NorEventGen: generative event extraction from Norwegian news

Huiling You[1], **Samia Touileb**[2], **Erik Velldal**[1], and **Lilja Øvrelid**[1]
[1]University of Oslo
[2]University of Bergen
{huiliny, erikve, liljao}@ifi.uio.no
samia.touileb@uib.no

Abstract

In this work, we approach event extraction from Norwegian news text using a generation-based approach, which formulates the task as text-to-structure generation. We present experiments assessing the effect of different modeling configurations and provide an analysis of the model predictions and typical system errors. Finally, we apply our system to a large corpus of raw news texts and analyze the resulting distribution of event structures in a fairly representative snap-shot of the Norwegian news landscape.

1 Introduction

Event extraction is a central information extraction task that is aimed at extracting structured representations of real-world event information provided in unstructured texts, commonly expressed in terms of an event trigger and its arguments in the text. While modeling approaches to this task have traditionally been based on sequence-labeling at the token level (Ji and Grishman, 2008; Du and Cardie, 2020; Lin et al., 2020), more recent approaches have allowed for a structure decoding that is less constrained by the exact input string. In particular, the widespread adoption of pre-trained language models based on encoder–decoder architectures have allowed for the formulation of this task as text-to-structure generation (Lu et al., 2021; Wang et al., 2023).

Current event extraction systems typically focus on English, with noteworthy exceptions for other large languages like Chinese and Arabic. This focus is largely due to the availability of manually annotated datasets in these languages (Doddington et al., 2004; Song et al., 2015). The newly released Norwegian event detection dataset EDEN (Touileb et al., 2024) contains manual annotation of news texts from newspapers as well as transcribed news broadcasts and enable large-scale event extraction from Norwegian news sources.

In this paper, we present the NorEventGen system for Norwegian event extraction, which builds on recent developments in the formulation of event extraction as text-to-event structure generation, mapping sentences into linearized event structures. While developing this system using the recently released EDEN dataset, we also evaluate a number of modeling choices related in particular to the format of the input data and the task formulation. Specifically, we analyze the choice of pre-trained Norwegian language model, the localization of event labels using translation and the reliance on explicit trigger word identification for event argument extraction. We provide a detailed analysis of the generated event structures and examine typical errors of our system. Finally, we apply our system to a large collection of news texts from a range of different sources and provide a preliminary analysis of the extracted event structures.

The paper is structured as follows. The next section presents related work, before section 3 presents a system description for our approach. We further describe experimental set-ups in section 4, and discuss the results in section 5. Section 6 presents a use case for our system on a large Norwegian news corpus, before we summarize our finding and contributions section 7.

2 Related work

2.1 Event detection

Event extraction has commonly been approached as a supervised classification task approached through sequence labeling. Classification-based methods typically perform event extraction via several more specific subtasks (trigger detection and classification, argument detection and classification), and either solve these separately with a pipeline-based ap-

Proceedings of the Joint 25th Nordic Conference on Computational Linguistics and 11th Baltic Conference on Human Language Technologies (NoDaLiDa/Baltic-HLT 2025), pages 801–811
March 3-4, 2025

proach (Ji and Grishman, 2008; Li et al., 2013; Liu et al., 2020; Du and Cardie, 2020; Li et al., 2020) or infer these subtasks jointly at the token-level (Yang and Mitchell, 2016; Nguyen et al., 2016; Liu et al., 2018; Wadden et al., 2019; Lin et al., 2020). Moving beyond sequence-labeling, event extraction has also been approached as structured prediction into graph structures (You et al., 2022).

More recently, however, approaches that solve the event extraction task as a generation task have received more attention, mapping a text into a linearized event structure or even a natural language representation of the event. For a recent survey of generative approaches to event extraction, see (Simon et al., 2024). Of particular relevance to our work, however, is the Text2Event system of Lu et al. (2021) which pioneered the text-to-structure approach to event extraction, jointly modeling event detection and argument extraction using a T5 encoder–decoder model (Raffel et al., 2020): Given an input sentence, the model generates a structured representation of an event in the form of an S-expression (i.e., an associative dictionary of labels and values), constrained decoding is enforced to restrict the output vocabulary to valid tokens at each step. The latter is show to be particularly helpful for small training sets. Their ablation study also includes curriculum learning and shows that using natural language tokens for argument roles is preferable to arbitrary tokens.

In an effort to further generalize the text-to-structure approach, Lu et al. (2022) introduce UIE – unified information extraction. UIE formalises a unified "structural extraction language" for encoding different information elements for different IR tasks, and includes IE-specific pre-training that removes the need for constrained decoding. Inspired by instruction tuning, Wang et al. (2023) further build on this to propose InstructUIE, where different IE tasks are reformulated into the task of natural language generation with instructions that include a description of the output format.

2.2 Event datasets

There are several manually annotated datasets for event extraction for English and a few other resource-high languages, such as Arabic and Chinese. The Automatic Content Extraction (ACE) program (Doddington et al., 2004) was an early effort in this space that resulted in several richly annotated datasets including entities, relations, and events for English, Arabic, and Chinese. The English ACE dataset has been widely used for development of event extraction systems and annotates 8 distinct event types (e.g. `Life`, `Conflict`, `Transaction`), along with 33 subtypes (e.g. `Conflict.Attack`) and 22 event-specific subtypes that adorn specific event trigger words in the text along with their event arguments (e.g. `Attacker`, `Agent`, and `Recipient`).

The ERE (Song et al., 2015) dataset, also referred to as Light ERE comprises the same event types and subtypes as ACE. Compared to ACE, ERE adopts a more simplified scheme by merging tags (Aguilar et al., 2014). ERE also comes in a version with richer annotations, dubbed Rich ERE (Song et al., 2015), which is aimed at enabling document-level event co-reference and extends on the ACE event ontology by incorporating 9 event types and 38 event arguments (You et al., 2023).

The MAssive eVENt detection dataset (MAVEN) (Wang et al., 2020), was introduced to cover more general event types, compared to ACE and ERE. It comprised 4,480 Wikipedia documents, containing 168 event types covering 118,732 event mentions. This dataset is only annotated for event types, which are derived from FrameNet (Baker et al., 1998). In MAVEN, first candidate event triggers were semi-automatically identified, followed by an automatic labeling phase, before human annotators provided the final annotations.

3 NorEventGen: text to event records

Our system is built upon Text2Event (Lu et al., 2021), with inspiration from InstructUIE (Wang et al., 2023), as described in Section 2.1 above. Our system differs from Text2Event by applying no constraints on generation and from InstructUIE by using the input sequence only without instructions. This means approaching event extraction as a text-to-structure problem. Given the input sequence $x = x_1, ..., x_{|x|}$, NorEventGen directly generates the event records in a linearized, structured format with a pretrained Norwegian encoder–decoder model.

3.1 Structured event records

Event records are represented in a structure similar to a linearized parse tree, where multiple event records are just sub-trees. As shown in Figure 1, an event record is structured as

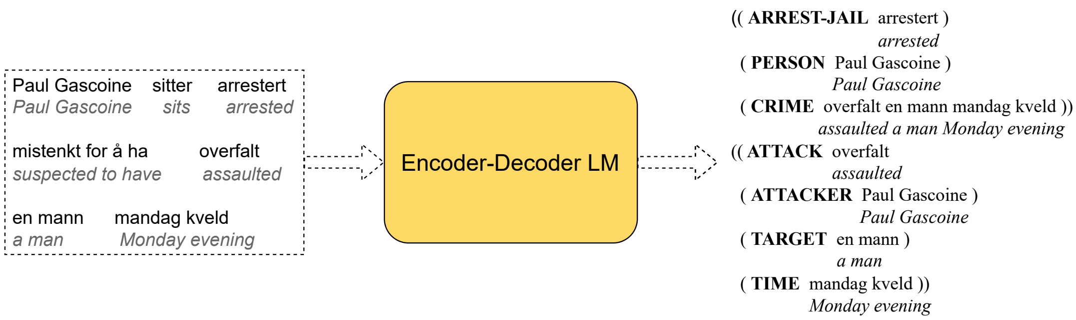


Figure 1: The architecture of NorEventGen. The model takes raw text as input and generates event records in a structured format. In this example, there are two events: (i) an `ARREST-JAIL` with its trigger "arrestert" and arguments `Person` and `Crime` and (ii) an `ATTACK` event with its trigger "overfalt" and associated `Attacker,` `Target` and `Time` arguments.

`((Event_type trigger (arg_role₁ arg₁)(arg_role₂ arg₂))`, and there are two events (`ARREST-JAIL`, `ATTACK`) from the example sentence. For a sentence that does not describe any event, empty event records are simply "()". To differentiate from text snippets and labels for event records, during implementation, the structure indicators "()" are replaced with special tokens `<extra_id_0>` and `<extra_id_1>`, which are trained together with the model. Event records can easily be retrieved via reading structured event records as trees.

3.2 Text to structure framework

With the above mentioned structured representations, NorEventGen generates structured event records via a transformer-based encoder–decoder T5 model (Raffel et al., 2020). For an input sequence $x = x_1, ..., x_{|x|}$, NorEventGen outputs structured event records $y = y_1, ..., y_{|y|}$. First, the raw text sequence x is processed by the encoder into hidden states $\mathbf{H}$:

$$\mathbf{H} = \mathrm{Encoder}(\mathrm{x}_1, ..., \mathrm{x}_{|\mathrm{x}|}) \tag{1}$$

With encoded input tokens, the decoder predicts the output structure token-by-token in an auto-regressive manner. At each generation step i, the i-th token y_i of the output and the decoder hidden state $\mathbf{h}_i^d$ are generated as following:

$$\mathrm{y}_\mathrm{i}, \mathbf{h}_\mathrm{i}^\mathrm{d} = \mathrm{Decoder}([\mathbf{H}; \mathbf{h}_1^\mathrm{d}, ..., \mathbf{h}_{\mathrm{i}-1}^\mathrm{d}]) \tag{2}$$

Decoder$(\cdot)$ predicts the conditional probability $p(y_i|y < i, x)$ for the token y_i. Prediction terminates once the end symbol (`<eos>`) is generated.

Split	#Sents	#Tokens	#Events	#Arguments
Train	20,968	326,145	4,584	7,416
Dev	1,919	35,668	387	626
Test	3,365	57,413	834	1,257

Table 1: Statistics of the Norwegian EDEN dataset.

Compared with some previous studies which treat labels (event ontology) as specific symbols or enforce various constraints during the decoding process, our text-to-structure framework treats labels as natural language tokens and employs greedy decoding during the generation stage. By verbalizing and generating the labels, the model learns event schema knowledge during training.

4 Experiments

In the following, we present the details of our experimental setup, and the specific experiments conducted as evaluation of our model.

4.1 Experimental setup

EDEN The recently released Event DEtection for Norwegian (EDEN) dataset (Touileb et al., 2024) generally adopts the ACE annotation schema and further adapts it to the annotation of news data and transcribed news broadcasts in Norwegian. The event ontology of EDEN defines 34 event types and 28 event argument roles. In total, it contains data from 630 documents containing over 500k tokens and almost 6,000 unique events. Detailed statistics can be found in Table 1.

Pre-trained LMs As mentioned above, we will be using the T5 architecture (Raffel et al., 2020) for the underlying base model. We experiment with two different versions pre-trained for Norwegian, named North-T5[1] and NorT5 (Samuel et al., 2023). Both come in several sizes, and we here use North-T5 base (220 million parameters) and large (770M), and NorT5 base (228M) and large (808M). The main difference is that while the NorT5 models were trained from scratch for Norwegian, the North-T5 models are based on the multilingual mT5 (Xue et al., 2021) (including the tokenizer) with further fine-tuning for Norwegian.

Evaluation Event extraction is evaluated on two key elements: 1) an *event trigger* is correctly predicted if the event type and trigger word(s) match a reference trigger; 2) an *event argument* is correctly predicted if its role type, event type, and argument word(s) match a reference argument. We report F measure (F1) for the following four metrics: Trg-I (trigger identification), Trg-C (trigger classification), Arg-I (argument identification), and Arg-C (argument classification). Since our system directly generates event records, the offset of the generated tokens in the input sequence is unknown; when evaluating trigger and argument identification, we therefore require an exact match towards a substring of the input text.

System comparison We compare our NorEventGen with JSEEGraph (You et al., 2023), a semantic-graph-parsing approach with previously reported results for the EDEN dataset. JSEEGraph differs fundamentally from our NorEventGen, since it is essentially an extract-and-classify approach.

Implementation detail All the reported models were trained on a single node of Nvidia RTX3090 GPU. We adopt AdamW (Loshchilov and Hutter, 2019) to optimize model weights with the learning rate of $6e-6$. We train all the models with batch size of 16 for 25 epochs. All the hyper-parameters are tuned on the development set of EDEN.

4.2 Experiments on label translation

Most event ontologies are formulated in English, including that of EDEN, which adopts the ACE annotation schema in English for the annotation of Norwegian texts. As such, the serialized event structures contain a mixture of Norwegian and English (see Figure 1). When monolingual models are used on non-English datasets, this language mix might affect model performance. To examine the influence of English labels on Norwegian event generation, we translate the ontology (event types and argument roles) into Norwegian, so that both labels and texts are in Norwegian. By comparing the results on original and translated datasets, we can evaluate to what extent the event structure language influences the results.

Model	Trans	Trg-I	Trg-C	Arg-I	Arg-C	PLM
JSEEGraph	—	69.1	**68.0**	52.4	51.5	XLMR-large
NorEventGen	—	61.8	47.4	48.5	47.4	NorT5-base
	✓	69.0	66.0	55.4	52.7	
	—	63.1	61.1	51.8	50.1	NorT5-large
	✓	**69.4**	66.8	**56.8**	**54.9**	
	—	61.3	57.9	44.4	42.0	North-T5-base
	✓	61.7	58.1	45.2	42.9	
	—	66.7	64.2	54.7	52.6	North-T5-large
	✓	67.6	65.7	56.0	54.3	

Table 2: Experimental results on EDEN (F1-score, %). Trg-I and Trg-C correspond to event trigger identification and classification; Arg-I and Arg-C correspond to event argument identification and classification. Trans indicates whether the labels are translated into Norwegian.

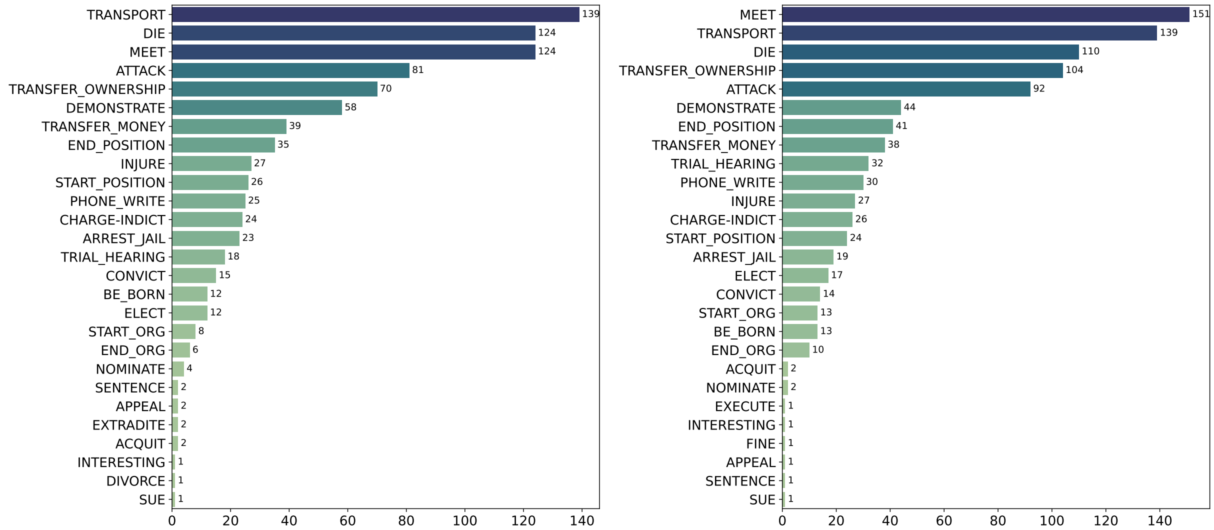
4.3 Experiments on trigger essentiality in structured event generation

As mentioned above, event extraction has traditionally been approached as a token-based classification task, which explicitly anchors the event structures to tokens in the input. This means that the classification of the event type is explicitly related to the event trigger word. For the current approach, this relation is less constrained, and it is therefore possible to evaluate the extent to which event extraction performance relies on the generation of the event triggers. Although the task of event extraction includes both event detection and argument extraction, the evaluation of arguments is exclusive of the trigger words, and is only affected by event type prediction. With our NorEventGen framework, it is convenient to re-structure the output by excluding the trigger text generation, by simply updating the structured event record to $\texttt{((Event\_type (arg\_role}_1\ \texttt{arg}_1\texttt{) (arg\_role}_2\ \texttt{arg}_2\texttt{)))}$. Together with the change of task formulation, we introduce "Evt-C" (event type classification) as the metric to evaluate event type prediction; an event type is correctly predicted if it matches a gold event type. The evaluation metric for event arguments remains the same.

[1]For access and more information about the North-family of models, please see; `https://huggingface.co/north`

PLM	Trans	Top 5 difficult event types
NorT5-base	✓	END-ORG, TRIAL-HEARING, START-POSITION, START-ORG, ELECT
	—	END-ORG, START-ORG, CHARGE-INDICT, CONVICT, TRIAL-HEARING
NorT5-large	✓	START-ORG, TRIAL-HEARING, END-ORG, BE-BORN, TRANSFER-MONEY
	—	START-ORG, TRIAL-HEARING, CONVICT, CHARGE-INDICT, END-ORG
North-T5-base	✓	START-ORG, BE-BORN, END-ORG, TRIAL-HEARING, CHARGE-INDICT
	—	INJURE, END-ORG, TRIAL-HEARING, PHONE-WRITE, START-ORG
North-T5-large	✓	END-ORG, BE-BORN, TRIAL-HEARING, CONVICT, START-ORG
	—	END-ORG, START-ORG, TRIAL-HEARING, CONVICT, BE-BORN

Table 3: Top 5 difficult event types for our models to predict, mesured by F1 scores of Trg-C (event trigger classification). Trans indicates whether the labels are translated into Norwegian.



(a) EDEN test set.

(b) Our best model's predictions.

Figure 2: Event type distributions in EDEN test set versus our best model's predictions.

PLM	Evt-C	Arg-I	Arg-C
NorT5-base	61.3	30.4	28.5
NorT5-large	63.6	21.5	20.6
North-T5-base	54.7	28.0	25.3
North-T5-large	59.0	22.2	21.2

Table 4: Experimental results with trigger text extraction excluded (F1-score, %). "Evt-C" refers to event type classification; Arg-I and Arg-C correspond to event argument identification and classification

5 Results and discussion

We here present the results of NorEventGen on Norwegian event extraction with generative modeling. We first present the overall performance for different model configurations, before discussing the role of label translation and trigger generation, as described above. We then provide a more in-depth analysis of the generated event structures with a specific focus on invalid generations and present an error analysis for the best performing system.

5.1 Overall performance

As shown in Table 2, our results align quite closely with those of JSEEGraph. Compared with previous work, our system shows better performance on event argument extraction; our best-performing system presents an improvement of around 4 percentage points on both argument identification and classification F1 scores. However, on trigger extraction, only large models are on par with previous work.

In terms of the choice of pretrained LMs, NorT5 generates better results than North-T5 across different model sizes, which is especially true for base models. For model size, moving from a base model to a large model, we find that the results improve

PLM	Trans	Event type			Trigger			Argument role			Argument		
		#Invalid	#Gold	#Pred	#Invalid	#Gold	#Pred	#Invalid	#Gold	#Pred	#Invalid	#Gold	#Pred
NorT5-base	—	0	881	614	2	881	614	1	1,524	1010	7	1,524	1,010
	✓	1	881	803	8	881	883	1	1,524	1,374	16	1,524	1,374
NorT5-large	—	0	881	656	0	881	656	1	1,524	1,030	4	1,524	1,030
	✓	2	881	956	10	881	956	1	1,524	1,595	10	1,524	1,595
North-T5-base	—	0	881	856	1	881	856	0	1,524	1,527	4	1,524	1,527
	✓	1	881	939	3	881	939	0	1,524	1,649	5	1,524	1,649
North-T5-large	—	0	881	1,065	5	881	1,065	0	1,524	1,835	10	1,524	1,835
	✓	0	881	997	5	881	997	1	1,524	1,698	3	1,524	1,698

Table 5: Invalid generations. Valid tokens are the event ontology (event types and argument roles) and the input sequence. For each item, the number of invalid instances are listed; "#Gold" and "#Pred" refer to the number of reference and predicted instances. "Trans" refers translated ontology into Norwegian.

considerably.

In terms of event types, as shown in Table 3, difficult event types to predict are largely shared across all of our models, and these event types are somewhat less frequent (as shown in Figure 2a). In particular, three event types (`END-ORG`, `START-ORG`, `TRIAL-HEARING`) are always among the top 5 difficult event types. Under different experimental setups, certain event types can also be difficult to predict; for instance, `INJURE` event even ranks as the most difficult event type for North-T5-base model trained on EDEN with translated labels.

5.2 Label translation

We further find that translating the language of the event ontology is beneficial for all models, in particular for the NorT5 model. The fact that the gain for North-T5 is less could be due to the fact that the model is continually trained from a multilingual T5 model, so it has substantial knowledge of English. In contrast, as a monolingual model trained from scratch for Norwegian, NorT5 is able to benefit more from the translated labels.

5.3 The importance of trigger generation

From Table 4, it is clear that excluding trigger generation (in both training and testing) dramatically affects the performance negatively for both event type prediction and argument extraction, in particular the latter. The scores for argument identification and classification are almost halved across all models. For event type classification, the F1 scores are also considerably lower. To sum up, trigger word(s) generation lies at the core of structured event record generation, since it is the strong indicator of event types, which further affect the evaluation of event arguments.

In terms of pretrained LMs, NorT5 performs better than North-T5 in both base and large variants. Considering the individual subtasks, the large models tend to perform better than the base versions on event type generation, but worse on argument generation, in this particular set-up.

5.4 Analysis of generated event structures

The task of event extraction relies on extraction and classification, namely extracting text spans (event trigger / argument) from the input sequence and labelling (event type / argument role) them. As such, in the context of generation, only tokens from the event ontology and the input sequence are valid generations. Since we do not apply additional decoding constraints during generation, the model is forced to learn the event ontology knowledge and attend to input tokens. Table 5 presents statistics for the generated event type labels, trigger words, argument role labels and argument words for the various model configurations. In general, models trained with NorT5 tend to under-predict, while models trained with North-T5 tend to over-predict. The number of predicted arguments is strongly influenced by the number of predicted event triggers, i.e., more predicted triggers come with more predicted arguments.

When it comes to the generation of invalid event triggers or arguments, as shown in Table 5, such invalid generations are minimal. In terms of event ontology, across all settings, the model rarely generates event type or argument role labels outside the ontology knowledge contained in the training data. There are maximum 2 cases out of hundreds of instances, for both event type and argument role. When it comes to extracting text spans from the input tokens for event triggers and arguments, we find that there are more cases of invalid generations. In general, the number of invalid trigger words is consistently lower than that of invalid arguments for the same model. We also find that the models using label translation seem to generate a higher proportion of invalid arguments than the models trained on non-translated event structures. The last rows of Table 6 provides an example of invalid trigger/argument generations.

5.5 Error analysis

There are various errors made by our model, as summarized in Table 6. Similar to classification-based models, our model predicts either wrong event type or argument role, and can extract wrong text spans for trigger or argument, e.g. in the case of the partially overlapping triggers "statlige tilskudd". Errors are also prevalent in cases of nested event arguments, which is a common challenge for event extraction systems (You et al., 2023). In Table 6, we see that the `Entity` argument of the `End-Position` event is nested within the `Position` argument, a relation that the system does not accurately predict.

Generation-based methods also introduce some new error types, namely invalid generations, as discussed above. These errors commonly occur in generated trigger word(s) or argument word(s) where the model generates words that do not occur in the original input text. We find that our model would generate synonyms of the gold tokens, like the listed example; "frijent" and "frifinn" are synonyms, both meaning "acquit". We also find that it is possible for the model to output just part of a token, like "sør" from "sørover".

6 Use case: event extraction from Norwegian news

One of the main use cases for event extraction systems is the automated analysis of large collections of news texts. An interesting question is whether the distribution of event types in newer news sources is similar to that found for the EDEN dataset (based on the somewhat dated news sources from the Norwegian Dependency Treebank(Øvrelid and Hohle, 2016)). We here apply our best model[2] on a newly collected news corpus dubbed the Norwegian MediaCorpus[3]. The MediaCorpus collects millions of news articles in 2010s from three major media houses in Norway: Amedia, Schibsted, and TV 2. Given its size, the corpus provides a representative sample of the Norwegian news landscape. Table 7 provides detailed statistics of the corpus. We randomly select a smaller set from the entire MediaCorpus to test our model; specifically, we select 200,000 articles from each media house. Detailed statistics are shown in Table 7.

6.1 Event types distribution in MediaCorpus

As shown in Figure 2b, on the test set of EDEN, the event types produced by our model share a similar distribution with the gold event types (Figure 2a). The distribution of predicted event types for the selected subset of MediaCorpus is shown in Figure 3, which also resembles the one on the EDEN test set, with a long tail. Even though the most frequent event type is still `MEET`, the proportion is much larger, and none other event types are on par. As shown in Table 8, among the top 10 trigger words for `MEET` event, apart from explicit words related to meetings, half of them are related to sport matches and Word Cup even ranks as the top 10. The event ontology of EDEN does not cover sports event types, though they are often news-worthy, but those events are predicted into the closet event type in the ontology, namely `MEET`. This phenomenon may indicate that frequent event types reported in the news will still be predicted, though not covered by the ontology itself.

Other frequent event types are `TRANSPORT`, `TRANSFER-OWNERSHIP`, `TRANSFER-MONEY`, `ATTACK`, and `INJURE`. Similarly, the least frequent event types in the MediaCorpus overlap with those in EDEN, such as `SUE`, `ACQUIT`, and `DIVORCE`. In summary, EDEN represents the Norwegian news landscape relatively well, and our NorEventGen model trained on the same dataset has value in real-life application.

6.2 Article tag vs event types

Each article in MediaCorpus has one or more custom tags. These are tags that have been manually assigned by journalists to the article in question. There are 287,687 unique tags in the entire MediaCorpus. Such a large set of article tags can be attributed to the authors' creativity and the lack of a consistent tag set. The most frequent tag `nyheter` ("news") is incredibly vague, and about 20% of the articles would be assigned this tag. Sports related tags are also among the most

[2]Our best model is trained with NorT5-large on EDEN with translated event ontology.

[3]The corpus can be accessed online on: `https://clarino.uib.no/korpuskel/corpora`

<table>
<tr><th>Error type</th><th>Gold</th><th>Pred</th></tr>
<tr><td rowspan="2">Wrong event type</td><td colspan="2">Input: Skole evakuert etter trusler på Internett
School evacuated after threats on the Internet</td></tr>
<tr><td>Event type: TRANSPORT
Trigger: evakuert
Artifact: Skole</td><td>Event type: ATTACK
Trigger: evakuert
Target: Skole</td></tr>
<tr><td rowspan="2">Wrong trigger</td><td colspan="2">Input: nye St. Olavs hospital ikke kan forvente flere statlige tilskudd.
new St. Olavs hospital cannot expect more government grants.</td></tr>
<tr><td>Event type: TRANSFER-MONEY
Trigger: tilskudd</td><td>Event type: TRANSFER-MONEY
Trigger: statlige tilskudd</td></tr>
<tr><td rowspan="2">Missing argument</td><td colspan="2">Input: Ledelsen av EU skifter fortsatt hvert halvår.
The leadership of EU changes still every six months.</td></tr>
<tr><td>Event type: END-POSITION
Trigger: skifter
Position: Ledelsen av EU
Entity: EU</td><td>Event type: END-POSITION
Trigger: skifter
Position: Ledelsen av EU</td></tr>
<tr><td rowspan="2">Invalid trigger</td><td colspan="2">Input: Han tilsto det ene drapet, men ble frikjent for drapet på sløgedal Paulsen.
He confessed the one murder, was acquitted of the murder of Sløgedal Paulsen.</td></tr>
<tr><td>Event type: ACQUIT
Trigger: frijent</td><td>Event type: ACQUIT
Trigger: frifinn</td></tr>
<tr><td rowspan="2">Invalid argument</td><td colspan="2">Input: · · · å selge trålfartøy med konsesjon sørover, mens det er helt kurant å selge andre veien.
· · · to sell trawlers with license in the south, while it is normal to sell the other way</td></tr>
<tr><td>Event type: TRANSFER-OWNERSHIP
Trigger: selge
Artifact: trålfartøy med konsesjon</td><td>Event type: TRANSFER-OWNERSHIP
Trigger: selge
Artifact: trålfartøy med konsesjon
Place: sør</td></tr>
</table>

Table 6: Typical errors made by our best-performing model trained with NorT5-large.

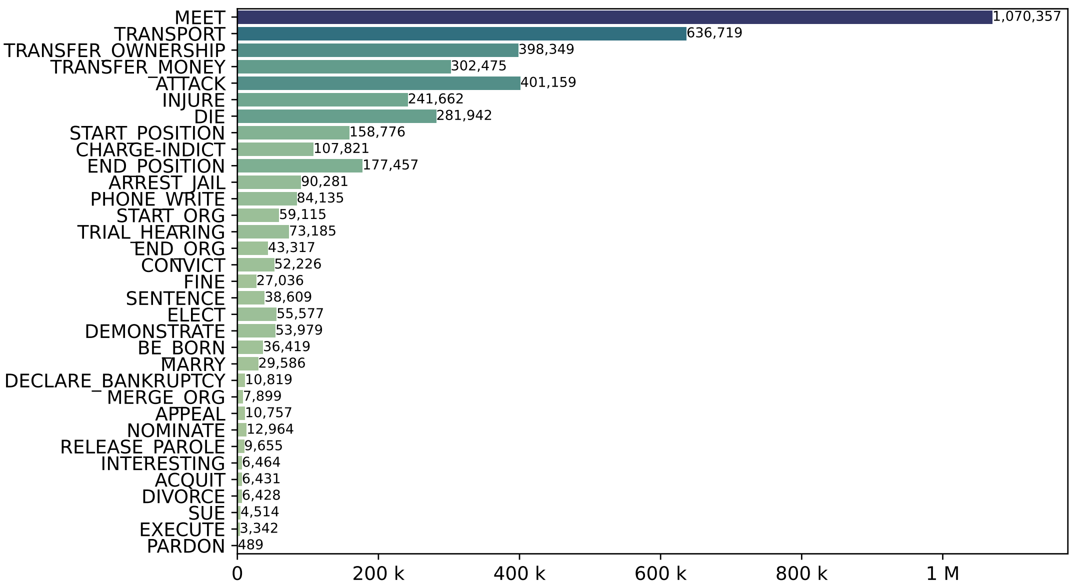


Figure 3: Predicted event types distribution on selected set of MediaCorpus.

frequent tags, and football stands out from other sports as `fotball`("football") is the third most frequent tag. In real life, sports is an important news-worthy topic, but the related event types are not covered in the event ontology of EDEN.

To better evaluate the relationship between

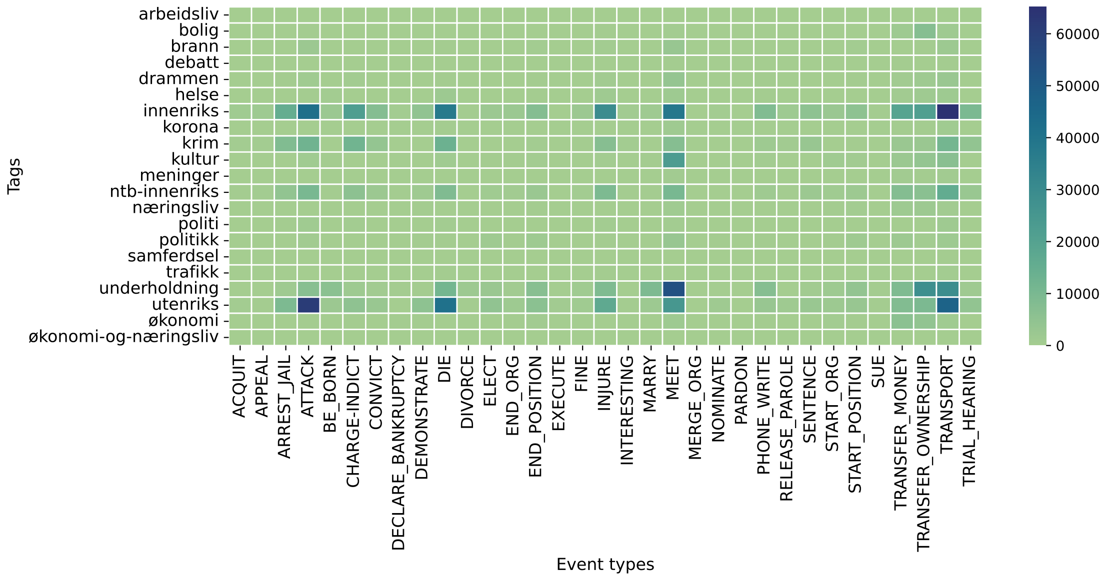


Figure 4: Frequencies of document tag and event type (predictions on selected set of MediaCorpus)

Source	**#Docs**	**#Sents**	**#Tokens**
		Entire corpus	
Amedia	5,263,591	139,482,285	2,218,694,185
Schibsted	2,710,885	67,802,274	1,089,613,721
TV 2	585,772	12,741,165	192,327,865
		Selected set	
Amedia	200,000	3,767,797	58,515,290
Schibsted	200,000	5,572,248	91,012,216
TV 2	200,000	3,914,595	55,298,721

Table 7: Statistics of MediaCorpus.

kampen	match
møte	meeting
kamp	match
kamper	matches
møter	meet
møtet	the meeting
møtte	met
besøk	visit
kampene	the matches
VM	World Cup

Table 8: Top 10 trigger words for `MEET` event in the predictions of the selected MediacCorpus.

article tags and event types, tags similar to `nyheter` and sports-related tags are excluded. The frequencies of article tag vs event type are shown in Figure 4. In general, a strong correlation between article tag and event type is not clear. There are several tags that frequently co-occur with events: `innenriks` ("domestic"), `krim` ("crime"), `utenriks` ("abroad"), and `underholdning` ("entertainment"). These tags often occur together with `ATTACK`, `DIE`, `MEET`, `TRANSPORT`, `TRANSFER-MONEY` and `TRANSFER-OWNERSHIP` events. It is clear that events about violence and economic activities are news-worthy both domestically and abroad.

7 Conclusion

In this paper, we address event extraction from Norwegian news with a generation-based method. Our experiments on the Norwegian EDEN dataset show that our NorEventGen model is able to acquire event ontology knowledge and generate tokens from the input sequence for event triggers and arguments, thus it is not necessary to implement constraints during the generation process. In our experiments, we also find that it is highly beneficial to localize the event ontology to the target language, in our case Norwegian, and using a monolingual Norwegian model is more beneficial. Beyond the EDEN dataset, we extend our system to process a large corpus of raw Norwegian news texts. By applying our model to this broader dataset, we analyze the predicted event distribution, providing insights into the types of events prevalent in Norwegian news. This analysis serves as a snapshot of the Norwegian news landscape and illustrates the potential applications of our approach for large-scale event analysis in less-resourced languages.

Acknowledgments

This work was supported by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the centers for Research-based Innovation scheme, project number 309339.

References

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 45–53, Baltimore, Maryland, USA. Association for Computational Linguistics.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ACE) program-tasks, data, and evaluation. In *International Conference on Language Resources and Evaluation*, volume 2, pages 837–840. Lisbon.

Xinya Du and Claire Cardie. 2020. Event extraction by answering (almost) natural questions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 671–683, Online. Association for Computational Linguistics.

Heng Ji and Ralph Grishman. 2008. Refining event extraction through cross-document inference. In *Proceedings of ACL-08: HLT*, pages 254–262, Columbus, Ohio. Association for Computational Linguistics.

Fayuan Li, Weihua Peng, Yuguang Chen, Quan Wang, Lu Pan, Yajuan Lyu, and Yong Zhu. 2020. Event extraction as multi-turn question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 829–838, Online. Association for Computational Linguistics.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 73–82, Sofia, Bulgaria. Association for Computational Linguistics.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7999–8009, Online. Association for Computational Linguistics.

Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. Event extraction as machine reading comprehension. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1641–1651, Online. Association for Computational Linguistics.

Xiao Liu, Zhunchen Luo, and Heyan Huang. 2018. Jointly multiple events extraction via attention-based graph information aggregation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1247–1256, Brussels, Belgium. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. 2021. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2795–2806, Online. Association for Computational Linguistics.

Yaojie Lu, Qing Liu, Dai Dai, Xinyan Xiao, Hongyu Lin, Xianpei Han, Le Sun, and Hua Wu. 2022. Unified structure generation for universal information extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5755–5772, Dublin, Ireland. Association for Computational Linguistics.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 300–309, San Diego, California. Association for Computational Linguistics.

Lilja Øvrelid and Petter Hohle. 2016. Universal Dependencies for Norwegian. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1579–1585, Portorož, Slovenia. European Language Resources Association (ELRA).

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

David Samuel, Andrey Kutuzov, Samia Touileb, Erik Velldal, Lilja Øvrelid, Egil Rønningstad, Elina Sigdel, and Anna Palatkina. 2023. NorBench – a benchmark for Norwegian language models. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 618–633, Tórshavn, Faroe Islands. University of Tartu Library.

Étienne Simon, Helene Bøsei Olsen, Huiling You, Samia Touileb, Lilja Øvrelid, and Erik Velldal. 2024. Generative Approaches to Event Extraction: Survey and Outlook. In *Proceedings of FuturED 2024: Workshop on the Future of Event Detection*, Miami, USA.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant, and Xiaoyi Ma. 2015. From light to rich ERE: Annotation of entities, relations, and events. In *Proceedings of the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 89–98, Denver, Colorado. Association for Computational Linguistics.

Samia Touileb, Jeanett Murstad, Petter Mæhlum, Lubos Steskal, Lilja Charlotte Storset, Huiling You, and Lilja Øvrelid. 2024. Eden: A dataset for event detection in norwegian news. In *Proceedings of the Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation ((LREC-COLING 2024)*. European Language Resources Association (ELRA).

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5784–5789, Hong Kong, China. Association for Computational Linguistics.

Xiao Wang, Weikang Zhou, Can Zu, Han Xia, Tianze Chen, Yuansen Zhang, Rui Zheng, Junjie Ye, Qi Zhang, Tao Gui, et al. 2023. Instructuie: Multi-task instruction tuning for unified information extraction. *arXiv preprint arXiv:2304.08085*.

Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. Maven: A massive general domain event detection dataset. *arXiv preprint arXiv:2004.13590*.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Bishan Yang and Tom M. Mitchell. 2016. Joint extraction of events and entities within a document context. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 289–299, San Diego, California. Association for Computational Linguistics.

Huiling You, David Samuel, Samia Touileb, and Lilja Øvrelid. 2022. Eventgraph: Event extraction as semantic graph parsing. In *Proceedings of CASE: The 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*.

Huiling You, Lilja Vrelid, and Samia Touileb. 2023. JSEEGraph: Joint structured event extraction as graph parsing. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023)*, pages 115–127, Toronto, Canada. Association for Computational Linguistics.