Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 126–132
Marseille, 11–16 May 2020

Modelling Narrative Elements in a Short Story: A Study on Annotation Schemes and Guidelines

Elena Mikhalkova, Timofei Protasov, Polina Gavin, Anastasiia Bashmakova, Anastasiia Drozdova
Tyumen State University
625003 Volodarskogo Street, 6
Tyumen, Russia
{e.v.mikhalkova, t.a.protasov}@utmn.ru

Abstract

Text-processing algorithms that annotate main components of a story-line are presently in great need of corpora and well-agreed annotation schemes. The Text World Theory of cognitive linguistics offers a model that generalizes a narrative structure in the form of world-building elements (characters, time and space) as well as text worlds themselves and switches between them. We have conducted a survey on how text worlds and their elements are annotated in different projects and proposed our own annotation scheme and instructions. We tested them, first, on the science fiction story "We Can Remember It for You Wholesale" by Philip K. Dick. Then we corrected the guidelines and added computer annotation of verb forms with the purpose to get a higher raters' agreement and tested them again on the short story "The Gift of the Magi" by O. Henry. As a result, the agreement among the three raters has risen. With due revision and tests, our annotation scheme and guidelines can be used for annotating narratives in corpora of literary texts, criminal evidence, teaching materials, quests, etc.

Keywords: cognitive linguistics, natural language processing, literature, text annotation, corpus, Text World Theory

1. Introduction

Text-processing algorithms that annotate main components of a story as a human reader would do, are presently in great need of corpora and well-agreed annotation schemes. Although the task of processing story-lines is yet at an early stage of research, it is studied from fundamental and practical sides in the following areas:

1. cognitive linguistics, to model readers' (Gavins, 2000; Whiteley, 2010; Whiteley, 2011; Giovanelli, 2011) and movie-watchers' (Lugea, 2013) cognition of stories and narratives;
2. forensics, to present and compare crime evidence (Wang et al., 2016; Somaraki and Xu, 2016; Ho et al., 2018a; Ho et al., 2018b);
3. philology, to analyze the structure of literary works (Gavins, 2000; Whiteley, 2010; Whiteley, 2011), poetry (Giovanelli, 2011) and movies (Lugea, 2013), political speeches (Gavins, 2007; Browse, 2018) and advertisement (Downing, 2003) as well as legal texts (Wang et al., 2016; Ho et al., 2018a; Ho et al., 2018b) and fan studies;
4. teaching (Giovanelli et al., 2015; Giovanelli and Mason, 2015; Cushing, 2018), to present language material (e.g. grammar) in a clearer and more understandable way;
5. humour analysis (Downing, 2000);
6. computational linguistics (Elson, 2012).

One of the current theories used in computational linguistics to model readers' perception of a text is the Text World Theory (TWT) suggested in the cognitive linguistics (Werth, 1999). This field focuses on mental mechanisms that people employ in their mind when processing or constructing an event. These mechanisms have been called *scripts, schemata, mental models, cognitive models, frames, mental spaces* and *conceptual frames*, as well as simply *worlds* (Gavins, 2007, p. 3). The titular term of TWT — *text worlds* — refers to "a deictic space, defined initially by the discourse itself and specifically by the deictic and referential elements in it" (Werth, 1999, p. 20). This definition heavily relies on the notion of *deixis* which "refers to all those universal features of language which orientate or 'anchor' our utterances in the context of proximity of space [...] and time [...], relative to the speaker's viewpoint" (Wales, 2014).

From the computational point of view, a *text world* can be understood as a stretch of text perceived by a reader as a whole part of a narrative and characterized by the union of world-building elements, i.e. limited by a time period and space boundaries that hold all of its participants (characters). A text consists of text worlds that change when a major shift in place and/or time happens. E.g. the main character enters or leaves the room, time passes so that this passage is verbalized as something significant ("Years passed by..."), etc. A shift in the narrative from one text world to another is indicated by special words or phrases, *world-switches* (Gavins, 2007).

TWT offers a model that generalizes readers' experience in form of world-building elements (characters, time and space) as well as text worlds themselves and switches between them. On the one hand, annotating these elements, readers reflect their subjective perception; hence, text worlds should have fuzzy edges, and the agreement between annotators should be very low. On the other hand, the text is a finished combination of words that was made such on purpose by the author; hence, the text has to have indication of how to be perceived. Unless the author's intent is to confuse the reader and foster multiple interpretations of the story-line (like in postmodernist novels). The latter

is usually not the case in classic examples of narratives like folk tales, love stories, science fiction, etc. That is why they are often used in experiments on text world annotation.

Presently, there are few living projects on annotation of text worlds. None of them have led to creation of a consistent open-access corpus. Also, the annotation schemes and guidelines have not been discussed in much detail to draw enough effort to create such corpora. Beside that, the TWT terminology describes some phenomena that are very TWT-bound. It would be hard for a specialist with a general knowledge of linguistics and literature to understand them without going deeper into the theory. The morphological and syntactical markup, as a rule, does not need such a deep understanding of parts-of-speech as much more effort has been spent on selection of linguistic phenomena to annotate. TWT also includes such intuitively understandable terms as 'character', 'place', 'time', a combination of which, in our opinion, can lead to a more prominent collective effort in annotation of narratives.

In this two-stage research, we propose our own TWT-based annotation scheme and instructions. At the first stage, six readers annotated the science fiction story "We Can Remember It for You Wholesale" by Philip K. Dick. The agreement among them proved to be very low, so, at the second stage, we corrected the guidelines with the purpose to get a higher inter-rater agreement and tested them again on the short story "The Gift of the Magi" by O. Henry (see Section 3.2.). We evaluated agreement among three readers who annotated *the both* novels and found out that it rose to a more reliable level. We will further describe our experiment as follows. First, we give a brief account of how text worlds are annotated in other projects; we then describe our two-stage experiment, the annotation scheme and the guidelines; we conclude about risks and perspectives of creating larger collections of annotated narratives.

2. A Review of Corpora with Annotation of World-Building Elements

We know of the following software previously used to annotate text worlds and their elements:

1. Worldbuilder (Wang et al., 2016).
2. Concordance tools: CLIC (Mahlberg et al., 2016) and WordSkew (Barlow, 2016). These are mainly used to study the author and characters' speech.
3. Visualisation tools: VUE (Kumar, 2007), Epicurus (Somaraki and Xu, 2016).
4. Tools for annotation of semantic roles: UCREL Semantic Annotation System, or USAS[1] (Hardie et al., 2007).

Among others, Worldbuilder is probably the most TWT-grounded annotation service [2]. It was written by a group of linguists from the University of Huddersfield and Sheffield Hallam University (Wang et al., 2016; Ho et al., 2018a; Ho et al., 2018b) in HTML5 and JavaScript. The input is POS-annotated .txt files. The annotation scheme of Worldbuilder includes worlds, world-building elements (people, time, location, etc.), function-advancing propositions and switches; see Figure 1. The tags are assigned to sentences and not words and phrases. Worldbuilder also has instruments to calculate statistics and visualize data. Such visualisation should help annotators to locate "world-building elements" and text events (Wang et al., 2016). Unfortunately, Worldbuilder does not support large text files and has not been updated since around 2017. Later, the same authors used Visual Understanding Environment (VUE) instead of their system for TWT annotation (Ho et al., 2018a; Ho et al., 2018b).

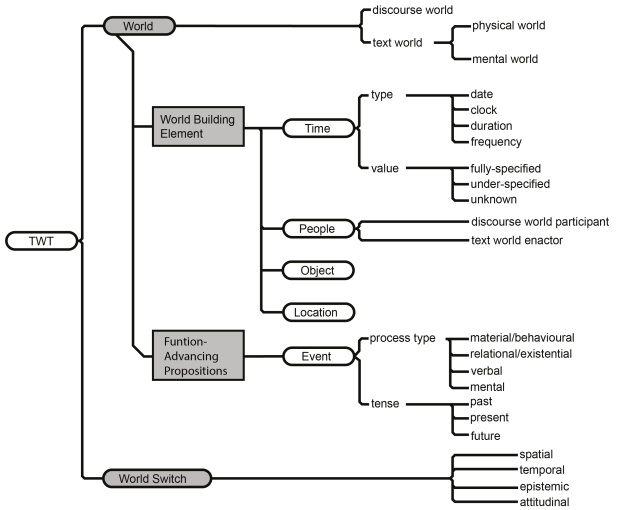

Figure 1: Worldbuilder's annotation scheme.

Somaraki and Xu (2016) used linguistic annotation of a text in XML-format on a corpus of criminal evidence. The annotated files are parsed by the authors' visualization tool "Epicurus" and turned into HTML imagery, allowing for different kinds of statistics and visual analysis. The authors also describe in detail logical structures of texts, existing, as of 2016, tools for their analysis and how TWT improves data extraction in their system. However, Somaraki and Xu (2016) focus more on their visualization tool rather than the issues of linguistic annotation. They provide only glimpses of their XML mark-up which is a summary of text worlds and world-building elements found within a stretch of a text; see Figure 2.

Also, we should mention CLiC (Corpus Linguistics in Cheshire) created by a team led by Prof. Michaela Mahlberg at the University of Nottingham (Mahlberg et al., 2016). It is not closely connected to TWT. However, it includes the web interface to link quotations to a particular character in a literary text and, hence, helps to identify a character as a world-building element.

TWT corpora are quite scarce due to, first, an early stage of practical research that connects TWT to computational linguistics and, second, copyright and security issues regarding contemporary literary works, teaching materials, criminal evidence, etc. For the said reasons, our own corpus of P.K. Dick's text (Mikhalkova et al., 2019) is only available

[1] Available at `http://ucrel.lancs.ac.uk/usas/`.

[2] Available at `http://viv-research.info/TWT/system`.

```
<root xml_tb_version="3.1">
    <twt_node>
        <world id="1" type="physical">
            <time>present time</time>
            <location>at the Police station, Perugia,
            Italy</location>
            <enactor>Amanda</enactor>
            <enactor>Patrick</enactor>
            <event type="material">wish to relate
            spontaneously what happened</event>
            <event type="mental">because these events
            deeply bothered me</event>
            <event type="">I am really afraid of
            Patrick, the African boy who owns the pub
            called "Le Chic" located in Via Alessi
            where I work periodically</event>
        </world>
        <world id="2" type="physical">
            <time>1, 2007, at about 21:00</time>
            <location>at the basketball court of Piazza
            Griamana</location>
            <enactor>Amanda</enactor>
            <enactor>Patrick</enactor>
            <event type="material">sending him a reply
            message ["I will see you"], I met him</event
            >
            <event type="material">we met</event>
        </world>
```

Figure 2: An example of annotation of criminal evidence.

by contacting the corresponding author.

3. An Experiment on Text World Annotation in a Literary Narrative

To annotate text worlds, in our experiment we used XML format and Sublime Text editor[3]. Hence, our annotated texts can be processed by any XML parsing software. In our experiment, we used the ElementTree XML library for Python.

3.1. "We Can Remember It for You Wholesale" by Philip K. Dick

At the first stage of our experiment, several experts with higher education in literature studies and linguistics (including the authors of this work) discussed the main terminology of TWT and suggested a general annotation scheme of world-building elements. They did not propose a set of strict rules for TWT-annotation, but conducted two 90-minute-long seminars on how world-building elements are usually manifested in a literary text with no regard to the to-be-annotated novel. After the discussion, six of these experts annotated world-building elements in a short science fiction story "We Can Remember It for You Wholesale" by P. K. Dick (1993). The novel was in Russian (translated and published in 1993). The annotation scheme included:

- Text world: `<T[No.]><c1>`He`</c1>`awoke—and`</T[No.]>`... [4]
- Switch: `<s>`descended`</s>`.
- Character: `<c[No.]>`Douglas Quail`</c[No.]>`.
- Place: `<p[No.]>`Room D.`</p[No.]>` ("Take `<c1>`me`</c1>` there, to `<p6>`Rekal, Incorporated`</p6>`").
- Time: `<t>`now`</t>` [5].

Table 1 illustrates tags assigned to eight text world elements, in order of appearance in the text. The main character of the novel Douglas Quail and the most mentioned place Mars, that both occurred in the very beginning, got similar tags. However, with every new line differences build up. For example, half of the annotators decided that valleys on Mars are a different location than the planet as a whole.

Text elements	1	2	3	4	5	6
D. Quail	c1	c1	c1	c1	c1	c1
Mars	p1	p1	p1	p1	p1	p1
Government agents	c2	c2	c2	c2	c2	-
high officials	c3	c3	c3	c3	c3	-
a clerk	-	-	-	c4	-	-
Kirsten	c4	c4	c4	c5	c4	c2
Valleys on Mars	-	p2	-	p2	-	p2
Quail's home	p2	p5	p2	-	-	-

Table 1: Tags assigned to the first eight elements in Philip K. Dick's novel. Numbers 1-6 in the head row denote annotators.

Table 2 contains the number of elements annotated in our collection. Although switches should depend on the number of worlds, some annotators marked the return of characters to a place that they had recently left as the same world. Hence, in four cases the number of switches is bigger than that of the worlds. However, in the two other cases the annotators did not mark returns as separate worlds and did not find switches between some worlds or forgot to mark them. There are more inexplicable distributions. Annotator 4 identified the maximum number of text worlds, switches and characters, and Annotator 6, who identified the smallest number of text worlds, tagged the maximum number of space and time indicators.

With all these peculiarities in annotation, the agreement between raters could not be reliable. We have calculated the agreement[6] among the three annotators who also took

An.	T	s	c	p	t	Sum
1	*20*	17	28	21	*61*	147
2	12	9	45	34	41	141
3	*14*	*18*	*37*	*41*	*56*	*166*
4	**56**	**72**	**48**	43	41	260
5	22	*22*	*43*	*37*	77	201
6	8	28	23	**44**	**92**	*195*
m	14, 20	18, 22	37, 43	37, 41	56, 61	166, 195

Table 2: Number of elements annotated in the corpus. An. — Annotator; T, s, c, p, t — tags; m — median. In bold: highest values. Italics: median values.

[3] Available at `https://www.sublimetext.com/`.

[4] "No." indicates a unique number in order of appearance of a world-building element.

[5] As each mention of time was almost always unique, we decided it was no use to enumerate them.

[6] We used NLTK package "metrics.agreement". See (Bird et al., 2009).

part in the second stage of the experiment as concerns their mark-up of the three world-building elements (characters, place and time) using Krippendorf's alpha: $\alpha = 0.580$. We have disregarded the number assigned to elements, considering only the alphabetical designation (c, p, t). According to (Krippendorff, 2004, p. 241), tentative conclusions are acceptable if $\alpha \geq 0.667$. Hence, in the experiment with "We Can Remember..." the agreement among our annotators was either low or hard to evaluate. After analysis of the annotation scheme and annotators' feedback, we introduced some changes and tested them on another text.

3.2. "The Gift of the Magi" by O. Henry

As our aim is to extract text worlds from the text, the text being a fixed composition of certain elements, and not to study the readers' fuzzy perception, we have to strive for a more agreed annotation. Analysis of the first stage of our experiment showed that:

- The longer the text is, the more variations will happen in the order of elements. Hence, numbering the elements is only necessary to collect all stretches of a text referring to the same element, i.e. it functions as entity linking and anaphora resolution. We do not consider numbers in the assessment of inter-rater agreement.
- As changes in text worlds are actually changes in time and/or place, instead of tagging worlds and switches, we should tag these changes.
- The guidelines should specify some details about world-building elements that cause much discrepancy. Consider the following examples:
  - In "He awoke—and wanted Mars. The valleys, he thought" should the valleys be considered a separate place?
  - "Reaching into his coat pocket": Are a pocket, drawer, doorway large enough to be marked as a place? And vice versa, are Earth, Mars, a city too large to mark them?
  - "He dialed his own conapt. And after a pause found himself confronted by a miniature but chillingly realistic image of Kirsten on the small screen." Does telecommunication like video and audio phone calls, television, telepathy create a new place which unites several others?
  - "a cab will leave you off at your conapt": Should automata and organizations be considered as characters if they perform actions, participate in narration?
  - "Quail followed the two technicians": What is the mark-up for groups of characters?
  - "had you picked Pluto or wanted to be Emperor of the Inner Planet Alliance" Should single mentions of places and characters be treated as repetitive instances, e.g. assigned a separate number?
- Time shift is indicated not only by words and phrases, but also by a change in the tense of the verb.

With regard to the first issue above, we decided to focus on a shorter narrative. Besides, we chose a story (in English) the copyright of which allows us to openly publish our annotated text: "The Gift of the Magi" by O. Henry (Henry, 2005) was first published in 1906 (more than 70 years ago). The elements annotated in the second part of our experiment [7] and their tags are as follows:

- Character: <c[No.]>Della</c[No.]>. "So <ts>now</ts> <c4>Della</c4> ' s *VBD* beautiful hair fell *VBD* about..."
- Character, a group: <cg>they</cg>. "in which <cg>they</cg> both took *VBD* a mighty pride..."
- Character, single mention: <cx>the queen of Sheba</cx>. "<ts>Had *VBD* <cx>the queen of Sheba</cx> lived *VBD*</ts> <px>in the flat across the airshaft</px>..."
- Place: <p[No.]>home</p[No.]>. "While <c4>the mistress of <p2>the home</p2></c4> is..."
- Place shift: <ps>out</ps>. "<c4>she</c4> fluttered *VBD* <ps>out</ps>..."
- Place, single mention: <px>in the flat across the airshaft</px>: "<ts>Had *VBD* <cx>the queen of Sheba</cx> lived *VBD*</ts> <px>in the flat across the airshaft</px>..."
- Time: <t>Now</t>. "<t>Now</t>, <t>when</t> the income..."
- Time shift: <ts>the next day</ts>. "And <ts>the next day</ts> would be *VB* <t>Christmas</t>."

First, we have removed switches from the annotation as they are bound to text worlds, being fillers between them: mark-up of a new text world entails mark-up of a switch just before it. Although sometimes if the next world opens right after the previous one, the transition happens without a switch. Second, we have removed text worlds as there was much theoretical confusion about the possibility of going back to a world that closed earlier. We have concluded that whenever there is a change in time and/or place, the new world opens and there is no coming back. From where it follows that there is no need to mark a new text world when this change (a significant shift) has been marked. To mark a change in text worlds, we added two new sub-classes: place and time shifts. We added NLTK-assigned POS-tags (Bird et al., 2009) to verbs which can be now annotated together with the verb as a "time-shift". We also introduced tags for single mentions of characters and places and a tag for groups subsuming several characters or just a general indication of a crowd.

[7] Our small annotated corpus of this story is available at https://github.com/evrog/TextWorlds. The full annotation guidelines are available at https://github.com/evrog/TextWorlds/blob/master/Annotation_guidelines.md.

We wrote an instruction for the new annotation scheme, discussed it at a 90-minute seminar with the three experts who also participated in annotation of "We Can Remember..." (annotators No. 2, 3, 4) [8]. These experts (now No. 1, 2, 3, correspondingly) annotated "The Gifts...".

Text elements	An. 1	An. 2	An. 3
(the) grocer	c1	c1	c1
(the) vegetable man	c2	c2	c2
(the) butcher	c3	c3	c3
one ' s	c4	-	cx
Della	c5	c4	c4
on the shabby little couch	-	-	p1
(the) mistress (of)	c6	c5	c4
home	p1	p1	p2

Table 3: Comparison of tags assigned to first eight elements in O. Henry's novel. An. — Annotator.

Similar to results in Table 1, the annotation of "The Gifts..." (Table 3) demonstrates differences in tagging that start with the fourth element (*one's*) and influence numbering from the very beginning. This issue is bound to such NLP challenges as entity linking and anaphora resolution. For example, Annotator 3 linked the word "mistress" to the previously mentioned character Della, while the other two annotators did not. This choice can be considered as an allowable interpretation of a narrative. However, the annotation of "one's" as a character and "the shabby little couch" as a place contradicts the guidelines, which stresses the necessity of measures that lead to a higher agreement.

As for the number of annotated elements (see Table 4), all the experts are fairly close in their evaluation of the number of characters, places and time. However, annotations of time shifts seem to differ much: 35, 148 and 106 instances in each annotated text. With the new annotation scheme and guidelines, agreement among the three annotators as concerns these three elements rose to: $\alpha = 0.694$. [9] If we exclude time, the agreement is much higher: 0.846, and reaches the acceptable level. [10]

[8] Although, at the first stage, Annotator 3 tagged the median number of all the elements, the other two annotators are closer to other annotators in the number of annotated elements than to each other (see Table 2). Hence, their mark-up can be considered independent.

An.	1	2	3
c	16	26	12
cg	14	21	18
cx	16	0	12
c (total)	46	47	42
p	15	13	14
ps	9	8	5
px	1	0	3
p (total)	25	21	22
t	36	22	39
ts	35	148	106
t (total)	71	170	145
Total No. of elements	142	238	209

Table 4: Number of elements annotated in "Gifts...". An. — Annotator; c, cg, cx, p, ps, px, t, tx — tags.

3.3. Analysis and Errors

We have registered rise in the level of inter-rater agreement at the second stage of our experiment, however, with only two texts annotated it is hard to rank factors that improved annotation. We tend to attribute this rise mainly to the second novel being shorter and more realistic. The readers might have formed a better understanding of what elements are more important in narration and, hence, worth annotating.

As for the removal of such elements as the text world and switch, annotating them might have distracted the readers and caused lower agreement at the first stage. The theoretical premise that we agreed on is that the text world itself is not a frame within which the action takes place, but rather a unity of space, time and characters that is broken by significant changes in time and/or place. Hence, there is no need to annotate text worlds and their borders with a separate tag, thinking of them as of a separate text element.

Also, at the second stage, annotation of time caused much disagreement, `ts` (time shift) being the most 'disagreeable' tag. The second novel contains weak supervision: automatic markers of verb forms that can also be marked as elements of time and time shifts. However, we do not attribute higher agreement to this change, as there is the mentioned big difference in the number of elements annotated as a time shift. Besides, the agreement calculated without them rises as much as to reach the acceptable level. Which is not the case with the first novel where the level also rises, but very slightly (up to $\alpha = 0.597$). Presently, the time shift appears to be the most subjective category as it is based on the individual interpretation of what we earlier called a *significant* change.

Although the two annotated novels differ in size (45,728 symbols in the first one and 11,169 in the second), the number of annotated elements came out to be rather close: cf. the median values of 166 and 195 in the first novel and 142, 238 and 209 in the second. It may be that literary novels generally tend to a certain proportion of characters, locations and text worlds. However, this hypothesis needs a large corpus of novels.

As mentioned earlier, we plan to automate annotation of text worlds with the help of such NLP tools as morphological and syntactic mark-up, anaphora resolution, named entity linking. This seems possible as world-building elements usually include such easily extracted markers as pronouns and names (in case of characters), verb tenses, lexical markers of time and place whose meaning can be extracted from dictionaries or modelled with embeddings. However, the described XML-based annotation rules need refinement. For example, the numbers in tags should be added as labels. Otherwise, the number of XML elements will differ from text to text.

[9] Again, we only consider the type of tags, not the numbering.

[10] According to (Krippendorff, 2004, p. 241), $\alpha \geq 0.800$ is required as an acceptable level.

The remaining disagreement between annotators includes understanding of borders of a phrase that indicates a text element. For example, some annotators do not include *the* in the phrase. As concerns place, one annotator marked the phrase "<p4>out the window</p4>" as a whole, but split "<ps>out</ps> <p11>the door</p11>" into two elements and also tagged the door and the phrase "<p11>near the door</p11>" as a separate place (as regards [No.]). We have also noted that all the annotators used such newly introduced tags as cg and ps, but Annotator 2 did not use cx and px at all.

4. Conclusion

We have described a two-stage experiment on the manual annotation of narrative elements in a literary work. The experiment revealed several issues and possible pitfalls in the task itself, annotation scheme and instruction.

The problematic field of automatic processing of story-lines is quite new. There are many disparate attempts to annotate and automatically process literary works and other types of texts that include narratives. The field is closely connected to entity linking, anaphora resolution, event-detection and several other branches of computational linguistics. However, there are yet few related projects that focus on large text collections and systematically annotate them. We consider that the Text World Theory of cognitive linguistics can be helpful here as it binds readers' perception which is very obscure and flexible to the text which is a ready-made unity of symbols. The text fosters an interpretation. However, the interpretation does not change the text as concerns classic literary art.

In our experiment the annotators were not only readers of the text, but also experts in literary studies and linguistics. Their reading background included analytical research on the categories of the author, character, time and place (chronotope). Hence, what they were reflecting in their annotation is also their theoretical view of the text structure. It is unclear whether annotation of a literary narrative should be based on the expertise or common readers' perception. However, at this early stage of research the annotators' help in shaping guidelines and annotation scheme was very valuable.

As concerns the annotation scheme and guidelines suggested above, we are conducting more experiments and planning to involve not literary experts, but common readers aware of the text elements as much as their school education allows. We plan to focus on short narratives and elaborate on the guidelines, but the general scheme that we introduced in the second part of our experiment seems to be sufficient. It is still unclear whether this approach can be applied to non-literary stories — this is yet to be tested.

5. Acknowledgements

We would like to thank LREC2020 reviewers for their valuable comments and questions that we addressed in the final version of this paper.

The reported study was funded by RFBR according to the research project No. 18-302-00001.

6. Bibliographical References

Barlow, M. (2016). WordSkew. *International Journal of Corpus Linguistics*, 21(1):105–115.

Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. "O'Reilly Media, Inc.".

Browse, S. (2018). *Cognitive Rhetoric: The cognitive poetics of political discourse*. John Benjamins.

Cushing, I. (2018). 'Suddenly, I am part of the poem': texts as worlds, reader-response and grammar in teaching poetry. *English in Education*, 52(1):7–19.

Dick, P. K. (1993). We Can Remember It for You Wholesale. In *Alien*, pages 405–428. GviDo, Riga, Latvia.

Downing, L. H. (2000). How to do things with contradiction: exploring humour in Joseph Heller's "Catch-22". *Atlantis*, pages 107–135.

Downing, L. H. (2003). Text world creation in advertising discourse. *Círculo de lingüística aplicada a la comunicación*, page 3.

Elson, D. K. (2012). Detecting story analogies from annotations of time, action and agency. In *Proceedings of the LREC 2012 Workshop on Computational Models of Narrative, Istanbul, Turkey*.

Gavins, J. (2000). Absurd tricks with bicycle frames in the text world of the third policeman. *Nottingham Linguistic Circular*, 15:17–33.

Gavins, J. (2007). *Text World Theory: An Introduction*. Edinburgh University Press.

Giovanelli, M. and Mason, J. (2015). 'Well I don't feel that': Schemas, worlds and authentic reading in the classroom. *English in Education*, 49(1):41–55.

Giovanelli, M., Macrae, A., Titjen, F., and Cushing, I. (2015). *A/AS Level English Language and Literature for AQA student book*. Cambridge University Press.

Giovanelli, M. (2011). *Dreams, desire and nightmares in the poetry of John Keats: A Text World Theory account*. Ph.D. thesis, University of Nottingham.

Hardie, A., Koller, V., Rayson, P., and Semino, E. (2007). Exploiting a semantic annotation tool for metaphor analysis. In *Corpus Linguistics Conference (CL2007)*.

Henry, O. (2005). *The Gift of the Magi*. Project Gutenberg.

Ho, Y.-F., Lugea, J., McIntyre, D., Xu, Z., and Wang, J. (2018a). Text-world annotation and visualization for crime narrative reconstruction. *Digital Scholarship in the Humanities*.

Ho, Y.-F., Lugea, J., McIntyre, D., Wang, J., and Xu, Z. (2018b). Projecting (un) certainty. *English Text Construction*, 11(2):285–316.

Krippendorff, K. (2004). Content analysis: An introduction to its methodology thousand oaks. *Calif.: Sage*.

Kumar, A. (2007). Visual understanding environment. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*, pages 510–510. ACM.

Lugea, J. (2013). Embedded dialogue and dreams: the worlds and accessibility relations of inception. *Language and Literature*, 22(2):133–153.

Somaraki, V. and Xu, Z. (2016). Epicurus: A platform for the visualisation of forensic documents based on a lin-

guistic approach. *2016 22nd International Conference on Automation and Computing (ICAC)*, pages 95–101.

Wales, K. (2014). *A dictionary of stylistics*. Routledge.

Wang, J., Ho, Y.-F., Xu, Z., McIntyre, D., and Lugea, J. (2016). The visualisation of cognitive structures in forensic statements. In *2016 20th International Conference Information Visualisation (IV)*, pages 106–111. IEEE.

Werth, P. (1999). Text worlds: representing conceptual space in discourse: Textual explorations. *Harlow: Longman*.

Whiteley, S. (2010). *Text world theory and the emotional experience of literary discourse.* Ph.D. thesis, University of Sheffield.

Whiteley, S. (2011). Text world theory, real readers and emotional responses to the remains of the day. *Language and Literature*, 20(1):23–42.

7. Language Resource References

Mahlberg, M., Stockwell, P., Joode, J. d., Smith, C., and O'Donnell, M. B. (2016). CLiC Dickens: novel uses of concordances for the integration of corpus stylistics and cognitive poetics. *Corpora*, 11(3):433–463.

Mikhalkova, E., Protasov, T., Drozdova, A., Bashmakova, A., and Gavin, P. (2019). Towards annotation of text worlds in a literary work. In *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue". Supplementary volume*, volume 18, pages 101–110.