Man vs. Machine

Konovalova, Aleksandra; Toral, Antonio

Published in:
Proceedings of the 6th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature

Man vs. Machine: Extracting Character Networks from Human and Machine Translations

Aleksandra Konovalova
University of Turku
aleksandra.a.konovalova@utu.fi

Antonio Toral
University of Groningen
a.toral.ruiz@rug.nl

Abstract

Most of the work on Character Networks to date is limited to monolingual texts. Conversely, in this paper we apply and analyze Character Networks on both source texts (English novels) and their Finnish translations (both human- and machine-translated). We assume that this analysis could provide some insights on changes in translations that could modify the character networks, as well as the narrative. The results show that the character networks of translations differ from originals in case of long novels, and the differences may also vary depending on the novel and translator's strategy.

1 Introduction

Character Networks (CNs) building can be considered as a part of Social Networks Analysis (SNA) research. The main difference between SNA and CNs extraction has to do with the type of datasets to which these methods are applied: CNs extraction is typically used for different works of art (mainly literary texts of different genres as well as films), while SNA is usually performed on more structured datasets, e.g. online social networks, such as YouTube or LiveJournal (Mislove et al., 2007).

Most of the work on CNs to date applies them to monolingual texts. The main novelty of our work stems from the fact that we apply these techniques not only to original texts but also to their translations (both by human translators and by Machine Translation (MT) systems). In doing so, we aim to unveil whether the connections between characters (represented by CNs), modified by human or machine translator, differ, which can point to narrative differences between original texts and translations.

There are different tasks that could benefit from the extraction of CNs in this bilingual setting, namely MT. MT could be enhanced with the contextual information, namely the global context of the whole text that could be in a form of a graph (Xu et al., 2020). We consider the task of CNs extraction in the scope of enhancing MT of literary texts. In this framework, we consider the extraction of CNs as a valuable first step, so that we can find out how the CNs of Human Translation (HT) and MT compare to the CN of the original.

In this paper we take a look at the CNs of English originals and Finnish translations thereof. The structure of the paper is as follows: first we provide an overview of the related work (Section 2), subsequently we describe our data (Section 3), after which we discuss the creation of the list of the characters' names that we will use for our methods (Section 4). We continue the paper with describing the method (Section 5) that we used. Finally, we present our results of both qualitative analyses and quantitative assessment and analyze them (Section 6). We conclude our paper in Section 7.

2 Related work

Character Networks extraction is a broad problem, so there have been many attempts to tackle it from different angles. It can be the main focus of the research (John et al., 2019; Kubis, 2021), or only a step towards a broader goal, e.g. learning representations of stories based on character networks (Lee and Jung, 2019, 2020). It can be automated (Chen et al., 2019) or not (Moretti, 2011). The data for character networks extraction can also vary from movies (Agarwal et al., 2014) to novels (Agarwal et al., 2012) and fairytales (Schmidt et al., 2021). The most thorough overview of character network extraction so far has been done by Labatut and Bost (Labatut and Bost, 2019). Also Schmidt et al. (2021), aside from their original topic of research, raised an issue regarding the evaluation of character networks: according to them, currently there is no standardized approach for such evaluation and most research on this topic evaluates the extracted networks by proxy or using SNA metrics (Schmidt et al., 2021).

Proceedings of the 6th LaTeCH-CLfL Workshop, pages 75–82

The novelty of our research with respect to previous work lies mainly in three points: firstly, we are taking a look at CNs of translations; secondly, we are looking at CNs of machine-translated texts; and thirdly, we are looking at an uncommon language pair (English-Finnish), since, as far as we know, there is no related work for character networks based on Finnish texts to date.

3 Data

The main dataset is made up of corpora of English and Finnish literary texts. The English part of the dataset was gathered from Project Gutenberg (https://www.gutenberg.org/), while the Finnish human-translated subcorpus is available at the Language Bank of Finland as The Downloadable Version of Classics of English and American Literature in Finnish (https://www.kielipankki.fi/corpora/ceal-2/).

The English subcorpus contains two novels (*Pride and Prejudice* by Jane Austen, *Bleak House* by Charles Dickens) and a short story (*The Washington Square* by Henry James). The Finnish human-translated corpus contains the corresponding Finnish translations of these works carried out by Kersti Juva, while the Finnish machine-translated subcorpus was created by DeepL Translator (https://www.deepl.com/translator) from the English originals on May 26, 2022. Table 1 contains some statistics about the corpora.

English and Finnish human-translated subcorpora and English and Finnish machine-translated subcorpora were sentence-aligned for consistency and for the purpose of doing close reading. The alignment was done using InterText (Vondricka, 2014). In case of Human Translations, the alignment was done semi-automatically (we had to go through the whole texts and align problematic sentences manually), but for MT it was done automatically, because the sentence splitting of the output translations of DeepL corresponded to the one of the original texts.

4 Creation of character names' list

Before applying our methods (see Section 5), we had to create a character names' list, so that we could use this list for implementation of our methods. To perform this task, we got the information from different internet sources that contain information about characters from the novels in our dataset (see Appendix A).

While creating a list of characters' names as a basis for our CNs, we also faced many questions about characters, such as: what is a literary character? Who do we consider a character from the point of the narrative? Do we take into consideration off-screen characters (characters that are only mentioned in the text and do not participate in the plot)? To answer these questions, we needed to define what / who the character is.

The literary character can be seen as a construct which definition and features depend on the study area (Margolin, 1990). Jannidis (2013) considered a character "a text- or media-based figure in a storyworld, usually human or human-like" or "an entity in a storyworld". Overall, characters are interconnected with both narrative and storyworld and contribute to their development from many aspects.

Based on this notion, we considered a literary character every figure that was relevant for the narrative development (thus, e.g. names of famous persons that are mentioned but do not appear in the novel directly were not included). So we decided to include both onscreen (entities that are actively participating in the storyworld) and offscreen (entities that are passively contributing to the construction of the storyworld) characters (e.g. in case of *Washington Square*, it was the mother of the main character that was mentioned only twice, but never participated in the story herself).

We also included all possible names that can be used for naming a certain character by splitting the full name (e.g. *Elizabeth Bennet* would also get versions *Elizabeth* and *Bennet*) and by analyzing possible versions (*Lizzy* for *Elizabeth Bennet*) that were mentioned in the internet sources (see Appendix A). We also included full names (if applicable) even if they were not used for naming a character in the text just for reference (e.g. in case of *Catherine Sloper*). So *Elizabeth Bennet* would get the following names: *Bennet*, *Eliza*, *Eliza Bennet*, *Elizabeth*, *Elizabeth Bennet*, *Lizzy*, and *Catherine Sloper* would get the names *Catherine*, *Catherine Sloper* and *Sloper*. The creation of the characters' names list was carried out only by one annotator.

As a result, we would have a list of all possible characters' names. For this research we decided not to link different names of the same characters, because there were relatives and namesakes which were impossible to distinguish from the context.

Corpus	Characters (without newline characters)	Sentences	Words
English subcorpus	2,956,068	28,360	549,383
Finnish subcorpus (Human Translations)	2,861,687	23,153	387,734
Finnish subcorpus (Machine Translations)	3,069,732	23,518	410,767

Table 1: Statistics for the corpora used in our study

5 Methods

We extracted the character networks from English, Finnish human-translated and Finnish machine-translated texts using the same workflow. The workflow was implemented using Python with the help of the NetworkX library (https://networkx.org/) for CN-related quantitative metrics and visualization.

The workflow proceeds sequentially as follows:

1. Splitting texts by chapters (we searched in the text for the expressions that contained the chapters' names and split the text by them) and transforming each novel into a list of chapters;

2. Searching for the names from characters' names list in every chapter and producing a list of character relationships for each chapter; Iterating through a list of chapters and producing the final results for character relationships in the novel.

We decided to use chapters as the units to build the CNs for several reasons. Firstly, using smaller units (e.g. paragraphs) may have led to unexpected results, since the texts also contained dialogues and letters which may have zero characters in one paragraph despite having a clear link between the characters outside the dialogue or the letter. Secondly, we consider a novel chapter as an autonomous part of narrative which may provide a more finalized view into characters' relationships.

We consider our approach to be semi-automated, since we had to build lists of characters' names manually. We also decided to introduce some limitations to our research.

For this paper, we decided not to link different versions of the names and their references, such as pronouns, due to the complexity of such a task, especially regarding Finnish translations. For example, in *Pride and Prejudice* we would have 5 characters that could be linked to "Miss Bennet", namely all five Bennet sisters: Jane, Elizabeth, Mary, Catherine and Lydia. Moreover, it is used in plural - there are "younger Miss Bennets" and "older Miss Bennets". As per our knowledge, there is also no coreference model for Finnish language, and training such a model would require from us creation of the annotated corpus, which could be a topic for a paper on its own. For a similar reason we also decided not to use Named Entity Recognition (NER) state-of-the-art tools, because our previous research has shown the need to further refine the results of NER pipeline: namely the results of lemmatization step for foreign names in Finnish texts needed to be polished further (Konovalova et al., 2022).

6 Evaluation and results

After implementing our workflow, we used it on our corpus, producing CNs for English original texts, Finnish Human Translations and Finnish MT outputs. For our assessment, we performed both qualitative and ative analyses. Quantitative analysis was done by analysing the CNs' metrics and qualitative analysis was done to provide some insights and possible reasons for such results of quantitative analysis.

6.1 Qualitative Analysis (close reading)

We performed close reading for English originals and Finnish Human Translations while doing sentence alignment. We also performed close reading for Finnish MT outputs.

We grouped the changes in translations in two groups: changing the pronoun into the proper noun and changing names completely.

6.1.1 Changing the pronoun into the proper noun

Human Translations

Close reading showed that in some cases in

Finnish Human Translations names of the characters were used instead of the pronouns in English originals: for example, in different interactions between *Elizabeth Bennet* and *Mr. Darcy* in *Pride and Prejudice* translation (see Table 2).

We assume that there could be two possible reasons for this: firstly, the translator's own style, and secondly, the nature of the target language (in Finnish there is one third-person pronoun, *hän*, that corresponds both to *he* and *she*, so the use of the character's name could be the attempt to avoid ambiguity. The other way to avoid ambiguity is to use notions like *mies* (*man*) and *nainen* (*woman*)). Since the last reason is tied to the target language, it could also be linked to the normalization strategy used by the translator (namely, adapting the translations to target language norms (Baker and Somers, 1996)).

We assume that such translator's decisions affected the quantitative results for the Character Networks that we present in the next subsection (6.2) on Quantitative Assessment.

Machine Translations

Similarly to Human Translations, there were cases when the pronoun would be replaced with the name. The possible reason for it could also be connected to the normalization principle that was learnt and used by the MT model during MT.

In Table 2 we present several examples where, surprisingly, both Human Translation and MT use the same normalization technique.

There was also an interesting example where it seems that coreference went wrong in the MT output, as "she" in original text is another character, *Mrs. Phillips* (referenced by *"täti"* (*aunt* in Finnish) in Human Translation):

Original: She received him with her very best politeness, which he returned with as much more, apologising for his intrusion, without any previous acquaintance with **her**, which he could not help flattering himself, however, might be justified by his relationship to the young ladies who introduced him to **her** notice.

MT: Jane otti miehen vastaan parhaalla mahdollisella kohteliaisuudellaan, ja mies vastasi kohteliaasti ja pyysi anteeksi tunkeutumistaan, koska hän ei tuntenut **Janea** aikaisemmin, mutta hän ei voinut olla imartelematta itseään, että hänen suhteensa niihin nuoriin neitoihin, jotka esittelivät miehen **Janeen**, saattaisi kuitenkin oikeuttaa tämän tunkeutumisen.

Human Translation: Täti otti herran vastaan kaikin tavoin kohteliaasti, mihin mies vastasi samalla mitalla ja pannen paremmaksi, pyysi anteeksi, että tunkeutui näin kylään ilman aikaisempaa tuttavuutta, mutta tahtoi kuitenkin uskoa sukulaisuussuhteen hänet esitelleisiin nuoriin naisiin antavan sille oikeutuksen.

6.1.2 Changing names completely

Human Translations

There was one case when the name that has a distinctive meaning in the original text has to be changed in the Human Translation to save and convey this meaning to the reader. On the contrary, it was not changed in MT. Compare:

Original: <...>, Mr. Snagsby mentions to the 'prentices, "I think my little woman is a-giving it to **Guster**!" This proper name, so used by Mr. Snagsby, has before now sharpened the wit of the Cook's Courtiers to remark that it ought to be the name of Mrs. Snagsby, seeing that she might with great force and expression be termed a **Guster**, in compliment to her **stormy** character.

HT: <...>, herra Snagsby sanoo oppipojilleen: "Siellä taitaa pikkurouva kurittaa **Mollya**!" Herra Snagsbyn mainitsema etunimi on aikaa sitten saanut naapurit letkauttamaan, että se olisi sopiva nimi rouva Snagsbylle, sillä nimi **Möly** istuisi hänelle kuin nyrkki silmään hänen **äänekkään** luonteensa ansiosta.

MT: <...>, herra Snagsby sanoo apulaisille: "Luulen, että pikku naiseni antaa sen **Gusterille**!". Tämä herra Snagsbyn käyttämä nimi on ennenkin saanut Cookin hovimiehet huomauttamaan, että sen pitäisi olla rouva Snagsbyn nimi, koska häntä voisi nimittää hyvin voimakkaasti ja ilmeikkäästi **Gusteriksi**, kohteliaisuutena hänen **myrskyisälle** luonteelleen.

Guster has a *stormy* personality, and her name sounds like *gust* - sudden rush of the wind. In Finnish Human Translation the translator faced two problems: first - how to convey the name's meaning to the Finnish reader and second - how to still have the English name in the translation. It was solved by linking the existing name - Molly - to Finnish word *möly* (*noise*) and saying that Molly/Möly has a *noisy* character.

Machine Translations

We noticed that in machine-translated texts there were also changes in the names: some of them were sometimes domesticated, for example, *Catherine* would be changed to *Katariina* (Finnish version

Original	Machine Translation	Human Translation
Elizabeth listened with delight to the happy, though modest hopes which **Jane** entertained of Mr. Bingley's regard, and said all in her power to heighten **her** confidence in it.	Elizabeth kuunteli ihastuneena **Janen** iloisia, vaikkakin vaatimattomia toiveita herra Bingleyn kunnioituksesta ja sanoi kaiken voitavansa vahvistaakseen **Janen** luottamusta siihen.	Elizabethille oli ilo kuulla onnellisista joskin kainoista haaveista, joita **Janella** oli herra Bingleyn tunteisiin nähden, ja hän sanoi kaiken, mitä pystyi sanomaan, vahvistaakseen **Janen** luottamusta niihin.
She could not help frequently glancing her eye at **Mr. Darcy**, though every glance convinced her of what she dreaded; for though **he** was not always looking at her mother, she was convinced that **his** attention was invariably fixed by her.	Hän ei voinut olla vilkaisematta usein **herra Darcyyn**, vaikka jokainen vilkaisu sai hänet vakuuttuneeksi siitä, mitä hän pelkäsi; sillä vaikka **Darcy** ei aina katsonutkaan äitiä, hän oli vakuuttunut siitä, että **Darcy** kiinnitti hänen huomionsa aina äitiinsä.	Hän ei voinut olla vähän väliä vilkaisematta syrjäsilmällä **herra Darcyyn** siitä huolimatta, että jokainen vilkaisu vahvisti hänen pelkonsa; sillä vaikka **Darcy** ei katsonut äitiin koko ajan, hän oli varma, että äiti oli **Darcyn** hellittämättömän huomion kohteena.
So I thought it a good opportunity to hint to Richard that if he were sometimes a little careless of himself, I was very sure he never meant to be careless of **Ada**, and that it was a part of his affectionate consideration for **her** not to slight the importance of a step that might influence both their lives.	Ajattelin, että nyt oli hyvä tilaisuus vihjata Richardille, että vaikka hän oli joskus hieman huolimaton itseään kohtaan, olin aivan varma, ettei hän koskaan aikonut olla huolimaton **Adaa** kohtaan ja että oli osa hänen hellästä huomaavaisuudestaan **Adaa** kohtaan, ettei hän vähättelisi sellaisen askeleen merkitystä, joka saattoi vaikuttaa heidän molempien elämään.	Niin minä katsoin tilaisuuden sopivaksi vihjata Richardille, että jos hän joskus olikin hiukan huoleton oman itsensä suhteen, hän ei toki koskaan voisi olla huoleton **Adan** suhteen, ja että hänen kiintymykseensä **Adaan** kuului osana se, ettei hän vähätellyt minkään askelen merkitystä, millä saattaisi olla vaikutusta heidän yhteiseen elämäänsä.

Table 2: Examples of changing pronouns into proper nouns in Human and Machine Translations.

of the name) or *Elizabeth* would be changed to *Elisabet* (also Finnish version of the name). So in the MT output there could be two versions for the same name: *Catherine* would correspond to both *Catherine* and *Katariina*, and *Elizabeth* - to *Elizabeth* and *Elisabet*.

Original: It pleased **Catherine** to think that she should be brave for his sake, and in her satisfaction she even gave a little smile.

MT: **Katariinaa** ilahdutti ajatus siitä, että hänen piti olla rohkea hänen vuokseen, ja tyytyväisyydessään hän jopa hymyili hieman.

Compare to:

Original: "It will be easy to be prepared for that," **Catherine** said.

MT: "Siihen on helppo valmistautua", **Catherine** sanoi.

Overall the results of close reading show that both Human and Machine Translation tend to use normalization techniques. In the case of MT, the domestication of the names was used sporadically, which created several versions of one name in translation.

6.2 Quantitative Assessment (metrics)

We used the following metrics for assessing and comparing our results:

1. Different centrality metrics which are the main ones for the analysis of character networks (Newman, 2010):
   (a) Betweenness centrality (how much each character connects other characters between themselves); we took a look at the first 5 results with the highest values for this metric;
   (b) Degree centrality (how many connections one node (character) has to others); we also took a look at the first 5 results with the highest values for this metric;
2. Density (what is the level of connections of the whole graph)
3. Diameter (how big the network is).

Text type	Betweenness centrality (max, n=5)	Degree centrality (max, n=5)	Density	Diameter
Original / Human Translation / Machine Translation	Almond: 0.037, Catherine: 0.037, Penniman: 0.037, Sloper: 0.037, Morris Townsend: 0.22	Almond: 1.0, Catherine: 1.0, Penniman: 1.0, Sloper: 1.0, Morris Townsend: 0.94	0.77	2

Table 3: Results of different metrics for *Washington Square*.

We present our results in the tables below (one table per novel) and we also provide visualization for the characters that have the highest values for the metrics. The nodes of the graphs represent characters, and the edges represent relationships between characters.

Washington Square (Table 3)

Probably because of the size of the *Washington Square* (35 chapters, length of the original text: 354,440 characters) and because we split the texts by chapters, the CNs were the same for all three versions (original, HT and MT). The five characters with maximum betweenness centrality and degree centrality correspond to the main characters: *Mrs. Almond*, *Catherine Sloper*, *Mrs. Penniman*, *Dr. Austin Sloper* and *Morris Townsend*.

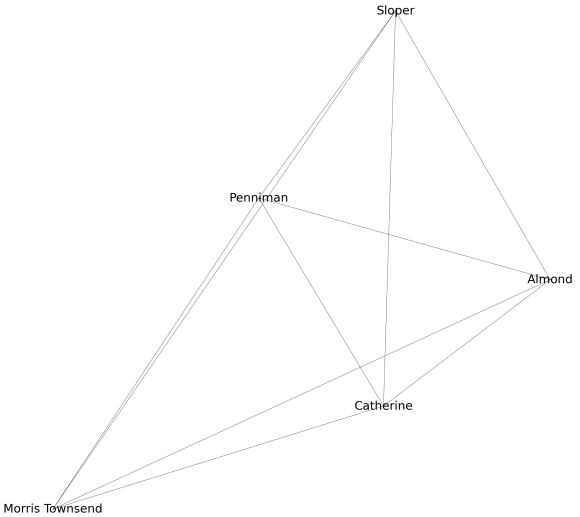

Figure 1: Example for the CN for *Washington Square* for 5 characters with maximum scores.

Pride and Prejudice (Table 4)

The main characters, according to the metrics, are *Bennet* (could be anyone from the family, but most probably it is either *Mr. Bennet* or *Mrs. Bennet*), *Bingley* (most probably *Mr. Bingley* than his sister, *Miss Bingley*), *Elizabeth Bennet* and *Mr. Darcy*. The fifth character varies: in original and machine-translated text it is *Mr. Wickham*, in human-translated text it is *Jane Bennet*. The difference in the mentions could also be attributed to the aforementioned translation strategy to use a character's name instead of the pronoun for better clarity. It is also interesting that in Human Translation *Jane Bennet* becomes a more important character than *Mr. Wickham* which could be attributed to the translator's strategy of using more proper nouns in the *Jane Bennet-Mr.Bingley* or *Jane Bennet-Elizabeth Bennet* interactions.

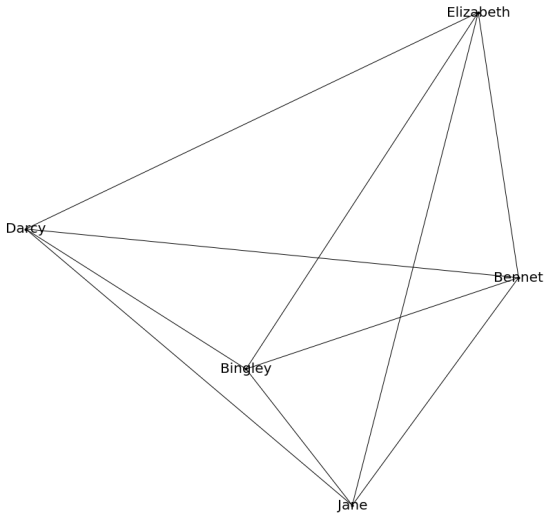

Figure 2: Example for the CN for *Pride and Prejudice* for 5 characters with maximum scores.

Bleak House (Table 5)

There are two narratives in *Bleak House*: one is done from third-person perspective, and the other is narrated by *Esther Summerson*, which may affect her appearance in the text. According to the metrics, the main characters are *Dedlock* (probably *Lady Dedlock*), *Esther Summerson*, *John Jarndyce*, *Richard Carstone* and *Mr. Tulkinghorn*. It is also

Text type	Betweenness centrality (max, n=5)	Degree centrality (max, n=5)	Density	Diameter
Original	Bennet: 0.013, Bingley: 0.013, Elizabeth: 0.013, Darcy: 0.013, Wickham: 0.012	Bennet: 1.0 Bingley: 1.0, Elizabeth: 1.0, Darcy: 1.0, Jane: 0.98	0.79	2
Human Translation	Bennet: **0.015**, Bingley: **0.015**, Elizabeth: **0.015**, Darcy: **0.015**, Jane: **0.013**	Bennet: 1.0 Bingley: 1.0, Elizabeth: 1.0, Darcy: 1.0, Jane: 0.98	**0.76**	2
Machine Translation	Bennet: 0.013, Bingley: 0.013, Elizabeth: 0.013, Darcy: 0.013, Wickham: 0.012	Bennet: 1.0 Bingley: 1.0, Elizabeth: 1.0, Darcy: 1.0, Jane: 0.98	0.79	2

Table 4: Results of different metrics for *Pride and Prejudice*. Scores in translations that differ from the original text shown in bold.

Text type	Betweenness centrality (max, n=5)	Degree centrality (max, n=5)	Density	Diameter
Original	Dedlock: 0.02, Summerson: 0.03, Jarndyce: 0.05, Tulkinghorn: 0.02, Richard: 0.02	Dedlock: 0.79, Summerson: 0.86, Jarndyce: 0.94, Richard: 0.76, Tulkinghorn: 0.77	0.4	**3**
Human Translation	Dedlock: 0.02, Summerson: 0.03, Jarndyce: 0.05, Richard: 0.02, Tulkinghorn: 0.02	Dedlock: 0.79, Summerson: 0.86, Jarndyce: 0.94, Richard: **0.78**, Tulkinghorn: **0.76**	**0.39**	2
Machine Translation	Dedlock: 0.02, Summerson: 0.03, Jarndyce: **0.04**, Lady Dedlock: 0.02, Tulkinghorn: 0.02	Dedlock: 0.79, Jarndyce: **0.93**, Richard: 0.76, Summerson: **0.85**, Tulkinghorn: **0.76**, Lady Dedlock: **0.76**	0.4	2

Table 5: Results of different metrics for *Bleak House*. Scores in translations that differ from the original text shown in bold.

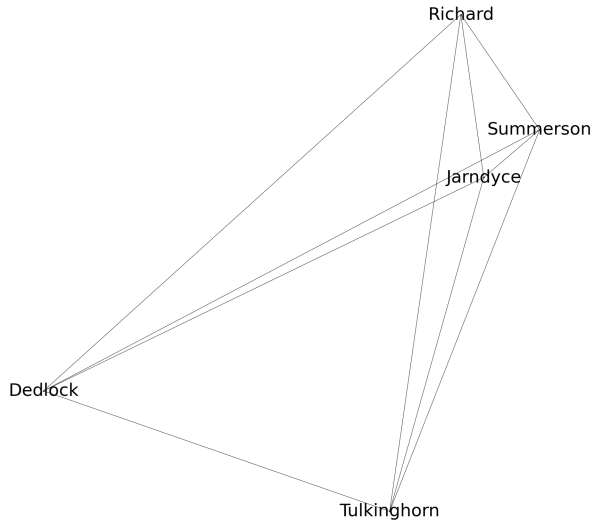

Figure 3: Example for the CN for *Bleak House* for 5 characters with maximum scores.

interesting that the diameter of the original network changes in both translations (original diameter was 3, while in translations it was reduced to 2). *Bleak House* also has the lowest density compared to other texts, which could be due to the size of the text (359,426 words, with the whole subcorpus being 548,383 words).

7 Conclusion

We have created Character Networks for original texts, for Human Translations and Machine Translations for three novels. Results show that for longer novels there are changes in Character Networks both in Human and Machine which may be attributed to the translator style or the target language features in human translations and to the models used in machine translations. One of the most interesting results is that the main 5 characters of *Pride and Prejudice* change in human translations with *Jane Bennet* replacing *Mr. Wickham*. We assume that our research could be enhanced further e.g. by using coreference which would require the creation of an annotated corpus, by grouping different versions of character names together (either manually or automatically) and by studying differ-

ent language pairs as source-target languages for originals and translations.

References

Apoorv Agarwal, Sriramkumar Balasubramanian, Jiehan Zheng, and Sarthak Dash. 2014. Parsing screenplays for extracting social networks from movies. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, pages 50–58, Gothenburg, Sweden. Association for Computational Linguistics.

Apoorv Agarwal, Augusto Corvalan, Jacob Jensen, and Owen Rambow. 2012. Social network analysis of alice in wonderland. In *CLfL@NAACL-HLT*.

Mona Baker and Harold Somers. 1996. *'Corpus-based Translation Studies: The Challenges that Lie Ahead'*. John Benjamins Publishing Company, Netherlands.

R.H.-G. Chen, C.-C. Chen, and C.-M. Chen. 2019. Unsupervised cluster analyses of character networks in fiction: Community structure and centrality. *Knowledge-Based Systems*, 163:800–810.

Fotis Jannidis. 2013. Character. In Peter Hühn et al., editor, *the living handbook of narratology*. Hamburg University, Hamburg.

Markus John, Martin Baumann, David Schuetz, Steffen Koch, and Thomas Ertl. 2019. A visual approach for the comparative analysis of character networks in narrative texts. In *2019 IEEE Pacific Visualization Symposium*, IEEE Pacific Visualization Symposium, pages 247–256, Piscataway, NJ. IEEE.

Aleksandra Konovalova, Antonio Toral, and Kristiina Taivalkoski-Shilov. 2022. Dr. Livingstone, I presume? polishing of foreign character identification in literary texts. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, pages 123–128, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

Marek Kubis. 2021. Quantitative analysis of character networks in Polish 19th- and 20th-century novels. *Digital Scholarship in the Humanities*, 36(Supplement$_2$) : $ii175 - -ii181$.

Vincent Labatut and Xavier Bost. 2019. Extraction and analysis of fictional character networks: A survey. *CoRR*, abs/1907.02704.

O-Joun Lee and Jason J. Jung. 2019. Integrating character networks for extracting narratives from multimodal data. *Information Processing Management*, 56(5):1894–1923.

O-Joun Lee and Jason J. Jung. 2020. Story embedding: Learning distributed representations of stories based on character networks. *Artificial Intelligence*, 281:103235.

Uri Margolin. 1990. The what, the when, and the how of being a character in literary narrative. *Style*, 24:453–68.

Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. 2007. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC '07, page 29–42, New York, NY, USA. Association for Computing Machinery.

Franco Moretti. 2011. Network theory, plot analysis. *Stanford Literary Lab 2*.

M. E. J. Newman. 2010. *Networks: an introduction*. Oxford University Press, Oxford; New York.

David Schmidt, Albin Zehe, Janne Lorenzen, Lisa Sergel, Sebastian Düker, Markus Krug, and Frank Puppe. 2021. The FairyNet corpus - character networks for German fairy tales. In *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 49–56, Punta Cana, Dominican Republic (online). Association for Computational Linguistics.

Pavel Vondricka. 2014. Aligning parallel texts with intertext. In *LREC*, pages 1875–1879. European Language Resources Association (ELRA).

Mingzhou Xu, Liangyou Li, Derek. F. Wong, Qun Liu, and Lidia S. Chao. 2020. Document graph for neural machine translation.

A Sources

1. The 5 Least Important Characters in Pride and Prejudice, accessed 09.01.2022. https://theseaofbooks.com/2016/04/29/the-5-least-important-characters-in-pride-and-prejudice/

2. Austenopedia, accessed 09.01.2022. http://austenopedia.blogspot.com/p/entry-number-1.html

3. Bleak House Characters | Course Hero, accessed 09.01.2022. https://www.coursehero.com/lit/Bleak-House/characters/

4. Washington Square Character Analysis | LitCharts, accessed 09.01.2022.https://www.litcharts.com/lit/washington-square/characters