Abstract

Existing datasets and methods that aim at the identification of time expressions in natural language text do not pay particular attention to expressions that are imprecise and that cannot be easily represented on a timeline. We call these vague time expressions (VTEs). We present an analysis of existing time extraction approaches and steps towards a novel scheme for the annotation of VTEs, developed using a corpus of German news articles. To the best of our knowledge, this work is the first to suggest an extension of the ISO standard TimeML with the goal of enabling the annotation of VTEs. In addition, we present a collection of 339 German VTEs as well as classification experiments on the news corpus with results from 60 up to 77 macro-avg. F1 score.

1 Introduction

Time is critical to the meaning of language, for understanding events, cause-effect relations and narratives, etc. In NLP, temporal expression analysis is a key issue which has been receiving a lot of attention in recent years, especially for English news texts (e.g., UzZaman et al., 2013; Caselli and Vossen, 2017; Strötgen et al., 2018). Existing approaches mostly focus on time expressions which can be more or less easily and specifically represented on a timeline with an accuracy of different granularity levels. Given an expression like at 6 o clock, the hour can be pinpointed, this Sunday or tomorrow refer to a specific day. Such time expressions can be annotated using standardized machine-readable expressions. This process is typically referred to as normalization.

Nevertheless, there is a large proportion of time expressions that are inherently vague – they are neither exact nor precise and, i.e., they cannot be readily normalized. Examples for vague time expressions (VTEs) are in the future or lately. VTEs are typically not taken into account by existing annotation schemes; some simply normalize them as a reference to the past or future. According to Tissot et al. (2019), around 13% of time expressions in news articles can be considered vague. In our German corpus, almost 30% of all time expressions are VTEs (Section 3). The annotation of time expressions in TimeML (Saurí et al., 2006), arguably the most well known scheme, is possible only if we can fully and precisely interpret the expression (Mazur and Dale, 2011). As VTEs cannot be fully and precisely interpreted, we are unable to represent them using TimeML, which is why an annotation scheme needs to be developed that is able to capture VTEs. Given that a large number of time expressions tend to be overlooked or oversimplified, our goal is the annotation and normalization of VTEs by extending TimeML; we concentrate on German documents and the German extension for TimeML. According to our research, no corpora exist that cover VTEs in a substantial way, neither for English nor for German.

While VTEs cannot be easily normalized and expressed on a timeline (Schilder and Habel, 2001), we argue that, based on our analysis, it is possible to describe their meaning systematically by their semantic and syntactic properties, which enables us to normalize VTEs more precisely than existing annotation schemes. Our main contributions are:

• We provide an overview of schemes for time expressions and their ability to express VTEs.
• We present a list of over 300 categorized German VTEs.1
• Building upon Tissot et al. (2019) and Mazur and Dale (2011), we develop a study about

1The full list is available under: https://live.european-language-grid.eu/catalogue/lcr/7975 (last access: 2021-08-13)
the possibilities of normalizing and classifying VTEs by expressing the closest or most precise meaning. To the best of our knowledge, we are the first to present such a study on VTEs for German. Our methods can be adapted to other languages.

- We present an annotated sample dataset and preliminary classification experiments.

2 Background and Related Work

Our approach is primarily based on the categorization of precise time expressions according to TimeML and LTIMEX (Section 2.1) as well as on the categorization of VTEs provided by Tissot et al. (2019) (Section 2.2). Channell (1983) and Dinu et al. (2017) describe approaches on vague expressions in domains other than time.

2.1 Categories of Precise Time Expressions

The ISO standard for the annotation of time expressions is TimeML (Pustejovsky et al., 2010). Temporal expressions are marked up using TimeML’s TIMEX3 tag to capture their meaning. Important attributes of this tag are type and value: Type records whether the expression is a duration, a point in time (either a specific date, or a time of the day) or a set of points in time (Sauri et al., 2006). The type of an expression determines how the expression is normalised in the value attribute. Temporal expressions with a modifier that cannot be expressed using the value attribute, e.g., “in about 3 days” are handled by the optional attribute mod, which was adapted from TIDES (Ferro et al., 2001).

With LTIMEX, Mazur and Dale (2011) attempted to modularise the normalisation process of a temporal expression. This annotation scheme extends TIDES to capture partial meanings of time expressions. It differentiates between local and global meaning. The local meaning is the same for each occurrence of a word and determining this meaning requires no contextual information. For example, “yesterday” has the local meaning “the day before today” (Mazur and Dale, 2011). The global meaning, on the other hand, is gained through the context about the utterance time of the expression and, based on the local meaning, the date of what is referred to as “yesterday” can be concluded.

LTIMEX distinguishes 12 categories of time expressions. Similar to TimeML, they include points, durations and sets. Additionally, offsets are functions that normalize a time expression relative to the document creation time (dct) or a given reference time (ref). An example for an offset would be “in 3 days” meaning “3 days after the dct”. Another class is ordinarily specified, which are expressions based on numbers, like “the first|last|every second Monday in July”. The categories that indicate VTEs are modified point and modified duration. The annotation scheme provides no other methods for normalising VTEs.

2.2 Categories of Vague Time Expressions

Using the definition of the word vague from Devos (2003), a time expression is vague, when it is not clear which date the expression refers to or which dates limit the referenced time period. Even an expression like “in 2010” can be used without referring to the whole year 2010 but to a specific, yet unknown point or period (Strötgen, 2015). Only few approaches or schemes deal with VTEs in more detail (e.g., Devos et al., 1994; Rong et al., 2017), from which Tissot et al. are the only ones to present a classification especially for VTEs (Tissot et al., 2019).2 The six categories are based on an evaluation of clinical corpora. The first category, present reference, includes temporal references related to the present, such as “now”, “recently” and “currently”. Modified precise time expressions, like “in approximately 10 days”, belong to the category modified value. Imprecise value refers to expressions built up around an imprecise period of time, such as “a few days” or “several weeks”. This category also contains expressions with an indefinite period of time, in which the granularity is usually represented in plural form without numeric values.3 An example would be “years” in “It took years to finish the job.”. The fourth category, range of values, describes time spans defined by limits, such as “every 3 to 4 months”. Partial period covers time spans that are part of a larger time frame, such as “mid-January”. The last class, generic expressions, includes a general period or duration, like “this time” or “at the same time”. Although these categories are relevant, Tissot et al. do not present further methods for normalization. We adapt and extend the classes for our own categorization (Sec-

---

2 Instead of “vague”, Tissot et al. use the term “imprecise”.
3 The granularity describes how precise a time expression is. It can, for example, be one of the values millennium, century, decade, year, month, day, hour, minute or second (Caselli and Sprugnoli, 2015).
3 Dataset

Part of the dataset we used for our primary annotation experiments including the annotations for VTEs is taken from KRAUTS (Strötgen et al., 2018). KRAUTS is a German corpus consisting of 50 articles from the newspaper Die Zeit, annotated using the TimeML guidelines by Minard et al. (2017). We also incorporate 1037 documents we call, for the remainder of this work, Short_News because they consist of short articles from various news media.\(^4\) The Short_News articles consist, on average, of fewer tokens than the articles in KRAUTS and only briefly and factually state the most important facts. One news document can contain several short articles on various topics while in KRAUTS, one document always covers exactly one topic. The KRAUTS articles are more detailed and explain the background of events or an opinion on a subject. The corpus includes comments, opinion pieces, reports, interviews, reviews, and excerpts from a book or film but also fictional types of texts, such as short stories and poems.

To narrow down the size of the dataset for the scope of this work, we selected 100 articles from Short_News with about every tenth text being chosen. We filtered near-duplicate articles which left us with 69 documents with 96,721 tokens in total.\(^5\) Table 1 shows key corpus statistics.

| Articles       | Sentences per file in total | Tokens per file in total |
|----------------|----------------------------|--------------------------|
| KRAUTS         | 52.2                       | 2,609                    | 1,005.0                  | 50,250                  |
| Short_News     | 22.3                       | 1,536                    | 673.5                    | 46,471                  |
| Both           | 37.2                       | 4,145                    | 839.2                    | 96,721                  |

Table 1: Key data set statistics (after reducing Short_News to 100 articles).

4 Data Annotation

The category adaptations and resulting TimeML extensions are described in Sections 4.1 and 4.2, respectively. Based on a list of German VTE (Section 4.3) we developed the attribute meaning which enables a more precise normalization of VTEs. The categories postulated by Tissot et al. (2019) indicate a variation in the level of precision and vagueness, which we utilize for our normalization of VTEs (Section 4.4). While the normalization was adapted to each of these categories, the categories themselves do not appear in the annotation. Furthermore, Section 4.5 describes additional vague time expressions, while Section 4.6 presents statistics about the annotated dataset.

4.1 Inferring a Classification

Tissot et al. (2019) do not describe the category modified value in much detail, which is why we interpret it as described in TimeML, where a (precise) temporal expression is modified by a modifier. Since partial period also includes modified time expressions, like “mid-January”, we decided to merge it into the modified value category. As mentioned, time expressions of the modified value category are precise time expressions that are made vague using a modifier, such as “approximately 10 days”. Here, the intended time span can be narrowed down to a few days. The category range of values contains expressions that give specific boundaries, like “in 2-3 days”. The exact point in time or time period is unknown yet somewhere in between. Expressions in the imprecise value category still reveal their granularity. For example, “in a few days” most likely refers to days after the utterance time and not weeks or months. Nevertheless, Tissot et al. (2019) do not distinguish between points in time, time periods or sets. We, therefore, took the categories modified value, imprecise value and range of values and subdivided them respectively for normalization according to the TimeML types date, time, duration and set. The categories that could not be subdivided according to TimeML types (present reference and generic expression) were not included in our annotation scheme.

4.2 TimeML Extensions

Our annotation scheme for VTEs builds upon TimeML and ensures compatibility. We realise TimeML-compliant normalization of precise time expressions by keeping the attribute value and by adding the new attribute meaning to cover the interpretation of VTEs.

4.2.1 Normalising the Attribute “meaning”

The mechanisms for capturing normalizations in the attribute meaning is based on the type of the VTE. We used the formalizations for the offset in...
the LTIMEX-scheme of Mazur and Dale (2011) as a starting point. Similar to Mazur and Dale (2011), a + or a − in the normalization means that the expression describes a point in time that lies before or after a reference point. For example, “+0000-00-0X” means “in einigen Tagen” (in a few days). The + indicates that the referenced point in time is after the dct, while “0000-00-0X” represents the number of years, month and days that are between the dct and the described point in time. The “X” placeholder indicates, in this case, a number of days between 1 and 9. In contrast to Mazur and Dale (2011), we included < and > as comparison operators. A_1 < A_2 means that an expression A_1 is temporally before an expression A_2, and A_1 ≤ A_2, that A_1 happens before or at the same time as A_2.

One of the placeholders A_1 or A_2 can be replaced by dct or ref to refer to the document creation time (dct) or another reference point (ref). While dct is important for factual text types (e.g., news articles), ref is helpful especially for narrative texts where the utterance time of the text is not necessarily the time when the document was created. P stands for period and indicates a normalization of type duration, Y stands for years and can be replaced by D (days), M (months), DE (decades) or by a leading T (time) and h (hours), m (minutes) or s (seconds) (Saurí et al., 2006). We expand the use of X in our scheme so that it can be used to indicate one or more decimal places. Therefore, in addition to “PXDE” (representing at most nine years), it is now also possible to use “PXXY” (representing at most 99 years). The largest range probable in a given context should always be used. We did not further modify the attribute value of TIMEX3.

The value of the attribute meaning resembles a function. It can take one of two forms. In the first form, it contains Z as a symbol for the time expression, see the examples in the rows for date, time, duration and set in the range of values category in Table 2. The second form is used to describe a expression that refers to a point that lies a specific time before or after ref or dct. In this, a number of units is subtracted from or added to dct or ref to represent a specific point in time. The number of units is specified in ISO format: YYYY − MM − [WW]DD − Thh : mm : ss, where zero represents an empty position and X represents an unknown position. A digit can be omitted if it is zero and followed only by zeros. For the expression “vor Jahrzehnten” (decades ago) in example (1) we derive the meaning dct − 00XX, which indicates that the expressed point in time must lie a two-digit number of years before the dct. The hundreds and thousands digits are 0. All units more specific than the year are left out. Appendix A contains more examples.

(1) Er ist vor Jahrzehnten ausgewandert.
He is before decades emigrated.
‘He emigrated decades ago.’

In TimeML (Saurí et al., 2006), only “Jahrzehnten” (decades) would be marked as a time expression while ignoring the preposition “vor”. We, however, consider prepositions as well as adverbs to be an inherent part of the time expression because they can convert one type of time expression into another one. In this example, the preposition converts a duration into a point in time, so that the value changes from “PXDE” to “PAST_REF”. The same applies to reverse cases, when a preposition or an adverb converts a point in time into a duration.

4.2.2 Additional Adaptations

In addition, we circumvent empty tags by specifying values directly in the appropriate attributes instead of creating and linking another point in time with the help of references. For example, instead of creating two empty tags to represent the begin and end points of a duration, these times are directly annotated in the begin and end fields of an expression of type duration. In the sentence “Die Expedition beginnt am 4. April 2022 und dauert ungefähr 10 Tage” (The expedition starts at April 4, 2022 and takes about 10 days) the start point will be “2022-04-04” and the end will be “2022-04-14” with the mod-value “approx” (see also row 5 in the modified value section of Table 2). In addition to numerical values, the normalization of the label set for irregular or unclear intervals can also include the values low, normal, high, increasing or decreasing in the attribute freq. For example, “Ich treibe selten Sport” (I rarely do sports), yields the value “low” for the attribute freq.

We foresee the attribute vague to distinguish VTEs from precise time expressions. It is true whenever a time expression cannot be normalized to an exact value, i.e., whenever value contains the placeholder X or is past_ref or future_ref. It is also true when a modifier is used. Every example in table 2 has this attribute set to "true".
4.3 Vague Time Expressions in German

We collected 338 German VTEs in total. The seed entries of our inventory were based on an analysis of the KRAUTS corpus (Strötgen et al., 2018) as well as various brainstorming sessions among the authors. The list was expanded using synonyms found in the DWDS and Duden online search. Similar expressions are summarised using placeholders. In “Anfang Monat” (begin month), “Monat” should be replaced by a specific month, e.g., “Januar” (January). Granularity expressions like days, weeks, etc. are indicated by a capitalised “G” (e.g., “in einigen G”) (in a few G)). Additionally, numbers are represented by an x. For example “in x G” can be replaced by “in 3 Tagen” (in 3 days).

We structure the time expressions into different categories (see Section 4.1). Not all expressions were assigned to a category since, as pointed out in Section 4.5, there are other types of time expressions which are challenging to describe with the given categories. Our list served as an initial basis for the development of the normalization approach. Appendix B shows an excerpt of the full list.

4.4 Description of the Classes

Table 2 illustrates our classification (Section 4.1) including the categories modified value, imprecise value and range of values and additional sub-categories according to the four types used in TimeML. We included two additional types based on the offset category (Mazur and Dale, 2011). These are offset-like date and offset-like time which use a time interval and a reference point to refer to a date or a time respectively. While offset-like time can take a time granularity like seconds, minutes or hours (e.g., “in 5 hours”) an offset-like date can take any other granularity like days, month or years (e.g., “in 5 days”). Additional example expressions and their normalizations are shown in Table 2. Although the offset types are eventually converted to dates or times when the local representation of TIMEX is turned into a global annotation, we listed them separately to show the semantic difference between points and offsets. While point expressions, like “Mitte Januar” (mid January), consist mostly of nouns, offset expressions seem to always contain a preposition, e.g., “in” (in) in the expression “in 6 Tagen” (in 6 days) or “vor” (ago) in “vor 6 Tagen” (6 days ago) or an adverb “danach” (after that) in the expression “10 Tage danach” (10 days after that).

4.5 Other Vague Time Expressions

There are types of time expressions that are difficult to classify in the way described above. In contrast to the examples in Table 2, expressions such as “bald” (soon) or “kurz danach” (shortly afterwards) do not inherently indicate a specific granularity. For example, “früher” (back then) in “früher war alles besser.” (Everything was better in the good old days.) does not refer to a duration with a certain start and end point, but to an unspecified span in the speaker’s past. It is probably valid to assume that a period of time is meant that is at least a decade in the past (depending on the age of the speaker), so that the granularity can be narrowed down to "dct - 00XX". The example shows that for the annotation of VTEs, world knowledge as

---

6https://www.dwds.de (last access: 2021-05-30)
7https://www.duden.de (last access: 2021-05-30)
8https://live.europarl.europa.eu/catalogue/lct/7975 (last access: 2021-08-13)
| TimeML-Type Subcategory | Example (English) | Example (German) | Normalization |
|-------------------------|------------------|-----------------|---------------|
| VTE category: modified value | date | mid-January* | Mitte Januar | mod="mid" value="xxxx-01" |
|                          | offset-like date | after about 10 days | nach ungefähr 10 Tagen | mod="approx" value="2021-05-11" |
|                          | time | around 1 p.m. | ungefähr 13 Uhr | mod="approx" value="2021-05-01-T13" |
|                          | offset-like time | after about 10 hours | ungefähr 10 Stunden | mod="approx" value="2021-05-01-T22" |
|                          | duration | about 10 days* | ungefähr 10 Tage | mod="approx" value="P10D" |
|                          | set | approximately every 3rd day | ungefähr jeden 3. Tag | mod="approx" value="P3D freq="1x" |

| VTE category: imprecise value | date | – | – | value="future_ref" meaning="dct + 0000-00-0X" |
| offset-like date | a few days earlier | vor ein paar Tagen | – | value="future_ref" meaning="dct + 0000-00-06" |
| time | – | – | – | value="future_ref" meaning="dct + 0000-00-00-TOX" |
| offset-like time | a few hours earlier | vor ein paar Stunden | – | value="PXD" |
| duration | a few days* | ein paar Tage | alle paar Tage | value="PXD freq="1x" |
| set | every few days | alle paar Tage | – | |

| VTE category: range of values | date | between August 13th and 15th | zwischen dem 13. und 15. August | meaning="2021-08-13 ≤ Z ≤ 2021-08-15" |
| offset-like date | 5 to 6 days later | 5 bis 6 Tage später | value="future_ref" meaning="dct + 0000-00-05 ≤ Z ≤ dct + 0000-00-06" |
| time | between 1 p.m. and 3 p.m. | zwischen 13 und 15 Uhr | meaning="2021-05-01-T13 ≤ Z ≤ 2021-05-01-T15" |
| offset-like time | 5 to 6 hours later | 5 bis 6 Stunden später | value="future_ref" meaning="dct + 0000-00-00-T05 ≤ Z ≤ dct + 0000-00-00-T06" |
| duration | between 8 and 10 years* | zwischen 8 und 10 Jahren | meaning="P8Y ≤ Z ≤ P10Y" |
| set | every 3-4 months* | alle 3-4 Monate | meaning="P3M ≤ Z ≤ P4M freq="1x" |

Table 2: VTE categories (taken from Tissot et al., 2019) with TimeML-type extensions and examples. Where possible, examples from Tissot et al. (2019) were used and marked with *.

The assumed document creation time (dct) is 2021-05-01-T12:00. Like Mazur and Dale (2011), a lowercase x represents a value that has to be determined from the context of an expression.

well as additional contextual knowledge are crucial and that the meaning cannot always be determined unambiguously and directly from the text.

A VTE can also be used anaphorically when another time expression is provided as context. In “2003 bin ich 6 geworden. Damals war die Welt noch in Ordnung.” (I turned 6 in 2003. Back then, the world was still alright.) “damals” receives the value 2003. With regard to future tense, in “Ina wird in zwei Jahren 18. Dann kann sie ihren Führerschein machen.” (Ina will turn 18 in two years. Then, she can get her driver’s license.) “dann” (then) gets the (local) meaning in two years which would, depending on ref/dct, result in a specific year. In both cases, the date of the otherwise vague time expression can be identified as such.

Expressions such as “künftig” (in future) and “in letzter Zeit” (lately) refer to a period of time anchored in the utterance time and facing towards the future or the past. The example “Ich werde künftig vorsichtiger sein.” (I’ll be more careful in the future.) suggests that the proposition “Ich bin vorsichtiger” (I will be more careful) applies to the speaker at any point after the utterance time. The expression is of type duration and receives the meaning "PXXY" with a beginpoint "dct" and an endpoint "future_ref". In “Peter hat in letzter Zeit sehr hart gearbeitet.” (Peter has been working very hard lately.) “in letzter Zeit” (lately) refers to a time span from a point in time in the near past to the utterance time. This period of time can be days, weeks, or months, depending on the context.

There are some idiomatic expressions or phrases in German (as well as in English) which contain a precise time expression but are used for expressing an undefined short time duration, and should be therefore regarded as VTEs, like “Hast du eine
Minute?" (Do you have a minute?) or “Eine Sekunde!” (Just a second!).

For expressions like “inzwischen” (meanwhile), in the example “Inzwischen hat es Rücktritts-forderungen gegen sie [...] gegeben.” (In the meantime, there have been calls for her resignation [...] ) (Strögen et al., 2018) we define the following framework. There is a given time in the past from which an implicit time span is drawn up to the utterance time. We reason that the expression is of type date because a call for resignation is an event that takes place at a specific time and/or date and is within a specified period. In the example, the starting point of the implicit time span is a police operation on an unspecified day, but probably several days prior to the dct. Between the operation and the statement, there has been at least one call for resignation. The expression can therefore be normalised to "dct - 0000-00-XX < Z < dct".

4.6 Dataset Statistics after Annotation

The annotation, performed by one of the authors, shows that the corpus includes 1910 time expressions, of which 568 are VTEs, i.e., about 29.74% of the time expressions can be considered vague, with 44.15% in KRAUTS and 18.09% in Short_News. The majority is of types date and duration. The highest ratio of VTEs to all time expressions has a book review (from KRAUTS) with 87.5% VTEs. The highest ratio in a Short_News article is 50%. The largest total number of time expressions (11.3%) as well as VTEs in one article can be found in a weather report with 5 300 tokens and 49 VTEs (Short_News). The largest total number of time expressions in KRAUTS is 46 and can be found in a report (2.43% of tokens) and in a newspaper column (2% of tokens).

Table 3 shows a summary of the annotation results. The statistics show that texts with a narrative structure, which appear more frequently in KRAUTS, contain more VTEs than texts limited to the most important facts, like the articles in Short_News. A possible reason for the increased use of VTEs in columns, comments or fictional texts in KRAUTS is that an exact point in time is neither relevant nor known, or that VTEs fit better into the flow of the text. The fact that there are more precise time expressions on average in the Short_News articles than in KRAUTS suggests that precise time expressions are more suitable to support the facts in short articles.

Table 4 presents the number of classes for each label and reveals a class imbalance in the corpus. The most striking imbalance can be observed for the labels vague and anchorType in Short_News. The labels type of Short_News and vague of KRAUTS are the ones with the most similar distribution of classes.

5 Experiments

For our classification experiments, all tested labels should have at least a limited number of values. This excludes labels like value and meaning because their values are not limited to a fixed set. The labels we tested are type, anchorType and vague.

We used the classifiers RandomForest, DecisionTree, softVoting and ExtraTrees from scikit-learn. The softVoting classifier uses the highest probability from the sum of the predicted probabilities. It combines the classifiers DecisionTree, RandomForest and LinearSupportVectorClassifier. Two types of classification tasks were tested. On the one hand, we used multiclass algorithms that can predict a label with multiple classes. For example, type can be predicted, which contains date, time, duration and set. On the other hand, there are multitask algorithms that can predict several classes, as well as several labels, i.e., predictions for type, vague and anchorType as well as their values can be made at the same time instead of one after the other.

6 Results and Discussion

The results for the full dataset show scores from 0.6 up to 0.75 for the soft voting classifier and up to 0.77 for Extra Trees for the binary classification of vague. Given the size of the corpus and the amount of classes and labels, we consider these results decent. RandomForest and DecisionTree, respectively for multiclass and multitask, achieved slightly lower scores (with up to 0.02 difference for KRAUTS and 0.003 for Short_News). There are no strong deviations between the multiclass and the multitask algorithms. We utilized a macro-averaged F1-score metric to weigh our metric towards the smallest class. Due to the label imbalance, this slightly lowers our score but more precisely represents the results of the experiments.

Table 5 shows the results for each label for the two best algorithms. The classifiers achieve better results on Short_News than KRAUTS. ExtraTrees performs best for two of three labels.

10https://scikit-learn.org (last access: 2021-05-12)
### Table 3: Overview of the annotation results (TE = time expression) – almost 30% of all TE are vague TE.

|              | TE Tokens | %TE of Tokens | VTE Tokens | %VTE of Tokens | % VTE of TE |
|--------------|-----------|---------------|------------|----------------|-------------|
| KRAUTS       | 854       | 1.7           | 7.5        | 0.75           | 44.15       |
| Short_News   | 1056      | 2.27          | 2.8        | 0.41           | 18.09       |
| Both         | 1910      | 1.97          | 5.2        | 0.59           | 29.74       |

On KRAUTS, the algorithms achieve low results on type, mainly because its class set has a low accuracy of 0.22 F1-score for the full dataset because there are only 80 annotated labels in KRAUTS and 29 in Short_News (Table 4). The same problem appears for anchorType with the infrequent ref label. The macro-averaged F1-score clearly demonstrates this because it weighs each class equally so the smaller classes with lower scores equally count to the overall score. In future work, we need to annotate additional data to achieve reasonable classification results. The remaining classes with more samples for type are slightly better with F1-scores from 0.42 for time up to 0.71 for date.

The results show that classifiers with small training sets are capable of achieving F1-scores of up to 0.77. We can assume that more sophisticated approaches will yield better results. In terms of future work, we plan to combine such machine learning-based and rule-based systems, such as Heideltime (Strötgen and Gertz, 2010), which achieves an F1-score of 93.8 on German narrative texts (Strötgen and Gertz, 2015) for precise time expressions. It remains to be explored if a rule-based system can provide similar results for VTEs.

## 7 Conclusion

We concentrate on the annotation of vague time expressions, borrowing especially from Tissot et al. (2019), whose categorization we modified and adapted to classify and normalize VTEs. We describe methods for the normalization of VTEs and annotated a data set of German news documents. Determining the meaning of a VTE proved to be difficult, because it is context-dependent and may require empirical knowledge if no temporal granularity (year, day, hour, etc.) is given. Although our annotation scheme was developed using German documents, we believe it to be applicable to English, too, because English VTE work in a similar way. Finally, we carried out preliminary classification experiments.

In terms of future work, we plan to label the data set with additional annotators to determine the inter-annotator agreement, to expand the data set and to improve the classification results. Another aspect for expanding our work would be to include an evaluation of time span representation of our normalizations. We also plan to explore additional possibilities of classifying different categories of VTE automatically, which are, as of now, only implicitly included. In that regard, it is worth exploring if a regular expression-based approach, like HeidelTime (Strötgen and Gertz, 2010), is able to derive normalised values of VTEs.

### Acknowledgments

The work presented in this paper has received funding from the German Federal Ministry of Education and Research (BMBF) through the project QURATOR (Wachstumskern no. 03WKDA1A).

### References

Tommaso Caselli and Rachele Sprugnoli. 2015. It-TimeML, TimeML Annotation Guidelines for Ital-
ian. Technical report.

Tommaso Caselli and Piek Vossen. 2017. The Event StoryLine Corpus: A New Benchmark for Causal and Temporal Relation Extraction. In Proceedings of the Events and Stories in the News Workshop, pages 77–86. ACL.

Joanna Mary Channell. 1983. Vague language: some vague expressions in English. Ph.D. thesis, University of York.

Filip Devos. 2003. Semantic Vagueness and Lexical Polyvalence. Studia Linguistica, 3(57):121–141.

Filip Devos, Nancy Van Gyselghem, Ria Vandenberghe, and Rita De Caluwe. 1994. Modelling vague lexical time expressions by means of fuzzy set theory. Journal of Quantitative Linguistics, 1(3):189–194.

Anca Dinu, Walther v Hahn, and Cristina Vertan. 2017. On the annotation of vague expressions: a case study on Romanian historical texts. Proceedings of the LT4DHICSEE in conjunction with RANLP, pages 24–31.

Lisa Ferro, Inderjeet Mani, Laurie Gerber, Beth Sundheim, and George Wilson. 2001. TIDES Temporal Annotation Guidelines. Technical Report MTR 01W0000041, The MITRE Corporation.

Paweł Mazur and Robert Dale. 2011. LTIMEX: representing the local semantics of temporal expressions. In 2011 Federated Conference on Computer Science and Information Systems, pages 201–208. Institute of Electrical and Electronics Engineers (IEEE).

Anne-Lyse Minard, Manuela Speranza, Sara Baino, and Martina Coser. 2017. Examples and Guidelines for Annotation of Temporal Expressions (<TIMEX3>) in German.

James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. ISO-TimeML: An International Standard for Semantic Annotation. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10), pages 394–397.

Xin Rong, Adam Fournier, Robin N. Brewer, Meredith Ringel Morris, and Paul N. Bennett. 2017. Managing Uncertainty in Time Expressions for Virtual Assistants. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, page 568–579. Association for Computing Machinery.

Roser Saurí, Jessica Littman, Bob Knippen, Robert Gaizauskas, Andrea Setzer, and James Pustejovsky. 2006. TimeML annotation guidelines.

Frank Schilder and Christopher Habel. 2001. From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages. In Proceedings of the Workshop on Temporal and Spatial Information Processing. ACL.

Jannik Strötgen. 2015. Domain-sensitive Temporal Tagging for Event-centric Information Retrieval. Ph.D. thesis, Heidelberg University.

Jannik Strötgen and Michael Gertz. 2010. HeidelTime: High quality rule-based extraction and normalization of temporal expressions. In Proceedings of the 5th International Workshop on Semantic Evaluation, pages 321–324. ACL.

Jannik Strötgen and Michael Gertz. 2015. A baseline temporal tagger for all languages. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 541–547.

Jannik Strötgen, Anne-Lyse Minard, Lukas Lange, Manuela Speranza, and Bernardo Magnini. 2018. KRAUTS: A German Temporally Annotated News Corpus. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan. European Language Resources Association (ELRA).

Hegler Tissot, Marcos Didonet Del Fabro, Leon Derczynski, and Angus Roberts. 2019. Normalisation of imprecise temporal expressions extracted from text. Knowledge and Information Systems, 61(3):1361–1394.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. SemEval-2013 task 1: TempEval-3: Evaluating time expressions, events, and temporal relations. In Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pages 1–9. ACL.
A Annotation examples

The assumed document creation time (dct) is 2021-05-31.

1. Wir haben uns zuletzt bei unserer Abschlussfeier getroffen. <Timex3 type="date" anchorType="dct" value="past_ref" meaning="dct - 000X &lt; Z &lt; dct" vague="true"> Inzwischen </Timex3> habe ich mir ein Auto gekauft.11

2. Er ist <Timex3 type="date" anchorType="dct" value="past_ref" meaning="dct - 00XX" vague="true"> vor Jahrzehnten </Timex3> ausgewandert.

3. Die Kampagne startet <Timex3 type="date" anchorType="dct" value="2021-06" mod="mid" vague="true"> Mitte Juni</Timex3>.

4. Das Gesetz wird <Timex3 type="date" anchorType="dct" value="2021-08" meaning="2021-08-12 &lt;= Z &lt;= 2021-08-15" vague="true"> zwischen dem 12. und 15. August </Timex3> verabschiedet.

5. Die Lebensspanne dieser Schmetterlingsart beträgt <Timex3 type="duration" value="P10D" mod="approx" vague="true"> circa 10 Tage </Timex3>

6. Der Umbau dauert nur noch <Timex3 type="duration" value="PXD" vague="true" beginPoint="dct" endPoint="dct + 0000-00-0X"> wenige Tage </Timex3>

7. Die Post kommt hier nur <Timex3 type="set" value="PXD" freq="1x" vague="true"> alle paar Tage </Timex3>

8. <Timex3 tid="t1" type="date" value="2003" vague="false"> 2003 </Timex3> bin ich 6 geworden. <Timex3 tid="t2" type="date" anchorType="ref" anchorTimeID="t1" value="2003" vague="false"> Damals </Timex3> war die Welt noch in Ordnung.

9. Es tut mir leid, dass ich dich verletzt habe. Ich werde <Timex3 type="duration" value="future_ref" anchorType="dct" meaning="PXXY" beginPoint="dct" endPoint="future_ref" vague="true"> künftig </Timex3> besser aufpassen.

B List of Vague Time Expressions

Table 6 is an excerpt from the list of over 300 VTEs which can be found in: https://live.european-language-grid.eu/catalogue/lcr/7975 (last access: 2021-08-13).

| Time Expression | Type       | Vague Type | Informal Meaning | Example               |
|-----------------|------------|------------|------------------|-----------------------|
| Pi mal Daumen   | depends on the context | MV | approximately | Pi mal Daumen 1,5 Jahre |
| überschlägig    | depends    | MV         | approximately    | überschlägig 1,5 Jahre |
| annähernd      | depends    | MV         | approximately    | annähernd 1,5 Jahre   |
| ca.             | depends    | MV         | approximately    | ca. 1,5 Jahre        |
| circa           | depends    | MV         | approximately    | circa 1,5 Jahre       |
| in etwa         | depends    | MV         | approximately    | in etwa 1,5 Jahre     |
| praktisch       | depends    | MV         | approximately    | praktisch 2 Jahre     |
| rund            | depends    | MV         | approximately    | rund 1,5 Jahre        |
| schätzungsweise | depends    | MV         | approximately    | schätzungsweise 1,5 Jahre |
| so ziemlich     | depends    | MV         | approximately    | so ziemlich 1,5 Jahre |
| um              | depends    | MV         | approximately    | um die 1,5 Jahre; um 1 Uhr |

11The meaning of this syntax becomes more apparent when the &lt; macros are expanded: meaning="dct - 000X &lt; Z &lt; dct".
| den einen Tag | date | GE | undefined reference | Den einen Tag auf der Bühne, den anderen vor der Kamera, dann noch auf den Kabarettbrettern, wo sie Lieder ihres geliebten [...]. |
| einst | date | not defined | undefined reference | Einst war das anders bei uns. |
| dereinst | date | not defined | past-ref, distant | Ich weiß nur mehr: ich küßte es [das Gesicht] dereinst. |
| einstmals | date | not defined | undefined reference | Einstmals war das anders bei uns. |
| vordem | date | not defined | <anaphoric point | Das Bild hatte vordem im Zimmer seiner Großmutter gehangen. |
| dann | date | not defined | anaphoric point | In einem Jahr steht die Abstimmung über die Abspaltung Schottlands vom Vereinigten Königreich an, spätestens dann muss sich die EU Gedanken machen, ob ein so wichtiger Teil Europas wie Schottland ausgeschlossen werden kann [...]. |
| nunmehr | date | not defined | dct | Der Streik dauert nunmehr schon einen Monat. |
| sofort | date | not defined | future-ref, approx. reference point | Ich habe dich sofort erkannt, als du aus dem Zug stiegst. |
| umgehend | date | not defined | future-ref, approx. reference point | einer Behörde, Instanz umgehend von etw. Mitteilung machen |
| hinterher | date | not defined | future-ref, anaphoric | Die Bedeutung dieser Worte wurde ihm erst hinterher klar. |
| Jahreszeit-monate | duration | GE | date-like | Am meisten liebe ich die Herbstmonate, wegen der vielen Farben. |
| warmelkalte Jahreszeit | duration | GE | date-like | Die warme Jahreszeit ist in dieser Region wirklich schön. |
| in diesen Tagen | duration | IV | anaphoric | Er hat in diesen Tagen viel gelacht. |
| G-lang | duration | IV | PXG | Er musste stundenlang darauf warten |
| all diese G | duration | IV | PXG | All diese Tage gehören der Vergangenheit an. |
| innerhalb von G | duration | IV | PXG / ref <Z <ref + G | Er hat sich innerhalb von Wochen davon erholt. |
| den ersten G | duration | IV | ordinal specified | In den ersten Tagen wird sie noch Probleme damit haben. |
| die nächsten G | duration | IV | DCT/ref + XG | Es wird die nächsten Tage wehtun. |
| in den nächsten G | duration | IV | duration, ordinal specified, future-ref, G | In den nächsten Jahren wird sie viel lernen. |
| die damaligen G | duration | IV | duration, past-ref, distant | Die damaligen Wochen waren wunderschön. Wir sehen uns spätestens in 3 Stunden wieder. |
| spätestens in x G | offset | MV | dct <Z <= dct + X G | frühestens in einem Monat |
| frühestens in x G | offset | MV | >= number G | Er geht nie weg, wenn er tags drauf arbeitet. |
| tags drauf | offset | not defined | anaphoric point +1 | |
| German Word | Offset | Set | Definition | Translation |
|-------------|--------|-----|------------|-------------|
| alsbaldig   | offset | not defined | future-ref, relatively close | Die Ware ist zum alsbaldigen Verbrauch bestimmt. |
| später      | offset | not defined | future-ref, relatively distant | Wie soll das erst einmal später werden? |
| zeitnah     | offset | not defined | future-ref, relatively close | Es gab zeitnah vor und nach dem Brief [von Dr. R.], Gespräche mit Dr. R[...]., daher hielt man eine schriftliche Antwort nicht für nötig. |
| alsbald     | offset | not defined | future-ref, relatively close | Narziß wendete sich zu ihm um, und alsbald fühlte er sich erlöst. veraltet; wie alsbald |
| alsobald    | offset | not defined | future-ref, close | Der Chef drängt auf eine schnellstmögliche Erledigung der Arbeit. |
| schnellstmöglich | offset | not defined | future-ref, very close | Ich komme gleich. |
| gleich      | offset | not defined | future-ref, very close | Die Haut sollte alle paar Tage gründlicher gereinigt werden, um Ablagerungen zu entfernen. Beim Haare waschen mit Shampoo [...]. |
| alle paar G | set | IV | set, irregular G | Heute soll es nur gelegentlich Niederschläge geben. |
| gelegentlich | set | not defined | set, irregular, rarely | [...] und Vokabular, sondern altersgemäß intuitiv durch den alltäglich stattfindenden Gebrauch der englischen Sprache im Betreuungskontext. |
| alltäglich   | set | not defined | everyday | reguläre Mahlzeiten |
| regelmäßig   | set | not defined | set, regularly | eine turnusmäßige Sitzung. Kontrolle |
| turnusmäßig  | set | not defined | set, regularly | etw. läuft zyklisch ab, verläuft zyklisch |
| zyklisch     | set | not defined | set, regularly | Die Beschwerden kehrten periodisch wieder. |
| periodisch   | set | not defined | set, undef | Der vor längerer Zeit errichtete und zwischendurch verfallene Zaun ist jetzt repariert. |
| zwischendurch | set | not defined | set, undef | Er lebt unregelmäßig. man muß das öfters üben, sagen |
| unregelmäßig | set | not defined | set, irregularly, rarely | Wir sehen uns nur ganz sporadisch. [... an der in der jeweiligen Prospektergänzung angegebenen Adresse vor 12 Uhr (irische Ortszeit) an dem dem betreffenden Handelstag vorangegangen [...]. |
| öfters       | set | not defined | set, irregularly, often | die Arbeit muss bis spätestens 12 Uhr fertig sein |
| sporadisch   | set | not defined | set, irregularly, rarely | Darüber hinaus sind Personen, die sich nach 20 Uhr auf dem DESY-Gelände aufhalten, verpflichtet, sich auf Verlangen den [...] |
| vor x Uhr    | time | MV | < number o'clock | |
| spätestens x Uhr | time | MV | <= number o'clock | |
| nach x Uhr   | time | MV | > number o'clock | |
Die SBB RailCities bieten täglich bis mindestens 23.00 Uhr bis maximal x Uhr 25. Juni 2012 Abbauende - ein verlängerter Abbau bis maximal 12.00 Uhr am Dienstag, den 26. Juni 2012 kann in Ausnahmefällen bis zum [...].

Ab 16:30 Uhr gibt es ein buntes Animationsprogramm und ab ca. 18:30 Uhr wird ein Filmhit nach Besucherwünschen gezeigt. Es ist gleich 12 Uhr.

Eben hat es fünf Uhr geschlagen.

| Time                | Type      | Value  |
|---------------------|-----------|--------|
| mindestens x Uhr    | time      | MV     | >= number o’clock |
| bis maximal x Uhr   | time      | MV     | until <= number o’clock |
| ca. x Uhr           | time      | MV     | approx. number o’clock |
| gleich x Uhr        | time      | MV     | approx. number o’clock (<) |
| eben                | time      | not defined | past-ref, very close |

Table 6: Excerpt of the list of VTE. Type values are taken from TimeML. Vague Type borrows from the categories described by Tissot et al. (2019) (without Partial Period). G means granularity and G + 1 means one granularity lever higher. For example, if G = month, then G + 1 = year. X represents a number.