Aligning the Norwegian UD Treebank with Entity and Coreference Information

Tollef Emil Jørgensen, Andre Kåsen
Norwegian University of Science and Technology, National Library of Norway
tollef.jorgensen@ntnu.no, andre.kaasen@gmail.com

Abstract

This paper presents a merged collection of entity and coreference annotated data grounded in the Universal Dependencies (UD) treebanks for the two written forms of Norwegian: Bokmål and Nynorsk. The aligned and converted corpora are the *Norwegian Named Entities* (NorNE) and *Norwegian Anaphora Resolution Corpus* (NARC). While NorNE is aligned with an older version of the treebank, NARC is misaligned and requires extensive transformation from the original annotations to the UD structure and CoNLL-U format. Here, we demonstrate the conversion and alignment processes, along with an analysis of discovered issues and errors in the data, some of which include data split overlaps in the original treebank. These procedures and the developed system may prove helpful for future work on processing and aligning data from universal dependencies. The merged corpora comprise the first Norwegian UD treebank enriched with named entities and coreference information, supporting the standardized format for the CorefUD initiative.

Keywords: Universal Dependencies, Coreference Resolution, Named Entity Recognition

1. Introduction

Resources for the Norwegian language have drastically increased in the last few years. Large text corpora such as the Norwegian Newspapers Corpus (Hofland, 2020) and the Norwegian Colossal Corpus (Kummervold et al., 2022) supported the development of pre-trained transformer-based models like *NB-BERT* (Kummervold et al., 2021) and *NorBERT* (Kutuzov et al., 2021). Moreover, there are task-specific resources, for example document-level and fine-grained sentiment analysis (Velldal et al., 2018; Barnes et al., 2019; Øvrelid et al., 2020), dependency syntax, part-of-speech, morphological features, lemmatization (Solberg et al., 2014; Øvrelid and Hohle, 2016), named entity recognition (Jørgensen et al., 2020), coreference resolution (Mæhlum et al., 2022), and question-answering (Ivanova et al., 2023).

In addition to UD Norwegian Bokmål and UD Norwegian Nynorsk (Solberg et al., 2014), there are two other available treebanks: 1) *Language Infrastructure made Accessible* (LIA) (Øvrelid et al., 2018) and 2) *Norwegian Dialect Corpus* (NDC) (Kåsen et al., 2022). These are based on speech transcripts rather than written sources like the former two. LIA is also converted to UD with the procedure from Øvrelid and Hohle (2016). This work makes use of The preliminary results by Mæhlum et al. (2022) were, before the alignment processes described in this paper, the only baselines for Norwegian coreference resolution on the NARC dataset.[1] This work was motivated by including Norwegian as a supported language in the CorefUD initiative (Nedoluzhko et al., 2022) to unify coreference corpora to a standardized CoNLL-U format. The presented corpora were used for all submissions to the shared task on Multilingual Coreference Resolution (Žabokrtský et al., 2023). The following sections describe related work, an overview of data sources and statistics, conversion, alignment with UD, error analysis, conclusions, and limitations. Code for reproducing experiments is hosted on GitHub.[2]

2. Related Work

NARC is annotated using the BRAT annotation tool (Stenetorp et al., 2012). While conversion scripts are available for the resulting pairs of *.ann* and *.txt* files, such as the official from BRAT[3], none sufficed for the annotation scheme used in NARC, due to cases like discontinuous mentions, validation checks for self-referring clusters and more. We can find an example of BRAT outputs and CoNLL in the Litbank corpus (Bamman et al., 2019). However, the initial annotations used in BRAT are unlike those used in NARC, and there is no available code for parsing.

We set up a conversion pipeline to the commonly used JSON line format for coreference resolution, as popularized by Lee et al. (2018), and finally to

[1] There is, however, an earlier effort for Norwegian coreference found in Borthen (2004), Nøklestad and Johansson (2006), Holen (2007), Johanson and Nøklestad (2008) and Nøklestad (2009).

[2] https://github.com/tollefj/UD-NARC

[3] https://github.com/nlplab/brat/tree/master/tools

LREC-COLING 2024, pages 704–710
20-25 May, 2024.

CoNLL-U[4], conforming to the CorefUD standards and validation requirements (Nedoluzhko et al., 2022). The procedures were validated throughout the alignment process using tools from UD[5] and Udapi (Popel et al., 2017).

3. Data

Three key data sources are involved in this project: UD treebanks for Bokmål and Norwegian, NARC, and NorNE. Following are brief descriptions along with statistics on the merging process.

3.1. Norwegian Dependency Treebank

The current UD treebank is based on NDT – the Norwegian Dependency Treebank (Solberg et al., 2014), one of the first widely used resources for Norwegian, initially developed within an in-house framework corresponding to the theories and practices described and documented by Faarlund et al. (1998). The inventory of part-of-speech tags follows those defined for the Oslo-Bergen tagger (Hagen et al., 2000). The treebank was later converted and included in Universal Dependencies (Øvrelid and Hohle, 2016). It is structured in the CoNLL-U format with unique sentence identifiers but lacks corresponding document-level references, which is the main issue for the alignment of NARC. As of April 2023, the UD treebank for both Bokmål[6] and Nynorsk[7] have been updated to the latest version of UD (version 2.12). Both NARC and NorNE are built upon the contents of NDT, consisting mainly of news texts (85%), government reports, parliamentary transcripts, and blog posts.

3.2. NARC

NARC (Mæhlum et al., 2022) is the first openly available corpus for Norwegian coreference resolution. Its annotations include markables, either as singleton mentions or as referred relational mentions, the latter subdivided into four types: anaphoric, cataphoric, split antecedent, and bridging relations. There are three major issues regarding conversion: 1) NARC is annotated per document, lacking sentence identifiers for direct alignment with UD. 2) It is annotated on a character-level basis, whereas the CoNLL-U format requires word-level annotations. 3) Some documents do not exist in the UD treebanks. We will revisit the issues in Section 4.

[4] https://universaldependencies.org/format.html

[5] https://github.com/UniversalDependencies/tools

[6] https://github.com/UniversalDependencies/UD_Norwegian-Bokmaal#changelog

[7] https://github.com/UniversalDependencies/UD_Norwegian-Nynorsk#changelog

3.3. NorNE

NorNE (Jørgensen et al., 2020) is one of the most extensive corpora for Norwegian named entities, annotated with persons, organizations, locations, geo-political entities, products, and events, in addition to a separate *derived* class for nominals derived from a name. While the NorNE corpus is already an enrichment of the UD treebank, UD has since received updates, mainly regarding corrected token HEADs. The alignment process only included extracting the CoNLL-U *MISC* field (the named entities) from NorNE, placing them with their matching token indices in UD. For an experimental exploration of NorNE, the reader is advised to consult Aasmoe (2019). Earlier efforts for Norwegian concerning NER can be found in both Johannessen et al. (2005), Haaland (2008) and Johansen (2019). The mentioned update of UD ensures that NorNE, through the conversion processes described in this paper, inherits all updated values.

3.4. Statistics

As annotated documents in NARC contain a subset of the existing UD documents, there is a noticeable information loss. Table 1 shows the information loss per category in the aligned corpus. Upon inspection, some lost data were unrelated terms preceding the document, likely sourced from metadata and related articles. We remind the reader that the corpus contains 85% news texts, which often include topics, categories, and other text that may not be related to the article's main body. As such, the raw numbers may not represent an equal loss regarding usability and realistic use cases.

Property	Bokmål	Nynorsk	Total
Sentences	789 (4.8%)	281 (2.2%)	1,070
Tokens	13,510 (5.2%)	6,562 (3.1%)	20,073
Markables	2,410 (4.4%)	1,071 (2.3%)	3,483
Mentions	3,582 (4.6%)	1,522 (2.4%)	5,104
SplitAnte	6 (4.3%)	1 (1.2%)	7
Bridging	35 (3.4%)	27 (3.1%)	62

Table 1: Information loss during the alignment of NARC. SplitAnte: Split antecedent clusters. Bridging: Bridging clusters.

Complete statistics on the number of sentences, tokens, and more can be seen in Tables 2 and 3. All numbers are extracted using Udapi (Popel et al., 2017), both its command-line tool and the Python integration[8], using the`corefud.MiscStats` and `corefud.Stats` modules. The *NARC*-column represents converted CoNLL-U formatted NARC, whereas the *Aligned*-column

[8] https://github.com/udapi/udapi-python

represents the aligned train/test/dev splits. While the statistics differ from those presented in the original paper (Mæhlum et al., 2022), the categories are described as follows:

- Markables are all unique entities in the document (including singletons)
- Mentions are all occurrences and references to the markables
- Bridging- and split antecedent clusters refer to the count of grouped clusters of each respective mention type – not the number of relations within each group.

Appendix A includes examples of how these terms are counted.

Bokmål	UD	NorNE	NARC	Aligned
Sentences	20,044	20,045	16,461	15,672
Tokens	310,221	310,222	257,646	244,136
Entities	-	20,134	-	16,271
Markables	-	-	55,225	52,815
Mentions	-	-	77,565	73,983
SplitAnte	-	-	140	134
Bridging	-	-	1,060	1,025

Table 2: Statistics of the Bokmål corpora

Nynorsk	UD	NorNE	NARC	Aligned
Sentences	17,575	17,575	12,762	12,481
Tokens	301,353	301,353	213,222	206,660
Entities	-	20,087	-	15,520
Markables	-	-	45,918	44,847
Mentions	-	-	63,137	61,615
SplitAnte	-	-	81	80
Bridging	-	-	868	841

Table 3: Statistics of the Nynorsk corpora

4. Coreference Conversion and Alignment

The initial part of aligning NARC is converting the original annotation files (*.ann/.txt* pairs) to the CoNLL-U format. A natural step along the way was to parse these files into the JSON line format with sentence, token, and clustering information. The JSON line files are converted to CoNLL-U and aligned with the UD treebanks.
The steps involved are:

4.1. Ann to JSON conversion

(a) Extract markables and mentions, bridging and split antecedents, group discontinuous mentions

(b) Find connected clusters by building a graph of coreference links

(c) Map character-based indices to word indices

(d) Restructure word-indexed markables and clusters into a JSON line (one .jsonl per .ann)

4.2. JSON to CoNLL-U conversion

(a) Adjust markables spanning tokens not in their equivalent UD spans

(b) Iteratively add markables and mention clusters token-wise, ensuring correct ordering of multi-entity spans according to UD standards.[9]

(c) Restructure according to the CoNLL-U format guidelines, populating the MISC column, leaving out empty fields to be filled by the UD treebank.

4.3. NARC to UD alignment

A highly compressed overview of the alignment process can be described as follows:

(a) Map UD sentence text $\rightarrow$ UD index

(b) Map UD index $\rightarrow$ train/test/dev split

(c) Process NARC documents and extract UD index candidate sentences (one-to-many)

(d) For every sentence with multiple candidates, extract its sentence identifiers in both NARC (N) and UD (U) and build a cost matrix based on the distances to neighboring indices:

$$C_{i,j} = \text{sent\_to\_UD\_dist\_score}(N_i, U_j)$$

We then disambiguate by minimizing sentence distances by solving the linear assignment problem for C (Jonker and Volgenant, 1988).

(e) Verify whether a sentence index is part of more than one UD split. If so, discard the document.

5. Analysis

We discovered several issues and error patterns throughout the conversion and alignment processes, some of which are already mentioned in the steps above. The following error analysis documents problems with the current treebanks and annotated corpora. The developed system may aid future alignment tasks in detecting errors, especially if one has a corpus managed and annotated by multiple parties.

5.1. Sentence mismatch and tokenization issues

A typical error in NARC is an inserted pipe character '|' preceding commas and following sentences,

[9] Validation code from the official UD GitHub repository: https://github.com/UniversalDependencies/tools/blob/master/validate.py#L2190

which is not the case for its alignments in UD data. The extra character is often included in involved markable spans, and its end-index must be decremented accordingly. A total of 2057 spans were corrected for 561 documents. Another issue is two aligned sentences having different tokens (see Table 4). In this case, we map 1:1 sentences to the UD tokens. In the same analysis, four documents in NARC Bokmål (klassekampen_{01,02,03,04}) were not found in UD Bokmål, but had matches in UD Nynorsk and should thus be moved.

NARC sentence	UD sentence
Illustrasjonsfoto .	Illustrasjonsfoto
Illustrasjonsfoto |	Illustrasjonsfoto
Illustrasjonsofoto |	Illustrasjonsfoto .
Nei !	- Nei ?
Nei !	- Nei .
- Ja .	Ja .

Table 4: Examples of tokenization mismatch

5.2. Duplicates and multiple sentence matches

Most commonly occurring in dialogue-based texts, we may observe recurring sentences like "illustrasjonsfoto" (illustration photo), "les også" (read also), interjections, and entity names included multiple times throughout a document. Pure string matching would fail in these cases, such as in the following example, where two people (*Elling* and *Espen*) have several mentions in a dialogue setting. The numbers are sentence indexes where the sentence itself is either *Elling* or *Espen*.

```
'Elling': [15, 26, 41, 56, 63, 79, 87,
    97, 103, 108, 114, 119],
'Espen': [33, 45, 65, 74, 91, 99, 106,
    110, 117]
```

Example 1: Elling and Espen mentioned in a dialogue setting (doc: kknn~20030124-27894)

There are, in total, 597 ambiguous sentences across 234 documents. These are resolved by the sentence disambiguation process in step (d) above.

5.3. Lemma injection

In rare cases, sentences have no symmetric match (even after preprocessing for tokenization issues) in both NARC and UD. Two of these were found to have a lemma injected in place of their original entry.

1. vtbnn~20090625-4275, sentence 23. "kostar **vi** mykje" (costs we a lot) where **vi** (we) is **oss** (us) in UD Nynorsk test, ID 017342.
2. firdann~20100305-5007021, sentence 15. "ordførar" (mayor) is "ordføraren" (the mayor) in UD Nynorsk train, ID 005311.

vtbnn~20031111-1592 has a unique error, where the conjunction "at" (that) is in place of the adposition "ved" (by), token 26 of UD Nynorsk train, ID 012440.

5.4. Data split overlap

Eleven documents were found to span train, test, and dev splits in the original treebanks (6 for Bokmål, 5 for Nynorsk). Although comprising one coherent text, these documents have two parts (with no logical separation), each in a different split in UD. This was perhaps not an issue for the original NDT/UD and NorNE datasets, but a problem occurs for coreference resolution concerning relational dependencies across sentence spans. The suggested correction is to update the original treebanks to contain the entire document. Details are found below in Table 5.

Document	Train	Test	Dev
ap~20081210-2445517 (BM)		✓	✓
ap~20091016-3323000 (BM)	✓		✓
bt~BT-20120916-2765289b (BM)		✓	✓
db~20081128-3858534b (BM)		✓	✓
kk~20110829-59221 (BM)		✓	✓
vg~VG-20121219-10048819 (NN)		✓	✓
firdann~20100118-4812178 (NN)		✓	✓
firdann~20110916-5739806 (NN)	✓		✓
kknn~20030804-23304 (NN)		✓	✓
vtbnn~20070403-3233 (NN)	✓		✓
vtbnn~20090625-4275 (NN)		✓	✓

Table 5: Documents with parts corresponding to multiple data splits in the Universal Dependencies treebanks.

6. Conclusions

We have presented the merging and alignment of NARC, NorNE, and UD for Norwegian Bokmål and Nynorsk, along with statistics of the final corpora. The processes are fully modular; any updates to external data will be properly aligned with their closest match in UD. With the developed system supporting the conversion of BRAT annotation files and the alignment of treebanks, we have been able to maximize the included data throughout the merging process. Future work involves 1) correcting the data split overlaps in UD and 2) adjusting the NARC annotation files according to the findings here to avoid future errors. Finally, we hope the developed system can aid future endeavors for processing UD corpora.

7. Limitations

While the system may be applied to other UD projects, e.g., expansion and alignment, task-specific details must be customized in the pipeline for this to work. Furthermore, there are likely more

UD alignment errors to uncover for data sources besides those described here.

8. Ethics Statement

Working with larger corpora often involves personally identifiable information, typically by full names. These have either been removed or appropriately anonymized in referenced data. There are, however, references to specific documents for replication. All data sources have been previously published and are openly available.

9. Acknowledgements

Thanks to Michal Novák and Daniel Zeman for valuable feedback throughout the conversion and alignment process.

10. Bibliographical References

Tobias Langø Aasmoe. 2019. Named entity recognition for norwegian-experiments on the norne corpus. Master's thesis.

Kaja Borthen. 2004. Annotation scheme for bredt. Technical report, Version 1.0. Tech. rep., University of Bergen.

Åsne Haaland. 2008. *A maximum entropy approach to proper name classification for Norwegian*. Ph.D. thesis.

Kristin Hagen, Janne Bondi Johannessen, and A Noklestad. 2000. A constraint-based tagger for norwegian. *ODENSE WORKING PAPERS IN LANGUAGE AND COMMUNICATIONS*, 1:31–48.

Gordana Ilić Holen. 2007. Automatic anaphora resolution for norwegian (arn). In *Anaphora: Analysis, Algorithms and Applications: 6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2007, Lagos, Portugal, March 29-30, 2007. Selected Papers 6*, pages 151–166. Springer.

Janne Bondi Johannessen, Kristin Hagen, Åsne Haaland, Andra Björk Jónsdottir, Anders Nøklestad, Dimitris Kokkinakis, Paul Meurer, Eckhard Bick, and Dorte Haltrup. 2005. Named entity recognition for the mainland scandinavian languages. *Literary and Linguistic Computing*, 20(1):91–102.

Bjarte Johansen. 2019. Named-entity recognition for norwegian. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 222–231.

Christer Johanson and Anders Nøklestad. 2008. Improving an anaphora resolution system for norwegian. In *Proceedings of the Second Workshop on Anaphora Resolution (WAR II), NEALT*.

Roy Jonker and Ton Volgenant. 1988. A shortest augmenting path algorithm for dense and sparse linear assignment problems. In *DGOR/NSOR: Papers of the 16th Annual Meeting of DGOR in Cooperation with NSOR/Vorträge der 16. Jahrestagung der DGOR zusammen mit der NSOR*, pages 622–622. Springer.

Per E Kummervold, Javier De la Rosa, Freddy Wetjen, and Svein Arne Brygfjeld. 2021. Operationalizing a national digital library: The case for a Norwegian transformer model. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 20–29, Reykjavik, Iceland (Online). Linköping University Electronic Press, Sweden.

Andrey Kutuzov, Jeremy Barnes, Erik Velldal, Lilja Øvrelid, and Stephan Oepen. 2021. Large-scale contextualised language modelling for norwegian.

Kenton Lee, Luheng He, and Luke Zettlemoyer. 2018. Higher-order coreference resolution with coarse-to-fine inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 687–692, New Orleans, Louisiana. Association for Computational Linguistics.

Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, Amir Zeldes, and Daniel Zeman. 2022. CorefUD 1.0: Coreference meets Universal Dependencies. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4859–4872, Marseille, France. European Language Resources Association.

Anders Nøklestad. 2009. *A machine learning approach to anaphora resolution including named entity recognition, pp attachment disambiguation, and animacy detection*. Ph.D. thesis.

Anders Nøklestad and Christer Johansson. 2006. Detecting reference chains in norwegian. In *Proceedings of the 15th Nordic Conference of Computational Linguistics (NODALIDA 2005)*, pages 156–162.

Martin Popel, Zdeněk Žabokrtskỳ, and Martin Vojtek. 2017. Udapi: Universal api for universal

dependencies. In *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, pages 96–101.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. Brat: a web-based tool for nlp-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107.

Zdeněk Žabokrtský, Miloslav Konopik, Anna Nedoluzhko, Michal Novák, Maciej Ogrodniczuk, Martin Popel, Ondrej Prazak, Jakub Sido, and Daniel Zeman. 2023. Findings of the second shared task on multilingual coreference resolution. In *Proceedings of the CRAC 2023 Shared Task on Multilingual Coreference Resolution*, pages 1–18, Singapore. Association for Computational Linguistics.

11. Language Resource References

Bamman, David and Lewke, Olivia and Mansoor, Anya. 2019. *An Annotated Dataset of Coreference in English Literature*. arXiv. [link].

Barnes, Jeremy and Touileb, Samia and Øvrelid, Lilja and Velldal, Erik. 2019. *Lexicon information in neural sentiment analysis: a multi-task learning approach*. Linköping University Electronic Press. [link].

Faarlund, Jan Terje and Lie, Svein and Vannebo, Kjell Ivar. 1998. *Norsk referansegrammatikk*. University of Illinois Press.

Knut Hofland. 2020. *Norsk Aviskorpus*. The National Library of Norway. [link].

Ivanova, Sardana and Andreassen, Fredrik and Jentoft, Matias and Wold, Sondre and Øvrelid, Lilja. 2023. *NorQuAD: Norwegian Question Answering Dataset*. University of Tartu Library. [link].

Jørgensen, Fredrik and Aasmoe, Tobias and Ruud Husevåg, Anne-Stine and Øvrelid, Lilja and Velldal, Erik. 2020. *NorNE: Annotating Named Entities for Norwegian*. European Language Resources Association. [link].

Kåsen, Andre and Hagen, Kristin and Nøklestad, Anders and Priestley, Joel and Solberg, Per Erik and Haug, Dag Trygve Truslew. 2022. *The Norwegian Dialect Corpus Treebank*. European Language Resources Association. [link].

Kummervold, Per and Wetjen, Freddy and de la Rosa, Javier. 2022. *The Norwegian Colossal Corpus: A Text Corpus for Training Large Norwegian Language Models*. European Language Resources Association. [link].

Mæhlum, Petter and Haug, Dag and Jørgensen, Tollef and Kåsen, Andre and Nøklestad, Anders and Rønningstad, Egil and Solberg, Per Erik and Velldal, Erik and Øvrelid, Lilja. 2022. *NARC – Norwegian Anaphora Resolution Corpus*. Association for Computational Linguistics. [link].

Øvrelid, Lilja and Hohle, Petter. 2016. *Universal Dependencies for Norwegian*. European Language Resources Association (ELRA). [link].

Øvrelid, Lilja and Kåsen, Andre and Hagen, Kristin and Nøklestad, Anders and Solberg, Per Erik and Johannessen, Janne Bondi. 2018. *The LIA Treebank of Spoken Norwegian Dialects*. European Language Resources Association (ELRA). [link].

Øvrelid, Lilja and Mæhlum, Petter and Barnes, Jeremy and Velldal, Erik. 2020. *A Fine-grained Sentiment Dataset for Norwegian*. European Language Resources Association. [link].

Solberg, Per Erik and Skjærholt, Arne and Øvrelid, Lilja and Hagen, Kristin and Johannessen, Janne Bondi. 2014. *The Norwegian Dependency Treebank*. European Language Resources Association (ELRA). [link].

Velldal, Erik and Øvrelid, Lilja and Bergem, Eivind Alexander and Stadsnes, Cathrine and Touileb, Samia and Jørgensen, Fredrik. 2018. *NoReC: The Norwegian Review Corpus*. European Language Resources Association (ELRA). [link].

Appendices

A. Coreference Terms

The following examples illustrate how Markables, Mentions, Bridge clusters and Split antecedent clusters are counted. Only Token and MISC columns included.

A.1. Bridge Example

- Markables: 3
- Mentions: 5
- Bridge Clusters: 1

```
Kidnapperne Entity=(1)
kom _
seg Entity=(1)
senere _
unna _
```

```
fordi _
kystvakten Entity=(2)
var _
redd _
de Entity=(1)
ville _
senke _
båten Bridge=2<3|Entity=(3)
. _
```

A.2. Split Antecedent Example

- Markables: 6
- Mentions: 6
- SplitAnte clusters: 1 (only one cluster, but two *mentions* within the cluster)

```
Hennes Entity=(1(2)
fraseparerte _
ektemann SpaceAfter=No|name=O
, _
som _
har _
hentet _
barnet Entity=(3
deres SplitAnte=1<4,2<4|Entity=(4)3)
noen Entity=(5
dager Entity=5)
tidligere SpaceAfter=No|Entity=1)
, _
er _
ikke _
å _
få _
tak Entity=(6
i SpaceAfter=No|Entity=6)
. _
```