from 2020 would need to be done on a total of 100,545 titles (translations included) in the fiction category, which at this current point in time lies beyond the scope of this study.

[111] One question I asked myself was if I could combine the DNB dataset and Austrian and Swiss catalogue data. However, in the case of the Austrian National Library, no large extraction of translation data is possible because the collection Austriacae includes all Austria-related publications. Additionally, "Belletristik" is not annotated. See Österreichische Nationalbibliothek. "Katalog." Accessed January 28, 2021. https://labs.onb.ac.at/de/dataset/catalogue/.

sums per year reveals cataloguing practices, raising the question of whether the extracted dataset can support a longitudinal analysis. In order to show cataloguing practices and how the extraction date affects the number of translated titles, I compare the title frequency distribution per year for two different extraction dates of the dataset (April 2021 and January 2019).
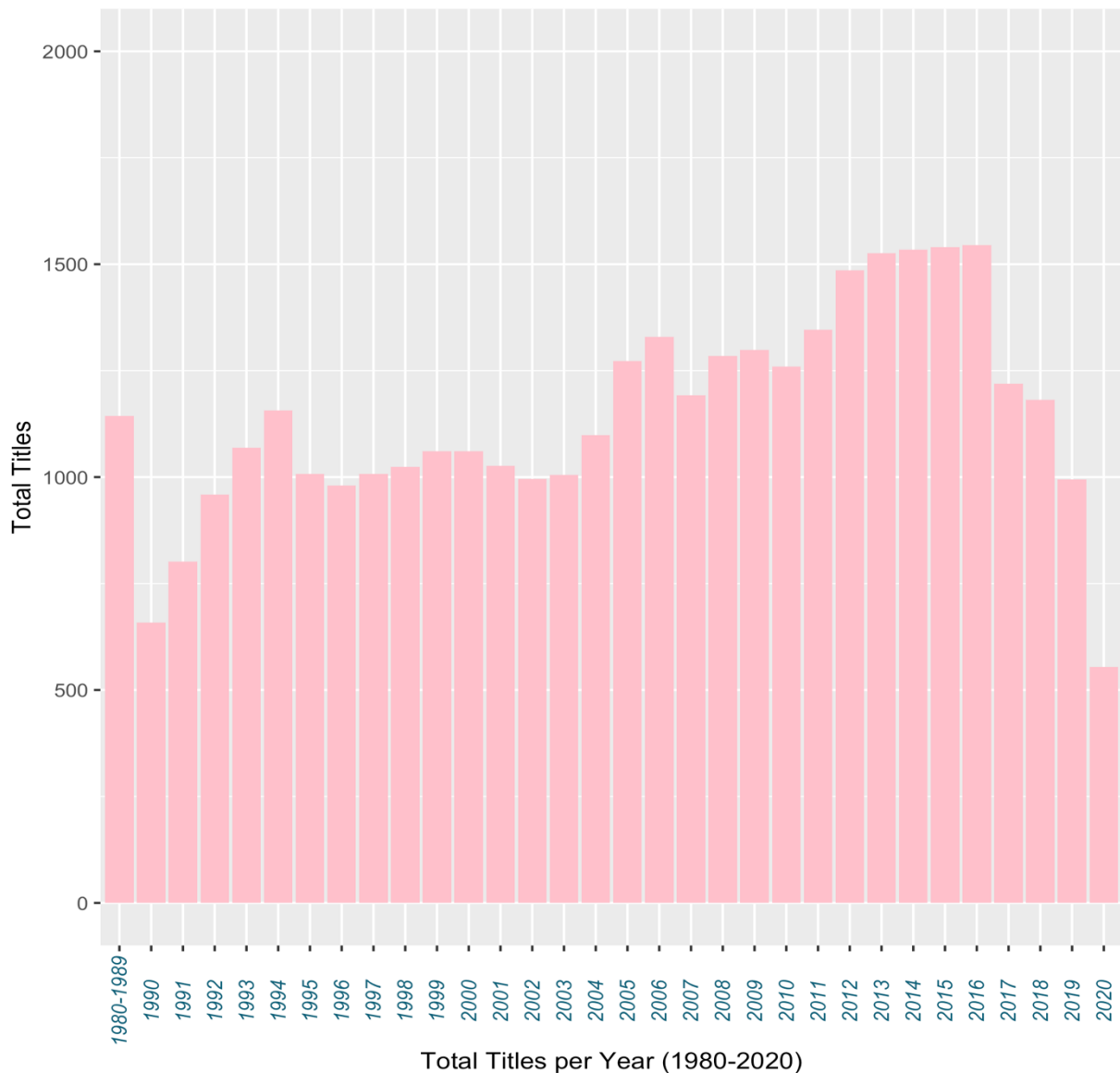
Figure 2. Title sums per year (1980 until 2020) from the German National Library (all translated works with German as the original language). 1980 until 1989 are combined due to the low number of titles (under 500). Data extracted: April 15, 2021

Comparing figure 2 and 3 shows that the date of extraction affects especially the years closest to that date. For example, figure 2, which shows the frequency distribution of translated titles per year in the DNB catalogue, reveals how the title count for 2020 is still very low, even though the data was extracted in 2021.
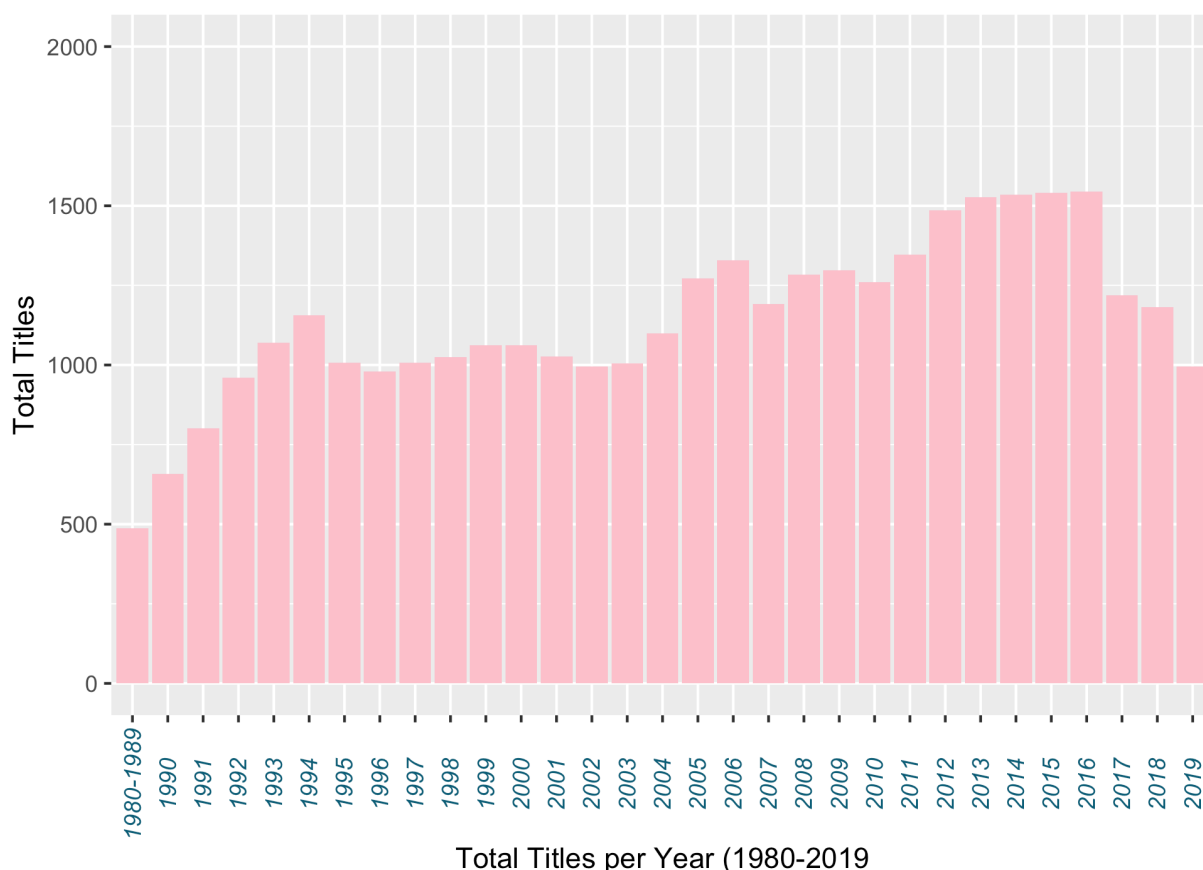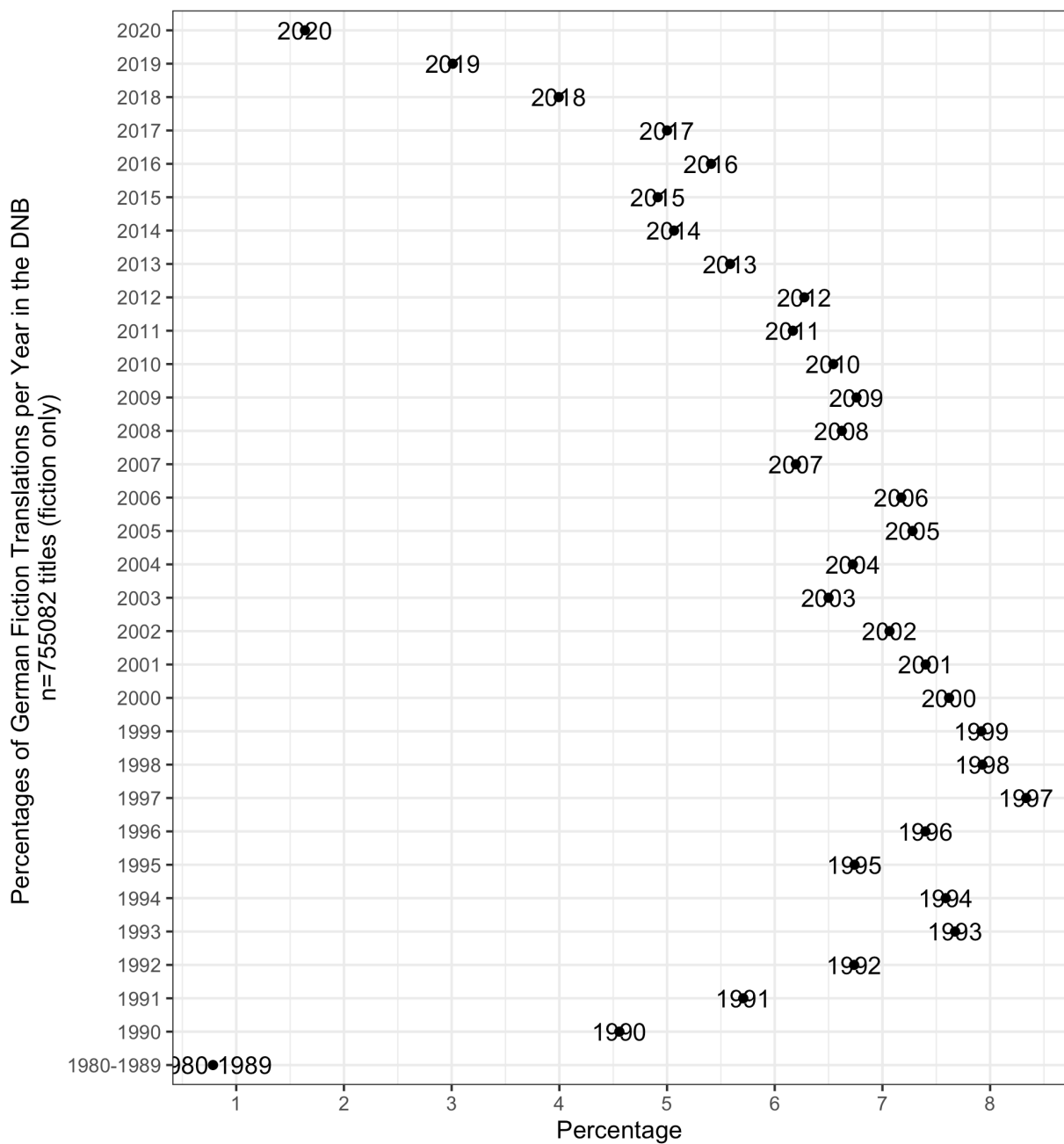
Figure 3. Title sums per year (1980 until 2019) from the German National Library (all translated works with German as the original language). 1980 until 1989 are combined due to the low number of titles (under 500). Data extracted: January 17, 2019

Figure 3 shows that, compared to the data extracted in 2020, we can see how cataloguing practices have affected specific years. For example, as figure 2 shows, more titles were added for the years 1980-1989 after the extraction in 2019. On both figures, the overall number of titles is increasing, especially from 2012 until 2016. Lower numbers before 1994 may both be correlated to lesser submissions to the German National Library collection or to a decrease in translations from German overall. For both extraction dates 2016 appears to be the year with most translations which again only shows that there were significant efforts to catalogue this year.[112] In summary, this comparison shows that the years closer to the extraction date appear to be underrepresentative of translated titles; these titles are then added in the successive years. Additionally, specific year ranges (especially pre-2000) appear to receive a retroactive boost in their catalogued translations, it is important to take these fluctuations into account and hence also complicate any longitudinal analysis—which I address in the introduction and discussion chapters as part of this thesis' limitations.

---

[112] This has been confirmed by the librarians of the Datenshop.

A second method to assess possible challenges related to cataloguing practices is to measure how the number of translations compare to the overall number of translations in the catalogue, or in other words, how large is the share of translations in the DNB catalogue? A related question is how large is the share of fiction in translation vs non-fiction? This was done by calculating percentages of German fiction in translation in the library catalogue for each year. By comparing percentages of fiction per year with the frequency distribution of fiction and non-fiction, the space that translations occupy in the catalogue becomes visible. As mentioned above, the number of outgoing translations in the catalogue is at 3.5% of all fiction titles in the DNB catalogue. However, this number also changes according to cataloguing practices. Again, when comparing the percentage of translated German fiction to non-fiction titles in the catalogue, we can see especially which years are over- and underrepresented.

Figure 4. Title percentages of fiction per year (1980 until 2020) from the German National Library (all translated works with German as the original language). 1980 until 1989 are combined due to the low number of titles (under 500). Data extracted: April 15, 2021

Figure 4 shows that at the time of the last extraction in April 2021, translated fiction has the highest percentage of all fiction in the DNB catalogue in 1997 and the lowest in the 1980-90 range and 2020 (the year closest to the extraction date). Again, considering that after 1997 cataloguing improved for fiction, this comparison appears to also illustrate cataloguing practices, and a longitudinal analysis of this data therefore further lies beyond the scope of

this study.[113]

Data quality of variables

Cataloguing practices affect not only the title frequencies and representability of the dataset, but also the data quality of specific variables. This section is dedicated to the data quality and specific challenges related to each variable in the dataset.

Variables

Before measuring the data quality of each variable, it is necessary to introduce what type of metadata is contained in the catalogue and which variables are central to my analysis. By applying the expert search, all titles were extracted as csv's containing the following metadata: subject, identifier, type, creator, title, volume, edition, publisher, year, format, ISBN, ISSN, ISMN, EAN.UPC, binding.price, language, country, date.of.publication, description, collective.title, links, relation, rights, subject.headings, connected.titles, uniform.title, frequency, and wv.number.

subject	Genre. A subset by subject according to Deutsche Literatur ; B Belletristik (Fiction) has been extracted
identifier	URN, URL, IDN or ISBN Contains information about the binding
type	e.g. "online resource", NA for print
Creator	name of the author or translator. Also includes associates ("Mitwirkende") and artist ("Künstler") Not all titles include a translator
title	work's title in translation and in the original, may also include the editor
volume	book volume
edition	noted in the target language (1. Auflage, First printing, Achtste druk etc.)
publisher	publishing house and location
year	year when the translation was published
binding.price	contains information about print (hardcover, etc.), price in Euros not always included
language	language of translation (target language)
country	country code of publication
date.of.publicationand description	mostly missing

---

[113] When comparing figures 4 with a figure for fiction and non-fiction combined (see appendix: figure_ch1.2_dnball_percentages_fiction_v2), we can see that for specific years (such as 1997 as compared to 1993), publications of translations of fiction versus translations of non-fiction do not overlap. 1997-2000 have by far been the best years for cataloguing translations of fiction, while for non-fiction that period came earlier (1992-1994).

collective.title	n cases when the work was published as part of a series (e.g. Linskog Quality Classics)
links	URLs to content and chapters
relation	links to entries within the catalogue of the German National Library
rights	only missing values
subject.headings	thematic categories
uniform.title	German title of the original[114]

Table 2. Metadata fields for CSV (Comma-Separated Values) formatted data in the German National Library catalogue.[115]

For the data collection and analysis, "language" and "country" serve as the main variables. The language variable serves as the main category for language distribution. The publisher field includes the publishing place, which is another variable for mapping and spatial analysis (see chapter 2) in addition to the country variable. Selecting the variables for analysis after extraction also aids in evaluating how reliable those variables are in terms of completeness and error rates, which I discuss in the following two sections.

Completeness

Measuring missing values across variables also shows that my main variables don't raise major challenges. Completeness is assessed by counting missing values for each category to assess which variables can be used for analysis. Some of the most complete categories are publisher (6.9% NAs),[116] author (6.5% NAs), country (7.5%), ISBN (6.3%), and format, which is the most complete category with only 34 missing values.

---

[114] In addition, there are several categories for internal use, such as links, relation, and unique identifier numbers, which are not central to the research questions.
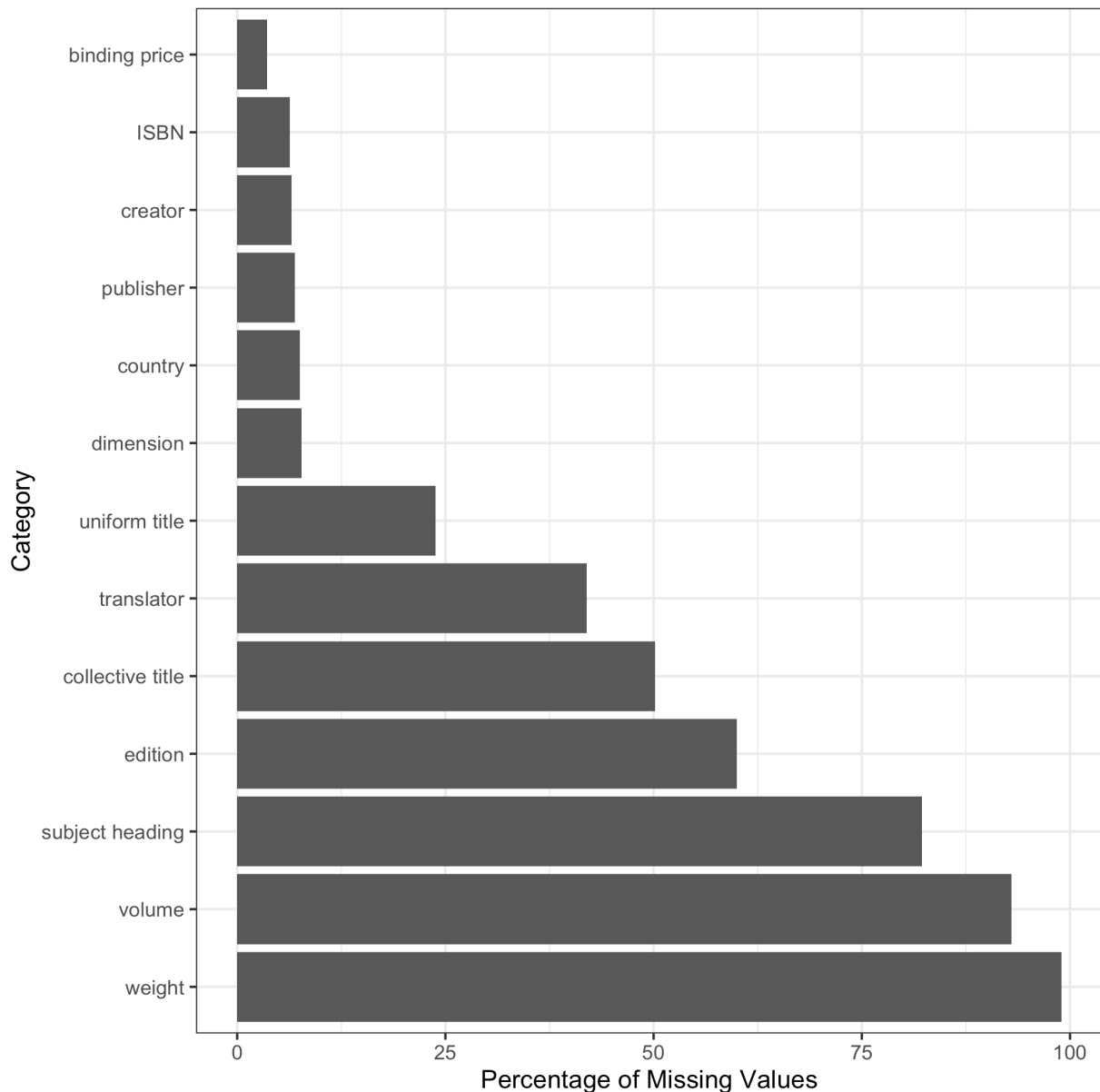
[115] DNB Metadatendienste. "Available Fields." Accessed January 28, 2021. https://www.dnb.de/EN/Professionell/Metadatendienste/Exportformate/CSV/csv_node.html.

[116] NA stands for not available and therefore for missing values.

Completeness in the DNB translation dataset
Percentages across categories

Figure 5. Title percentages of missing values for each variable in the German National Library catalogue. Data extracted: April 15, 2021

However, several categories have an increased number of missing values, such as translator (42% missing values), some of which can be extracted from other fields like title where the translator is mentioned following a notation in the target language. Volume and edition appear as the two categories with the most missing values, with 93% of missing values for volume and 60% for edition. Dimensions are missing for 7.7% of all entries; in contrast, for weight, data is missing for 99% of entries. Binding price appears to be a more complete category, with missing values for only 3.6% of all works. Additionally, collective title (i.e., translated title) with 50.2% missing values, subject headings with 82.2% missing

values, and uniform title/original German title with 23.8% missing values are also categories with an increased number of missing values. Language and year (which is the publication date) are the most comprehensive categories with no missing values. These numbers allow the evaluation of the limitations of each variable, and thus which variables can be reliably included in the analysis.

In summary, after evaluating the completeness for the variables central to my analysis—title, author, language, country, and publishing place—some additional data cleaning and aggregating needed to be done, especially in regard to the publishing place. Additionally, uniform.title required further investigation into how the missing titles could be added from other fields such as the title field. As a workaround for this challenge, I mainly focus my analysis on authors and not on original titles. As I discuss in the following section, author names are a variable that raises specific challenges among other variables.

Sources of errors across variables (accuracy)

In this section I measure error rates (accuracy) per variable and address sources of common errors in bibliographic data identified in scholarly literature, which allows me to establish whether my main variables for analysis pose major challenges for their analysis. To assess data quality and bias, accuracy is defined as the ratio of correct and incorrect values with a focus on sources of errors mentioned by previous studies on bibliographic data (Olensky 2014, 22-23). According to Olensky, areas of concern in data accuracy for bibliographic data are: inconsistent and erroneous spellings of author names, author names with accented characters, names with prefixes, double middle initials with or without punctuation, misspelling of author names with many adjacent consonants, various abbreviations and punctuations in journal titles, and lack of journal title standardization of the numeric bibliographic fields (publication year, volume number, pagination). Author name is one of the main identified sources of errors according to Olensky (2014), while standardization in each bibliographic field is also an issue.

To get an estimate of all unique values, authors' names have been aggregated to identify possible ambiguities. The process wielded 6,457 unique author names, which are coherent in orthography and not ambiguous. Only three authors' names were not identifiable, which suggests a high object-level accuracy for the author field. According to the German National Library's cataloguing practices, authors have a standardized format which ensures that they do not include variations in spelling. Accordingly, author name appears to not pose

significant challenges when modeling this variable.

For titles, however, there is only a rough standardized format and 5% of all titles appear to be not unique, meaning that they are repeatedly used.[117] Considering that the title variable also includes information on the translator and the original title, this appears to be unlikely. Upon closer examination, it becomes apparent that most of these 5% are information on edition, suggesting that instead of the titles, only the edition has been included in the catalogue.

For all other variables, I measure accuracy across a random sample of 100 catalogued items with attention to publication place, language, and country codes.[118] When looking at the accuracy for each variable, I found that sometimes two publishing places are included, only for one of which the country code matches (typically the first one). The publisher field also includes place names, which are erroneously or inconsistently used, such as Barcellona and Barcelona. Additionally, several place names include the country of publication as well, while others do not, posing additional serious issues.[119]

Similarly, language is also a category with very few errors, whereby only eight entries are marked with the ambiguous term "xxx". Language is also the most complete category, with no missing values.[120] However, some categories have specific properties. Here is an example which shows language as "ger" (German) while it is in fact in Dutch:

---

[117] There is no procedure to assess the accuracy of work titles, since they are a unique set of strings which would need to be validated with an external dataset.

[118] Several irrelevant variables for my analysis have problematic levels of accuracy, such as volume, edition, and publisher. Volume, edition, and publisher are categories that include more ambiguity compared to language, title, and author. For volume and edition, erroneous values mostly result from ambiguous categorizations, especially of editions. Edition and volume information is included in each work's target language (e.g., Bd. 1, ed. 1, 1. Aufl, etc.). Additionally, volume also includes the original titles and translated titles. Data of these categories is too inconsistent to give an estimate of accuracy.

[119] Additional cleaning steps and solutions for place name ambiguities are discussed in chapter 4.

[120] Besides the relevant variables, format is also one of the most accurate and complete categories, including information on page numbers, dimensions, and weight. However, the orthography is not consistent and for each unit, there are a number of variants (e.g., Seiten, pages, S., gr., g, Gramm).

creator	title	publisher	year	ISBN	language	country
Goethe, Johann Wolfgang von [Verfasser] ; Stassijns, Koen [Herausgeber] ; Claes, Paul [Übersetzer]	De mooiste van Johan Wolfgang von Goethe / samengesteld. door Koen Stassijns en Ivo van Strijtem. Vertaald door Paul Claes ... Ingeleid door Heidy Margrit Müller	Tielt : Lannoo; [Amsterdam] : Atlas	2011	978-90-774-4133-6	ger	XA-BE

Table 3. One record of a translation and its fields in CSV format.[121]

Here, we have a multilingual publication by a publisher located in the Netherlands and in Belgium; however, only the latter appears in the country field. Similarly, for language, the catalogue only includes German and not Dutch. This appears to be a bilingual publication, based on the publication place and country information.[122] These types of entries make up approximately 3.6% of the whole dataset for which the language is German, and henceforth will be treated as their own language category.[123] Additionally, my definition of translation—a work with differing target and original language—also naturally includes any bilingual and multilingual ("mul") publications for which German is at least one of the source languages.[124]

In summary, author name, language, and country appear to be very accurate and consistent categories, which promises that any further analysis will be rather representative of the translational field. Assessing possible error sources for the main variables of analysis was not

---

[121] https://d-nb.info/1030294305
[122] Same for: http://www.worldcat.org/oclc/847838967
Additionally, some online resources list Germany as country: http://www.worldcat.org/oclc/1027479955.
[123] According to statements from researchers at the Datenshop, the category "ger" in the language field has only been used after 2007, which is why it only affects catalogue entries from after that date.
[124] In collaboration with the Datenshop, I worked on ways to extract and add the target languages; however, in my final model it has not been included.

only necessary to develop strategies to tackle related challenges and limitations, but also to document how representative, complete, and accessible the German National Library catalogue is regarding translation data.

Limitations and challenges

Consistency, accuracy, completeness, and sample bias provide insight into the limitations and challenges of the dataset, as well as which variables appear to be reliable sources of information for my analysis. By measuring consistency, accuracy, completeness and sample bias, the following can be concluded: first and foremost, the DNB is one of the most comprehensive resources for bibliographic translation data for German fiction, but in order to extract translation data, a clear definition of what constitutes a translation needs to be formulated and an expert search designed accordingly. Secondly, author name ambiguity does not appear to be a major challenge since the library catalogue follows orthographic standards. In other data categories, such as language, erroneous data are below 0.1% of all cases. Third, incomplete data is a minor challenge that can be solved by substituting some categories such as country with information from other fields. The most problematic variable in terms of missing value is the translator field, which again can be extracted from the title field. Lastly, self-published works, multilingual works, and translations from dialects appear to be an underrepresented category of translations and need to be further examined.

As the aforementioned numbers show, in order to support claims about a library collection, it is necessary to assess the representativeness of the dataset by documenting biases, errors, and scope of the analyzed data (Erlin et al. 2021; Mocnik et al. 2017; Pechenick et al. 2015; Mäkelä et al. 2020). For bibliographic datasets of translation, the bias is often defined by the dataset selection or sampling criteria on the one hand—for example, prestigious authors, such as prizewinners (see Erlin et al. 2021, 43; Pechenick et al. 2015)— and collection-specific bias on the other. The latter, for example, has resulted in previously "unreported bias" (Lahti et al. 2020, 287) within Eighteenth-Century Collections Online (ECCO), which in turn lead to the repeated underrepresentation of specific aspects of literature of this period within scholarship using this dataset. Especially for widely available resources, such as Google Books, which are biased toward prestigious authors and published alongside tools such as the ngram viewer to visualize keyword frequencies in the corpus, the question of representativeness in relation to the dataset quality and bias is important to take into consideration. Mäkelä et al. (2020) point out regarding projects in the digital humanities,

"without facilities for acknowledging, detecting, handling and correcting for such bias, any results based on the material will be faulty" (82). This is why scholars of the digital humanities, bibliographic data analysis,[125] and literary studies recommend assessing, documenting, and most importantly accounting (Pechenick et al. 2015, 12) for the data quality, sample bias, and the overall representativeness of bibliographic datasets.

However, there are no best practices or standardized methods established yet for translation data specifically, and methods proposed by scholars working on bibliographic data are mainly focused on scientific journal data (Van Kleeck et al. 2017; Olensky 2014; Demetrescu et al. 2018),[126] which raises its own challenges regarding the applicability of the proposed data quality assessment methods to translation data. The common point of these approaches is to first identify the possible sources of bias. For bibliographic data, for instance, some sources of bias, inconsistencies, and inaccuracy have been identified as author name ambiguity due to MARC cataloguing standards, multiple editions not being linked, and collection bias, alongside cataloguing practices which create gaps in certain periods (see Mäkelä et al. 2020 and Lahti et al. 2020). These and other documented challenges can then be considered when curating the dataset, and possible solutions can be proposed.

The workflow presented in this chapter for tackling these challenges documents the extraction and curation process, as well as possible solutions to working with translation data from the German National Library Catalogue. Hence, the proposed methods range from formulating an expert search based on a definition of translations to simple frequency counts, and any findings shall be disclosed in a field guide for other researchers interested in the dataset. This is especially important since, too often, "decisions about how to handle missing data, impute missing values, remove outliers, transform variables, and perform other common data cleaning and analysis steps may be minimally documented" (Borgman 2015, 27). In line with Borgman's call for more transparent documentation, Tolonen et al. (2019) especially point out the lack of a standardized workflow and documentation of data harmonization for library catalogue datasets. Therefore, Tolonen et al. present an open, collaborative framework

---

[125] Lahti et al. (2019) define Bibliographic Data Science as follows: "Bibliographic data science derives from the already established field of data science. It associates this general paradigm specifically with quantitative analysis of bibliographic data collections and related information sources. While having a specific scope, BDS is opening up pragmatically oriented and substantial new research opportunities in this area, as we have aimed to demonstrate" (18).

[126] Van Kleeck et al. propose a comparison with another database as reference and to assess completeness (e-resources only). Olensky assesses differences in data quality between journal platforms such as WOS and Scopus, while Demetrescu only focuses on journal articles.

that would enable scholars to analyze bibliographic data across national library catalogues by developing a standardized workflow with partially automated processes to enhance data consistency and accuracy and make translations visible in the library catalogue.[127] Clearly defining what a translation is in the context of the catalogue, documenting the data extraction process, and assessing challenges and reliability of data quality are necessary steps toward building a dataset for analysis. With this thesis, I hope to contribute toward making translation data more visible, accessible, and transparent.

---

[127] Additionally, they emphasize to highlight aspects of data by visualizing document dimensions, publication years, author life spans, and gender distribution based on timelines, scatterplots, and histograms to identify outliers and possible biases in the dataset. Besides documenting data harmonization and analysis efforts, they propose natural language processing, feature selection, clustering, and classification as methods to control for duplicate entries, inaccuracies, and errors in bibliographic data. Some of these methods find application throughout chapter 2.

2. Language, title, and author concentration in the DNB

Coming back to my initial example in my introduction, in which I describe the observation that there is an imbalance of translated titles, the languages they are translated into, and the author's role in the national canon as compared to their translationalism, I found that the sum of Ilse Aichinger's translated titles and the languages she had been translated into compared to Herta Müller showed that the latter not only has more translated titles, but is also distributed across more languages than Aichinger. Based on the German National Library catalogue, which lists Aichinger's 30 titles in 16 languages, I argue that Aichinger's position in the national canon is not reflected in the number of translated works, while Herta Müller has a total of 198 translations in 39 languages and hence is well-situated in the translational field.[128] Comparing both authors' linguistic distributions to their translated titles shows that each author occupies a different position in the catalogue and, therefore, in the field of translation.

The case of Aichinger and Müller are not unique in that we can also observe this imbalance for other authors and their translated titles. The once-banned novel *Am Rande der Nacht* (*At the Edge of the Night*) by Frieda Lampe, published in 1934, has only been translated into English in 2019. Although Lampe is a well-known author in Germany and *Am Rande der Nacht* has been translated into French (1970), Dutch (1974), and Italian (2001), only due to the publisher Hesperus' initiative did it get translated into English with the aim of "bringing back to light crucial and important classics that have been sadly neglected or forgotten."[129] This is a perfect example of how a novel can be well established in the national canon and still not circulate beyond the original language until much later or only to a very limited extent. On the other hand, German author Marion Poschmann—next to another five female authors—has recently been added to the Man Booker International Prize 2019. The novel for which she has been listed, *Die Kieferninseln* (*The Pine Islands*, published in 2017 in German) was translated in the same year it was nominated. It was followed by 12 translations between 2017 and 2020, despite only having one edition (2017). Thus, as these

---

[128] See appendix for frequency table: alldnb_aichinger_lang_titlefreq.csv
alldnb_muller_lang_titlefreq.csv

[129] Flood, Alison. "Acclaimed German Novel Banned by Nazis Gets First English Translation." *The Guardian*, January 23, 2019. https://www.theguardian.com/books/2019/jan/23/acclaimed-german-novel-banned-by-nazis-gets-first-english-translation?fbclid=IwAR3sT2UXwgNSG2st8LFODFoInIPCBXO9w4yG_Wp-UQXPIMMM77c59TM4U_M.

two examples demonstrate, translations of a given novel may differ significantly from their editions and thus their status in the national canon. Comparing Poschmann and Lampe and the translational trajectory of two of their central novels, we can see that an author's position in the field of translation is conditioned not only upon the languages they are translated into, but also the role specific works play in their translationalism.

The observation that the category of the national canon does not necessarily overlap with the authors' translationalism, has only recently been addressed in literature and urges scholars of literature to re-evaluate the role translations and languages play in the circulation of the national canon beyond its linguistic boundaries. The translational turn brought about translation with a focus on global distribution as an analytical category in order to make visible the networks of circulation and transfer that stretch across languages, literatures, and cultures beyond the national, monolingual definition of a national literature. For scholars such as Walkowitz and Ivanovic and Seidl, translationalism—as it is apparent in born-translated works—also implies its multilingualism, its global circulation in different languages, and readerships.[130] In this regard, Walkowitz argues that born-translated literature is, among other factors, literature that is translated into many languages in a short time (see Walkowitz 2017, 47), which is also apparent in the examples mentioned above. Hence, we can argue that the number of languages and the rate of translation are a defining factor for the global circulation of authors and literary works as is the case for the translations of Poschmann's *Die Kieferninseln* and works by Müller.[131] These arguments thus show that the relationship of an author, their translations, and their target languages constitute a core feature of their translationalism by which they move beyond the boundaries of the national canon and position themselves in the field of translation.

To better understand how works are categorized as representative of the national canon on a global scale, let us turn back to former world literature canonization efforts by Hesse. As I mentioned in chapter 1.1., Hesse argues that there are few titles which are already acknowledged, included in the national canon, and thereby listed as representative of German literature on a global literary field. Hence, canonicity often has been defined as a category of representative works of a national literature or belonging to one source language.

In turn, the strength of Walkowitz's argument is that it proposes to re-evaluate the

---

[130] Walkowitz applies a similar definition to born-digital literature, that is, works that "appear in multiple languages and multiple versions of language at the same time" (Walkowitz 2017, 47). She further argues that "these works reflect on the relationship between language and citizenship, testing common definitions of monolingual, domestic and foreign" (ibid.).

[131] Asking whether *Die Kieferninsel* is born-translated would entail measuring the rate of translation in a longitudinal analysis as well as text inherent translationalism (as suggested by Ivanovic and Seidl).

existing categories of canonical in translational terms and formulate definitions based on a work's relationship to languages and editions. Building on her definition of what sets born-translated works apart[132] and treating that as a sign of an author's translationalism, we can take language as a proxy for an author's translationalism, while edition can serve as a unit of analysis for canonicity. Accordingly, in this study, canonicity is measured by the concentration of titles per author (the total number of editions), and translationalism is measured by linguistic distribution (the total number of target languages). Following this definition, *Die Kieferninseln* appears to belong to the translational category with having been translated into several languages before internationally reaching the status of a representative work of the national canon by its numbers of editions. By positioning authors according to their translated titles and linguistic distribution, I test the hypothesis that there is an imbalance of translated titles by author across languages which sets canonical (authors representative of the national canon) and translational authors (authors widely distributed linguistically) apart.

As I argue in this chapter, authors can be positioned in the field of translation by their title concentration and language distribution. As my observation of an imbalance in translated titles for authors such as Ilse Aichinger, Herta Müller, Marion Poschmann, and Frieda Lampe shows, an author's concentration does not always overlap with their language distribution. An author may have many editions and be highly canonical even though their translations are not distributed across many target languages, or vice versa. For instance, even though some authors did not publish many novels, they can still be globally distributed in many target languages. This would place them high on the canonical scale and low on the translational scale in the field of translation.

How does that affect how we think about the national canon in an international space? Before any work of fiction can move from its original language into other languages, it needs to be published as a translation in few target languages. Accordingly, we can deduce that translation is a pre-condition of the national canon to move into other languages and these titles will also be over-represented in translations into other languages. In other words, the more languages a work is published in, the more likely it is to be categorized as translational and the more space it takes in the field of world literature. In that perspective, *Die Kieferninsel* has entered the world literary field by means of getting international recognition through its timely translations into central languages such as English. One might even argue

---

[132] For my model, I restrict my analysis to linguistic distribution as a measure for translationalism and therefore draw on Walkowitz's definition in order to understand the nature of translationalism.

that translation into central languages accelerated the global distribution of the novel, which is documented in the library catalogue—a common argument in sociological translation studies. Languages therefore have an important role in positioning of authors in the field of translation.[133]

Therefore, this chapter addresses the observed imbalance in the library catalogue by looking at which languages and authors it is representative of and by mapping the positions of authors in the field of translation by their title concentration and their language distribution. Accordingly, in this chapter I ask: Which languages do most translated titles and authors represent in the DNB catalogue? Do we see a distinguished group of authors that set themselves apart in their translationalism by means of linguistic distribution? Which authors can be categorized as canonical and/or translational and what distinguishes them? And lastly, what is left out by canonicity that could be made visible by considering translationalism, or, vice versa, how does the application of the translational category relate to the literary canon?

While these questions guide the empirical part of this chapter, the overarching question of how the translational category challenges the classificatory structures of the national collection, and thus the collection's role in the making of world literature, accompanies the analysis of bibliographic data from the German National Library. Most importantly, what does measuring linguistic concentration and distribution tell us about the literary field and the role of translations in the circulation of the national canon through the library catalogue?

In this regard, Venkat Mani, who has extensively described the relationship between translations and canonization in and beyond libraries, identifies representativeness as one of the main deciding factors for the categorization of world literature, which thus "pushes us to consider translation, circulation, and distribution" (Levine and Mani 2013, 144). These "new" world literatures, he argues, are "often measured against the timeless value ascribed to representative works of a national or a linguistic canon assembled under the rubric 'world literature'" (Levine and Mani 2013, 144). Not only do Levine and Mani suggest here that the translational category, when compared to the national canon, reveals how the world literature rubric is formed, but they also stress the library's role in curating, preserving, and representing this category. Hence, Levine and Mani speak of a "world literary corpus" and the role that educational institutions, publishers, and also scholars play in its curation (Levine and Mani 2013, 147). Along those lines, Hesse's list of world literature, encyclopedias, and

---

[133] The role of authors connecting languages and the role of central versus peripheral languages will be addressed in detail in the following chapter.

anthologies, as well as library catalogues, documents the corpus of works that have transgressed boundaries of the national to the translational literary field.[134] Both Levine and Mani and Hesse see translation as a predecessor to canonization through curated lists and repositories constructed by official institutions such as the library. In a way, translation is therefore already implied in the formation of a translational (world literary) canon, while the role of the archive or library catalogue in this process is to catalogue, curate, and reinforce the national canon. In other words, the library catalogue is representative of combined lists and anthologies focused on the national canon, but it extends its boundaries by also listing any translations that are collected as bibliographic data.

However, as I argue in the previous chapter, in the German National Library catalogue, translations represent an invisible subgroup, since there is no standardized category applied to all records of translated works. This lack of visibility of translations caused by the cataloguing practices of national collections underlines the role of national libraries as repositories of national bibliographies, which in turn represent the national canon. Hence, even though Mani defines the library as a "crucial institution in which world literature is defined, imagined, and redefined" (Mani 2013, 240), the status of translations within national collections raises the question then of "How does one imagine and gain access to world literature through the library?" (Mani 2013, 321). Taking up Walkowitz's and Mani's point, I argue that national library collections are representative of vernacular paradigms of a literary history and that translational literature, by placing itself in between applied categories (linguistic and other), challenges the underlying premises of the very catalogue it is or is not part of. Hence, as I show throughout this thesis, library collections are representative archives of canonical and translational tendencies. Accordingly, I expect to find that translations within the catalogue are concentrated in a specific set of authors representing the translational extension of the national canon. However, despite this general tendency, I also expect to find a different set of authors that only partially overlap with canonical authors in their language distributions but set themselves apart by their translationalism.

By presenting the findings in this chapter regarding the overlap of spaces of the national canon with translational works, their authors, and boundaries, I want to open a discussion regarding the library as a curator or facilitator of world literature, with the goal to critically evaluate the literary canon based on translationalism and its linguistic and geographic networks. In tracing the focus and tendencies of translations within the catalogue,

---

[134] This raises questions regarding their role in the making of world literature which I shall return to in the discussion chapter.

this and the successive chapters are dedicated to making translations visible as part of literary history, canonization processes, and multiple language communities.

Mani's perspective highlights the important question of how canonicity has been measured across these repositories of national and translational literatures and draws attention to the difficult task of developing and testing a model for analyzing the relationships between canonical and translational works of German fiction, the question that this chapter is devoted to. Even though several scholars have developed a quantitative model to measure canonicity with bibliographic data in terms of overrepresented titles or authors, the datasets are primarily based on pre-curated lists of canonical authors.[135] Gonzalez et al. (2021) measure canonicity by overall representation across their dataset, defining canonical authors as those who appear in at least 50% of all the lists. Similarly, in their study on world literature on Wikipedia, Hube et al. (2017) measure the authors' presence across language editions of the online encyclopedia. Both models are based on the increased presence of canonical authors in a dataset extracted from an educational institution or encyclopedia which are representative of the canon of world literature.

As discussed in the previous chapter, Heilbron and scholars in the field of sociological analysis, such as Pym and Bachleitner and Wolf, suggest modeling translations based on their distribution across target languages. Pym and Chrupała (2005), for instance, use UNESCO data to present some evidence on the English case, confirming its outlier position by measuring the percentage of translations by books published in each language.[136] They also underline the central position of Western European languages as compared to other languages.[137] However, as I discuss in detail in chapter 1.2., a model accounting for both

---

[135] In the sense of Poupaud (2009), "research filter" has been applied prior to analysis based on the expertise and knowledge of the authors.

[136] Pym and Chrupała (2005) designed their own model to demonstrate the influence of the size of language on its translation rate (See Pym and Chrupała 2005, 29-30).

[137] See Figure 2.3 (Pym and Chrupała 2005, 34). Heilbron makes a similar point on the role of English as compared to other languages when describing the functioning of the world system of translation, showing that translation flows move from the core to the periphery. More than half of the books in the world are translated from English, which thus occupies a hyper-central position (the rate was close to 60% in the 1990s). Translations from French, German, and Russian each made up 10-12% of this market until 1989. These could therefore be considered as central languages according to Heilbron's analysis, but the number of translations from Russian drastically decreased in the 1990s, leaving only two central languages apart from English. Eight languages have a semi-peripheral position, with a share that varies from 1% to 3% of the international market (Spanish and Italian, for example) according to the UNESCO dataset. The other languages all have a share of less than 1% of the international market and may thus be considered peripheral (See Sapiro 2008, 158). In other words, Heilbron argues that the positions of translations in the network of literary transfer are mainly defined by their connecting function between languages. Languages can take different positions in the network and based on that, works and authors appear as translational. Heilbron especially points out that concentration is a proxy for the position of the language in the network. Languages with a concentration of titles therefore occupy a central position, while languages with fewer titles will be situated semi- or peripherally. Heilbron conceptualizes

canonicity and translationalism has not yet been developed, even though Mani, Heilbron, Pym and Chrupała, as well as Bachleitner and Wolf, propose to formulate a new model for a sociological analysis of translations and their distribution in the global literary field by focussing on the relationship between translations and their languages.[138]

In line with the theoretical groundings presented by Heilbron, Pym and Chrupała, Bachleitner, and Walkowitz, I propose a model to measure the observed (im)balance between translated titles and languages by looking at the title distribution across languages and then using edition to position authors in the translational field by correlating it to language count.[139]

Building on the definition of canonicity as linguistic concentration of titles and authors, or representation as in Gonzalez and Hube et al., I propose measuring canonicity in the library catalogue by concentration—the sum of titles and authors per language—which is indicative of how canonical my dataset is and how (un)equally titles are distributed among languages. I also measure linguistic concentration for languages by titles and authors to analyze which languages are central in the library catalogue. By looking at the concentration of titles, I visualize the roles that few central languages and representative authors play in the field of translation and the dimensions of the canon in translation within a national collection. For translationalism, I measure the sums of languages a given author has been published in and directly compare them to the title count.[140] In other words, canonicity is a measure of editions and translationalism a measure of languages. I calculate the correlation between language and title count per author to locate authors that lie in and beyond the translational or canonical fields. This allows me to determine which and to what extent a work or author (as based on this model, categorized as translational) may or may not overlap with canonical sets of works, providing an entry point into discussing the overall tendencies of the library catalogue toward representing specific groups of authors or languages. My model builds on

this idea under his model of language networks with central and peripheral languages, arguing that the translatability of work depends on its position in the language network. This is discussed in more detail in chapter 3.

[138] As has already been mentioned, Mani calls for a method of world systems theory and patterns of global circulation. His point directly corresponds to Heilbron's suggested model for analyzing translations as an international system (Heilbron 1999, 431).

[139] The translational field denotes the space of linguistic distribution an author can be positioned in, while the field of translation is used to define the space translations occupy, as well as their network of languages and authors in the library catalogue.

[140] As stated in chapter 2.0, I only count outgoing translations of fiction originally published in German. When I refer to languages of a given publication or author, I imply target languages or languages of the published translation. Additionally, I take the total number of editions, which means that they may include different editions of the same work, e.g., re-translations. For my analysis, I treat editions as equally contributing to an author's concentration and therefore did not differentiate between first, and successive editions. The more editions an author has, the higher their concentration.

both canonicity and translationalism in order to map out the space where they overlap and to open a discussion on how these two categories challenge the structures that form and sustain the national literary canon within the library catalogue.

The following three sections investigate the space between the field of the national canon in translation in one of its keepers and sustainers, the national library, and the translational literary field. Section one of this chapter is dedicated to measuring language concentration in titles with the expectation that titles are unequally distributed across languages—accounting for the Eurocentric tendencies in the catalogue—while section two looks at the concentration of authors, exploring the hypothesis that the translational and canonical fields do not perfectly overlap. I expect to find a set of authors that are concentrated in terms of their count of editions in few central languages and a set of translational authors that are spread out through many languages which thereby stand out in their translationalism. The third section zooms in on the authors that surpass the threshold of title or language concentration and, by close reading the trajectories of these author's works, contextualizes their positions within the canonical and translational fields. This last part not only serves as a validation of the proposed model in this chapter, but also as a way to investigate its implications for the representation of translations in the library catalogue and the limitations of the applied measures.

Language concentration

How concentrated is the catalogue in specific titles across languages and language families? How equally are titles distributed among languages? If we assume that the number of languages as well as their sizes are decisive of a work's circulation, the first measure requires testing their concentration. The concentration of translations per language then can be analyzed as a factor for the categorization of canonical and translational literature. Following scholars such as Damrosch and Etiemble, who have drawn attention to the Eurocentric focus in the field of world literature, as well as Heilbron's argument regarding the role of central languages in the circulation of translation, we can expect a highly skewed distribution toward few, central languages. Hence, titles are assumed to be overrepresented in few main languages, while smaller peripheral languages will have less than a certain threshold.
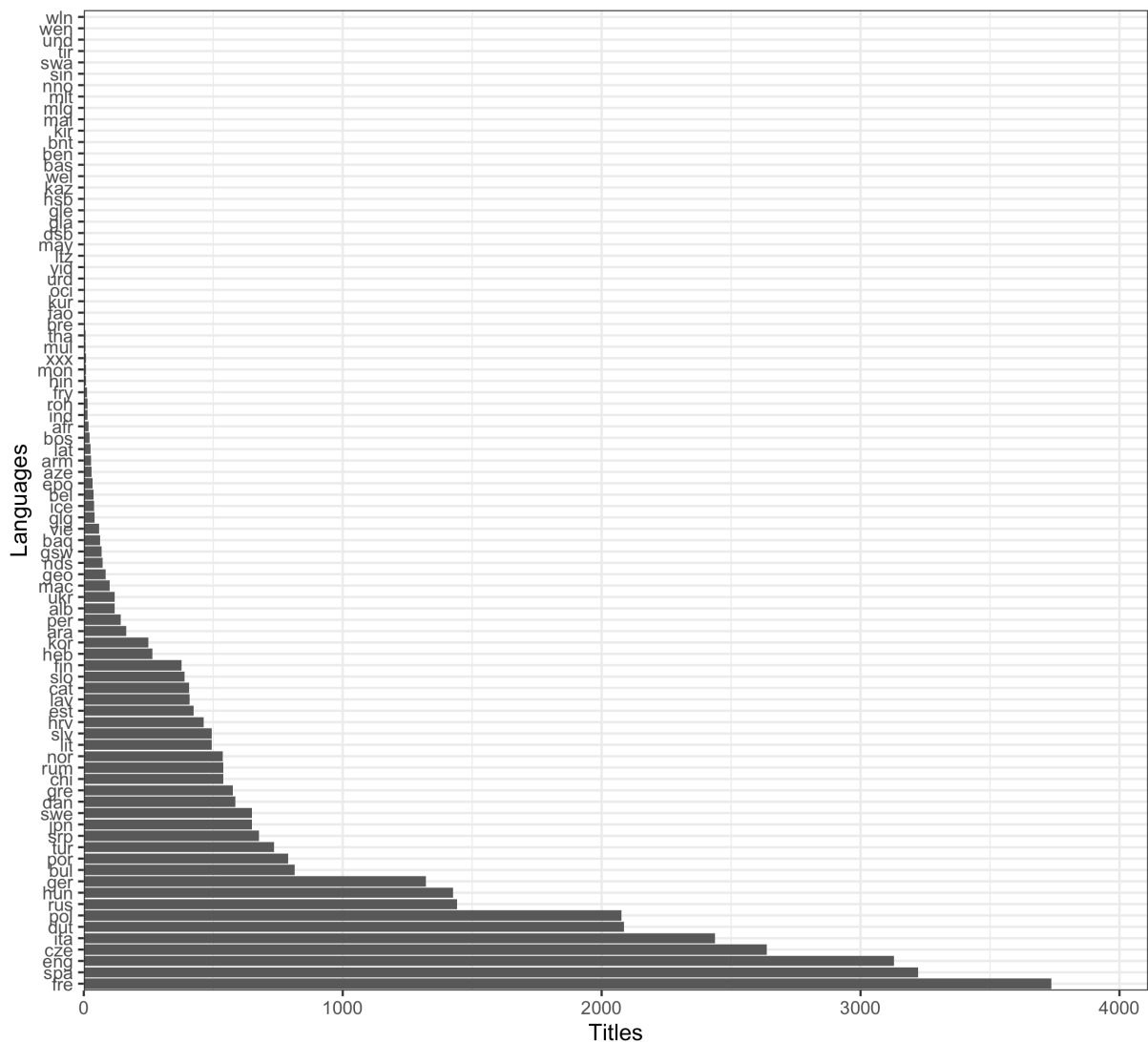
Figure 2.1. Distribution of title sums (raw count of titles) per language.[141]

Looking at the frequency distribution of titles per language in figure 2.1, we can see that half of the languages have the majority of titles, which again supports the claim of a centralized or, rather, concentrated field of translation in terms of language.[142] Out of 86 languages, almost half of them have less than 1000 titles, with a mean at 403 (titles per language) and a standard deviation of 758. Additionally, the frequency graph shows a skewed

---

[141] Some publications have German (ger) as a language—which are bilingual publications for which only one language has been extracted from the library's Datenshop, same for multilingual publications (mul).

[142] To further stress the initial argument that translations in the library catalogue are concentrated in few target languages and language families, I also calculated the Gini-coefficient (Schulenberg 2018) for title counts per language for all languages in my dataset. Since the Gini-coefficient is a measure of inequality, it can be used to get an estimate of the degree of concentration of titles per language. The result of a Gini of 0.76 (whereby 1 is the highest possible inequality) represents a 76% percent concentration in a language's title distribution, which accounts for a high inequality of title shares. The Lorenz curve (see appendix: figure_ch1.2_dnball_gini_lorenz_title_langfam) shows that 25% of all languages have the majority of titles.

distribution. This supports the point that titles appear to be concentrated in few central languages, but does it also indicate that there are many titles in few languages or the other way around? Calculating the pnorm[143] gives us the percentages of variables in a certain range based on the area above or below the z-score. For languages, this returns the following: 54% of languages are below 500 titles (as the histogram shows as well), the area above the z score of 500 is approximately 45% (suggesting that the distribution is skewed to the right), and, most importantly, 23% of all languages have between 1000 and 4000 titles, which shows that very few languages have the most out of all titles. In other words, the top 20 languages account for 83% of all titles, which, again, points toward a highly concentrated language network of translations. The high standard deviation (758.9108) confirms that the language distribution is not spread out evenly in the dataset, with a mean of title count at 403.6118. Contrary to an equal distribution of titles across languages, these findings show that there are many translations into few languages such as English, French, and Czech, which supports the argument for a high linguistic concentration of titles in the catalogue. Additionally, the thin, long tail of the distribution indicates that translations of German fiction do not display a high level of linguistic diversity, but a focus on few central languages.

Calculating the overall percentage of titles in the catalogue of the top 20 languages also shows that only few languages pass the 5% threshold.[144] Most translations of the top languages are accumulated in Roman and Slavic languages, while over one third of all languages are underrepresented. We can see a clear separation or clustering of four groups: languages below 2.5%; a small group around 3.5% composed of German, Russian, and Hungarian; another smaller group between 5.5% and 7.5% composed of Dutch, Polish, Italian, and Czech; and then the last group, with first French (more than 10%) followed by English and Spanish at a similar level. According to Heilbron's classification of language positions, the findings of this section map out the hyper-central position of Indo-European languages in contrast to the hyper-peripheral position of all other language families. Further, figure 2.1 illustrates, that among the top 20 languages, there seems to be a clear grouping of central (French, Spanish, English), mid-central (Czech, Italian, Dutch, and Polish), and mostly North- and Eastern European semi-peripheral languages.[145]

While the distribution of titles per language on figure 2.1 reveals which languages are

---

[143] Based on the distribution function by Cody 1993, the Pareto Chart shows which languages are in the 20% that contain 80% of all titles. See appendix: alldnb_language_title_pareto_ch2.1_v2.csv.pdf

[144] See appendix: figure_ch1.2_dnball_language_percentages_top20_scatterplot_v2.png

[145] In chapter 3, I separate publishing places into European and non-European, the results of which further supports the Eurocentric focus of the catalogue, but also the transcontinental role of languages like English, Spanish, Portuguese, and bilingual texts.

most central and underline the focus on specific groups of languages, language families give a general picture of the Eurocentric focus of the catalogue. As figure 2.1 shows, more than half of all languages are under 500 titles, with mostly non-European languages such as Hindu alongside many other languages, which supports the claim that languages are centralized in Europe (other than Spanish, which also includes Latin America; English, which also includes North America, Australia, and New Zealand; and French, which also includes parts of North Africa and North America).
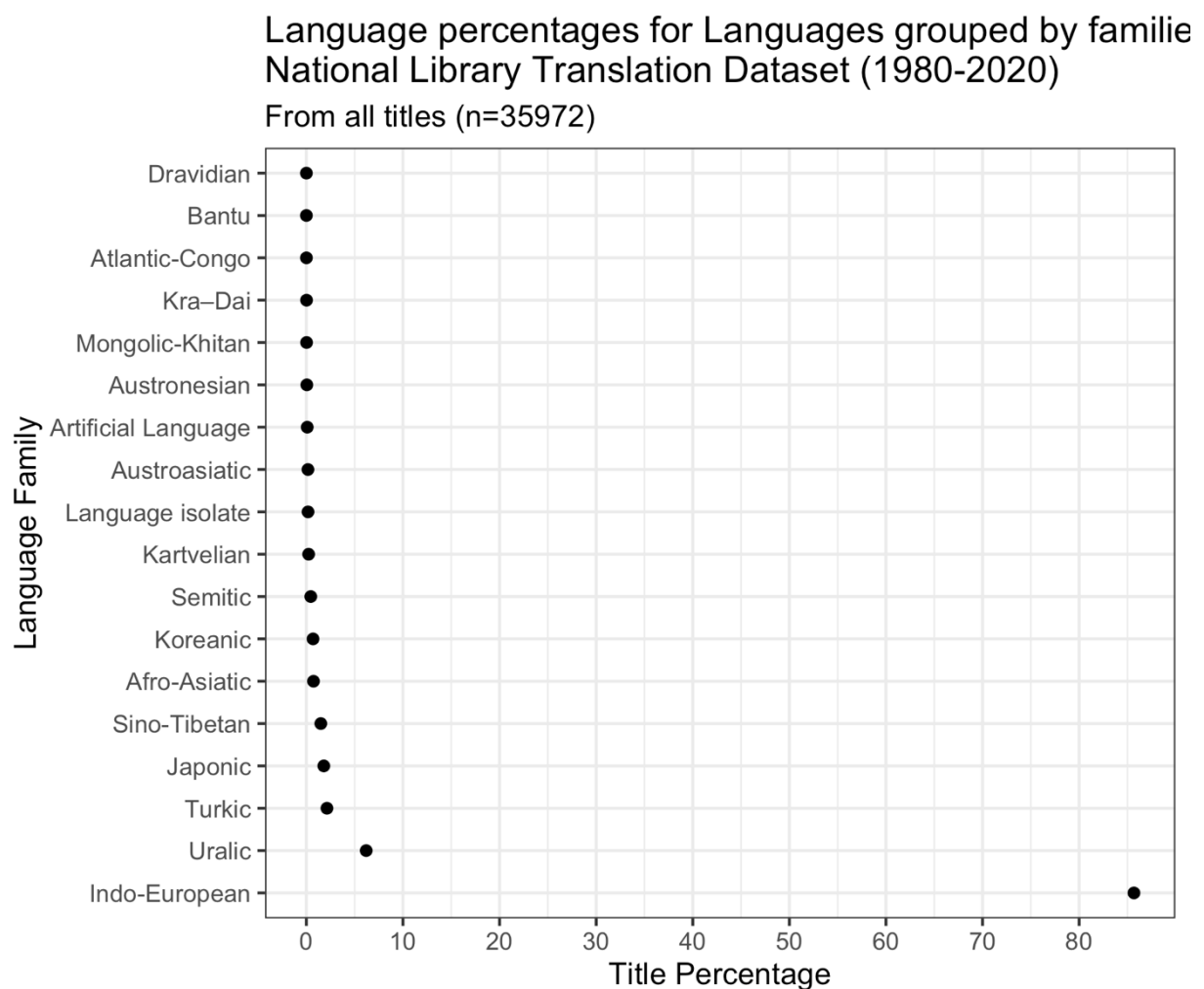
Figure 2.2. Percentages of titles per language family.[146]

Figure 2.2 visualizes the title percentage for each language family and gives a picture of to what extent titles are concentrated in mostly Indo-European languages (85% of all titles),[147] revealing the large gap between other languages which are all under 7% and are

---

[146] See appendix for the table with languages in each group: alldnb_lang_stats_ch2-1_v2_langfamilies-csv.csv
[147] I will discuss this finding in relation to publishing places in detail in chapter 4.

comparatively underrepresented in translation.[148] Along those lines, 79% of all titles have been published in European countries, supporting the argument of a Eurocentric focus. From all translations into Spanish (3222 titles),[149] 89% have been published in Europe, while only half of all translated titles into English have been published in Europe; similarly Portuguese also has a 55/45 ratio of European to non-European publishing places.[150] The catalogue clearly favours Indo-European languages and European publishing places.

From these findings, we can conclude that the library catalogue consists of a large amount of concentrated titles for a few central languages, which appear to play a major role in the distribution and circulation of German fiction in translation according to the DNB. What these findings suggest is that from concentration measures, we can infer that the library catalogue is highly concentrated on a few languages and favour Indo-European languages, which, when analyzing the data, can in turn lead to a Eurocentric view of the translation field. This brings us back to the previous discussions on the Eurocentric focus of national library collections and the question of how this is reflected in the dataset, and which languages are more concentrated than others. As I discussed in chapter 1.1, the Eurocentric focus has raised concerns in debates regarding world literature and its canon. After taking the catalogue's inherent Eurocentrism into account, German literature might not be as "worldly" as previously assumed. On the contrary: it appears to be quite concentrated, specialized, and centralized in terms of languages. If, according to the catalogue, only a small number of mainly European languages have published the majority of translations, that would mean that the field of translation is representative of few dominant literary cultures. When describing the library catalogue as a representative archive of canonical and translational tendencies, these findings illustrate how the catalogue focuses on and documents a specific set of languages in large numbers, which in turn play important parts in the formation of the field of translation.

However, the tail of the distribution is long, and the 17% of titles that appear across the smaller language groups suggests that there are translations which are spread out

---

[148] Looking at a comparison of Gini-scores for language families reveals that among the most concentrated language groups are Indo-European, Turkic, and Afro-Asiatic languages, which are also the three largest groups in the dataset. 0.7384724 Indo-European, 0.6591479 Afro-Asiatic, Atlantic-Congo and Austroasiatic 0, 0.7265319 Turkic, and 0.3138839 Uralic. See appendix: figure_ch1.2_dnball_language_percentages_top20_scatterplot_v2.png

[149] The "country" field in the catalogue data contains country codes, for which all European countries contain the prefix "XA-". I subset my dataset by country codes containing this sequence of characters in order to calculate the ratios.

[150] In comparison, only 4% of French translations have been published in non-European countries. Chapter 4 is dedicated to an in-depth analysis of publishing places.

linguistically. Can we observe the same tendency for authors in that they are concentrated in few languages, and what is the share of authors that are spread out among the smaller languages?

Title and author concentration in languages

In the previous sections, I measure title concentration by language which indicates the imbalance in the field of translation and the catalogue's focus on only a few languages, while in this section I focus on the author in the Foucauldian sense of its function within a language field[151] that "serves as a means of classification" (Foucault 1998, 210) of editions; in Foucault's words, "a name can group together a number of texts and thus differentiate them from others" (ibid.). For translations, this entails that an author can function to group all editions in various languages together. Thus, the author has a multilingual dimension and can be put in direct relation to other authors by translated editions in the same languages. In the context of a library catalogue or archive, an author name will serve as the main category by which to associate editions. As discussed in chapter 2.0, author names are unique, standardized units which apply globally to the complete catalogue, and therefore, under Foucault's definition, can be used as a means of positioning them in the translational field according to their concentration and linguistic distribution. Following my initial observation for Aichinger and Müller and the findings of the previous section, we can expect titles to be highly concentrated in few authors, which are published in central languages.

To test the hypothesis that most authors are concentrated in few target languages—meaning that they will have many editions in few central languages—and only few authors spread out across many languages, I first collected the frequency counts of languages per author and tested for normal distribution. A skewed distribution would support the argument that most authors have a high language concentration (translated in few central languages) while few authors are translated in a broad number of languages and hence fulfilling the premise of translationalism by linguistic distribution.

Measuring the language count per author gives an idea of how many languages the authors have been translated into. From the frequency distribution of languages per author and measuring the area below the z-score of five languages, we can infer that approximately

---

[151] "[A]n author's name is not simply an element of speech (as a subject, a complement, or an element that could be replaced by a pronoun or other parts of speech). Its presence is functional in that it serves as a means of classification. A name can group together a number of texts and thus differentiate them from others" (Foucault 1998, 210).

90% of authors are published in under 10 languages, while 72% of all authors have between 0 and 20 languages, which means the majority of authors are translated into few languages.[152] This again supports the argument of a highly concentrated, imbalanced field of translation in the German National Library catalogue, indicating that many authors are translated into few languages, and few authors publish in many languages.[153] Consequently, we can deduce that the translational space, where authors are distributed across many languages, is very small if we set the threshold as an author with more than 10 languages. This would indicate that only a very small and exclusive group of authors fulfils the premise of translationalism by being translated into many languages.
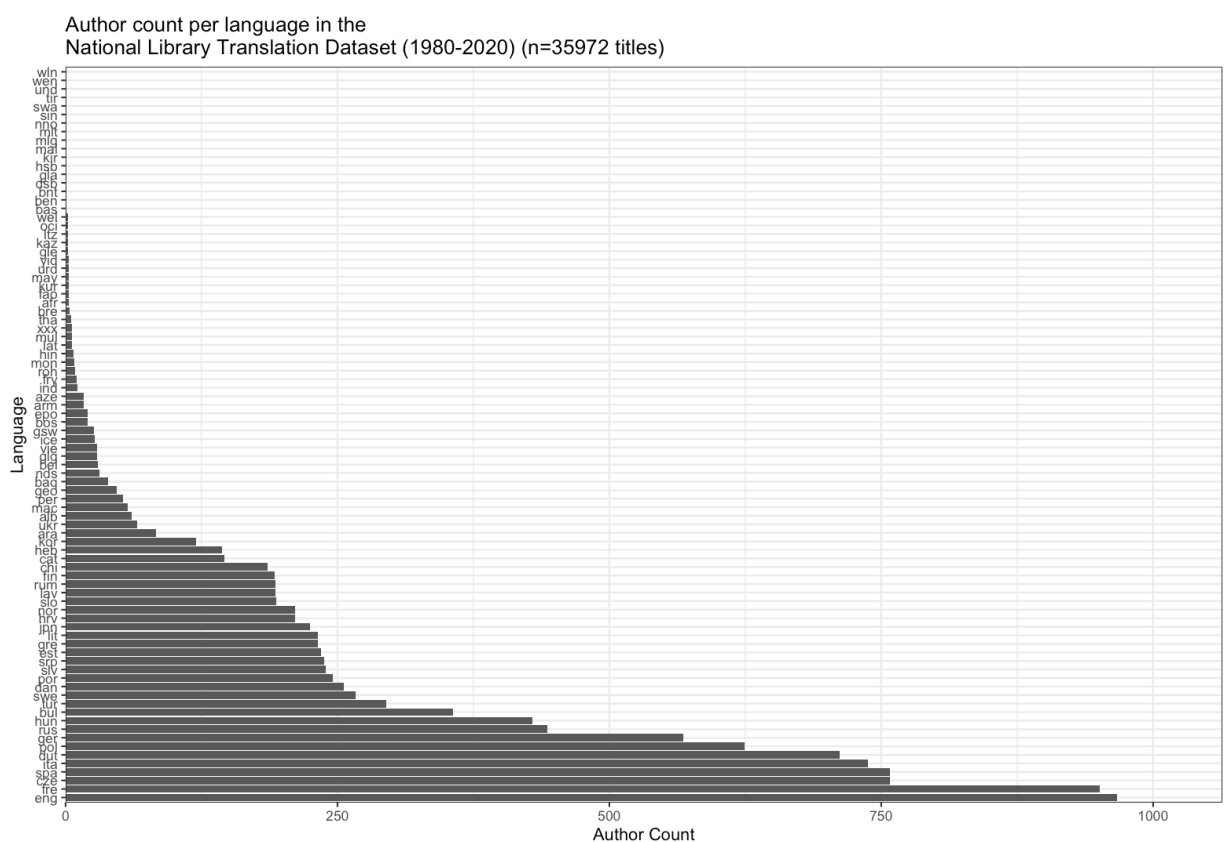

Figure 2.3. Distribution of author sums (raw count of authors) per language.

While authors appear to be highly concentrated in few, central languages, by calculating the number of authors per language, we can get a better sense of which languages authors are concentrated in. The distribution of figure 2.3 clearly resembles the title count

---

[152] Additionally, similar to the title distribution, normality cannot be assumed based on the increased number of languages close to the mean (3.0) with a high standard deviation of 5.2, again suggesting that authors publish on average around three translations, while the frequency counts are spread across on a vast number of values. See appendix: figure_ch1.2_dnball_allauthors_freq_meanhist_v2.png
[153] According to the Gini-coefficient for languages per author (0.56), authors appear to be slightly more equally distributed across languages than titles. See appendix: figure_ch2.1_dnball_language_author_gini_v2.png

distribution, where the top 10 languages have a very high count of authors and the majority of languages have less than 250 representing the long tail. Same as for title count, some languages occupy a more central position based on their concentration of authors. Only a few languages have an increased number of authors; the further down the tail, the fewer authors are translated. Again, the unequal distribution and concentration of authors in few languages supports the argument that only few authors are linguistically widely distributed and therefore positioned in the translational field.
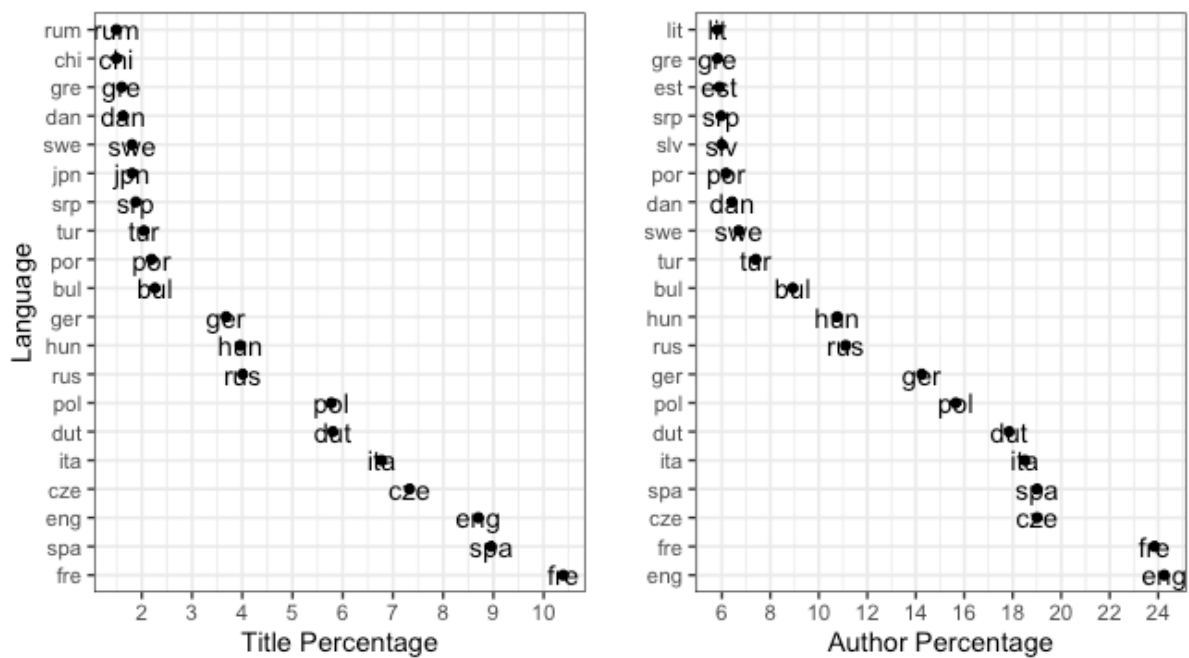
Figure 2.4. Percentages of translated titles and authors per language for the top 20 languages in the German National Library catalogue.

Additionally, except for English, the same few central languages appear to be concentrated in titles and authors alike. However, when directly comparing the figures of title and author frequency distributions (figure 2.4), some differences become visible. First, the language with the highest author count is English and not French, followed by a clear gap between the next group, with under 20% of all authors in the dataset consisting of the same languages as for title counts (Czech, Spanish, Italian, Dutch, Polish, and bilingual). Again, Pym and Chrupała's analysis, which points out the specific position of English and the fact that it is commonly referred to as a bottleneck language for translations, resonates with this finding. English therefore acts as a comparatively diverse target language in terms of the number of authors.

While the position and roles of languages will be analyzed in detail in chapter 3, for this

section the most important finding is that the translational field (as defined by linguistic distribution) appears to be smaller than the canonical field (as defined by title concentration within target languages for both authors and titles). In other words, if authors are not spread out across languages and not equally distributed, it cannot be expected that there is a translational tendency, and therefore an increased number of translational authors, in the catalogue.

The positions of authors in the canonical and translational fields

While title and author count per language give a global picture of concentration with some indication of a large canonical and smaller translational field in the German National Library catalogue, it does not address where the authors are positioned in the translational field according to their title concentration and linguistic distribution. For the findings presented in the rest of this section, I focus on the initial questions of this chapter: Do we see a distinguished group of authors that set themselves apart in their translationalism by means of linguistic distribution in the German National Library catalogue? In other words, which authors are situated in and beyond the overlap of canonical and translational?

In order to further zoom in on the canonical and translational fields in the language space and to investigate this overlap, I measure how an author's title count correlates with their language count, asking if authors who have a high title count also publish in many languages, or if they concentrate in few specific languages. Calculating the correlation between language and title count per author further pinpoints where overrepresented (possibly canonical) and spread out (possibly translational) authors are positioned. We can expect to identify distinct groups of authors with a high title count that also have a high language count (highly correlated), or authors with a high title and low language count and authors that have a high language and low title count (lesser correlated). A high language, but low title count positions an author farther in the translational field while authors concentrated in titles which are not linguistically distributed across many languages are positioned more in the canonical field. We can also expect some authors with a balanced title and language count, which are equally distributed linguistically, and therefore positioned at the overlap of the translational and canonical fields.

To measure the language-title correlations for all authors in the dataset, I created a data frame with author names as rows, and columns with the title frequency counts per author and the language frequency count per author. I then measured the correlation between these

two variables, which are highly correlated and significant when measuring correlation with Kendall with a tau of 0.8644 and a p-value of < 2.2e-16. Plotting the correlation for all authors in the dataset clearly shows the few authors who are overrepresented, but also illustrates that for most authors in the lower quartile, title and language count are strongly correlated. The majority of authors (98.9%) are located under the 125 title and 30 language thresholds,[154] suggesting that only few authors are concentrated in titles and distributed across many languages.
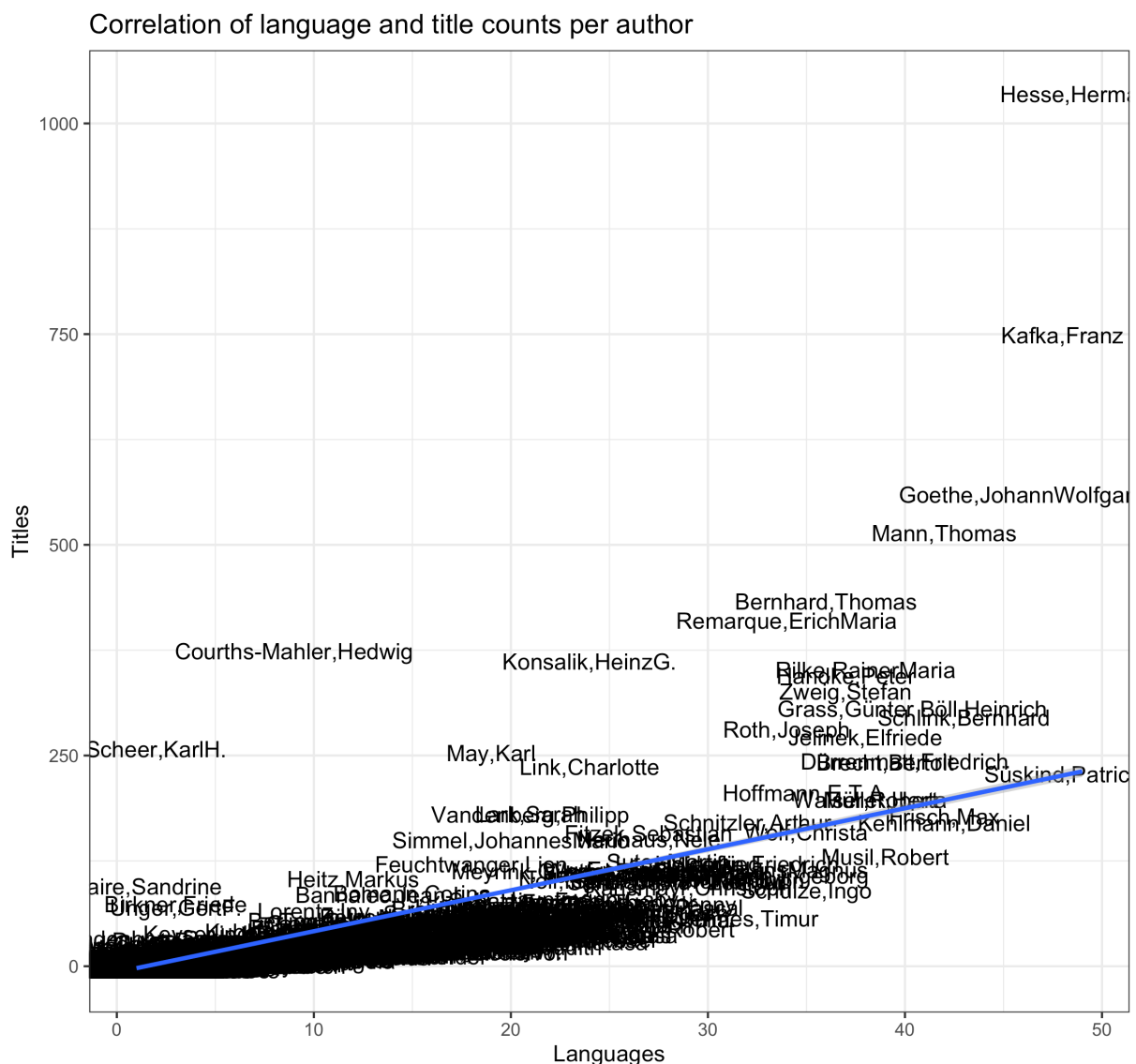
Figure 2.5. Correlation of translation title and language counts per author in the German National Library catalogue

As mentioned in the previous section, the correlation plot on figure 2.5 shows that

---

[154] The finding that only few authors have translations in many languages corresponds to Wischenbart et al.'s finding for bestselling authors that "only a remarkably small number of authors succeed in finding broader readerships across multiple languages" (Wischenbart et al. 2018, 10).

only few authors and titles reach a certain threshold of distribution among many languages (>10 languages). There are only 19 authors beyond the 250-title threshold and 33 authors passing the 39-language count. Hence, we see a large group of authors in the German National Library catalogue that are positioned on the threshold between canonical and translational.

Looking at the author names beyond that threshold reveals highly canonical names, such as Kafka, Hesse, Goethe, Bernhard, and Rilke. Especially in the field with high concentration (>250 titles), we can clearly see that the most concentrated are authors deemed representative of German literature, which confirms that high concentration is an indicator of the canonical space in the field of translation. This is not unexpected, since canonization and distribution are centralized endeavours through cataloguing practices in national collections, which also make certain authors more likely to be distributed among many languages once they are concentrated or canonized—as Hesse's lists so perfectly illustrates.[155] In other words, if an author has many translated titles, it is more likely for them to be distributed among many languages which makes them part of the world literary field while representing German literature on a global scale. However, as figure 2.5 shows, there are some authors, such as Hedwig Courths-Mahler, that appear to have many titles but translations in few languages and are thus positioned in the canonical more so than in the translational field. Max Frisch and Daniel Kehlmann are situated on the other end of the spectrum with many languages and fewer titles. This imbalance sets Frisch and Kehlmann apart in their translationalism[156] from other authors in the same language range that have more titles. Hence, which authors are comparatively more or less concentrated in their titles, and which are more or less translational based on their language distribution?

In order to compare authors in terms of title count and language count, and then visualize the overlap, I rank and subset my data by the top 20 authors for title count and then for language count. By comparing the two sums, I identify the authors who either have many titles, but fewer language or vice versa.

---

[155] It remains an open question if Hesse had the slightest clue that he would be one of the most translated authors (at least according to the catalogue) as he put together his list of representatives of German literature.
[156] As I have pointed out at the beginning of this chapter, scholars like Walkowitz also set apart works of fiction which originate in many languages and enter the literary field as global, born-translated works.

Authors ranked by title count	Languages	Titles	Authors ranked by language count	Languages	Titles
Hesse,Hermann	49	1035	Hesse,Hermann	49	1035
Kafka,Franz	48	749	Kafka,Franz	48	749
Goethe,JohannWolfgangvon	47	560	Süskind,Patrick	48	228
Mann,Thomas	42	514	Goethe,JohannWolfgangvon	47	560
Bernhard,Thomas	36	433	Böll,Heinrich	44	306
Remarque,ErichMaria	34	410	Schlink,Bernhard	43	295
Courths-Mahler,Hedwig	9	374	Frisch,Max	42	178
Konsalik,HeinzG.	24	362	Kehlmann,Daniel	42	170
Rilke,RainerMaria	38	352	Mann,Thomas	42	514
Handke,Peter	37	345	Dürrenmatt,Friedrich	40	243
Zweig,Stefan	37	326	Brecht,Bertolt	39	242
Böll,Heinrich	44	306	Müller,Herta	39	198
Grass,Günter	37	306	Musil,Robert	39	130
Schlink,Bernhard	43	295	Jelinek,Elfriede	38	272
Roth,Joseph	34	281	Rilke,RainerMaria	38	352
Jelinek,Elfriede	38	272	Walser,Robert	38	198
Scheer,KarlH.	2	258	Grass,Günter	37	306
May,Karl	19	253	Handke,Peter	37	345
Dürrenmatt,Friedrich	40	243	Zweig,Stefan	37	326

Table 2.1. Top 20 ranked unique author names with title and language counts (1990-2020).

From comparing language and title count of the top 20 ranked authors (table 2.1), we can infer that the majority of authors with the most languages are also the ones with the most titles—with major exceptions of Daniel Kehlmann, Max Frisch, Robert Musil, Herta Müller, Robert Walser, and Patrick Süskind who are not in top 20 titles, but top 20 languages. Additionally, Hedwig Courths-Mahler, Erich Maria Remarque, Heinz K. Konsalik, Joseph Roth, and Karl H. Scheer do not appear in the top 20 languages, but in the top titles.[157] Out of the top 20 authors with the most translated titles, 70% are also in the top 20 authors with most languages, meaning that 30% of the top authors are comparatively more canonical or translational. In summary, this suggests that there is a difference between the title and

---

[157] Interestingly, all these authors count among the most commercially distributed among all other listed authors.

language distribution for authors. In other words, some authors with many titles will be widely represented across languages, while some authors have few titles, but are translated across many languages.
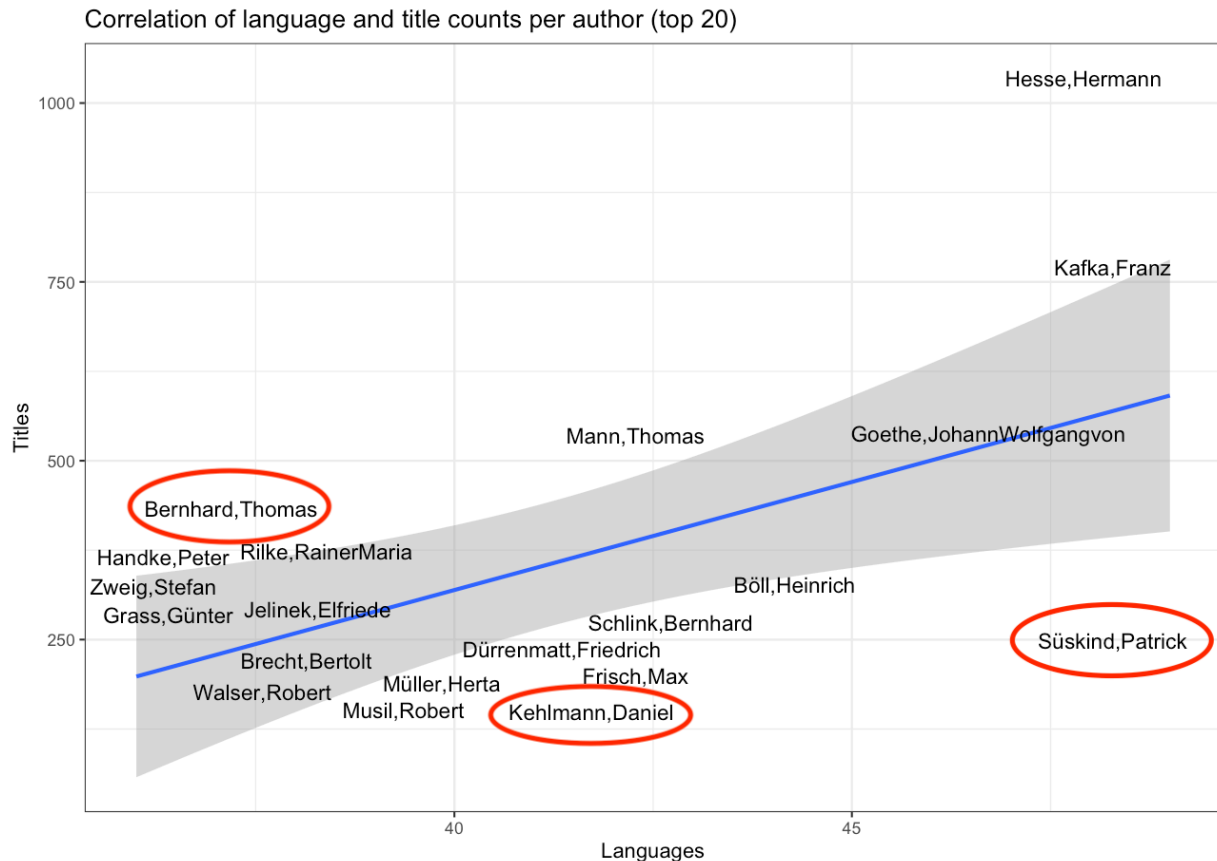
Figure 2.6. Correlation of translation title and language counts for the top 20 authors in the German National Library catalogue

Figure 2.6, which shows the correlation of the top 20 authors placed according to their title sum (y-axis) and language sum (x-axis), illustrates the difference in title and language count apparent on table 2.1, and it further allows us to pinpoint outliers as well as the authors' similarities in their positions. On figure 2.6, we can see clearly which authors are positioned beyond the highly canonical threshold (>500 titles), in a balanced field of title and language count (close to the line of best fit), and the authors in the translational field (<250 titles and >40 languages). The following paragraphs discuss in detail which specific titles contribute to the authors' positions in order to contextualize these findings and validate the proposed model.

On the correlation plot, we can see the exceptional positions of Hesse, Kafka, and Goethe, which illustrate the overall tendency in the library catalogue to overrepresent highly canonical authors. This is especially apparent in the increased title counts for Hesse, Kafka,

and Bernhard, who are, together with Goethe (among others), the most representative figures of the national German literature canon. In the top right corner of the graph is Hesse, who exceeds all other authors in terms of his title count (1055 translated titles) while also having translations in the most languages (49). His highly canonical novel *Siddhartha* has 127 translated editions in 47 languages, 14 of which are in Spanish, nine in English, four in Hungarian and Bulgarian, with single editions in Armenian, Thai, Malayalam, and Kurdish among others. Hesse hence appears to be not only overrepresented, but also translational due to his increased language count. On the contrary, in the left margin of the graph, we can find Bernhard, who has the least languages in the top 20 authors (36) even though he exceeds the 300-title threshold (438 titles) and therefore sets himself apart from the rest of the authors.[158] For Bernhard, we don't see the same tendency as for Hesse: no single title comes close to the number of languages of all other translated titles. In that regard, Bernhard does not appear to be translational, but more so canonical in that he has more titles than the majority of the top 20 authors.

In the lower left quartile, where title and language count are balanced, we can also observe the tendency toward male authors. The top 20 for both language and title count only includes two female authors: Herta Müller and Elfriede Jelinek, both of which share similar positions with Berthold Brecht, Robert Walser, Friedrich Dürrenmatt, and Günter Grass—all of which have a balanced title and language count ratio. Besides their gender, what sets Müller and Jelinek apart from these writers is that they are contemporaries as well as Nobel Prize winners with major international successes. With close to 200 titles and 39 languages, Müller is positioned slightly more in the translational field (x-axis) than the canonical (y-axis) compared to Jelinek, who has 275 titles in 38 languages. Müller's most prominent novel *Atemschaukel* (44 titles) has been translated into 33 languages according to the catalogue, and Jelinek's *Die Klavierspielerin* (59 titles)—which was one of her first novels to get translated into English—has a language count of 36. Not only do these cases highlight the overall male-dominated, canonical tendency, but they also highlight the space where title concentration and linguistic distribution are either balanced or unbalanced to the extent where single titles can have an influence on the position of the author in the field of translation.

If we move further up the x-axis into the translational field, at the same title range as Herta Müller and the language range of Goethe, we find Süskind. This is the perfect example

---

[158] Bernhard himself did not seem to have been a huge fan of translations, as this verse in his 1978 play *Der Weltverbesserer* proves: "Translations are always disgusting, but they brought me a lot of money" (Honegger 2002).

showing his translational position as compared to Goethe who has a high title count—highly canonical—but a similar language count to Süskind. Süskind hence fulfils the definition of translational as linguistically distributed even though he is not concentrated in titles. This is by no means surprising, considering Süskind's success with his novel *Das Parfum,* which is one of the most translated titles in the German National Library catalogue with translations into 42 languages and a title count of 71. When comparing these counts with all other translated titles by Süskind, it turns out that *Das Parfum* has almost half (44%) of all his translated titles and more unique languages than all his other titles (39 unique languages), including Icelandic, Hindu, and Galician. We can see from this example that an author's canonicity and translationalism partially depends on the distribution of one central title. Süskind's case in comparison with Goethe's position in the field also illustrates that comparing concentrated (with many titles) with translational (distributed across many languages) positions distinguishes specific authors or groups them together.

In a similar range on the y-axis (<250 titles) with more than 40 languages, we also find Frisch and Kehlmann, who match almost perfectly, indicating that both authors are equally represented below the threshold with comparatively high language counts and low title counts. Similar to Süskind, Frisch and Kehlmann are also suitable candidates for the translational category. Again, this is not surprising considering Kehlmann's international success of *Die Vermessung der Welt*[159] which appeared on bestselling lists as a successor to Süskind's *Das Parfum*. *Die Vermessung der Welt* has 44% of all titles translated by Kehlmann in the catalogue but has more languages than all other titles with 39 languages (all other titles have 30), including Belarusian, Bosnian, Catalan, and Faroese. Similar to *Das Parfum*, one title by Kehlmann has reached more languages in translation than all his other titles, setting him apart from other authors in the translational field. Translationalism hence appears to be strongly related to few titles that reach an increased number of languages in addition to the overall title concentration and linguistic distribution.

However, for Frisch this is not the case; he has slightly more translated titles (182) than Kehlmann and 50 unique titles, while Kehlmann only has 31 unique titles, implying that Kehlmann's translationalism is dependent on few successful titles which have been translated into many languages. Frisch's title with the highest edition count (44 or 25% of all titles), *Homo Faber,* has translated editions in 28 of the 39 languages of all his titles, which shows

---

[159] Rebecca Braun discusses Kehlmann and his success novel the *Measurement of the World* as an example for a network celebrity, or as Braun puts it: "the way the author begins to function as a 'node': to borrow again from Latour, his authorship becomes a larger node on the map of actor—network relations, as ever more cultural and economic initiatives that mark out the contemporary moment are routed through him" (Braun 2016, 461).

that Frisch's translationalism is tied to the distribution of many titles across languages and not to one title that has been translated into many languages. Hence, even though Kehlmann and Frisch are closely positioned in the translational field with a lower title count and high language count, they differ in the titles that drive their translationalism.

Plotting the language and title count of the aforementioned titles allows for visualizing the proximity between each title's position and further pinpoints which titles are not balanced in concentration and linguistic distribution.[160]
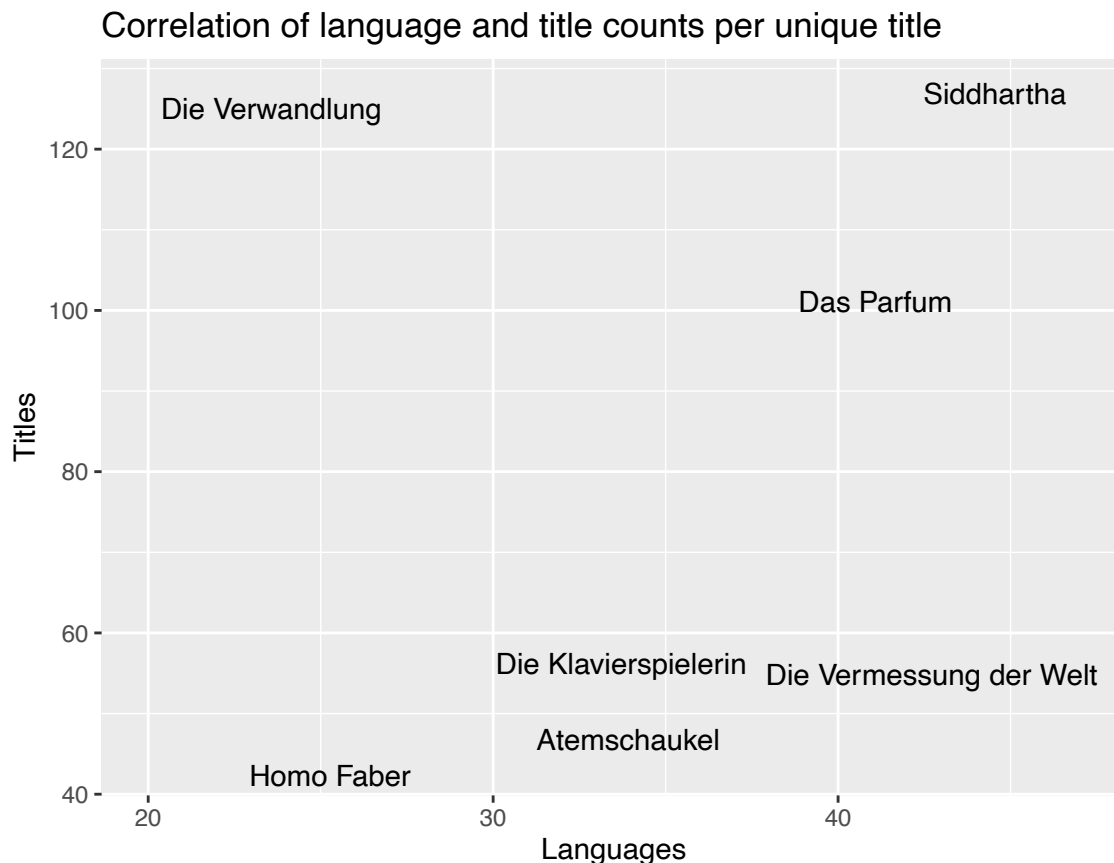
Figure 2.7. Correlation plot of language count and sum of editions for *Die Verwandlung, Siddhartha, Das Parfum, Die Klavierspielerin, Die Vermessung der Welt, Atemschaukel, and Homo Faber.*

Looking at the correlation plot in figure 2.7, we can see significant differences between the titles in terms of language and title count, which show precisely for which titles

---

[160] Calculating the correlation coefficient between title and language count for unique original titles revealed that they are significantly and highly correlated (tau 0.9221168). Additionally, a Chi-squared test of independence returned a p-value of 0.98 meaning they match almost perfectly, which would mean that language and title count overlap highly (I removed all titles with less than 10 translations since this slightly improves the significance of the test). We can describe titles with unexpectedly higher language counts as being positioned in a translational field, while partially also in the overlapping canonical field of titles with high title counts. However, when testing which titles are the ones that contribute mostly to the result of the Chi-square test based on residuals, the titles with the highest percentage are canonical classics (*Die Verwandlung* 6.7%, *Faust* 6.5%, *Der Prozeß* 5.2%, *Die Leiden des jungen Werther* 4.7%, followed by *Der Steppenwolf* and *Der Zauberberg* 1.4%, *Siddhartha* with 1.3%, *Das Parfum* 0.9% and all other titles under 1%).

their frequency count and the language count overlap. We can already tell by the positions that three of the titles (*Siddhartha, Die Verwandlung*, and *Das Parfum*) have significantly higher title counts (>100) than all others, making them highly concentrated translations. *Die Verwandlung* stands out with a surprisingly low language count as compared to the other titles but closely positioned to *Homo Faber* on the x-axis. With a large gap, the other titles are situated in the lower quartile of the graph (<40 titles), while nevertheless having between 20 and 40 translations. When comparing *Die Vermessung der Welt* with *Das Parfum*, we can see how the latter is situated at the same language range as *Das Parfum* with more than 100 titles. Interestingly, the same relationship between Jelinek's *Die Klavierspielerin* and Müller's *Atemschaukel* is visible as in figure 2.6, with both titles positioned around 35 languages, further illustrating that it does not take an overrepresentation in terms of title in order to be distributed across many languages. Both novels, as well as *Die Vermessung der Welt*, for which title and language count almost perfectly match, underline the argument that an author's position in the fields is also related to the distribution of their titles. Hence, in addition to the visualization of position of authors (figure 2.5) in the field of translation, these examples illustrate the balance and imbalance between concentration and linguistic distribution—the gap as well as the overlap of the canonical and translational fields in the library catalogue—based on an author's or work's position in the field of translation.

To sum up the findings of this section, correlating language and title count reveals that first, the most linguistically distributed authors are also those with most editions, supporting the observation that there is an imbalance of translated titles and their linguistic distribution. However, this is only the case for a few authors, while the majority of authors are positioned on the overlap of the canonical and translational fields. Secondly, it reveals a group of (mostly male) authors exceeding the threshold of concentration or linguistic distribution, which represents outliers in the translational field. Last, it shows that within the translational field, some author's positions are tied to single title major successes driving their translationalism. Hence, translationalism has been measured by correlating title count and language count for each title and author, revealing authors with unexpectedly higher language counts as being positioned in a translational field, authors in the canonical field with a high title count, and authors positioned on the overlap of canonical and translational fields. As this exploratory model and chapter underline, from juxtaposing and correlating concentration and language distribution, we can see two distinct groups of titles and authors appear, that, even though they statistically seem to overlap, differ when factoring in how many languages they

have been translated into. This underlines my argument that not only is the library catalogue concentrated on few languages and mainly Eurocentric, but also that only a few authors are significantly more linguistically distributed and highly translational.

In this chapter, I propose a model to analyze which authors and languages translations in the DNB catalogue represent. By correlating title concentration and linguistic distribution, the imbalance of specific languages and the position of authors in the translational field become visible, also revealing the positions of languages and, especially, the role of central languages, just as Heilbron suggests. The findings show that a small number of central languages play a significant role in the field of translation and that only a few titles and authors seem to reach smaller languages.

So, how does factoring in translationalism change or expand our understanding of the function of the library catalogue as an archive of the national literary canon in the world literary space? The findings presented in this chapter imply that the library catalogue is highly focused on few languages and authors which correspond to the representative authors of the national canon. In response to the question of which authors are most represented in translation in the DNB, the library catalogue displays an overall canonical tendency, favouring canonical authors—the gap between *Siddhartha* and *Die Vermessung der Welt* and their title counts is an excellent example of that. Accordingly, the function of the library catalogue appears to be not much different than Hesse's list of world literature for which he chose the term *Bibliothek der Weltliteratur,* hinting at the function of a library in compiling representative works of national literature.[161] While the translational field in the DNB appears to be small, canonical works and authors have a high visibility within the catalogue, which highlights the role of the DNB as a representative archive of a national canon and its extension through translation in the world literary space.

As I hope to show in this thesis, making translations visible in the catalogue and investigating the relationship between titles and languages for authors of German fiction challenges the classification frameworks for the canon of world literature by documenting cataloguing practices in the German national library collection. It investigates which works of fiction and authors are favoured and which ones are "left out" from translation networks and circulation. It also emphasizes the role and significance of a language model whereby authors can be positioned in the literary field. Additionally, the correlation of title versus language

---

[161] Mani has pointed this out as well. In *Eine Bibliothek der Weltliteratur*, he writes, "the library becomes an agent of the idea of world literature, which it in turn both contains and disseminates" (Mani 2013, 149).

count allows us to identify which authors are comparatively more or less translational. Furthermore, the authors beyond the threshold give first insights into which authors are connecting languages, which serve as a base to model the network and identify regional subgroups of authors.

However, the descriptive statistical measures applied in this chapter are limited to giving a general overview of language concentration, while disregarding any connections and ties between the respective languages and authors. Beyond concentration, which languages have influential positions in the translation network by connecting to other languages by shared authors? Are highly concentrated languages also highly connected in that they share the majority of authors with other languages? Can we identify sub-groups of languages based on the authors they share? And, most importantly, which authors are connecting concentrated and less-concentrated languages? Can we see a set of canonical or translational authors connecting language communities? While this chapter forms the basis to explore these questions, the next chapter is dedicated to modeling the language network of translations in the German National Library catalogue.

3. The language network of German fiction in translation

As I discuss in chapters 1.2 and 2, within the field of sociological translation studies, Johan Heilbron formulated a model to analyze the dynamics within the field of translation by centering it on language as unit of analysis, focusing on their positions on the international map of translations.

Drawing on Heilbron's network model and his central hypotheses about the cultural field of translation, this chapters explores the use of social network analysis to describe and measure the positions of languages in the translation network as a method to make visible translational communities in the DNB catalogue. This chapter further builds on the hypothesis of a concentrated and centralized language field, raised in the previous chapter, by first applying social network analysis and secondly, investigating the translational and canonical authors' connecting function in the language network.

Within Heilbron's approach to translation as a transnational cultural field (Heilbron 1999) of languages and their connecting values and categories (e.g., peripheral versus central), he raises two main hypotheses which are of importance for this study.[162] First, he states that the language network is hierarchical based on the centrality of a language according to its share in the market of translations. For a model of a translation network, this is important because it allows us to identify the language network's main actors as central or peripheral based on the authors they are connected by. For instance, central languages like French or English will have a large share of connecting values or edges in the network. Even though the national catalogue is not representative of the market share of each target language, centrality can be measured, as suggested by Heilbron, by the language's connecting values—translated authors that languages have in common.[163] Accordingly, Heilbron's model is applied to measure centrality and connectivity in the language-author network with the expectation of a highly centralized network into central and peripheral languages based on shared authors.

---

[162] He also raised the hypothesis that the language network is longitudinal ("dynamic constellation"), meaning that the position of language groups changes with time, gradually (due to cultural reorientation) or suddenly (due to the shift of political power to a strong centre or regime). However, testing this hypothesis lies beyond the scope of this study since longitudinal data extracted from the German National Library Catalogue alone is not reliable enough (see chapter 2.0).

[163] While Heilbron takes into consideration prestige (literary prizes) which may also shift a language to a more central position (e.g., international recognition may prompt shifts in an author's popularity in peripheral literature), he does not include it in his model.

Secondly, Heilbron states that literary transfer is dependent on each language's position in the network and "what is translated from one peripheral language into the other depends on what is translated from these peripheral languages into the central languages." (Heilbron 1999, 435). Within the framework of this study, this hypothesis finds application in investigating the relationship of centre to periphery and which authors they are connected by, as well as the roles of intermediary or vehicular languages (such as the role of English as gateway to global market distribution) based on their positions in the network. Depending on each language's position, it is possible to describe their influence and reach in linking different parts of the network, as well as the groups of authors linking different fields in the network. Hence, analyzing the connecting value that bridges large distances in the network also allows for investigating the function of canonical and translational authors.[164]

Even though the type of network analyzed in this section differs from Heilbron's—in that it is catalogue-specific, focused on authors as connecting values, and only including outgoing translations[165]—his central hypotheses provide the pillars of a theoretical and methodological framework to model translations as a network based on the language positions, their roles as central, intermediary, or peripheral nodes of transfer, and the relation between their positions and literary categories. Modeling the language network accordingly can then conceptualize how each language's position shapes canonicity and translationalism.

Further, Heilbron's model is only based on titles as the connecting values between languages and not authors, while the model proposed in this chapter is built on the author as the connecting value as well as the central social actor. The theoretical goal of the proposed model is to make visible areas of canonical and/or translational flow, looking at the translation network in terms of authors connecting languages operationalizing Heilbron's network model and building on his central hypotheses.

Based on Heilbron's model, I propose three main hypotheses. First, I discuss the notion of a centralized language field, questioning whether there is a clear division of central versus peripheral languages. By measuring degree centrality—how many connections/authors a language has with other languages—I follow the argument that language centrality is based

---

[164] Heilbron also argues that "centrality implies Variety": "The more central a language is in the international translation system, the more types of books are translated from this language" (Heilbron 1999, 438). Here he refers to book categories such as "religion," "law," and "leisure." Since the library catalogue data only includes sparse genre descriptors, testing this hypothesis lies outside of the scope of this study.

[165] According to Heilbron "it is not so much the national tradition, but rather the international position of national cultures, which determines the level of cultural importation" (Heilbron 1999, 440). From German's proportionally high percentage of translation into German, we can assume its central position for outgoing translations.

on the share of canonical and translational authors in the translation network. Thus, I expect the network to consist of a few central, large languages, which, however, may share significant amounts of authors with smaller, more peripheral languages. Hence, centrality is a measurement for identifying centre and periphery, but also for assessing overall connectivity in the network.[166] In other words, the more connections authors make between few languages, the more separated centre and periphery appear, which accounts for a highly connected, centralized language network.

Secondly, I raise the claim that the positions of languages (intermediary, central, peripheral) are decisive for the formation of the translational and canonical field of authors. Following Heilbron's second central hypothesis, I map the translational relationships between languages based on their positions as central or peripheral and their connecting roles in the network. Looking at the language's positions, I argue, serves to identify which authors connect the highly influential or marginal languages, the underlying assumption being that if an author is among the smallest languages, it is highly canonical, considering that only few highly canonical, representative authors reach the periphery of the translation network. In order to measure a language's influence and reach, I measure betweenness to identify languages that lie on paths between other languages and therefore have a position of influence in the flow of translations in the network.[167] As I explain in detail in the second section of this chapter, degree centrality indicates how well connected the node is to other nodes, while betweenness gives a sense of a node's connecting function as an important intersection of flows of translation. In addition, I also measure eigenvector centrality in order to assess which languages are connected to important languages. This is especially interesting when looking at peripheral languages, which may only be connected to central languages like French or English by few connections. These connections are then examined in order to see which authors they represent.

The third notion explored in this chapter is the function of authors in the author-language network as bridges between central and peripheral languages, asking if they constitute communities of canonical and translational authors. In order to split the network

---

[166] I rely on Kilduff's and Brass' definition of centrality as "the extent to which an actor occupies a central position in a network by having many ties to other actors (i.e., degree centrality), by being able to reach many other actors (i.e., closeness centrality), by connecting actors who have no direct connections (i.e., betweenness centrality), or having connections to centrally located actors (i.e., eigenvector centrality)" (Kilduff and Brass 2010, 355-56). Closeness centrality has not been measured because it only applies to connected graphs, while the language-author network graph is disconnected.

[167] According to Newman, betweenness is not related to connectedness like other centrality measures, but "it measures how much a vertex falls 'between' others" (Newman 2010, 189). Likewise, a node can have a high betweenness and a low degree centrality and vice versa.

into central and peripheral languages based on their connections, I use community detection algorithms which divide "the vertices of a network into groups, clusters, or communities according to the pattern of edges in the network" (Newman 2010, 254). Based on the assumption that there is a set of authors unique to each group of languages and a set of authors connecting those groups, I investigate which authors are exclusive to peripheral or central languages and which authors have a connecting role between language groups. My expectation is that highly canonical authors are more likely to connect the central to peripheral languages. On the other hand, I expect to find a set of authors unique to only peripheral languages that sets them apart from the central language group. From this approach, canonicity can hence be defined not only in terms of concentration, but also in its tendency to connect central and peripheral language groups in the network. Following these three main notions, by applying different network centrality measures and community detection algorithms, I map the network of literary transfer between languages and the authors these languages are connected by in order to explore the role of canonical and translational authors in the German National Library catalogue.[168]

Categories of canonical or translational hence are defined in terms of positions of authors and languages and may form distinct communities within the network. For example, in a network model the canonical author Daniel Kehlmann and the author Juli Zeh both have translations into some of the central languages (English, French, Spanish) which are connected by their common authors. Walser, however, has several translations into peripheral languages, such as Armenian, Lithuanian, and Belarus, and hence also functions as a connection between the centre and periphery. At the same time, he also belongs to a separate group of authors that connect peripheral languages. Thus, not only is Walser an author represented in both centre and periphery, but he also bridges language groups, which can be indicative of the author's translationalism based on their linguistic distribution.

As this example illustrates, the proposed network model analyzes authors in terms of their function in the network and their role in bridging target languages, to form distinct groups (canonical or translational), and to transgress their boundaries. As Butts argues, "to represent an empirical phenomenon as a network is a theoretical act" (Butts 2009, 416) which resonates with the decision to represent the translation network through the information contained in the bibliographic metadata (languages, titles, publishers, editions, or authors) and modeling it by applying social network analysis.

---

[168] While Heilbron defines centrality of a language according to the share in the market of translation (Heilbron 1999), my definition is limited to the share in the German National Library catalogue.

In summary, social network theory is the application of a network model to analyze an actor's positions (nodes) in a network by measuring their connections (bridges).[169] In order to construct the network, a set of edges (connections) needs to be identified, whose elements are termed nodes, which in this case represent languages. For the language network, connections are based on translated authors. Therefore, a connection is defined as shared authors between languages. If a title by a certain author has been published in English and French, for instance, then those nodes are connected. In the network, authors therefore represent the relationship between languages (edges). The connecting values of all relationships (edges/authors) between all languages (nodes) is visualized as a network graph. A network model therefore reveals how and by whom languages appear to be connected and contributes toward an understanding of connectivity between languages by authorial circulation, while also helping to identify which authors cross linguistic boundaries and form the channels of transfer.

From here, one may further ask "how do the generic networks of canonical novels compare to those that are more obscure?" (Dewitt 2015, 176). Dewitt makes an excellent point that a network approach to periodicals makes visible the invisible patterns in how genre categorization is a process of readers putting novels into context with each other. A similar point can be made in terms of translations and the library catalogue. Unlike genres, translations are not categories set by readers, but their connections between languages reveal specific communities of authors that represent the canonical and translational categories beyond national boundaries. Additionally, as Dewitt's study on genre communities of authors shows, community detection within the network further illustrates the importance of specific nodes in connecting communities (or bridges between canonical and translational). While in each language community there are different actors that make these connections, such as editors, translators, and authors themselves, language communities represent literary cultures, which can have significant influence in the circulation of translations, e.g., English as the bottleneck language. In the field of translation of German fiction, a language network model allows for analyzing the function of authors in denominating and bridging different literary cultures by means of their translationalism.

In the previous chapter, canonicity in the translational field is defined as concentration of specific authors in large, central languages, while transnationalism refers to authors spread out over several smaller languages. This gives a static image of the translational field, while

---

[169] See Butts 2009: "The network is represented by a graph, which is defined as the set of nodes together with the set of pairwise relationships among them" (325).

an underlying dynamic between languages exists that shapes the movement of authors in the field. Languages share connections by publications of the same set of authors, and authors share connections by languages. Based on those connections, a network is visible, in which we can locate the areas of canonical classics and contemporary works of fiction.

Additionally, the last chapter asks whether descriptive statistics alone can serve as a sufficient measure of canonicity or translationalism. Clearly, descriptive statistical measures such as juxtaposing and correlating concentration of titles and authors according to target languages are limited methods to address these types of questions in that they describe predominantly which languages the catalogue is focused on. Accordingly, such inquiries require looking at the dynamics within the networks, the central actors, and their positions, since descriptive statistical measures disregard the relationships between the central actors in the translation network. By applying a network model to the catalogue data of translations extracted from the German National Library, this chapter presents first-time network visualizations of the relationships between languages and authors by which we can see how densely connected translations are in their function to linking cultures through co-occurrence of canonical and translational authors.

Social network models within digital humanities research and translation studies have been mostly limited to single authors or a set of representative authors, character relationships, or specific genres and periods, and have not drawn on the DNB catalogue as a data resource and object of analysis. Multiple digital humanities projects mostly focus on character networks (Alexander 2019, Elson et al. 2010), periodicals, and literary exchange with an emphasis on visualization. Most prominent is the Stanford Republic of Letters project, a network of letter correspondence between famous authors that explores the use of visualization tools to reveal new perspectives on literary exchange (see Chang et.al 2009). Additionally, several studies have applied network models on characters in fiction in order to confirm or refute a pre-existing genre-classification in literary theory by modeling character networks and factoring in authorship (see Ardanuy and Sporleder 2015 or Park et al. 2013). Hence, studies have not specifically addressed the literary canon in translation and their categorization even though their data is mainly extracted from a selected number of canonical works representative of a national literature (e.g., nineteenth-century British novels, see Elson et al. 2010).[170]

---

[170] "Our corpus of 60 novels was selected for its representativeness, particularly following categories: authorial (novels from the major canonical authors of the period), historical (novels from each decade), generic (from the major sub-genres of nineteenth-century fiction)" (Elson et al. 2010, 138-139).

Additionally, network models have been theoretically but not extensively quantitatively explored in translation studies. In their book *The Network Turn: Changing Perspectives in the Humanities*, Ahnert et al. give an overview of the network approaches in the humanities and, specifically, digital humanities, but do not mention a single study on literary translations (see Ahnert et al. 2021). Nevertheless, there is an increasing interest in analyzing translations of mainly canonical authors, such as the project Worldliterature@UCLA,[171] for which bibliographic data is used on translations to show world literary networks of circulation. However, up until now their network analysis is limited to one author at a time (see Kim and de Carlo 2015) which is the case for multiple projects on translation networks, where a dataset only includes one publisher, one author, or one work. Şehnaz Tahir-Gürçaglar (2007), who did a network analysis of the Turkish publisher Altin Kitaplar, points at the current lack in scholarly literature to address questions about the canon's representation in translation. Library catalogues and the bibliographic data of translations have been proven to be a rich resource of data for network analysis, as Katrin Dennerlein's study (Dennerlein 2020) illustrates: she uses the Austrian National Library data to create a theater play network analysis of authors that connect specific collections in the library catalogue. For the German National Library catalogue, however, to date network analysis has not found extensive application that is not limiting the network analysis to a single publisher, author, or time period, especially for analyzing translated German fiction, its linguistic communities, and the authors connecting or signifying them.

Mapping translation networks of library catalogue data therefore calls for exploring social network methods and theoretical models to make visible dynamics between languages and their clusters, as well as the different functions of authors. By taking up on Dewitt's contribution and expanding its object of study to translations within the DNB library catalogue, the network model explored in this chapter addresses this lack of network analysis by exploring translation data of German outgoing fiction with special attention to the role of authors in connecting linguistic communities as a sign of their translationalism and/or canonicity. Modeling the network of transfer sheds light on the role of languages and authors as central or peripheral, as bridging or secluding. Again, when thinking about the visibility of translations and specific authors within the library catalogue, their positions are defined by their relations and marginal, peripheral positions in the network. Mapping the languages and authors' positions allows us to look beyond the centre of the catalogue and come closer to

---

[171] "WorldLiterature@UCLA: Translational Network Analysis in the Digital Humanities." Accessed August 11, 2022. https://worldlit.cdh.ucla.edu.

making visible what constitutes the catalogue. Based on the definition proposed in previous chapters of translational authors as being present in many languages, what connects linguistic communities can be indicative of locating this set of authors. The communities of authors and their connecting functions not only situate them in the network, but also constitute an additional method to locate canonical and translational authors. In other words, making visible the translation networks between languages offers a way to explore translations in the DNB by challenging the concept of national library catalogues as representative collections of a literary history based on a national canon. Modeling translations, as I argue in this thesis, makes visible the dynamics and sub-communities present in library collections and allows for re-evaluating what a national canon constitutes within a national collection.

A centralized language field

In chapter 2, I measure the degree of concentration of titles per language and infer that the library catalogue is highly concentrated on few languages and favours Indo-European languages, arguing that languages with a concentration of titles occupy a central position while languages with fewer titles are situated semi- or peripherally. In this section, I test this argument and build on Heilbron's first central hypothesis, measuring the extent to which languages occupy central positions in a network by having many ties to other languages and being highly connected, and which have a peripheral position having few connections, by applying degree centrality measures. By applying social network measures for centrality and thereby giving a sense of "the extent to which a network is centralized around one or few actors" (Kilduff and Brass 2010, 356), it can be tested whether the catalogue is as concentrated on few central languages as is shown in chapter 2 and how the centre connects to the periphery. Some languages may be more central in sharing authors with peripheral languages. In that case, we may ask, which members form a distinct cluster by which a core set of influential, highly connected languages can be identified? Do periphery members connect only to core members and not to other members of the periphery, as Kilduff claimed? In order to discuss these questions, I generate an undirected—the flow of information between nodes is equal and has no specific direction, e.g., from English to French—and unimodal network graph, where languages (nodes) appear connected if they share translations by the same authors.
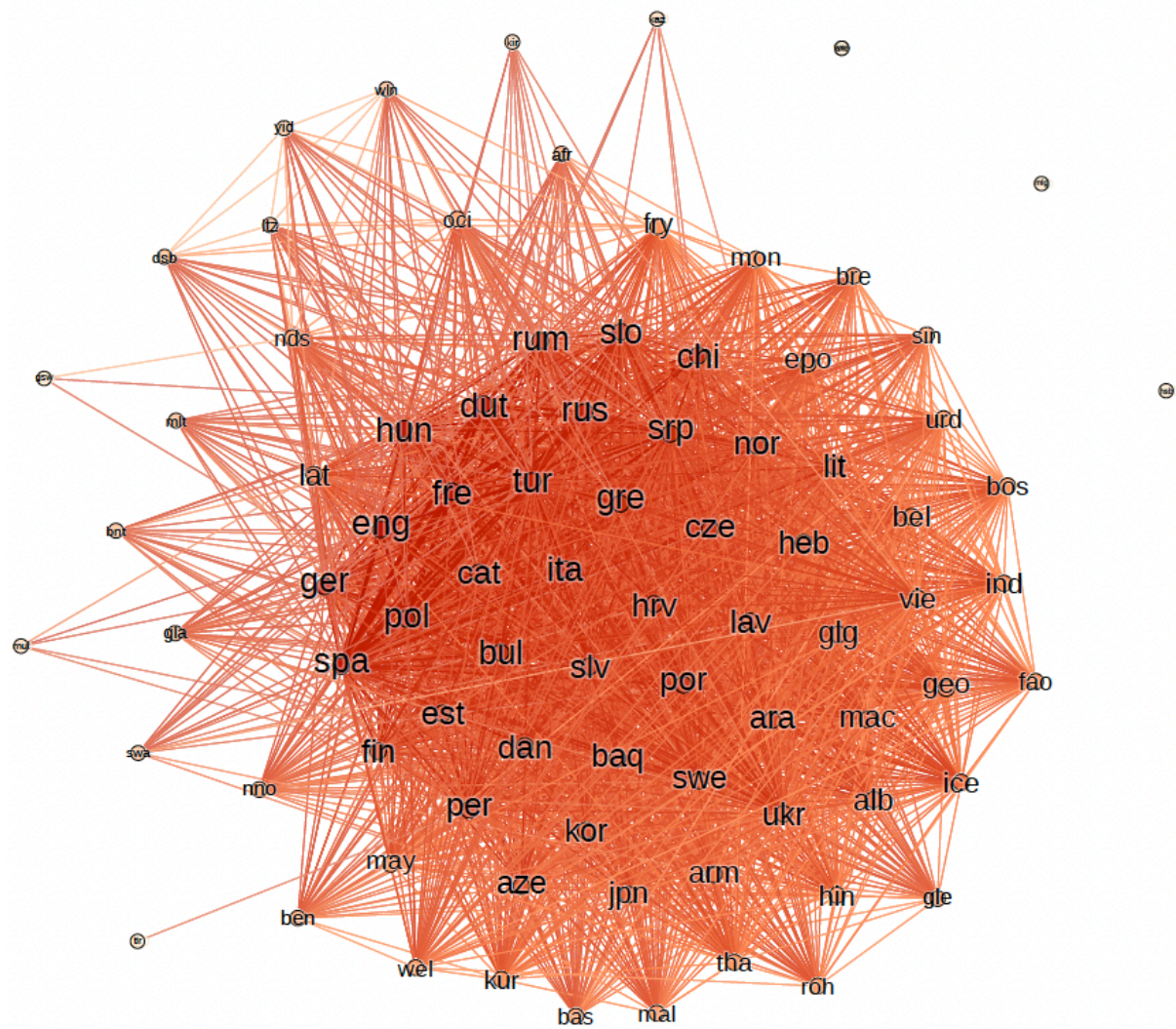
Figure 3.1. Network graph of languages with shared authors as edges[172] produced with Gephi. The size of the node labels and color—the darker the more connections—has been adjusted to the number of connections (degree).[173]

Visualizing the full network graph (figure 3.1) shows a highly connected network of languages with few isolates and lesser connected languages. The languages that appear to be most central clearly stand out in the mid-left centre of the graph. The central position of Spanish (spa), English (eng), Bilingual (ger), Polish (pol), French (fre), Hungarian (hun), but also Turkish (tur), Greek (gre), Russian (rus), Dutch (dut), Bulgarian (bul), and Catalan (cat) is visible. The overall shape is star-like with peripheral languages especially on the left side, including Norwegian (nno), Gaelic (gla), Swahili (swa), and Bantu (but), being highly connected to the central languages, but not with each other. Looking at the overall color

---

[172] I used the Fruchterman Reinhold layout to avoid overlapping nodes.
[173] In undirected networks, there is just one type of degree (one type of connection between nodes), and hence only one number for the degree centrality. If the network would be directed, meaning that it would matter from which to which node the translations go and in what succession, we would need to calculate in-degree and out-degree (see Newman 2010, 9).

shade of the edges, we can also see that Spanish, English, Hungarian, Polish, and Dutch share strong connections and therefore have many authors in common that are published in translation in these target languages. While figure 3.1 gives a picture of connections and the language's position in the network, it is necessary to further measure the number of connections to establish how connected the network is and which languages have an influential, central position by sharing many translated authors to other languages.

To map languages in central versus peripheral positions, this section focuses on degree centrality in line with Heilbron's first central hypothesis stating that the language network is organized in centre and periphery. Social network theory often describes degree centrality as an influence measure or a spatial map of static relations between a centre and a periphery. For Folaron and Buzelin:

> the notions of *centrality* and *periphery*, … define the relative power of each component of a particular network. A central node—or *sociometric star*—is one that has many relationships and that is in a position of control with respect to other nodes. Reversely, a peripheral position is one that is loosely connected to a limited number of nodes, the extreme case being that of the *isolate*, i.e., a node that is not connected at all. (Folaron and Buzelin 2007, 612-613)[174]
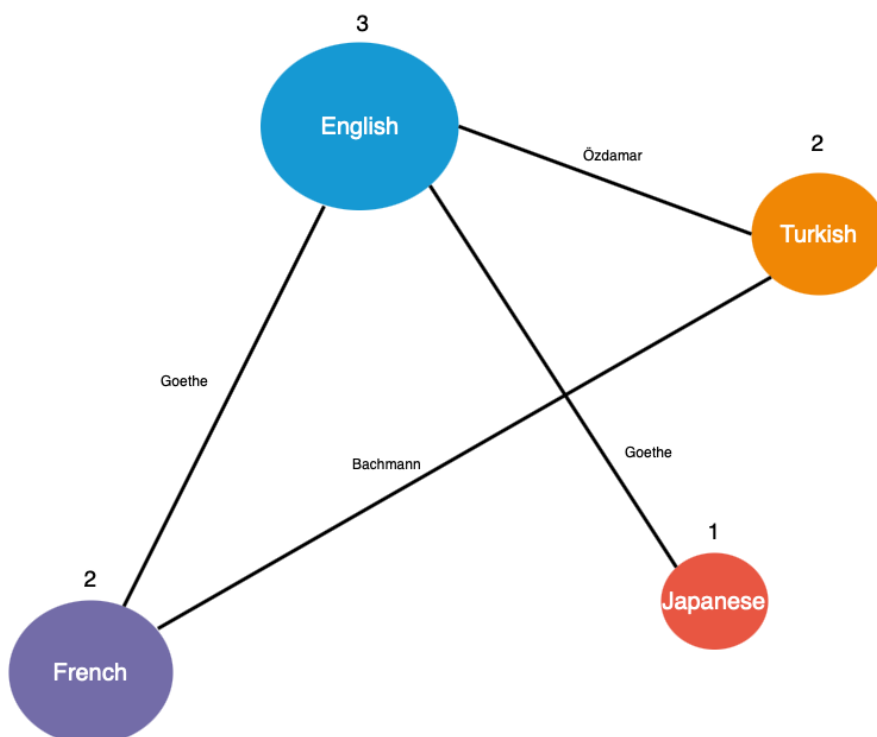
Figure 3.2. Example of a network graph with authors as edges and languages as nodes. The degree is

---

[174] In line with Folaron and Buzelin (2007), Borgatti 2005 also applies centrality measures as the number of ties incident upon a node or a measure of immediate influence.

indicated by the number on top of each node. French has two authors that are also translated into English and Turkish, based on which French has a degree centrality of 2. Turkish also has two edges representing shared authors with French and English and a degree of 2. English has a degree of 3 and Japanese of 1. English, having the highest degree, appears to be the most connected and hence central language in the network, while Japanese appears to be the most peripheral, least connected by its degree.

In the first section I ask: Is there a clear division of central versus peripheral languages? Focusing on degree centrality, defined as the number of connections by shared authors between languages, I argue that language centrality is based on share of canonical/translational authors in the translation network and based on the findings in chapter 2, I expect the network to consist of few central, large languages which however may share significant amounts of authors with smaller, more peripheral languages. For each language the number of edges (representing authors) connected to it are calculated. In the case of the translation network for which a connection is defined as shared authors between languages, for central languages, we can expect a high degree centrality, while for peripheral languages, only a few ties to other languages.
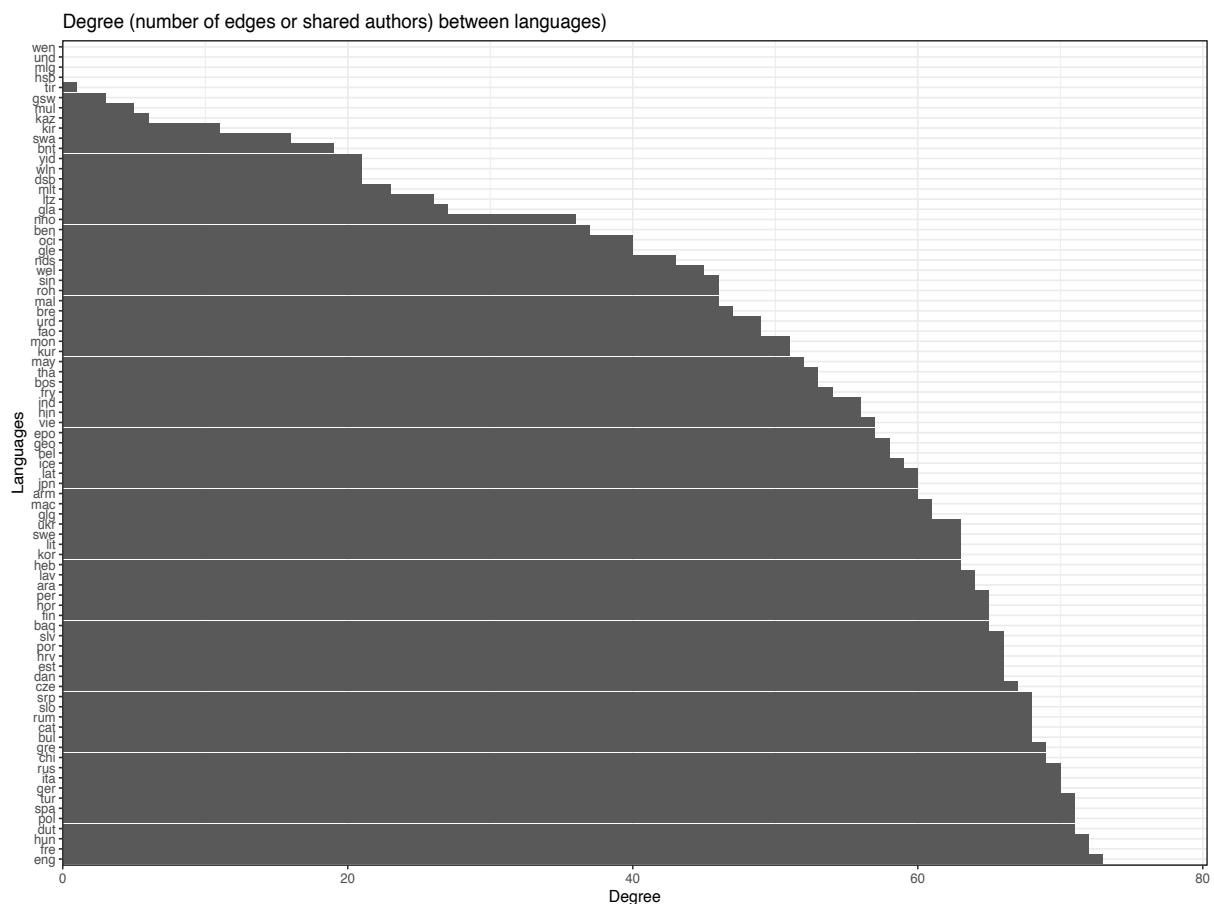

Figure 3.3. Degree distribution of edges (shared authors) per language.

Measuring degree centrality[175] shows that the network is highly connected, and, besides a few isolates, there is no clear distinction of central and peripheral languages. The network consists of 81 nodes and plotting the degree distribution (figure 3.3)—the number of edges per language—shows a normal distribution with only a few isolates that do not connect to any other language (Sorbian languages, Undetermined, Malagasy, Upper Sorbian). The mean (49.75) appears to be high which supports the argument that the network is highly connected.[176] The "extent to which a network is centralized around one or few actors" (Kilduff and Brass 2010, 356) appears to be lower than expected in that the translation network is highly connected, implying that the periphery too is highly connected and therefore plays a significant role in the transfer of authors.

In the same line, the degree distribution in figure 3.3 shows a normally distributed network, not clearly separating central and peripheral language by the number of ties they connect by. While few central, well-connected languages were expected by measuring degree centrality, the language network appears to be well-connected in peripheral languages as well. The most central and well-connected languages according to degree distribution reveal the role of non-Indo-European languages in the network. While measuring the frequency distribution for author count per language in chapter 2 shows that the language with the highest author count is English, followed by French and Czech, Spanish, Italian, Dutch, Polish and bilingual, these results partially match up with the degree distribution. As figure 3.3 illustrates, English is also the most central language connecting to 73 other languages by shared authors, again followed by French (72) but also by Hungarian (72), Dutch, Polish, Spanish, and, particularly, Turkish (all 71). While the role of English again underlines its function as a bottleneck language, the positions of Hungarian and Turkish challenge the notion of a Eurocentric translation network. Compared to the language-author distribution presented in chapter 2, the role of less-concentrated languages such as Turkish or Hungarian is apparent. The comparable position of Hungarian and Turkish to English highlights the connecting role of languages which, according to their title concentration, would appear in a less favourable position for the formation of the translational field of authors.

In chapter 2, author concentration revealed a clear separation of languages with a long tail of peripheral languages, while here, degree centrality reveals a large group of central and a small group of peripheral languages, without a clear separation between central and semi-

---

[175] Using the degree() function in the R package "sna" (see Butts 2020).
[176] Additionally, transitivity (how many closed loops there are as a percentage of length 2 paths) also accounts for a highly connected network with a score of 0.86.

peripheral languages. Hence, measuring degree centrality of languages that share authors challenges the hierarchical separation between central versus peripheral languages which stands against the chapter 2 findings that the library catalogue is highly concentrated on few languages as compared to all other languages. A network approach therefore reveals that even though by author concentration certain languages prevail over others, those other languages are still highly connected by shared authors and play significant roles in shaping the translation network.

From figure 3.3 of the degree distribution, it can only be observed how connected the language network overall is by shared authors, which does not provide insight into the types of connections. Hence, what does the network for highly connected languages reveal and what does that imply about the role of central language for the transfer of the translational canon? Which are the most connected languages, and which other languages are they connected to? The aim here is to investigate the relationship between central, peripheral, and semi-peripheral languages. Further, by looking at less-connected languages with less than 100 shared authors, I test Heilbron's hypothesis that peripheral languages are not well-connected to each other because they mainly connect to central languages. In order to visualize the network of only highly connected languages, I use the edge weights—the sum of edges a node has, representing the sum of authors they share with other languages—to subset the network nodes by languages with more than 100 edge weight and languages with less than 100 edge weight, and then re-calculate the degree centrality.

**Language Network
by shared authors (>100)**
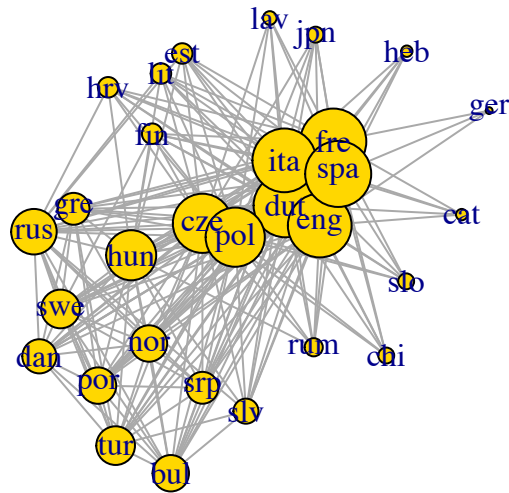
Degree Centrality

Figure 3.4. Network graph of degree centrality for the subset of languages with an edge weight of more than 100. Node size is adjusted to degree.[177]

As shown in figure 3.4, within the subset of highly connected languages (>100 edge weight), a clear separation between peripheral and central languages becomes visible.[178] We can see that while central languages cluster together in the middle of the graph, peripheral languages are arranged in a star-like composition around them on the right side. As expected, the central languages are mostly European: French, Spanish, Italian, Dutch, and English. On the other side, there is another semi-peripheral cluster of languages with fewer connections, including Portuguese, Bulgarian, Swedish, and Danish. This cluster appears to be more

---

[177] This and the successive graph visualizations use the Kamada Kawai force-based algorithm which enables the display of all nodes with the least overlap possible while still taking into account their positions. See Kamada, Tomihisa, and Satoru Kawai. "An Algorithm for Drawing General Undirected Graphs." *Information Processing Letters* 31, no. 1 (April 1989): 7–15. https://doi.org/10.1016/0020-0190(89)90102-6.

[178] It is important to mention that node degree and centrality depend on the number of nodes and edges in the network. If the size of the network changes, it will also affect which languages appear as central or peripheral. We see this in the histogram of the degree distribution for languages with more than 100 edge weight (see appendix: alldnb_lang_lang_author_above100_degree_ch2.2_v2), which does not correspond to the previous (see position of Hungarian). The maximum number of ties (29) and mean (15.20) is less than for the complete network and looking at the top languages, we can observe a shift with French, followed by Spanish, Dutch, English, and Italian as the most central languages.

connected as compared to the peripheral languages like Japanese, Slovenian, and Chinese. We can also observe that Czech, Polish, and Hungarian take connecting positions between the central and the semi-peripheral languages. We can hence argue that for strongly connected languages, central and peripheral languages are distinguishable, and the connecting role of semi-peripheral languages is more apparent.

Visualizing the node sizes according to degree centrality and observing the shapes of the edges on the graph gives a sense of whether languages connect to mainly central or peripheral languages as well as their grouping. By looking at each language's node size, we can see clusters of central (the cluster around Spanish), semi-peripheral (the cluster around Swedish and Portuguese in the left side of the graph), and peripheral languages (the star-like cluster surrounding the centre on the right side of the graph). We can observe patterns of central languages forming the centre of the star, semi-peripheral languages as their own group which is highly connected within themselves, and languages with a lower degree, such as Catalan, Chinese, Danish, Bulgarian, or Estonian, that occupy peripheral positions, indicating that they are highly connected to few central languages. Even though peripheral languages may have many shared authors, in the case of Catalan or Chinese they only do so with few central languages and do not connect to each other, which supports Heilbron's argument that peripheral mainly connect to central languages and have few connections with other peripheral languages.

Especially looking at the peripheral languages in the subset, the roles few central languages play in connecting them become apparent. However, is this still the case when decreasing connectivity in the network by removing the strongest ties? What happens when we only look at lesser connected languages? How do only peripheral languages connect to each other by shared authors? Do they appear to be highly connected to each other in that their degree centrality increases when removing the strongest connections? Based on Heilbron's hypothesis, we can expect peripheral languages to not be highly connected to each other, but only to central languages. In order to visualize the network of lesser connected languages, I subset my language-author network by edges with less than 100 shared authors for which I used the edge weight, the sum of shared authors per connection between one language and another.

**Language Network
by shared authors (<100)**
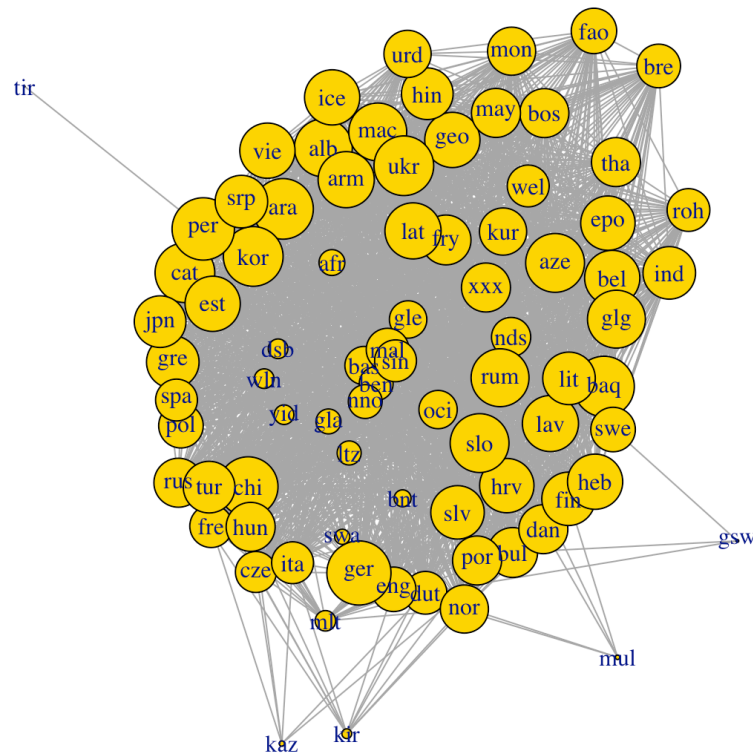
Degree Centrality

Figure 3.5. Network graph of degree centrality for subset (<100 edge weight).

As we can see in figure 3.5, modeling the degree centrality network of languages with an edge weight of less than 100 visualizes a highly connected, de-centralized network with few isolates.[179] Compared to figure 3.4, the focus shifts, showing a set of different languages' connecting roles. Languages with the highest degree for this subset are bilingual (72 degree centrality), Persian (70), Arabic and Basque (68), Catalan, Chinese, Korean, and Ukrainian (67).[180] As the node size indicates, most languages have a similar degree centrality, with the exceptions of Tigrinya (1), Swiss German (3), "mul" (multiple languages) (5), and Kazakh (6), which have only very few connections to central languages but none to each other. Additionally, there is no clear distinction of central and semi-peripheral languages, while similar to figure 3.4, there is a cluster of peripheral languages in the middle, which only have

---

[179] Including a transitivity of 0.738742.
[180] See appendix for full table: alldnb_lang_lang_author_degree_under100_ch2.2_v2.csv

a few connections based on their node size. While they do not have a high degree centrality, they appear to be highly connected and assume an important role in connecting different areas of the network by a few shared authors; indeed, their position in the network appears to be defined by strong ties of few frequently shared authors. Hence, this underlines the role of peripheral and semi-peripheral language in the library catalogue by forming few, but important connections by shared authors.

To summarize the findings of this section, comparing the complete network with a section of the most and least connections between languages indicates that the network is highly connected with distinct clusters, while a network with only the strongest connections visualizes the separation between peripheral and central languages. Zooming in on languages with more than 100 edges (shared authors) as compared to the complete network therefore supports Kilduff's argument that central languages with a high degree centrality connect to most other languages, while peripheral languages only share few ties with central languages. Increasing connectivity further shows that the transfer of translated authors appears to connect languages in a star-like shape, with a few central, large languages sharing significant amounts of authors with smaller, more peripheral languages. Additionally, when removing the strongest connections (by edge weight) from the graph, a highly connected network between peripheral languages emerges. This network is balanced in degree centrality with only one group of very small languages in the middle, which are highly connected by few authors. Therefore, to the previous argument we need to add that peripheral languages form strong connections to especially semi-peripheral languages, which contests Heilbron's hypothesis of peripheral languages only being highly connected to central languages. The role of central languages appears to be a strong factor in shaping the network in regard to their relationship with peripheral languages, while peripheral languages are highly connected as well.

However, what does this tell us about translationalism and the translational field in terms of linguistic distribution and shared authors? From looking at the type of connections, we can describe the flow of authors between languages in terms of their connecting function. Coming back to my initial argument, that centrality is based on the share of canonical/translational authors in the translation network, we can see that especially peripheral languages, even though they do not have a high centrality degree, can be highly connected by shared authors, and are therefore an area of transfer. While the findings from chapter 2 regarding author concentration per language suggest that the translational field is

small in that few authors are distributed across many languages, the findings of this section indicate otherwise. We can see that even though especially peripheral languages have few authors, they appear to be highly connected to both central languages and semi-peripheral languages and represent an area of strong translational transfer of some important authors. Again, we can assume, based on the findings in chapter 2 on the linguistic distribution of authors, that the share of highly canonical authors may link periphery to centre in the translation field. What connects core to periphery therefore can serve as an indicator for areas of translational exchange.

Language positions and influence

While degree centrality gives some indications about the roles of languages based on their connections, it does not offer insight into what types of roles each language occupies in direct relation to other languages, such as English's role as a bottleneck and broker language, Hungarian as connecting several semi-peripheral languages, and peripheral marginal languages as places of highly canonical exchanges with central languages. In this section, I apply two other centrality measures to the language-author network, eigenvector and betweenness, to further investigate the positions of influence of languages on the one hand, and, on the other, explore the specific roles of languages in connecting the network's centre, periphery, and semi-periphery. Building on Heilbron's second central hypothesis, my main argument here is that through the extent of influence (connecting different parts of the network) and reach (stretching into the periphery of the network), we can determine the position of each language in the network (centre, periphery, semi-periphery) and what connects them. This allows for exploring which authors connect the centre and periphery and which authors have the widest reach in the network. Based on my previous findings, I expect highly concentrated (canonical) authors to connect central to peripheral languages, while translational authors will be present in subgroups of peripheral languages, bridging different parts of the network. The underlying claim, therefore, is that the positions of languages are a way to locate the translational and canonical authors that connect the network. Hence, in addition to author concentration as a factor for canonicity, the number of shared authors between peripheral and central languages can also be identified as a factor for the circulation of the national canon in translation.

Eigenvector centrality measures the extent to which specific actors hold a position of influence in network in reaching many actors and is calculated by the degree of the

neighbours (how connected the neighbours are) and gives a sense of the extent to which actors are "having connections to centrally located actors" (Kilduff and Brass 2010, 355-56).[181]
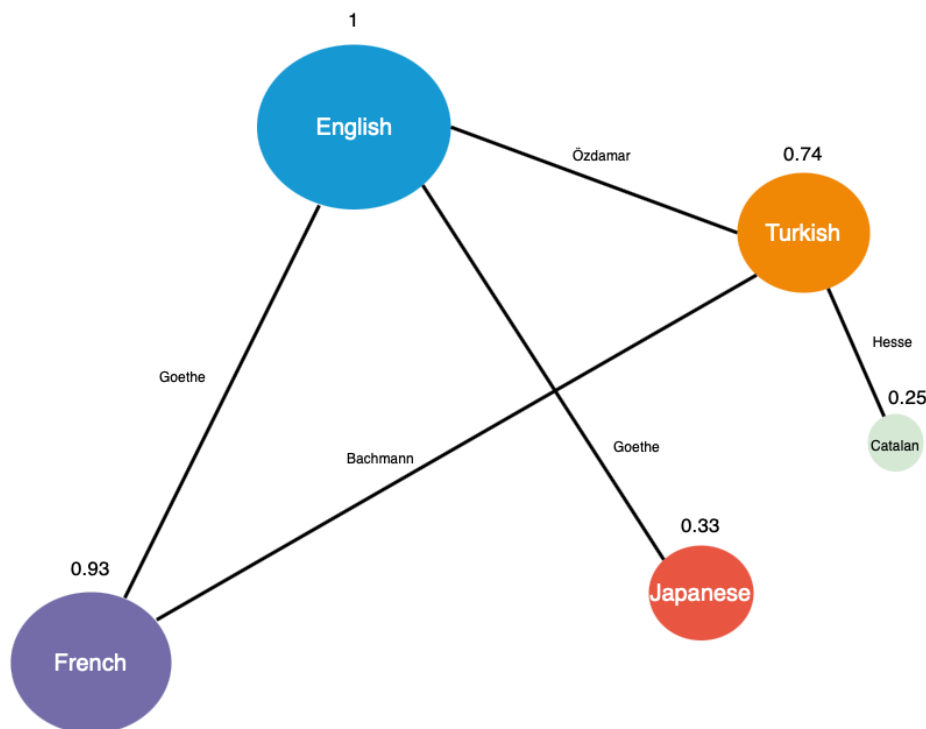
Figure 3.6. Example network graph of eigenvector centrality with eigenvalues indicated on top of the nodes.

For instance, as the basic example in figure 3.6 shows, if Japanese shares an edge with well-connected and centrally positioned English (eigenvalue of 1 since it is connected to all other well-connected neighbours), it will have a higher eigenvector centrality and hence position of influence in the network. This is even more apparent when comparing the case of Japanese to the eigenvalue of Catalan, which only shares one connection with Turkish and hence has a much lower eigenvalue. From comparing French and Turkish, it is apparent how the eigenvalue depends on the nodes' neighbours. Turkish is also connected to Catalan, which only has one edge and has therefore a lower eigenvalue than French, which is connected to the languages with the strongest neighbours (English and Turkish). Hence, French occupies an influential position by sharing edges with powerful neighbours, even though it has the same degree as Turkish which occupies a less central position in the

---

[181] Newman describes eigenvector centrality as follows: "We can think of degree centrality as awarding one 'centrality point' for every network neighbor a vertex has. But not all neighbors are equivalent. In many circumstances a vertex's importance in a network is increased by having connections to other vertices that are themselves important. This is the concept behind eigenvector centrality. Instead of awarding vertices just one point for each neighbor, eigenvector centrality gives each vertex a score proportional to the sum of the scores of its neighbors" (Newman 2010, 169).

transfer. As this exemplifies, the eigenvector centrality as a measure of influence therefore allows for examining the core-periphery dynamics: how central and peripheral languages relate to each other and influence the translational transfer. In other words, the well-connected languages condition the transfer of translated authors to other languages and the dynamics between centre and periphery.

When calculating the eigenvector centrality score for the language network,[182] languages that have a high degree centrality also occupy the top languages with the highest eigenvalues, although they do not follow the same order. When calculating the correlation of degree and eigenvector centrality for the complete network of languages, both values appear to be correlated (0.68 correlation coefficient[183]), indicating that languages that occupy a central position—French, Spanish, Italian, English, Dutch, Czech, Polish, Russian, Hungarian—also occupy the most influential position. Central languages, therefore, form tight relations with each other and appear to also have many relations to peripheral languages, which implies that they have a broad influence on the transfer of translated authors and represent fields with an increased translational flow.
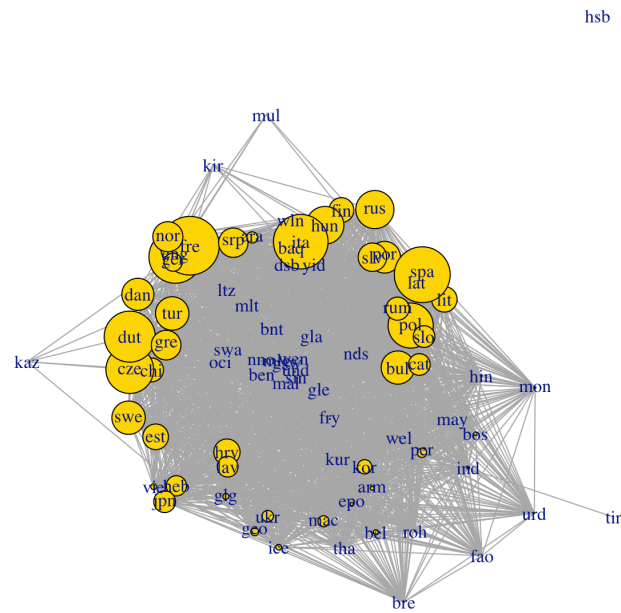
However, degree and eigenvector centralities show differences in the positions that specific languages occupy. For instance, in the language network, French has an eigenvalue of 1, meaning that it is connected to all nodes who themselves have high eigenvalues, while English has a lower eigenvalue as compared to its degree centrality. Spanish, too, has a higher eigenvalue than its degree centrality score and appears to occupy a more influential position than English. Since eigenvalue reveals which languages are positioned in a favourable place in the network, next to other influential languages, this means that highly connected languages per degree also assume an influential position in that they connect to other highly connected languages.

---

[182] Using the evcent() function in the R package "sna" (see Butts 2020).
[183] See appendix for correlation coefficients: alldnb_lang_lang_author_centralities_correlation_ch2.2_v2.csv

**Language Network
by shared authors (>100)**

Eigenvector Centrality

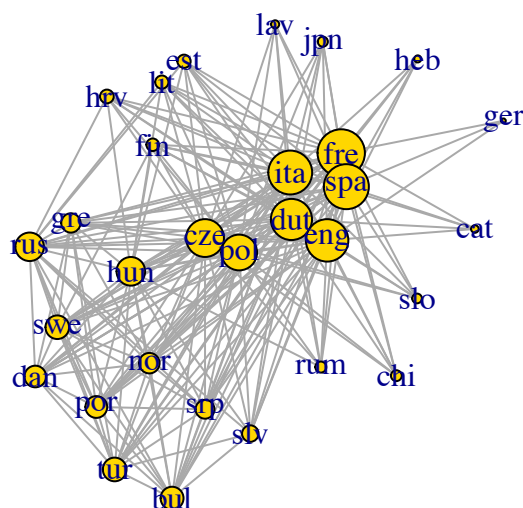**Figure 3.7.** Network graph of eigenvector centrality.

However, if compared with degree centrality, which showed a well-connected network of the majority of languages, in figure 3.7 we can clearly see that most languages are not in proximity to many highly connected neighbours. As figure 3.7 shows, central and mainly Indo-European languages have a large node size (representing the eigenvalue) while peripheral languages in the middle and right lower part of the graph have lower values. More than half of all languages have an eigenvalue of less than 0.1, meaning that they form relations with few connected neighbours. This would support Kilduff's definition of core-periphery dynamics, where periphery members do not connect to each other but mostly to central languages.[184] Again, a clear distinction between core and periphery becomes visible with the same core of central languages by degree clustered together by eigenvector as well.

Additionally, from figure 3.7 it is apparent that the semi-periphery—Turkish, Greek, Chinese, Swedish—also assumes positions of influence by being connected to the core languages. These languages therefore further link centre and periphery while also being highly connected to each other. This finding underlines the role of semi-peripheral language

---

[184] "[C]ore members connect to everyone and periphery members connect only to core members and not to other members of the periphery" (Kilduff and Brass 2010, 355-56).

in linking centre to periphery.

**Language Network
by shared authors (>100)**

Eigenvector Centrality

Figure 3.8. Network graph of eigenvector centrality for subset (>100 edge weight).

Measuring eigenvector centrality for a subset of highly connected languages with more than 100 shared authors presents a similar image, with a distinct core of central languages and a highly connected cluster of semi-peripheral languages. In figure 3.8, the same core group of languages—French, Spanish, Italian, Dutch, English, Czech, Polish, and Hungarian—as in the complete network have comparable node sizes by their eigenvalues (1-0.60), while a second group in the semi-periphery, consisting of mostly Slavic and Nordic languages, have a lower eigenvalue. A third group of peripheral languages with an eigenvalue under 0.3 appear to have shifted the position of Chinese, Rumanian, Slovenian, and Catalan into the periphery as compared to their degree centrality score. Among most connected languages, a set of central languages, led again by French, are the most influential in connecting with the periphery, while the semi-peripheral languages appear to form a homogenous, strongly connected group.

As the differences in eigenvector values of centre, semi-periphery, and periphery show, the type of roles languages take in forming connections through authors depends on their position in the network, which again is shaped by the interrelations between languages. In the previous paragraphs, I present the argument that central languages shape the network by forming links with the periphery. Comparing degree and eigenvector centrality shows that the type of role each language occupies in the network changes according to the applied centrality measure and reveals the importance of specific connections as well as intersections in the translational flow.

So far, this section focuses on the roles of highly connected languages. As the results on eigenvector centrality suggest, there are specific languages that are positioned between centre and periphery or on the borders of the network. What happens on and in between the fringes of the network and how it makes canonical authors visible as bridges between languages will therefore be discussed in the following paragraphs.

A common method to measure which roles actors have in mediating between core and periphery, "by connecting actors who have no direct connections" (Kilduff and Brass 2010, 355-56) in a social network, is betweenness centrality. Betweenness describes the role of an intermediary actor in connecting with the peripheral, more isolated parts of the network by measuring "the extent to which a vertex lies on paths between other vertices" (Newman 2010, 185). In social network theory, betweenness is also described as brokerage, or when an actor occupies a "position lying between other actors who can only reach each other by way of the actor between them" (Risku et al. 2016, 12).[185]

---

[185] Risku et al. (2016) apply the definition of betweenness as a typology for the translation market. In their study, the client and technical writer both occupy a broker-like positions. Their study, however, does not measure betweenness for all actors and limits it to a conceptual framework for the translation network.
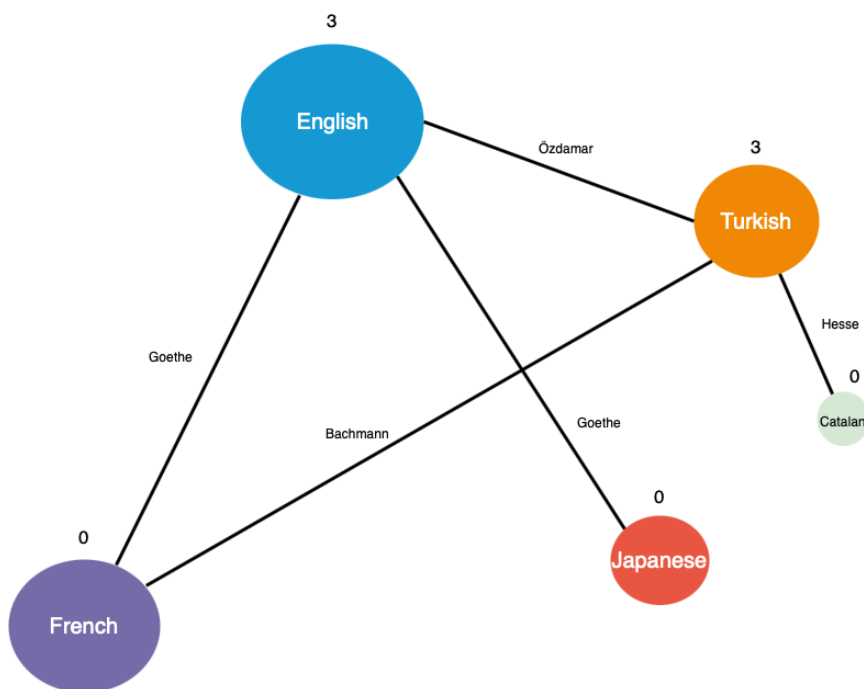
Figure 3.9. Example network graph with betweenness centrality score indicated on top of the nodes.

Coming back to our example network, on figure 3.9 we can see that only English and Turkish lie on the paths to other disconnected nodes (Catalan and Japanese) and hence have a betweenness score of 3, signifying the number of paths Catalan needs to cross to visit all other connecting nodes. For the language network, we focus on which languages are vehicular in connecting centre to periphery to identify which authors they are connected by, further revealing which authors reach the fringes of the network.

Hence, for each language in the network, I calculate the betweenness score,[186] revealing the exact opposite of degree centrality. According to betweenness centrality, Singhalese has the highest score (359), followed by Malayalam (341), Bengalese (199), Norwegian (160), Irish (158), and Occitan (126),[187] which are positioned on short or boundary paths acting as connectors of the periphery.[188] As expected, central languages like English, Dutch, Italian, and French have betweenness scores between 30 and 20, much lower than peripheral languages.[189] The languages with a high betweenness score, positioned on the fringes of the network, then can be investigated to identify authors with a far reach.

A closer look at which authors peripheral languages are connected by reveals that the ties consist of a few highly canonical or translational authors. It is notable that the number of

---

[186] Using the betweenness() function in the R package "sna" (see Butts 2020).

[187] For the complete table of betweenness values see the appendix: alldnb_lang_lang_author_betweenness_ch2.2_v2.csv

[188] "Conceptually, high-betweenness vertices lie on a large number of non-redundant shortest paths between other vertices; they can thus be thought of as 'bridges' or 'boundary spanners'" (Butts 2020, 14).

[189] See appendix: alldnb_lang_lang_author_betweenness_ch2.2_v2

shared authors per tie never exceeds 1, which in all cases represents one author. Singhalese for instance, only connects to other languages through Patrick Süskind, which goes in line with the findings in chapter 2 and Süskind's translational function in bridging a wide range of languages in the field.

Likewise, the second highly ranked language, Malayalam, has a total of 48 edges (connections to other languages) for only Hermann Hesse, similar to Bengalese, which has 38 edges for only Bertolt Brecht. Norwegian is also highly ranked with 38 as well and just one author: Robert Musil. Irish shares two authors with its connected languages: Erich Fried and Max Frisch. Among the authors that Occitan shares are Wilhelm Busch and Joseph Roth. With 56 edges, Frisian shares a maximum of four authors (with central languages like French), including Patrick Süskind, Wilhelm Busch, Elias Canetti, and Wolfgang Borchert and 25 of its edges only by one author with other peripheral languages (Malayalam, Yiddish, Hindu, etc.). When going further down the languages ranked by betweenness centrality, it becomes more and more apparent that what links central to peripheral languages is a set of canonical authors. The fewer canonical authors a language has, the higher is their betweenness, which shows how strong these authors are in establishing connections between the centre and periphery. As these examples illustrate, the literary canon in translation becomes visible in peripheral languages with a high betweenness centrality and consists of connecting actors between different parts of the network. Hence, canonicity is not only visible in the concentration of specific authors in the library canon, but also in the roles and positions languages occupy in the network of transfer. In other words, the further we go from the centre, the more the canonical connections become visible.

While betweenness can give an indication about the language that connects short paths, we can also go even further to the boundaries of the network and examine the languages that are positioned on the very fringe, only connected by few edges. Therefore, I measure the constraint of the network to estimate which languages bridge the most disconnected parts of the network (also called structural holes).[190] The combined values give an indication of the roles and types of connections at the very periphery.

For the language-author network, the languages with a low constraint overlap mostly with the languages with the highest betweenness score, meaning that they share authors that connect centre and periphery. Malayalam with 0.089 and Singhalese with 0.091, followed by languages with comparably lower betweenness centrality such as Latin, Malay, Urdu, and

---

[190] See Burt 2004.

Kurdish, therefore assume connecting roles at the boundaries of the network, sharing authors with languages that would otherwise become isolates. Constraint therefore confirms the findings of betweenness centrality, highlighting the positions that connect the fringes of the network to the centre.

Conversely, high constraint is an indicator of languages positioned at the very fringes of the network. Tigrinya, for example, only connects to one other language and has therefore the highest value when measuring constraint (1) and is only connected to Persian by the common author Marianne Herzog and her children's book *Lily, Ben und Omid* which does not appear in any other languages.[191] If Tigrinya did not sustain this one connection, it would become an isolate,[192] which is the language's constraint with its neighbour. Swiss German has the second highest constraint (0.47), which shares only one author (Rolf Sperling) with Bulgarian, Dutch, and Low German.[193] It would not become an isolate if it did not share the connection with Dutch, for example, which again emphasizes the dependency of peripheral languages on central ones when only one author connects them to the network.

Together with the findings from measuring betweenness centrality, this indicates that a set of central languages is strongly linked to peripheral languages by few but important connections. These important connections represent a set of translated authors and the extent to which centre and periphery are connected by a canon of a few overrepresented authors in the library catalogue. Where, then, is the canon of authors in translation in the network most visible, and what type of inter-relation between language positions is it following?

When further examining the authors by which they are connected, a set of canonical authors, such as Hesse, alongside contemporary authors, such as Fabian Williges, arises. This latter author of theological texts, poems, and songs has been translated into central languages including French, English, Italian, and Dutch, but also into Vietnamese, Azeri, and Armenian. Hence, the position and type of role each language occupies has an effect not only on the transfer of translated authors, but also which authors they are connected by and where the connection occurs in the network. The lower the constraint combined with a high betweenness centrality, the more likely it is for a language to connect different parts of the

---

[191] In the current catalogue, there might be more titles, since they might have been added after the date I extracted the dataset. This applies especially to books published in the same or the previous year from the extraction date.

[192] For the complete table of constraint values see the appendix: alldnb_lang_lang_author_constraint_ch2.2_v2.csv

[193] After Kazakh, multilingual, Kyrgyz, Swahili, and Maltese, central languages like English, Dutch, French, and Italian follow. Therefore, some peripheral languages appear to have a similarly low constraint to some very central ones (although they differ in betweenness).

network by few canonical authors. In other words, highly canonical authors appear to be the main connector between periphery and centre.

**Language Network
by shared authors (>100)**
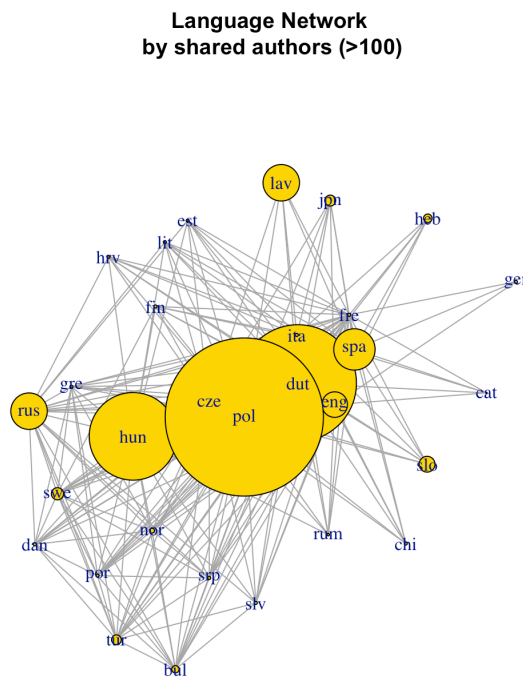
Betweenness Centrality

Figure 3.10. Network graph of betweenness centrality for subset (>100 edge weight)

Do these positions shift if we zoom in on a subset of the most connected languages? By modeling betweenness centrality for the subset composed of languages with more than 100 shared authors or edge weight (see figure 3.10), we can further observe the connecting role of semi-peripheral languages (Polish, Dutch, Hungarian, Spanish, Latvian, Russian, and English). Slavic and Nordic languages appear to occupy especially important roles alongside English, which goes against the argument of English's broker-like position. In this network, French has a betweenness score of 0 and, therefore, no role in connecting distant parts of the network. Again, an inverse of the degree centrality network graph becomes visible which further supports the role of semi-peripheral languages as brokers for the transfer of authors.

In order to identify which authors connect the languages with high betweenness scores, I then calculate the correlation coefficient between those languages. This tests which languages have connections through a similar set of authors, with the expectation that canonical authors bridge centre to periphery. For instance, the highest-ranking language in

betweenness centrality, Polish, correlates mostly to other peripheral languages, such as Turkish, Croatian, Norwegian, Greek, and Russian.[194] Looking at which authors Polish shares with the most highly correlated languages again provides a list of mostly canonical authors[195] such as Kafka, Hesse, Goethe, and Mann. However, the authors for which Polish has the most translations are not represented across the most correlated languages. For instance, the dime novel author Hedwig Courths-Mahler[196] links Polish to Bulgarian, Czech, and Hungarian, but not English or French; the contemporary author Charlotte Link connects Polish with central languages such as French and Italian, but also with Latvian, Lithuanian, and Korean. For another dime novel author, Patricia Vandenberg, Polish only shares translated titles with Czech. Based on the case of Polish, it is apparent that canonical authors mostly connect the centre to periphery, with dime novel authors also playing an important role.[197]

What does betweenness centrality, in addition to degree and eigenvector, then imply for canonicity and the place of canonical authors in the language network of the library catalogue? I raised the claim that highly canonical authors represent the connections between central and peripheral languages, becoming more visible toward the fringes of the network. The findings of the previous sections imply that only few highly canonical authors reach the periphery of the translation network. Hence, I argue that if an author is among the smallest languages but at the same time representing a bridge to the central languages, they are likely to be highly canonical. In other words, the further toward the boundaries of the network we move, the more visible the national canon in translation gets.

Network communities of central and peripheral languages

In the first section of this chapter, I claim that a set of core languages occupying central positions share a set of canonical authors and are actively shaping the transfer of translations. However, looking at the peripheral network, reveals a set of authors that only connect peripheral languages. Additionally, as shown in section two, while certain authors connect the different clusters and represent bridges between core and periphery, other authors may exclusively appear in central or peripheral languages without connecting the two. For instance, when comparing authors of Hungarian and English, both languages share 2850 authors, while English has 894 other authors not shared with Hungarian and Hungarian has

---

[194] See the appendix for correlation coefficients: alldnb_lang_lang_author_cor_ch2.2_v2.csv
[195] See appendix for the full list of ranked authors per highest correlated language with Polish: alldnb_lang_lang_authortitlesum_polish_ch2.2_v2.csv
[196] Ernestine Friederike Elisabeth
[197] A pseudonym for Elke Zumbusch-Stieber and Gery Schiede.

244 additional authors that English does not have. They share the same amount of ties for 49 authors: Daniel Kehlmann (9 ties), Ingo Schulze (6), Robert Menasse (4), Martin Walser (4), Gaby Hauptmann (3), Daniel Glattauer (3), Hans Hellmut Kirst (2), Wilhelm Busch (2), and Martin Buber (2), plus 40 more authors with a single tie. Moreover, there are 2,801 authors for which both languages have zero ties. When comparing which authors Hungarian does not share with English, though, we get an interesting result: Sandrine Jopaire, the author functioning with the most ties in English, has 0 translations in Hungarian, same as Petra Durst-Benning.[198] We can observe that specific authors form bridges between the two languages, while some authors only appear in one of the two languages. Hence, we can also assume that some authors are exclusive to central or peripheral languages.

As the example of English and Hungarian illustrates, languages can be connected but also separated by authors. What is the function of authors in the author-language network as bridges between languages, and do they constitute communities of canonical or translational authors shared across or unique to language groups? Where is the border between translational and canonical? Are they clearly distinguishable in the network? In this section I further explore the function of connecting or exclusive authors by applying community detection. I follow the hypothesis that there are some authors that are regionally specific to distinct communities in the language network, with the objective to find sub-communities of languages who are brought together by authors.

Although centrality measures give indications on groups and clusters of languages based on their positions, without applying a meaningful threshold it is not possible to split groups up into centre and peripheral communities of languages. Therefore, I applied the fast-greedy optimization of modularity algorithm[199] for community detection[200] to the complete language-author network graph in order to identify language groups which I expect to reveal a clear separation of central and peripheral languages observed in chapter 2. The community detection algorithm serves as a method "to separate the network into groups of vertices that have few connections between them" (Newman 2010, 371).

---

[198] See appendix for a complete list of authors: alldnb_lang_lang_author_degree_match_hunvseng_ch2.2_v2.csv
[199] See Clauset et al. 2004.
[200] In comparison with community detection based on edge betweenness (see Newman and Girvan 2004), the fast-greedy algorithm produced the best results comparative to the results from the centrality measures. The central and peripheral language groups corresponded to the cluster in the centrality network graphs.
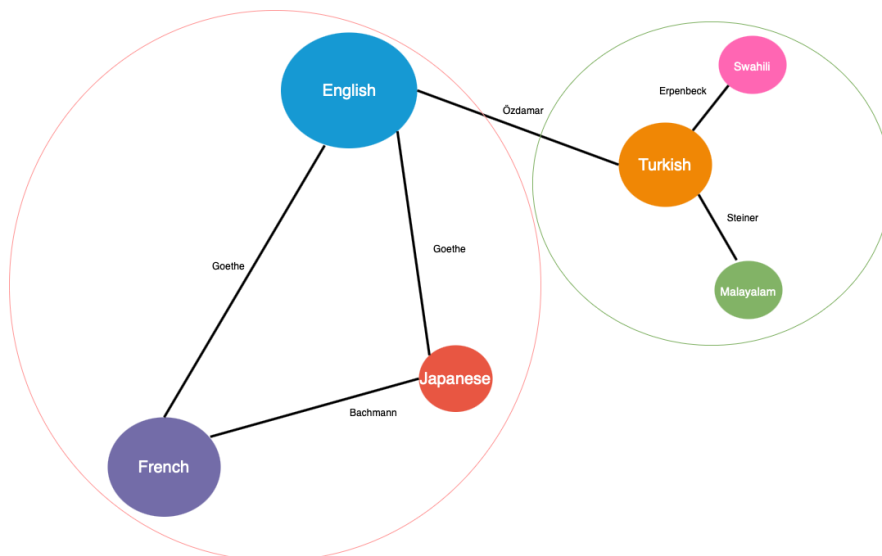
Figure 3.11. Example network graph of communities.

Consider figure 3.11 for example, which is an example graph of nodes representing languages connected by translated authors. From looking at the edges connecting English, French, and Japanese, which are all connected, and Turkish, Swahili, and Malayalam, which are only connected by the edge between English and Turkish, we can separate these two into distinct groups based on the pattern of edges they are connected by. The applied algorithm takes all connections of the network and based on modularity "produces *some* division of the vertices into communities, regardless of whether the network has any natural such division" (Newman 2004, 1).[201] Hence, as a result of running the algorithm on the language network, we can expect each language to fall into a group based on the edges and their positions in the network, meaning that central languages can be expected to be in one group and peripheral in another. This method is used to not only differentiate centre and periphery, but also to investigate which authors they are connected by and see whether they are representative of the national canon and which authors represent regionally specific groups.

---

[201] This is done by using modularity (Q), as Newman describes: "Starting with a state in which each vertex is the sole member of one of n communities, we repeatedly join communities together in pairs, choosing at each step the join that results in the greatest increase (or smallest decrease) (Q)" (Newman 2004, 2).
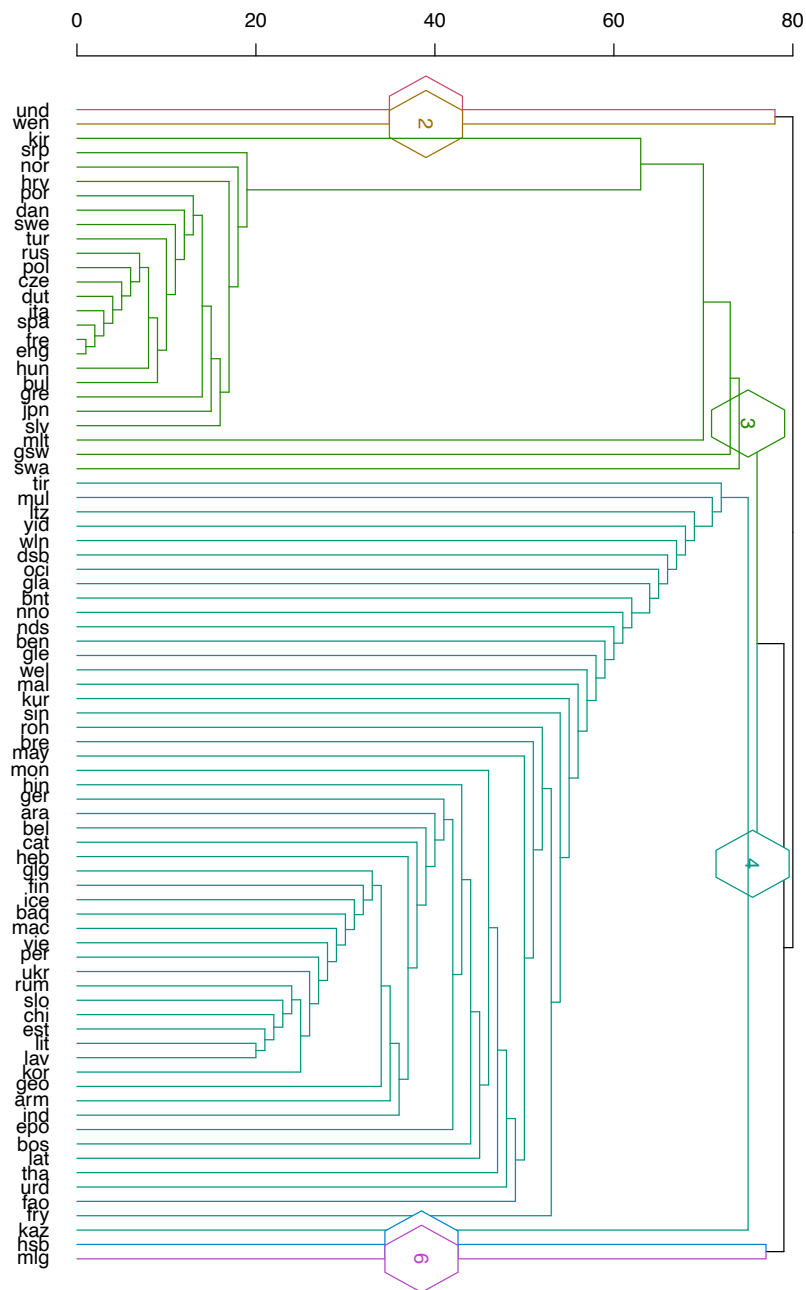
Figure 3.12. Dendrogram of communities detected by fast greedy modularity optimization

After running the fast greedy community detection algorithm on the network graph, the resulting communities (figure 3.12) are clearly split into two larger groups (label 3 and 4) of central and peripheral languages, while a total of six groups were detected,[202] four of which only consist of one language each, representing languages with only one edge such as Sorbian and Malagasy. The second, much larger peripheral group also contains subgroups of

---

[202] Modularity score 0.023.

semi-peripheral languages. To assess which authors form connections between centre and periphery, I focus on the two largest groups (3 and 4 on figure 3.12).

With this method, it is possible to categorize a set of authors which connect central and peripheral languages in the translation network, but also to detect authors that are unique to centre and periphery. To compare the network edges and determine which authors are connecting the central and peripheral language groups, I created two author-author adjacency matrices for each group and measured the intersection by the number of overlapping and therefore shared edges. Accordingly, 1,610 authors are shared between central and peripheral languages and 3,176 are not shared, meaning they are either only in the peripheral or the central language group.[203] Hence, only 33% of all authors assume a connecting role between centre and periphery in the language network, which points toward a small canon of connecting authors.

The top 50 authors according to the number of edges that peripheral and central languages[204] connect include representative authors of the German national canon such as Patrick Süskind (798 edges)—who in many cases has been identified as the sole connecting author of peripheral languages —Heinrich Böll (794), Thomas Mann (794), Johann Wolfgang von Goethe (792), Franz Kafka and Rainer Maria Rilke (788), Friedrich Dürrenmatt (784), Elfriede Jelinek (777), Hermann Hesse (775), Stefan Zweig (774), Max Frisch (771), and Günther Grass (767), followed by Ingo Schulze (762), an award-winning author of novels translated into 35 languages in the catalogue. Similarly, with slightly less edges than Robert Walser (684), we can find contemporary authors like Juli Zeh (680), followed by Jenny Erpenbeck (668) alongside the canonical author Friedrich Hölderlin (666). Hence, some contemporary authors with a certain level of prestige and an increased number of translations in recent years appear alongside canonical authors.

What these contemporary authors have in common is that they have an increased number of single titles translated into many peripheral languages. Ingo Schulze for instance has single translations into 13 different languages such as Albanian, Arabic, Estonian, and Lithuanian. Juli Zeh, in turn, has translations into 32 languages, the same number as Hölderlin, with the difference that Zeh has more single translations into peripheral languages such as Ukrainian, Hebrew, and Korean, while Hölderlin has an increased number of translations into central languages and fewer single connections. Hence, Zeh's network of

---

[203] A Pearson's product-moment correlation between the two matrices revealed a score of 0.64 indicating that the author network of central and peripheral languages is positively correlated.
[204] See appendix for table: alldnb_author_author_center_periphery_intersection_ch2.2_v2.csv

translated titles is less concentrated and more spread out, reaching both centre and periphery; based on this, she can be identified as a translational author. In the same way, Jenny Erpenbeck (the daughter of the translator Doris Kilias and prizewinner of the Independent Foreign Fiction Prize) has 12 titles translated into English, and of the 27 languages her books have been translated into according to the catalogue, 11 have single translations into Ukrainian, Rumanian, Macedonian, Slovenian, and others. While her works focus on English, she also appears to have a spread-out network including few translations into central and peripheral languages alike.

These examples illustrate that the same connecting roles as canonical authors representative of classics of German literature are also occupied by some contemporary authors with a large network spreading from central to peripheral languages. These authors correspond to a translational category in that they are translated into many languages. In other words, besides canonical authors, other translational authors function as intermediaries in the language network. This implies that canonical and translational authors constitute similar categories in how they weave the threads by which translations connect linguistic and literary communities.

Detecting communities helps to identify authors that occupy a connecting role between centre and periphery. However, the groups of authors which only appear in either the central or peripheral language group raise the question which language group includes predominantly authors that do not occupy a connecting role. These authors would supposedly not belong to a set of canonical authors, since findings in this chapter suggest that the more canonical an author, the further into the periphery their reach is. Hence, authors only shared among periphery or centre then would not constitute the representational canon of German literature in translation.

Following this claim and to address the number of authors not connecting the central and peripheral language group, I subset my dataset by authors with no connecting edges between centre and periphery and calculated the percentage of authors unique to each group. Of the 3,176 authors not shared across centre and periphery, 78.5% (2,496) of authors appear in the central language group, meaning that the periphery has few unique authors (680). There is an increased number of authors that signify the central language group, while the 21.5% of authors exclusively in the periphery cluster therefore represent an independent group of authors.

Now, we can assume that canonical authors, transferred between central peripheral languages, occupy a connecting role, while specific, possibly translational authors only occur

in smaller communities. However, where is the border between translational and canonical, and do authors form distinct communities for either of these categories? Running the fast-greedy community detection algorithm on the complete author-language network graph clusters authors into 2,508 groups (modularity score 0.34), five of which have the majority of authors, while 2,503 are single author groups.[205]

The authors with an increased number of single titles translated into many peripheral languages—Patrick Süskind, Juli Zeh, Jenny Erpenbeck, Ingo Schulze—are in group 2 with 904 other authors. Accordingly, this group appears to consist of mainly contemporary, translational authors. In contrast, group 1, with 542 authors, includes authors such as Friedrich Hölderlin and Johann Wolfgang von Goethe, so it could be representing canonical authors. However, Franz Kafka, Heinrich Böll, and Thomas Mann are also in group 2, which suggests that it includes some authors that are both highly canonical (concentrated) and translational (represented across many languages). In conclusion, while distinct communities can be identified that carry canonical and translational connotations, they appear to overlap and not be exclusive to one category. This brings us back to the similar function canonical and translational authors occupy in the language network. However, by looking at shared authors only among peripheral languages, we can see a set of authors challenging this notion. Hence, in addition to assuming a connecting role between centre and periphery, translational authors also appear to have the function to signify a certain language-community to which they are exclusive.

Coming back to the initial question of whether the function of authors in the author-language network is to form bridges between languages, by applying community detection it can be observed that languages form distinct groups based on the authors they do or do not share. This further illustrates the role of canonical and translational authors in connecting linguistic communities. My findings show that most authors are shared within the centre, while the periphery also has a portion of authors not shared with central languages, representing regional sub-communities of authors. This challenges the notion that centre and periphery are highly connected by shared authors and emphasizes the connecting function of canonical and translational authors. If we assume (as stated in chapter 2) that canonical authors are concentrated in languages and translational authors are spread out across languages, looking at authors unique to a language community underlines the importance of regional sub-communities of authors in the field of translation.

---

[205] See appendix for table: alldnb_author_membership_ch2.2_v2.csv

In conclusion, the findings of this chapter show that a network approach challenges definitions of the canon in translation and provides a framework to expand its definition to one that considers the role of translational authors. Additionally, a network model of the German National Library catalogue allows us to re-define the world literary canon as a category in terms of its connecting function in the language network, while also stressing the importance of authors that are significant to literary communities.

In the first section of this chapter, I argued that the language position and the extent to which the network is centralized are indicative of the areas of translational transfer. In section two, I found that central and peripheral languages, especially those located at the fringes of the network, are connected by few significant authors. Accordingly, I raised the claim that authors connecting the centre to periphery and supporting reach to marginal languages can be identified as belonging to a set of representatives of the national canon. Hence, canonical authors have been described by their function of forming bridges between distinct parts of the network. Finally, in the last section, by means of community detection, the authors connecting centre to periphery and supporting reach to marginal languages could be identified as belonging to a set of canonical classics and forming bridges between distinct parts of the network. Community detection set apart another group of contemporary authors that permeate the network in specific literary communities and account for their translationalism. These arguments and related findings show that the relationship of a translation to its target languages constitutes a core feature that sets translational literature apart from what we would formerly simply categorize as canonical or classical. Mapping the canonical space in the library catalogue as a network makes visible the non-canonical spaces or literary transfer.

Hence, a social network model is a fruitful approach for visualizing the hidden patterns in the transfer of translations between languages as represented by the German National Library catalogue. Adapting Heilbron's model made it possible to identify the relationship of centre versus periphery and positioning languages in the network. Precisely because this approach is centred on a language-author network instead of being limited to the social aspects of the network, as has been the focus in prior studies, dynamics of transfer have been revealed that have formerly not been accounted for. Further, the definition of the world literary canon could be extended by attributing it to the connecting function of authors between centre and periphery. Translationalism hence bridges and at the same time distinguishes linguistic and literary communities.

4. Mapping publishing centres of translations and author communities in the DNB

In the fourth and final chapter of this thesis, a last important dimension of translations is modelled: the geographic representation of translations by publishing place in the German National Library catalogue. While social network analysis is an approach to visualize the connections and communities within the library catalogue by languages, it does not reflect the spatial relationships between authors and publishing places. Thus, this chapter is dedicated to mapping the geography of the translation network and further investigates how translational works and authors challenge our understandings of the national and the world literary canons of German fiction, as well as explores the role of the national collection in shaping German fiction in translation beyond its national boundaries.

Scholars of translation studies have increasingly raised awareness for the necessity of a combined study of translations and geographic space. Susan Bassnett has stressed that "any study of translation necessarily involves a geographic dimension" (Bassnett 2014, 235) and calls the interconnectedness of literary translation studies and geography "radical shifts of perception" (ibid.) in which "methodologies from translation, literature, and geography can assist one another" (Bassnett 2014, 241). The geographic "shift," or "new perspective" within translation studies as Federico Italiano calls it, has put forth questions of "'where' translation happens" (Italiano 2016 ,4) and "[H]ow do books travel? How do they become vessels of stories and migrate from one geographical area to another?" (Mani 2017, 1). This shift in translation studies—the call for a "geography of translation" (Italiano 2016 ,4)—and its methodology has been echoed especially by scholars interested in world literature. Mani trails the focus of analysis of world literature back to geography, an approach echoed by Laachir et al. who argue that "we need a richer spatial imagination of the 'world'" (Laachir et al. 2018, 290) to re-evaluate world literature by visualizing "the movement of literary texts in space and time" (ibid.). The shift toward a geographic study of translation thus has been viewed as a framework to re-evaluate the world literature paradigm, as well as the monolingual categories of national literatures.

For Rebecca Walkowitz, geography is an intrinsic dimension of born-translated literature. She argues that it therefore pushes against the idea of a single origin of a national geography and language of literature and calls for a decentralized study of translation that is not limited to one geography.

> [L]iterary studies will have to examine the global writing of books, in addition to their classification, design, publication, translation, anthologizing, and reception across multiple geographies. (Walkowitz 2006, 528)

Walkowitz proposes a pluralistic, decentralized approach to translation in geographic space as opposed to the national category of the literary canon. In *Born Translated,* she picks up this point, formulating another core feature of novels that set themselves apart by their geographic dimensions:

> Once literary works begin in several languages and several places, they no longer conform to the logic of national representation. Many born-translated novels signal this departure by blocking languages, invoking multiple scales of geography, and decoupling birthplace from collectivity. (Walkowitz 2017, 30)

Here, for Walkowitz, born-translated literature does not only break with common categories used to describe the canon, but also challenges the way we think about literature spatially. We can take Walkowitz's definition of born-translated fiction as a feature of translationalism and formulate a similar hypothesis for translational literature in that besides being translated into many languages, translational literature also connects many places and therefore challenges national categorization.[206] In this light, the representative national canon of German literature in translation can also be made visible in terms of its geographic dimensions by tracing how literary works or authors transgress its national and linguistic boundaries. This follows the hypothesis that canonical and translational authors do not only bridge linguistic communities, but also geographic places, and their network is inscribed into a geography of literary exchange.

Besides Walkowitz's concept of born-translated literature, the focus on geographic dimensions has encouraged spatial concepts in translation studies such as translation sites or "zones" (see Simon 2018; Cronin 2003; Apter 2006), transfer (Kershaw and Saldanha 2013; Pym 2010), and third space (see House 2021; Kershaw and Saldanha 2013), which also aim at challenging the binary systems of classification (e.g. in world literature of national versus world) and point toward what has been left out in translation studies by focusing on the national canon. In line with Walkowitz's concept of multiple geographies and languages of born-translated literature, Kershaw and Saldanha argue that "the geography of translation posits translational activity as a natural and political force in the transformation of landscapes; the geography of the landscape posits meaningful spaces which do not respect the

---

[206] "Refusing to match language and geography, many contemporary works will seem to occupy more than one place, to be produced in more than one language, or to address multiple audiences at the same time. They build translation into their form" (Walkowitz 2017, 6).

delimitations national and linguistic of borders" (Kershaw and Saldanha 2013, 137). The idea that translations transgress the categories of national—also in terms of language—and generate "geographies that are significant to the actors rather than generic meta-categories such as 'world' or 'global'" (Kershaw and Saldanha 2013, 297) has expanded the conventional theoretical and methodological space of translation studies to develop new models that include the geographic dimension. Through the concepts of third space as well as born translated, we can observe the continuation of the spatial turn in literary studies and the sociological focus on fields, networks, and movements which serve to challenge the classification systems formerly dominant in literary studies—national versus world—by introducing new models of analysis. This study is an attempt to tie these theoretical and methodological shifts together in order to design a model to put translations on the map and build on their main arguments.

By applying concepts like the zone, transfer, third space, and born translated, scholars of translation studies have put forth the core argument of an unequal, uneven distribution of translations in geographic space with cultural centres. In line with Heilbron's centralization model (core-periphery) and Homi Bhabha's concept of cultural centres and metropolitanism in the location of culture,[207] scholars such as Walkowitz, Bachleitner and Wolf, and Reynolds and Vitali have put forth the argument that translations are centralized, but unevenly distributed within the "literary systems of world circulations"[208] because of the role of cultural publishing centres.[209] Bachleitner and Wolf, in line with Heilbron, argue that what makes a language central is not solely related to its spread in terms of speakers, but more so its symbolic capital which is accumulated as a result of its role as a metropolitan, cultural centre, and its prestige of authors.[210] Reynolds and Vitali, in line with translation studies arguments on cultural centres, argue in favour of an unequal distribution which they use to challenge the notion of one canon, saying that the "uneven distribution of translations complicates both, the

---

[207] "[T]ranslation studies are borrowing mainly from Homi Bhabha's understanding of hybridity, led to the development of the metaphor of translators inhabiting an in-between space that is outside both the source and target cultures but where those cultures meet and interact in conflicting but creative ways" (Kershaw and Saldanha 2013, 139).

[208] "Books are no longer imagined to exist in a single literary system but may exist, now and in the future, in several literary systems, through various and uneven practices of world circulation" (Cited in Mani 2017, 38 from Walkowitz 2006 "The Location of Literature: The Transnational Book and the Migrant Writer").

[209] Franco Moretti, for instance, looks at distribution markets and translations in library circulation catalogues, arguing that the novel is a centralized genre; not only does it mostly cover the centre as a space (the city) but also it is distributed from few centers in large quantities (see Moretti, 1998).

[210] "Aus der Sicht der Hierarchien des literarischen Feldes sind New York oder London zwar Zentren, die Konsekration durch Übersetzung verleihen können, aber nicht die einzigen Zentren. Daneben sind Paris, das Casanova kurzerhand zur literarischen Welthauptstadt erklärt, Barcelona, Lissabon/Sãσo Paulo oder Berlin zu nennen – Städte, die in Geschichte und/oder Gegenwart ihren Sprachraum dominieren. Das globale literarische Feld ist somit polyzentrisch strukturiert" (Bachleitner and Wolf 2004, 3).

universality and the canonicity of what we have hitherto been calling a 'global and 'canonical' novel" (Reynolds and Vitali 2021, 8). In Reynolds and Vitali's words, we can hear an echo of Walkowitz's critique of a national monolingual and mono spatial definition of the literary canon and its limitations for the analysis of translations.

Another core argument which was born out of the cultural centres concept is the role of regional sub-communities of authors in specific countries (peripheral languages, the migrant canon) or what Apter calls a "translational transnationalism (a term I use to emphasize translation among small nations or minority language communities)" (Apter 2006, 4). Bachleitner and Wolf hint at the significance of sub-communities within a linguistic field regarding the Austrian translation publishing market, which they argue is susceptible to the import of peripheral literature due to its peripheral position and close geographic proximity to other peripheral publishing markets.[211] Reynolds and Vitali also apply Heilbron's centrality theory to the geographic space of translation, referring to core-periphery, but they also show how their model's outliers drew their attention to channels, connections, and interventions and "complicating the idea that translations radiate from the source text at the centre to ever more peripheral contexts of reception" (Reynolds and Vitali 2021, 9). They do this by using a time series analysis to show the importance of publishing centres in a specific point in time, which reveals a variety of centres outside of Europe.[212] In other words, they observe similar dynamics in the geographic translation space to my findings for the language space of a centre-periphery in the previous chapter, which also takes us back to my discussion on the importance of sub-communities and their role in the field of translation. These observations further open up the question of whether the uneven distribution or concentration in specific languages is also the case for cultural centres in geographic space, and how the linguistic network of literary communities is imprinted on the map of publishing centres.

In response to the geographic shift in translation studies, the aim of this chapter is to map translations and project the field and network model onto the geographic space in order to challenge the mono-spatial and monolingual notion of the national German literature paradigm in the catalogue. In continuation with the geographic turn in translation studies outlined above, geographic maps serve as an explorative tool to test the hypothesis of an

---

[211] "Dies wirft die Frage auf, ob Österreich, das in der deutschsprachigen Verlagslandschaft eine periphere Position einnimmt, prädestiniert ist für den Import einer ihrerseits peripheren Literatur, die geographisch nahe liegt und zu der historische Beziehungen bestehen" (Bachleitner and Wolf 2004, 21).
[212] "Each translation of Jane Eyre is not a passive act of reception but an active engagement; and each culture where this happens is not merely peripheral to a centre that lies elsewhere but remains in many respects a centre to itself" (Reynolds and Vitali 2021, 10).

unequal distribution across publishing centres and investigate the role of peripheral and sub-communities in connecting geographic areas and creating pathways for the global circulation of German fiction in translation. Additionally, in line with Walkowitz, I also test the hypothesis that translational authors invoke multiple geographies, bridging places in the circulation network by means of language. Hence, in this chapter I present a model for mapping translation data from the German National Library to contribute toward a comparative literary geography discourse in translation studies, German literary studies, comparative literature, and digital humanities by bringing visibility to the spatial imagination of translations.

Mapping translations and the use of the map as an exploratory tool for investigating centralization, networks, metropolitanism, and author communities have only had marginal and very recent attention. For German literature in translation within a library catalogue that is not restricted to a certain author, period or genre, mapping has hardly found any application to date. However, in digital humanities, mapping is more and more applied as a common approach to model metadata of translations (Cheesman 2016,[213] Reynolds and Vitali 2021, Zhai et al. 2020).[214] One of the most recent applications of maps to model translation by publishing places is by Reynolds and Vitali of a single novel. In line with the concept of metropolitanism and cultural centres, they also argue that "mapping translations by cities reveals a world in which they—in line with the publishing industry in general—are variably concentrated or dispersed" (Reynolds and Vitali 2021, 6), further stressing that translations tend to be concentrated in a city as compared to a language or state. They also point out that maps reveal how the canon is challenged by looking at the geographic distribution of translations. Arguing that translations contribute to canonization while not being canonical themselves, they put forth the concept of "acts of translation": "as soon as you start making a map, what strikes you most about translation is that it regenerates the text in a new location," "emerging from place after place, city after city" (Reynolds and Vitali 2021, 5). By focusing on places and cities which acts of translation emanate from and are concentrated in, they re-evaluate the category of a national canon—one source text and language—echoing Walkowitz's decentralized model of born-translated literature.

Not only do Reynolds and Vitali's findings correlate with Walkowitz, Bassnett, Heilbron, and Bachleitner and Wolf, but they also use maps of publishing places as an

---

[213] See also Cheesman, Tom. "Othellomap." Accessed June 30, 2022. https://othellomap.nand.io.

[214] Core points in translation studies were however focused on places OF translation, "translation zones," and transnationalism (Italiano, Simon, Cronin, Sapiro), rather than mapping translation flows in geographic space.

explorative tool and model for analysis. Building on Reynolds and Vitali's study, in this chapter, maps are both a visualization of the language distribution and concentration and an explorative tool to investigate the relationships between publishing centres, centre-periphery dynamics, and sub-communities of authors. Additionally, geographic mapping is used to further test the findings of previous chapters on concentration (chapter 2), network centralization (centre-periphery) of authors, and sub-communities (chapter 3).

The unit of analysis in this chapter for the geographic mapping of authors and titles in the catalogue is the publication place extracted from the "publisher" column in the dataset.[215] Besides the name of the publisher, this column also includes at times multiple cities associated with the publisher. According to the glossary of Library & Information Science, the place of publication is "the place associated with the publication, release, or issuing of a resource or document."[216]

Link zu diesem Datensatz	https://d-nb.info/1224616944
Art des Inhalts	Fiktionale Darstellung
Titel	Anyone who utters a consoling word is a traitor : 48 stories for Fritz Bauer / Alexander Kluge ; in collaboration with Thomas Combrink ; translated by Alta L. Price
Person(en)	Kluge, Alexander (Verfasser) Price, Alta L. (Übersetzer)
Werk(e)	Wer ein Wort des Trostes spricht, ist ein Verräter
Verlag	London ; New York ; Calcutta : Seagull Books
Zeitliche Einordnung	Erscheinungsdatum: 2020
Umfang/Format	123 Seiten : Illustrationen ; 21 cm
ISBN/Einband/Preis	978-0-8574-2-782-3 (falsch) Festeinband 978-0857427823 (falsch)
Sprache(n)	Englisch (eng), Originalsprache(n): Deutsch (ger)
Sachgruppe(n)	830 Deutsche Literatur ; B Belletristik
Literarische Gattung	Erzählende Literatur: Gegenwartsliteratur ab 1945
Weiterführende Informationen	Inhaltsverzeichnis

Leipzig	Signatur: 2021 AA 6767 Bereitstellung in Leipzig

Figure 4.1. Screenshot of the catalogue entry of one of the titles in the dataset showing the cataloguing fields containing geographic names of publishing places.[217]

---

[215] The publishing place is used as a signifier of the geographic localization of a cultural product (a translation). It represents the target language's distribution in geographic space since publishing place largely correlates with language and therefore makes visible dynamics that have not been apparent in the language model of chapter 2 and 3. One might object that publishing place is closely linked to publishers and therefore also visualizes publisher distribution. This is certainly one dimension which is not analyzed in this chapter and needs to be kept in mind when looking at the data. However, in this chapter, publishing place is analyzed as a dimension of language distribution and therefore limited to this focus.

[216] Librarianship Studies & Information Technology. "Place of Publication," July 3, 2019. https://www.librarianshipstudies.com/2016/02/place-of-publication.html.

[217] https://d-nb.info/1224616944

For instance, the translation in figure 4.1 published by Seagull has three distinct publishing places (London, New York, Calcutta) listed in the field "Verlag" (publisher). The last represents the place where the publisher was first founded, while London and New York represent divisions of where the publisher is currently located or distributing from. For Seagull, the order and number of publishing places vary in the catalogue. Other titles for instance only include London or the locations appear in a different order (Calcutta; London; New York).[218] I limit my dataset for mapping to the first publication place, since it represents the place preferred by cataloguing practices[219] and supposedly the most recent place the publisher had been active in.[220] Additionally, the majority of titles have only one publishing place and only 3.5% of all titles have more than one publishing place. After extracting the first publishing place for each title[221] and moving it into a new column in the data frame, I geocoded each location with the Nominatim API and tidygeocoder package[222] in RStudio. Afterwards, the mapping was done with the Leaflet package.[223]

Place name ambiguity—the differences in an orthography of a place name, the language, or metonymy—is a common obstacle when working with geodata and place names (see Overell 2011, Gritta et al. 2018), hence working with the data on publishing places extracted from the DNB catalogue entailed a number of additional steps to standardize and clean the place names before passing it on to the API to ensure that coordinates could be assigned. Since geocoding works with an API to assign coordinates to a given string (the place name), especially for a multilingual dataset which includes toponyms in various names (e.g., Beograd, Belgrade, Belgrad), I had to group and verify if they have the same coordinates. This was done by fuzzy matching, an algorithm that works with a distance measure to group strings based on their similarity.[224] Based on the toponyms with the closest

---

[218] See: https://d-nb.info/1179247434

[219] It is common according to MLA and other citation styles to only cite the first place of publication.

[220] Suhrkamp's publishing place, for instance, has been catalogued as Frankfurt am Main prior to their relocation to Berlin in 2010 in the German National Library catalogue.

[221] Using ";" as a separator. From here on "publication place" is synonymous to first publication place.

[222] Cambon, Jesse, Diego Hernangómez, Christopher Belanger, and Daniel Possenriede. "Tidygeocoder: An R Package for Geocoding." *Journal of Open Source Software* 6, no. 65 (September 9, 2021): 3544. https://doi.org/10.21105/joss.03544.

[223] Cheng, Joe, Bhaskar Karambelkar, and Yihui Xie. *Leaflet: Create Interactive Web Maps with the JavaScript "Leaflet" Library*, 2021. https://CRAN.R-project.org/package=leaflet.

[224] Fuzzymatching was done in R with the package "fuzzyjoin" and the stringdist_left_join() function using the Levenshtein distance measure. I then filtered out the groups for which coordinates differed and manually grouped them by the largest toponym (e.g., Beograd to Belgrade, L viv to L'viv, Sofia to Sofija, Tihrān to Tehran, and Tōkyōto to Tōkyō).

distance (1)[225] I manually re-grouped place names and reassigned the coordinates. Additionally, I verified and re-grouped all places with more than five titles for which the API had not assigned coordinates (such as Yerûšālayim, Or Yehudah, Kūlūniyā, and Ere w an).[226] As a final cleaning step in order to avoid multiple place names, I extracted and verified all place names with the same coordinates to then group them together.[227] After these additional steps and manual verification on the map, it was possible to rule out most wrongly placed or represented publishing locations.

In order to layer highly concentrated to lesser concentrated publishing places, I calculate title counts and subset them in increments (see first section). First, I map the concentration of titles and authors in geographic space by their publishing place to identify and visualize cultural centres from where translations are circulated. By superimposing title counts as layers on the map of publication places for highest to least concentrated publishing places, I visualize central versus peripheral publishing places. In line with my findings in chapters 2 and 3, I expect titles to be highly concentrated in European cities. Secondly, I address the hypothesis that the network of centre and periphery observed in chapter 3 is also visible in the geographic map of publishing places. I expect to see a centralized, star-like shape between the central and peripheral publishing places with few smaller clusters in the periphery. Third, I use the subset of authors that are unique to either the central or the peripheral languages from chapter 3 to explore the notion of sub-communities that form channels of authorial circulation between publishing centres. I then compare the author's title count to their publishing place count to identify for which authors the title or place count is higher than expected. I expect to find a set of authors specific to a language group which are either concentrated in title count or geographically distributed. I then further examine their relationship to the publishing places in the context of their translated titles.

Mapping publishing centres by title concentration

In order to map the overall concentration of translations of German fiction in the DNB catalogue in publishing centres (figure 4.1), I calculate the frequency of titles per place of publication (n=1850 places) and size each circle marker on the map accordingly.

[225] See appendix for table: 20220604_alldnb_ch3_pubplace_stringdist_toponyms.csv
[226] A total of 95 places had no coordinates assigned to them, only four of them had more than five titles. See appendix for table: 20220608_alldnb_ch3_author_pubplace_geo_clean_nocoordinates_v2.csv
Additionally for 2,468 titles the publication place is missing in the data.
[227] See appendix for a list of duplicate coordinates:
20220613_alldnb_ch3_pubplace_duplicategeodata_toponyms_cleaning_v2.csv

Additionally, I create three layers for subsets of the data since, as chapter 2 has shown, the title distribution is highly skewed and concentrated in specific languages, which also was the case for the publication place distribution,[228] with most publishing places having less than 50 titles. This allows for exploring the map from most to least concentrated publication places and testing the hypothesis that the geography of translated German fiction displays an uneven distribution with few cultural centres located in European countries.
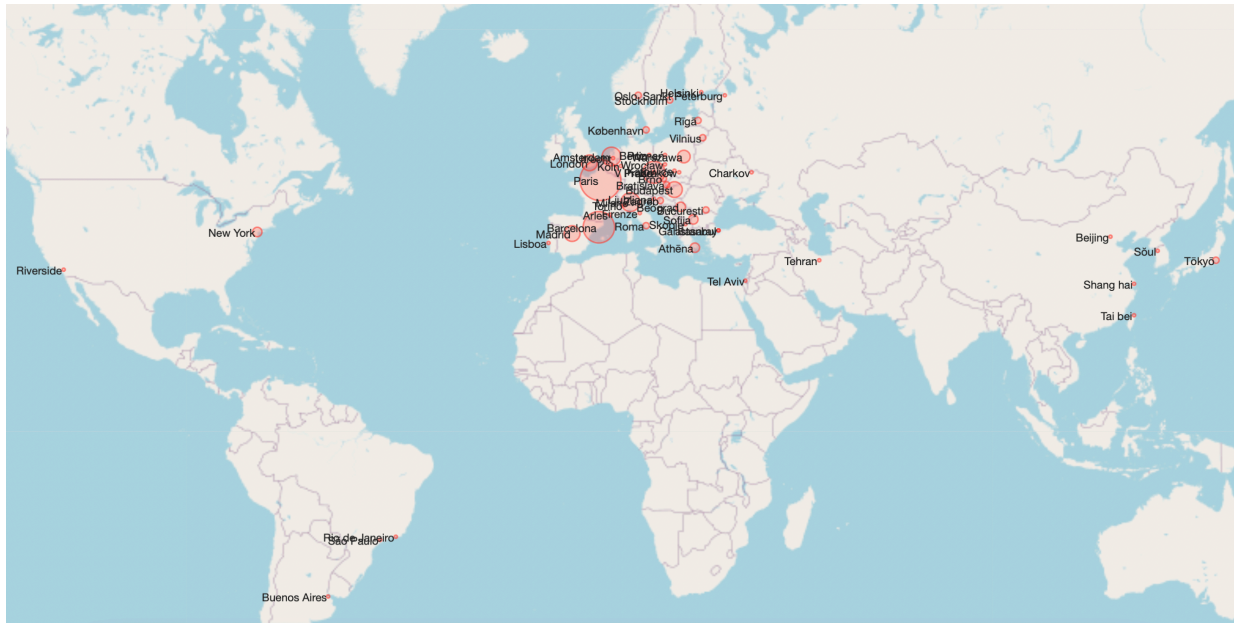
Figure 4.2. Title counts per first publication place with layers of >100 titles. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. Marker sizes correspond to title count. Places with more than 100 titles are marked in red.[229]

As figure 4.2 shows, at a high concentration with 100 and more titles, few large centres become visible, mainly in Western Europe with smaller centres in northern and medium large in south-east and smaller centres in northern South America and Asia. Out of 54 places with more than 100 titles, only 14 are outside of Europe such as New York (558 titles), Tōkyō (446 titles), Sankt Peterburg (277 titles), İstanbul (267), São Paulo (170), Tel Aviv (163), Tehran (133), Tai Bei (127), Rio de Janeiro (125 titles), Riverside (122), Beijing (119), Galatasaray (115), Buenos Aires (114), and Shang hai (105). With only 54 unique places in the subset of places with more than 100 titles, accounting for 2.6% of all places in the dataset, we can observe a high concentration in few metropolitan centres. The Eurocentric focus of the catalogue observed in chapter 2 and 3 is also apparent on the geographic map of concentrated publishing places.

---

[228] See appendix for density graph: 20220607_alldnb_ch3_author_pub_place_title_density_v2
[229] The map is an interactive html file, which cannot be attached here. I therefore included screenshots for each layer.
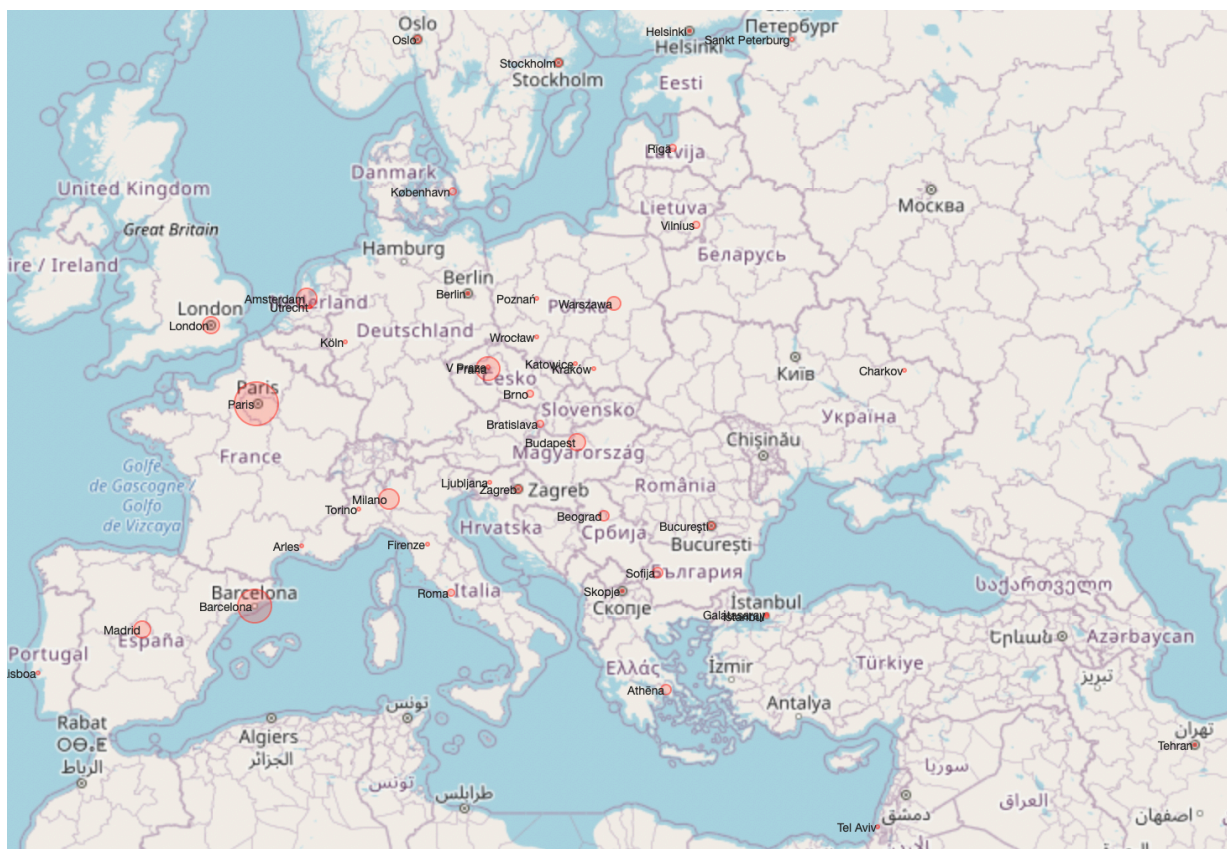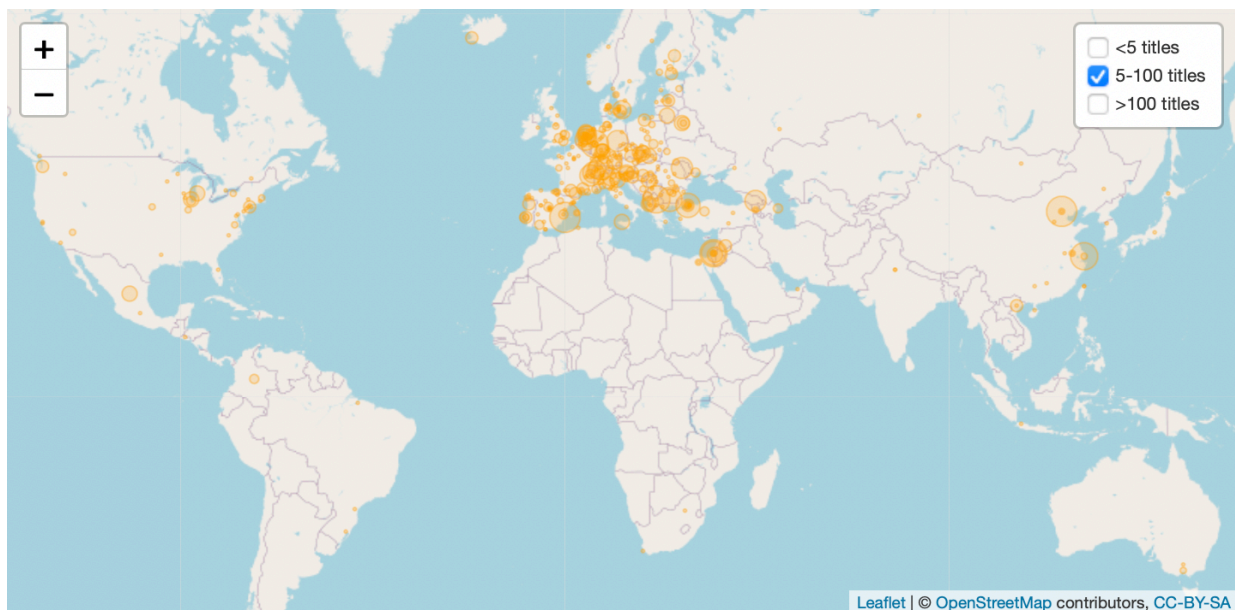
Figure 4.3. Title counts per first publication place with layers of >100 titles with zoom on Europe. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. Marker sizes correspond to title count. Places with more than 100 titles are marked in red.

Figure 4.3 (a zoom of figure 4.2 on Europe) shows that leading centres with more than 100 titles are Paris (2,676 titles), Barcelona (1,949), Prague (1,344), Amsterdam (1,152), Milan (1,112), Budapest (1,015), followed by Moscow, Warszaw, Madrid, London, Sofia, Athens (all less than 1,000 titles). While Western Europe is marked by few large centres, Eastern Europe has a variety of medium to smaller sized centres agglomerated along Poland, Czech Republic, and Slovenia, down to Turkey. This uneven distribution of a few large centres versus many smaller centres in specific regions densely clustered together aligns with Bachleitner's observation of a closer network of peripheral publishing centres. It also sheds light on the overrepresentation of the French language from chapter 2, underlining the importance of a strong centre for linguistic distribution located in Paris.

Besides France, most countries only include one or two centres (like the UK with only London, Portugal with Lisbon, Turkey with Istanbul,[230] Japan with Tokyo, or Norway with Oslo) which implies that the uneven distribution is mainly linked to few strong publishing

---

[230] Istanbul is an interesting example here since it differentiates between districts in the city (e.g., Çemberlitaş, Galatasaray, or Cağaloğlu). It shows how a city can be further stratified in terms of local publishing places.

centres representing a country. However, several countries display a diversity of multiple publication centres, such as Italy (with Vincenza, Milano, Rome, and Firenze), Poland (with Warszawa, Poznań, Wroclaw, Katowice, and Kraków), Spain, Denmark, Switzerland, and Russia, with most other countries having only two publication places maximum.[231] We can clearly see that the more concentrated titles are, the more centralized on few publishing places the translational geography is. Does the Eurocentric, centralizing tendency still hold if we decrease the title concentration? With less concentration we would expect more smaller centres to emerge as well as a lesser Eurocentric focus.

Figure 4.4. Title count per first publication place with layers of 5-100. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. Marker sizes correspond to title count. Places with 5-100 titles are marked in orange.

Figure 4.4 shows the map of publication places after subsetting the frequency table by titles in a range of 5-100 which reduces places to 398, accounting for 21% of all unique places. As compared to figure 4.2 and 4.3 of highly concentrated places, figure 4.4 adds more semi-concentrated centres to North America (Seattle, Columbia, Chicago, Los Angeles, Minneapolis), and Asia (Seo ul, Nan jing, Hà Nôi, New Delhi). While the map is still following a Eurocentric focus of the library catalogue, few semi-peripheral centres are now visible, which have not been captured by the language model of the previous chapters. In the range between five and 100 titles, we can see a number of new places pop up in Spain and the Netherlands (36 additional places), the US (34 additional places), France (28), Italy (27 additional places), China (14), and Poland (15). Germany also appears among the most

---

[231] See appendix for full table: 20220614_alldnb_ch3_pubplace_110titlesplus_countries_freq_v3.csv

diverse countries in terms of publishing centres with 27 places,[232] stressing the importance of bilingual editions. We can see that while decreasing title concentration did not change the Eurocentric tendency—most centres are still situated in Europe—it diversifies countries underrepresented in the previous figure, which stresses the role of semi-concentrated centres for the geographic distribution of translated German fiction. This observation becomes even more apparent when looking at the top centres with the most titles for that range, led by Gata de Gorgos Alicante (95 titles), Tiranë (90), Zürich (88 titles), Cağaloğlu (83 titles), or Tblisi (71 titles).[233] With decreased concentration, peripheral centres become apparent, shifting away from the Eurocentric focus.

Subsetting the frequency tables by places with less than five titles reveals the geographic distribution of the majority of places (1,397 places) accounting for 75.5% of all places. Hence, further decreasing concentration shows that there are only few concentrated centres and an increased number of places outside of those centres with very few titles.
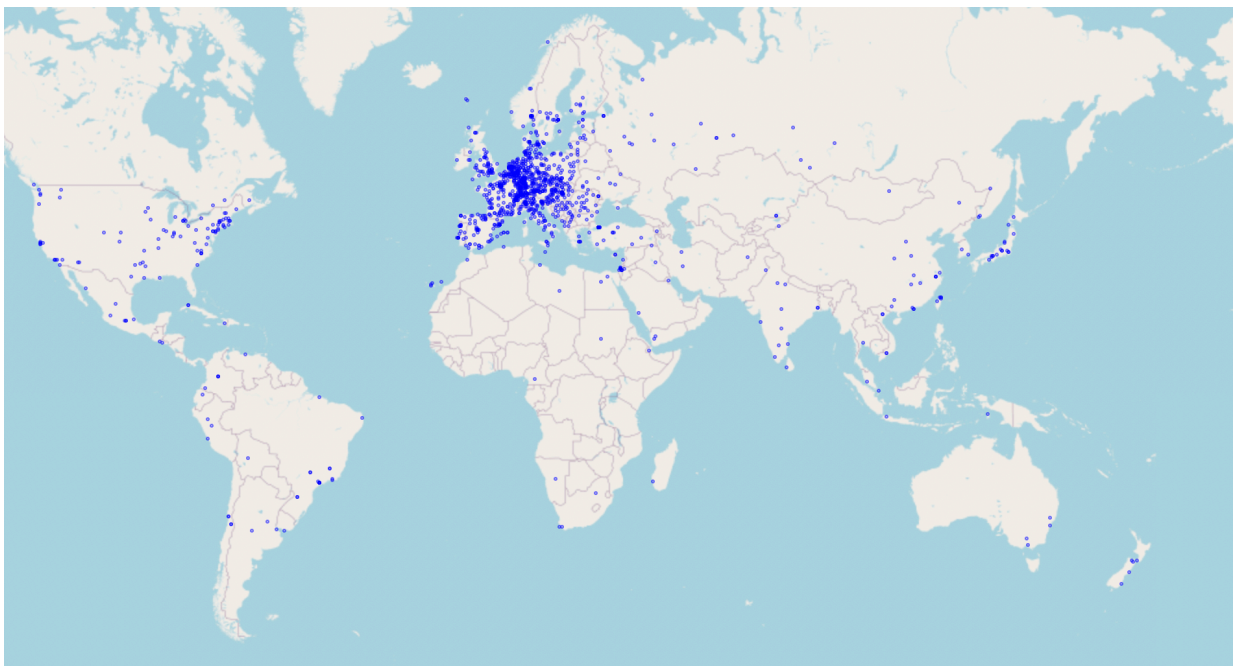
Figure 4.5. Title count per first publication place with layers by increments of < 5, 5-100 and >100 titles. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. Places under five titles are marked in blue.

As has been already discovered in chapter 2, figure 4.5 shows that titles are

---

[232] Köln, for instance, is the publishing place of a number of bilingual publishers such as Manšūrāt al-Ǧamal. See appendix for table: 20220614_alldnb_ch3_pubplace_5_100titles_countries_freq_v3.csv
[233] See appendix for table: 20220614_alldnb_ch3_pubplace_5_100titles_places_freq_v3.csv

concentrated in mainly Europe, while places with less than five titles (blue markers) also appear scattered across the map; continents such as South America, Africa, large parts of Asia, India, Australia, and New Zealand appear to be peripheral publishing places. While the Eurocentric focus persists even for very small publishing places—apparent in the blue cluster of points on the map in Europe—we also see that besides large centres there is an increased number of smaller publishing places in the same country. By counting unique places per country,[234] we can see that the countries with the most publishing places are France (133 places), the U.S. (90 places), Spain (80 places), Italy (70 places), Switzerland (60), and the Netherlands (52) followed by Germany, England, Poland, Czech Republic, Turkey, Denmark, and Hungary.[235] Decreased concentration hence reveals that some countries in the periphery, while they do not have large centres, are comparably important, as revealed by a diversified publishing landscape of certain countries composed of smaller centres.[236]

In chapter 3, I argue that peripheral languages reveal a high canonical focus. In other words, the further into the periphery we look, the more the canon becomes visible. Does the finding still hold that peripheral publishing places are marked by a high canonicity of authors? Based on the findings in chapter 3 that show that the periphery is composed of highly canonical authors and titles, we can assume that this is also the case for places with less than five titles. When looking at the titles that are most prevalent among least concentrated places (< five

---

[234] "Country" column. See appendix for full table
20220614_alldnb_ch3_pubplace_5mintitles_countries_freq_v3.csv
[235] A higher number of publishing places is also reflected in higher numbers of distinct publishers. France for instance has 152 publishers, the U.S. 113, Spain 89, Italy 88. See appendix: See appendix:
20220614_alldnb_ch3_pubplace_5mintitles_countries_publishers_freq_v3.csv
[236] A closer look at the publishing places reveals publishing places like Roskilde, Aix en Provence, al Ǧīza, Baġdād, Wu han, Austin, Amadora, Belgrad. While these cities do not have a concentration in titles, they do, however, display a diversity in their publishing market. Further zooming on each centre shows further how stratified and diverse these smaller peripheral centres are. Belgrad, for instance, has only four titles in the catalogue dataset, all of which published by a different publisher (Laguna, Samizdat B92, Platō, Clio). Same for Baġdād with three titles by three different publishers (aṭ-Ṭabʿa al-ūlā, aṭ-Ṭaʿba 1., aṭ-Ṭabʿa aṯ-ṯānīya). However, Wu han has the same publisher (Chang jiang wen yi chu ban she), Austin and Amadora with 2 distinct publishers. We can see that these peripheral publishing places display a diversity in publishers which in turn also influences the selection of translated titles, or the authors these publishers are specialized in. Amadora for instance has a more canonical focus with a publication series of "Grandes génios da literatura universal" including Goethe's *Die Leiden des jungen Werthers*, or Nietzsche's *Also sprach Zarathustra*, while Belgrade's focus is on canonical (Bernhard Schlink and Robert Walser) as well as contemporary literature including Richard Schwartz's autobiographical book *Der andere nebenan: Eine Anthologie aus dem Südosten Europas* that appeared in five languages simultaneously. Baġdād appears to have a similar focus with Walser, Süskind and Michael Krüger, while Austin seems to be specialized in contemporary authors such as Ursula Krechel and Brigitte Kronauer. Wu han only has titles by one author, Elfriede Jelinek all published by the same publisher. These few case studies not only show that peripheral places have a high diversity in publishers, but also in titles and authors appearing to be specialized in their selection.
See appendix for table: 20220607_alldnb_ch3_pubplace_5_titles_places_freq_v2.csv

titles),[237] we can see titles such as Goethe's *Faust* (nine places) and *Die Leiden des jungen Werthers* (six places), Hesse's *Siddhartha* (five places), and Kafka's *Die Verwandlung* (six places) and *Der Prozess* (four places), but also Rudolf Steiner's *Anthroposophischer Seelenkalender* (six places), Karl May's *Winnetou* (five places) and, not surprisingly, Süskind's *Das Parfum* (four places).[238]

Looking at the distribution of titles in geographic space for the complete corpus reveals that the canon is not only concentrated but also widely distributed in geographic space. Counting unique places per title for the complete dataset shows that with 90 distinct places, Goethe's *Faust* appears to have the highest coverage (56 places), followed by *Der Prozess*, *Siddhartha*, *Die Leiden des jungen Werthers* (45), Schlink's *Der Vorleser* (41), Hesse's *Der Steppenwolf* (37) and *Demian*, *Das Parfum* (35) and *Die Verwandlung* (34).[239] Only 50 out of all unique titles (n=12350) are represented in more than 20 places,[240] which comes to show that only few titles have a far reach geographically. Most of these works are canonical, which supports the argument that the canon is not only concentrated but also widely distributed in geographic space, for which lesser concentrated, peripheral publishing places play a significant role.

As compared to language distributions, mapping publication places shows that while there is a persistent Eurocentric focus in the catalogue, peripheral places do contribute toward the global circulation of German literature in translation, especially of canonical and popular authors. Karl May, for instance, has translations published in 48 different places, 33 of which appear among the places with less than five titles (e.g., Tjumen, Koloszvár, Santafé de Bogotá, Stara Zagora). This makes Karl May stand out as a translational author and partially confirms the prior finding while also highlighting that not only strictly canonical titles are shared among peripheral places. The appearance of translational authors such as Richard Schwartz, Patrick Süskind, or Elfriede Jelinek support the argument that the periphery is both representative of the literary canon as well as of translational literature, which both have a far reach linguistically and geographically.

These findings confirm the previous observations and findings of chapter 2 and 3, that the translation field is concentrated in few centres with a Eurocentric focus. However, mapping publishing places by increments of title concentration reveals a diverse, stratified

---

[237] I calculated the total sum of places per title in the subset of places with less than five titles (using the uniform.title column).
[238] See appendix for table: 20220614_alldnb_ch3_pubplace_5mintitles_freq_v3.csv
[239] See appendix for table: 20220614_alldnb_ch3_pubplace_geo_cleantitles_placecount_v3.csv
[240] Most of these titles are by few authors such as Hesse, Goethe, and Kafka. I will discuss authors in more detail in the last section of this chapter.

literary geography of translation and highlights a diverse network within countries and publishing centres—namely that translation is not vernacular or national but more so regional—which the language model overlooks. With decreasing concentration, the dominant role of linguistic centres and the centre-periphery relationship are challenged by the multitude of centres within each linguistic community, which goes in line with Reynolds and Vitali's argument of a concentrated or dispersed translation field.

From looking at the distribution of specific titles and authors, it is apparent that canonical titles have a connecting role due to their far reach and spread geographically. As the top titles with most publishing places listed above illustrate, authors such as Goethe, Hesse, or Schlink, but also Karl May, have an exceptionally wide geographic distribution and assume a connecting role between different regions on the map. In the following sections I further expand on the argument that the translation network is connected by shared authors and testing whether it is centralized, following a star-like shape defining centre and periphery by its linguistic connections in geographic space.

The role of canonical authors in connecting publishing places and languages

In chapter 3 I raise the argument that the author network is highly connected in terms of languages, while increased connectivity also increases centralization and the separation of centre and periphery. In light of the last section's findings on the role of literary centres, in this chapter I explore the question if we see a similarly connected network of places than we did for language. And further, do we observe a similar tendency, as in chapter 3, that canonical works are defined by connecting the centre to the periphery?
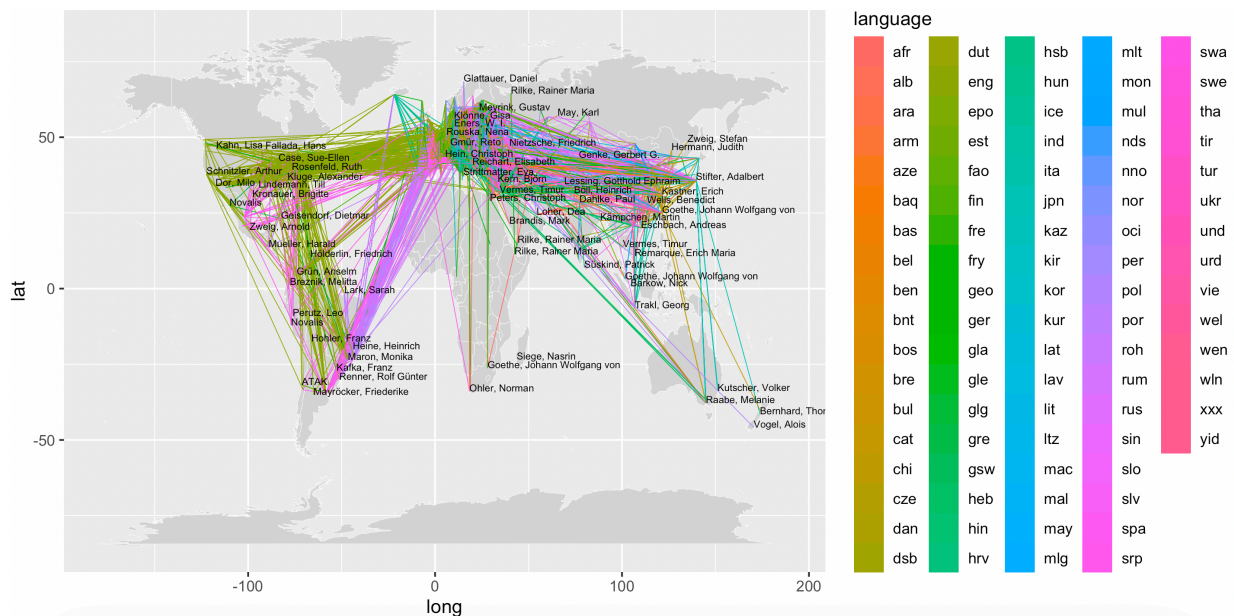
Figure 4.6. Publication places connected by lines of shared authors. Labelled by most frequent authors per place. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet.[241]

As chapter 3 has shown, the majority of authors are shared between central and peripheral languages, visible as a star-like network. Figure 4.6 gives an overview over the main connections between continents through common languages and the most common connecting authors. Publication places are connected by shared authors; in other words, two places have a connection if they have a publication by the same author. The geographic network of translated authors and their publishing places also shows a connected network of literary centres with a star-like shape, centered around Europe.[242] Indeed, 1,437 of the 1,850 publication places are located in Europe, while only 465 of them are in countries outside of Europe,[243] confirming the Eurocentric focus of the catalogue. In comparison to figure 4.2, we can see that the most concentrated places appear to be also the most connected ones by common authors. While the argument that translations appear both unevenly distributed geographically and highly concentrated in larger publishing centres has been confirmed by the findings in the previous section, figure 4.6 challenges this argument by confirming the findings of the previous chapters that the network of translations appears to be highly

---

[241] For a detailed view see appendix,
20220614_alldnb_ch3_author_pub_place_geo_map_lines_languages_places_authorlabels_v3.pdf
[242] While on the Western Hemisphere English, Spanish, and Portuguese dominate the connections, in the Eastern hemisphere the languages are much more scattered. The dominance of green coloured connections highlights the importance of English between Europe, North and South America. With 395 publishing places, English is—besides Germany—the most geographically dispersed language, followed by French with 218 places which only shows in few connections with Africa. Purple lines between Brasil and Portugal, and pink lines between Argentina, Mexico, Lima, and Spain also illustrate linguistic paths and connections of translations, with Spanish having 138 publishing places. See appendix for table:
20220614_alldnb_ch3_author_langplace_frequency_v3.csv
[243] Subsetting the dataset by the prefix "XA" in the country column.

connected by authors with a wide geographic distribution.
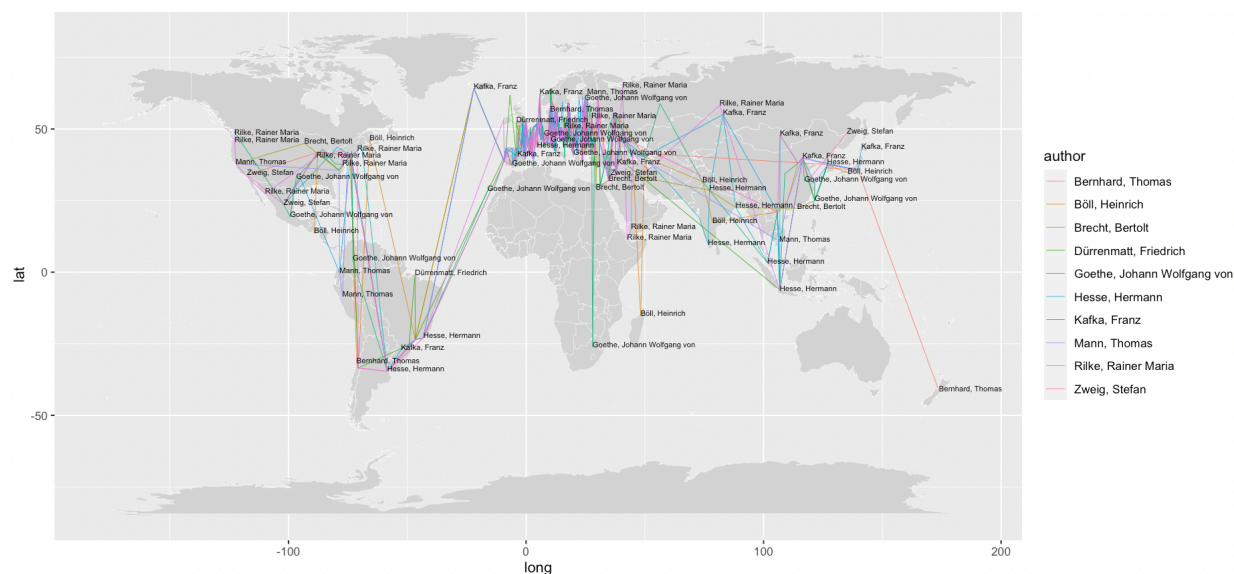
Figure 4.7. Top 10 authors with most publication places connected by lines of shared authors. Labelled by author names most frequent per place. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet.[244]

Mapping the top 10 authors with the most publications places further stresses the wide reach of canonical authors. Figure 4.7 reveals the same names of canonical authors that have been identified as connecting centre to periphery in chapter 3. From the map, we see especially Hesse, Rilke, Goethe, Bernhard, and Kafka connecting peripheral centres to other centres. The centre clusters the authors together in Europe with mostly north-south connections, while toward the periphery connections and centres are sparser, only revealing specific names of authors, such as Rilke, Goethe, Böll, or Kafka and fewer connections. Especially when looking at the places with only few connections, specific authors stand out, such as Rilke connecting publishing places on the West coast of North America, or Hesse bridging Depok, Kuala Lumpur, Calcutta, and Mumbai, demonstrating that for peripheral publishing places, these authors appear to have a connecting role which supports the argument that canonical authors bridge distinct parts of the network.

The role of canonical authors in connecting centre to periphery is also evident when looking at the authors connecting European to non-European publishing places, supporting the finding that the translation network is centralized toward Europe. 52% of authors appear to be shared across European and non-European publishing places and the main connecting authors are Hesse, Kafka, Goethe, Böll, Mann, Schlink, Bernhard, Rilke, Grass, Handke, and

---

[244] See appendix for detailed map:
20220614_alldnb_ch3_author_pub_place_geo_map_lines_topconnected_authors_colored_legend_v3.pdf

other mostly canonical authors.[245] These authors also appear among the most highly connected.

Counting unique places per author reveals that only 6.8% of all authors have more than 10 places and that 66.7% of all authors have only one publishing place. With a mean of 3.3 publishing places, same as for titles, we can see that only a few authors are distributed evenly in translation on the map of global circulation. Hence, few authors contribute toward the global distribution of German fiction to a great extent, while the majority of authors stay in specific centres and regions and in turn expedite the uneven geographic distribution.

Ranked by title count	Languages	Titles	Ranked by language count	Languages	Titles	Ranked by place count	Places
Hesse,Hermann	49	1035	Hesse,Hermann	49	1035	Goethe, Johann Wolfgang von	180
Kafka,Franz	48	749	Kafka,Franz	48	749	Rilke, Rainer Maria	139
Goethe,JohannWolfgang von	47	560	Süskind,Patrick	48	228	Kafka, Franz	135
Mann,Thomas	42	514	Goethe,JohannWolfgang von	47	560	Hesse, Hermann	120
Bernhard,Thomas	36	433	Böll,Heinrich	44	306	Mann, Thomas	95
Remarque,ErichMaria	34	410	Schlink,Bernhard	43	295	Bernhard, Thomas	83
Courths-Mahler,Hedwig	9	374	Frisch,Max	42	178	Dürrenmatt, Friedrich	80
Konsalik,HeinzG.	24	362	Kehlmann,Daniel	42	170	Böll, Heinrich	78
Rilke,RainerMaria	38	352	Mann,Thomas	42	514	Zweig, Stefan	78
Handke,Peter	37	345	Dürrenmatt,Friedrich	40	243	Brecht, Bertolt	76
Zweig,Stefan	37	326	Brecht,Bertolt	39	242	Hoffmann, E. T. A.	74
Böll,Heinrich	44	306	Müller,Herta	39	198	Frisch,	73

---

[245] See appendix for table: 20220616_alldnb_ch3_pubplace_geo_euvsnoneu_sharedauthors_freq_v3.csv

						Max	
Grass,Günter	37	306	Musil,Robert	39	130	Handke, Peter	73
Schlink,Bernhard	43	295	Jelinek,Elfriede	38	272	Jelinek, Elfriede	72
Roth,Joseph	34	281	Rilke,RainerMaria	38	352	Grass, Günter	69
Jelinek,Elfriede	38	272	Walser,Robert	38	198	Schlink, Bernhard	68
Scheer,KarlH.	2	258	Grass,Günter	37	306	Süskind, Patrick	67
May,Karl	19	253	Handke,Peter	37	345	Hölderlin, Friedrich	65
Dürrenmatt,Friedrich	40	243	Zweig,Stefan	37	326	Schnitzler, Arthur	62

Table 4.1. Top authors ranked by title, language, and place count. Authors are marked in blue, orange, and green if their title, language, and place count do not correlate.

As table 4.1 shows, not surprisingly, the top geographically distributed authors mostly correlate with the top concentrated authors in title count. Goethe, with 180 publishing places, is the author with the most reach, followed by Rilke (139), Kafka (135 places), Hesse (120), Mann (95), Thomas Bernhard (83), Dürrenmatt (80), Böll, Zweig (78), Brecht (76), Hoffmann (74), Frisch (73), Handke (73), Jelinek (72), Grass (69), Schlink (68), and Süskind (67).[246] We can infer that highly canonical authors also appear to be the most distributed authors geographically.

Comparing the count of publishing places with the previous results on title and language counts highlights the role of Goethe, who has the highest place count, but lower language and title counts than, for instance, Hesse. The same goes for Rilke, who has comparatively lower language and title counts but the second highest place count. Kafka contrarily balances all three counts and appears to be concentrated as well as linguistically and geographically distributed. Surprisingly, Hesse has a lower place count compared to his title and language counts, for which he is at the top. However, as remarked above, he appears to have a high reach, connecting peripheral places. Three new author names appear in the top authors by place count which are not included in the other top ranked authors by title and language count: E.T.A Hoffmann, Friedrich Hölderlin, and Arthur Schnitzler. Out of Hoffmann's 187 titles, 19 of them are published in 12 different places outside of Europe, which shows that he is more widely distributed across Europe, but also stresses the

[246] See appendix for table: 20220614_alldnb_ch3_author_pubplace_frequency_v3.csv

importance of his single titles that reach various peripheral publishing places. Similarly, for Hölderin, almost all his titles (12 titles out of 122 in total) published outside of Europe are single title publications stretching across North America (New York), Asia (Beijing), South America (Lima), and the Middle East (Tehran, Tel Aviv, Istanbul). Within Europe, more than half of his places only have a single publishing place. As compared to Hoffmann and Hölderlin, Schnitzler appears to have the widest reach, with 18 non-European publishing centres and not just single titles. North American publishing centres stand out especially, including Chicago (4 titles), Studio City (3) and Riverside (5) in California, alongside single titles in New York and Los Angeles. Schnitzler also has more than one title in São Paulo and Rio de Janeiro. 27 of his 44 European publishing places have single titles, with most titles in Barcelona, Paris, and Milano. From these few outliers, the connecting role of specific canonical authors is apparent while underlining that concentration does not necessarily lead to a wider reach geographically. Specific authors such as Schnitzler can link distinct places by single titles and not be as concentrated as Goethe.

To cycle back to the initial argument of this section, that the author place network is highly connected by authors with a wide reach in a star-like shape of central and peripheral publishing centres, the results presented above further support the findings of the previous chapter. The geographic network appears to correlate with the centre-peripheral network of the languages in chapter 3. Accordingly, it could also be confirmed that in a geographic dimension, canonical authors bridge distinct parts of the network, forming connections between centre and periphery.

Additionally, looking at the spatially most widely distributed authors and comparing their title and language counts also reveals that authors can be distributed across many publishing places, even if they have a lower concentration in title counts. Hence, even though specific centres appear to be highly concentrated in titles, as I show in the first section of this chapter, my findings align with Reynolds and Vitali in that these outliers challenge "the idea that translations radiate from the source text at the centre to ever more peripheral contexts of reception" (Reynolds and Vitali 2021, 9). Some authors' importance appears to be in that they are distributed across many publishing places as well as linguistic communities, an importance that is emphasized in Walkowitz's definition of born-translated literature. While Reynolds and Vitali highlight a de-centered view on literary geography by replacing boundaries and categories of national literatures and languages with regional translational "channels, connections, and interventions" (Reynolds and Vitali, 9), what their model does not capture is the linguistic fields and the communities they form in connecting as well as

distinguishing publishing centres.

Author sub-communities in peripheral vs. central languages and their geographic distribution

The previous sections show that the centre-periphery model is challenged by the geographic model with its multiple centres and multiple peripheries. As Bachleitner and Wolf argue, geographic proximity of peripheral centres can indeed create communities that share authors; however, as Reynolds and Vitali have claimed, they are not related to proximity due to the importance of channels, connections, and relations. A geographic centre-periphery based on spatial proximity, as suggested by Bachleitner and Wolf, however, does not take into consideration that centres are also connected by language (e.g., the US to England to Australia by English), forming distinct linguistic communities. When looking at figure 4.5-7, we can clearly see that the centre and periphery are not two disconnected fields, but that they form a network of languages and shared authors, an argument presented in chapter 3. Channels—as I show in this section—are hence defined by languages shaped by authorial communities, and by tracing them in linguistic and geographic space, the author sub-communities, as well as the relations between publishing centres, become visible. From there it can be investigated which authors the centre and periphery represent and which authors connect the two.

In the last section of this chapter, I follow this notion of channels, connections and interventions tracing the connecting role of authors in geographic and linguistic space by looking at author communities. I use the results obtained in chapter 3 for author communities in central and peripheral language groups and superimpose them unto geographic space, visualizing connections of linguistic communities in order to explore author communities in geographic and linguistic proximity. First, I subset the dataset by authors identified in section three of chapter 3 from the community detection algorithm that are shared across or unique to either centre or periphery.[247] I then look at authors unique to centre and periphery by subsetting the dataset by authors that do not connect centre to periphery, but only appear in either of these groups.

The results of the previous chapters have repeatedly made the point that only few authors connect centre and periphery. According to chapter 3, only 33% of all authors are shared across centre and periphery, meaning that the majority of authors appear in either

---

[247] Source data frames: 20220627_alldnb_common_authors_peripheral_ch2.2_v2.csv
20220627_alldnb_notcommon_authors_peripheral_ch2.2_v2.csv

centre or periphery and form sub-communities. For shared authors especially, the most-connected ones are highly canonical authors, which supports the argument that the canon connects places and forms pathways of translations linking central and peripheral languages but also publishing centres.[248] In many ways, figure 4.5 displays similarities to the visualization of the language network in chapter 3; however, when having a closer look at the authors in each group and their geographic reach (numbers of places), the defining role in connecting centre to periphery of few canonical authors is even more apparent.[249] The authors ranked by the highest number of publishing places in the shared author subset is identical with the most geographically distributed authors in table 4.1.[250] Thus, the point can be made that not only do these highly canonical authors bridge places but also linguistic communities in geographic space.[251]

This raises the question regarding the other 67% of authors that only appear in either central or peripheral languages and whether centre and periphery are also highly connected within themselves by authors with a wide reach across different publishing centres as observed in the language network. In other words, can we observe a similar centre-periphery structure as modeled in section three of chapter 3 with the language network, in that central and peripheral languages reveal different author communities? In chapter 3, I argue that translational authors have the function to signify a certain language-community to which they are exclusive, while highly canonical authors link the centre to the periphery. Accordingly, the next paragraphs describe the linguistic channels that link sub-communities in geographic space and the authors they are connected by.

---

[248] Additionally, I calculated the percentage of shared, peripheral, and central authors most concentrated in titles—represented in the first three figures. 95% of shared authors appear to be highly concentrated with more than 100 titles and 77% of all shared authors are increasingly concentrated (5-100 titles), which supports the finding that canonical (highly concentrated) authors connect the centre and periphery. In comparison, 65% of all central authors are highly concentrated (>100 titles) and 35% moderately concentrated (5-100 titles). While for shared and central authors we can observe increased concentration, this is not the case for peripheral languages. Only 30% of all peripheral authors are highly (>100 titles), 44% moderately, and 25% not concentrated (<5 titles).

[249] The most connected authors geographically per group by counting unique publishing places per author clearly reflect the same author names as on figure 4.6. See appendix for the map of shared authors between central and peripheral languages:
20220621_alldnb_ch3_author_pub_place_geo_map_lines_sharedauthors_langcolored_v3.pdf

[250] See appendix for table: 20220623_alldnb_ch3_sharedauthors_place_freq_v3.csv

[251] When looking at transcontinentally shared authors only (subset of languages English, Spanish, Portuguese, and bilingual), we can see that from the most shared and highly canonical which is even more apparent when looking at trans-continental languages. Hence, figure 4.8 point toward the argument that more so than geographic proximity, linguistic proximity appears to play a significant role. However, there are only a few large, overrepresented trans-continental languages that seem to dominate that tendency. See appendix for table and map:
20220621_alldnb_ch3_sharedauthors_transcontinentallang_authorplace_freq_v3.csv
20220621_alldnb_ch3_author_pub_place_geo_map_lines_sharedauthors_langcolored_transcontinentallang_v3.pdf

In chapter 3, I present the following findings showing an imbalance of authors within central and peripheral languages: 78.5% (2,496) of authors appear uniquely in the central language group and only 680 authors in the peripheral. Hence, the central language group has almost twice as many authors as the sum of shared authors between centre and periphery (1,610). From there we can expect the central language group to include authors with a wider reach across publishing places and a strong concentration in the most concentrated publishing centres.
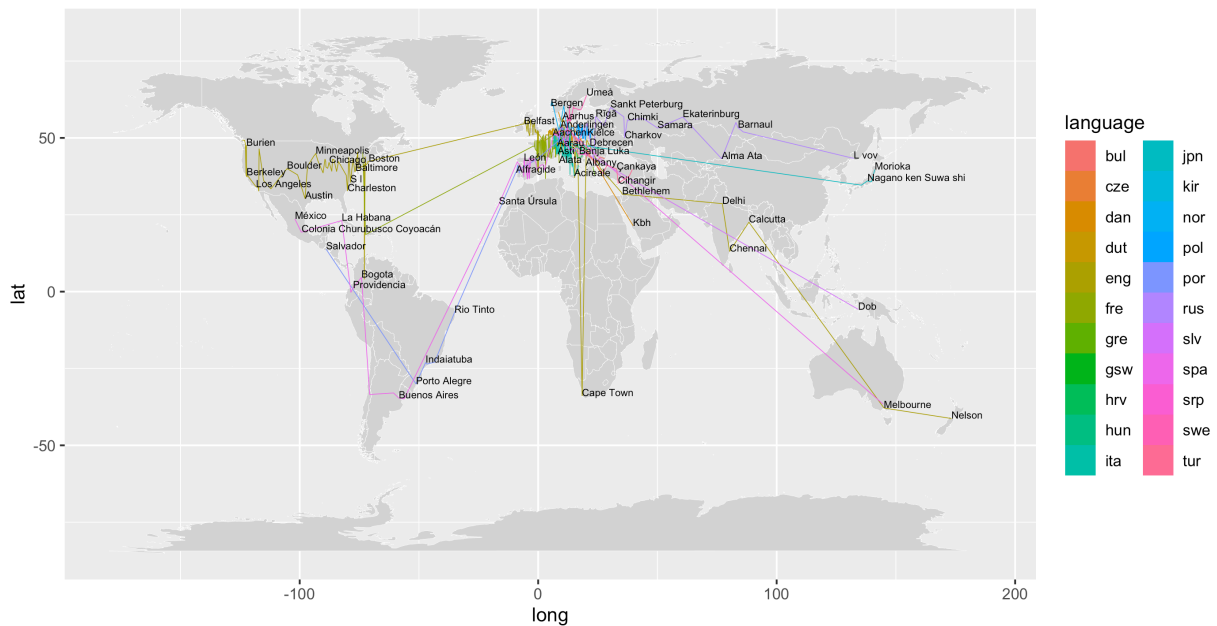
Figure 4.8. Map of publication places of the subset of authors exclusive to central languages grouped by community detection in chapter 3 connected by lines of shared authors. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet.[252]

As we can see on figure 4.8, for central languages, the places with the highest number of authors are Paris (299 authors), Prague (238), Amsterdam (171), Barcelona (161), Budapest (137), Milano (119), Warszawa (118), London (108), and Sofija (197).[253] The role of highly concentrated publishing centres in figure 4.2 also applies very visibly to the central author group. Looking at the languages, central authors outline paths between places connected by languages, such as a Russian path connecting several places, English connecting Europe and North America, and Spanish and Portuguese connecting Latin America and Europe.

---

[252] See appendix for complete map:
20220623_alldnb_ch3_author_pub_place_geo_map_lines_centralauthors_langcolored_v3.pdf
[253] Publishing places overlap with the top languages in terms of author count. See appendix for table:
20220627_alldnb_ch3_centralauthors_geo_authorsperlanguage_freq_v3.csv
20220627_alldnb_ch3_centralauthors_geo_authorsperplace_freq_v3.csv

When ranking authors unique to the central language group by place count, authors reveal a different set of names[254] than we have seen previously for most concentrated titles: Jacob Grimm with 30 places, followed directly by Akif Pirinçci, Ernst Weiß (16 places), Georg Martin Oswald (13 places), and Veza Canetti, Ernst W. Heine, Ludwig Hohl, Uwe Johnson, Gerhard Roth, and Hermann Ungar (12 places). The top 10 central authors with the farthest reach reveal the role of migrant, emigrated and exiled authors. After the brother of Wilhelm Grimm, Jacob Grimm with 30 publishing places, appears Akif Pirinçci, an author of Turkish descent who has titles published across 17 places. Besides Turkey, he also has multiple translations published in Paris, Amsterdam, London, Oslo, Prague, and Warszaw, most of which are translations of his 1989 detective novel *Felidae*, and his novels *Tränen sind immer das Ende*, *Der Rumpf* and *Salve Roma!*. Pirinçci is followed by Ernst Weiß, a Jewish Austrian writer, doctor and translator who had lived in Prague, Berlin, and Paris (DLL[255] vol. 29, 2014, 329). His translations reach back to 1985 and are distributed across 16 places, mostly in Prague.[256] Additionally, Martin Oswald, a columnist of several newspapers who worked for the Bayrischer Rundfunk, has 12 titles each published in a different place; 10 of them are his novel *Alles was zählt* (2000) whose translations all date between 2001 and 2002.

Oswald is followed by several authors with 12 publishing places. One of them is Veza Canetti, an author and translator of Jewish decent born in Vienna in 1897 who later was exiled to London, and who was also the wife of Elias Canetti (DLL EB II, 1994, 568). Out of her 17 translations in the catalogue, two are anthologies published in Riverside, with five translations of her sole novel *Die Schildkröte*. She has several translations published in single publishing places such as Amsterdam, Paris, and Venezia, most frequent titles being two of her short stories, *Die gelbe Strasse* and *Geduld bringt Rosen,* and her novel. In the same geographic distribution as Canetti is the architect and writer of "Unterhaltungsliteratur" (entertainment literature) Ernst W. Heine, who lived most of his life in South Africa and later in Saudi Arabia (DLL EB IV, 1996, 574) and who has titles mostly published in Athens, Brno, Paris, Tokyo, and Turkey. His 1994 novel *Das Halsband der Taube* is his most prominent title in the catalogue with multiple translations in Athens and Turkey. The case of Heine exemplifies how an author can have hotspots of translations that then spur successive translations. Ludwig Hohl is another author who was published in 12 different places in

---

[254] See appendix for table: 20220623_alldnb_ch3_centralauthors_place_freq_v3.csv
[255] DLL stands for Deutsches Literaturlexikon. EB stands for Ergänzungsband.
[256] Ernst Weiß also has an extensive collection of postcards and prints of his works in the library catalogue.

translation according to the library catalogue. The author of Swiss origin, who had lived in Den Haag, Vienna, Paris, and Genf (DLL vol. 8, 1981, 13) has translations published in Genf, Milano, Paris, Barcelona of mainly his two short stories "Bergfahrt" and "Nächtlicher Weg." With multiple titles in Paris and Barcelona, he also appears to have few hotspots. The German author and member of Gruppe 47 Uwe Johnson was also active in several places between 1962 and his death in 1984 after leaving the DDR and East Berlin, such as Rome, New York and Kent and Sheerness in England (DLL vol. 8, 1981, 662). This is reflected in his 12 publishing places of two titles, a collection of writings and his novel *Jahrestage,*[257] which takes place in the exact same address he used to live at in New York. The Austrian author Gerhard Roth also has several translations published in 12 places, such as Riverside, Athens, Istanbul, and London, and single titles in Providence, Saint Petersburg, Barcelona, and Milan. With 12 unique out of 17 titles, most of his novels got translated, including *Wille zur Krankheit* (1973) or *Der Stille Ocean* (1980) (DLL, vol. 13, 1991, 358). Gerhard Roth's work has also been released in the form of films and documentaries, for which he travelled to the U.S., where his crime novel *Der große Horizont* (1979) follows the story of Daniel Haid, an Austrian fleeing to the U.S. With 22 titles and also 12 publishing places, the Jewish-Czech author Hermann Ungar has hotspots in Toulouse, Barcelona, and Boskovice, as well as København and Amsterdam. For the translation of his second most prominent title in the catalogue, his novella *Knaben und Mörder* (1921), he travelled to Paris from Prague (DLL, vol. 24, 1991, 523). His novel *Die Verstümmelten* (1923) is his most prominent title with translations into French, English, Dutch, and Spanish. Surprisingly, only two titles are translations into Czech. Bronislava Volkova has noted that "Ungar's work was immediately noted in France, yet rather neglected in Germany and Czechoslovakia" (Volková 2021, 52), a tendency which appears to be also visible in the catalogue. What the majority of the authors with a wide reach in the central language group show is the tendency of having been active or moving between places. Additionally, Ungar's case also draws our attention to the gaps in translation in an author's work, but also highlights the linguistically specialized focus of the central language author community. Additionally, what ties these authors together, other than being published in central languages, is their biographic tendency to have lived and worked in several places, which also is reflected in their multitude of publishing places.

Taking a closer look at authors where this imbalance of titles and publishing places is present, by comparing the authors with the widest geographic reach with the most

---

[257] A 2018 translation of the novel published by the Review of Books.

concentrated ones in terms of title in the central language group, reveals additional names that have few publishing centres but the highest title counts.[258] The writer of mainly romantic novels Friede Birkner, for instance, has translated titles published only in Brno, Budapest, and Katowice, but is one of the top authors by title count with 74 titles. Similarly, Patricia Vandenberg,[259] another romantic serial novel writer, has 34 translated titles (more than Akif Pirinçci) published only in Warszaw and Praze. Brigitte Riebe has the same title count as Ernst Weiß but only half the publishing places. Similar cases are Gilbert Haefs and Marie Cordonnierová,[260] the writer of romantic, serial, and youth novels and a translator, who has 25 translated titles solely in Brno.[261] From these examples we can see a different group of authors, mostly serial novel writers, who are highly specialized in a few places and are not representative of the previous group of authors who are geographically dispersed. Additionally, the role of female authors in the central group is more apparent, with half of the top 10 authors by title count in the central language group being women writers. Comparatively, only Veza Canetti is included in the top 10 authors by place count. In this group of authors, we observe the tendency of spatial specialization, meaning that they have an increased number of titles concentrated in few publishing places.

As these examples illustrate, when moving away from connecting authors across language communities, sub-communities with specific characteristics emerge. While we can clearly see that the shared authors group also represents highly canonical authors, the further we move into the centre, the more spatially specialized and non-canonical it gets. Accordingly, we can expect the same tendency in the peripheral language group, considering that the periphery is connected by few canonical authors to the centre, therefore when removing connecting authors, sub-communities of authors can be made visible.

As I show in chapter 3, in the peripheral language group, authors have the tendency to be even more dispersed and less centralized. I claim that the 21.5% of authors exclusively in the periphery cluster include an independent group of possibly translational authors; however, the geographic distribution is comparatively smaller than in the central language group. In the peripheral group, only 6% (40 of 646) of authors have more than one publishing place with a

---

[258] I ran a linear regression model on title and place count for central authors and it showed that few outliers fall outside of the regression line with expectedly more places or titles. See appendix for figure. Both the p-values (<2e-16) for the intercept and the predictor variable (title frequency) are highly significant (t value 44.45).
[259] Pseudonym for Gerty Schiede (DLL vol. 14, 523).
[260] Pseudonym for Gaby Schuster (DLL vol. 16, 693).
[261] These authors had some of the lowest residuals in the linear regression model, meaning that they scored unexpectedly high on titles as compared to places. See appendix for table of place, title count and residuals: 20220623_alldnb_ch3_centrallang_author_place_title_freq_v3.csv

maximum of four places, which in comparison to the central language group does not set apart few distinct outliers with comparatively many places.[262]
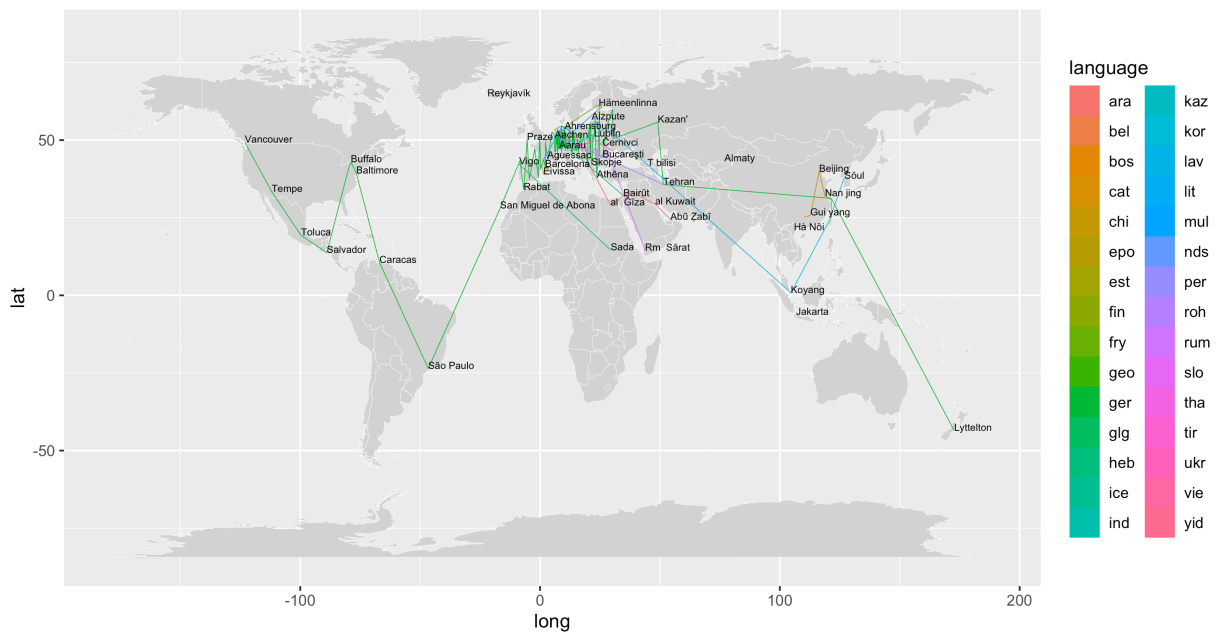
Figure 4.9. Map of publication places of the subset of authors exclusive to peripheral language group grouped by community detection in chapter 3 connected by lines of shared authors. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet.[263]

On the map for authors that only appear in the peripheral language group, visualized in figure 4.9, we clearly see the role of Arabic, Chinese, and Persian beside bilingual publications that are mainly located in Europe and North America, connecting also to Africa and Russia. With concentrated centres in Rīgā (36 authors), Berlin (32), Frankfurt (32), Vilnius (31), Bratislava (24), Munich and Zürich (19), Bucureşti (17), and Helsinki (15), peripheral authors appear to be mostly concentrated in Germany, Northern and Eastern Europe, and East Asia.[264] From the most prominent languages, we can see that bilingual editions stretching across 50 different countries play an especially important part in the periphery, making up 55% of all titles. Figure 4.9 hence visualizes the wide reach of bilingual publications across the map.

When looking at the table of peripheral authors and comparing their title and place count, we can see that they are mostly balanced, meaning that the top authors have each of

---

[262] The regression does not appear linear. See appendix for figure:
20220628_alldnb_ch3_peripherallang_author_place_title_freq_lm_graph_v3.png
[263] See appendix for complete map:
20220623_alldnb_ch3_author_pub_place_geo_map_lines_peripheralauthors_langcolored_v3.pdf
[264] See appendix for table: 20220628_alldnb_ch3_peripherallang_author_place_freq_v3.csv

their titles published in a different place.[265] 89% of all authors in this group have a balanced title and place and 85% have a 1:1 title and place count ratio, which clearly shows that most authors in this group are not concentrated or distributed geographically.[266] Manfred Peter Hein, who is a translator of Czech and Finnish texts and author of children's literature who resides in Helsinki (DLL vol. 7, 2014, 688), has the highest place count in the peripheral group with four titles in four places (Berlin, Reykjavík, Germersheim, Thonon les Bains). One of his titles published in Berlin is a collection of poems translated into 12 languages[267] and one of the six only multilingual (language="mul") titles in the catalogue. Two of his other titles are bilingual: a collection of poems from 2017

L' érable contre la maison = Der Ahorn am Haus[268] published in Thonon les Bains in 2017 in French and German, and another collection of poems with German alongside the Icelandic translation.[269] Ranking with the highest place count are Nick Barkow, who has translations of his period novel inspired by a movie *Das Lied vom traurigen Sonntag*[270] into Korean and Chinese in Koyang, Shang Hai, and Tai Bei. Manfred Chobot also has two bilingual titles in the catalogue, *Schwarze Lava*[271] in French and German which is published in Brunn am Gebirge (Austria), and *Die Enge der Nähe* in Polish and German published in Wroclaw[272] alongside a title published in Kiev. Among the authors with three titles and publishing places is also the bilingual author Ursula Heinze de Lorenzo, who lives and works in Spain; two titles of her collected poems are also bilingual (German-Galician) published in Kassel[273] and La Coruña[274] and one other title is a Galician translation. Ulla Lenze is a contemporary author who also has three titles at three different places published in translation (Dimašq, Helsinki, and Tai Bei) according to the catalogue, two of which are translations of her novels *Schwester und Bruder* (2003)[275] and *Der Empfänger* (2020) which is translated into Finnish.[276] Although none of her translations are bilingual, they both describe the story of a traveller, one to India (in *Schwester und Bruder*) and a German exile in New York (in *Der Empfänger*).

---

[265] See appendix for table: 20220628_alldnb_ch3_peripherallang_author_place_title_freq_v3.csv
[266] As a comparison in the central language group, 62% of authors have a balanced title and place by a 1:1 ratio, which is a tendency that both groups share, but not to the same extend.
[267] https://d-nb.info/1057976911
[268] https://d-nb.info/1161124942
[269] https://d-nb.info/956034128
[270] https://d-nb.info/957527039
[271] https://d-nb.info/1009779001
[272] https://d-nb.info/960490620
[273] https://d-nb.info/974377287
[274] https://d-nb.info/1007913940
[275] https://d-nb.info/989667545
https://d-nb.info/98184488X
[276] https://d-nb.info/122036617X
Interestingly, she got nominated with the Uwe Johnson award for *Der Empfänger.*

She was a writer in residency in Dimašq, Mumbai, Istanbul, and Venetia and spent extended periods of time in India.[277] The last author with the highest place count in the peripheral group and an equal title count of three translations is Peter Völker, who only has bilingual (language="ger") titles in the catalogue, all of which are published in Germany (München, Gelnhausen, and Berlin) and represent collected poems in German and Greek.[278] Peter Vögler is founder and collaborator in several projects in Greece, Uruguay, and Bosnia-Herzegowina.[279] These authors with balanced place and title counts also show similar characteristics as the central authors in that they are transnationally active with the added role of bilingual and multilingual titles. They also appear to be spatially specialized, in which we can see a close relation between the language and the publishing place of bilingual editions.

For spatially highly specialized authors with low place count and proportionally high title count, this tendency still holds. To see which authors are disproportionally concentrated compared to geographically distributed, I rank authors by highest title count, comparing it to the place count. From there, we can find a few authors with a higher title count[280] as compared to other authors, such as the poet Oliver Roland[281] who has eight titles, seven of which are bilingual and published in Windflecken in Persian, Rumanian, Frisian, Russian, and Portuguese, as well as the sculptor Guido Häfner with seven bilingual titles of graphic prints of poems by Ingo Cesaro published in Nürnberg in English-French, English-Portuguese, and English-Maori. Other authors with an increased title count in the peripheral language group also include the contemporary dramaturgist Winnie Abel with five titles of theatre plays and only one place (Swiss Germany),[282] followed by Ingo Cesaro, who has five bilingual poetry collections in English, Indonesian, Dutch, and Arabic published in Nürnberg, and Johannes Häfner who has three translations in Arabic and one bilingual collection of his poems published in Nürnberg. In addition, there are also several other authors in the peripheral language group who rank the highest in titles but are only published in one publishing place. With four novels into Latvian published in Riga, Samantha Roosen has the

---

[277] Lenze, Ulla. "Vita – Ulla Lenze." Accessed June 29, 2022. https://www.ullalenze.de/vita/.
[278] https://d-nb.info/988077612
https://d-nb.info/973956569
https://d-nb.info/975326651
[279] Vögler, Peter. "Peter Völker – Schriftsteller Für Romane, Lyrik Und Belletristik." Accessed June 29, 2022. http://www.petervoelker.de.
[280] In addition to comparing title and place count per author, I ran a linear regression model. According to negative residuals from the linear regression model, the t value for the title count is lower than for central languages, (18.44) but still highly significant (p value <2e-16), meaning that there is a significant association between the predictor and the outcome variables.
[281] With the lowest residual of -1.59445494 of the linear regression model.
[282] The second lowest residual, -1.06702597, of the linear regression model.

same ratio as Carsten Schreier, who has four of his theatre plays published in Mühltal in low German, and Edith Sommer with four bilingual titles of poems in French published in Nice. After Karl Stejn, with four bilingual translations into Czech published in Vienna, all other authors have less than three titles and two publishing places. Thus, from the authors with higher title count as compared to place count in peripheral languages, we can observe a more drastic spatial and therefore also linguistic specialization in a few German centres, which again appears to be influential in bilingual publishing. The role of German dialects, especially in theatre, is also more apparent, as well as the importance of poetry collections as a bilingual genre. As the case of Häfner and Cesaro illustrate, some translations are in various media forms such as graphic prints; therefore, bilingual publishing not only transgresses linguistic categorical boundaries of the catalogue and the translation network, but also the definition of a translated literary work as a set of texts translated by one author from an original into a target language.

To revisit the main points in this section, a centre-periphery model of author communities and their publishing places reveals that the central language community of authors represents the translational category in the sense of being spread out across several publishing centres, with an emphasis on serial novels and authors that are themselves moving in geographic space. In comparison, the peripheral group of authors is characterized by bilingual and multilingual poetry translations with a high spatial specialization in few publishing centres mainly in Germany, with theatre plays translated into German dialects also being significant.

Furthermore, coming back to the map of central and peripheral languages, while a close look at the authors in each group shows that not only are authors connecting different places translationally—as in the case of central authors—but are also doing the opposite in that they specialize in few publishing places but still connect to other linguistic areas through bilingual translations. Hence, while they bridge different linguistic spaces, they are highly concentrated locally in geographic space. Especially in the peripheral language author group, these tendencies are visible since, as bilingual works illustrate, texts can be tied to one publishing place but represent several languages.

The findings for the peripheral and central language groups call for a re-evaluation of the main points of this chapter. In the first section results have shown that publishing places of translations of German fiction in the German National Library catalogue are highly

concentrated in few European publishing centres, which correlates with the findings in previous chapters for the language distribution. Additionally, with decreased concentration more peripheral places are visible. I also further pursued the observation of the previous chapter that peripheral languages are marked by high canonicity. As the results of the second section reveal, this is the case for authors connecting the centre to the periphery, which also follow a Eurocentric star-like shape similar to the language network. Highly concentrated publishing centres hence also appear to be highly connected by shared authors. Further, the most concentrated authors also turn out to be the most geographically distributed, representing a set of canonical authors with outlier cases of authors with a comparatively lower title count and a high place count. In section three, I then found that the authors of this small canonical set are the main bridges between the central and peripheral language groups and publishing places, while the majority of authors are in the peripheral and central language groups and cluster in specific publishing places, making distinct author communities visible. Closely examining the relationships of authors to their publishing places, languages, and genre highlights the cases that transgress the language and geographic model of publishing places and target languages.

Mapping publishing places and visualizing the role of cultural centres for the geographic distribution of translations partially confirms Heilbron's, Bachleitner and Wolf's, and Reynolds and Vitali's arguments of an uneven distribution of translations from few literary centres. However, the role of canonical authors in connecting publishing places by languages further visualizes the underlying network of this uneven distribution and its dimensions in geographic space.

The geographic model presented in this chapter hence still follows the implication that languages form connections in geographic space. As the maps in the second section have shown, the role of transcontinental languages as well as bilingual publications in section three challenges the notion of geographic proximity presented by Bachleitner and Wolf and Reynolds and Vitali. Each translation is situated in linguistic and geographic space simultaneously and can also include multiple languages, therefore forming connections and addressing a multiplicity of communities. The examples of authors in central and peripheral language groups resonate with Walkowitz's point that once "literary works begin in several languages and several places, they no longer conform to the logic of national representation" (Walkowitz 2017, 30). In addition to that, as section three illustrates, publishing places may be strongly linked to the author's own translational occupation, their location, their travels, and their projects. It is those instances in translation that question the implications and

classification systems that rely on national identifiers such as "German" literature or writer. In Walkowitz's words, who challenges the monolingual understanding of literature "viewed from the perspective of migration, the concept of literary belonging may have outlived its usefulness" (Walkowitz 2017, 25).

5. Discussion

The findings presented in this thesis for German fiction in translation in the German National Library catalogue can be summarized under three notions: concentration, connectivity, and sub-communities. This chapter is dedicated to discussing each notion in the light of the literature reviewed in chapter 1.1 and 1.2, presenting my main arguments and the overarching questions that have guided this project in order to revisit my main hypotheses: that there is an imbalance of translated titles by author across languages, which sets canonical and translational authors apart on the one hand, and, on the other, that authors bridge linguistic communities and publishing places, therefore making the connections between different literary cultures visible.

I hope to show how this proposed model of descriptive statistics, network analysis, and geographic mapping offers a methodological framework for gaining a deeper understanding of the role translations play in the formation of the national, translational and/or world literature canon through the library catalogue and how the concept of a translational canon challenges these paradigms. In the last part of this chapter, I expand on how my model and findings contribute toward making bibliographic translation data more visible in the field of literary studies, German studies, and digital humanities. By outlining the limitations of this thesis project, I hope to point toward areas where future research could venture.

Translations are concentrated in few languages and authors

My results of title and author concentration revealed the dominance of few authors and few languages in the translation field. I measured concentration to address the argument that translations are not evenly distributed across languages and publishing centres, meaning that certain languages and authors have the majority of titles in the catalogue. My findings point toward the question of: what does the catalogue's concentration on languages and authors tell us about translation and world literature? Studies about world literature have noted the uneven distribution of translations as Eurocentrism/Germanocentrism, especially critiquing the representability of world literary authors, who are highly concentrated in anthologies, canonical lists, and national collections. Accordingly, I used concentration as a proxy for canonicity in order to position authors in the field of translation. I measured linguistic

distribution as a proxy for an author's position in the translational field with the underlying assumption of an even distribution, which would imply that the translations by authors are equally distributed across languages.[283] However, when measuring language and title concentration in the library catalogue, we can see that the catalogue is concentrated in few languages and authors, making the translational field small, where only a few authors reach smaller languages. To come back to my point in chapter 1.1, that the national canon is an exclusive category to a few dominant literary cultures, I found that only a small number of mainly European languages are included in the catalogue in which the majority of translations are published.

While my findings of a highly concentrated catalogue in a few languages and authors align with Damrosch's, Bassnett's, and Bachmann-Medick's argument of a Eurocentric world literature paradigm, they clearly challenge world literature's claim of universality highlighted by Zhenduo and Apter, showing that German fiction in translation is linguistically highly concentrated and canonical. The results for concentration show to what extent translations of German fiction in the DNB catalogue are not as worldly as presumed, but instead highly focused on a few large languages and a few canonical overrepresented authors. Hence, we can ask, how does the uneven distribution of translations across languages and authors challenge the notion of a global, representative canon of world literature, as Bassnett and Bachmann-Medick have argued?

In search for the translational category, I claimed that my model made it possible to determine which and to what extent a work or author can be categorized as translational—measured by linguistic distribution—a task which, however, turned out to be only possible to a limited extent since translationalism is intertwined with the field of the national canon in translation. I looked at the overlap of concentrated and distributed authors and found that the most concentrated are also the most linguistically distributed authors, and these most-concentrated authors also turned out to be representatives of the national canon. Accordingly, I argue that world literature is where canonical and translational fields overlap. My results showed a distinct translational category, where certain authors had a stronger translational tendency in the form of balanced title and language counts, meaning that even though they had few titles, they were linguistically and geographically well distributed. This implies that both translationalism and canonicity are equally important for the formation of the field of

---

[283] We could also ask if we can expect an even distribution considering the nature of publishing markets with their dynamic of large versus independent publishers. However, this question lies beyond the scope of this study and will have to be addressed in successive work.

translation and the global distribution by means of translation, or what has been commonly categorized as world literature.

This finding led me to ask whether world literature is a global extension of the national canon *in* translation by means of a high concentration and wide linguistic distribution. As I showed in chapter 1.1, especially in the dawn of the making of world literature canons, world literature is strongly tied to the representativeness of selected works in the national canon with translation as a pre-condition which is only marginally addressed. I argued that besides Hesse's list, encyclopedias and anthologies as well as library catalogues are classification tools of canonization within the framework of world literature. The top concentrated and linguistically distributed authors reflect the names on the mentioned lists of world literature, with Goethe and Hesse as the top two translational and canonical authors in the collection, both of whom have significantly contributed to the making of world literature as a canon of representative works. We can argue that the library catalogue as a national collection documents the global canonization of the national canon by cataloguing works in translation by authors deemed representative of its initial collection in the source language (German). The national and translational/transnational canon hence appear to be inseparable and, as my results suggest, a distinct translational category depends on the national canon *in* translation. If we follow the argument that looking at translations in the catalogue and where the canonical (concentrated) and translational (linguistically distributed) authors and titles overlap, highlights the space of world literature in the national collection, world literature appears to be a mere descriptor of the national canon *in* translation rather than a distinct category of works and authors with its own modes of classification.

This observation resonates with the counter-model of translational transfer for the world literature paradigm as proposed in translation studies, with Damrosch's argument that world literature is always national and transnational, and with Etiemble's claim that national literatures are not distinct from a "world" literature.[284] Bassnett goes as far as to say that when we study translation, we also study world literature.[285] Looking at my findings for concentration, I would add that when we study translation, we also study the global canonization of the national canon as well as the national library's agency in this process.

When looking at the language concentration and distribution of specific authors, as well as the concentration in specific publishing centres, we can see how translations provide a

---

[284] "[M]ay we candidly admit that the totality of all national literatures simply makes up literature, without adjective?" (Etiemble 2012, 87).
[285] One might ask: Why not just study translation instead of world literature, then?

map to locate a work in a global context as Damrosch has argued, or a map of language geopolitics as Apter calls it, that documents the transfer or systems of transactions (Bassnett). As I argued in chapter 1.1, translation is a pre-condition for the circulation of literary works beyond the boundaries of their source language(s) and therefore is the driving force behind the global canonization which we call world literature: in other words, the extension of the national canon into multiple literary cultures by forming new linguistic connections within a network of literary transfer.[286]

Translations reveal a highly connected network of shared authors across language communities

By applying network analysis based on Heilbron's centre-periphery language model, I found that the language network is highly connected by shared authors and, furthermore, that a set of canonical authors connects the central and peripheral language groups detected by community detection. I found similar central positions of languages identified by Heilbron, Wischenbart, and Bachleitner and Wolf, as well as Sapiro, which contribute to the uneven distribution (concentration) of authors both linguistically and geographically. I also found that canonical authors were most visible in connecting peripheral languages to central languages.

Hence, comparing my findings for connectivity to concentration also revealed the connecting function of translations between language communities. Again, in defining translationalism in terms of linguistic distribution, the translational and canonical fields appear to overlap in that the most concentrated are also the most connecting authors, which are representative of the national canon. I argued that language centrality is based on the share of canonical/translational authors in the translation network and that the farther toward the boundaries of the network we move, the more visible the national canon in translation gets. In other words, what distinguishes canonical authors in the field of translation is an increased concentration in combination with bridging central and peripheral languages.

My findings are supported by applying geographic mapping of centre vs periphery with connecting languages, highlighting the role of geographic centres with high title concentration—with a special case of bilingual translations—in connecting languages by

---

[286] As mentioned in chapter 2, Pym also argues that translations are a necessary measure of the "degree of language diversity" in the network of the circulation of one national literature globally (see Pym and Chrupała 2005).

shared authors. Going back to the literature on the sociology of translation I discussed in chapter 1.1 and 1.2, this represents a novel finding for the field in that even though the notion of literary transfer and the core-periphery system of languages has received a significant amount of attention in studies on translation as a counter-model to the canonical, Eurocentric study of world literature,[287] the connecting role of authors through translation has not been addressed.

The finding that languages are highly connected by shared authors and that they form distinct communities went against my initial hypothesis that the library catalogue is Eurocentrically concentrated on few languages and canonical authors. However, it allowed me to further investigate the role of the canon in connecting linguistic communities and make visible the non-canonical spaces of shared authors in language communities. My findings can also be seen as a counterargument to the Eurocentric world literary paradigm by narrowing the canon's importance to connecting central and peripheral language groups.

In chapter 1.2, I discussed several studies that use bibliographic translation datasets, among others such as publishing reports, and I pointed out that even though they stress the concept of transfer or flow of translations, their model is limited to frequency distributions per nation or languages, as is especially the case in market-driven studies. While the core claim is that translations transgress national and linguistic boundaries, most studies investigate each variable and how they contribute toward the circulation of translations, but do not look at the connections and networks of transfer. Hence, the proposed language network model responds to the main challenge with bibliographic data which former studies did not address: that translation represents a network of languages and therefore not a binary, but rather a one-to-many relationship of variables (one edition) and, in the case of bilingual or multilingual editions, a many-to-many relationship (a German-Japanese poetry collection in translation). Coming back to the transfer (Sapiro) and flow (Wischenbart et al.) concepts in translation, we can see that they have found limited methodological application in modeling flow of translations of German fiction and by describing distributions of variables and factors, which are oftentimes limited to a single author, genre, or period, without modeling

---

[287] Transfer being one of the central concepts of Bassnett as a part of the translational shift toward the study of "interconnectedness, on global literary and cultural flows on the one hand, and on question of agency on the other" (Bassnett 2011, 239), which has also found application in Reynolds and Vitali's study "tracing interconnections and variances among the many co-created versions which constitute the global *Jane Eyre*" (Reynolds and Vitali 2021, 3). Long and So (2016) analyzed the dispersion of stream of consciousness across a corpus of randomly selected novels in different languages with a focus on English and Japanese in order to test a computational model of SOC style to study its diffusion as a world literary form. They too apply a field terminology tracking the movement of text-intrinsic features from a world literary system's core to its semi-periphery and periphery.

their relationships. As I argue, taking into account how translations transgress national boundaries by forming linguistic communities and how the position of specific languages can indicate the role of specific authors in the global translational circulation of fiction. In that regard, my model is a direct answer to Heilbron's central question "how can one account for the uneven flow of book translations between various language groups? And how can one explain the varying role of translations within different language groups?" (Heilbron 1999, 431). The underlying hypothesis that shared authors connect language communities opens up this concept of flow and transfer, offering a way to model relationships within the translation network without solely relying on a source-target concept, instead using connections between languages by shared authors and language positions to identify canonical and translational fields in the catalogue.

Finding sub-communities of authors and mapping the translational network of authors geographically challenged the source-target dynamic inherent to the concept of literary transfer of world literature—that text radiates from centre to periphery—which Reynolds and Vitali critiqued by introducing the idea of channels, connections, and interventions. Hence, if translationalism and canonicity are equally important in driving global canonization of world literature, so are the language communities and positions of languages in which works of translations are circulated. In that regard, translationalism as distinguishing and connecting language communities challenges the common understanding of world literature in that it is not limited to a wide linguistic distribution and a high concentration.

Translations reveal sub-communities of authors

In response to the transfer concept—translations transgressing linguistic and geographic boundaries—central to the geographic shift in translation studies (discussed by Bassnett, Mani, and Italiano), investigating the authors that are shared between central and peripheral languages visualized the channels of canonical authorial circulation in linguistic and geographic space. While my findings via network analysis accounted for connections and relations between linguistic communities, geographic mapping contributed to the understanding of channels between publishing centres, which are formed by authors connecting language groups.

On the one hand, the finding of canonical authors connecting publishing places by language communities reinstated Heilbron's centre periphery model and went in hand with my previous findings for connectivity. On the other hand, mapping and then closely

investigating the spatial distribution of authors unique to central and peripheral languages revealed author communities beyond the canonical set of authors the catalogue is focused on. I found that the authors exclusive to the central and peripheral language groups represent the majority of authors as compared to the shared authors, which suggests that only a few authors connect centre to periphery, while the author sub-communities make up a significant amount of the catalogue's translations.

Since central language authors represent highly concentrated languages with large publishing centres, they are also highly connected in terms of geographic distribution and can therefore be considered as translational. Peripheral language authors do not show the same tendencies in terms of being distributed across many publishing centres and therefore well connected, but, on the contrary, are regionally specialized. Mapping bilingual publications, however, showed how highly connecting, both linguistically and geographically, translations of specific authors are. Furthermore, the role of bilingual translations also challenges Wolf and Bachleitner's argument that geographic proximity in peripheral publishing centres forms communities of authors, since these translations connect multiple language communities and publishing centres on multiple continents. These findings go against the understanding of translation transfer or flow that texts radiate from centre to periphery but still stress the importance of regionally specialized and multilingual authors in shaping the channels in the translation network.

A closer reading of the authors in each group with special attention to their titles and publishing places showed that the centre and periphery reveal sub-communities of authors, which are genre-specific and spatially specialized. While central authors have serial novels distributed across many publishing centres, peripheral authors are theater and poetry-based, localized with few titles published in different places. These characteristics of the most spatially distributed authors in the centre and peripheral language fields set the author communities apart from the canonical authors connecting centre and periphery.

While I assumed that the canon is mostly visible in peripheral languages (based on the assumption that canonical authors have a wide reach), looking at authors only shared among peripheral languages showed that this is not solely the case. The canon appeared to be connecting centre and periphery, while peripheral and central languages revealed a different set of authors, which challenges our understanding of world literature. If German fiction in translation circulates in peripheral languages only and is regionally specific, can it be considered world literature even though it lacks geographic distribution across multiple publishing centres and languages, but rather demonstrates the wide geographic and linguistic

reach of German fiction authors in the world? In other words, if we assume that the world literary canon is an extension of the national canon in translation, connecting central and peripheral languages, translational sub-communities of authors represent their own category of translational transnationalism.

My findings for peripheral and central language groups in that regard directly respond to Laachir's call for a richer spatial imagination of the world in order to re-evaluate the world literature paradigm by visualizing the translational networks that are oftentimes merely described as "generic meta-categories such as 'world' or 'global'" (Kershaw and Saldanha 2013, 297). While the sub-communities not only highlight the universality of the imagination of the "world" in world literature by mapping the global channels of translations, they simultaneously also call to re-evaluate canonicity and categories of a "global" and "canonical" novel (Reynolds and Vitali 2021, 8). My findings highlight how geographic mapping of the translation network challenges the boundaries of the national categorization of the canon by discussing cases that "no longer conform to the logic of national representation" (Walkowitz 2017, 30), invoking multiple geographies, bridging places in the circulation network, and hence making visible the geography of literary transfer.

Contributing toward the translational shift in German Studies

The question resonating throughout these findings and discussion of the literature on translation in translation studies as well as German studies is: What defines the world literary canon and what is the role of translations in the global circulation of the national canon? Despite not having set out to investigate the national canon, mapping translations led me further into looking at the relationship between the canonical authors and translation. Walkowitz's definition of born translated and the role of translations in global circulation therefore represent the starting point of this thesis, because it purposefully brought translation into the world literature paradigm by using the translational category to challenge the view on single language national vernacular canonical works by its multi-spatial and multilingual features.

While Walkowitz does see translation as the driving force behind world literature canonization, it is still not clear if world literature describes the translational canon as an extension of the national canon, or a different thing altogether, which her concept of born-translated literature implies in that it sets specific authors and works apart. In defining translationalism by linguistic and geographic distribution, following Walkowitz's "born

translated," I raised the hypothesis that specific authors bridge linguistic communities and publishing places despite this imbalance, and, therefore, they shape the network of literary transfer. In following this hypothesis and testing it by mapping the translational space, the findings presented in this thesis showed that within the library catalogue, the most connecting authors and titles are also the most canonical (concentrated) ones. More specifically, the canon appears to connect central and peripheral language groups.

On my journey of identifying translational authors by network analysis and geographic mapping, the overlap and borders of translational and canonical successively became blurred, meaning that they are not clearly distinctive categories. However, by applying the definition of translational, a definition of canonical could also be delivered: connecting centre and periphery.

The underlying hypothesis in my model of descriptive statistics, network analysis, and geographic mapping is that languages connect places by shared authors. Looking at their relationship then revealed the connecting roles of translations as well as the role of authors transgressing the boundaries of the canon by forming sub-communities. The further the analysis moved into the sub-communities away from the canon, the more visible translational authors became in central and peripheral languages. In other words, mapping the canonical space in the library catalogue makes visible the non-canonical spaces of literary transfer.

Coming back to the role of translations in shaping and transgressing "the world" in world literature, we can reply to Wai Chee Dimock's question whether *"the transnational"* is "always symmetrical to the national, a replay of its exclusionary form on a spatially extended register?" (Dimock 2006, 221) by suggesting that translations make visible both the extent of the national canon in a global space and where it is not always symmetrical. We saw this especially in the set of authors only circulated within specific language groups. In the case of translations and their role in driving global canonization, we can see that transnationalism is intertwined with translingualism,[288] which was especially apparent in shared authors across language groups and the role of the national canon in connecting them. However, as bilingual publications showed, though translationalism does not always have to be transnational, it can still invoke multiple scales of geography and language communities. Furthermore, we can ask how translations transgress the national boundaries of the national collection, or as Mani puts

---

[288] Drawing on Helgesson and Kullberg who define translingualism as "the linguistic condition under which writers, publishers, and editors make use of language" (Helgesson and Kullberg 2018, 137), I use this term to define the phenomenon of translations crossing linguistic boundaries and invoking different linguistic dimensions as a result of the "uneven translingual events in which linguistic tensions are manifested either at the micro level of the individual text or at the macro level of publication and circulation—or both" (ibid.).

it, "Was the 'national' library of any European nation, whose acquisitions reached far beyond the political boundaries of a nation-state, necessarily national?" (Mani 2017, 4). The simple answer would be no, obviously, as the role of bilingual and translational works in the catalogue illustrate; however, as I discussed in chapters 1.1 and 1.2, mapping translations visualizes to what extent the catalogue transgresses the national into the transnational. We can even argue that translations to a certain degree challenge the "National" in the German National Library and question how useful the categories of national and transnational, as well as world literature, are in describing translations that clearly go against these fixed concepts. Translations appear to represent both an extension of the national canon on multiple geographic and linguistic scales as well as its own category altogether.

Reflecting on the translational and canonical categories thus raised the question: why are translations not considered part of a national canon, and can we speak of a translational canon instead? Coming back to the claim that what we call world literature is simply a transnational extension of the national canon, world literature hence appears to be simply the national canon *in* translation, or as I call it, a translational canon. In the German National Library catalogue, this translational canon plays a major role in the global circulation of German-language fiction, which underlines the role of the national collection in documenting, but also actively shaping, the national canon in translation. My findings regarding sub-communities of translational authors that signify specific language groups parallel to the translational canon point toward the limitations and challenges of existing categories, especially with these categories dominating literary studies and the lack of a conceptualization of a translational canon based on theoretical frameworks with translation as its own category. In this effect, this thesis is a response to Dimock's claim that "literary studies can significantly transform itself simply by suspending its default (which is to say, nation-based) form of aggregation and engaging its database as material not yet classified" (Dimock 2006, 226), or "data that fall below the line, that fail to rise to the level of national histories and national literatures" (Dimock 2006, 225). Translations without any doubt represent material not yet classified, which supports the premise of this thesis: that translations are not a visible category in the German national collection of the DNB, and thus translations in the DNB catalogue are located at the threshold of national, transnational, and translingual.

In order to work with the data not yet classified, I based my classification of translations on the definition that source and target languages differ: if a text's original edition is in German while also having editions in other languages, it can be identified as a

translation and hence becomes detectable and extractable. As my findings show, besides some exceptions, this definition has proven to be effective in extracting translations from the catalogue and thus, in making them visible in the national collection. However, as I discussed in chapters 1.1, 1.2, and 2.0, there is no consensus and standardized method to create a translation dataset from a library catalogue, let alone apply it across different catalogues. As I outlined in chapters 1.1 and 1.2, the position of translations in digital humanities and literary studies regarding world literature is marginal or limited to case studies, especially for the German context, just as the availability and accessibility of translation data is marginal and limited. The reason why translations take up a marginal role in quantitative literary studies and digital humanities seems to be tied to this methodological and conceptual lack of translations as their own literary category beyond the world, national, and transnational literary paradigms. While some datasets and quantitative studies of translations exist, a framework and model that can be used on translation data in various cultural and literary contexts to date is yet to be developed.

The challenges outlined by Zhou and Sun, Norrick-Rühl, and Vitali and Reynolds that make it difficult to extract translations, such as the inconsistencies and incoherence in different catalogue fields (e.g., Library of Congress as compared to the British National Library as compared to the Index Translationum), can be used to develop a comparative model for mapping translations across catalogues. Even for the German-speaking context, the Swiss, Austrian and German National Library catalogues display vast differences in the way translations are catalogued and their fields are not consistent, which entails that more work will need to be done in order to compare translations across these collections.

By presenting my model, documenting the data quality as well as extraction and providing a definition of translations in the context of the library catalogue, my work goes toward developing a model for the comparative analysis of translations. I hope my definition for the extraction of translation, which is based on languages, and my model, which is based on the relationships between languages and authors, can both fill the lack of quantitative studies of German fiction in translation and help develop a comparative model that can be adapted for other catalogues.

Limitations

My model contains a number of limitations, which are important to discuss in the context of future work:

First, my definition of translation that I rely on to identify and extract translations from the DNB catalogue raises challenges in using it for other catalogues. As mentioned above and in chapter 2.0, metadata standards such as MARC and the Dewey Decimal System are part of harmonization efforts across national library catalogues. In addition to the Dewey Decimal System lacking a field for translated languages, not all catalogues include source and target languages. Both the Swiss and Austrian National Library catalogues do not include a separate field for source and target language, which significantly limits the application of the same method of extraction based on differing source and target languages to their bibliographic data.

In that case, alternative methods need to be developed to extract translations, such as extracting by common phrases that identify a translation, such as "trad. par," "translated by," and "trans. by.". However, since these phrases are in various target languages and are not always contained in the same catalogue field (it can be in the title or notes fields), such a method would require a significant amount of work and experience with the specifics of each catalogue and target language. The same applies to the WorldCat database, where these phrases will often be in the "Notes" or "Responsibility" fields. For VIAF, however, my definition seems to be applicable, since for each original edition, all translations are catalogued by target languages in the "Expressions" field, which makes it most compatible with my definition of translation and thus possible to apply my model with some adaptions to their data structure. As we can see, pointing out these possible limitations in extracting translations from different catalogues in order to develop a comparative model of analysis again points to the precarious position of bibliographic translation data. There is much work to be done before my model can be tested for robustness in a comparative analysis. As I argue above, pointing out catalogue-specific features that limit extracting and analyzing translations can open up our discussion on the (in)visibility of translations in literary studies, library sciences, critical archival studies, and digital humanities.

Secondly, preliminary work has been conducted as part of this thesis project to analyze the role of the publisher by asking to what extent publishers contribute to the concentration of titles and authors. As Sapiro argues, "this brings us to the economy of symbolic goods and field theory. The international book market can be regarded as structured, like the national markets, around the opposition between large-scale and small-scale circulation" (Sapiro 2008, 159), and hence goes into the direction of market-related research of translations which is not the focus of this thesis. Initial questions that were marginally explored within the data on publishers in DNB were: are few publishers driving

the translation circulation? How are smaller publishers represented across the dataset? How does this relate to linguistic diversity? While the results were not included in the final version, they suggested that for each language the title concentration also went in hand with the publisher concentration, meaning that a language with a high title count also has a high publisher count. However, the languages Czech, French, Italian, and Dutch, which, according to a Chi-square test, had unexpectedly fewer publishers as compared to their titles, underline the role of few large publishers such as Éditions du Seuil and Gallimard for language concentration. On the other hand, English, Persian, and Korean had more publishers than expected according to their residuals from the Chi-square test, stressing the importance of smaller languages and English in representing a variety of publishers. Accordingly, publisher size in future work can be an indication of how diverse a language is in terms of publishers. Pierre Bourdieu's study on principal component analysis of the literary field (see Bourdieu and Johnson 1993) also points out the direct correlation between publisher size and language. By applying multiple correspondence analysis, Bourdieu distinguishes seven leading publishers (Gallimard and Éditions du Seuil included). He found that all of them translated predominantly from English, while the smaller publishers almost never did. From here, it is clear that Bourdieu's findings cannot directly be compared with those presented in this study because he only considers French as a target language; therefore, he makes the exact opposite case as this study, in which I look at the target languages with German as the sole source language. However, my findings suggest that less centralized publishers occupy most of the non-European language fields, which is also the case in Bourdieu's study.

Additionally, the role of independent publishers in the field of translation remains to be explored by using the DNB translation dataset compiled in this study building on Bourdieu's field theory in which he makes a similar point: "these small, innovative publishers may not exercise much influence in the field, but they are nonetheless its raison d'eˆtre. They justify its existence, represent its ideals. For this reason, they play a crucial role in its transformation" (Bourdieu and Johnson 1993, 135). Results could then be compared with Pickford's findings, which demonstrate how independent and university press publishers factor into translation, a result which calls for an analysis of a multivariant publisher typology.[289] In future work, my dataset and language model could be used for developing a

---

[289] Sapiro contends this argument by making a point about the specialization of smaller publishers that contributes toward the diversity of the translation market. Bourdieu argues that an analysis of the structural field of publishers is based on the interactions of small- and large-scale publishers, however he "[e]xcluded as well . . . those small- scale publishers who have not yet made a name for themselves and have yet to exercise any real influence in the field (and that are difficult to analyze statistically due to insufficient data)" (Bourdieu and

publisher typology in analyzing their role in language networks of translation.

Third, a longitudinal analysis of the translation data also lies beyond the scope of this project. As I have mentioned in chapter 2.0 and as my findings for title distributions per year suggest, cataloguing practices have an immense influence on which years are over- and underrepresented. Additionally, there is no gold standard of translation rates with which the DNB can be compared to establish if the longitudinal distribution of title counts is representative of the overall fluctuations in the translation market or predominantly cataloguing practices. Since the DNB to date represents the most comprehensive resource for translation data of German fiction, finding a comparable dataset is challenging and consulting market-related datasets and studies might be necessary to conduct a longitudinal study of translations.

Additionally, due to the observed differences in cataloguing practices and their effect on title distributions per year in between the initial data collection in 2017 and the final data collection in 2021, we can expect that the results presented in this thesis are different if the analysis is reproduced at a later point in time. I address the differences in title frequency in chapter 2.0 for the years 2021 and 2019 and show that especially for the years close to the date of extraction, fewer translations will be included in the catalogue data than for the other years. Hence, the results presented here correspond to the state of the catalogue at the time of extraction and therefore represent a snapshot of the field of translations in the DNB. Re-running the analysis as part of future work based on the model presented here can further contribute toward documenting translations and their evolution through cataloguing practices.

Fourth, the role of the translator as compared to authors also has been part of preliminary analysis but has not been pursued further, since it would imply a different set of questions and theoretical groundings. Preliminary results have shown that translators are highly specialized in specific languages and authors. Especially for canonical authors, such as Hesse, we can see that only a few translators translate most of his titles, which suggests that they have a central role in the field of translation and a strong collaboration with specific publishers.

In addition, I also found a number of translators that are authors—406 out of 7,336, which accounts for 5.5% unique translators—and hence play different roles that go beyond

Johnson 1993, 128). He argues extensively in favour of the small-scale publishers and draws attention to their innovative and especially importance for translation field diversity of minor authors and literatures. Heilbron clearly argues that authoritative and established publishers are a central factor to the success of a translation to reach other languages. In that way, the role of bigger publishers cannot be refused, however the data hints at further work being needed regarding their role in the network. Pickford also stresses this point in mentioning the role of university presses, which suggests that publisher typology is another important feature to be explored.

the translation network. Some of these authors publish under multiple pseudonyms (e.g., Clark Darlton aka Walter Ernsting) and some of them are translators of non-German literature and therefore not included in my dataset as translators since I only include outgoing translations. As we can see, the translator field poses various challenges, but also provides a rich source of possible research projects that can profit from the dataset compiled as part of this thesis project.

Lastly, the amount of untranslatables is not measured or addressed in this study, for which however there exists great interest in translation studies (see Walkowitz 2006). Measuring the ratio of translations of German fiction and untranslated works would either require a gold standard that needs to be outsourced (similar to longitudinal studies) or a comprehensive analysis of the complete catalogue. Even though the results presented in chapter 2.0 go into that direction, they only allow for estimating the size of the space that translations occupy in the catalogue. It could be assumed that any work that is not identified as a translation according to the applied definition of source and target language(s) could be an untranslatable. Hence, further definition of what constitutes an untranslatable within the DNB catalogue is required, either in terms of authors (the percentage of works translated versus not translated for a given author) or by titles (the percentage of titles not translated per language). As with the previous points, the role of untranslatables is a fruitful topic for discussion for future studies on the DNB, which can draw on my dataset.

In outlining these limitations, I hope to provide pointers to future research that could emerge from this thesis project and contribute toward designing a comparative translation research model and framework. By providing the DNB translation dataset and model of analysis, I hope to open up this rich resource to the scholarly community.

Conclusion

Let us come back to the initial starting point of this thesis, the question as to why some authors are representative of a national canon and translated while others are not. While my project did not bring me closer to the question as to why, it did enable me to identify the representative authors of the national canon in translation and discuss their roles in the language network by focusing on the question of whose, where, and into which languages translations are published.

My results show that within the German National Library catalogue, this imbalance—I use the term concentration—is visible in specific languages that take a central position in the network of transfer. From a critical archival perspective, we could also argue that the catalogue's focus on few central languages lets us gain insight into the literary cultures the catalogue is invested in shaping, and therefore contributing to the creation of the world literary space.

By applying methods of quantitative analysis, my results also reflect my initial observation that there is an existing imbalance toward a small set of authors that have an increased number of translations. Regarding my question about the role of canonical and of non-canonical (translational) authors and my hypothesis that authors connect linguistic communities in geographic space, I found that while the canonical and translational author fields appear to overlap, meaning that authors with high concentration are also distributed across many languages, the author's roles are dependent on how they connect specific language communities. The role of canonical authors, according to my findings, is to connect peripheral and central language groups, while translational authors are highly connected within sub-communities of languages.

Going back to my central hypothesis, we can see that not only increased concentration sets canonical authors apart, but also the language communities that these authors connect. In addition, my findings indicate that their role depends not only on whether they connect languages, but also which language communities they constitute channels for according to their positions in the translation network. Thus, language position, in addition to publishing centres, are crucial indicators of author communities which are transgressing national boundaries. Moving away from highly canonical authors allowed the exploration of these smaller author communities that exist within language groups and, as I argue, form literary communities, as was visible especially in peripheral languages which appear spatially

specialized as well as specialized in specific genres.

In summary, from my findings, we can distill three arguments regarding the role of translations in the national and global canonization of an author or work of fiction. First, translations in the library catalogue make visible the global canonization dynamics within the framework of world literature for which translations are a pre-condition for its global circulation. We can hence define the world literary canon as the national canon *in* translation. Secondly, the canon in translation is characterized by its connecting function of central and peripheral language groups, which further specifies the relationship of the national with the transnational and the canonical with the translational. And third, the roles of sub-communities of translational authors challenge our understanding of a world literary canon in its concept of "the world" in that translationalism extends far beyond this set of authors which we categorize as world literature.

In this thesis project, I propose a model to analyze the underlying dynamics in the translation network, which consists of methods drawn from social network analysis and geographic mapping to visualize how literary fields are connected by people, places, languages, and works according to the DNB catalogue. The initial aim was to develop a model to map these relationships in order to gain a deeper understanding of the national canon in translation, translationalism, and authors that transgress linguistic and geographic boundaries and hence challenge these classifications.

While the starting point of data exploration, with its focus on measuring concentration, aimed at testing for the imbalance of specific languages and authors mentioned in previous studies of translations, it was limited in visualizing the connections between languages. By applying a language model (following Heilbron), it was possible to use the variables contained in the German National Library's bibliographic data and its metadata fields and construct the network with the languages as nodes and authors as edges. The value of this approach is that it is possible to detect author communities and further identify the author's roles in connecting linguistic communities.

Geographic mapping then further served to visualize the networks of translation transfer and highlight publishing centres and their role in linguistic communities. It was a crucial step in visualizing the connections by languages and superimposing the observed structure of the social network graphs in geographic space. The resulting maps visualize how the network transgresses national boundaries via language communities that are connected by authors; in addition, the maps highlight the important role of bilingual publications in the global translational transfer, which has not been visible in the previous network graphs. Even

though geocoding and data cleaning involved several additional methods such as fuzzy matching, geographic maps of publishing centres offered a rich method to look beyond categories, such as national versus world literatures, by putting the world on the map and showing that it is highly concentrated in few publishing centres and therefore not as worldly as often assumed. Mapping also provides the necessary context to understand what the world consists of when speaking of world literature. This is where the act of making translations visible can significantly contribute toward rethinking existing paradigms in literary studies; by doing so, translations emerge as an invaluable analytic category that has so often been overlooked in German studies.

Besides making translations visible, I see the strength of my contribution to the fields from which quantitative translation research emerges in the methodological exploration and the design of a model tailored to bibliographic translation data extracted from the DNB, as well as in making the dataset available to the research community. In that sense, the research output in the form of the dataset tables and the R scripts has an equally important role as discussing the results.

Regarding the limitations I listed in the discussion chapter as well as my introduction, there are several directions future work can take when building on my model and dataset. First, behind the three-fold model with mixed methods are several additional statistical tests and trials that either replicated results or did not deliver the results responding to the research questions.[290] While numerous studies in quantitative translation research offer various methods and models, I aimed at exploring methods that have not been exploited in terms of bibliographic translation data, such as centrality measures in the language-author network and geographic mapping of linguistic routes of transfer through translation. It was important for me to stay as close as possible to the catalogue; therefore, I limited my use of resources external to the catalogue to geodata in order to make it possible to apply my model to different datasets extracted from the DNB catalogue and to account for reproducibility.

Secondly, in order to develop the model tailored to the specifications of the catalogue, it was necessary to formulate a definition based on which the translations could be extracted, which followed the condition that translations have an original language (which for all cases is German) and a different language as a target. This definition and designed query to find and extract translations of German fiction has proven to be an efficient method to compile the

---

[290] The results of additional tests can be found in the appendix.

dataset for analysis.

However, the extraction method and model are limited to the specifics of the DNB metadata fields and structure, and adapting them to other catalogues in order to develop a model for comparing translations across different national collections will require additional steps. However, once extracted, my model can be applied to bibliographic datasets that include the main variables of analysis (author name, target language, publishing place). Future work can therefore test my model with other national library catalogues in order to develop a comparative framework to map translation across different national collections.

Additionally, as the date of extraction has a significant influence on the data completeness, future work needs to be conducted on the effect of cataloguing practices on the analysis's results. The general trend appears to be that titles are added retroactively to the catalogue, which results in fluctuations for specific years. However, once a certain time has passed, those fluctuations appear to stabilize, meaning that only the years closest to the extraction date will have significant changes in catalogued titles. Since metadata standards are (usually) not subject to change within the catalogue, it is possible to re-run the same analysis and model on a new dataset. For future projects, this would make it possible to track how the DNB catalogues translations over the course of the next years and to construct a timeline of "Momentaufnahmen," such as the one presented in this thesis project.

Lastly, as stated in my discussion, the role of the publisher and the translator as well as a longitudinal analysis of translations lies beyond the focus of this study. Considering that there were over 80 Russian publishers in Berlin prior to the war (Schiff 2001, 25) which are not considered part of the national literature of German, but who published important writers such as Nabokov in original and in translation, such factors offer an interesting starting point for future research into the publishing landscape that challenges our understanding of a national literature category. For example, Nabokov is a representative author of the Russian national literary canon who spent a significant portion of his life in Berlin with his wife Vera translating his works, while also being part of a large network of Russian authors and publishers. Following this interesting case, future work should also address the role of bilingual, multilingual authors who write in several linguistic communities (e.g., Yoko Tawada, Emine Sevgi Özdamar) and what their position according to the catalogue represents in reflecting translationalism, hence building on studies by Ivanovic and Seidl and Walkowitz.

To conclude this thesis, it is important to mention that this project is the result of a fruitful

collaboration with the German National Library's Datenshop. I cannot stress enough the importance of researchers to work together with the people responsible for curating, maintaining, and constantly improving the library catalogues. As this project illustrates, a catalogue is a complex construct, highly dependent on the local cataloguing practices and the context of collections contained within the catalogue. While moving toward harmonizing catalogues by metadata standards like MARC and the Dewey Decimal System, it is also important to document the information that helps reconstruct and document the archive's historical and cultural connotations that are buried in the bibliographic data. Despite metadata harmonization efforts, we as scholars of literature should work toward capturing and preserving the archival history contained in our national collections. Hence, I believe that the geographic shift and translational turn in digital humanities need to go in hand with the developments in cataloguing practices and respond to the changing nature of our subjects of analysis.

Appendix[291]

Chapter 2.0
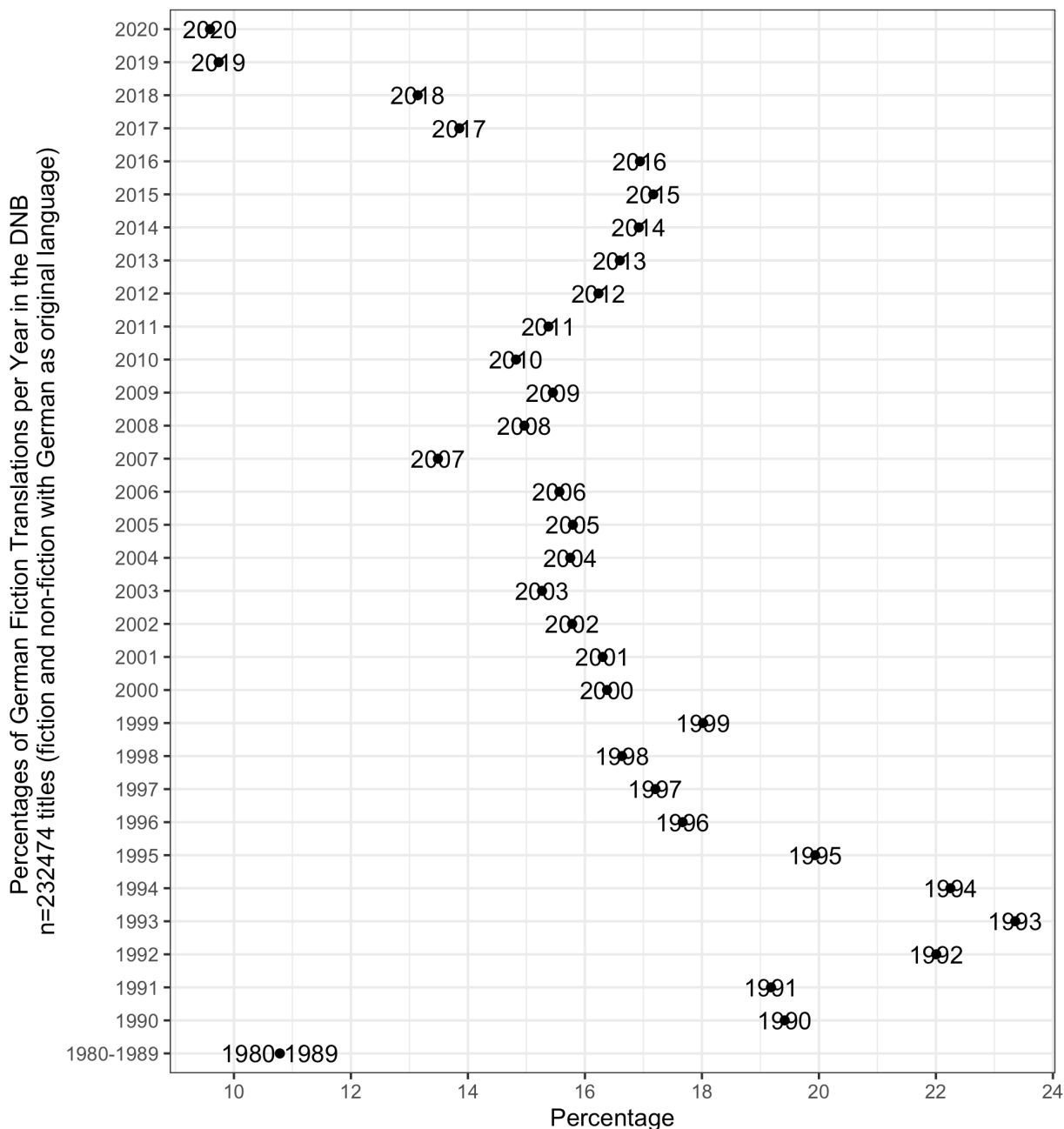
Figure A.1. Title percentages per year (1980 until 2020) from the German National Library (all translated works with German as the original language). 1980 until 1989 are combined due to the low number of titles (under 500). Data extracted: Data extracted: April 15th, 2021 File: figure_ch1.2_dnball_percentages_fiction_v2.png

---

[291] Files of additional figures and tables are in the following Github repository: https://github.com/lisateichmann/Mapping-German-Fiction-in-Translation

Chapter 2.1

Distribution of languages for title sums with mean (418.2791)
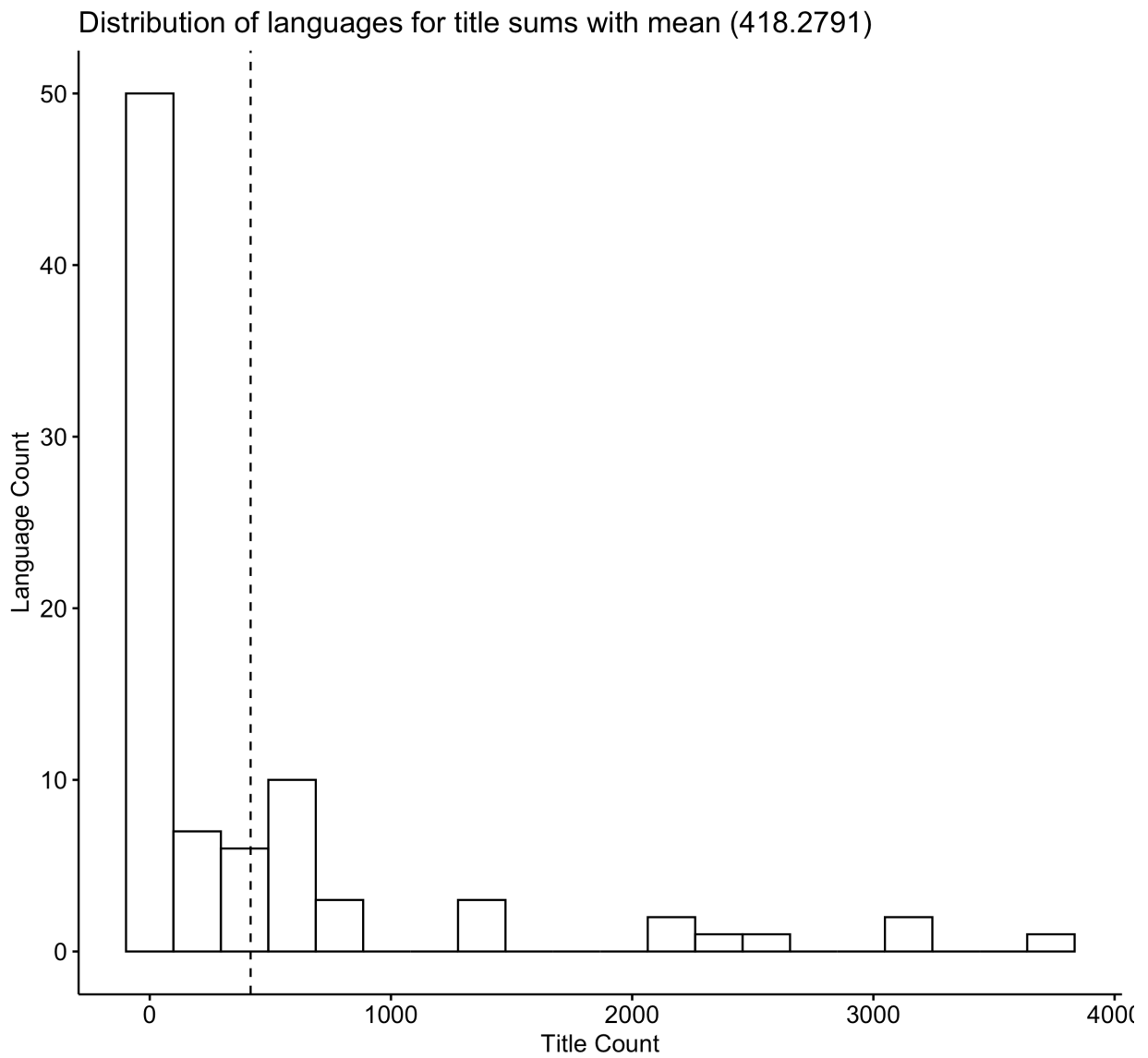
Figure A.2. Distribution of languages for title sums (n=35972) with mean. File: figure_ch1.2_dnball_alllanguages_freq_meanhist_v2.png
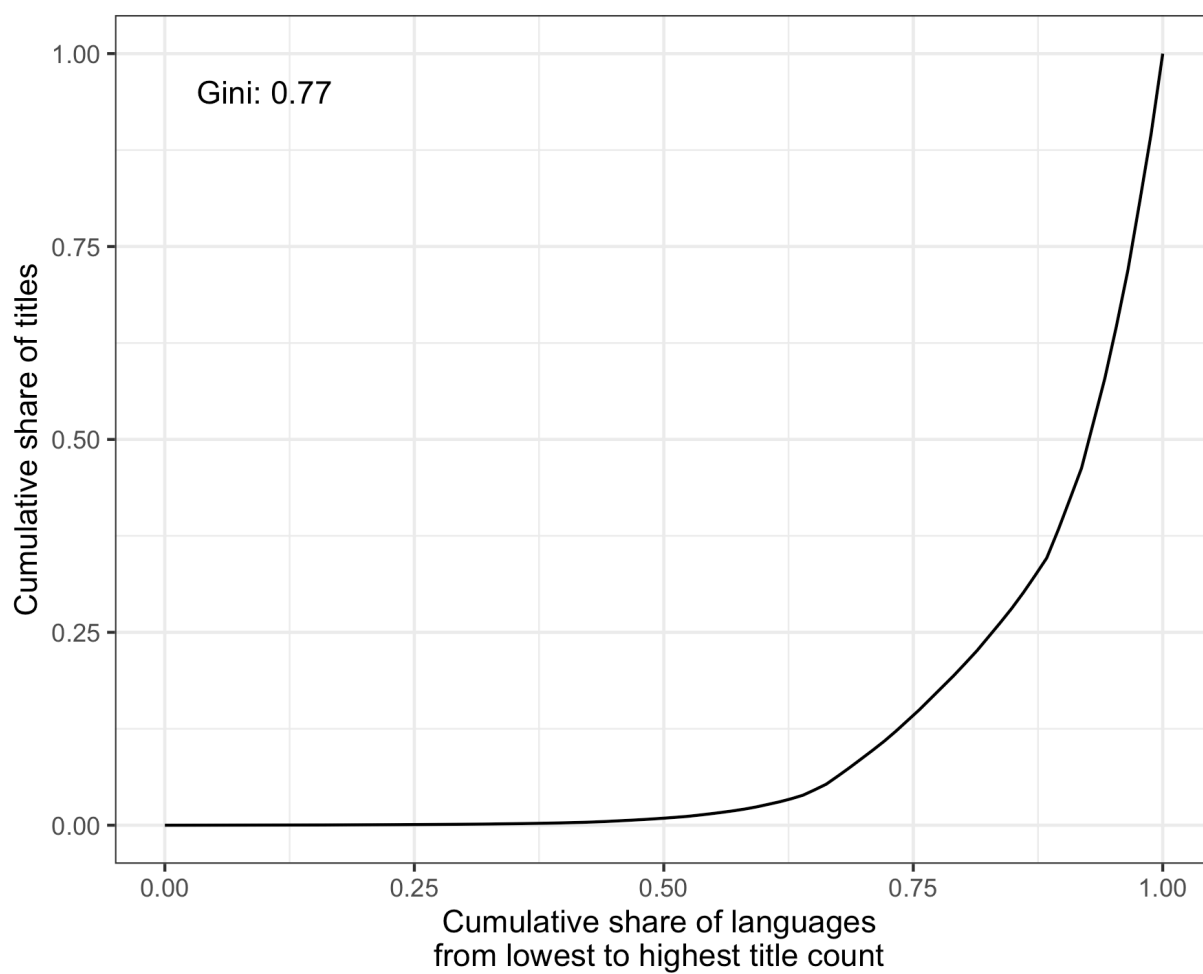
Figure A.3. Gini-coefficient for title counts per language (n=86) extracted from the German National Library catalogue. Filename: figure_ch2.1_dnball_language_title_gini_v2.png
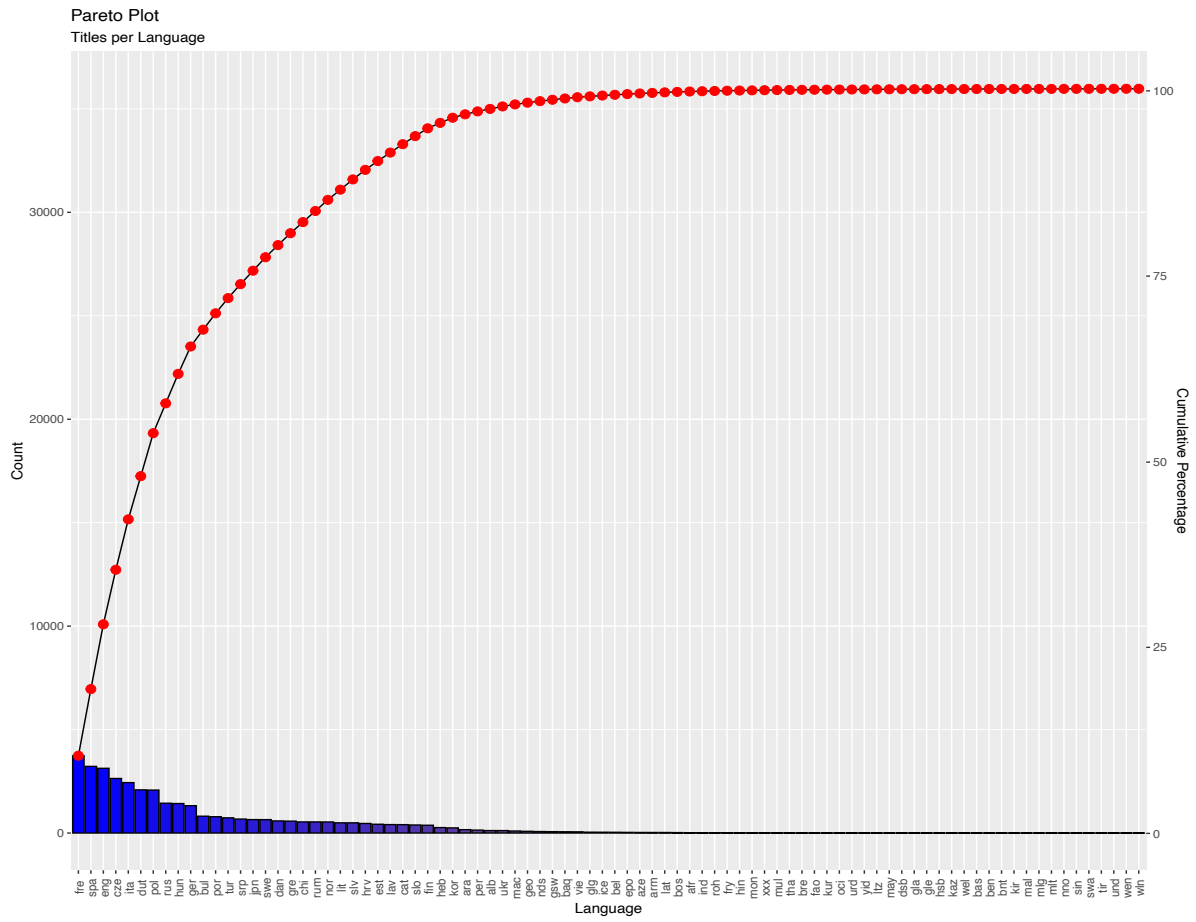
Figure A.4. Pareto chart of languages and title counts. File:
alldnb_language_title_pareto_ch2.1_v2.csv.pdf

Language percentages for the top 20 languages in the
National Library Translation Dataset (1980-2020)
From all titles (n=35972)
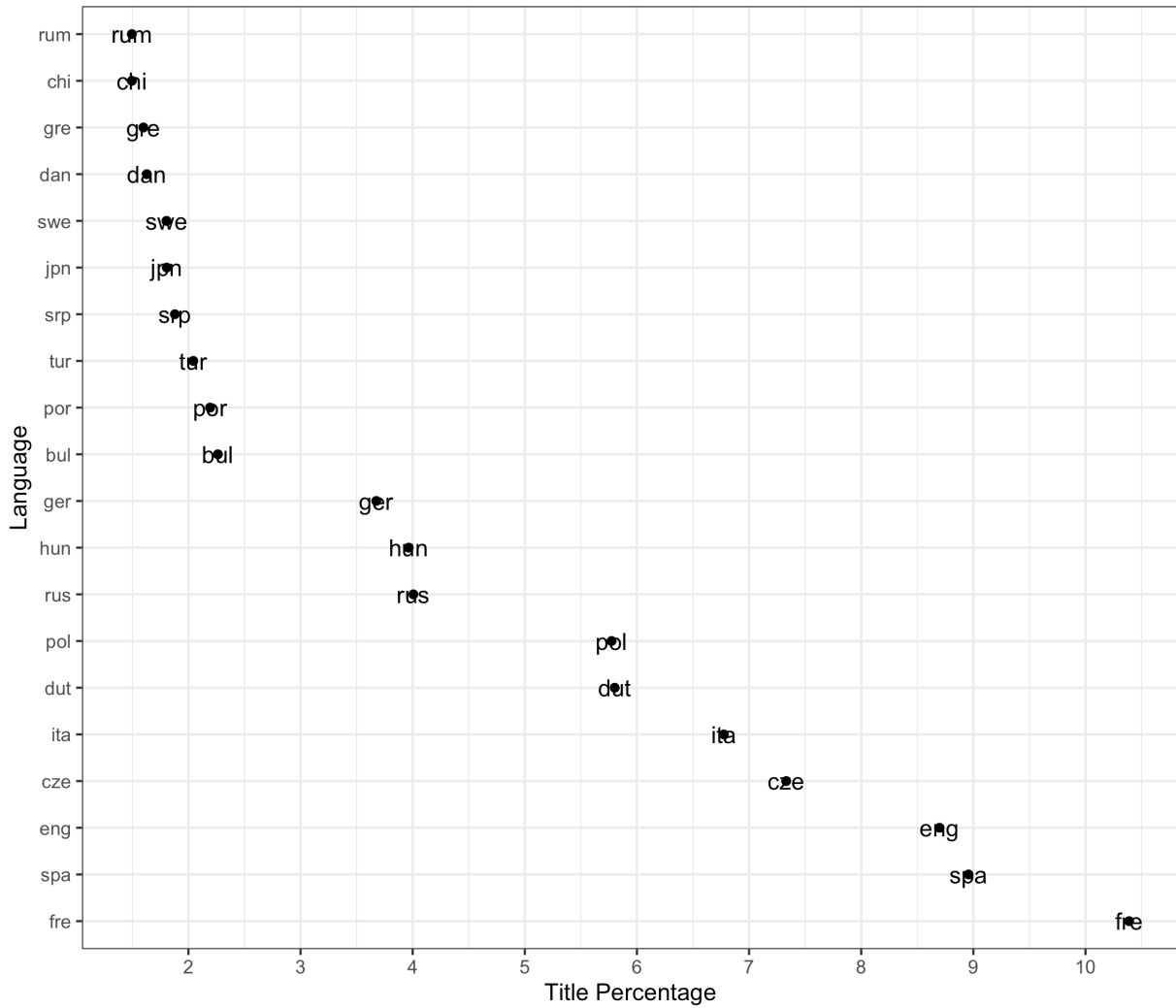
Figure A.5. Percentages of translated titles per language for the top 20 languages in the German National Library catalogue. See appendix for the full table with values. File: figure_ch1.2_dnball_language_percentages_top20_scatterplot_v2.png
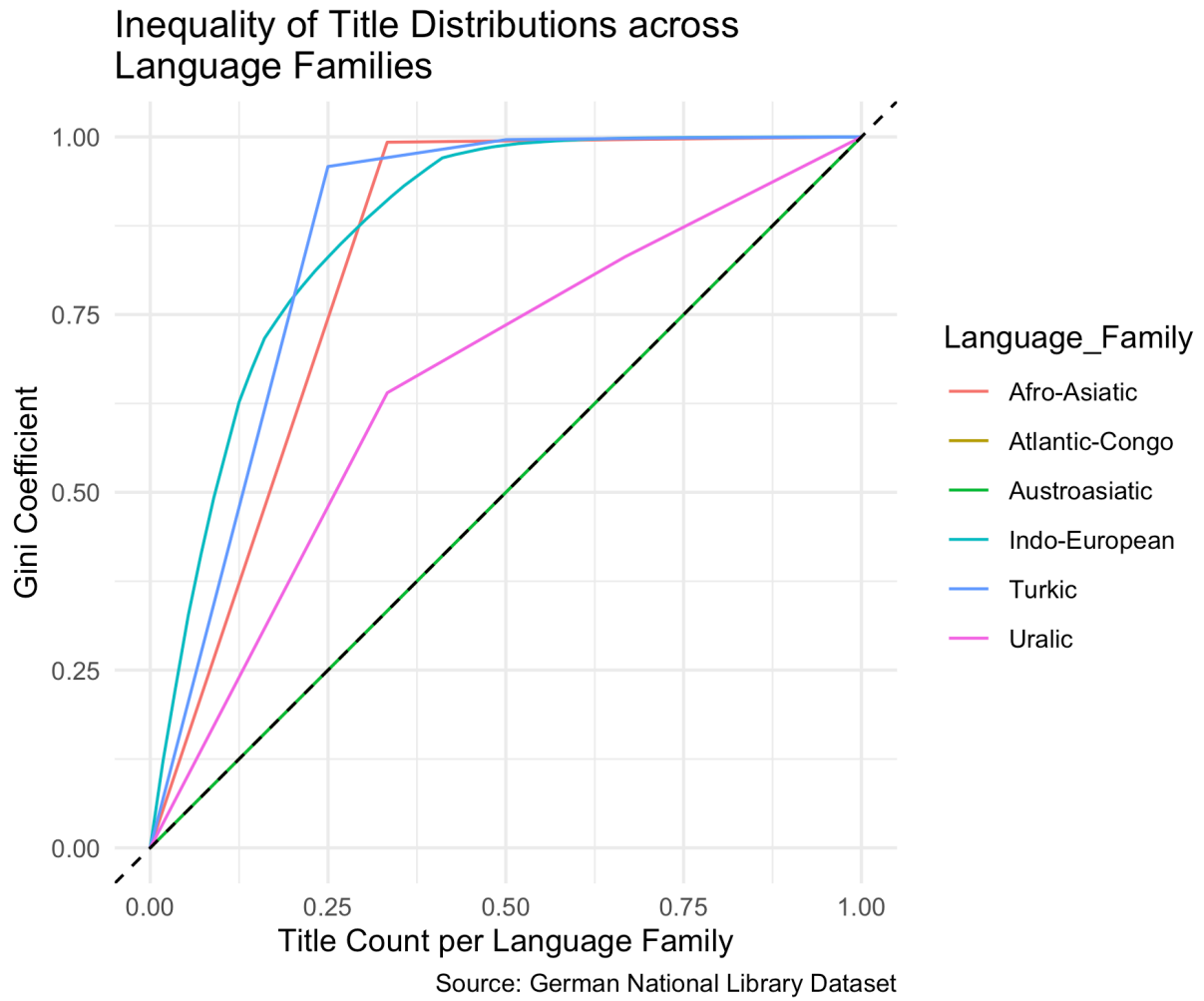
Figure A.6. Gini-coefficient for title counts per language families extracted from the German National Library catalogue. File: figure_ch1.2_dnball_gini_lorenz_title_langfam.png
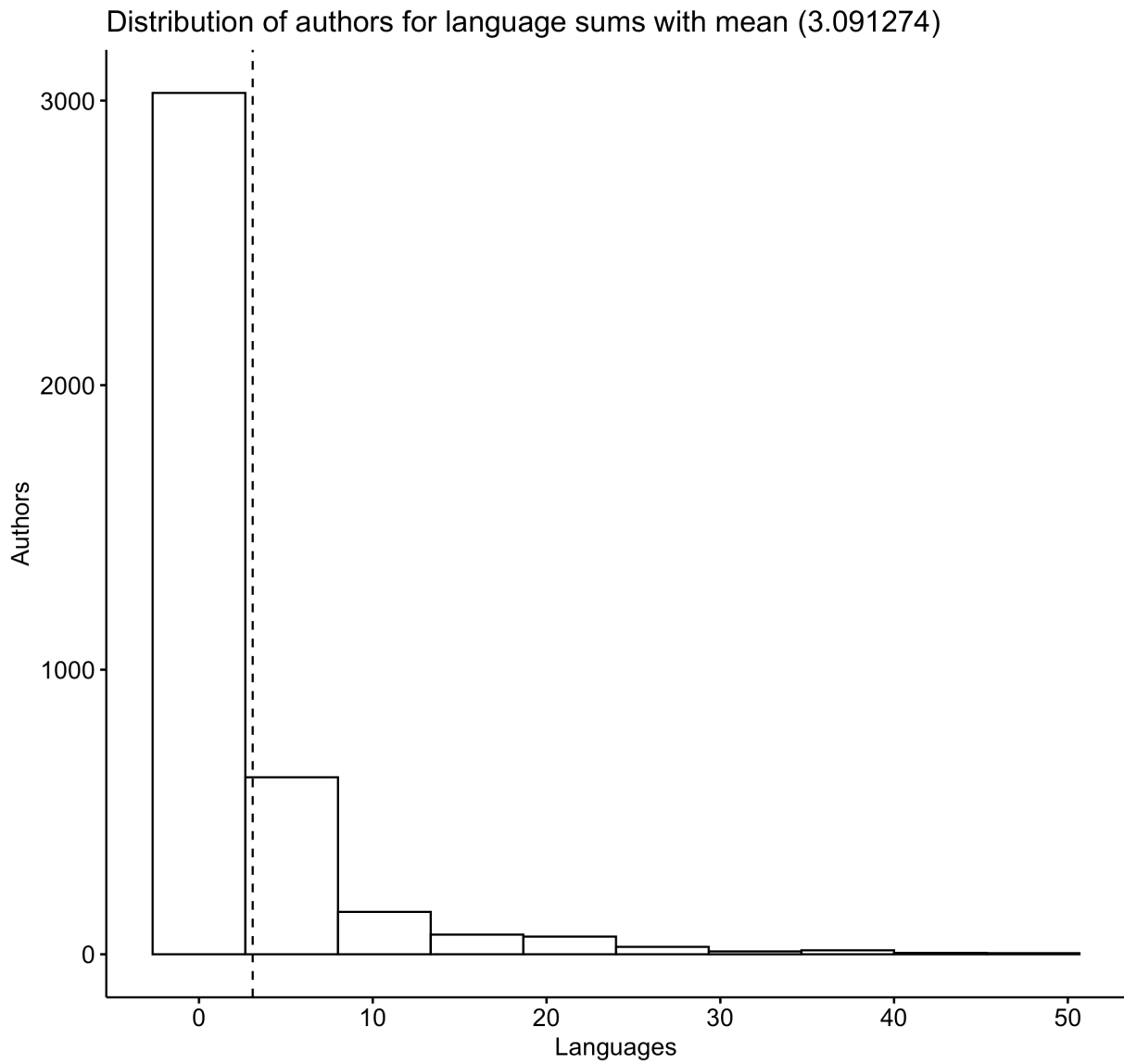
Figure A.7. Distribution of authors for language sums in the German National Library catalogue. File: figure_ch1.2_dnball_allauthors_freq_meanhist_v2.png

Figure A.8. Gini-coefficient for language count per author. File: figure_ch2.1_dnball_language_author_gini_v2.png

Figure A.9. Language sums of translations (raw count of titles) per author (top 20). File: figure_ch1.2_dnball_top_authorlang_count_distribution_v2.png

Top 20 Authors with title sums of translation in the
National Library Translation Dataset (1980-2020) (n=35972 titles)
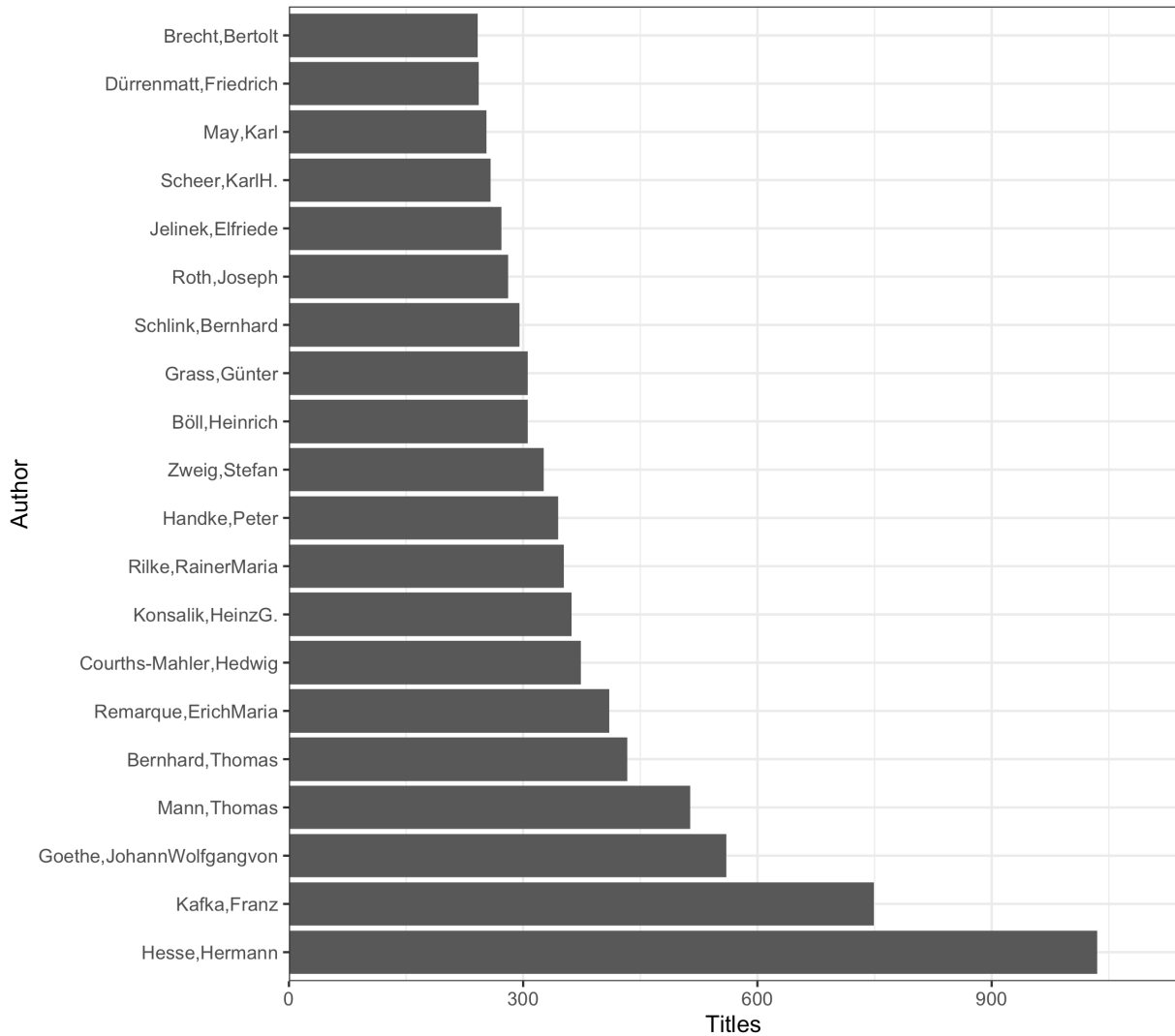
Figure A.10. Title sums of translations (raw count of titles) per author (top 20). File:
figure_ch1.2_dnball_topauthors_distribution_v2.png

uniform.title	language_count	title_count
Siddhartha	37	91
Der Vorleser	34	66
Der Steppenwolf	31	81
Das Parfum	30	71
Faust	30	133
Die Vermessung der Welt	29	35
Atemschaukel	27	32
Demian	27	59
Die Klavierspielerin	27	41
Benedict Wells Vom Ende der Einsamkeit	24	25
Er ist wieder da	24	31
Das Lavendelzimmer	23	26

Der Prozess	23	104
Die Blechtrommel	23	42
Die Mittagsfrau	23	31
Franz Kafka Der Prozess	23	29
Im Krebsgang	23	25
Adler und Engel	22	24
Das Glasperlenspiel	22	50
Die Heimkehr	22	28

Table A.1. Top titles ranked by language count versus title count. File: alldnb_uniformtitle_count_perlanguage_ch2.1_v2.csv

Chapter 2.2

Language Network by shared authors
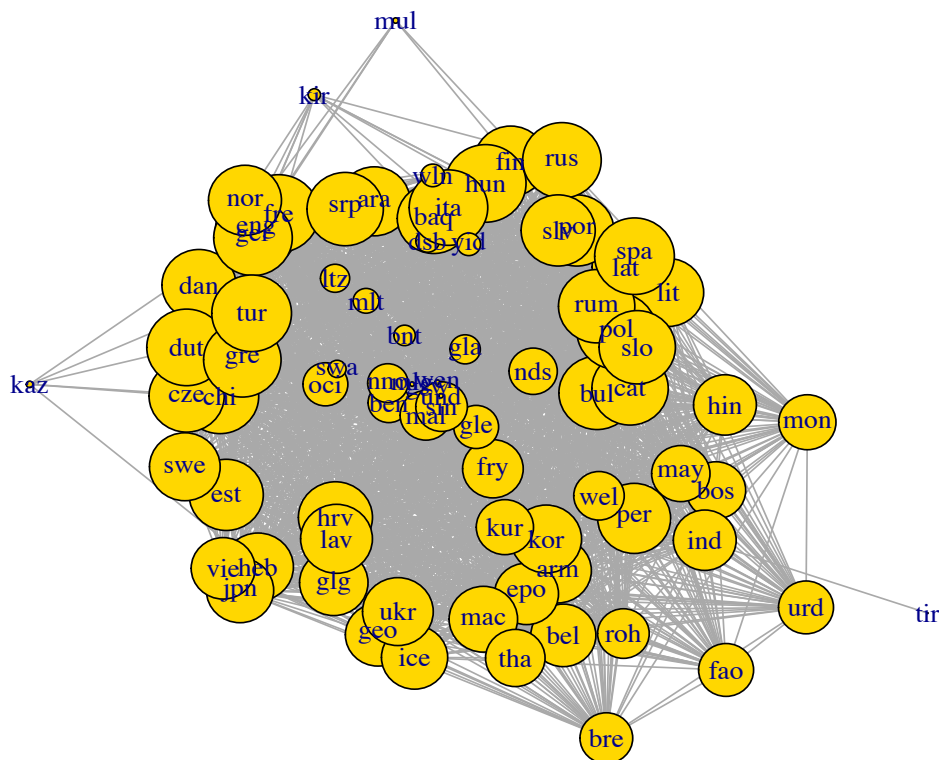
Figure A.11. Network graph with node size for degree. File: lang_lang_author_network_degree.pdf
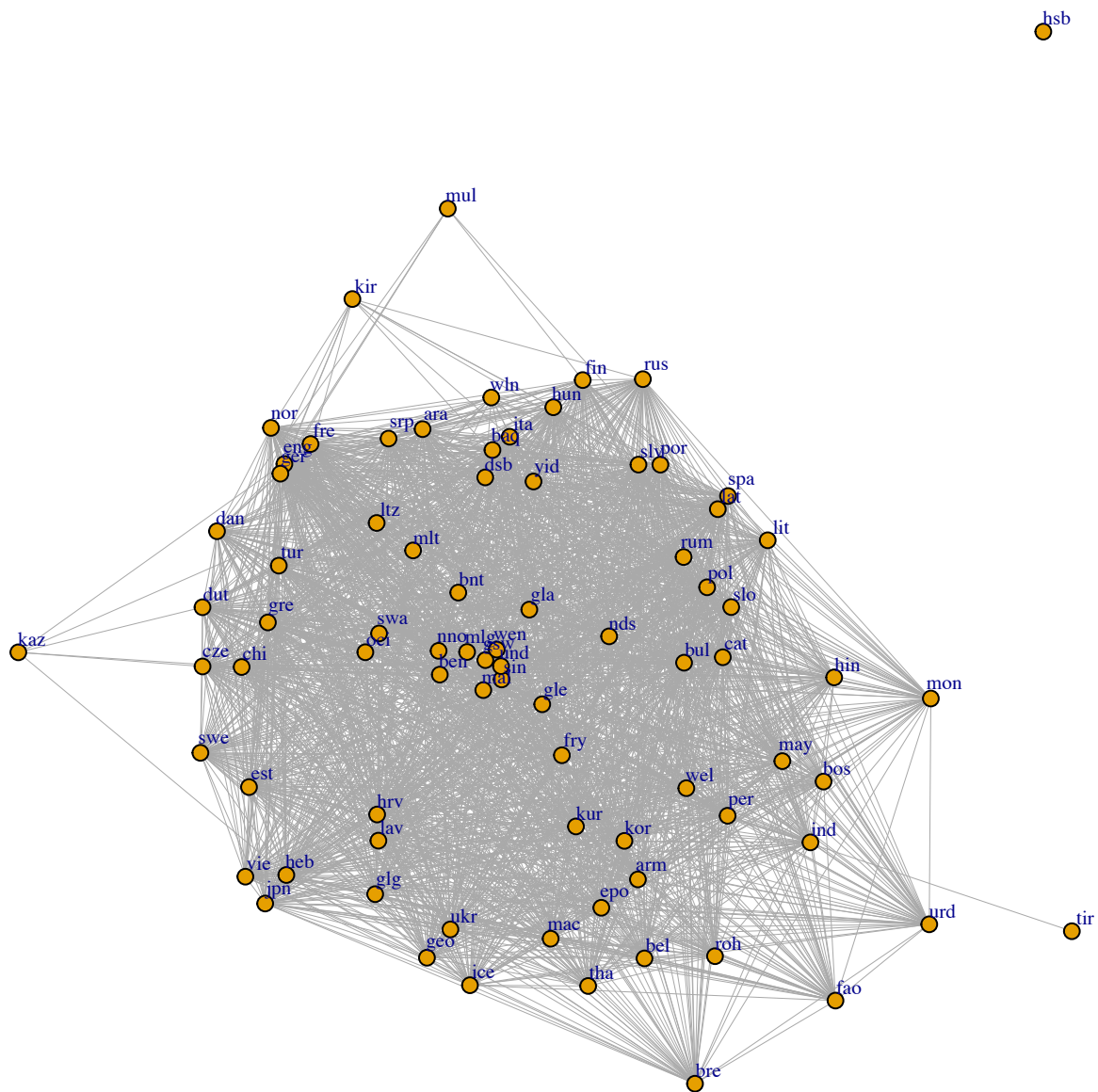
Figure A.12. Network graph without node size. File: lang_lang_author_network_all_kamada.pdf
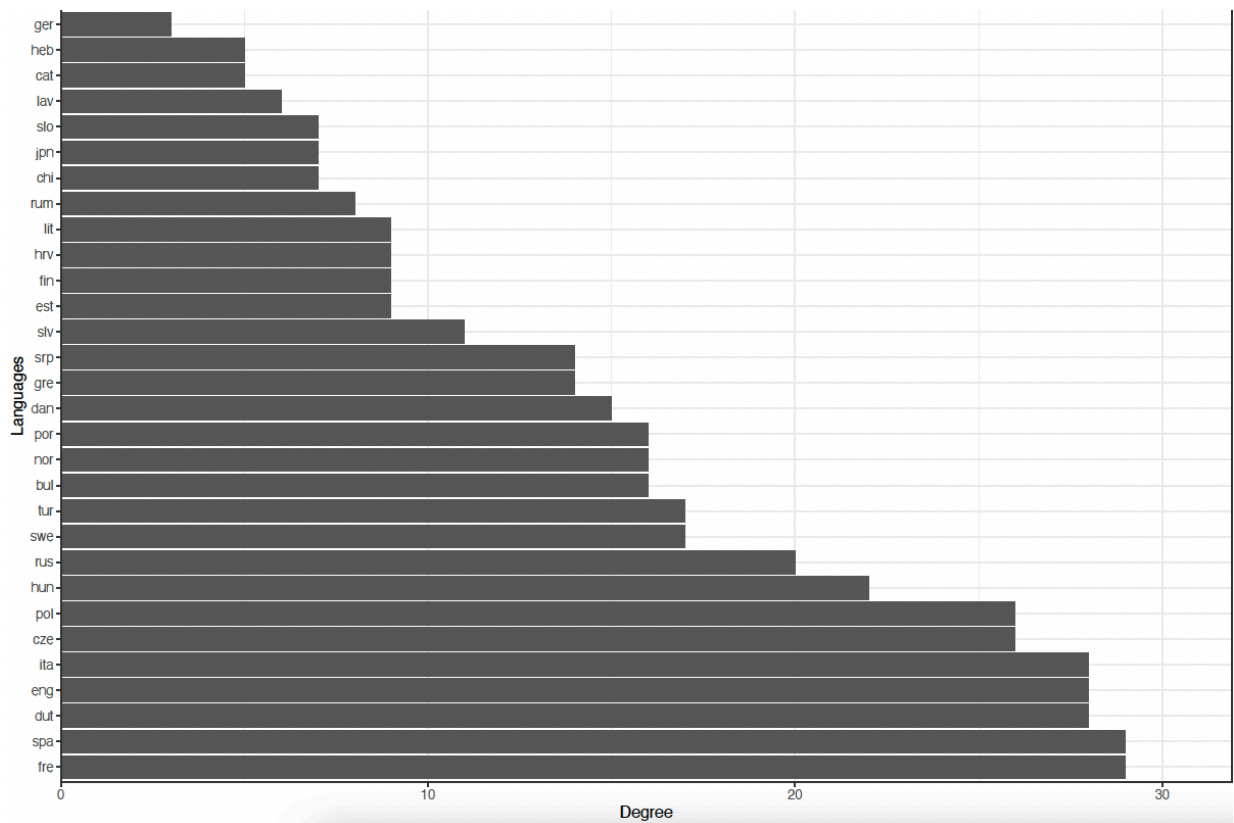
Figure A.13. Degree distribution for a subset (>100 edges per language). File: alldnb_lang_lang_author_above100_degree_ch2.2_v2.pdf

**Language Network
by shared authors**

Betweenness Centrality
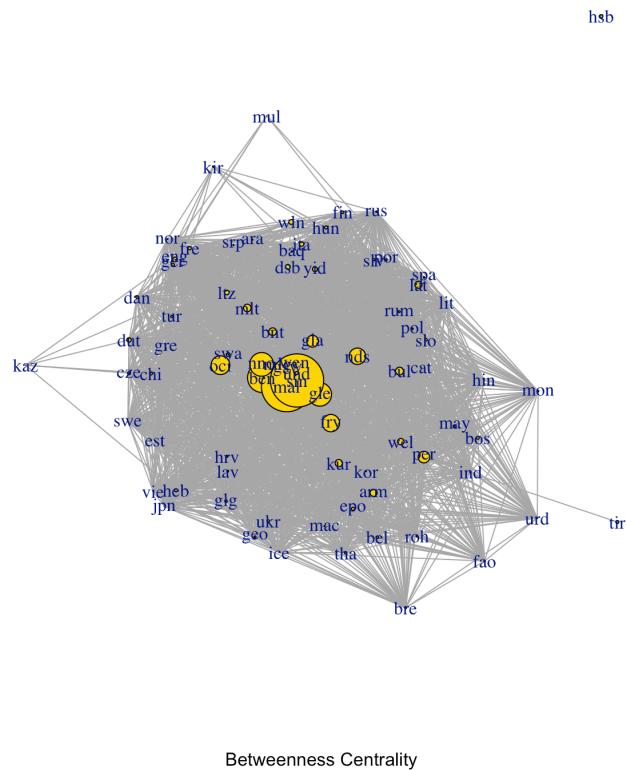
Figure A.14. Network graph of betweenness centrality. File:
alldnb_lang_lang_author_betweenness_ch2.2_v2.png

Language	Degree
eng	73
fre	72
hun	72
dut	71
pol	71
spa	71
tur	71
ger	70
ita	70
rus	70
chi	69
gre	69
bul	68
cat	68
rum	68

Table A.2. Degree per node (number of edges) for top 15 languages. File:
alldnb_lang_lang_author_degree_ch2.2_v2.csv

Language	Eigenvalue
fre	1
spa	0.95238637
ita	0.93034589
eng	0.90034392
dut	0.87089775
cze	0.81373596
pol	0.77008454
rus	0.65118198
hun	0.64273088
tur	0.5702413

Table A.3. Eigenvector centrality score for top 10 languages. File:
alldnb_lang_lang_author_eigenvec_ch2.2_v2.csv

Language	ISO code	Betweenness
Singhalesisch	sin	359.68139
Malayalam	mal	341.680402
Bengali	ben	199.042609
Nynorsk	nno	160.437525
Irisch	gle	158.10636
Okzitanisch	oci	126.648065

Friesisch	fry	117.496521
Niederdeutsch	nds	110.925558
Gälisch-Schottisch	gla	79.2485078
Bulgarisch	bul	52.9413281

Table A.4. Betweenness centrality score for top 10 languages. File: alldnb_lang_lang_author_eigenvec_ch2.2_v2.csv

	deg	eigen	close	bet	deg_g2_above 100	eigen_g2_above 100	close_g2_above 100	bet_g2_above 100
deg	1	0.68	0.68	0.46	0.1	0.68	-0.02	0.07
eigen	0.68	1	0.35	0.82	0.25	1	-0.03	-0.1
close	0.68	0.35	1	0.28	0.21	0.35	-0.06	0.03
bet	0.46	0.82	0.28	1	0.39	0.82	-0.09	-0.15
deg_g2_above100	0.1	0.25	0.21	0.39	1	0.25	-0.08	-0.59
eigen_g2_above100	0.68	1	0.35	0.82	0.25	1	-0.03	-0.1
close_g2_above100	-0.02	-0.03	-0.06	-0.09	-0.08	-0.03	1	0.07
bet_g2_above100	0.07	-0.1	0.03	-0.15	-0.59	-0.1	0.07	1

Table A.5. Degree, eigenvector, betweenness, closeness centrality correlation coefficients. File: alldnb_lang_lang_author_centralities_correlation_ch2.2_v2.csv

Author	degree(int)
Süskind,Patrick	798
Böll,Heinrich	794
Mann,Thomas	794
Goethe,JohannWolfgangvon	792
Kafka,Franz	788
Rilke,RainerMaria	788
Dürrenmatt,Friedrich	784

Jelinek,Elfriede	777
Hesse,Hermann	775
Zweig,Stefan	774
Frisch,Max	771
Grass,Günter	767
Schulze,Ingo	762
Handke,Peter	753
Bernhard,Thomas	725
Hoffmann,E.T.A.	714
Musil,Robert	711
Schnitzler,Arthur	702
Enzensberger,HansMagnus	696
Walser,Martin	690
Kehlmann,Daniel	689
Schlink,Bernhard	687
Walser,Robert	684
Zeh,Juli	680
Erpenbeck,Jenny	668
Hölderlin,Friedrich	666
Bachmann,Ingeborg	665
Celan,Paul	663
Müller,Herta	663
Wolf,Christa	658
Schirach,Ferdinandvon	654
Lark,Sarah	648
Noll,Ingrid	633
Vermes,Timur	631
Remarque,ErichMaria	625
Brecht,Bertolt	625
Lenz,Siegfried	623
Schiller,Friedrich	618
Sebald,W.G.	618
Werfel,Franz	609

Fontane,Theodor	600
Safier,David	594
Ransmayr,Christoph	593
Feuchtwanger,Lion	589
Canetti,Elias	586
Meyrink,Gustav	580
Neuhaus,Nele	579
Roth,Joseph	576
Bürger,GottfriedAugust	566
Mercier,Pascal	565
Hauptmann,Gerhart	562

Table A.6. Author edges by intersection between central and peripheral language groups (top 50). File: alldnb_common_authors_centralandperipheral_ch2.2_v2.csv

hun	eng	eng_hun_match	author
0	95	greater	Jopaire,Sandrine
12	66	greater	Kafka,Franz
11	53	greater	Goethe,JohannWolfgangvon
24	47	greater	Mann,Thomas
4	47	greater	Brecht,Bertolt
12	41	greater	Bernhard,Thomas
7	41	greater	Böll,Heinrich
33	40	greater	Hesse,Hermann
11	29	greater	Rilke,RainerMaria
1	29	greater	Roth,Joseph
2	27	greater	Handke,Peter
3	26	greater	Schnitzler,Arthur
0	25	greater	Durst-Benning,Petra
6	23	greater	Zweig,Stefan
7	22	greater	Grass,Günter

Table A.7. Authors that have a significantly higher title count in English than in Hungarian. File: alldnb_lang_lang_author_degree_match_hunvseng_ch2.2_v2.csv

| hun	eng	eng_hun_match	author
156	0	lower	Courths-Mahler,Hedwig
30	0	lower	Birkner,Friede
23	19	lower	Remarque,ErichMaria
22	4	lower	May,Karl
12	0	lower	Fischer,MarieLouise
11	10	lower	Hoffmann,E.T.A.
11	2	lower	Kästner,Erich

Table A.7. Authors that have a significantly lower title count in English than in Hungarian. File: alldnb_lang_lang_author_degree_match_hunvseng_ch2.2_v2.csv

Chapter 3
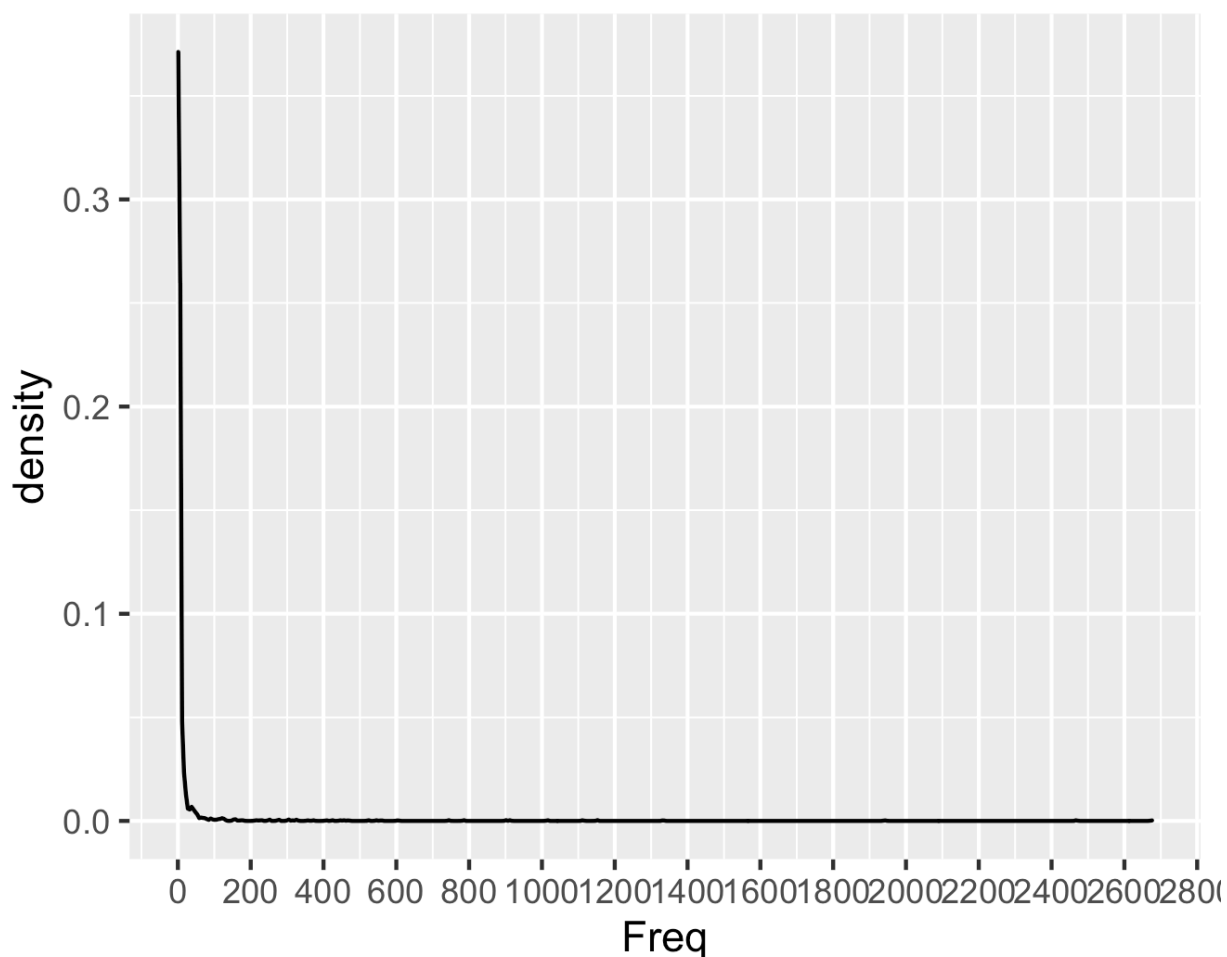
Figure A.14. Density graph of titles per publishing place. "Freq" stands for title sums. File: 20220607_alldnb_ch3_author_pub_place_title_density_v2.png
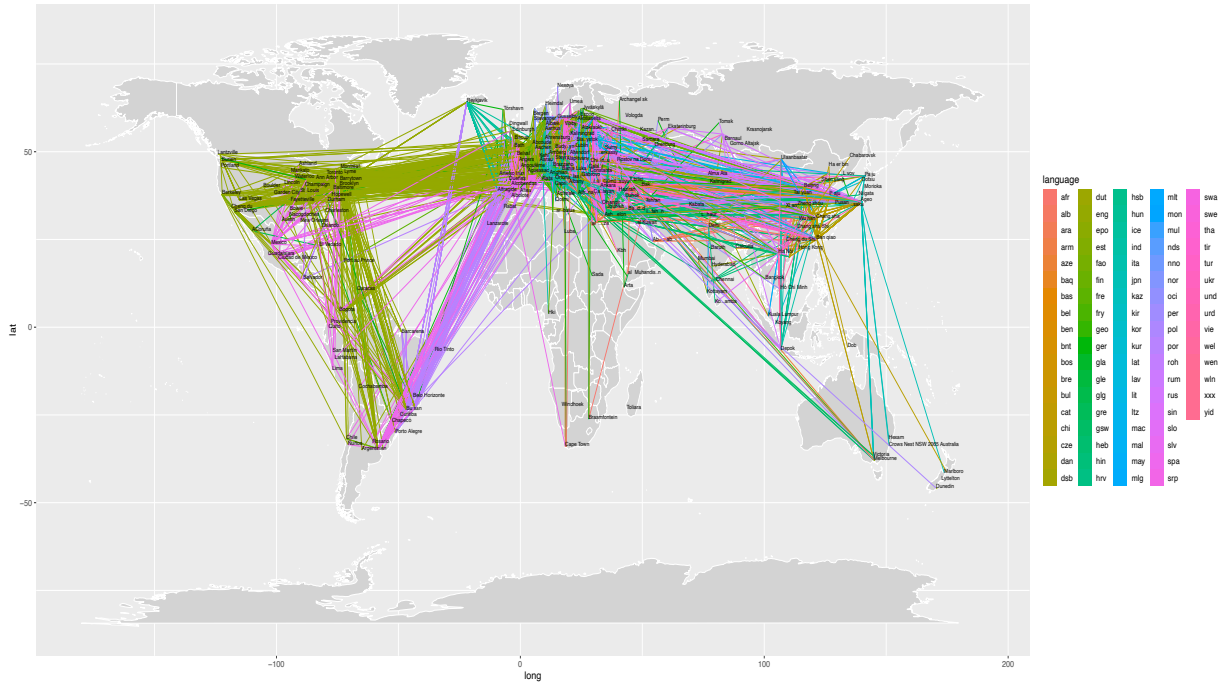
Figure A.15. Publication places connected by lines of shared authors. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. File:
20220614_alldnb_ch3_author_pub_place_geo_map_lines_languages_places_v3.pdf
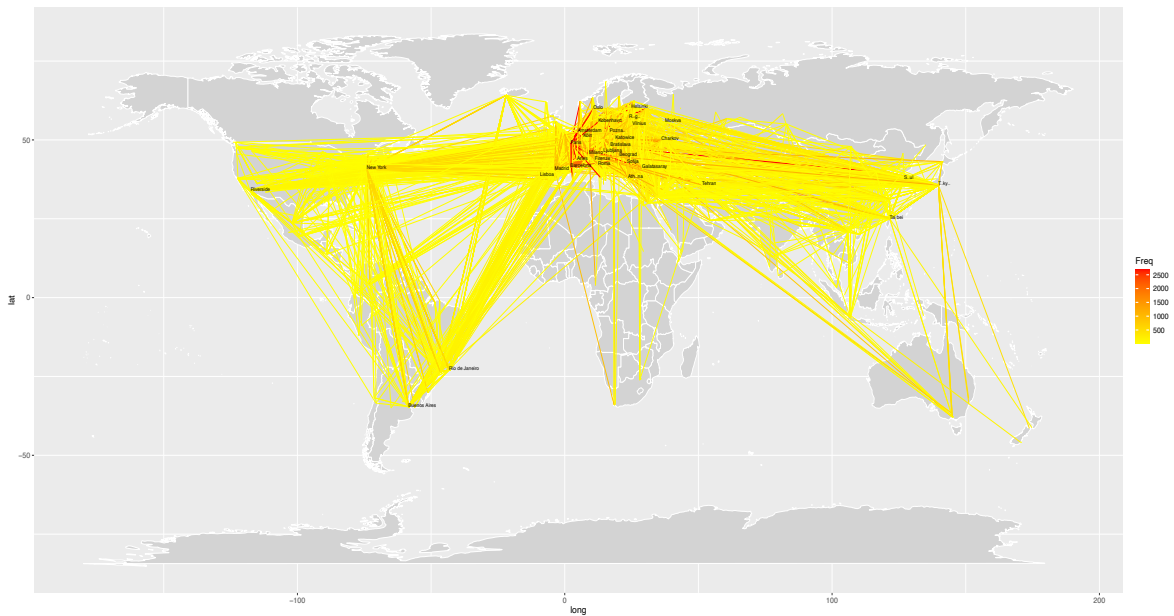
Figure A.16. the 16 most concentrated cities and their number of titles (see legend). Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. File:
20220602_alldnb_ch3_author_pub_place_geo_map_lines_topconnected_colored_v1.pdf
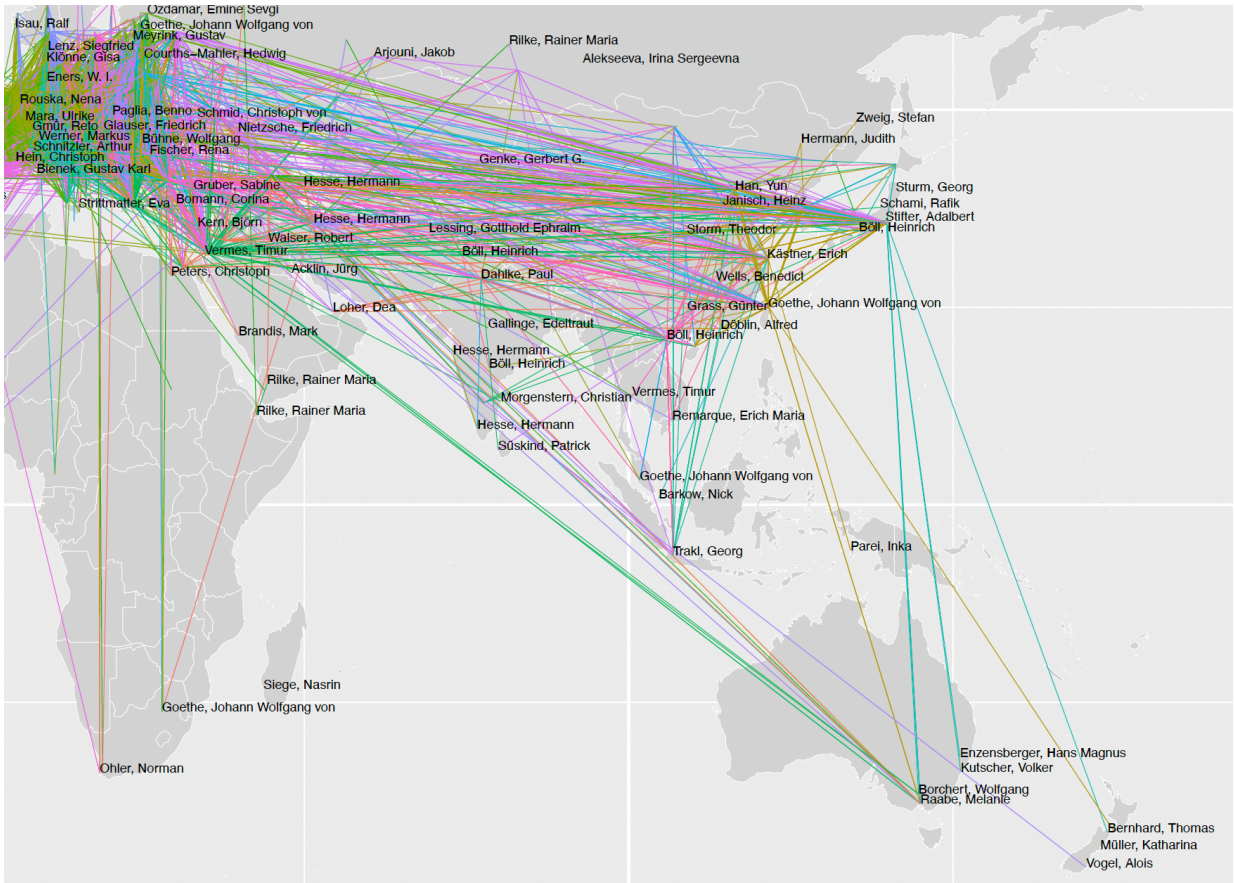
Figure A.17. Zoom on the Eastern hemisphere showing publication places connected by lines of shared authors. Labelled by author names most frequent per place. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. File: 20220601_alldnb_ch3_author_pub_place_geo_map_lines_languages_places_v1.pdf
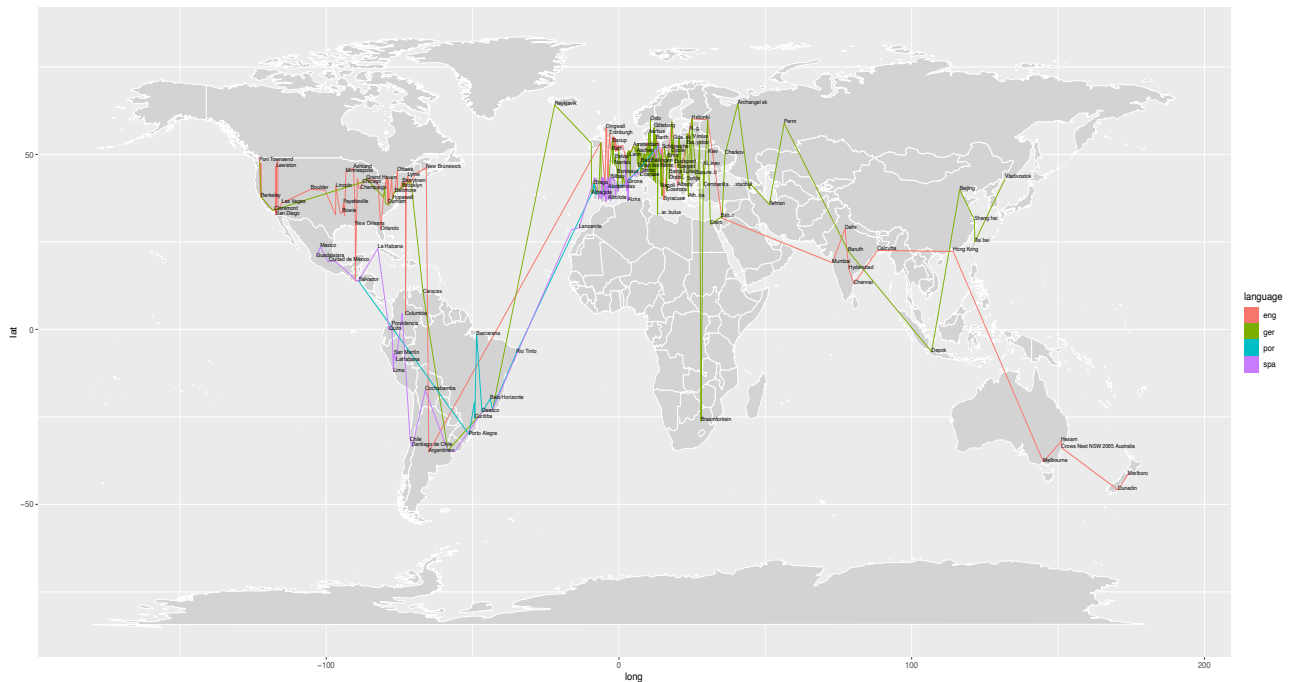

Figure A.18. Map of publication places of the subset of authors shared in central and language group grouped by community detection in chapter 2.2. connected by lines of shared authors. Subset of transcontinental languages. Map produced with geocoding by Nominatim API (tidygeocoder) and

Leaflet. File:
20220623_alldnb_ch3_author_pub_place_geo_map_lines_peripheralauthors_langcolored_v3.pdf
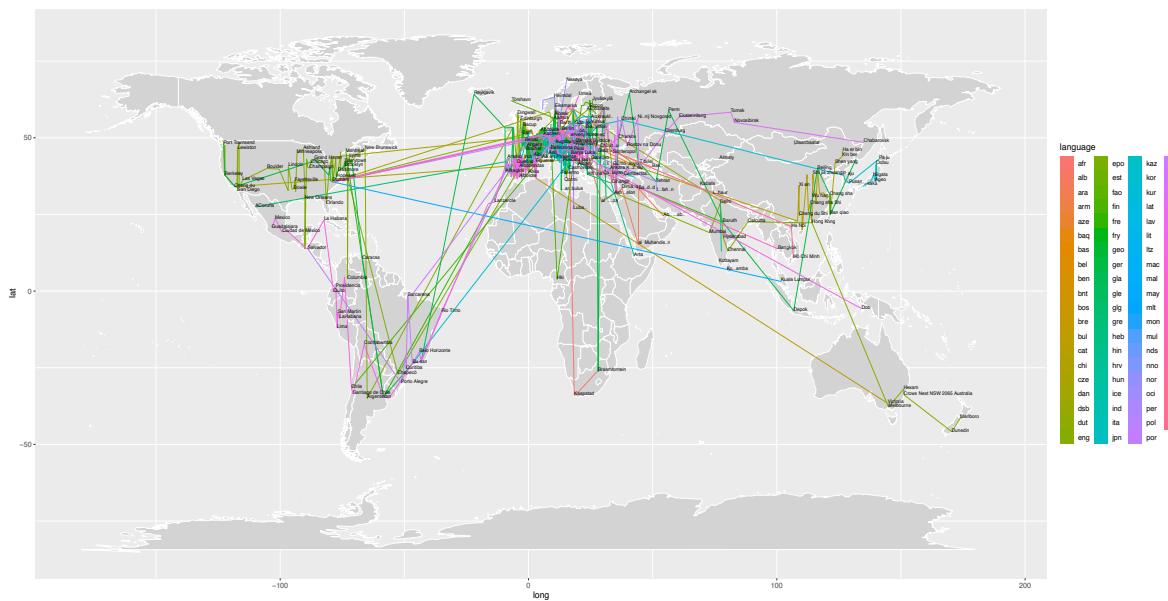
Figure A.19. Map of publication places of the subset of authors shared in central and language group grouped by community detection in chapter 2.2. connected by lines of shared authors. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. File:
20220621_alldnb_ch3_author_pub_place_geo_map_lines_sharedauthors_langcolored_v3.pdf
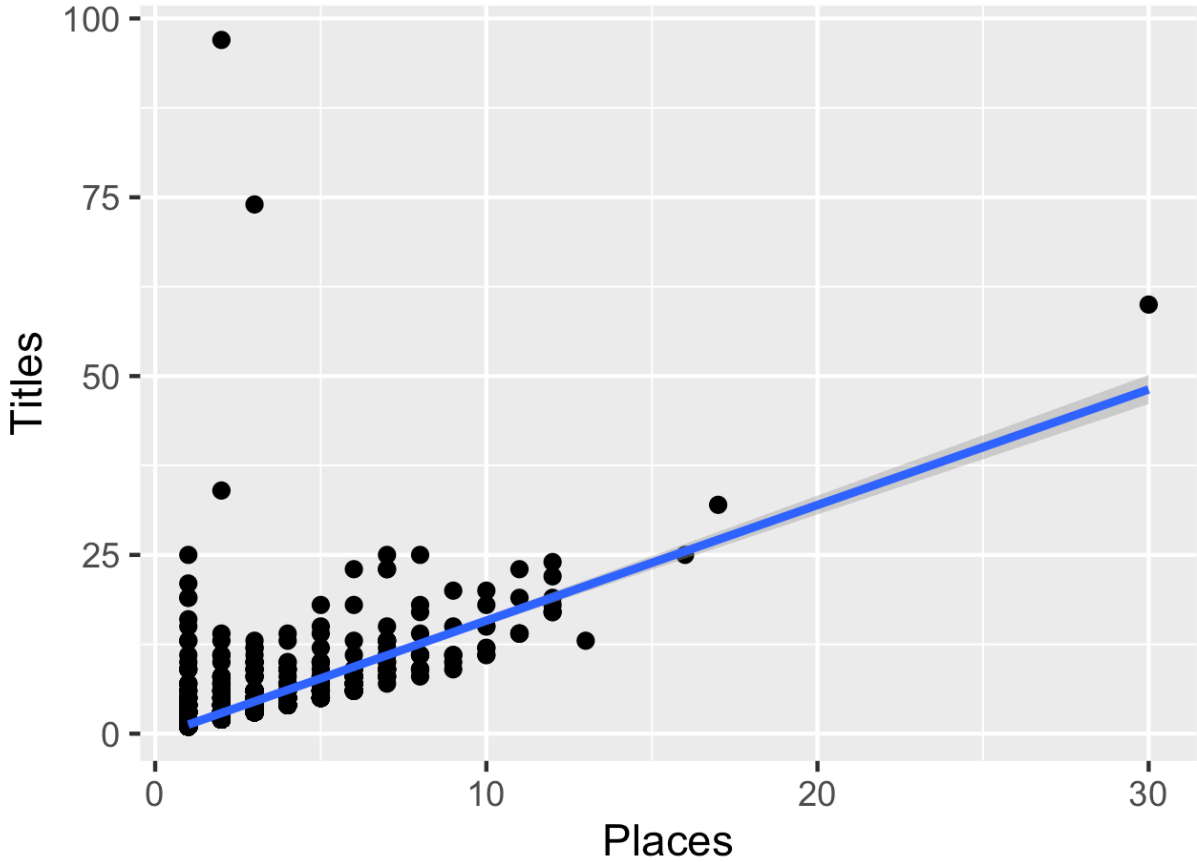
Figure A.20. Linear regression plot for titles per publishing places. File:
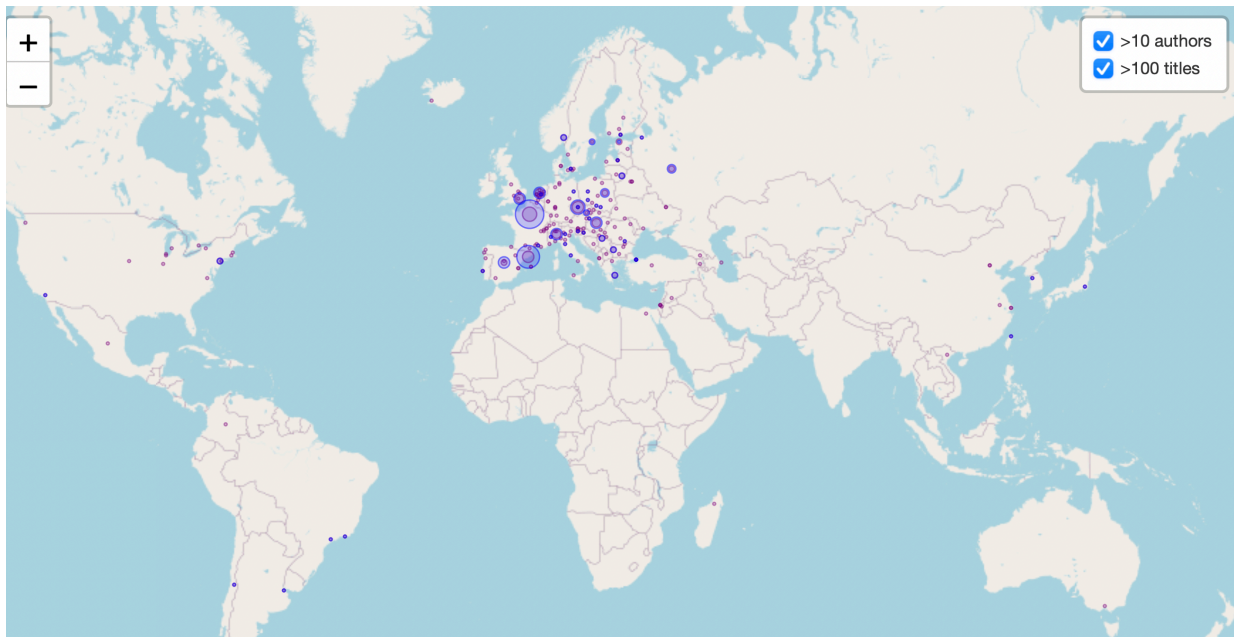20220628_alldnb_ch3_centrallang_author_place_title_freq_lm_graph_v3.png
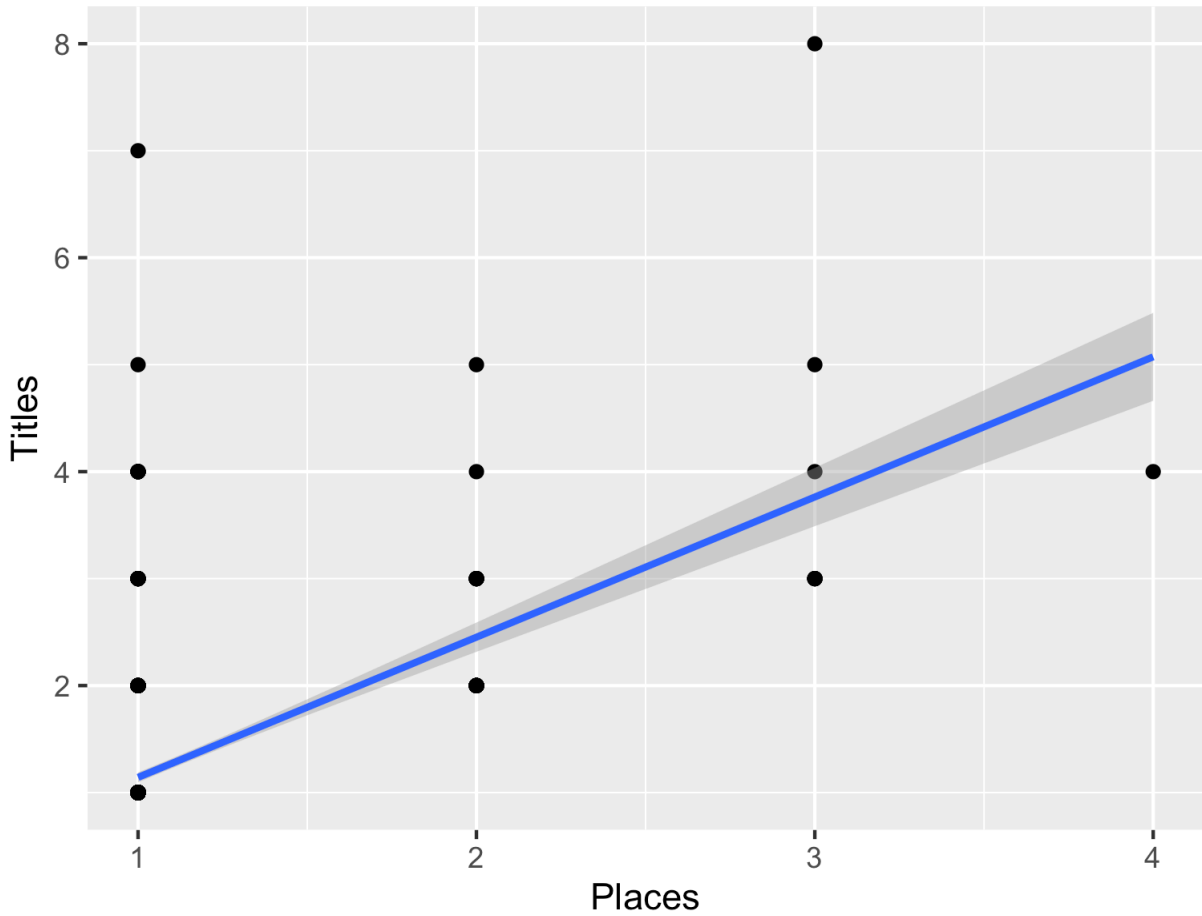
Figure A.21. Linear regression plot for authors exclusive to the peripheral language group per publishing places. File: 20220628_alldnb_ch3_peripherallang_author_place_title_freq_lm_graph_v3
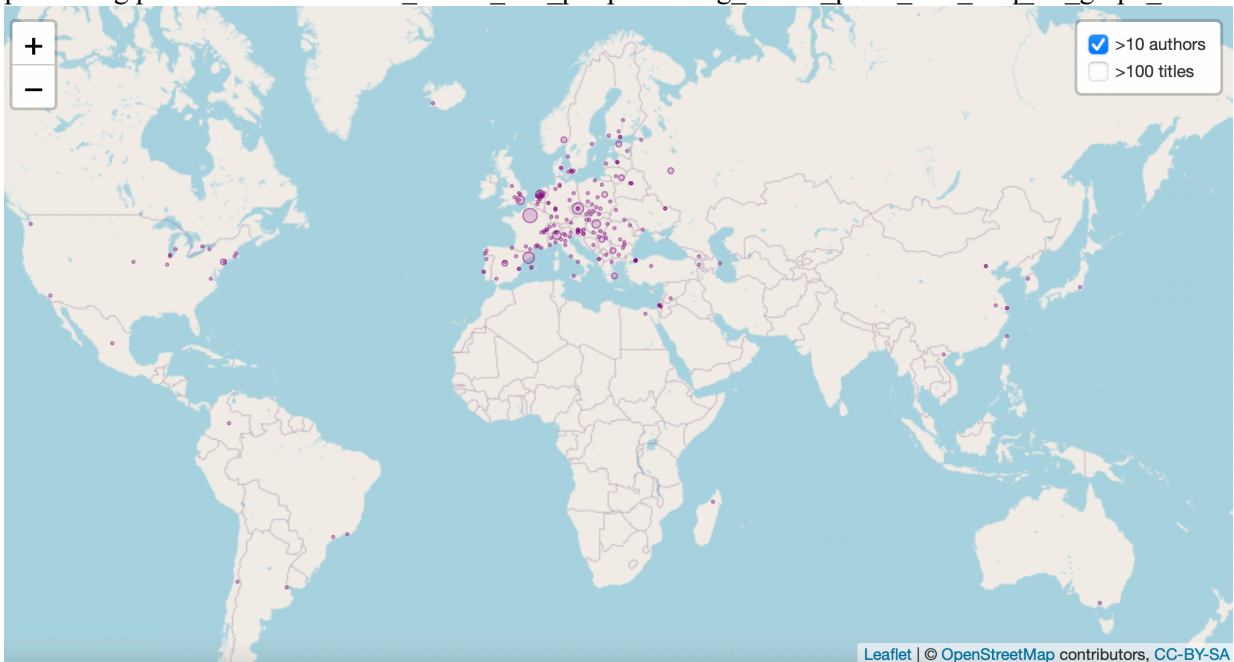
Figure A.22. Author count per first publication place with places of >10 authors. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. Marker sizes correspond to title and author

Figure A.23. Author count per first publication place with places of >10 authors and >100 titles. Map produced with geocoding by Nominatim API (tidygeocoder) and Leaflet. Marker sizes correspond to title and author count. File:
20220513_alldnb_ch3_author_pub_place_geo_map_author5min_title100min_v1.html

Bibliography

"A Literary Map of Turkey | Franz-Benjamin Mocnik, Lisa Maria Teichmann," 2015. http://projects.mocnik-science.net/literary-map-of-turkey/.

Ahnert, Ruth, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. *The Network Turn: Changing Perspectives in the Humanities*. 1st ed. Cambridge University Press, 2020. https://doi.org/10.1017/9781108866804.

Alexander, Sam. "Social Network Analysis and the Scale of Modernist Fiction." Edited by Rebecca Walkowitz. *Modernism/Modernity Print Plus* 3, no. 4 (February 1, 2019). https://doi.org/10.26597/mod.0086.

Apter, Emily. *The Translation Zone: A New Comparative Literature*. Princeton University Press, 2006. https://doi.org/10.1515/9781400841219.

Apter, Emily S. *Against World Literature: On the Politics of Untranslatability*. London ; New York: Verso, 2013.

Arac, Jonathan. "The Worldliness of World Literature." In *The Routledge Companion to World Literature*, edited by Theo d'Haen, David Damrosch, and Djelal Kadir. Milton Park, Abingdon, Oxon ; New York: Routledge, 2012.

DNB Metadatendienste. "Available Fields." Accessed January 28, 2021. https://www.dnb.de/EN/Professionell/Metadatendienste/Exportformate/CSV/csv_nod e.html.

Bachleitner, Norbert, and Michaela Wolf. "Auf dem Weg zu einer Soziologie der literarischen Übersetzung im deutschsprachigen Raum." *Internationales Archiv für Sozialgeschichte der deutschen Literatur (IASL)* 29, no. 2 (December 2004): 1–25. https://doi.org/10.1515/IASL.2004.2.1.

VIAF. "Bachmann, Ingeborg, 1926-1973. | Malina." Accessed August 23, 2022. http://viaf.org/viaf/182465068.

Bachmann-Medick, Doris. "Introduction: The Translational Turn." *Translation Studies* 2, no. 1 (January 2009): 2–16. https://doi.org/10.1080/14781700802496118.

Baker, Mona. *Translation and Conflict: A Narrative Account*. London: Routledge, 2006.

Bassnett, S. "Chapter 8 The Translation Turn in Cuiturai Studies," 2011.

Bassnett, Susan. "From Cultural Turn to Translational Turn: A Transnational Journey." In *World Literature in Theory*, edited by David Damrosch, 234–45. Chichester, England:

Wiley-Blackwell, 2014.
https://ebookcentral.proquest.com/lib/nottingham/detail.action?docID=1566390.

Bergenmar, Jenny, and Katarina Leppänen. "Gender and Vernaculars in Digital Humanities and World Literature." *NORA – Nordic Journal of Feminist and Gender Research* 25, no. 4 (October 2, 2017): 232–46. https://doi.org/10.1080/08038740.2017.1378256.

Kritikatur – Das Kulturprotal. "Bibliothek Der Weltliteratur (BDW)." Accessed July 3, 2022. https://www.kritikatur.de/Bibliothek_der_Weltliteratur_(BDW).

Borgatti, Stephen P. "Centrality and Network Flow." *Social Networks* 27, no. 1 (January 2005): 55–71. https://doi.org/10.1016/j.socnet.2004.11.008.

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Massachusetts: The MIT Press, 2015.

Börsenverein des Deutschen Buchhandels. *Buch und Buchhandel in Zahlen 2020: Zahlen, Fakten und Analysen zur wirtschaftlichen Entwicklung*, 2020. https://proxy.library.mcgill.ca/login?url=http://campus-content.library.mcgill.ca/eresources/csmono/BuBiZ_2020_E-Book.pdf.

Bourdieu, Pierre, and Randal Johnson. *The Field of Cultural Production: Essays on Art and Literature*. New York: Columbia University Press, 1993.

Braun, Rebecca. "The World Author in Us All: Conceptualising Fame and Agency in the Global Literary Market." *Celebrity Studies* 7, no. 4 (October 2016): 457–75. https://doi.org/10.1080/19392397.2016.1233767.

Burt, Ronald S. "Structural Holes and Good Ideas." *American Journal of Sociology* 110, no. 2 (September 2004): 349–99. https://doi.org/10.1086/421787.

Butts, Carter T. "Revisiting the Foundations of Network Analysis." *Science* 325, no. 5939 (July 24, 2009): 414–16. https://doi.org/10.1126/science.1171022.

———. *Sna: Tools for Social Network Analysis*, 2020. https://CRAN.R-project.org/package=sna.

Cambon, Jesse, Diego Hernangómez, Christopher Belanger, and Daniel Possenriede. "Tidygeocoder: An R Package for Geocoding." *Journal of Open Source Software* 6, no. 65 (September 9, 2021): 3544. https://doi.org/10.21105/joss.03544.

Chang, Daniel T., Yu Ge, Shiwei Song, Nicole Coleman, Jon Christensen, J. Heer, and Stanford. "Visualizing the Republic of Letters," 2009.

Cheesman, Tom. "Five Maps of Translations of Shakespeare." Edited by Tom Cheesman, Kevin Flanagan, Stephan Thiel, and Jan Rybicki. Evanston, IL: Northwestern University Press, 2016.

———. "Othellomap." Accessed June 30, 2022. https://othellomap.nand.io.

Cheng, Joe, Bhaskar Karambelkar, and Yihui Xie. *Leaflet: Create Interactive Web Maps with the JavaScript "Leaflet" Library*, 2021. https://CRAN.R-project.org/package=leaflet.

Chotiwari-Jünger, Steffi. *Georgische Literatur in Deutschen Übersetzungen: Bibliographie-1887 Bis 2017*. Wiesbaden: Reichert Verlag, 2017.

Cima, Rosie. "Bias, She Wrote. The Gender Balance of The New York Times Best Seller List." Accessed August 23, 2022. https://pudding.cool/2017/06/best-sellers/.

Clauset, Aaron, M. E. J. Newman, and Cristopher Moore. "Finding Community Structure in Very Large Networks." *Physical Review E* 70, no. 6 (December 6, 2004): 066111. https://doi.org/10.1103/PhysRevE.70.066111.

Cody, W. J. "Algorithm 715: SPECFUN–a Portable FORTRAN Package of Special Function Routines and Test Drivers." *ACM Trans. Math. Softw.* 19, no. 1 (March 1993): 22–30. https://doi.org/10.1145/151271.151273.

"Creative Commons Public Domain." Accessed August 23, 2022. https://creativecommons.org/publicdomain/zero/1.0/deed.

Cronin, Michael 1960-. *Translation and Globalization*. 1 online resource (x, 197 pages) vols. London: Routledge, 2003. https://doi.org/10.4324/9780203378816.

Damrosch, David. "Toward a History of World Literature." *New Literary History* 39, no. 3–4 (2008): 481–95. https://doi.org/10.1353/nlh.0.0045.

———, ed. *World Literature in Theory*. Chichester, West Sussex [England] ; Malden, MA: Wiley Blackwell, 2014.

Damrosch, David, and Gayatri Spivak. "Comparative Literature/World Literature: A Discussion with Gayatri Chakravorty Spivak and David Damrosch." *Comparative Literature Studies* 48 (2011): 455–85.

Demetrescu, Camil, Andrea Ribichini, and Marco Schaerf. "Accuracy of Author Names in Bibliographic Data Sources: An Italian Case Study." *Scientometrics* 117 (October 29, 2018). https://doi.org/10.1007/s11192-018-2945-x.

Dennerlein, Katrin. "Netzwerke Medialer Formationen Der Dramen- Und Theaterhistoriographie. Eine Analyse Großer Dramensammlungen Im 18. Jahrhundert." Edited by Kirsten Dickhaut and Gabriel Viehhauser. *Comparatio. Zeitschrift Für Vergleichende Literaturwissenschaft*, Themenheft „Theaterhistoriographie und Digital Humanities", 12/2 (2020): 145–61.

Deutsches Literatur-Lexikon, Band 7, Haab - Hogrebe: DE GRUYTER SAUR, 1979. https://doi.org/10.1515/9783110974744.

Deutsches Literatur-Lexikon, Band 8, Hohberg- Kober: DE GRUYTER SAUR, 1981.
https://doi.org/10.1515/9783110974737.

Deutsches Literatur-Lexikon, Band 13, Rill - Salzmann: DE GRUYTER SAUR, 1991.
https://doi.org/10.1515/9783110954364.

Deutsches Literatur-Lexikon, Band 14, Salzmesser - Schilling: DE GRUYTER SAUR, 1992.
https://doi.org/10.1515/9783110954357.

Deutsches Literatur-Lexikon, Band 16, Schobel - Schwaiger: DE GRUYTER SAUR, 1995.
https://doi.org/10.1515/9783110955408.

Deutsches Literatur-Lexikon, Band 24, Tsakiridis - Ursinus: Berlin, Boston: DE GRUYTER
SAUR, 2004. https://doi.org/10.1515/9783110956221.

Deutsches Literatur-Lexikon, Band 29, Wedekind - Weiss: Berlin, Boston: DE GRUYTER
SAUR, 2009. https://doi.org/10.1515/9783110949704.

Deutsches Literatur-Lexikon, Ergänzungsband II, Bernfeld - Christen: Berlin, Boston: DE
GRUYTER SAUR, 1994. https://doi.org/10.1515/9783110948691.

Deutsches Literatur-Lexikon, Ergänzungsband IV, Fraenkel - Hermann: Berlin, Boston: DE
GRUYTER SAUR, 1996. https://doi.org/10.1515/9783110948455.

Dewitt, Anne. "Advances in the Visualization of Data: The Network of Genre in the
Victorian Periodical Press." *Victorian Periodicals Review* 48, no. 2 (2015): 161–82.
https://doi.org/10.1353/vpr.2015.0030.

Dimock, W. C. "Scales of Aggregation: Prenational, Subnational, Transnational." *American
Literary History* 18, no. 2 (March 1, 2006): 219–28.
https://doi.org/10.1093/alh/ajj011.

Donahaye, Jasmine. "Three Percent? Publishing Data and Statistics on Translated Literature
in the United Kingdom and Ireland." Making Literature Travel. Mercator Institute for
Media, Languages and Culture, Aberystwyth University, Wales, UK, December 2012.
https://www.lit-across-frontiers.org/wp-content/uploads/2013/03/Publishing-Data-
and-Statistics-on-Translated-Literature-in-the-United-Kingdom-and-Ireland-A-LAF-
research-report-March-2013-final.pdf.

Eckermann, Johann Peter. *Gespräche Mit Goethe in Den Letzten Jahren Seines Lebens*.
Leipzig: Brockhaus, 1837.

Elson, David, Nicholas Dames, and Kathleen McKeown. "Extracting Social Networks from
Literary Fiction." In *Proceedings of the 48th Annual Meeting of the Association for
Computational Linguistics*, 138–47. Uppsala, Sweden: Association for Computational
Linguistics, 2010. https://aclanthology.org/P10-1015.

Erlin, Matt, Andrew Piper, Douglas Knox, Stephen Pentecost, Michaela Drouillard, Brian Powell, and Cienna Townson. "Cultural Capitals: Modeling 'Minor' European Literature." *Journal of Cultural Analytics*, February 17, 2021. https://doi.org/10.22148/001c.21182.

Etiemble, René. "Should We Rethink the Notion of World Literature?" In *The Routledge Companion to World Literature*, 1st ed. Routledge, 2011. https://doi.org/10.4324/9780203806494.

"FEP Publications." Accessed January 28, 2021. https://fep-fee.eu/-Publications-.

Flood, Alison. "Acclaimed German Novel Banned by Nazis Gets First English Translation." *The Guardian*, January 23, 2019. https://www.theguardian.com/books/2019/jan/23/acclaimed-german-novel-banned-by-nazis-gets-first-english-translation?fbclid=IwAR3sT2UXwgNSG2st8LFODFoInIPCBXO9w4yG_Wp-UQXPIMMM77c59TM4U_M.

———. "Herta Müller Takes Nobel Prize for Literature," October 8, 2009. https://www.theguardian.com/books/2009/oct/08/herta-muller-nobel-prize-literature.

Folaron, Deborah, and Hélène Buzelin. "Introduction: Connecting Translation and Network Studies." *Meta* 52, no. 4 (March 7, 2008): 605–42. https://doi.org/10.7202/017689ar.

Foucault, Michel. *Aesthetics, Method, and Epistemology: Essential Works of Foucault 1954-1984.* Place of publication not identified: Penguin Books, 1998.

Wikipedia. "Frauenliteratur." Accessed February 26, 2020. https://de.wikipedia.org/wiki/Frauenliteratur.

Gerber, Leah. "Building Bridges, Building a Bibliography of Australian Children's Fiction in German Translation 1854–2007." *New Review of Children's Literature and Librarianship* 14, no. 2 (September 2008): 141–57. https://doi.org/10.1080/13614540902835186.

González, José Eduardo, Elliott Jacobson, Laura García García, and Leonardo Brandolini Kujman. "Measuring Canonicity: Graduate Reading Lists in Departments of Hispanic Studies." *Journal of Cultural Analytics*, March 19, 2021. https://doi.org/10.22148/001c.21599.

Gritta, Milan, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. "What's Missing in Geographical Parsing?" *Language Resources and Evaluation* 52, no. 2 (June 2018): 603–23. https://doi.org/10.1007/s10579-017-9385-8.

Haen, Theo d', David Damrosch, and Djelal Kadir, eds. *The Routledge Companion to World Literature*. Milton Park, Abingdon, Oxon ; New York: Routledge, 2012.

Hass, Annika. *Der Verleger Johann Friedrich Cotta (1764-1832) als Kulturvermittler zwischen Deutschland und Frankreich: Frankreichbezüge, Koeditionen und Übersetzungen*. Zivilisationen & Geschichte, Band 33. Frankfurt am Main ; New York: Peter Lang Edition, 2015.

Hassan, Waïl S. "Agency and Translational Literature: Ahdaf Soueif's *The Map of Love*." *PMLA/Publications of the Modern Language Association of America* 121, no. 3 (May 2006): 753–68. https://doi.org/10.1632/003081206X142869.

Heilbron, Johan. "Towards a Sociology of Translation: Book Translations as a Cultural World-System." *European Journal of Social Theory* 2, no. 4 (November 1999): 429–44. https://doi.org/10.1177/136843199002004002.

Helgesson, Stefan, and Christina Kullberg. "Translingual Events." *Journal of World Literature* 3, no. 2 (June 1, 2018): 136–52. https://doi.org/10.1163/24056480-00302002.

Hesse, Hermann. *Schriften Zur Literatur*. Frankfurt am Main: Suhrkamp, 1972.

Honegger, Gitta. "Language Speaks. Anglo-Bernhard: Thomas Bernhard in Translation." In *A Companion to the Works of Thomas Bernhard*, edited by Matthias Konzett, 169–86. Boydell & Brewer, 2002. https://www.cambridge.org/core/books/companion-to-the-works-of-thomas-bernhard/language-speaks-anglobernhard-thomas-bernhard-in-translation/173D3C481C88BA55BCFECD440FEFA7EC.

House, Juliane. "Towards a Linguistic Theory of Translation as Re-Contextualisation and a Third Space Phenomenon." *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 7 (October 25, 2021). https://doi.org/10.52034/lanstts.v7i.213.

Hube, Christoph, Frank Fischer, Robert Jäschke, Gerhard Lauer, and Mads Rosendahl Thomsen. "World Literature According to Wikipedia: Introduction to a DBpedia-Based Framework." *ArXiv:1701.00991 [Cs]*, January 4, 2017. http://arxiv.org/abs/1701.00991.

"Index Translationum." Accessed March 20, 2019. http://www.unesco.org/xtrans/bsstatexp.aspx.

Italiano, Federico. *Translation and Geography*. New Perspectives in Translation Studies. Milton Park, Abingdon, Oxon ; New York, NY: Routledge, 2016.

Ivanovic, Christine, and Barbara Seidl. "What Is Translational Literature and How to Classify It?: Crowd-Sourcing as a Starting Point for Corpus Building and Type Distinction in

Comparative Literature." In *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality - TEEM '16*, 957–63. Salamanca, Spain: ACM Press, 2016. https://doi.org/10.1145/3012430.3012632.

VIAF. "Kafka, Franz, 1883-1924. | Prozess." Accessed August 23, 2022. http://viaf.org/viaf/183949498.

Kamada, Tomihisa, and Satoru Kawai. "An Algorithm for Drawing General Undirected Graphs." *Information Processing Letters* 31, no. 1 (April 1989): 7–15. https://doi.org/10.1016/0020-0190(89)90102-6.

Österreichische Nationalbibliothek. "Katalog." Accessed January 28, 2021. https://labs.onb.ac.at/de/dataset/catalogue/.

Kershaw, Angela, and Gabriela Saldanha. "Introduction: Global Landscapes of Translation." *Translation Studies* 6, no. 2 (May 2013): 135–49. https://doi.org/10.1080/14781700.2013.777257.

Kilduff, Martin, and Daniel J. Brass. "Organizational Social Network Research: *Core Ideas and Key Debates.*" *Academy of Management Annals* 4, no. 1 (January 2010): 317–57. https://doi.org/10.5465/19416520.2010.494827.

Kim, David D, and Nickolas de Carlo. "PROJECT ONE: WorldLiterature@UCLA Tracking International Publics with Goethe." *Transit* 10, no. 1 (2015). https://doi.org/10.5070/T7101029575.

Kovać, Miha., Wischenbart, Ruediger Wischenbart., Jursitzky, Jennifer., Kaldonek, Sabine., Coufal, Julia.,. "Diversity Report 2010 : Literary Translation in Current European Book Markets : An Analysis of Authors, Languages, and Flows." [Wien?]: Wischenbart.com, 2010. /z-wcorg/. http://www.wischenbart.com/diversity/report/Diversity-Report_2010.pdf.

Laachir, Karima, Sara Marzagora, and Francesca Orsini. "Significant Geographies: In Lieu of World Literature." *Journal of World Literature* 3, no. 3 (August 10, 2018): 290–310. https://doi.org/10.1163/24056480-00303005.

Lahti, L., E. Mäkelä, and M. Tolonen. "Quantifying Bias and Uncertainty in Historical Data Collections with Probabilistic Programming." In *CHR*, 2020.

Lahti, Leo, Ville Vaara, Jani Marjanen, and Mikko Tolonen. "Best Practices in Bibliographic Data Science." In *Proceedings of the Research Data And Humanities (RDHUM) 2019 Conference : Data, Methods And Tools*, 57–65. Oulu, 2019.

Lenze, Ulla. "Vita - Ulla Lenze." Accessed June 29, 2022. https://www.ullalenze.de/vita/.

Levine, C., and B. V. Mani. "What Counts as World Literature?" *Modern Language Quarterly* 74, no. 2 (January 1, 2013): 141–49. https://doi.org/10.1215/00267929-2073043.

Long, Hoyt, and Richard Jean So. "Turbulent Flow: A Computational Model of World Literature." *Modern Language Quarterly* 77, no. 3 (September 2016): 345–67. https://doi.org/10.1215/00267929-3570656.

Mäkelä, Eetu, Krista Lagus, Leo Lahti, Tanja Säily, Mikko Tolonen, Mika Hämäläinen, Samuli Kaislaniemi, and Terttu Nevalainen. *Wrangling with Non-Standard Data*, 2020.

Mani, B. V. "Borrowing Privileges: Libraries and the Institutionalization of World Literature." *Modern Language Quarterly* 74, no. 2 (January 1, 2013): 239–60. https://doi.org/10.1215/00267929-2073007.

Mani, B. Venkat. *Recoding World Literature: Libraries, Print Culture, and Germany's Pact with Books*. First edition. New York: Fordham University Press, 2017.

May, Markus, and Evi Zemanek, eds. *Annäherung - Anverwandlung - Aneignung: Goethes Übersetzungen in Poetologischer Und Interkultureller Perspektive*. Würzburg: Königshausen & Neumann, 2013.

Mocnik, Franz-Benjamin, Amin Mobasheri, Luisa Griesbaum, Melanie Eckle, Clemens Jacobs, and Carolin Klonner. "A Grounding-Based Ontology of Data Quality Measures." *Journal of Spatial Information Science*, no. 16 (June 23, 2018): 1–25. https://doi.org/10.5311/JOSIS.2018.16.360.

Moretti, Franco. *Atlas of the European Novel 1800 – 1900*. London: Verso, 1998.

Newman, M E J. "Fast Algorithm for Detecting Community Structure in Networks." *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 69, no. 6 Pt 2 (2004): 066133.

Newman, M. E. J. *Networks: An Introduction*. Oxford ; New York: Oxford University Press, 2010.

Newman, M. E. J., and M. Girvan. "Finding and Evaluating Community Structure in Networks." *Physical Review E* 69, no. 2 (February 26, 2004): 026113. https://doi.org/10.1103/PhysRevE.69.026113.

Norrick-Rühl, Corinna, and Melanie Ramdarshan Bold. "Crossing the Channel: Publishing Translated German Fiction in the UK." *Publishing Research Quarterly* 32, no. 2 (June 2016): 125–38. https://doi.org/10.1007/s12109-016-9458-3.

Olensky, Marlies. "Testing an Automated Accuracy Assessment Method on Bibliographic Data." *Journal of Library and Information Studies* 12 (December 1, 2014). https://doi.org/10.6182/jlis.2014.12(2).019.

Olson, Jack E. *Data Quality: The Accuracy Dimension*. San Francisco: Morgan Kaufmann, 2003.

Overell, Simon. "The Problem of Place Name Ambiguity." *SIGSPATIAL Special* 3, no. 2 (July 2011): 12–15. https://doi.org/10.1145/2047296.2047299.

Papastergiadis, Nikos. *The Turbulence of Migration: Globalization, Deterritorialization, and Hybridity*. Cambridge, UK : Malden, Mass: Polity Press ; Blackwell Publishers, 2000.

Österreichische Nationalbibliothek. "Parfjumăt : Istorijata Na Edin Ubiec . 3. Izd." Accessed August 23, 2022. http://data.onb.ac.at/rec/AC08832393.

Park, Gyeong-Mi, Sung-Hwan Kim, Hye-Ryeon Hwang, and Hwan-Gue Cho. "Complex System Analysis of Social Networks Extracted from Literary Fictions." *International Journal of Machine Learning and Computing*, 2013, 107–11. https://doi.org/10.7763/IJMLC.2013.V3.282.

Pechenick, Eitan, Christopher Danforth, and Peter Dodds. "Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution." *PloS One* 10 (January 5, 2015). https://doi.org/10.1371/journal.pone.0137041.

Pickford, Susan. "The Factors Governing the Availability of Maghrebi Literature in English: A Case Study in the Sociology of Translation." *CLINA* 2–1 (June 2016): 77–94.

Piper, Andrew. "Paraphrasis: Goethe, the Novella, and Forms of Translational Knowledge." *Goethe Yearbook* 17, no. 1 (2010): 179–201.

Librarianship Studies & Information Technology. "Place of Publication," July 3, 2019. https://www.librarianshipstudies.com/2016/02/place-of-publication.html.

Poschmann, Marion. *Die Kieferninseln: Roman*. Erste Auflage. Berlin: Suhrkamp, 2017.

Poupaud, Sandra, Anthony Pym, and Ester Torres Simón. "Finding Translations. On the Use of Bibliographical Databases in Translation History." *Meta* 54, no. 2 (July 16, 2009): 264–78. https://doi.org/10.7202/037680ar.

"Publishing Translations in Europe. Trends 1990-2005. Based on Analysis of the Index Translationum Database." Making Literature Travel. Mercator Institute for Media, Languages and Culture, Aberystwyth University, Wales, UK, 2011. https://www.lit-across-frontiers.org/wp-content/uploads/2013/03/Publishing-Translations-in-Europe-Trends-1990-%E2%80%93-2005.pdf.

Pym, A., and Grzegorz Chrupała. "The Quantitative Analysis of Translation Flows in the Age of an International Language," 2005.

Pym, Anthony. *Translation and Text Transfer: An Essay on the Principles of Intercultural Communication*. Tarragona: Intercultural Studies Group, 2010. http://usuaris.tinet.cat/apym/publications/TTT_2010.pdf.

Queneau, Raymond. *Pour Une Bibliothèque Idéale*. Gallimard, 1956.

Raven, James. "An Antidote to the French?: English Novels in German Translation and German Novels in English Translation 1770-99." *Eighteenth-Century Fiction* 14, no. 3–4 (2002): 715–34. https://doi.org/10.1353/ecf.2002.0032.

Renn, Joachim. *Übersetzungsverhältnisse: Perspektiven Einer Pragmatistischen Gesellschaftstheorie*. 1. Aufl. Weilerswist: Velbrück Wissenschaft, 2006.

Reynolds, Matthew, and Giovanni Pietro Vitali. "Mapping and Reading a World of Translations: Prismatic Jane Eyre." *Modern Languages Open* 1 (December 17, 2021): 23. https://doi.org/10.3828/mlo.v0i0.375.

Risku, Hanna, Regina Rogl, and Christina Pein-Weber. "Mutual Dependencies: Centrality in Translation Networks." *The Journal of Specialised Translation*, February 1, 2016.

Sälzer, Christian, and Martin Schmitz-Kuhl. "Jahresbericht." Deutsche Bibliothek (Frankfurt Am Main; Leipzig): Jahresbericht. Frankfurt am Main, Leipzig: Deutsche Nationalbibliothek, 2020. https://d-nb.info/1212372646/34.

Deutsche Nationalbibliothek. "Sammlung Körperlicher Medienwerke." Accessed January 28, 2021. https://www.dnb.de/DE/Professionell/Sammeln/Koerperliche_Medienwerke/koerperliche_medienwerke_node.html.

Sapiro, Gisèle. "How Do Literary Works Cross Borders (or Not)?" *Journal of World Literature* 1, no. 1 (2016): 81–96. https://doi.org/10.1163/24056480-00101009.

———. "The Transnational Literary Field between (Inter)-Nationalism and Cosmopolitanism." *Journal of World Literature* 5, no. 4 (November 4, 2020): 481–504. https://doi.org/10.1163/24056480-00504002.

———. "Translation and the Field of Publishing: A Commentary on Pierre Bourdieu's 'A Conservative Revolution in Publishing.'" *Translation Studies* 1, no. 2 (July 2008): 154–66. https://doi.org/10.1080/14781700802113473.

Schiff, Stacy, Hermann Kusterer, and Stacy Schiff. *Véra: ein Leben mit Vladimir Nabokov*. rororo 22991. Reinbek bei Hamburg: Rowohlt-Taschenbuch-Verl, 2001.

Schulenberg, René. *Dineq: Decomposition of (Income) Inequality*, 2018. https://CRAN.R-project.org/package=dineq.

Simon, Sherry. *Sites of Translation A Field Guide.* 1 online resource vols. New Perspectives in Translation and Interpreting Studies. Routledge, 2018. https://www.taylorfrancis.com/books/e/9781315311098.

Snell-Hornby, Mary. "What's in a Turn? On Fits, Starts and Writhings in Recent Translation Studies." *Translation Studies* 2, no. 1 (January 2009): 41–51. https://doi.org/10.1080/14781700802496225.

"Special Eurobarometer 386. Europeans and Their Languages," June 2012.

Spemann, Adolf. *Vergleichende Zeittafel Der Weltliteratur Vom Mittelalter Bis Zur Neuzeit (1150-1939)*. Engelhornverlag, 1951.

Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Author." Accessed January 28, 2021. https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=5&nTyp=min&topN=50.

Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Original Country." Accessed January 28, 2021. https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=1&nTyp=min&topN=50&lg=0.

Index Translationum. "Statistics on Whole Index Translationum Database: Top 50 Original Language." Accessed January 28, 2021. https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=3&nTyp=min&topN=50&lg=0.

VIAF. "Süskind, Patrick 1949- Das Parfum." Accessed August 23, 2022. http://viaf.org/viaf/220436158.

Tahir-Gürçağlar, Şehnaz. "Chaos Before Order: Network Maps and Research Design in DTS." *Meta* 52, no. 4 (March 7, 2008): 724–43. https://doi.org/10.7202/017694ar.

Tashinskiy, Aleksey. *Literarische Übersetzung als Universum der Differenz: mit einer analytischen Studie zu deutschen Übersetzungen des Romans Oblomov von I.A. Gončarov*. Ost-West-Express, Band 33. Berlin: Frank & Timme, 2018.

Tatlock, Lynne. *German Writing, American Reading: Women and the Import of Fiction, 1866-1917*, 2012. http://ezproxy.uniandes.edu.co:8080/login?url=https://muse.jhu.edu/book/23962/.

Teichmann, Lisa. "Migration and Literary Education: A Comparative Study of the Turkish American Community and the Turkish Austrian Community, San Francisco and Vienna." Marshallplan Foundation, 2016.

University of Rochester. "Three Percent Translation Database." Accessed July 3, 2022. http://www.rochester.edu/College/translation/threepercent/translation-database/.

Tolonen, Mikko, Hege Roivainen, Jani Marjanen, and Leo Lahti. "Scaling up Bibliographic Data Science." In *Digital Humanities in the Nordic Countries : Proceedings of the Digital Humanities in the Nordic Countries 4th Conference*, 450–56. Aachen, 2019. https://researchportal.helsinki.fi/en/publications/scaling-up-bibliographic-data-science.

Turrin, Enrico. "The Book Sector in Europe." Federation of European Publishers, 2017. http://fep-fee.us11.list-manage2.com/track/click?u=f3e97360a9e613a0b533827b2&id=39a85d6b7d&e=70ce53a2fb.

Tymoczko, Maria. *Enlarging Translation, Empowering Translators*. Manchester, UK ; Kinderhook, NY: St. Jerome Pub, 2007.

Deutsche Nationalbibliothek. "Unser Sammelauftrag." Accessed January 28, 2021. https://www.dnb.de/DE/Professionell/Sammeln/sammeln_node.html.

Vaicenoniene, Jurgita. "Literary Translation in Anglo-American and Lithuanian Contexts (1990-2008)." *KALBU STUDIJOS*, STUDIES ABOUT LANGUAGES, 16 (2010).

Van Kleeck, David, Gerald Langford, Jimmie Lundgren, Hikaru Nakano, Allison Jai O'Dell, and Trey Shelton. "Managing Bibliographic Data Quality in a Consortial Academic Library: A Case Study." *Cataloging & Classification Quarterly* 54, no. 7 (October 2, 2016): 452–67. https://doi.org/10.1080/01639374.2016.1210709.

Vögler, Peter. "Peter Völker – Schriftsteller Für Romane, Lyrik Und Belletristik." Accessed June 29, 2022. http://www.petervoelker.de.

Volková, Bronislava. *Forms of Exile in Jewish Literature and Thought: Twentieth-Century Central Europe and Migration to America*. Academic Studies Press, 2021. https://doi.org/10.1515/9781644694060.

Walkowitz, Rebecca. *Born Translated*. S.l.: Columbia University Press, 2017.

Walkowitz, Rebecca L. "The Location of Literature: The Transnational Book and the Migrant Writer." *Contemporary Literature* 47, no. 4 (2006): 527–45.

Wischenbart, Rüdiger. "Rüdiger Wischenbart Conent & Consulting." Accessed August 23, 2022. https://www.wischenbart.com.

Wischenbart, Rüdiger, Miha Kovać, and Michaela Anna Fleischhacker. "Diversity Report 2020. Trends in Literary Translation in Europe. Authoring and Publishing Models, Market Developments and Sponsorship Policies in Austria, Czech Republic,

Germany, France, Italy, Poland, Slovenia, Spain, and Sweden." Diversity Report.
Vienna: Verein für kulturelle Transfers. CulturalTransfers.org, 2020.

Wischenbart, Rüdiger, Miha Kovać, Yana Genova, and Michaela Anna Fleischhacker.
"Diversity Report 2018. Trends in Literary Translation in Europe Market
Developments. Traditional and New Publishing Models. Funding Schemes."
Diversity Report. Vienna: Verein für kulturelle Transfers. CulturalTransfers.org,
2019.

Wischenbart, Rüdiger, Miha Kovać, Jennifer Jursitzky, Sabine Kaldonek, and Julia Coufal.
"Diversity Report 2010. Literary Translation in Current European Book Markets. An
Analysis of Authors, Languages, and Flows." Diversity Report. Vienna: Verein für
kulturelle Transfers. CulturalTransfers.org, 2010.
http://www.wischenbart.com/DiversityReport2010.

"WorldLiterature@UCLA: Translational Network Analysis in the Digital Humanities."
Accessed August 11, 2022. https://worldlit.cdh.ucla.edu.

Zhai, Alex, Zheng Zhang, Amel Fraisse, Ronald Jenn, Shelley Fisher Fishkin, and Pierre
Zweigenbaum. "TL-Explorer: A Digital Humanities Tool for Mapping and Analyzing
Translated Literature." In *Proceedings of the 4th Joint SIGHUM Workshop on
Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and
Literature*, 167–71. Online: International Committee on Computational Linguistics,
2020. https://aclanthology.org/2020.latechclfl-1.20.

Zhenduo, Zheng. "A View on the Unification of Literature." In *World Literature in Theory*,
edited by David Damrosch. Chichester, West Sussex [England] ; Malden, MA: Wiley
Blackwell, 2014.

Zhou, Xiaoyan, and Sanjun Sun. "Bibliography-Based Quantitative Translation History."
Perspectives 25, no. 1 (January 2, 2017): 98–119.
https://doi.org/10.1080/0907676X.2016.1177100.