Date and Time in Universal Dependencies

Daniel Zeman
Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University, Prague
zeman@ufal.mff.cuni.cz

Abstract

We attempt to shed some light on the various ways how languages specify date and time, and on the options we have when trying to annotate them uniformly across Universal Dependencies. Examples from several language families are discussed, and their annotation is proposed. Our hope is to eventually make this (or similar) proposal an integral part of the UD annotation guidelines, which would help improve consistency of the UD treebanks. The current annotations are far from consistent, as can be seen from the survey we provide in appendices to this paper.

1 Introduction

The label of the UTC timezone\(^1\) suggests that time can be coordinated and universal. Unfortunately, date and time expressions in the world’s languages are not universal, and their current annotation in the various corpora in Universal Dependencies (UD) (de Marneffe et al., 2021)\(^2\) is far from coordinated. One likely reason is that from the point of view of a grammarian, date and time expressions are a rather marginal phenomenon. Similarly, they are not the first thing to be covered by corpus annotation guidelines; and sometimes there are no guidelines for them at all. To the best of our knowledge, this is the case of Universal Dependencies, at least of the universal part of the UD guidelines\(^3\) (we cannot exclude that one of the language-specific sections discusses these expressions). The issue has been discussed in the UD Github issue tracker\(^4,5\) but the discussion did not result in a concrete specification in the guidelines. No coherent proposal seems to have emerged, neither on the website nor in UD-related papers (de Marneffe et al., 2021; Nivre et al., 2016; Nivre et al., 2020). A noteworthy exception is the recent proposal by Schneider and Zeldes (2021, Section 5), who try to solve dates in English.

The main research question is whether (or to what extent) date and time expressions have internal syntactic structure. UD is a syntactic framework, so in clear cases of syntactic structure we should annotate it analogously to similar constructions elsewhere in the language. On the other hand, date and time expressions are frequent in certain genres across the world’s languages, with globally understood semantics, so it would be beneficial for language-understanding applications to always organize the corresponding items (year, month, day, hour, minute…) the same way. Ideally, we would like to find a rule that is language-independent, yet it does not clash with morphosyntax when applied to concrete languages.

Date and time expressions are difficult not only because of their (understudied) grammatical peculiarity, but also because of the way they are encoded in written language. If the expression involves digits and symbols, the general guiding principle in UD is that the analysis should be parallel to how the expression is pronounced and how it would appear in a treebank of spoken language (de Marneffe et al., 2021, p. 285). Therefore, another research question of this paper is to what extent this principle can actually be followed, as in some cases there are multiple possible readings.

\(^1\)UTC = Coordinated Universal Time
\(^2\)We work with UD 2.8, http://hdl.handle.net/11234/1-3687.
\(^3\)https://universaldependencies.org/guidelines.html (all webs retrieved on October 4, 2021)
\(^4\)https://github.com/UniversalDependencies/docs/issues/113
\(^5\)https://github.com/UniversalDependencies/docs/issues/210
In the present study, we survey various time-related expressions in a selection of the UD languages and their internal syntactic structure (if any). We then propose a solution: a set of guidelines that — if added to the UD documentation — would make annotation of these expressions easier and hopefully more consistent. The current annotation in UD treebanks is not discussed directly in the survey, but for basic date expressions, an overview is provided in the appendices. With typological diversity in mind, we conduct the research on 5 languages from 4 different families.

2 Tokenization

If a date refers to the month by its name, then the day and year numbers are separate tokens, too. However, if the date consists entirely of numbers and punctuation, some UD treebanks prefer to treat the date as a single token. This is not exactly wrong (and it helps avoid some of the issues we are going to discuss in this paper), yet we would argue that splitting the date into multiple tokens can increase the parallelism with dates where the month is named. Furthermore, the orthographic rules in some languages allow writing numerical dates with or without spaces, which would lead to inconsistent annotation if the dates without spaces are not split.

Punctuation separators such as slashes or hyphens should be independent tokens, too. However, if there are periods that, according to the language-specific rules, mark the numbers as ordinal numerals, we recommend to keep them in the same token as the number. The token will then be recognizable as an ordinal numeral, and it will be parallel to English 1st, 2nd, 3rd, 4th...

It is less obvious whether a similar argument should be made for time expressions. The core part, hours and minutes, are typically spelled as one string, looking like a decimal number (although they actually use the sexagesimal system, and sometimes a different punctuation symbol). There are multiple options how to pronounce it, but it is not uncommon that the hour and minute parts are simply read as a sequence of two numbers. We suggest to keep such numbers as one token. However, if the string contains unit names or abbreviations, it should be better split into multiple tokens: 19h15m would become 19h15m.

3 Tags and Features

Names of months and days of week are considered proper names (PROPN) in some languages and common nouns (NOUN) in others. Sometimes the distinction is just in language-internal orthographical rules (whether or not the word is written capitalized), sometimes there may be deeper consequences, e.g., whether the word is used with a definite article. The decision has to be made on a per-language basis; if there are no reasons supporting PROPN, we suggest to use NOUN as the default.

Tokens consisting entirely of digits (i.e., years and sometimes days) should always be tagged as cardinal numerals (NUM). However, sometimes the form of the number clearly indicates that it should be pronounced as an ordinal numeral (5th, 5.). Ordinals are a subclass of adjectives in UD, hence the ADJ tag should be used. Note that the POS category is not changed to noun when the numeral heads a nominal. According to the UD tagging principles, the nominal function is encoded in the incoming dependency relation but the POS tag stays the same.

4 Dates

In the UD taxonomy of syntactic units, dates are nominals. Most often they function as temporal oblique modifiers in clauses (obl:tmod), as in The data will be released on November 15, 2021, or even without a preposition (It took place last April). However, dates can appear in all other constructions where nominals can. They can be non-temporal modifiers (What did they say about June 30?), subjects (June 30 suits me perfectly) etc.

---

6 For instance, in Czech, the default is with spaces after periods but the standard admits an alterative format without spaces in business and technical documents (https://prirucka.ujc.cas.cz/?ref=160&id=810).

7 E.g. in Czech, German and Finnish.

8 It could be argued that the read-out-loud rule leads to multiple tokens for 7:15 because seven fifteen is written as two tokens. However, we do not think it would be useful to extend this rule to purely numerical strings because then we would have to split also numbers that do not denote time.
4.1 English

Two date patterns are common in English: May 15, 2015 (pronounced May (the) fifteenth, twenty fifteen / two thousand (and) fifteen) and 15 May 2015 (pronounced the fifteenth of May twenty fifteen / two thousand (and) fifteen). In both cases, the day may be spelled so that the ordinal numeral is overtly marked: May 15th, 2015. The pronunciation of the day-month-year pattern with the preposition of shows that the month modifies the day and not vice versa; we should treat the written date as it is pronounced, even though the preposition is not visible (cf. (de Marneffe et al., 2021, p. 285)). The month is a temporal nominal modifier, nmod: tmod.9 Similar patterns occur in some other European languages, such as Spanish (el quince de mayo, lit. the fifteenth of May).

The situation is less clear with the month-day-year pattern. Schneider and Zeldes (2021) propose to make the day the head here, too, since the month can be omitted with sufficient context (I'll see you on the fifteenth); if we wanted to omit the day instead, the case marker would have to change (I'll see you in June). While this argumentation probably makes sense in English (no surface signals that the day modifies the month, parallel structure with the first pattern), note that the omission of the month could also be explained as ellipsis, and then the standard UD solution would be to promote the day to the head position (if the rule were that normally the month is the head).

An analogous argument can be made about the year. It can be omitted from the full date (on May 15, 2015)) but it cannot occur with the day and without the month (*on the fifteenth, 2015). We take this as evidence that the year modifies the month, rather than the day-month complex as a whole, and it should be attached to the month. Here we disagree with Schneider and Zeldes (2021), who attach the year to the day. In either case the relation is not nummod because the year is a label that does not express quantity (cf. (de Marneffe et al., 2021, p. 285)). It is a temporal nominal modifier, nmod: tmod. An optional era specifier (BC/AD) will be attached as nmod: tmod to the year.

Note that the year can appear with the month and without the day if the case marker is changed: in May 2015. Both in May and in 2015 are grammatical; however, in the right context, in May has the same meaning as in May 2015 while in 2015 refers to a longer period. Therefore we propose to attach the year as a dependent of the month. Furthermore, it is also parallel to expressions where of is overtly used: in October of 2002.

4.2 Czech

In Czech, the standard word order is day-month-year: 15. května 2015 (pronounced patnáctéhokvětna dva tisíc pátáct, lit. fifteenth Gen May. Gen 2015. Nom). The day is an ordinal numeral (ADJ) but unlike in English, it modifies the noun that denotes the month. This is semantically slightly odd (we are referring to the fifteenth day of May, not to a fifteenth May in a sequence of Mays), and it likely stems from a longer expression “the fifteenth day of May”, but the morphosyntactic behavior has developed to that of regular

---

9The current practice in English UD seems to be that the :tmod subtype is only used when there is no preposition, as a kind of justification why there is nmod without case. We believe that it is equally useful to use it for prepositional temporal modifiers. Similarly, advmod: tmod could be used instead of plain advmod for time-related adverbs, as it is currently used in some other UD languages. However, regardless of how broadly they are applied, relation subtypes are always optional in UD.
adjectival modifiers. The day adjective thus agrees with the month in gender, number, and case. In the example above both of them have the genitive form, which is the default for temporal oblique modifiers, but both of them will switch to the nominative if the date is used as a subject, accusative if used as an object etc. Similar patterns occur in some other European languages, such as German (am fünfzehnten Mai, lit. on-the.Dat fifteenth.Dat May.Dat).10

10There are also German examples where the month is in the genitive and modifies the dative day ordinal: am Fünfzehnten des Monats “on the fifteenth of the month”.

The above reasoning does not work so much when the month is encoded numerically: 15. 5. 2015. It is conventionally pronounced patnáctého pátý dva tisíc patnáct, lit. fifteenth.gen fifth.Nom 2015.Nom, that is, there are two ordinals (plus the year cardinal) and they no longer agree in case. We do not have an explanation for this reading; nevertheless, in the absence of morphosyntactic evidence for one of the possible analyses, we propose to use a parallel structure to that of spelled-out month, i.e., the month is the head. The only change is the label of the first dependency, nmod: tmod instead of amod: tmod, as the relation does not behave like standard adjectival modification in Czech.

4.3 Finnish

The most frequent date form in Finnish is 15. toukokuuta 2015, pronounced viidestoista toukokuuta kaksituhattaviisitoista, lit. fifteenth.Nom May.Par 2015.Nom. The day number is ordinal (ADJ) and it does not agree in case with the month name; instead, it forces the month name into the partitive form. We take this as a sign that the month depends on the day, like in English and unlike in Czech. The partitive case is used also when the month is spelled and pronounced as an ordinal number: 15.5.2015, pronounced viidestoista viidettä kaksituhattaviisitoista. A month name with year (but without the day number) is typically found in the inessive case (toukokuussa 2015 “in May 2015”) but can occur in other cases if they are required by the surrounding syntactic context. Their relation should be nmod: tmod because the number does not specify a quantity of months.

Finnish also has an alternative date pattern, common in informal and spoken situations, where the month precedes the day. Here the month must be in the genitive case and the day takes the essive form. The day ordinal is optionally followed by the word “day”, also in essive: toukokuun viidentänoitoista (päivänä). Another option is that the day ordinal and the word “day” are in nominative: toukokuun viidestoista päivä. Here the day ordinal modifies päivä and agrees with it in case.

The above reasoning does not work so much when the month is encoded numerically: 15. 5. 2015. It is conventionally pronounced patnáctého pátý dva tisíc patnáct, lit. fifteenth.gen fifth.Nom 2015.Nom, that is, there are two ordinals (plus the year cardinal) and they no longer agree in case. We do not have an explanation for this reading; nevertheless, in the absence of morphosyntactic evidence for one of the possible analyses, we propose to use a parallel structure to that of spelled-out month, i.e., the month is the head. The only change is the label of the first dependency, nmod: tmod instead of amod: tmod, as the relation does not behave like standard adjectival modification in Czech.

4.3 Finnish

The most frequent date form in Finnish is 15. toukokuuta 2015, pronounced viidestoista toukokuuta kaksituhattaviisitoista, lit. fifteenth.Nom May.Par 2015.Nom. The day number is ordinal (ADJ) and it does not agree in case with the month name; instead, it forces the month name into the partitive form. We take this as a sign that the month depends on the day, like in English and unlike in Czech. The partitive case is used also when the month is spelled and pronounced as an ordinal number: 15.5.2015, pronounced viidestoista viidettä kaksituhattaviisitoista. A month name with year (but without the day number) is typically found in the inessive case (toukokuussa 2015 “in May 2015”) but can occur in other cases if they are required by the surrounding syntactic context. Their relation should be nmod: tmod because the number does not specify a quantity of months.

Finnish also has an alternative date pattern, common in informal and spoken situations, where the month precedes the day. Here the month must be in the genitive case and the day takes the essive form. The day ordinal is optionally followed by the word “day”, also in essive: toukokuun viidentänoitoista (päivänä). Another option is that the day ordinal and the word “day” are in nominative: toukokuun viidestoista päivä. Here the day ordinal modifies päivä and agrees with it in case.

4.3 Finnish

The most frequent date form in Finnish is 15. toukokuuta 2015, pronounced viidestoista toukokuuta kaksituhattaviisitoista, lit. fifteenth.Nom May.Par 2015.Nom. The day number is ordinal (ADJ) and it does not agree in case with the month name; instead, it forces the month name into the partitive form. We take this as a sign that the month depends on the day, like in English and unlike in Czech. The partitive case is used also when the month is spelled and pronounced as an ordinal number: 15.5.2015, pronounced viidestoista viidettä kaksituhattaviisitoista. A month name with year (but without the day number) is typically found in the inessive case (toukokuussa 2015 “in May 2015”) but can occur in other cases if they are required by the surrounding syntactic context. Their relation should be nmod: tmod because the number does not specify a quantity of months.

Finnish also has an alternative date pattern, common in informal and spoken situations, where the month precedes the day. Here the month must be in the genitive case and the day takes the essive form. The day ordinal is optionally followed by the word “day”, also in essive: toukokuun viidentänoitoista (päivänä). Another option is that the day ordinal and the word “day” are in nominative: toukokuun viidestoista päivä. Here the day ordinal modifies päivä and agrees with it in case.
Finally, it is also possible to encounter date expressions that feature the word päivänä “day” but the preceding number is not properly marked as ordinal (the period is missing): 15 päivänä toukokuuta 2015. We acknowledge that it may then be tagged as a cardinal, although it may be pronounced as an ordinal (viidentenätoista “fifteenth.Ess” instead of viitenätoista “fifteen.Ess”). The relation from päivänä can be nmod:tmod but not nummod.

4.4 Indonesian

In Indonesian, dates are often introduced by the word tanggal “date”: pada tanggal 15 Mei 2015, pronounced pada tanggal lima belas Mei dua ribu lima belas, lit. on date five -teen May two thousand five -teen. The day and year numbers are cardinal (rather than ordinal) numerals. As always, they do not denote quantity of anything, so they should not be attached via nummod. There are no morphological clues that would enlighten the relation between the day and the month, so a flat analysis seems appealing. Nevertheless, we can at least repeat what Schneider and Zeldes (2021) said about English. The month can be omitted and we can say pada tanggal lima belas “on the fifteenth”. If that leads to making the day the head in English, we should make the Indonesian analysis parallel and attach the month to the day as nmod:tmod. The year, if present, will be attached to the month, and the optional era specifier modifies the year.

Modifiers such as lalu “last” and mendatang “next” are tagged ADJ in some treebanks and VERB in others (last = passed, next = coming). Of course, this should be harmonized, but the adjective-verb distinction is beyond the scope of this paper. If the modifier is a verb, it is a relative adnominal clause (“that has passed”) and should be attached as acl:relcl. Otherwise it is a simple adjectival modifier, amod. In tahun lalu “last year”, lalu modifies tahun. We could also say 15 Mei lalu “last May 15”, which may or may not be in the previous year, and it is typically both the fifteenth day of the last May, and the last fifteenth of May. Hence both the day and the month could serve as the parent node, and in the absence of other criteria, we propose to attach the modifier high:
4.5 **Chinese**

In Chinese, dates proceed from the least specific to the most specific item. Numbers are always accompanied by the nouns for “year”, “month” and “day”; there are no names for months, they are encoded simply as a number + “month”. Years are normally written using Western Arabic digits. Months and days either use Arabic digits, too, or they are written in Chinese characters. Examples: 2015年5月15日 (2015 nián 5 yuè 15 rì, pronounced èr líng yī nián wǔ yuè shíwǔ rì), lit. two zero one five year five month fifteen day; 五月十五日 (wǔ yuè shíwǔ rì) “May 15”.

The numerals are cardinal and modify the respective nouns, but their relation should be nmod:tm instead of nummod, as they do not denote quantity. Similarly to Indonesian, there are hardly any criteria that would favor one of the three items as the head. We therefore propose an analysis that is parallel to Indonesian and English, i.e., the less specific item depends on the more specific one. An era specifier, if present, modifies the year expression and is attached to its head noun. The year number may be substituted by an expression such as 同年 (tóngnián) “same year”, 次年 (cìnián) “following year” etc.

5 **Days of Week**

When the name of the day of week occurs together with the date, it can be understood as an apposition. Both expressions refer to the same day and they can be reordered. The first expression is treated as the technical head.

6 **Time**

We propose in Section 2 that time expressed using digits be one token, as in the following example (copied from Schneider and Zeldes (2021)). Schneider and Zeldes also propose that if the time is written as ten o’clock, the token o’clock should be considered an adverb and advmod of ten.
In general, verbose time expressions (as opposed to numbers) can vary substantially across languages, resulting in different analyses.

Note that in English, Czech, Finnish and Indonesian, the prepositions indicate which part is the modifier. In Chinese, we have two numbers with units. Unlike in dates, the nummod analysis is quite appropriate here, as we are counting hours (and minutes). This is not the case in Indonesian, where sepuluh jam “ten hours” is nummod (indicating duration) but jam sepuluh (lit. the hour ten) is nmod:tmod (labeling the hour). The relation between the Chinese hours and minutes cannot be characterized as subordination, hence we propose conj instead of nmod:tmod (nine hours [and] forty-five minutes; another possible candidate would be flat).

If time occurs together with date (it happened on May 15 at 9:45), they are often two independent modifiers — note that in the previous English example, each has its own preposition. However, there are situations where date and time have to be considered as one unit, denoting a point in time, which has a syntactic function in the sentence: What about July 12, between 1:30 and 4:00. It is not obvious whether the date or the time should serve as the head in such cases. In this particular English example, we could say that the preposition about belongs to the date, hence the following analysis:

7 Ranges

Dates and times often come in ranges, as in the sentence The festival takes place from May 15 to June 10. In writing, the range can be signaled by a dash (It takes place May 15 – June 10.) The range can occur at various levels of precision, e.g. It takes place May 15 – 16, or It takes place from May to November 2015.

There are several options how to analyze ranges. Two full date expressions with different prepositions (from and to) could be quite naturally annotated as two sibling oblique modifiers of the same clause. Cases where only a part of the date is ranged (e.g., the month in from May to November 2015) could be handled as ellipsis (i.e., the first modifier would be from May and the second would be to November 2015). Analogously with a dash, the first expression would be May and the second – November 2015.
This solution has the disadvantage that the partial first expression is detached from the shared part, so it is more difficult to infer that May actually refers to May 2015.

Another option is to attach the closing part of the range to the opening part; for partial ranges, only the ranged parts can be connected, and the shared, less specific part is attached to the head. This approach is currently taken in some UD treebanks; however, if the closing part has a preposition (either spelled out, or assumed to be encoded by the dash), the annotators mechanically pick the nmod relation, which seems wrong. The second date does not really modify the first date. Their relation is much closer to coordination: the event occurs on both the dates (as well as on all dates in between). In fact, some languages use conjunctions instead of (or in addition to) prepositions to express ranges: German Mai bis November, Czech květen až listopad “May to November”. Treating the prepositional (from–to) cases as coordination would have the advantage of better parallelism and it would solve the shared modifier problem (from May to November 2015 would be analogous to either May or November 2015). Coordinating May with November would allow 2015 to technically modify the first conjunct but be propagated in enhanced UD to the second conjunct.

We lean towards the coordination analysis as the most parallel and cross-linguistically applicable solution. If that is not accepted by the UD community, then we think that attaching the two endpoints as siblings is better than making one of them an nmod of the other.

8 How to Fix the Treebanks

We provide a survey of annotation patterns in the UD 2.8 treebanks of English, Czech, Finnish, Indonesian and Chinese in the appendices. Some treebanks are internally consistent, some less so, but there is very little consensus across treebanks of the same language. It is thus obvious that any improvement would be welcome, even if it cannot be done perfectly.

In fact, this approach also matches example (49) of (de Marneffe et al., 2021).

As one of the reviewers pointed out, the correlative expression from X to Y is certainly a grammaticalized construction (it cannot be paraphrased as to November from May), and the constituents need not be nominals (Heights range from tall to short, etc.), reminiscent of coordination. See (Reynolds and Pullum, 2013) for an argument that versus has grammaticalized from a preposition to a coordinator.
Fortunately, it is not necessary to re-annotate all UD treebanks manually. Language-specific patterns can be designed that will find (almost) all occurrences of date and time expressions in a treebank, and identify their parts. The annotation can then be harmonized using tree-rewriting systems such as Udapi (Popel et al., 2017) or Grew (Guillaume, 2021).

9 Conclusion

We have surveyed various date/time-related expressions in five languages from four different language families. We have shown that some of these expressions in some languages have internal morphosyntactic structure, which should be observed when constructing their UD analysis. The syntactic structure of semantically corresponding expressions is not always compatible across language boundaries, hence the annotation rules cannot be language-independent. However, in cases where no underlying syntax can be detected, we recommend one of the annotation options as the default solution.

We believe that it is necessary to add some guidelines for date and time expressions to the UD documentation, as it will greatly improve consistency of the UD data (these expressions are quite frequent in some genres). We also believe that with language-specific heuristics, the data can be fixed relatively easily, using existing tools for automatic modification of the dependency structures.

Acknowledgements

This work was supported by the grants 20-16819X (LUSyD) of the Czech Science Foundation; and LM2018101 (LINDAT/CLARIAH-CZ) of the Ministry of Education, Youth, and Sports of the Czech Republic.

We thank the anonymous reviewers for very useful comments.

References

Marie-Catherine de Marneffe, Christopher Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

Bruno Guillaume. 2021. Graph matching and graph rewriting: Grew tools for corpus exploration, maintenance and conversion. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 168–175.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajič, Christopher D. Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, Reut Tsarfaty, and Daniel Zeman. 2016. Universal Dependencies v1: A multilingual treebank collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia. European Language Resources Association.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, pages 4034–4043, Marseille, France. European Language Resources Association.

Martin Popel, Zdeněk Žabokrtský, 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, Göteborg, Sweden.

Brett Reynolds and Geoffrey K. Pullum. 2013. New members of ‘closed classes’ in English. 02.

Nathan Schneider and Amir Zeldes. 2021. Mischievous nominal constructions in Universal Dependencies. In *Proceedings of the Fifth Workshop on Universal Dependencies (UDW, SyntaxFest 2021)*, Sofia, Bulgaria.

---

13 Lists of language-specific names for months and week days are very useful in such heuristics.
A Survey of English Date Annotations in UD 2.8

We provide an overview of date annotations currently applied in the English UD treebanks. The survey demonstrates the variability of dependency relations. There are further differences in POS tags and features but they are not shown here. We only show dates, i.e., without days of week, without times, no ranges etc. Sometimes the same pattern receives multiple annotations within the same treebank. Other treebanks are more consistent internally, yet the approaches differ heavily across treebanks.

The total numbers of occurrences in the tables are approximate. Some marginal cases with extra dependents are either clustered with more generic patterns or omitted from the table completely.

A.1 EWT

| Pattern         | Proposal | Total | EWT Trees |
|-----------------|----------|-------|------------|
| 5/15/2015       |          | 183   | 5/15/2015  |
| 15 May          |          | 11    | 15 May     |
| 15 May 2015     |          | 20    | 15 May 2015 |
| 15th May        |          | 3     | 15th May   |
| 15th of May     |          | 2     | 15th of May |
| May 15          |          | 80    | May 15     |
| May 15, 2015    |          | 45    | May 15, 2015 |
| May 15, 2015 AD |          | 2     | May 15, 2015 AD |
| May 15th        |          | 23    | May 15th   |
| May 15th, 2015  |          | 5     | May 15th, 2015 |
### A.2 GUM

| Pattern | Proposal | Total | GUM Trees |
|---------|----------|-------|-----------|
| 15 May  | 15 May   | 4     | 15 May, 2015 |
| 15 May 2015 | 15 May 2015 | 16 | 15 May 2015, 15 May 2015 |
| 15th of May | 15th of May | 1 | 15th of May |
| May 15  | May 15   | 21    | May 15, 2015 |
| May 15, 2015 | May 15, 2015 | 82 | May 15, 2015, May 15, 2015 |

### A.3 LinES

| Pattern | Proposal | Total | LinES Trees |
|---------|----------|-------|-------------|
| 15 May  | 15 May   | 4     | 15 May     |
| 15 May, 2015 | May 15, 2015 | 2 | 15 May, 2015 |
A.4 ParTUT

| Pattern                  | Proposal | Total  | ParTUT Trees |
|--------------------------|----------|--------|--------------|
| 15 May                   | 15 May   | 5      | 15 May       |
| 15 May 2015              | 15 May   | 50     | 15 May 2015  |
| 15 May of this year      | 15 May   | 1      | 15 May of this year |
| May 15                   | May 15   | 1      | May 15       |
| May 15, 2015             | May 15, 2015 | 5 | May 15, 2015 |

A.5 PUD

| Pattern                  | Proposal | Total  | PUD Trees |
|--------------------------|----------|--------|-----------|
| 15 May                   | 15 May   | 3      | 15 May 15 May |
| 15 May 2015              | 15 May   | 4      | 15 May 2015 |
| 15th May 2015            | 15th May | 1      | 15th May 2015 |
| May 15                   | May 15   | 1      | May 15     |
May 15, 2015

B.1 CLTT

| Pattern | Proposal | Total | CLTT Trees |
|---------|----------|-------|------------|
| 15. května 2015 | 15. května 2015 | 12 |  |
| k 15. květnu kalendářního roku | k 15. květnu kalendářního roku | 1 |  |

B.2 FicTree

| Pattern | Proposal | Total | FicTree Trees |
|---------|----------|-------|---------------|
| 15. května 2015 | 15. května 2015 | 5 |  |
### B.3 PDT

| Pattern                              | Proposal                                      | Total PDT Trees |
|--------------------------------------|-----------------------------------------------|-----------------|
| 15. května tohoto roku               |                                               | 1               |
| 15. 5. 2015                          |                                               | 126             |
| 15. května                           |                                               | 782             |
| 15. května léta Páně 2015            |                                               | 1               |
| 15. května roku 2015                 |                                               | 1               |
| 15. května letošního roku            |                                               | 42              |
| 15. května 2015                      |                                               | 240             |

### B.4 PUD

| Pattern                              | Proposal                                      | Total PUD Trees |
|--------------------------------------|-----------------------------------------------|-----------------|
| 15. května                           |                                               | 8               |
| 15. května                            |                                               | 10              |
C. Survey of Finnish Date Annotations in UD 2.8

C.1 FTB

| Pattern          | Proposal           | Total | FTB Trees |
|------------------|--------------------|-------|-----------|
| 15.5.2015        | 15. 5. 2015       | 10    | 15.5.2015 |
| 15. toukokuuta   | 15. toukokuuta    | 12    |           |
| 15 toukokuuta    | 15 toukokuuta     | 1     | 15 toukokuuta |
| toukokuun 15.    | toukukuun 15.     | 1     |           |
| toukokuussa 2015 | toukokuussa 2015  | 4     |           |

C.2 OOD

| Pattern          | Proposal           | Total | OOD Trees |
|------------------|--------------------|-------|-----------|
| 15.5.2015        | 15. 5. 2015       | 18    | 15.5.2015 |
| 15. toukokuuta   | 15. toukokuuta    | 3     | 15. toukokuuta |
| 15. toukokuuta 2015 | toukokuuta 2015  | 3     | 15. toukokuuta 2015 |

C.3 PUD
| Pattern | Proposal | Total | PUD Trees |
|---------|----------|-------|-----------|
| 15. toukokuuta | 15. toukokuuta | 4 | 15. toukokuuta |
| 15. toukokuuta 2015 | 15. toukokuuta 2015 | 12 | 15. toukokuuta 2015 |
| 15. toukokuuta vuonna 2015 ekr. | 15. toukokuuta vuonna 2015 ekr. | 1 | 15. toukokuuta vuonna 2015 ekr. |
| toukokuun 15. päivä | toukokuun 15. päivä | 1 | toukokuun 15. päivä |
| toukokuun 15. päivänä vuonna 2015 | toukokuun 15. päivänä vuonna 2015 | 1 | toukokuun 15. päivänä vuonna 2015 |
| toukokuussa 2015 | toukokuussa 2015 | 15 | toukokuussa 2015 |

**C.4 TDT**

| Pattern | Proposal | Total | TDT Trees |
|---------|----------|-------|-----------|
| 15.5.2015 | 15. 5. 2015 | 23 | 15.5.2015 |
| 15 päivänä toukokuuta 2015 | 15 päivänä toukokuuta 2015 | 120 | 15 päivänä toukokuuta 2015 |
| 15. toukokuuta | 15. toukokuuta | 47 | 15. toukokuuta |
| 15. toukokuuta 2015 | 15. toukokuuta 2015 | 185 | 15. toukokuuta 2015 |
### D Survey of Indonesian Date Annotations in UD 2.8

#### D.1 CSUI

| Pattern         | Proposal          | Total | CSUI Trees |
|-----------------|-------------------|-------|------------|
| 15 Mei          | 15 Mei            | 12    | 15 Mei     |
| 15 Mei 2015     | 15 Mei 2015       | 32    | 15 Mei 2015 |
| 15 Mei lalu     | 15 Mei lalu       | 5     | 15 Mei lalu |
| tanggal 15 Mei 2015 | tanggal 15 Mei 2015 | 5 | tanggal 15 Mei 2015 |

#### D.2 GSD

| Pattern         | Proposal          | Total | GSD Trees |
|-----------------|-------------------|-------|-----------|
| 15 Mei          | 15 Mei            | 22    | 15 Mei    |
15 Mei 2015

tanggal 15 Mei

tanggal 15 Mei 2015

tanggal 15 Mei pada tahun itu

tanggal 15 Mei 2015 lalu

tanggal 15 Mei 2015 yang lalu

D.3 PUD
### E Survey of Chinese Date Annotations in UD 2.8

#### E.1 GSD

| Pattern | Proposal | Total | GSD Trees |
|---------|----------|-------|-----------|
| 2015年5月 | 2015 年 5 月 | 141 |  |
| 2015年5月15日 | 2015 年 5 月 15 日 | 236 |  |
E.2 PUD

| Pattern       | Proposal        | Total | PUD Trees |
|---------------|-----------------|-------|-----------|
| 2015年5月      |                 |       |           |
| 2015年5月15日 |                 |       |           |
| 5月15日        |                 | 67    |           |
| 五月十五日     |                 | 1     |           |
| 同年5月        |                 | 21    |           |
|               |                 |       |           |
|               |                 | 14    |           |
|               |                 | 12    |           |
2015年五月

5月15日

公元前2015年5月15日

去年5月15日