

# Natural Language Processing for Literary Text Analysis: Word-Embeddings-Based Analysis of Zofka Kveder's Work

Senja Pollak<sup>1</sup>, Matej Martinc<sup>1,3</sup>, and Katja Mihurko Poniz<sup>2</sup>

<sup>1</sup> Jožef Stefan Institute, Ljubljana, Slovenia  
{senja.pollak, matej.martinc}@ijs.si

<sup>2</sup> University of Nova Gorica, Nova Gorica Slovenia  
katja.mihurko.poniz@ung.si

<sup>3</sup> Jožef Stefan International Postgraduate School, Ljubljana, Slovenia

**Abstract.** In this paper we use word embeddings to analyse the corpus of work of a Slovenian modernist writer Zofka Kveder. For a selection of central concepts of her work, we compute the closest FastText embeddings. The interpretation shows that many of the word relations can be interpreted in line with the findings in literary studies (e.g., woman is discussed in relation to men and Catholicism), or open novel paths for interpretation (e.g., woman in relation to European). On the other hand we also point at some problems resulting from the lemmatization and from using FastText embeddings.

**Keywords:** Digital literary studies · Word embeddings · Distant reading · Gender · Zofka Kveder

## 1 Introduction

The computational methods have importantly enriched the studies of literary history in the last two decades. However, it can also be stated that the digital humanities depend in large part on literary studies [19]. Since 1980s, the feminist theory and gender studies have played an important role in the analysis of literary texts. Moreover, the feminist literary history has made visible a large number of forgotten women authors through digitalization projects, such as The Orlando Project (<https://www.artsrn.ualberta.ca/orlando/>), The Women Writers Project (<https://www.wwp.northeastern.edu/>) and the Virtual Research Environment NEWW Women Writers in History (<http://resources.huygens.knaw.nl/womenwriters/>), which is focusing not only on collection of

---

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). DHandNLP, 2 March 2020, Evora, Portugal.

women’s writings but especially on the collection of reception data. This illustrates the fruitful connection between feminist literary history and digital humanities. Projects that connect computational methods and feminist approaches with literary texts, enable better understanding of women’s roles, especially in the literary life of earlier periods. Through the use of approaches such as distant reading, developed by F. Moretti [15] in the Stanford Literary Lab, and by novel data visualization and data analysis methods applied to large textual corpora, new light can be shed on the work of forgotten or neglected women writers.

The writings of women have always been an integral part of the literary field although they encountered many obstacles caused by the male dominance [2]. Feminist digital literary studies have impacted the field of digital humanities, but regrettably also here, their contributions have not always been recognised and acknowledged [20]. Recent studies [17,4] prove how feminist approaches within digital humanities enable new findings and more complex understanding of the literary history. Also in our work, we investigate a female author, and position our work in the field of digital humanities in combination with with the feminist literary history and gender studies. We focus on the work of Zofka Kveder who was one of the most important female writers in the multicultural space of Habsburg Empire. Zofka Kveder’s work has been previously transformed to digitalized form and identified as a good source for digital humanities investigations [12], but the presented work is the first actual study analysing the work with digital humanities and natural language processing methods.

In this paper, we test how natural language processing methods can facilitate the investigation of a literary text, in particular how using word embeddings can reveal interesting relations between concepts in the work of Zofka Kveder. Word embeddings are vector representations of words, where each word is assigned a real-valued vector in a vector space. Due to the distributional hypothesis (see Zellig Harris [7]), which states that words used in similar context will have similar meanings, embeddings can capture a certain degree of semantics by translating semantic relations in text into vector space relations. Word embeddings have been previously applied to literary text corpora. For example, Grayson et al. [6] use them to compare the characters in 19th-century fiction by the authors Jane Austen, Charles Dickens, and Arthur Conan Doyle, while Wohlgenannt et al. [21] test several word embeddings methods on the task of extracting a social network for literary texts, specifically on the series of fantasy novels. In our experiments, we train FastText embeddings [1] on the corpus of works of Zofka Kveder and for selected concepts identify twenty nearest semantic neighbours according to the cosine similarity between the embedding vectors. These are then interpreted from a literary perspective, showing the potential of simple computational approaches to literary investigation, as well as some deficiencies of automated methods.

## 2 Corpus of Zofka Kveder

Zofka Kveder (1878-1926) was a multilingual and multicultural author who lived in three Central European capitals: Ljubljana, Prague and Zagreb and wrote in

three languages: Slovenian, Croatian and German. Many of her works were published in newspapers and literary magazines in the Central and South-Eastern Europe in Czech, Slovak, Bulgarian, Serbian, Polish and Serbian language. She was also a cultural mediator and an ardent feminist. In most of her stories, a female character in various roles is in the foreground. She also touched on the concept of free love, which was an important issue at the time, and acknowledged the problems of forced marriages, women's urges, illegitimate motherhood, abortion, suicide, prostitution, early death at childbirth and many other themes from the lives of women. As many authors from the late 19th and early 20th century, Zofka Kveder depicted the incompatibility of women's emancipation with marriage and motherhood, and criticized the double moral of the middle-class society, which could not accept an unmarried woman to enjoy sex without feelings of guilt [14].

She also looked for concrete possibilities that would allow women to overcome their position as the Other, to change their relationship with their own bodies and to overcome feelings of guilt and uselessness, which, as she demonstrated, could lead to the disintegration of identity or even death. Having opted for stylistic pluralism, Kveder shaped her narrative using naturalist-realistic stylistic devices, while also feeling an affinity for stylistic procedures typical of New Romanticism. Her works have been translated to many European languages, including Bulgarian, English, German, Polish, Czech.

The corpus in our study contains all Kveder's Slovenian writings: two novels, two plays, two one-act plays, some dramatic scenes, a large number of short stories and tales and articles (literary reviews, feminist writings, etc.). The corpus was published in five volumes of the Collected works of Zofka Kveder - as printed and electronic books edited by K. Mihurko Poniž [10,9]. In total, the corpus contains 1,217,517 tokens.

### 3 Selected concepts

For our analysis, we have selected 12 concepts (words): *ženska* [woman], *žena* [wife/woman], *moški* [man], *mož* [husband/man], *moderen* [modern], *duša* [soul], *pijanec* [drunkard], *ljubezen* [love], *otrok* [child], *emancipiran* [emancipated], *potovati* [to travel], *mati* [mother].

The concepts have been selected according to the prevailing topics in Kveder's works. The selection criterion was also Kveder's relationship to the concepts of the so-called Wiener Moderne (Viennese modern age). In the last decade of the 19th century and in the first decade of the 20th century in the majority of the European literatures different stylistic formations had been developed that are associated under the umbrella-term of the literary modernity (cf. Le Rider [11]). We wanted to find out how Kveder positioned herself and her literary figures towards the concept of "modern". The authors of the Viennese modern age re-visited the concepts of gender roles, therefore the concepts of woman, man and love were chosen. The femaleness was in the majority of writings of the modern age still connected with the motherhood, therefore we were interested how

Kveder connected different topics around this thematic field and consequently if we can conclude that she represented the conservative or progressive views on this problem (concepts of mother and child). The writers who belonged to the literary modernism (especially to the literary currents of symbolism and new romanticism) developed new views on the relationship between body and internal world - in Slovenian language the word soul was used at that time for the individual's psyche. In the feminist but also in the literary writings Kveder reflected upon women's emancipation (therefore we selected the concept of emancipation), which was also one of the central topics of the modern age. Since in the late 19th and early 20th century women traveled and explored new worlds more often than ever before, we were also interested, how the concepts of travel and migrations were coded in Kveder's texts (concept of travel). The concept of the drunkard was chosen because of Kveder's strong (autobiographical) interest in the topic of alcoholism.

## 4 Method

Even if usually word embeddings are trained on very large corpora, there are several studies that show that embeddings can also lead to useful results on smaller text collections. For example, a recent study by Diaz et al. [5] suggests that leveraging embeddings trained on a large general corpus for modelling semantic relations on a specialized corpus is problematic due to strong language use variation. The study shows that embeddings trained only on a small topic specific corpus outperform non-topic specific general embeddings trained on very large general corpora for a somewhat specific task of query expansion in specialized text. Similar, FastText embeddings trained on small domain corpora were used in Pollak et al. [18]. Since we are also dealing with a specialized historical text written by a single author known for its distinct writing style, we follow this line of research and train embeddings only on the author's text.

To counteract the negative affects of a small training corpus, we use FastText embeddings [1] as they are capable of capturing subword information and modeling of affixes and suffixes by representing a word as an average of its character n-grams of specific length. This allows the model to compensate for the lack of available semantic information due to small corpus size by leveraging also morphological similarity, which in many cases translates to semantic relatedness. Also, since Slovenian is a morphologically rich language, FastText embeddings are trained on a lemmatized corpus of Zofka Kveder's texts. Lemmatization was employed in order to prevent the scenario in which the majority of nearest semantic neighbours we want to obtain for each seed concept would be different forms of the input concept. Text was lemmatized using Lemmagen lemmatizer (Juršič et al., 2010). After that, for each of the chosen concepts we obtain 20 nearest semantic neighbours according to the cosine similarity between the embedding vectors. The selected sets are then used for literary investigation, presented in next section.

**Table 1.** 20 closest word embeddings for 12 seed concepts in Slovenian (asterisk \* denotes words which were incorrectly lemmatized; in some cases the correct lemma is given in square brackets).

Seed word	20 closest semantic neighbours (Slovenian)
ženska	ženeva, evropejec, moški, mošnja, mera, žena, moštvo, moš*, farizej, razmera, konkurent, klerikalec, klerikalizem, konkurenca, modistinja, konkurenčen, ovsenjaka, bošnjak, filolog, možitev
žena	mož, ženin, ženitev, možen, sina, zmožen, uboštvo, možitev, rodbina, ženska, neomožen, možev, ženeva, moštvo, omož* [omožen], družina, milodar, evropejec, naključje, vdova
moški	joški, refoški, ženski, filološki, kmetiški, ženska, loški, spol, miški, bloški, psihološki, kakršen, fanuški*, ženstvo, emancipiran, šiški [Šiška], evropejec, neženski, privilegiran, moštvo
mož	defraudant, vdova, žena, ženin, vnuk, sina, tesin*, nadlega, vnukinja, vdov, zbera, vnovič, ženitev, oče, sin, sorodništvo, služinčad, guvernanta, rojak, starš
moderen	monumentalen, operen, uporen, efekten, tipičen, uvoden, orientalen, obnošen, vzhoden, mody*, originalen, buben, ošpičen, patetičen, okusen, koncerten, teren, impozanten, asketičen, frapanten
duša	hrepeneje, dušeč, srce, hrepeneč, verujoč, čuteč, vesoljstvo, grenkoba, atom, slepota, toplina, suša, tesnoba, mrzlota, napetost, hrepeneži, omamljenost, gorkota, pesimizem, sočutje
pijanec	jeruža, žganec* [žganje], dedec, indijanec, galicijanec, žganjar, sodovka, pivec, umazanec, košir, kregar, debevčev, grunt, podgradar, hribovka, oštir, mešičkovka, pokora, matijec*, kristjan
ljubezen	zvestoba, vkljub, neljub, vdan, samoljuben, ljuta, samoljubje, preljubzniv, preljub, zaljubljenost, zvest, verujoč, sočutje, tesnosrčen, dobrota, lažniv, objetje, milina, poljuben, ponižujoč
otrok	sirotica, vnuk, vnukinja, angel, njivica, odrasel, šib, lenoba, črviček, ubožica, angela, dekletec, mezinec, snah, sinček, terezinka, rajnki, deca, sinka, sinek
emancipiran	talentiran, improviziran, absolviran, izoliran, privilegiran, karakterizirati, cimperman, kompliciran, civiliziran, filolog, tradicionalen, respekt, naturalizem, rezerviran, konvencionalen, tiranizirati, individualiteta, individualen, afektiran, manir
potovati	pestovati, službovati, izposlovati, posvetovati, odsvetovati, prorokovati, zborovati, posredovati, gostovati, poizvedovati, nazadovati, vasovati, učinkovati, obrekovati, obiskovati, obedovati, nadzorovati, napredovati, vojskovati, tekmovati
mati	obubožati, hči, obdržati, omehčati, poizkušati, svarilo, obljubovati, hčerin, dedov, starati, očenati* [očenaš], pokopavati, hčerkin, prikrajšati, vnuk, objokovati, odklepati, zblížati, sinek, obupati

**Table 2.** 20 closest word embeddings for 12 seed concepts in English translation (asterisk \* denotes words which were incorrectly lemmatized in Slovene).

Seed word	20 closest semantic neighbours (English translation)
woman	geneva, european, man, pouch, measure, woman/wife, manhood/team, husband*, pharisee, situation, competitor, cleric, clericalism, competition, milliner, competitive, oatcake, bosniak, philologist, wedlock
wife/woman	husband, groom, marriage, possible, son, capable, poorness, wedding, relatives, woman, unmarried, husband's, geneva, maleness, marry*, family, charity, european, coincidence, widow
man	boobs, refošk, female, philological, peasant, woman, loški, gender, mice, bloški, psychological, kind, fanuški, womanhood, emancipated, šiška, european, unmarried, privileged, maleness
husband/man	defalcator, widow, wife, groom, grandson, son, tesin*, problem, granddaughter, widow, gather, again, marry, father, son, kinship, domestics, governess, compatriot, parent
modern	monumental, opera, rebellious, showy, typical, introductory, oriental, worn-out eastern, mody*, original, bubna, naughty, pathetic, tasty, concert, terrain, imposing, ascetic, fascinating
soul	longing, suffocating, heart, yearning, believing, feeling, universe, bitterness, atom, blindness, warmth, drought, anxiety, coldness, tension, longing, stoned, bitterness, pessimism, compassion
drunkard	spirit, alcohol, buster, indian, galician, distiller, soda, drinker, dirty, košir, kregar, debevec's, estate, podgradar, highlander, innkeeper, mešičkovka, penance, matijec*, christian
love	fidelity, despite, undesirable, devoted, self-impetuous, angry, self-centeredness, kindest, dearest, infatuation, faithful, believing, compassion, coldhearted, kindness, lying, embracement, grace, arbitrary, humiliating
child	orphan, grandson, granddaughter, angel, field, adult, rod, laziness, worm, pauper, angel, girl, little finger, daughter-in-law, little son, terezinka, deceased, children, little son, little son
emancipated	talented, improvised, absorbed, isolated, privileged, characterized, cimperman, complicated, civilized, philologist, traditional, respect, naturalism, reserved, conventional, tyrannize, individuality, individual, affected, manner
to travel	hold in arms, serve, arrange, consult, advise against, prophesy, convene, mediate, host, inquire, regress, feast, effect, slander, visit, dine, control, progress, wage war, compete
mother	lose all possessions, daughter, retain, soften, try, warn, promise, daughter's, grandfathers, grow old, Lord's prayer, bury, daughter's, deprive, grandson, grieve, unlock, bring together, little son, despair

## 5 Analysis and discussion

In Table 1 we present the results of experiments and in Table 2 the English translations. In this section, we interpret selected concepts and relations from the perspective of literary studies, and reveal some limitations of the method.

The concept *woman* (see table rows 1 and 2) is placed in a relationship with *man* (*man, husband, groom*) and with Catholicism: *Pharisee, cleric, clericalism*. Words related to writer’s family situation (*family, widow, son*) also appear. In connection with the concept *woman*, there is also the word *milliner*, which indicates a common women’s profession in the modern age. The connection with the word *European* is also interesting because literary studies have not discussed it yet.

In the concept of *emancipated*, no word appears in direct connection with a woman. The concept of *mother* appears in connection with *daughter*, which is expected, given the frequent topic of the mother-daughter relationship in Kveder’s writings. The concept *mother* connects with words that express emotions associated with suffering: *to bury, to grieve, to despair*. The literary studies on the topic of motherhood in Kveder’s work connected Kveder’s work with negative representations of motherhood [13].

A *man* is positioned in the relationship to a woman (*wife, marriage*) and in his family role (*son, grandson, father, parent, kinship*). Interesting is the connection with the word *gender*, which does not occur in resulting words for the the seed concept *woman*. The concept *love* is associated with fidelity/loyalty and emotions. Surprisingly, it does not connect with the body, although this was discussed as a frequent topic in Kveder’s oeuvre in the field of literary and gender studies [8].

The concept of *modern* is associated with words such as *rebellious, effective*, original but also *pathetic, imposing, ascetic* and *fascinating*, which are the expected connections according to previous research of Kveder’s writings. The connection with the *Oriental* is interesting and calls for further research. The relation with the concept *soul* (*longing, heart, compassion*) are unsurprising since this concept was often discussed in the modern age. However, the connection with the word *atom* is interesting and unexpected.

Subword information, which influences the similarity in FastText embeddings, generally improves the semantic modelling in morphologically rich languages and small corpora, as it helps with finding semantically similar concepts with similar morphological structure. For example, similar lexemes in *žena* [wife], *ženin* [groom], *ženitev* [marriage] are reflecting also semantic similarity of words. On the other hand, the results indicate that this feature can also lead to strong correlations between semantically (mainly) unrelated words containing many common characters (e.g., strong correlation between *ženska* (woman in Slovenian) and *Ženeva* (Slovenian name for the city in Switzerland)). An interesting example is also *ljubezen* [love], where we can find many etymologically related words with common lexemes, which can be easily interpreted as semantically related, such as *samoljuben* [self-centred], *samoljubje* [self-centeredness],

*preljubezniv* [kindest], but on the other hand we have words with similar subword information that are not closely semantically related (e.g. *poljuben* [arbitrary]).

Another aspect related to string similarity is related to word types. We can notice that the part-of-speech (POS) of seed words and returned words is to a large extent preserved. In our case the majority of seed words are nouns, and corresponding returned embeddings return nearly only nouns, while when adjectives or verbs were selected as seeds, the related words also belong mainly to the respective POS categories. This is due to stronger association between words with the same characteristic word type affixes. For example, the seed adjective *emancipiran* has returned the set of semantically closest words *talentiran*, *improviziran*, *absolviran*, . . . , while the verb *potovati* is associated to verbs *pestovati*, *službovati*, etc. However, when a noun has a typical verb ending (e.g. *mati*), there are many verbs in the returned list. Even if in general grouping of words in relation to their POS can be interpreted as a feature, we found that in our study (on a small corpus) this could be a source of confusion and that nouns provided more interesting and reliable results than verbs and adjectives. Nouns would also be our first choice for selection of seed words for future analysis of literary works in Slovenian language with FastText embeddings. Another thing that made the interpretation of results harder are mistakes in the lemmatization. For example, first name *Joško* was lemmatized into *joške* (Slovenian for boobs), showing that named entities should be marked or better handled in lemmatization.

## 6 Conclusion and future work

In this paper, we investigate how natural language processing methods can facilitate the investigation of literary work. We have trained FastText embeddings on the corpus of the text of Zofka Kveder who was a principal Slovenian modernist, multilingual and multicultural author and a feminist. We have selected twelve concepts (words), according to the prevailing topics in Kveder's works and in relationship to the concepts of the so-called Wiener Moderne (Viennese modern age). For these words, we have computed twenty nearest semantic neighbours according to the cosine similarity between the embedding vectors. These were then interpreted from a literary perspective, showing the potential of simple computational approaches to literary investigation. From the point of view of the research of Zofka Kveder's work, the results of our research confirm the findings of literary history studies, and also open new research directions (e.g. the relation between woman and European, and modern and Oriental), which means that computer analysis provides interesting results that can be a useful approach in literary studies. In future work, we plan to enlarge the set of concepts, investigate how to remove noise by using improved lemmatization, testing the results without lemmatization and recognizing named entities. In addition to observing similarities between concepts, word embeddings allow for investigating analogies, which could be very interesting when applied to literary texts. Further, we plan to use the methods on larger literary corpora, which would allow for embeddings-based analysis of differences between authors, authors' gender

(e.g. male vs. female authors in the same period) and diachronic based studies, which would allow to analyse how certain concepts evolved during time. Last but not least, we will analyse different word embeddings methods (w2v, GloVe,...) to analyse strengths and weaknesses of different methods, and compare the results to statistical word association measures, such as PMI [3] and PPMI [16].

## Acknowledgements

The work presented in this paper has been supported by European Unions Horizon 2020 research and innovation programme under grant agreement No. 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media). The authors acknowledge also the financial support from the Slovenian Research Agency core research programme Knowledge Technologies (P2-0103) and the research programme Historical Interpretations of the 20th century (P6-0347). The authors also acknowledge the COST Action Distant Reading (Grant No. CA 16204).

## References

1. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. In: Transactions of the Association for Computational Linguistics. pp. 135–146 (2017)
2. Bourdieu, P.: La domination masculine. *Actes de la recherche en sciences sociales* **84**(4), 2–31 (1990)
3. Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. *Computational Linguistics* **16**(1), 22–29 (1990), <https://www.aclweb.org/anthology/J90-1003>
4. Coker, C., Ozment, K.: Building the women in book history bibliography, or digital enumerative bibliography as preservation of feminist labor. *Digital Humanities Quarterly* **13**(3) (2019), <http://digitalhumanities.org/dhq/vol13/3/000428/000428.html>
5. Diaz, F., Mitra, B., Craswell, N.: Query expansion with locally-trained word embeddings. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 367–377. Berlin, Germany (August 2016)
6. Grayson, S., Mulvany, M., Wade, K., Meaney, G., Greene, D.: Novel2Vec: Characterising 19th century fiction via word embeddings. In: Proceedings of 24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS'16). Dublin, Ireland (2016)
7. Harris, Z.S.: Distributional structure. *Word* **10**(23), 146–162 (1968)
8. Jensterle-Doležal, A.: Problem identitete in travma telesa v prozi Zofke Kveder (A Problem of identity and the trauma of the body in the prose of Z. Kveder. In: Jensterle-Doležal, A., Honzak Jahić, J. (eds.) *Zofka Kvedrová (1878–1926): Recepte její tvorby ve 21. století*, p. 81–99. Praha: Národní knihovna ČR–Slovanská knihovna (2008)
9. Kveder, Z.: The collected works of Zofka Kveder. 1. In: Poniž, K.M. (ed.) *The collected works of Slovene poets and writers*. Litera, Maribor (2006)

10. Kveder, Z.: The collected works of Zofka Kveder. 2-5. In: Poniž, K.M. (ed.) The collected works of Slovene poets and writers. Založba ZRC, Ljubljana (2014-2018)
11. Le Rider, J.: *Modernity and Crises of Identity: Culture and Society in Fin-de-siècle Vienna*. Continuum, New York (1993)
12. Mihurko Poniž, K.: Digitalisation of the literary creativity : a case study of Zofka Kveder. In: Vraneš, A. (ed.) Digital humanities and Slavic cultural heritage : international scientific conference, Belgrade, 6-7 May 2019 : [the book of summaries]. University of Belgrade, Faculty of Philology, Belgrade (2019)
13. Mihurko Poniž, K.: The depictions of motherhood in the writings of Zofka Kveder and German literary women in the 19th century. In: Kiiskinen, H. (ed.) ISCH 2010 Conference book (May 2010)
14. Mihurko Poniž, K., Parente-Čapková, V.: The new women from the margins. *Interlitteraria* **20**(2), 184 (Dec 2015). <https://doi.org/10.12697/il.2015.20.2.15>
15. Moretti, F.: *Distant reading*. Verso, London and New York (2013)
16. Niwa, Y., Nitta, Y.: Co-occurrence vectors from corpora vs. distance vectors from dictionaries. In: COLING 1994 Volume 1: The 15th International Conference on Computational Linguistics (1994), <https://www.aclweb.org/anthology/C94-1049>
17. Ozment, K.: From recovery to restoration: Aphra behn and feminist bibliography. *Early Modern Women: An Interdisciplinary Journal* **13**(1), 105–116 (2018)
18. Pollak, S., Repar, A., Martinc, M., Podpečan, V.: Karst exploration: Extracting terms and definitions from karst domain corpus. In: *Proceedings of eLex 2019*. pp. 934–956. Sintra, Portugal (2019)
19. Pressman, J., Swanstrom, L.: The literary and/as the digital humanities. *Digital Humanities Quarterly* **7**(1) (2013)
20. Wernimont, J.: Whence feminism? assessing feminist interventions in digital literary archives. *Digital Humanities Quarterly* **7**(1) (2013)
21. Wohlgenannt, G., Chernyak, E., Ilvovsky, D.: Extracting social networks from literary text with word embedding tools. In: *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*. pp. 18–25. Osaka, Japan (December 2016)