Made to Be a Woman: A case study on the categorization of gender using an individuation-based approach in the analysis of literary texts

Mareike Schumacher <mareike_dot_schumacher_at_ur_dot_de>, University of Regensburg https://orcid.org/0000-0002-7952-4194

Marie Flüh <marie_dot_flueh_at_uni-hamburg_dot_de>, Hamburg University https://orcid.org/0000-0002-1707-284X

Abstract

In this article we analyze Simone de Beauvoir's *The Second Sex* in order to come up with a category system including gender categories that do not only represent a binary notion but also include non-binary forms of gender. By manually annotating and then visualizing features of gender roles as a networked graph and analyzing the resulting "gender sphere" we come up with two category systems, one containing five, the other one four gender categories, each of which contain a number of gender roles. We compare the gender sphere Beauvoir creates in her text with one resulting from the analysis of narrative texts that she used as references for her analysis of gender roles. In this fiction corpus, we annotated character features, clothes and gender roles, evaluated them quantitatively and then matched them with the gender sphere. Thus it becomes visible that the literary corpus creates a scalar notion of gender ranging from a female to a male pole with a neutral center. Roles relevant for and characters showing traits of gender-diversity are set at the margins. Unlike this scalar notion Beauvoir works towards a more spherical understanding of gender where roles relevant for gender-diversity are well integrated into the center of the gender sphere. Although Beauvoir's gender sphere cannot be mapped onto the graph data of the fiction corpus, it does offer a good starting point for macro- and microanalysis of literary characters, the features, clothes and roles that are used to describe them and the resulting gender specific profiling.

1. Introduction

In the history of feminism and gender theories, there have been different attempts to categorize gender. From a stereotypical binary categorization of people as either male or female ([de Beauvoir 1992], [Bourdieu 2010]) up to a negation of biological indications of gender [Butler 2003], there have been many thoughts on what it could be that defines humans as belonging to one or another category of gender. A pioneering contribution that is still relevant today (cf. [Connell 2015, 61, 64–5]; [Kirkpatrick 2020]), is *The Second Sex* by Simone de Beauvoir. In this article, we analyze this approach because it represents and reproduces a large set of stereotypical female roles (e.g. "wife," or "working woman") by differentiating them from male counterparts. We use this binary, clearly defined dichotomy as a starting point for defining two opposite categories and questioning this opposition. We analyse this text in particular because it is still one of the most relevant and often cited texts with regard to gender aspects. Additionally, it marks a turning point in the study of gender: Although Beauvoir analyzes stereotypical roles, she shows some roles that do not fit into a binary understanding of gender. More importantly, she points out the social construction of all of these roles and the features most typically assigned to them. Especially the famous sentence "one is not born but made to be a woman" is often quoted as a key statement. However, a systematic analysis of all gender roles contained in the text and the transfer into a model that can be used for further analysis of gender roles does not exist at this point. In her study, Beauvoir extracts gender roles from literary texts, among others, to illustrate what kinds of gender roles exist. To transmit the gender roles presented and features assigned to them by Beauvoir into a transferable category system we annotated the second part of the text.[1] We first develop a gender model based on Beauvoir's theoretical text. Beauvoir herself quotes indirectly from literary texts or uses direct quotations in order to show which gender roles exist. In a second step, we apply the

gender model highlighted by Beauvoir to the literary texts she cites. Since gender roles alone are not the only important factor when it comes to gender aspects, we also look for features that are attributed to different gender roles in the texts. We do this to further differentiate the model and ultimately gain a more accurate understanding of how gender roles are constructed in literary texts. Our contribution can be understood as a transfer of a non-digital study to computational literary studies. As Beauvoir uses a common approach of non-digital humanities we wanted to redesign the attempt to a controlled setup in which exactly the same amount of text is used from each of the sources. Instead of listing some of the quoted features we gather all of the roles and features assigned to characters in these passages. We then compare the gendered characterizations in a network analysis. Moreover, we assemble all characters' roles, clothes and features in one graph in order to see whether there are common features assigned to certain gender roles.

Beauvoir ends up with an intrinsically binary system of female and male roles[2] and gives examples of a few alternative gender roles (such as "lesbian"). Social acceptance of certain features plays a key role for her attempt. She describes how certain features[3] (e.g. narcissism) are either supported (for female children in the case of narcissism) or oppressed (for male children in the case of narcissism) during youth depending on the gender identity that shall be formed. It is important to keep in mind that the focus here lies on the development of certain features according to social norms, whereas individuals can be born with any kind of feature combination. It is also crucial to note that Beauvoir does not end up with a fixed set of features for each role she describes. Instead gender roles turn out to be defined as fuzzy sets (cf. [Zadeh 1965]) because

2

1. gender roles are not clearly distinguishable categories but show permeable borders and sometimes even overlap, and
2. gender roles are not defined by a fixed set of features that all individuals embodying such a role show. Instead individuals can show a subset of features assigned to a gender role and even some features assigned to another role.

In addition gender roles are non-exclusive, which means that a character can show more than one gender role.

3

What holds true for the gender roles developed by Beauvoir is also true for the gender categories we come up with through analyzing her work: in each gender category various gender roles are included; some of them are more closely related to a top-level-category than others; roles that are not very closely related to the top-level categories can sometimes also be grouped into another top-level category. We do not exactly know which features form a category, where a category ends and a new one begins. We do not assume a closed, rigid theoretical gender model, but rather a network of commonalities that can overlap and intersect, thus yielding diverse and less sharply separated gender roles.

4

2. Methodology

We mainly use two methods from the spectrum of Digital Humanities: digital manual annotation and network analysis. With annotation, we chose a method that is central to hermeneutic text analysis and which is commonly used in the (Digital) Humanities (cf. [Jacke 2018]). However, annotation is most often used in the analysis of primary sources. In our approach, we annotate literary texts, as well as Beauvoir's *The Second Sex*. In order to make a list of all literary works she cited in the second part of her work and to connect all features named by Beauvoir to the gender roles she assigns them to, we annotated the second part of her work, in which she elaborates on her system of gendered roles. In a close reading process we thus annotated (a) literary works she cites and (b) features and typical clothes she describes for the different gender roles. We then did a context-analysis of the feature annotations in order to relate features and clothes to gender roles. Altogether we used an expert annotation approach in which one of us first did the whole annotation and then in a second runthrough the other one conducted the context-analysis and corrected annotations if necessary. In this second runthrough works, features, clothes and gendered roles were matched as the data was manually transferred to the network analysis software Gephi ([Bastian et al. 2009]). In a next step we turned the annotation of cited works into a list and built up a corpus containing some of these sources.

5

Having annotated Beauvoir's text we turned to these primary sources and annotated features, clothes and roles used to describe characters in a sample of 19 novels (see section 2.2). We used the same close reading expert annotation

6

approach as described above and built one database for each of the texts. To further analyze character features in the novels we used OpenRefine ([Krause 2021]) to disambiguate the features so that shared features of characters show more easily (see section 4). We then imported the data to Gephi to visualize it in one single networked graph.

Being a multidimensional network a graph can show different kinds of entities and relations. We ran a force-directed network algorithm that visualizes nodes with many relations to each other or with many shared relations to other nodes using proximity (cf. [Schumacher 2018]). Nodes that are neither connected to each other via direct nor indirect relations are shown as being located very far from one another. The algorithm was stopped when equilibrium was nearly reached, i.e. as the nodes and edges stopped changing their location. We thus compare the creation and functioning of a category system in two different research settings.

2.1 Corpus

We focused on two corpora. The first corpus (called non-fiction corpus) was made up of around 200,000 tokens taken from Beauvoir's theoretical text *The Second Sex*. The second corpus (called fiction corpus) was made up of 380,000 tokens taken from 19 novels she mentioned. In total in the second part of her work Beauvoir mentions 104 literary texts and text collections that can be categorized by genres including novels (84, some of which can be classified as autofiction), short stories (5), drama (7), and fairy tales (7). We annotated translations to German.[4] To create our corpora we followed an opportunistic approach (cf. [Schöch 2017]): we included all primary sources to the fictional corpus that could be found as German translations and in digital format.[5] We annotated the first 20,000 tokens per text, which led to ca. 6,700 annotations within the entire fictional corpus. The texts in this corpus cover a time span of 592 years ranging from 1353–1945 with a clear focus on 19th century literature. Twelve texts were written by men and seven by women. By using the same sources Beauvoir was using we ensure that our corpus consists of texts that were read in terms of analyzing gender roles and therefore are relevant for the topic. This means that in this corpus there are stereotypical representations of gender as well as uncommon depictions which illustrate roles that do not exactly fit into a binary system. In addition to that we use both types of texts to develop a transferable category system, which includes hints on gender roles and features from different sources. Using a digital approach to recreate Beauvoir's study provides us with the possibility to include all of the features, clothes and gender roles mentioned in a standardized excerpt of the primary sources. This enables us to work on the accuracy of the category system of gender roles before we transfer it to texts unknown to Beauvoir such as contemporary novels in future research.

2.2 Annotation

The annotation of both text types is aimed at differentiating and specifying gender roles and character features, so that we can speak of conceptualizing annotation (cf. [Pagel et al. 2020, 127]). Thus, the merging of both types of annotation data serves the development of theory.

While annotating the fiction as well as the non-fiction corpus we used a broad definition of our phenomenon of interest. In the case of both text types, we understand features as central, long-lasting characteristics of people and literary characters. They can be external such as beauty or internal such as intelligence. Emotions and actions can be understood as features if they are long-lasting characteristics. In the sentence "he is an angry young man" the emotion of anger is turned into such a character trait (cf. [Schwarz-Friesel 2007, 71]). In addition, we annotated descriptions of clothing. Finally we annotated gender roles as designations for characters. The gender roles were assigned to one of the gender categories "female," "male" and "neutral." The length of annotated passages in most cases includes single words (like "smart" or "famous"), a combination of two words (e.g. "long hair," "velvet neck"), four or five words that appear in the text as an enumeration ("sinister, stupid, stubborn, opinionated man"), or whole phrases, which often contain a comparison ("high as a tree"). When transferring the annotation data into a graph, we resolved enumerations into individual components. This means that for the properties "beautiful, round face" appearing together in the text, two entries were created, one for "beautiful face" and one for "round face." We annotated all features mentioned in the context of characterization, i.e. features a character shows, others project on it or would like it to develop more.

2.3 Graph Data

We used the network algorithm Force Atlas 2 to create graphs in which gender roles or literary characters are shown in their proximity to each other. This proximity is based on distinctive and overarching features, clothing and gender roles. We define distinctive features as features that occur only one time and overarching features as features that occur at least two times, i.e. in the profiles of at least two gender roles in the non-fiction corpus. Finally, we use the extrapolated gender categories to characterize gender spheres in Beauvoir's theoretical text as well as in a corpus of narrative fiction. To build up graphs representing the non-fiction corpus we manually extracted features and clothing from our annotations and traced them back to gender roles. In the fiction corpus we extracted features and clothing as well as gender roles used for characterization and traced them back to individual characters. In the graph showing Beauvoir's system of gender roles we thus came up with two types of nodes ("roles," and "features"). The nodes represent both the is-state ("wearing," "feeling") and the should-state ("shall have," "shall be," "shall renounce," "wants to have," "needs").The graph database we developed for the fictional corpus includes five types of nodes: "character," "feature," "role," "clothes,"[6] and "text." In addition the node type "role" has been assigned the property "gender" having the values "female," "male" and "neutral." We included the node type "text" in order to be able to see if certain literary trends might show or if characters of texts from a similar time period cluster together.

3. Beauvoir's System of Gendered Roles

In the following part, we describe the category system that emerged from the conceptual annotations of Beauvoir's theoretical opus magnum. In her text, Beauvoir describes gender roles by referring to sociological observations and by using examples from historical, psychoanalytic, and literary texts ([Botond 2017, 117]). In doing so, she undermines the division between disciplines and employs an approach that Sigmund Freud, for example, uses as well:[7] real-world oriented, person-related observations, phenomena, and roles substantiated through characters in numerous yet exemplary literary texts. Thus she develops a category system of gendered roles such as "mother," "father," "young girl," "young man," and so on. Her argumentation always follows descriptions of the female roles first. In most cases, male roles are developed accordingly. Comprehensibly, the focus on the female physique and psychology leads to a far more detailed description of female roles than of male roles. Further on, Beauvoir gives examples of roles that cross the border between male and female e.g. "virile woman," "tomboy," "feminine man" or "dandy." As these roles are not only characterized by traditionally female traits but also by male characteristics they incorporate diversity. As can be seen from Figure 1 Beauvoir indeed describes the roles "woman" and "man" as opposing each other. But some other, more unexpected particularities show when interpreting the graph data:

1. Although "woman" and "man" clearly oppose each other they do not show the largest distance in the graph. "Girl" is situated behind "woman" and "son" appears behind "man."[8]
2. By contrast, roles incorporating diversity are situated in the middle of the graph, but
    - linguistically female roles[9] ("lesbian," "virile woman," "tomboy") are situated towards the male pole of the graph, whereas
    - linguistically male roles[10] are situated closer to the feminine pole ("feminine man," "gay man," "dandy").
3. There are some cases of transcendence of feminine roles towards the male pole and vice versa:
    - "independent woman" is situated in close proximity to "man,"
    - "Wife" is situated in close proximity to "husband,"
    - "The crucified" is situated even behind the role "woman."
4. Neutral roles are underrepresented in the graph and are not situated exactly in the middle.

Altogether the graph representing gender roles as described by Beauvoir in *The Second Sex* appears to be quite spherical: male and female areas are transferring each other, especially female roles cover a wide stretched area inside the graph. As expected, female roles are more dominant in number than male roles are.
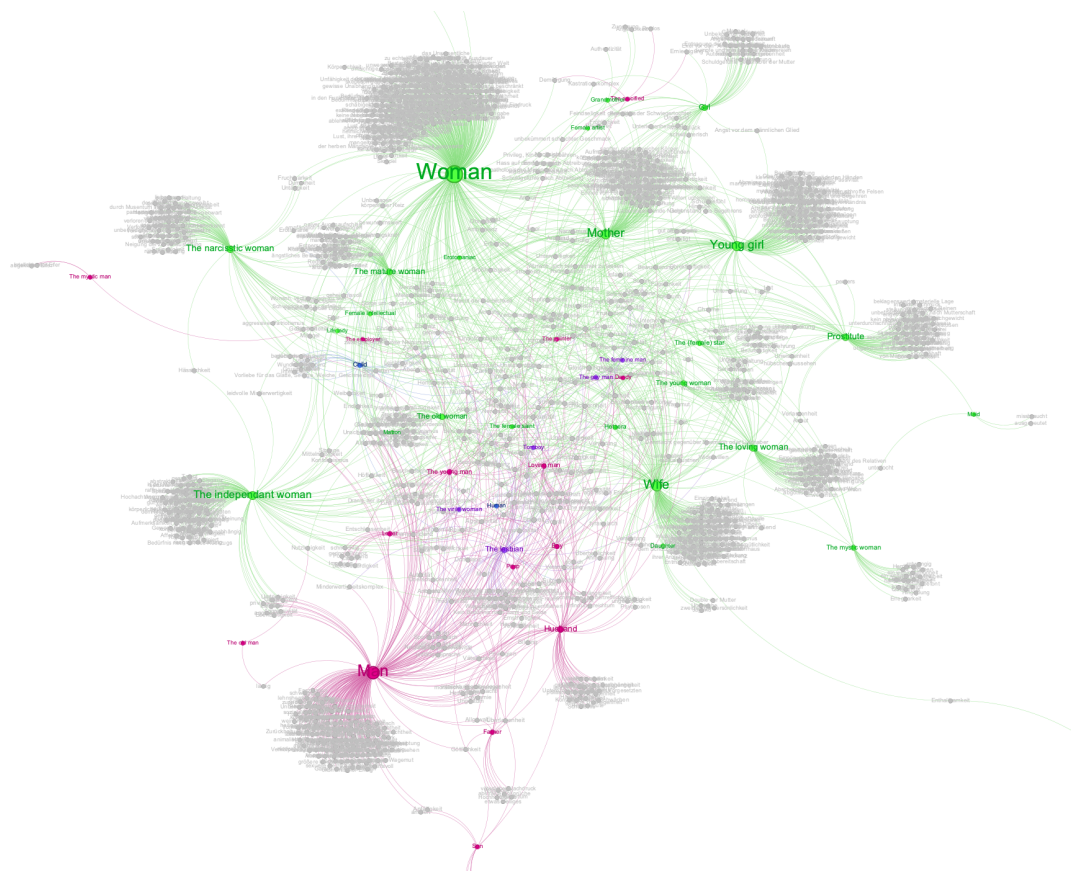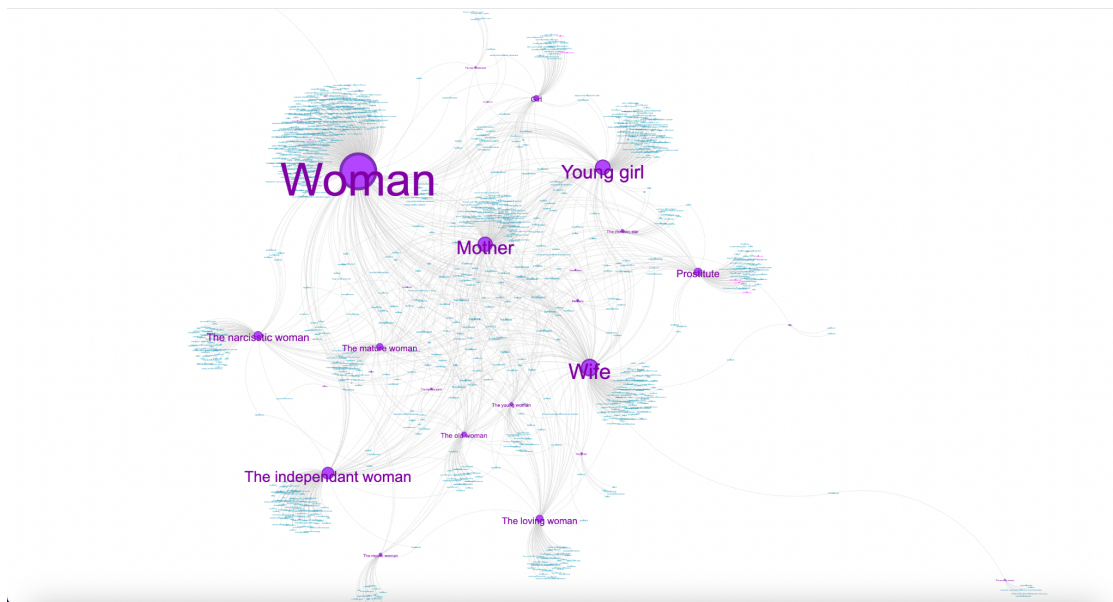
Figure 1. Gender sphere designed by Simone de Beauvoir in *The Second Sex*[11], green nodes represent female roles, blue nodes stand for neutral roles and red nodes designate male roles. Edges are coloured according to parenting nodes.

3.1 Mapping Beauvoir's Gender Sphere

In the following, we will first take a look at the presumed opposition of a binary system in which "male" and "female" are the crucial pair of opposites. In the second step, we analyze non-binary roles in Beauvoir's gender sphere (cf. Figure 1). Finally, we analyze distinctive and shared features of rather female, neutral and rather male roles. One interesting aspect of the graph shown in Figure 1 is that although "woman" and "man" oppose each other they do not inhabit the most extreme positions inside the graph. When excluding roles that only share a single feature with another role one ends up with having "girl" on one pole and "son" on the other. It is also noteworthy that the position of "girl" is further away from "woman" than "son" is from "man." In order to focus on shared features of linguistically female roles we modelled another graph only including female roles (cf. Figure 2).

Figure 2. Linguistically female roles (purple) and their features (blue) and clothes (fuchsia).

It is one characteristic of this graph that most roles are situated at the margins whereas shared features can be found in the middle with two exceptions: firstly, roles that are not connected to many features and that show a small number of individual features are also situated in the middle and secondly, the role "mother" is also positioned in the middle although it is indeed connected to many individual features. Why is that? The role "mother" is connected to 151 features altogether. 70 of these features are individual features whereas 81 features are shared with other roles, e.g. more than half of the features are shared features. The role "mother" is thus connected to 21 of the 28 other female roles inside the graph. Moreover the node representing the role "mother" with a score of 1 has the highest Eigenvector Centrality in this graph. The Eigenvector Centrality is a metric which informs about the influence of a node in a network. In this case it shows that "mother" is sharing numerous features with other roles and is also connected to many nodes representing features shared by many roles. It becomes apparent that the role "mother" in many ways is the most important female role. Zooming in on features shared by many female roles in this area (cf. Figure 3) we end up with the following top ten list (where *degree* refers to the number of connections to other nodes, in this case female gender roles):

1. Narcissism (degree 11)
2. Passivity (degree 10)
3. Jealousy (degree 10)
4. Coquetry (degree 10)
5. Beauty (degree 9)
6. Vanity (degree 9)
7. Dependency (degree 9)
8. Loneliness (degree 8)
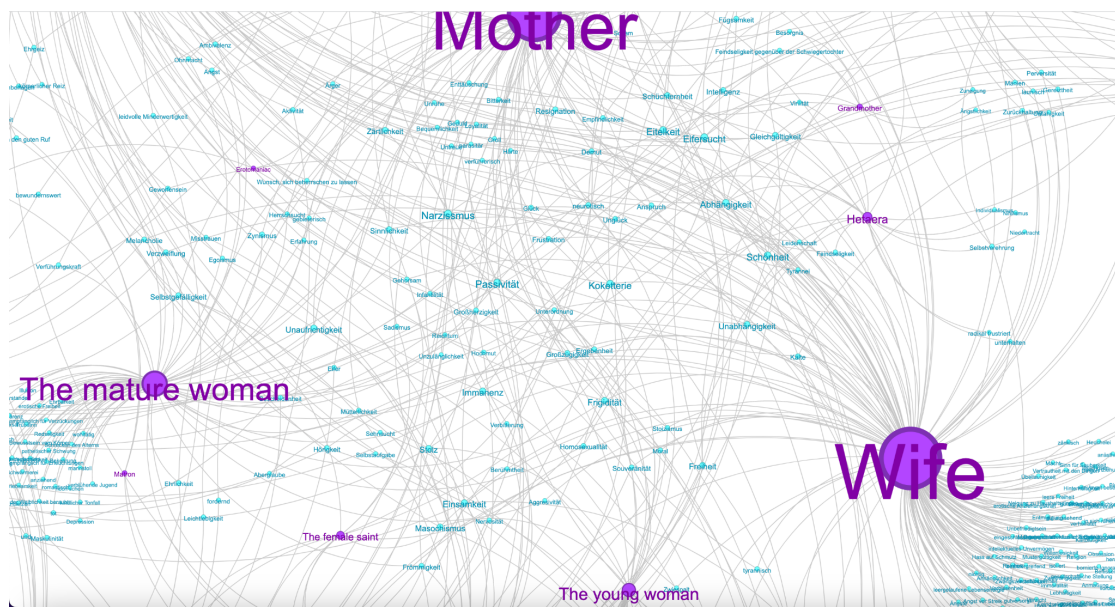9. Immanence (degree 8)
10. Frigidity (degree 8)

Figure 3. Zoom on the center of the female area of the graph representing the non-fiction corpus.

However, male and female roles that oppose each other inside the gender sphere do also share some features. The role "girl" e.g. shares features with "man," "husband," "pimp," "boy" "loving man," "young man," "dandy" and "crucified" (cf. Figure 4).
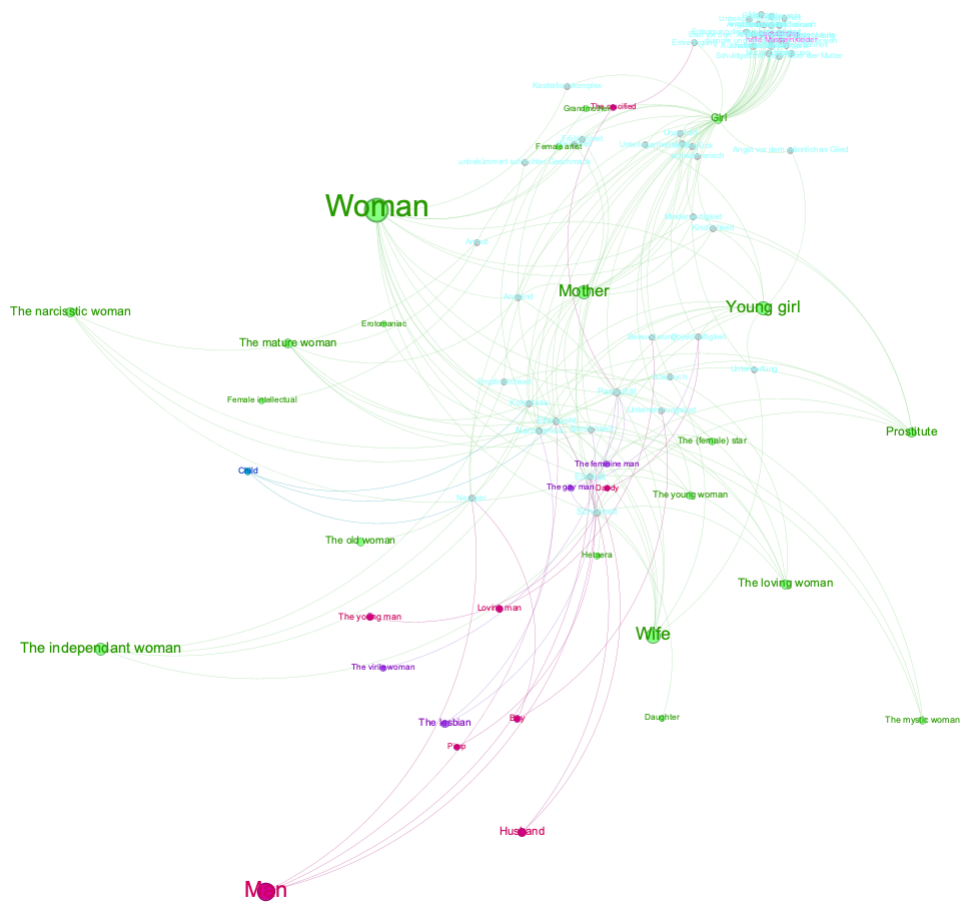
Figure 4. Ego-network of "girl" with a depth of 2, meaning that all features and clothes of "girl" show as well as all roles with whom "girl" shares features and clothes.

The "son" on the other side shares features with the female roles "independent woman," "old woman," "loving woman," "Hetaera," "(female) star" and "young girl" (cf. Figure 5).
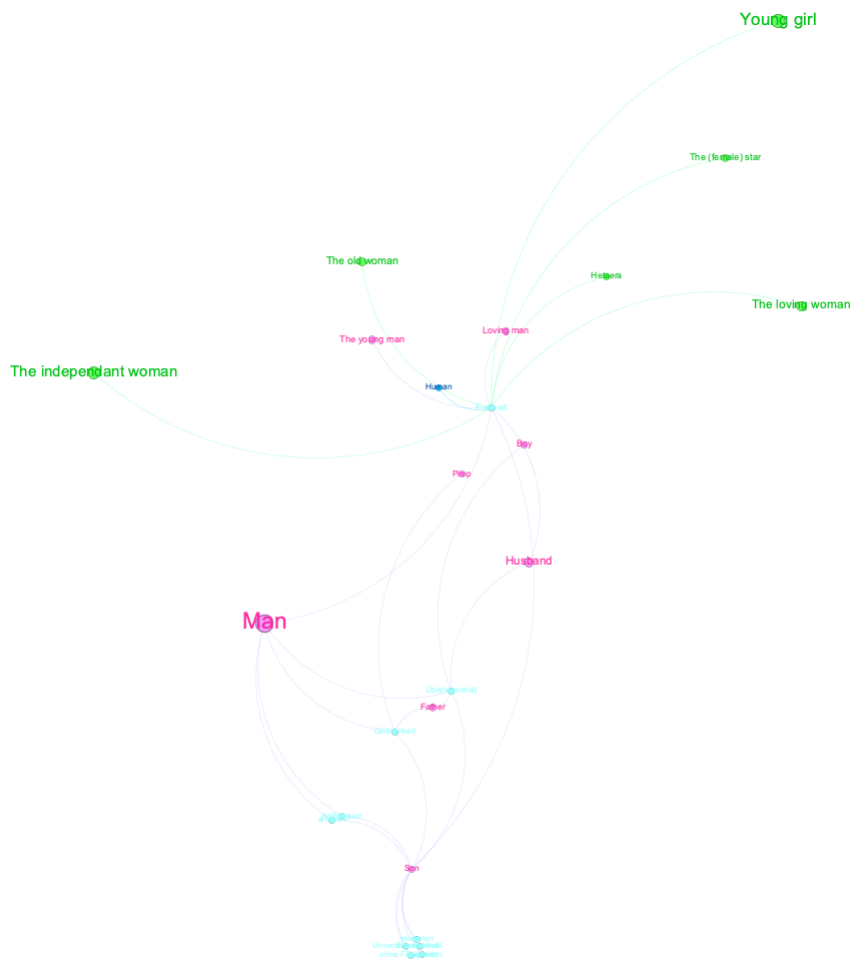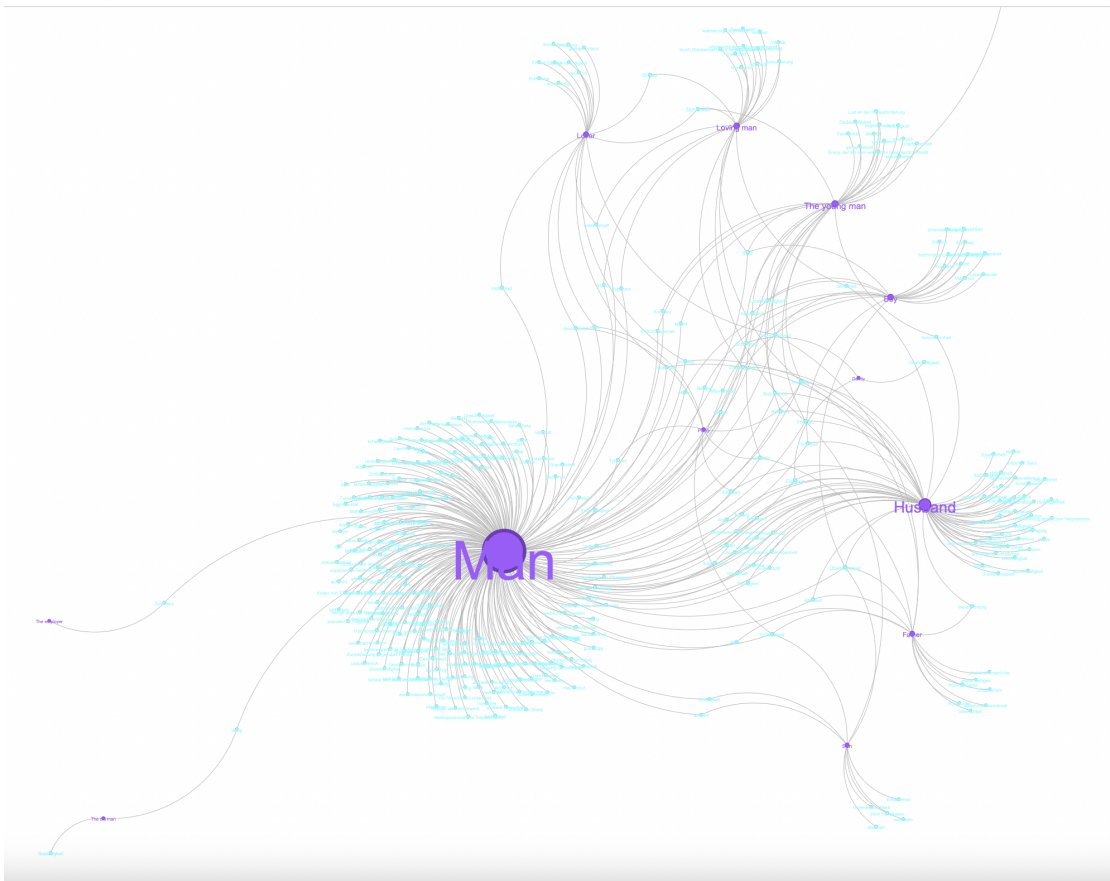
Figure 5. Ego network of "son" with a depth of 2.

As these two roles only mark the poles with the longest distance between them, obviously all other roles even share features with more roles altogether.

Figure 6. Linguistically male roles (purple) and their features (blue).

Now turning towards the male area of the graph we instantly see that it is much smaller, shows fewer roles that are connected to fewer features and no clothes (cf. Figure 6). Most roles are situated at the margins of the graph which means that their profiles include more individual than shared features. Solely "pimp" and "dandy" are situated towards the centre; the "dandy" showing at least one individual feature, the "pimp" being defined only by shared features. The top ten of shared features of the male sphere are:

1. Prestige (degree 6)
2. Superiority (degree 5)
3. Freedom (degree 5)
4. Transcendency (degree 5)
5. Pride (degree 4)
6. Divinity (degree 4)
7. Authority (degree 4)
8. Elegance (degree 3)
9. Beauty (degree 3)
10. Privileges (degree 3)

Except for the feature "beauty," which is top 5 for female and top 9 for male roles, the 10 most commonly shared features of female and male roles do not overlap. However in this unbalanced setting it is hard to compare the importance of these features as Beauvoir in general describes male roles in much less detail than female roles. We will therefore try to rebalance these findings in comparison with the outcome of the second part of our study.

We now have briefly outlined the male and the female areas of the graph that create opposing spheres. In the following we turn to roles that transcend binary approaches to gender.

19

20

21

3.2 Non-binary Areas in Beauvoir's Gender Sphere

Taking such transcending roles as a starting point we now want to shed some light onto those areas of the gender sphere that could possibly represent gender diversity. In the course of this, we will analyze roles that linguistically could be identified as male or female but appear inside the opposite area in our graph. We call those roles that appear inside the female area of the graph "rather female" and those that appear inside the male area "rather male." It is a characteristic of the graph representing Beauvoir's approach to gender that the category "rather female" includes only linguistically male roles whereas the category "rather male" includes only roles that are linguistically female. We want to emphasize again that proximity inside the graph is algorithmically created on the basis of shared features and clothing (although Beauvoir does not often operate with clothing), so that roles sharing a higher number of features appear closer to each other than roles sharing less features. The following transcending roles, which we classify as "rather female" and "rather male" (cf. Table 1) can be found. Finally we faced the question of what to do with neutral roles or such roles that appear on the borderlines between the female and the male area of the graph. For the time being we classify those as potentially neutral. While comparing the graph representing the non-fiction corpus with the one showing data gathered from the fiction corpus we will revisit this classification.

rather female (rf)	potentially neutral (n)	rather male (rm)
feminine man	child	lesbian
painter	employer	virile woman
gay man	old woman	independent woman
dandy	young man	
crucified	loving man	
	female saint	
	tomboy	
	human	
	matron	

Table 1. "rather male," "rather female" and potentially "neutral" (gender) roles

The classification shows that there is an almost equally small number of rather female and rather male roles, whereas comparatively many roles can be found in the borderlands of female and male areas. The category "potentially neutral" contains only two linguistically neutral roles being "child" and "human." The other six categories are more troubling but also more interesting cases. There are three linguistically female ("old woman," "female saint" and "tomboy") and three linguistically or grammatically male roles ("employer,"[12] "young man" and "loving man"). To shed more light on the characterization of roles that are potentially not fitting into a binary gender system we analyzed distinctive and shared features of roles according to this categorization.

First we sorted out features that were only used to characterize roles falling into one of the three categories mentioned above. In order to do so we filtered the graph data to ego networks. The ego networks focus on each of the aforementioned roles and show their next two neighbours (these are the features used to describe a certain role (neighbor 1) and the roles they share those features with (neighbor 2)). Second we found out which features are used to describe roles from two categories. Finally we compared features that are only shown by roles in one of the gender categories with those that are shared by two or all three categories. We thus came up with lists of features that were at least shared by two roles. If we take a closer look at distinctive and overarching features of potential top-level categories under this premise, the first thing to mention is that overarching features are more rare than distinctive features (cf. Figure 7).

Figure 7. Distinctive and overarching features of non-binary gender roles.

It is obvious that Beauvoir concentrates mostly on roles that can be defined as linguistically female. This explains – at least partly – why rather female roles show only two distinctive features. Potentially neutral roles show many more distinctive features (26) but here we have to keep in mind that this category includes nine roles altogether. Finally rather male roles show the most distinctive features – 84 in total. This is mostly due to the fact of the extraordinary amount of detail in which the independent woman is described by Beauvoir.

25

That the categories are indeed fuzzy sets (cf. [Zadeh 1965]) and permeable towards neighboring categories becomes very clear by taking a closer look at the shared features of roles from different categories. The categories "rather female" and "potentially neutral" share 6 features. On the other side the categories "neutral" and "rather male" are even closer and share a total of 24 features. Finally "rather female" and "rather male" as opposing categories of the non-binary gender sphere share 7 features. There are also five features shared by all categories of these three gender categories.

26

As one can see from Table 1, Figure 1, and Figure 7, an attribution of the center of the gender sphere with neutrality turns out to be questionable. Even though truly neutral gender roles such as "human" can be found here, there are also a variety of roles situated at the borders of the female and male areas that are associated with other factors such as religion (female saint), love (loving man), or age (old woman). Sometimes these factors can turn a role towards something one could call gender neutrality with a feature like "having no sex" attached to "old woman." In some cases an opposite association is constructed such as being the "desire" of the "loving man." What Beauvoir creates is more of a diversity sphere situated in the middle of the graph stretching from "rather female" to "rather male" roles.

27

3.3 Mapping the Gender Sphere or: From Areas to Categories

In this section, we will introduce a system for gender categories which involves five different main categories for gender, each containing several gender roles or sub-categories that are more or less firmly attached to one of the main categories. The five gender categories are: "female," "rather female," "neutral," "rather male" and "male." Starting from defining areas inside the graph we call Beauvoir's gender sphere we consolidated a category system in which we organized gender roles mentioned by Beauvoir according to these five top-level categories. Following Beauvoir's notion of not being born to have a certain gender but being made to show features of it we mainly ignored a biological notion of gender, although we do use linguistic information on gender such as pronouns. However we use this information very

28

carefully as we know that the use of "he" and "she" pronouns might only be due to a lack of neopronouns in our source text from 1949. Our approach is based on the previous data analysis in which we mostly focussed on shared and common features and used network algorithms that group nodes according to higher numbers of edges linked to them. We ignored roles and features that show insignificant numbers of connections (namely roles that are either linked to the main graph by a single shared feature only or not connected to the main graph at all). For our category system that means that we did not include all gender roles mentioned by Beauvoir. In the end, the categories and gender roles can be visualized as shown in Figure 8.
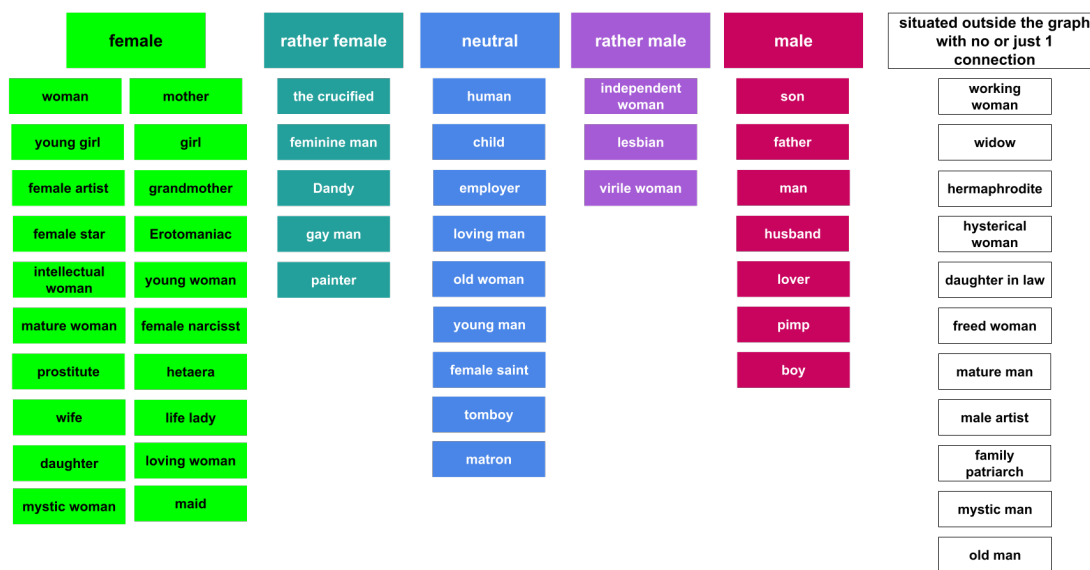
female		rather female	neutral	rather male	male	situated outside the graph with no or just 1 connection
woman	mother	the crucified	human	independent woman	son	working woman
young girl	girl	feminine man	child	lesbian	father	widow
female artist	grandmother	Dandy	employer	virile woman	man	hermaphrodite
female star	Erotomaniac	gay man	loving man		husband	hysterical woman
intellectual woman	young woman	painter	old woman		lover	daughter in law
mature woman	female narcisst		young man		pimp	freed woman
prostitute	hetaera		female saint		boy	mature man
wife	life lady		tomboy			male artist
daughter	loving woman		matron			family patriarch
mystic woman	maid					mystic man
						old man

Figure 8. Gender category system developed from Beauvoir's *The Second Sex*.

The gender roles inside a category vary in their connection to the top-level category, which we visualized by means of order. Gender roles named at the top are situated in a position inside the graph that can clearly be identified as belonging to an area dedicated to the top category. Gender roles named towards the bottom of the list are situated towards an area clearly defined as belonging to another top-level category or further towards the margin of the graph. This means that gender roles on top of a list always are more accurately connected to the top-level-category, i.e. the top-level-category usually is the only category this gender role fits into. This arrangement highlights that the above named categories are permeable (which is a typical feature of fuzzy sets as we mentioned in section 1) and have intersections (cf. [Zadeh 1965, 342] for intersections of fuzzy sets). Again representatives of the category we call "potentially neutral" can be seen as especially debatable cases. In these cases we have a rather eclectic collection of a) roles that are really neutral ("human," "child"), b) roles clearly influenced by certain factors namely religion, age, love and work ("female saint," "old woman," "young man," "loving man," "employer") and c) the special case of the "tomboy" and the "matron." This results in the category "neutral," unlike the others, not being a fuzzy set at all (because no clear relation between all of the roles can be found). Neither do the sub-categories show significant amounts of similarity. Here we deal with a loose collection of gender roles that are only connected by their central position right in the middle of the gender sphere. We would like to stress the fact that this is a preliminary attempt to gender categorization which we will overhaul in the second part of the study. However, another fact one must keep in mind when viewing the gender sphere graph and that can clearly be seen from Figure 8 is that Beauvoir's approach is an example of the female perspective on gender roles. The number of "female" roles is much higher than the ones of all the other categories. In the end, the analysis of Beauvoir's gender sphere has provided us with a preliminary category system containing five top-level categories of which four can be understood as fuzzy sets with a set of sub-categories with permeable borders. Each of the sub-categories in this system is associated with a (mostly) high number of features that can be either distinctive or overarching.

Using both roles and features Beauvoir's approach turned out to be an ideal starting point for building a category system that not only sticks to binary gender roles but also sheds light on the characterization of non-binary genders.

However, it was not possible to develop a system in which all top-level categories could be constructed according to the same basic characteristics of fuzziness. Instead, we have to operate with a certain degree of inconsistency in the categorization of genders. Nevertheless, the category system can be used to approach questions of gender representations in literary texts. But in order to gain a deeper understanding about the endowment and architecture of gender roles in literary texts and to refine the category system using our computational approach we turn to the literary sources Beauvoir used to build up the foundation of her gender sphere.

4. Test Case Literature: Gendered Character Descriptions

In this chapter we will describe the postprocessing of the annotations in our fiction corpus and shed some light on the distribution of features and roles within it. We mainly focus on collective and individual features and roles but this time they are bound to individual literary characters or groups of characters instead of the gender roles they might stand for. [13] To be able to interpret the large amount of annotation data, we disambiguated all annotations. We performed the removal of linguistic ambiguities semi-automatically as well as manually using OpenRefine.[14] Disambiguation involves different forms of unification at a grammatical and a semantic level.[15]

<div style="text-align:right">31</div>
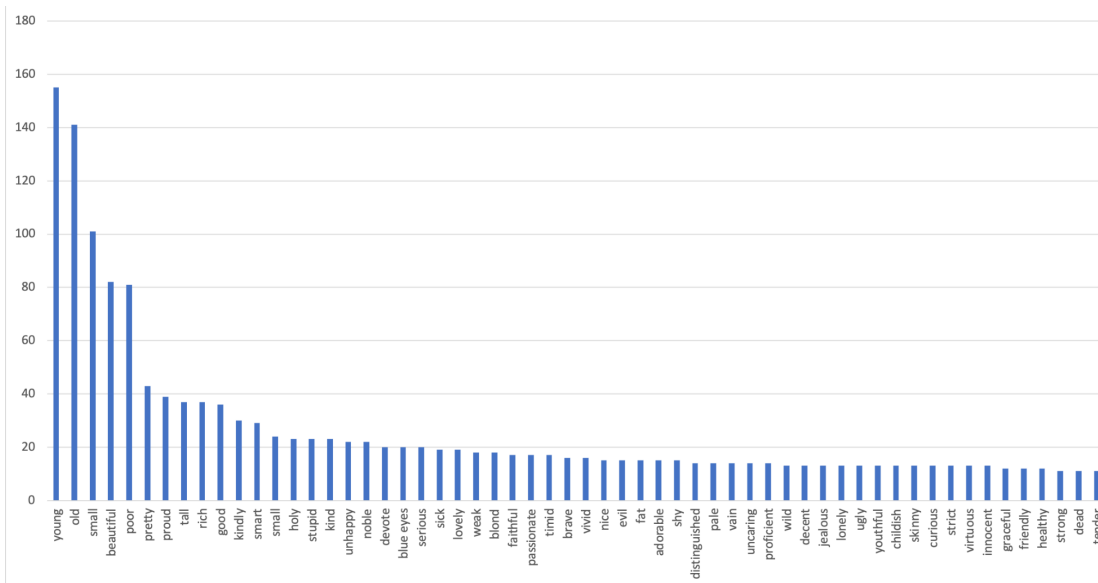
After disambiguation, a quantitative evaluation of the annotations shows that a total of 4,558 features,1,283 roles and 519 clothes were annotated in the entire corpus. Among these features, two broad areas can be identified: collective features and singular features. We define collective features as features that occur with high frequency (i.e., repeatedly and more than ten times, meaning that more than ten characters are described by these features). By singular features, we mean features that occur only once in the entire corpus and therefore are also used to describe one single character only. Both groups can be further categorized into internal and physical features. Internal features are not visible from the outside and describe the personality of a character. Physical features describe the appearance and the physiognomy of a character and are visible from the outside.

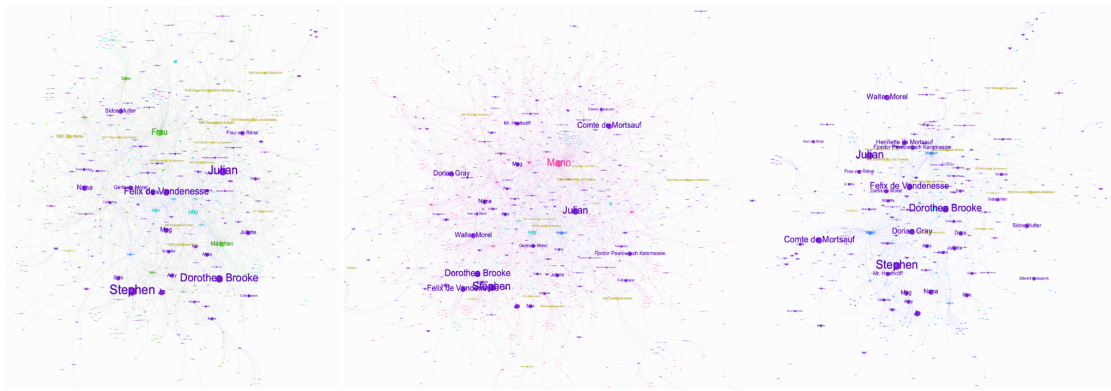<div style="text-align:right">32</div>

4.1 Collective Features

458 features occur more than 10 times and therefore can be seen as the most frequently occurring features or collective features (cf. Figure 9). In comparison with the graph data shown in Figure 10 and Table 2 it becomes clear that collective features like "young," "old" "good" or "friendly" are rather general, i.e. are used to describe characters of all gender categories, and are therefore gender non-specific. An exemplary analysis of the collective feature "young" shows that it is tied to characters mostly described using female roles in 63 cases, to characters mostly described using male roles in 79 cases, and to characters mostly described using neutral roles in 12 cases.[16] The hypothesis that can be derived from this distribution is that this feature is not gender-specific. The case is quite different for the likewise high-frequency feature "old." This attribute is assigned to characters that embody primarily (i.e. more than 50%) female gender roles in 44 cases, to characters that embody primarily male gender roles in 99 cases, and to characters that embody primarily neutral gender roles in 5 cases. Based on the unbalanced distribution, we can speak of a gender-specific pattern.A closer look at collective features also shows that they are primarily internal features, with few exceptions ("blue eyes," "tall," "skinny"). Unlike in the case of the non-fiction corpus, where the collective features are more common and singular features occur less frequently, with 1,514 different types of features in total, the group of collective features is comparatively small in the fiction corpus.

<div style="text-align:right">33</div>

Figure 9. Most frequent collective features.

When we filter the graph representing the fiction corpus in order to build neighbor networks with a depth of 1 focussing female, male and neutral roles (cf. Figure 10), the first thing that shows is that especially main characters show up in all networks. This is due to the fact that they are described in much more detail using many more features and clothes than are used for minor characters. This means that the features mostly used to describe the characters in the three filtered versions of the graph are not exclusively used for characters showing roles of one gender or the other. Nevertheless we see some patterns for gender-(non-)specific features when analyzing the top ten collective features.

Figure 10. Graph representing the fiction corpus filtered to show neighbor networks of female (left), male (center) and neutral (right) roles.

When we compare the top ten features in all three graphs we see which features tend to be used in a gender-non-specific way and which features are probably more gender-specific. In Table 2 we listed top ten features for the three filtered versions of the graph and colored them according to a heatmap: features written in red show up in all three top ten lists, features in orange are present in two of the graphs and green features in only one of them:

Top	Female filter	Male filter	Neutral filter
1	Young 34	Young 40	Young 38
2	Small 32	Old 21	Great 15
3	Poor 24	Pretty 15	Good 15
4	Pretty 23	Great 14	Sweet 10
5	Beautiful 21	Blue eyes 13	Smart 8
6	Old 11	Pride 11	Ugly 5
7	Rich 9	Vanity 9	Innocent 5
8	Great 9	Fury 7	Simple 4
9	Charming 9	Dignity 6	Male 3
10	Delightful 8	Holiness 6	Female 3

Table 2. Top ten features of the gender-filtered representations of the graph shown in Figure 10

Comparing this data to the top ten feature lists we generated from the non-fiction corpus we see that the overlap is marginal. For female roles only "beauty" is present in the top ten list created from the non-fiction corpus as well as the one generated from the fiction corpus. For male roles the overlap is higher with "pride" and "beauty" showing up in both lists. Also Beauvoir highlights "divinity" for male roles whereas in the fiction graph "holiness" shows a tendency of being used with male roles rather than with female roles. Although these terms are not synonymous they show at least some similarities. What we also see from Table 2 is that features with the highest degrees are rather used in a gender-non-specific way whereas more rarely used features show tendencies of gender-specific usage. Bringing together this characteristic with the fact that individual features are much more in number than collective ones we get the impression that an individuation-based approach in the form of character profiling could be most fruitful in terms of gender analysis.
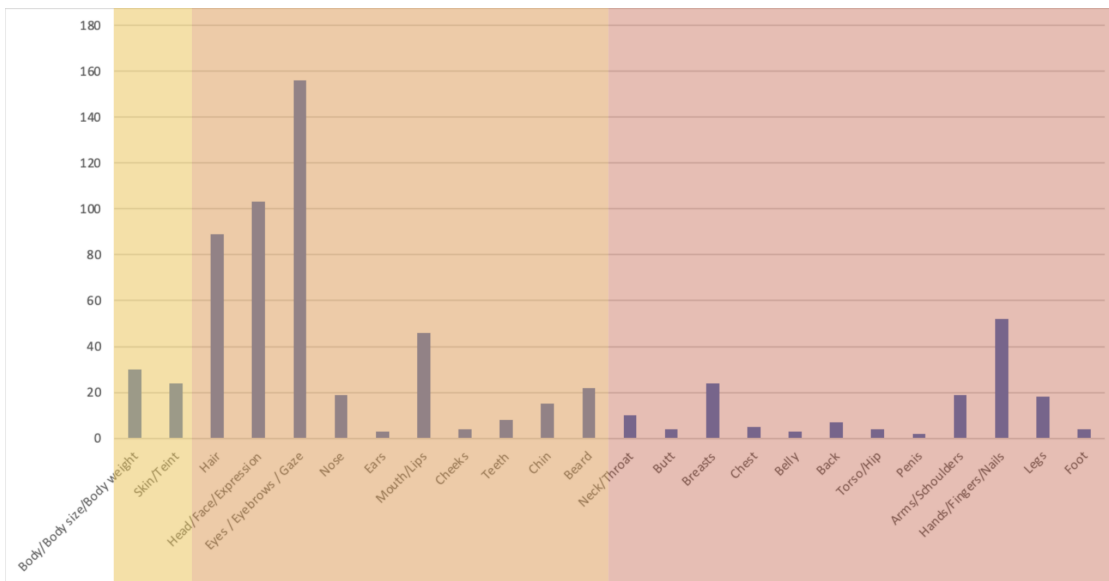
36

4.2 Singular Features

The feature types that occur less than ten times are mainly singular features: 2,613 features occur only once in the entire fictional corpus.[17] These hapax legomena among the features are of extraordinary importance in the feature profiles of the characters in our corpus. Especially main characters show a high number of features that occur only once in the entire corpus. They include specific, detailed attributes, which can only be listed here by way of example: "cynicism," "gypsy-like," "fragmented mind," "need for tenderness," "X-legs," "withered face," "restless gaze," "theater-obsessed," "magnificent teeth," "diplomat's face," or "thick eyebrows" are among them. Thus, the multifacetedness of internal singular features is much greater than for collective features. While a majority of internal features refer to a concrete inner character trait, they can also be combined with certain body regions. The physical features appear in kaleidoscopic diversity, as can be seen from examples like "eyes like burnt fried eggs," "snail green eyes," or "knees like pink shimmering islands." Most frequently singular-physical features are related to the eyes (156 different features), face (103), and hair (89 different features) of the characters. If we look at the total stock of singular-physical features (Figure 11), we see that they occur mainly in relation to the face (orange region), in some cases to the torso (red region) and much less frequently in general references to the body and the skin (yellow region).

37

Figure 11. Singular-physical features combined with body parts (yellow: general, orange: face, red: torso).

The analysis of the annotation data shows that characters in the fiction corpus are primarily defined by singular-internal features which are unique in content and appearance, while a small set of more general, gender-non-specific features can be interpreted as a fixed core of recurring features.

The quantitative evaluation of annotation data allows a differentiation of the two categories collective and singular features which can be used to annotate character features in literary texts. While the distinction between collective and singular features is a concrete result of the explorative annotation process, the analysis of annotation data shows that it seems to be useful to include the categories physical and intrasubjective right from the beginning in order to get more precise results. Another conceivable subcategory that we did not include, but that emerged during the annotation process and could be interesting for a more detailed analysis of character features, is "talent" or "aptitude." Assigning features to clearly defined categories is one way to look for gender-specific features. However, this approach lacks a reference to gender roles and clothes. Such a multi-faceted look on the data is the goal of the following section.

5. Analyzing a Literary Gender Sphere

In this part of the paper we bring together the annotation data from the fiction corpus (i.e. the features, clothes and roles of the characters) with the gender categories developed using the data gathered from Beauvoir's *The Second Sex*. In order to handle the amount of data and keep the setting comparable we use Gephi ([Bastian et al. 2009]) as network visualization software again and stick to the same algorithm used for the gender sphere of Beauvoir's *The Second Sex*, named Force Atlas 2. It is not possible in the scope of this paper to fully discuss roles, features and clothes and their respective use to build character profiles. Instead we use all three parameters to model the graph and to base positions of characters inside it on all three dimensions. Having already analyzed collective and singular features from the annotation data, in the analysis of the graph we highlight gender roles in order to proceed with the categorization of gender. From previous research (cf. [Flüh et al. 2022a], #fluh_etal2023, [Flüh et al. 2022b] we know each character in a literary text is usually associated with more than one single gender role. Protagonists are often characterized by a large number of different gender roles. So in addition to features and clothing we annotated all depictions of characters as potential gender roles. Features, clothes and roles were assigned to individual characters thus building an individual character profile. This assignment is based on close readings of the text chunks features, clothes and roles appear in. We classified the roles according to the basic gender categories "female," "neutral" and "male" in a first step as we did in the first part of this study. The color code is the same we used in Figure 1: green for "female," blue for "neutral" and red for "male." Features, clothes and roles are the most important basis for the positions of characters in this network. In addition to that work titles hold together characters of one text a little. In order to focus on gender categories we gave all but roles nodes a gray color here which turns them almost invisible. Focussing on the gender roles three things become

38

39

40

apparent:

1. In contrast to the data from the non-fiction corpus in the fiction corpus male gender roles are more dominant than female roles,
2. nevertheless female and male roles are more balanced in this corpus than in the non-fiction corpus,
3. in the center of the graph representing the fictional corpus a zone of neutral roles is situated.

We conclude that in contrast to the graph data of the non-fiction corpus the graph of the fiction corpus shows a scale. However we need to look more closely in order to see if and where diversity finds its place in this corpus. But we would like to stress the fact that neutral roles are way more important in the fiction than in the non-fiction corpus. This meets our general impression that neutrality is often overlooked in gender theories.[18] From the graph data representing the fiction corpus (cf. Figure 12) one can see that even in historical data one can frequently find roles associated with more than two genders, mostly being either female or male and neutral. But it is less common to find profiles that show an equal amount of or more or even exclusively neutral roles than such that include either male or female roles only.[19]
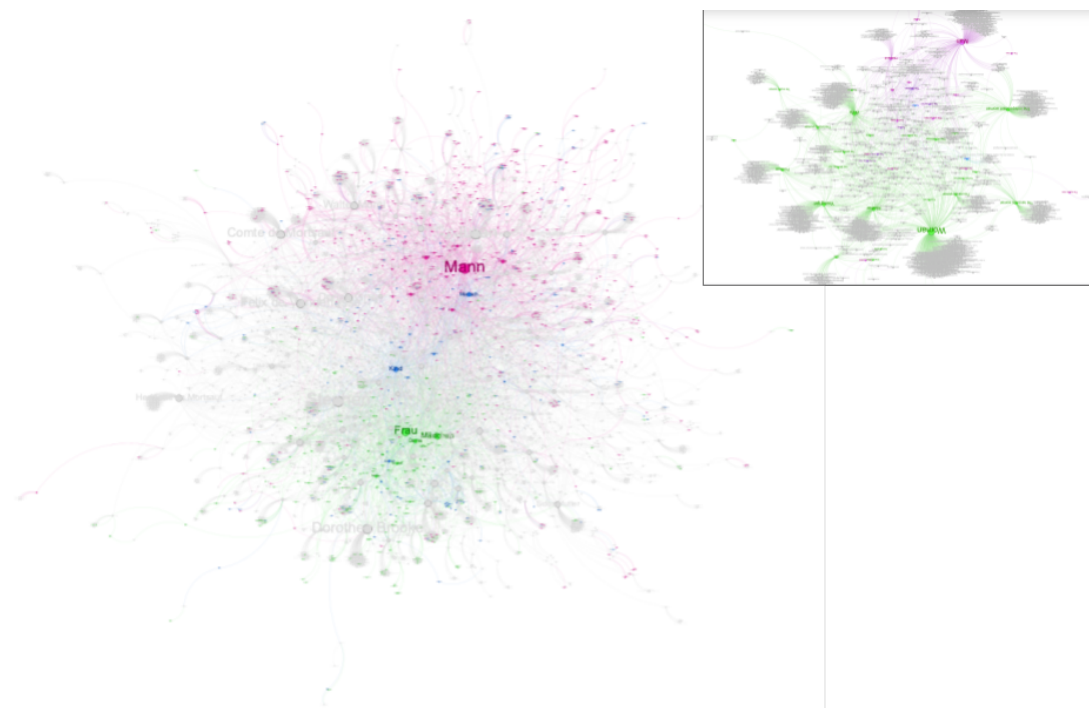

Figure 12. Gender sphere of the fiction corpus in comparison with the non-fiction corpus.

The fact that we are dealing with characters described using features, clothes and roles in the fiction corpus and not with roles being described by features leads to a different organisation of the graph: roles are situated less towards the margin and more towards the center. It is due to this characteristic that the most frequent male, female and neutral roles show in the middle of the area they are situated in. Zooming into the graph one can see that in the fiction corpus transcending roles, i.e. roles being classified as one gender category in an area mostly occupied by roles of another category, are appearing very frequently. Nevertheless, mostly these transcending roles are represented by very small nodes as they are only used to describe a single character. As we did before we exclude these cases from our categorization as we do not think this to be enough data to base a valid interpretation on. Once again the category "neutral" provides us with some special cases which we will analyze in some detail later on. But first we turn towards the category "male."

By zooming in on "man" as the most frequently used male role to describe characters in the fictional corpus, we can identify roles most closely connected to the center of the male area of the graph as well as characters mostly described using this or similar roles and works in which these highly frequent male roles are used especially often (cf. Figure 13).
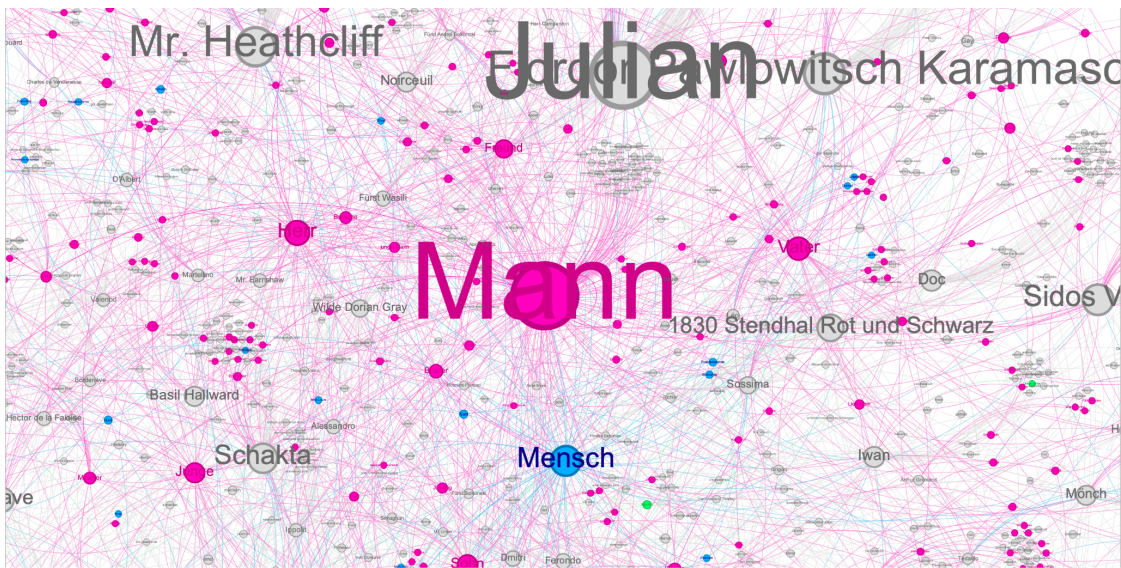
Figure 13. Zoom into the network for a closer look at the role "man."

Highly frequently used roles that can be found in close proximity to "man" are e.g. "brother," "father," "(male) friend," "mister" and "lad." Characters that show many such roles in their profile are e.g. Julian from *Red and Black* and Fjodor Petrowitsch Karamasow from *The Brothers Karamasow*. Novels in which a high number of these roles appear are – unsurprisingly – *Red and Black* and – more strikingly – *The Portrait of Dorian Gray*. By looking at graphs we modelled using data of single texts only we can validate that indeed most of the characters introduced in the first 20,000 tokens of these two novels show mostly male roles in their character profiles. Actually in both novels there is only one character that shows a high number of features as well as female gender roles in their profile as can be seen from Figure 14.
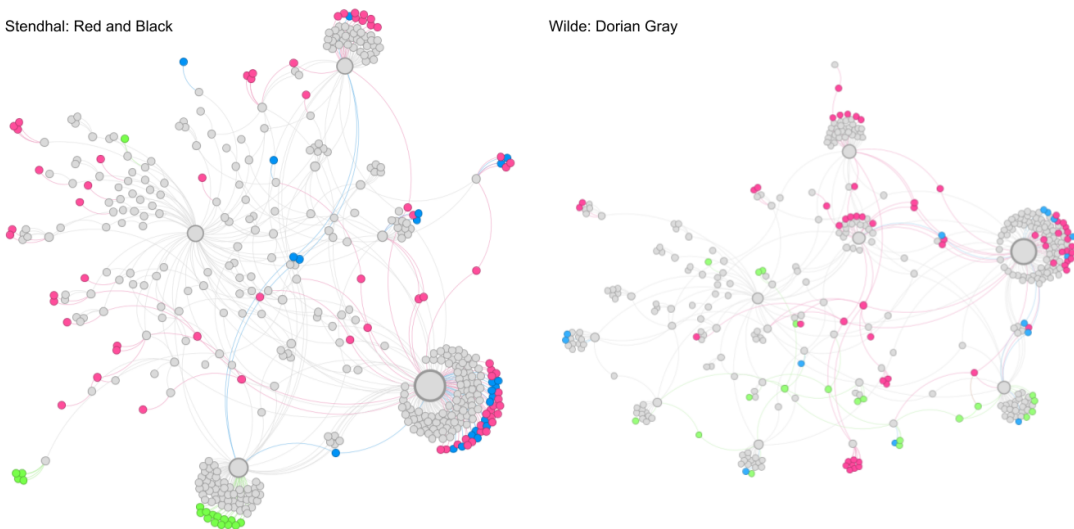
Figure 14. Networks of single works highlighting gender roles: *Red and Black* and *Dorian Gray*. (Male roles are colored in red, female roles in green and neutral roles in blue. All other node types are represented in grey. Edges are colored like their parenting nodes.)

By gradually zooming out of the graph we sorted the most frequent gender roles into the top level category "male."

What can already be seen from Figure 13 is that "human" as an intrinsically neutral category shows up in close proximity to the role "man." This means that in fact the role is more often used to describe characters that show a high number of male roles in their character profile than for roles that show a higher number of roles of other genders. We will see that there are equivalents of this phenomenon in the female area of the graph later. But first we would like to

zoom in on the center of the graph where neutral roles like "child" can be found (cf. Figure 15).
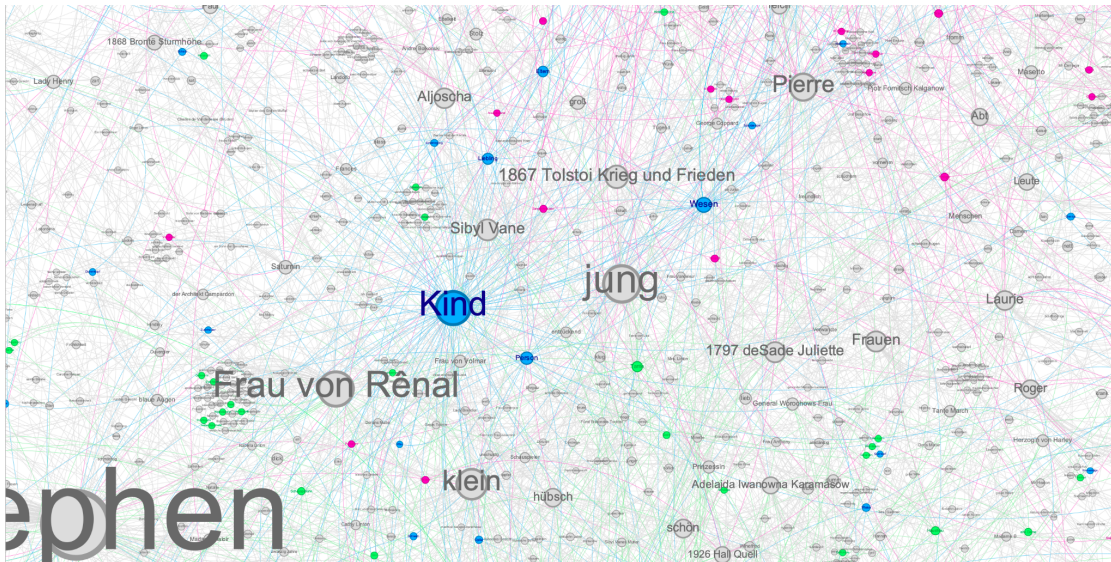

Figure 15. Zoom into the center of the graph for a closer look at neutral roles.

Again by gradually zooming out we identified roles for the top-level category "potentially neutral" from the preliminary category system we developed earlier. Roles that are situated here are "person," "being," "darling," "parents," "intruder" (being the neutral "Eindringling" in german), "guardian angel" and "audience." Roles in this category could be further grouped into terms used for single persons and groups, as nicknames or swear words. Characters that can be found in close proximity of these roles are Sibyl Vane from *Dorian Gray* or Madame Volmar from *Juliette*. Novels that are situated close by are *War and Peace* as well as *Juliette*. In contrast to the findings from the analysis of the male area this does not mean that there are especially many characters in these novels that show neutral roles in their profiles. As Figure 16 shows *Juliette* indeed features a main character that is attributed with many neutral roles whereas *War and Peace* has a comparatively more balanced number of male and female roles used to describe main characters.

Figure 16. Networks of single works highlighting gender roles: *Juliette* and *War and Peace*.

The most central and highly frequent female roles in addition to "woman" are "Lady," "girl," "daughter," "(female) friend," "widow," "nun," "mother" and "sister" (cf. Figure 17). Characters that can be found closely to the center of the female area of the graph are Atala and Juliette from the novels of the same names. There are no novels that are positioned so

47

48

closely next to the role "woman" as *Dorian Gray* and *Red and Black* are to "man," which we see as an indicator that there is no novel constructing some kind of "women's world" in the opening passage such as *Dorian Gray* takes its beginning in a clearly male-dominated setup. But there are *Pot Luck* by Balzac and *Well of Loneliness* by Hall as two novels that can be found not far from the center of the female area of the graph.
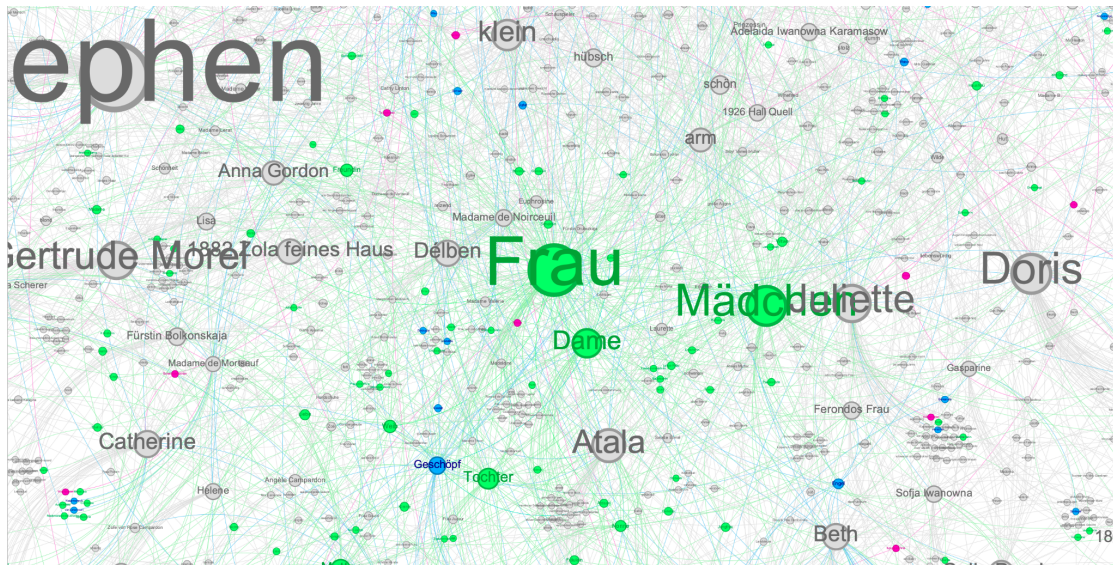

Figure 17. Zoom into the center of the graph for a closer look at female roles.

The graphs representing the data of these two novels show that *Pot Luck* indeed has more characters that are mostly characterized by female roles than characters mostly described using male roles. The *Well of Loneliness* is a very special case because it contains more characters with mostly female roles in their profile as well as one character that shows female, male and neutral roles in an almost balanced way. This is the main character Stephen who we found to be one of the very rare characters in this corpus that can be called "diverse" in terms of gender (cf. Figure 18).

Figure 18. Networks of single works highlighting gender roles: *Pot Luck* and *The Well of Loneliness*.

Altogether for male as well as neutral and female roles, family relations are an important factor. Similar to the male area yet less apparent is the transcendence of a frequently used neutral role into the female area of the graph. Whereas characters mostly described by male roles are highly frequently also referred to as "human," characters that are described using mostly female roles are often also depicted as "creature" (although the German word "Geschöpf"

actually found in our data is much less negatively connotated). Another gender reference transcending from the neutral area into the female sphere through frequent use in mainly female character profiles is "thing." So although all three terms – "human," "creature" and "thing" – are intrinsically not bound to a certain gender the frequency of use in combination with male and female roles differs significantly. However it can be seen from the whole graph that many neutral minor roles, i.e. less frequently used to describe characters, can be found all over the graph. Due to the focus of this study which does not lie on the analysis of gender neutrality we have to postpone further studies of the subject to future research.

Finally we turn to the question whether there are roles used inside this corpus that could be classified as rather female or rather male. If we use the preliminary category system developed in part one we come to a disappointing insight. Most of the roles suggested by Beauvoir are not used in this corpus to describe characters. No "gay man," no "lesbian," no "independent woman." Only two roles of the category "rather female" can be identified being "dandy" and "painter." This is not a very surprising finding, when we keep in mind that we are dealing with historical data from the timespan 1353–1945. But does this mean that not any gender diversity is present in this corpus? Bringing together our finding that Stephen from *Well of Loneliness* is profiled using a balanced amount of female, male and neutral roles with the fact that Beauvoir mentions the heroine of this novel in the section of her book called "the lesbian" (cf. [de Beauvoir 1992, 855]) we assume that gender diversity (except for the very rare cases in which "dandys" are named as such) might not be explicitly mentioned in our fiction corpus but profiled using a variety of roles of the other gender categories. Once again, we are facing the problem of fuzziness. Whereas we can see from the graph data of the fiction corpus that it is very common that characters show a majority of either male or female roles together with some neutral roles and one role of the opposite gender (when assuming a gender scale as we do from analyzing the graph data), roles that show at least two roles of the opposing gender in their profile are more rare. Finally characters like Stephen that show a profile with an almost balanced amount of roles of three gender categories are exceptions. While we are certain that such a profile can be used to describe a character that does not fit into the boundaries of a binary notion of gender without stating it explicitly, the question whether it would suffice for a character to show two gender roles of the opposing gender (in relation to the majority of gender roles in its profile) in the first 20.000 tokens of a narrative must be postponed to further research in which we analyze them as whole texts.[20] However there is one character in our corpus showing a very similar profile as Stephen, which is Jo from Alcott's *Little Women* (cf. Figure 19).

Figure 19. Ego networks with a depth of 1 of Stephen and Jo showing the roles, features and clothes they are characterized with.

Jo is not named as "lesbian," "virile woman" or "independant woman" by Beauvoir but rather described as a tomboyish young girl, who nevertheless gets married in a traditional constellation. Yet Jo shows a strikingly similar profile as Stephen. Both characters are linguistically constructed as female, i.e. referred to with she/her pronouns. Among others

both characters are described with the central and highly frequent male role "man." Stephen is also referred to as "boy" and Jo as "brother." Another interesting similarity is that both characters are referred to by using male historical figures, being Nelson and Wilhelm Tell in the case of Stephen and Shakespeare for Jo. Both characters are also described using names of male animals such as "cock" (Stephen) and "stallion" and "colt" (Jo). Although Beauvoir does not name Jo as an example for a character of a non-binary gender category we take from this that we are dealing with a character that clearly breaks the boundaries of such a notion of gender. This is especially remarkable as the novel dates from 1868 and thus shows that gender-profiles of characters that clearly do not fit into the boundaries of a binary notion of gender can be found as early as in 19th century fiction.

Coming back to the analysis of the whole graph we can now answer the final question of where to find these characters that contest binary notions of gender in the graph. Other than in the graph representing the non-fiction corpus, characters that show potential for gender diversity are not situated between the (neutral) center of the graph and the male and female areas towards the poles. Instead they show up beside the main axis of the scale situated towards the margins of the graph (cf. Figure 20).

Figure 20. Characters and roles showing traits of gender diversity and their position inside the graph.

Although this is not nearly enough data to come to a final conclusion about the categorization of gender diversity from here we can use the position inside the graph as a heuristic hint towards characters that are of potential use for further analysis of the topic. Such characters are for example Dorian Gray from the novel of the same name and Felix de Vendenesse from the *Lily of the Valley*. In order to proceed with the categorization of gender diversity one could annotate and analyze these novels as whole texts and analyze the resulting gender profiles.

The revisited category system looks very different from the preliminary one we developed in part one of this study (cf. Figure 21 to Figure 8).

53

54

55

Figure 21 content:

female	neutral	male	diverse	situated outside the graph with no or just 1 connection

female: woman, girl, lady, young woman, widow, prostitute, female friend, love, female, daughter, mother, old woman, nun, young girl, sister, female artist, wife, maid
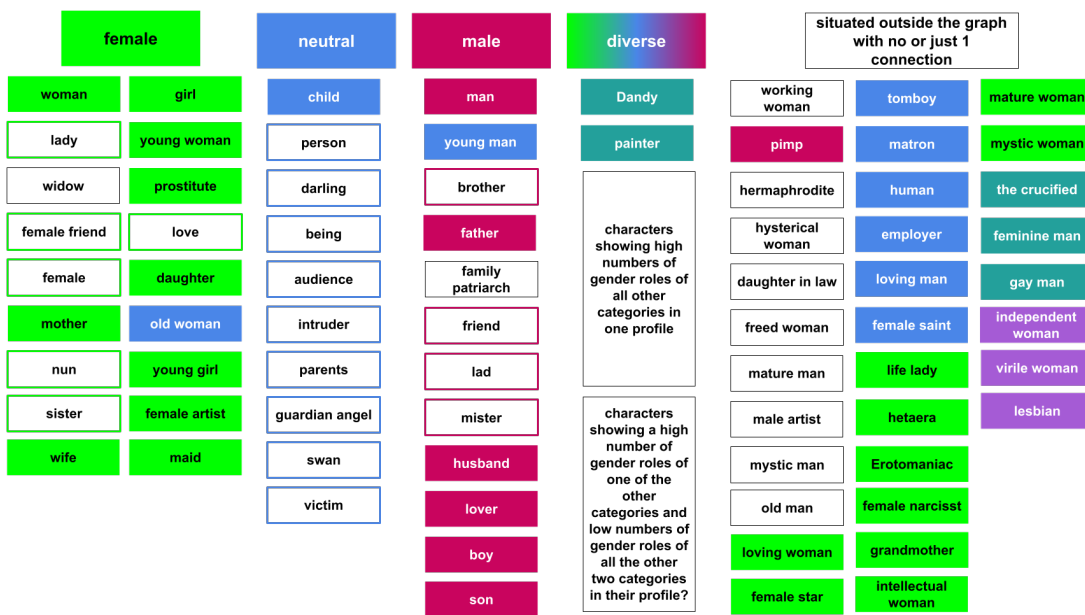
neutral: child, person, darling, being, audience, intruder, parents, guardian angel, swan, victim

male: man, young man, brother, father, family patriarch, friend, lad, mister, husband, lover, boy, son

diverse: Dandy, painter, characters showing high numbers of gender roles of all other categories in one profile, characters showing a high number of gender roles of one of the other categories and low numbers of gender roles of all the other two categories in their profile?

situated outside the graph with no or just 1 connection: working woman, tomboy, mature woman, pimp, matron, mystic woman, hermaphrodite, human, the crucified, hysterical woman, employer, feminine man, daughter in law, loving man, gay man, freed woman, female saint, independent woman, mature man, life lady, virile woman, male artist, hetaera, lesbian, mystic man, Erotomaniac, old man, female narcisst, loving woman, grandmother, female star, intellectual woman

Figure 21. Gender categories revisited.

We kept the categories "female," "neutral" and "male" in which we could sort many of the roles named and categorized by Beauvoir. However we did not find equivalents to the roles she depicted for "rather female" and "rather male." Actually only two of these roles were named in the fictional corpus being "dandy" and "painter." We realized that characters that do not fit into a binary gender system are not explicitly named with roles of "rather male" or "rather female" gender. Instead they show high numbers of roles of all three gender categories "female," "male" and "neutral" in their profile. We thus came up with another category which we named "diverse." As the former category "potentially neutral" it is very different from the other categories as it shows only a low number of specific roles. Again it is only a first attempt to gender diversity that can be used to further proceed along this path. It suggests that diversity should be analyzed at character level using an individuation-based approach rather than on the level of roles. As can be seen from Figure 21 neither the non-fiction nor the fiction corpus provided us with enough data to categorize all gender roles. By using the same colors as in Figure 8 we visualized categories we kept from the first attempt, those we moved to another category and those we added (colored in white but framed in the color of the top level category). The visualization also shows that once again many roles could not be classified as they were not named in the fiction corpus. Finally the visualization neither includes all the low frequency roles from the fiction corpus as it would simply be too much to include more than 1,000 roles in one visualization.

6. Summarizing Notes and Conclusion

By annotating theoretical as well as literary texts in terms of features, clothes and roles used for characterization we were able to reconsider Beauvoir's approach to gender roles and to implement it into a category system. We mapped the gender sphere she creates in her theory and were able to show that gender roles incorporating diversity are well integrated into the center of this sphere. In addition we could single out that neutral roles are underrepresented and almost entirely overlooked in her approach. However her notion of gender varies considerably from a scalar notion of gender ranging from male to female and having a neutral center. Such a notion can be modelled using the data of the fiction corpus. In this notion of gender diversity is set at the margins, i.e. appears as falling outside of the scale.

By analyzing distinctive/individual and overarching/collective features we realized that Beauvoir highlights features that according to her are gender-specific but that are not very common in the fiction corpus. We also found that highly frequent features are mostly used in a gender-unspecific way in the fiction corpus whereas features with a lower degree are more likely to be part of gender-specific patterns. However we conclude that gender roles are of higher importance for profiling gender in the fiction corpus. By analyzing individual gender profiles we found out that characters not fitting into a binary gender system are characterized using an almost balanced amount of female, neutral and male roles. But

56

57

58

we also found that it is very common for literary characters to show a majority of roles of one gender plus an indefinite number of neutral roles and a single role of the opposing gender (when following the notion of a gender scale). However our data was not sufficient to show the exact breaking point from characters that fit into gender norms to those that break them – although we have been able to single out characters and works that could be of interest for further research on this subject.

In the end we came up with two category systems that can be used to further investigate gender characteristics in literary texts. One closely follows Beauvoir whereas the other is based on gender roles, features and clothes frequently used in a corpus of fiction. By transferring Beauvoir's non-digital approach to a controlled computational research setup we could show the differences in modelling gender in one way or the other. After having tested Beauvoir's gender model on the primary sources she used herself and adjusted the resulting category system, we can now transfer this categorization of gender to other research areas such as contemporary literature.

Appendix of Primary Sources (Fiction Corpus) with English Short Titles

Alcott, Louisa May. (1868/69) *Kleine Frauen* (*Little Women*)

Balzac, Honoré de. (1835) *Die Lilie im Tal* (*Lily of the Valley*)

Boccaccio, Giovanni. (1353) *Das Decameron* (1–6) (*Decameron*)

Brontë, Emily (1847) *Sturmhöhe* (*Wuthering Heights*)

Chateaubriand, François-René de. (1801) *Atala*

Colette, Sidonie-Gabrielle Claudine. (1929) *Sido*

Dostojewski, Fjodor Michailowitsch. (1880) *Die Brüder Karamasow* (*The Karamazow Brothers*)

Eliot, George. (1871) *Middlemarch*

Keun, Irmgard. (1932) *Das kunstseidene Mädchen* (*Artificial Silk Girl*)

Hall, Radclyffe. (1928) *Quell der Einsamkeit* (*Well of Loneliness*)

Lawrence, David Herbert. (1913) *Söhne und Liebhaber* (*Sons and Lovers*)

Sade, Marquis de. (1797) *Juliette*

Steinbeck, John. (1945) *Die Straße der Ölsardinen* (*Cannery Row*)

Stendhal, Henry-Marie Beyle. (1830) *Rot und Schwarz* (*Red and Black*)

Tolstoi, Lew Nikolajewitsch. (1867) *Krieg und Frieden* (*War and Peace*)

Wilde, Oscar. (1890) *Das Bildnis des Dorian Gray* (*Dorian Gray*)

Woolf, Virginia. (1931) *Die Wellen* (*Waves*)

Zola, Émil. (1880) *Nana*

Zola, Émil. (1882) *Ein feines Haus* (*Pot Luck*)

Notes

[1] Beauvoir's text consists of two main parts or books. The first part, *Facts and Myths*, is more about the history of gender roles while the second part, *Lived Experiences*, includes an in-depth look at different gender roles. Our analysis is based on the second book.

[2] We use the term "gender roles" for the various gendered denominations given to characters such as "wife," "mother," "working woman" etc. These roles can be grouped into "gender categories" like female, male, neutral and diverse. Characters mostly can be assigned to one dominant gender category but are described using numerous gender roles.

[3] We use the term "feature" in a non-technical way to describe the different attributes which form different gender roles.

[4] This holds true for the theoretic text by Beauvoir as well as for the narrative texts. However we translated most of the features into English in this article for the sake of a better readability.

[5] A list of all literary texts we used for this case study can be seen in the bibliography. A full list of all literary sources named in the second part of Beauvoir's study can be accessed here: https://msternchenw.de/teilkorpus-simone-de-beauvoir/.

[6] In the early stage of annotation we did not distinguish between features and clothes, but then we realized that clothes are indeed a category of its own underlying much more trends and thus being less stable than bodily and mental features that are more deeply rooted in a character.

[7] The importance of literature in Freund's psychoanalysis is evident, among other things, in the numerous recursions to literary topoi, motifs, authors, and texts ([Alt 2008]). Examples can be found in [Lüdemann 1994], [Thomé 1998], or [Rohrwasser 2005].

[8] There seem to be some more peculiarities as e.g. the "hermaphrodite" being situated behind "woman" too, but a closer look shows that this position relies on one single shared feature. We ignore such cases assuming that this is not enough data to base an interpretation on.

[9] By linguistically female we understand roles that are referenced with the pronoun "she."

[10] By linguistically male we understand roles that are referenced with the pronoun "he."

[11] Datasets generated during this study can be found here: https://zenodo.org/record/7388960.

[12] As we worked with German translations some words are classified as linguistically male which are neutral in their English translation. "Employer" which translates to "Arbeiter" is one of these words.

[13] We use the terms "collective" and "individual" here instead of the terms "overarching" and "distinctive," which we used in the first part of this article. We thus highlight the fact that the focus of our study now lies on characters instead of roles.

[14] https://openrefine.org/ (Access: December 1st 2022).

[15] Grammatical level: nominalizations were disambiguated and returned in the adjective form. In the case of intensified forms, the positive was given. Typing errors or spelling variations (hyphenation and contraction) were corrected or standardized. Inflected verb forms were brought into the basic form. Different (historical) spellings were adapted to the current spelling. Age references have been standardized to numeric references. Negatives are, if possible, uniformly introduced with un- ("without" or "not" is deleted).Semantic level: Synonyms were resolved with regard to the preferred adjective form and, if possible, merged to existing properties ("no money" becomes "poor"/ "smart" and "clever" becomes "smart"). Different variants of a feature were merged into one feature. Specifications of a feature were deleted ("exceedingly beautiful" becomes "beautiful").

[16] We did this part of the study using the same starting point we used for the first part: a basic category system for gender roles including the categories "female," "neutral" and "male." Again a more sophisticated category system is our research goal.

[17] The frequency of multiple occurrence in descending order: 1,027 features occur twice, 180 features occur three times, 113 features occur four times, 51 types occur five times each, 37 different types occur six times each, 24 types occur seven times, 25 types occur eight times, twelve types occur nine times, and ten different types occur ten times.

[18] Nevertheless there are of course some attempts to the topic such as in Foucault's *The true sex* and Bourdieu's *Postscriptum – on domination and love*.

[19] Such characters can be found in six novels altogether ranging from 1835 (Balzac's *Lilly of the Valley*) to 1932 (Keun *Artificial Silk Girl*) and include e.g. Marat from *Lilly of the Valley*, who is only referred to with the role "angel," Susie Perkins from *Little Women*, who is only compared to a "castle dog," the jew from *Dorian Gray*, who is designated as "monster" and "creature," Paul from *Sons and Lovers*, whose profile shows an equal amount of male and neutral roles (neutral roles being "child," "lamb," "unborn" and "little one," male roles "man," "boy," "little boy" and "most beautiful," the last being linguistically male in German), Jinny ("child," "horse," "giant"), Rhoda (same as Jinny but with "empress" as female role on top), Susan ("giant," "horse," "child," "swimmer," "fox," "jay" and "marten" as neutral, "explorer" as male and "female swimmer" as female role) and Neville ("giant," "child," "horse" as neutral, "master" as male role) from Woolf's *Waves* or Pilli ("bar" and "rack") or the red moon ("idiot") from *The Artificial Silk Girl*.

[20] Characters showing this profile are Juliette from the novel of the same name by de Sade published in 1797 and Sido from the narrative with the same name by Colette published in 1931.

Works Cited

Alt 2008 Alt, Peter-André. (2008) "Einführung." In: Alt, Peter-André, and Anz, Thomas (Eds.): *Sigmund Freud und das Wissen der Literatur*. Berlin, New York: De Gruyter, pp. 1–17.

Bastian et al. 2009 Bastian, Mathieu, Simon Heymann, and Mathieu Jacomy. (2009) "Gephi: An Open Source Software for Exploring and Manipulating Networks."

de Beauvoir 1992 de Beauvoir, Simone. (1992) *Das andere Geschlecht. Sitte und Sexus der Frau*. new translation. Rowohlt-Taschenbuch-Verl. Hamburg: Rowohlt.

Botond 2017 Botond, Anneliese. (2017) "Simone de Beauvoir." In: Christian Bermes (Ed.): *Kindler Kompakt: Philosophie 20. Jahrhundert*. Stuttgart: Metzler, 217–219. DOI 10.1007/978-3-476-05539-2_22.

Bourdieu 2010 Bourdieu, Pierre. (2010) *Die männliche Herrschaft*. 1. Ed. Reprint. Frankfurt am Main: Suhrkamp.

Butler 2003 Butler, Judith. (2003) *Das Unbehagen der Geschlechter*. 1. Ed. Reprint. Frankfurt am Main: Suhrkamp.

Connell 1996 Connell, Raewyn. (1996) *Gender and power society, the person and sexual politics*. Reprint. Cambridge: Polity Press.

Connell 2015 Connell, Raewyn. (2015) *Der gemachte Mann. Konstruktion und Krise von Männlichkeiten*. Stuttgart: Springer VS (Geschlecht & Gesellschaft Band 8).

Data Commons 2022 Data Commons. (2022) Data Commons, viewed 29 Nov 2022, https://datacommons.org.

Flüh et al. 2022a Flüh, Marie, Horstmann, Jan, and Schumacher, Mareike. (2022) "Genderaspekte in Fantasy-Jugendromanen von 2008 bis 2020: Distant Gender Reading," in *Gender in der deutschsprachigen Kinder- und Jugendliteratur*. De Gruyter, pp. 457–482. Available at: https://doi.org/10.1515/9783110726404-025.

Flüh et al. 2022b Flüh, Marie, and Schumacher, Mareike. (2022) "Jung, wild, emotional? Rollen und Emotionen Jugendlicher in zeitgenössischer Fantasy-Literatur," in *DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2022)*, Potsdam: Zenodo. Available at: https://doi.org/10.5281/zenodo.6327983.

Flüh et al. forthcoming Flüh, Marie and Schumacher, Mareike. (Forthcoming) "Macht vs Emotion. Handlungstreibende Muster in Günderrodes Dramen digital, distant und scalable gelesen." In: Bogard, Roland, Middelhoff, Frederike, Wernli, Martina (Eds.) *Noch Zukunft haben*. Berlin, New York: De Gruyter.

Jacke 2018 Jacke, Janina. (2018) "Manuelle Annotation," in *forTEXT*. Available at: https://fortext.net/routinen/methoden/manuelle-annotation.

Kirkpatrick 2020 Kirkpatrick, Kate. (2020) *Simone de Beauvoir. Ein modernes Leben*. München: Piper.

Krause 2021 Krause, Todd B. (2021). "IE6.com. Cleaning data with OpenRefine" In: *Folia Linguistica*, vol. 55, no. 42, pp. 527–533. https://doi.org/10.1515/flin-2021-2038.

Lüdemann 1994 Lüdemann, Susanne. (1994) *Mythos und Selbstdarstellung. Zur Poetik der Psychoanalyse*. Freiburg: rombach.

Pagel et al. 2020 Pagel, Janis, Reiter, Nils, Rösiger, and Schulz, Sarah.. (2020) "Annotation als flexibel einsetzbare Methode." In: Pagel, Janis, Reiter, Nils, Rösiger, and Schulz, Sarah. (Ed.): *Reflektierte algorithmische Textanalyse: Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*, edited by Nils Reiter, Axel Pichler and Jonas Kuhn, Berlin, Boston: De Gruyter, pp. 125–142. https://doi.org/10.1515/9783110693973-006.

Rohrwasser 2005 Rohrwasser, Michael. (2005) *Freuds Lektüren. Von Arthur Conan Doyle bis zu Arthur Schnitzler*. Gießen: Imago.

Schumacher 2018 Schumacher, M. (2018) "Netzwerkanalyse," *forTEXT - Literatur digital erforschen.* [Preprint]. Edited by E. Gius et al. Available at: https://fortext.net/routinen/methoden/netzwerkanalyse (Accessed: 3 May 2022).

Schumacher 2020 Schumacher, Mareike. (2020) "Netzwerkanalyse mit Gephi," in *forTEXT*. Available at: https://fortext.net/routinen/lerneinheiten/netzwerkanalyse-mit-gephi.

Schwarz-Friesel 2007 Schwarz-Friesel, Monika. (2007) Sprache und Emotion. Narr Francke: Tübingen.

Schöch 2017 Schöch, Christof. (2017) "Aufbau von Datensammlungen." In: Jannidis, Fotis, Kohle, Hubertus, and Rehbein, Malte (Eds.): *Digital Humanities: Eine Einführung*, Stuttgart: J.B. Metzler, pp. 223–33. DOI: https://doi.org/10.1007/978-

3-476-05446-3_16.

Thomé 1998 Thomé, Horst. (1998) Freud als Erzähler. Zu literarischen Elementen im "Bruchstück einer Hysterie-Analyse." In: Danneberg, Lutz, and Niederhause, Jürgen (Eds.): *Darstellungsformen der Wissenschaften im Kontrast. Aspekte der Methodik, Theorie und Empirie.* Tübingen: Narr, pp. 471–493.

Zadeh 1965 Zadeh, L.A. (1965) "Fuzzy sets," *Information and Control*, 8(3), pp. 338–353. Available at: https://doi.org/10.1016/S0019-9958(65)90241-X.