Proceedings of ISA-18 Workshop at LREC2022, pages 27–32
Marseille, 20 June 2022

Levels of Non-Fictionality in Fictional Texts

Florian Barth[1], **Hanna Varachkina**[2], **Tillmann Dönicke**[1], **Luisa Gödeke**[2]
[1]Göttingen Centre for Digital Humanities, [2]Department of German Philology
University of Göttingen
florian.barth@uni-goettingen.de, hanna.varachkina@stud.uni-goettingen.de,
tillmann.doenicke@uni-goettingen.de, luisa.goedeke@uni-goettingen.de

Abstract

The annotation and automatic recognition of non-fictional discourse within a text is an important, yet unresolved task in literary research. While non-fictional passages can consist of several clauses or sentences, we argue that 1) an entity-level classification of fictionality and 2) the linking of Wikidata identifiers can be used to automatically identify (non-)fictional discourse. We query Wikidata and DBpedia for relevant information about a requested entity as well as the corresponding literary text to determine the entity's fictionality status and assign a Wikidata identifier, if unequivocally possible. We evaluate our methods on an exemplary text from our diachronic literary corpus, where our methods classify 97% of persons and 62% of locations correctly as fictional or real. Furthermore, 75% of the resolved persons and 43% of the resolved locations are resolved correctly. In a quantitative experiment, we apply the entity-level fictionality tagger to our corpus and conclude that more non-fictional passages can be identified when information about real entities is available.

Keywords: Named Entities, Fictionality, Semantic Web, Automatic Annotation, German

1. Introduction

One can easily distinguish three levels of (non-) fictionality in a text: First, a text might be classified as a work of fiction or a work of non-fiction according to whether it describes imaginary or actual events, people or places. Second, it is a common observation that a fictional text does not only consist of fictional discourse but also contains passages that suggest assertions or hypotheses about the real world.[1] And third, the people or places mentioned in a text may exist in the real world even if the text or the story is overall fictional.

Non-fictional passages within a fictional text are of special interest in literary studies because they often contain central messages of a work or correspond with specific statements or intentions of the author. However, although the automatic classification of texts into fiction and non-fiction can be considered a solved problem (e.g. Piper (2016)), the identification of non-fictionality within a fictional text remains an open task. While our ultimate goal is to identify non-fictional passages, we assume that the fictionality status of named entities can serve us as feature and we further consider the automatic annotation of (non-)fictional entities in a text to be a useful application on its own (cf. van Dalen-Oskam et al. (2014), Chu et al. (2020)).

In this paper, we briefly describe the theoretical background and working hypotheses on non-fictional passages (Sec. 2), our still-growing corpus with manual annotations (Sec. 3), the automatic annotation of fictionality for named entities (Sec. 4–6), and an analysis of the interplay between entity-level and passage-level fictionality (Sec. 7).

[1]Likewise, a non-fictional text might contain passages that make assertions about fictional people or events.

2. Theoretical Background

In fictional literature, fictional discourse builds the fictional world, e.g. introduces characters and describes actions and scenes. From a linguistic perspective, it has been repeatedly observed that fictional discourse challenges the semantic notion of truth and reference. This is because fictional discourse is obviously not true and does not refer to real entities in the real world. In (1), for example, for us readers it is clear that the character Gustav (the protagonist of the work) does not correspond to any real-world entity and all the other information concerning this character are not true, either.

(1) He [Gustav] loved the ocean for deep-seated reasons: because of that yearning for rest, when the hard-pressed artist hungers to shut out the exacting multiplicities of experience and hide himself on the breast of the simple, the vast; and because of a forbidden hankering—seductive, by virtue of its being directly opposed to his obligations—after the incommunicable, the incommensurate, the eternal, the non-existent. (Mann, 2021)

Therefore, the most influential view on fictional discourse is that fictional utterances are invitations to imagine things (cf. Currie (1990), Konrad (2017), Stock (2017), Maier (2017)).

This approach, however, neglects a certain macrostructural property of some passages in fictional texts that do not prima facie contribute to building the fictional world: so-called non-fictional-passages (NfPs). Let us consider the continuation of example (1):

(2) [...], the non-existent. To be at rest in the face of perfection is the hunger of everyone who is aiming at excellence; (Mann, 2021)

The last sentence poses problems for the imagination hypothesis on the one hand and for the truth notion on the other: The utterance, after all, may indeed be true beyond the fictional world. Thus, unlike a purely fictional utterance, it obviously does not serve exclusively to create the fictional world. Therefore, it is misleading to assume this utterance to exclusively serve the imagination. From this, we conclude that fictional works can consist of both fictional and non-fictional discourse.
Non-fictional discourse in itself can come in different varieties. In our example, it is a kind of generic, aphoristic wisdom (cf. Konrad (2017)). Another form of NfPs can refer to real places or people, sometimes with detailed descriptions apparently researched by the author, e.g. this description of a slaughterhouse in Berlin:

(3) In the northeast part of the city, from Eldenaer Strasse across Thaerstrasse across Landsberger Allee as far as Cotheniusstrasse along the Belt Line Railway, run the houses, halls, and stables of the slaughter- and stock-yards. They cover an expanse of 44.78 hectares, equal to 118.31 acres. Not counting the structures behind Landsberger Allee, 27,083,492 marks were sunk into its construction, [...]. (Döblin, 2003)

Konrad (2017) argues that these two forms of NfPs are characterised by certain linguistic features, including generalisation/abstraction, researched details and technical language. In addition to various forms of generalisation, we consider immigrant objects (Parsons, 1981) to be crucial, i.e. objects that migrated from the real world to the fictional world.[2] We therefore assume that non-fictional discourse mainly consists of generalisations and named entities referring to the real world.

3. Data and Annotation

We currently construct a diachronic corpus of German fictional literature from 1600 to 1950. Most of the texts originate from the KOLIMO corpus (Herrmann and Lauer, 2017), which is a subsample of prose texts extracted from TextGrid-Repository[3] and Project Gutenberg (Reu, 2013) encoded in TEI-XML and enriched with metadata such as identifiers from the Integrated Authority File (GND, German for "Gemeinsame Normdatei") for the author of each corpus record.[4]
As of now, we annotated 22 texts (6,555 sentences). Our annotation procedure is as follows: Each text is first annotated by four out of six student assistants (in varying constellations), all having a background in German philology. Two annotators each annotate either non-fictional passages (NfPs) or generalising passages (GenPs), where we define a passage to span at least one and potentially an open number of subsequent clauses. In a second step, the initial annotations are discussed and then confirmed, corrected or deleted by two researchers, yielding our gold standard.
The annotation of NfPs includes all passages that suggest assertions or hypotheses about the real world (considering the time when the text was written). An example is shown in (4), which is a free translation from May (1888). The boldfaced passage makes the assertion that confederate prisoners were interned in Fort Jefferson at the time of the story. Since this makes a reference to events during the American Civil War, which the author presumably had knowledge of, the assertion can be understood to be about real world's Fort Jefferson. Note that Fort Jefferson is not called by name in the boldfaced passage but referenced by the anaphoric pronoun *this*. In such cases, where some context is required to properly interpret an NfP, we additionally annotate a larger span (with a separate tag) that includes the NfP and the minimal reference context (underlined in the example). Although the first sentence mentions two real-world entities, *Tortuga* and *Fort Jefferson*, we do not consider it to be non-fictional discourse because it describes fictional events and does not suggest assertions about the real world.

(4) The storm had driven our ship against the Tortugas, against the island on which Fort Jefferson is located. **Confederate prisoners of war were interned in this at the time**. The fishermen took care of me in the friendliest way and provided me with fresh linen and the most necessary clothes, for I was only dressed in the way *in which one goes to bed during a sea voyage*.

The annotation of GenPs is independent of fictionality and driven by mainly linguistic criteria. For example, the italicised passage in (4) makes a generalising claim about how one used to be dressed for sleep during a sea voyage. We use the tagset of Dönicke et al. (2021) to annotate subcategories of GenPs but these are not relevant for this paper.
The average inter-annotator agreement measured with γ (Mathet et al., 2015) is shown in Table 1. NfPs and GenPs are annotated with substantial agreement.[5]

Tagset	Tags	γ (multi)	γ (binary)
NfP	2	.73	.79
GenP	6	.65	.68

Table 1: Inter-annotator agreement considering subclasses (multi) or merging all classes into one (binary).

4. External Resources

To use external knowledge about fictional and real entities, we integrate knowledge graph databases, Wikidata

[2]We use the terms "immigrant object" and "real(-world) entity" synonymously in this paper.

[3]https://textgridrep.de/

[4]https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html.

[5]Our corpus and annotation guidelines are published in Barth et al. (2021).

and DBpedia, that can be queried via SPARQL.

4.1. Wikidata

Wikidata is a free, multilingual, collaborative and open knowledge base developed by the Wikimedia Foundation that can be read and edited by both humans and machines.[6] Wikidata was launched in 2012 (Vrandečić and Krötzsch, 2014) and currently holds more than 97 million items. It consists mainly of items with a label (name), their description and aliases (alternative names). The structure of the data is the following: item – property – value, e.g. *Harry Potter* (Q3244512) – *instance of* (P31) – *literary character* (Q3658341) or *Globe Theatre* (Q272434) – *located in the administrative territorial entity* (P131) – *London Borough of Southwark* (Q730706). This structure corresponds to the graph format (semantic triples: Subject – Predicate – Object) and can be queried using a SPARQL query service that Wikidata provides.

4.2. DBpedia

DBpedia is a community-based platform that aims to extract structured information from Wikipedia articles so that Semantic Web techniques can be employed such as SPARQL queries or an interlinking of datasets (Auer et al., 2007). The DBpedia dataset currently consists of 850 million facts (RDF triples)[7] and it is interlinked with several open datasets from a wide range of domains such as lexical resources (WordNet), spatial knowledge bases (Geonames, LinkedGeoData), social networks (FOAF), literary resources (Project Gutenberg), and other encyclopedias (Wikidata).

5. Metadata Extraction and Enrichment

Based on the GND-identifier for a work's author within the KOLIMO corpus, we identify the author's Wikidata entry and, if existent, the Wikidata entry of the current text that we process. We employ this metadata later for the classification of fictionality and the linking of Wikidata entries to named entities, which is why we developed an own metadata structure to store and process metadata from the original corpus and own enrichments – enriched metadata can be, furthermore, serialised back to TEI-XML format.

6. Entity Classification and Linking

We parse our texts with spaCy, which also contains a named entity (NE) recogniser.[8] The NE recogniser assigns the labels PER, LOC, ORG and MISC to denote persons, locations, organisations and miscellaneous, respectively. We also use an advanced version of Krug et al. (2015)'s algorithm for coreference resolution on all noun phrases, including NEs.

Building on the preprocessing, we aim to solve two tasks: 1) determine if an NE is fictional or real and 2) assign a specific Q-identifier from Wikidata to the NE. For now, we only consider NEs tagged as PER or LOC, because persons and locations are the most relevant categories in novels, while organisations and miscellaneous are less common. In a first step, we create a set of variant forms for each NE and request information about them via the SPARQL interface of Wikidata and DBpedia. Beside the form of the NE that appears in the text, we add a variant based on the longest mention in the NE's coreference chain.[9] Furthermore, we add variants to queries by constructing the nominative form for NEs in genitive case, deleting function words using part-of-speech tags, considering only tokens with the suffix *-isch* (that might indicate locations) and normalising old spelling by substituting the *ß*-ligature.

To identify and differentiate PER entities (PERs), we check if one of the queried variants equals or is part of a description or an alias of an item (subject) that has an *instance of* property (P31) corresponding with an item (object) that we regard as either fictional or real. Relevant fictional items are among others *literary character* (Q3658341) and *fictional human* (Q15632617). Real items correspond especially with an instance of *human* (Q5). Since querying a large amount of *human* items exceeds the Wikidata API, we query DBpedia for *foaf:Person* entries that are supplied by a Wikidata identifier that has an *instance of* relation to the Wikidata item *human*. We further regard certain real-world concepts as immigrant objects, e.g. mythological PERs such as god in monotheistic religions (inter alia: Q190, Q2095353, Q2155501, Q825, Q5576009) or Greek deities (Q22989102, Q878099). Besides the direct identification of Wikidata items corresponding to the set of queries, we utilise Wikidata entries for author and work from our enriched metadata record to identify and scrape the Wikipedia article of the current text. If literary characters do not have an own Wikidata entry to link them, we can instead identify them within the Wikipedia article. In this case, no linking is applied, but corresponding PERs will be classified as fictional.

The Wikidata property that helps identify real locations the best is *coordinate location* (P625). If a query that contains this property does not yield any results, other properties are used: *located in the administrative territorial entity* (P131), *located in or next to body of water* (P206), *located in time zone* (P421), *country* (P17), *area* (P2046), *significant place* (P7153) and *located in the statistical territorial entity* (P8138). The search is conducted among labels of Wikidata items first, and — if it yields no results— continues among aliases of item labels.

The collected Wikidata entries serve as candidates for which we extract features for the classification of fictionality and the entity linking. These features are based on the requests for the query variants and the enriched metadata. Thereupon, we apply a scoring sys-

[6] https://www.wikidata.org/wiki/Wikidata:Main_Page

[7] https://www.dbpedia.org/blog/snapshot-2021-12-release

[8] https://spacy.io/api/entityrecognizer

[9] This variant consists of nominal phrases including adjectives and strips other tokens (like verbs, pronouns etc.).

Type	Tot.	Fict.	Real	None	A.	Link.	A.
PER	149	125 (.84)	6 (.04)	18 (.12)	.97	8 (.06)	.75
LOC	90	19 (.21)	36 (.40)	35 (.39)	.62	46 (.84)	.43
both	239	144 (.60)	42 (.18)	53 (.22)	.87	54 (.29)	.48

Table 2: Evaluation of classified and linked NEs: total number of NEs; number (percentage) of fictional / real / incorrectly recognised NEs; accuracy for classifying fictional and real NEs; number (percentage) of linked fictional and real NEs; accuracy for linking NEs.

Type	Gold	P.	R.	F1	Link.	A.
PER	Fict.	1.00	.97	.98	.02	.00
	Real	.60	1.00	.75	1.00	1.00
LOC	Fict.	.45	.47	.46	–	–
	Real	.71	.69	.70	.69	.80
both	Fict.	.92	.90	.91	–	–
	Real	.69	.74	.71	.74	.84

Table 3: Evaluation separated by fictionality status (as labelled in the gold standard): precision / recall / F1 for classifying NEs; percentage of linked NEs; accuracy for linking NEs.

tem for both tasks that weights indicators for either fictional or real items and assigns a Wikidata entry if possible. For PERs, the scoring considers the amount fictional and real items that have an *instance of* relation to the entry candidate. If an entry candidate is *instance of human* or is associated with other figurative concepts of the real world, the Wikidata sitelinks are applied in an item-class-adjusted manner to estimate the importance of this Wikidata entry. The higher scoring value for fictional or real determines the fictionality classification and if the corresponding Wikidata entry candidate holds a defined minimal value of sitelinks it will be linked to the NE. For LOCs, the scoring relies on sitelinks as well as the NE's context (its clause), which we compare with the Wikidata description of an entry candidate. The minimal amount of sitelinks to accept a linking can be lower for locations since locations seem less interlinked than persons or fictional characters.

7. Evaluation and Analysis

We test our classifier on one text from our corpus—Fontane (2012)—where we manually classified each LOC and PER (as found by the NE recogniser) as fictional or real and compared the manual annotation with the automatic one. As Table 2 shows, 239 entities are recognised in the text, from which we labelled 60% as fictional and 18% as real. The remaining 22% constitute errors by the NE recogniser, which we exclude from the evaluation. From the correctly recognised NEs, 87% are correctly classified as fictional or real, including all 6 real persons; 29% are linked to a Wikidata entry, where 48% of the links are correct. We achieve higher accuracies for PERs than for LOCs in both tasks. Table 3 shows separate results for fictional and real entities. Overall, fictional and real entities are identified with 91% and 71% F1, respectively. 74% of the real entities are linked to a Wikidata entry, where 62% of the links are correct. On the other hand, we do not link fictional LOCs to a Wikidata entry so far. For fictional PERs there are no NEs in the text that have a Wikidata entry and therefore no fictional PERs that should be linked. The accuracy for only correct links would therefore be 0/0=NaN (not 100%). Still, 2 fictional PERs are incorrectly linked to a Wikidata entry, which produces an accuracy of 0/2=0%.

	RE	GenP	RE ∪ GenP	RE ∩ GenP
$P(x\|\text{NfP})$	.14	.72	.78	.08
$P(\text{NfP}\|x)$	.15	.29	.23	.68

Table 4: Observed probabilities for cooccurrences of NfPs with immigrant objects (REs) and/or GenPs, based on 10 texts of our corpus.

Table 5 presents the results for the fictionality classification and the NE linking. For PERs, we correctly identify historic entities such as "Fiedrich Wilhelms IV" (Frederick William IV; Q57180) for which the alternative spellings help to formulate a query including the nominative ("Wilhelm"). Furthermore, simple forenames like Dubslav (the main character) can be identified as characters of the novel through Wikipedia scraping based on enriched metadata. "Berlin" (Q64) is correctly recognised as an instance of real location in text. Although multiple entities with this label are found on Wikidata, the number of sitelinks helps to assign it to the correct one as capital of Germany. "Schloß Stechlin" is a fictional castle. Even though the spelling was adapted to the modern grammar for a query, it could not be found among real locations on Wikidata, correctly so. Interestingly, the NE "Stechlin" appears multiple times in the novel: as a real lake, a fictional village, the main character by his surname and as a family name that is not linked to any entity, which poses a lot of problems to the NE recogniser and for disambiguation of the entities existing in the real world. The evaluation results suggest that the entity-level fictionality classification works adequately enough to perform a quantitative analysis of non-fictionality on the passage level in our corpus. Our hypothesis from Section 2 is that NfPs are usually generalising or contain a real entity (RE). The first row in Table 4 shows that indeed 72% NfPs (including the minimal reference context) overlap with a GenP, 14% contain an RE, and 8% do both. Returning all passages overlapping with an RE or GenP, would correctly find 78% of the NfPs. However, as the second row in Table 4 shows, both REs and GenPs occur far more often in purely fictional than in non-fictional passages,[10] which means that the returned

[10] Since we do not annotate "purely fictional passages", we

Named Entity	Entity Type	Queries	Count	Real	Wikidata item	Wikidata Description	Sitelinks
'Friedrich Wilhelms IV.'	PER	{'Friedrich Wilhelms IV.', 'Friedrich Wilhelm IV.', 'Regierungsantritt Friedrich Wilhelms IV.', [...] }	1	True	Q57180	King of Prussia (1795-1861)	65
'Dubslav'	PER	{'Dubslav', 'Regiment Garde du Corps'}	10	False	None	None	None
'Berlin'	LOC	{'Berlin'}	2	True	Q64	federal state, capital and largest city of Germany	410
'Schloß Stechlin'	LOC	{'Schloß Stechlin', 'Schloss Stechlin'}	1	False	None	None	None

Table 5: Examples of PER and LOC NEs in Theodor Fontane's *Der Stechlin* with query variants for Wikidata and query results (item, description, sitelinks)

passages still have to be filtered to get a good precision.

8. Conclusion and Future Work

We observe that the spaCy model recognises a solid number of named entities, but also makes a considerable number of mistakes. We plan to evaluate the NE recogniser used in spaCy on the domain of fictional literature. So far, we excluded wrongly recognised NEs from the analysis, in the future, we will also consider those NEs that were missed by the model. The mistakes are likely caused by the fact that the spaCy model was trained on data from a different domain, namely Wikipedia articles. For comparison, Jannidis et al. (2015) report a performance decrease of 65% F-score and as much as 74% recall for a system trained on newspaper texts and applied to a corpus of German novels, while precision remains similar. Therefore, we plan to adapt the existing NE recogniser for characters in fiction[11] by extending it to locations.

We have seen that 78% of non-fictional passages contain either a real-world entity or a generalisation. While this paper presents methods for the identification of real entities,[12] the identification of generalising statements (e.g. Friedrich et al. (2016), Gödeke et al. (to appear)) and the combination of both into a passage-level fictionality recogniser constitute another task which we have to solve in the future.

9. Acknowledgements

Our special thanks go to our fellow members of the MONA project group, who were actively involved in developing the foundations of this work: Benjamin Gittel, Anke Holler, Caroline Sporleder, and Anna Weimer. We further thank our student assistants for their engagement, and the anonymous reviewers for their valuable comments. This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 424264086 (Barth, Varachkina) and Volkswagen Foundation (Dönicke, Gödeke).

treat REs and GenPs that do not overlap with an NfP as purely fictional passages for the calculation.

[11] https://github.com/MarkusKrug/NERDetection

[12] Our implementation is available at https://gitlab.gwdg.de/mona/pipy-public/-/releases/v2.0.

10. Bibliographical References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., and Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer.

Barth, F., Dönicke, T., Gittel, B., Gödeke, L., Hofmann, A. M., Holler, A., Sporleder, C., and Varachkina, H. (2021). MONACO: Modes of narration and attribution corpus. https://gitlab.gwdg.de/mona/korpus-public.

Chu, C. X., Razniewski, S., and Weikum, G. (2020). ENTYFI: A system for fine-grained entity typing in fictional texts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 100–106, Online, October. Association for Computational Linguistics.

Currie, G. (1990). *The nature of fiction*. Cambridge University Press.

Dönicke, T., Gödeke, L., and Varachkina, H. (2021). Annotating quantified phenomena in complex sentence structures using the example of generalising statements in literary texts. In *Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, pages 20–32, Groningen, The Netherlands (online), June. Association for Computational Linguistics.

Döblin, A. (2003). *Berlin Alexanderplatz. The story of Franz Biberkopf [1929]. Translated by Eugene Jolas*. Continuum.

Fontane, T. (2012). Der Stechlin. In *TextGrid Repository*. Digitale Bibliothek.

Friedrich, A., Palmer, A., and Pinkal, M. (2016). Situation entity types: automatic classification of clause-level aspect. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1757–1768,

Berlin, Germany, August. Association for Computational Linguistics.

Gödeke, L., Barth, F., Dönicke, T., Weimer, A. M., Varachkina, H., Gittel, B., Holler, A., and Sporleder, C. (to appear). Generalisierungen als literarisches Phänomen. Charakterisierung, Annotation und automatische Erkennung. *Zeitschrift für digitale Geisteswissenschaften*.

Herrmann, B. and Lauer, G. (2017). Kolimo. a corpus of literary modernism for comparative analysis.

Jannidis, F., Reger, I., Weimer, L., Krug, M., and Puppe, F. (2015). Automatische Erkennung von Figuren in deutschsprachigen Romanen.

Konrad, E.-M. (2017). Signposts of factuality: On genuine assertions in fictional literature. *Art and Belief, London*, pages 42–62.

Krug, M., Puppe, F., Jannidis, F., Macharowsky, L., Reger, I., and Weimar, L. (2015). Rule-based coreference resolution in German historic novels. In *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pages 98–104, Denver, Colorado, USA, June. Association for Computational Linguistics.

Maier, E. (2017). Fictional names in psychologistic semantics. *Theoretical Linguistics*, 43(1-2):1–45.

Mann, T. (2021). *Death in Venice [1911]. Translated by Kenneth Burke*. Project Gutenberg.

Mathet, Y., Widlöcher, A., and Métivier, J.-P. (2015). The unified and holistic method gamma (γ) for inter-annotator agreement measure and alignment. *Computational Linguistics*, 41(3):437–479.

May, K. (1888). Winnetou II. In *Projekt Gutenberg*. Haffmans.

Parsons, T. (1981). Nonexistent objects. *Yale University Press*.

Piper, A. (2016). Fictionality. *Journal of Cultural Analytics*, 12.

(2013). Gutenberg-de edition 13 dvd-rom. klassische literatur in deutscher sprache. DVD-ROM.

Stock, K. (2017). *Only imagine: Fiction, interpretation, and imagination*. Oxford University Press.

van Dalen-Oskam, K., de Does, J., Marx, M., Sijaranamual, I., Depuydt, K., Verheij, B., and Geirnaert, V. (2014). Named entity recognition and resolution for literary studies. *Computational Linguistics in the Netherlands Journal*, 4:121–136.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Commun. ACM*, 57(10):78–85, sep.