

# Anotation guidelines for fine-grained German NER

Josef Ruppenhofer, Ines Rehbein

December 2, 2019

## Contents

<b>1</b>	<b>General instructions</b>	<b>2</b>
<b>2</b>	<b>Grammatical considerations</b>	<b>3</b>
2.1	Annotation depth . . . . .	3
2.2	Articles . . . . .	4
2.3	Names used outside of their normal category . . . . .	4
2.4	Name variants . . . . .	5
2.5	Names versus non-names . . . . .	5
2.6	Metaphorical uses of Names . . . . .	6
2.7	Metonymic uses of Names . . . . .	6
2.8	Foreign language material . . . . .	7
2.9	Interrupted entities, elaborations, repetitions and ranges . . .	7
2.10	Mistokenized material . . . . .	7
2.11	Syntactic coordination . . . . .	8
<b>3</b>	<b>Labels</b>	<b>9</b>
3.1	AGENTS . . . . .	9
3.1.1	ORG Organizations . . . . .	9
3.1.2	MED MEDIUM . . . . .	12
3.1.3	PER Persons . . . . .	13
3.1.4	CREAT . . . . .	13
3.1.5	TITLE . . . . .	14
3.1.6	NRP NationalityReligionPolitical affiliation . . . . .	15
3.1.7	GPE geopolitical entities . . . . .	16
3.2	SPATIAL INFO . . . . .	17
3.2.1	LOC Locations . . . . .	17

3.2.2	ADD Addresses . . . . .	18
3.2.3	URL URLs . . . . .	19
3.3	HUMAN CREATIONS . . . . .	19
3.3.1	FAC Facilities . . . . .	19
3.3.2	LAN Language . . . . .	20
3.3.3	LAW . . . . .	20
3.3.4	ART Works of Art . . . . .	21
3.3.5	PRODUCT . . . . .	21
3.4	TIMES DATES EVENTS . . . . .	22
3.4.1	EVT event . . . . .	22
3.4.2	PROJ project . . . . .	22
3.4.3	FREQ . . . . .	22
3.4.4	SORD . . . . .	23
3.4.5	TIME . . . . .	24
3.4.6	DATE . . . . .	25
3.4.7	DUR Durations . . . . .	27
3.4.8	AGE . . . . .	27
3.5	QUANTITATIVE . . . . .	28
3.5.1	MON Money . . . . .	29
3.5.2	CARDINAL . . . . .	29
3.5.3	ORDINAL . . . . .	30
3.5.4	PERC percent . . . . .	30
3.5.5	FRAC fractions . . . . .	30
3.5.6	QUANT quantity . . . . .	31
3.5.7	RATE . . . . .	31
3.5.8	SCORE . . . . .	31
3.6	OTHER . . . . .	32
3.6.1	MISC miscellaneous . . . . .	32

4	Notes	32
---	-------	----

# 1 General instructions

Many cases of Named entities (NEs) are clear and easy to identify but that is not true of all of them. There exist some core criteria for identifying named entities, whether they belong to a particular domain or not. However, these criteria do not always apply to everything one wants to tag under the umbrella of NER. Accordingly, these are not necessary and sufficient for all labels but more characteristic of prototypes such as PERson names, ORGanization names and certain subclasses of geographic names.

- **Referential uniqueness:** does the expression identify a unique referent, either across all possible words/domains/discourse contexts or just in a specific context? „Pope John Paul II” seems like a cross-world unique referent. „President Bush” is not unique in this ‘real’ world. Still, for most NLP purposes one would be content to treat this as an NE. (The metadata might disambiguate which President is referred to.)
- **Grammatical behavior:** does the expression exhibit grammatical behavior different from common nouns (at least in Standard German)? For instance, does it not combine with articles? Proper nouns are good candidates for NE status.
- **Referential autonomy:** is the expression self-sufficient to identify the referent, at least in a given communicative context, or does one need external knowledge? E.g. the (current) president of the US is uniquely referring but it is not autonomous.
- **Denominational stability:** a tendency that NEs have been claimed to exhibit is that they have fewer and more regular variations than other noun phrases.
- **Rigid designation:** the expression is a name / designation that references the same object in all possible worlds

**A note on these instructions:** the examples below usually discuss and mark up only a single relevant named entity. Example sentences may contain other instances or instances of other types of NE that are not marked up. Of course, in real annotation they should also be labeled.

## 2 Grammatical considerations

### 2.1 Annotation depth

**One layer only** When a named entity candidate is part of a larger span that also is a named entity, we only label the larger one. For instance, the NP in 1 representing the full name of the car company BMW, contains the NRP *Bayerische*, which goes untagged.

- (1) [Bayerische Motorenwerke ORG]

**No subword annotation** Named entity morphemes or words that are part of compounds cannot be annotated. Thus, in (2) the GPE *Bayern* which is the non-head element of the adjectival compound *bayernweite* cannot be captured. The token *bayernweite* receives no annotation.

- (2) Im Oktober werden bayernweite Streiks im Einzelhandel stattfinden.  
 ‘There will be Bavaria-wide strikes in the retail sector in October.’

## 2.2 Articles

Generally, we do not include definite articles with name-like entities except with some very rare cases when it is clear that they need to be included

- (3) die [Bayerischen Motorenwerke ORG]

- (4) der [SV Werder Bremen ORG]

vs

- (5) [Die Toten Hosen ORG] (name of band)

- (6) [Die Ärzte ORG] (name of band)

- (7) Es ist ihm wieder einmal gelungen: Hauptkommissar Richard Voss hat den ZDF-Krimi „[Der Alte ART]“ zum Quotensieg geführt.  
 (name of tv show)

## 2.3 Names used outside of their normal category

Person or place names may occur as part of NEs of other categories.<sup>1</sup>

- (8) mit einem rumänischen [Schiff Transylvania FAC]

- (9) [Sturmtief Heinz EVT]

- (10) [Orkan Niklas EVT]

- (11) [Niklas EVT] zieht über den Kudamm und bringt die Kräne zum  
 „Tanzen“.

---

<sup>1</sup>The substring representing a person or place name is not separately tagged in any way as we don’t have a second-layer of annotation.

## 2.4 Name variants

Certain types of entities often have multiple name variants. For instance, soccer clubs may have an official long form name (12) and various shorter variants (13–14) as well as potentially abbreviations used in social media contexts (15). We tag all these variants as instances of the NE as long as they are recognizable as referring to the entity.

- (12) [Fußballclub Gelsenkirchen-Schalke 04 e. V. ORG]
- (13) [FC Schalke 04 ORG]
- (14) [#Schalke ORG] löst Vertrag mit Ex-Trainer #Tedesco auf.
- (15) Auf Einladung des FC Sevilla: [S04 ORG] spielt um die Trofeo Antonio Puerta

Name variants may also occur with other types of NEs, for instance LAWs.

- (16) [Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken LAW]
- (17) [Netzwerkdurchsetzungsgesetz LAW]
- (18) [Facebook-Gesetz LAW]

## 2.5 Names versus non-names

Sometimes it is difficult to separate “true” names from popular / colloquial ways of referring to an NE. For instance, the German noun *Unterhaus* need not strictly be referring to the *British House of Commons*, as in (19). It can also refer to the lower chamber of other countries’ parliaments (cf. 20). Similarly, the simple noun *Bundesregierung* in a newspaper in Germany by default is interpreted to refer to the federal government of Germany. But Austria’s federal government can also be referred to in short as *Bundesregierung*. Accordingly, we should not label *Unterhaus* as an NE in 19. But we do label phrases headed by those words when they contain further modifiers or complements that help uniquely identify a particular instance.

- (19) Am Samstagmorgen kommt in London das Unterhaus zusammen.
- (20) Er brachte seiner Partei 2015 einen historischen Sieg ein, als die Wähler sie von 36 auf 184 Sitze im [kanadischen Unterhaus ORG] katapultierten.

## 2.6 Metaphorical uses of Names

Although in such cases the concept of the literally referred to entity may be activated, we do not annotate these cases as such. We leave them unlabeled.

- (21) [BoJo PER] ist der Donald Trump [Englands GPE]
- (22) das einstige Silicon Valley der [Schweiz GPE]
- (23) [Bucharest GPE] ist das Paris des [Ostens LOC]
- (24) Exklusiv-Studie: Wo liegen die Silicon Valleys in [Deutschland GPE]?
- (25) Und wir müssen aufpassen, dass die Donald Trumps dieser Welt, diese negativen Kräfte, nicht das bekommen, was sie wollen.

One reason for our practice is that doing otherwise would violate our usual guideline that we annotate what the expression refers to in context. Recall that with metonymies we also annotate the metonymic category rather than the basic/literal category. Another reason is that formally the metaphorical uses are no longer named entities. This can be seen from the fact that these cases accept the definite article like other kind-referring common nouns, e.g. *der Donal Trump* in (21). And finally, one can pluralize such cases because there may be multiple instances of the type (cf. 24–25).

## 2.7 Metonymic uses of Names

We treat these as belonging to their target category. One very frequent sub-case in news data is the use of place names to refer to an ORG as in 27. In German, unlike English, it is also very common to refer to sports teams in the form of demonyms (cf. ??). There are also cases of referring to the maker of a product instead of the name of the specific model, as in 28. (Cf. section 3.3.5).

- (26) [Köln ORG] greift über links an.
- (27) Die [Kieler ORG] spielen heute in Hamburg.
- (28) Hans fährt einen [VW PRODUCT]

## 2.8 Foreign language material

We do not annotate such expressions unless they constitute well-defined Named entities. E.g. below in (29), „Back to the 80s” , the „80s” gets no label DATE.

(29) Back to the 80’s mit Livemusik von „The Echo Police”

Foreign language song and movie title or organization names do get annotated:

(30) Als Jugendllicher stand ich auf [The Who ORG]-

## 2.9 Interrupted entities, elaborations, repetitions and ranges

In the spoken language data we label, we sometimes find cases where a named entity is not produced in one go but is interrupted in the middle by short pauses, which are potentially filled by *äh* and other fillers. In these cases, we also put a label on the filler or pause token to create a contiguous set of labels.

(neunzehnhundert B-DATE) [äh I-DATE] [dreiunddreißig I-DATE]

In other cases, two or more taggable expressions follow each other. While often it may seem as if the speaker was aiming to produce only a single true value, which might argue for using only a single label, it is usually hard to decide. Accordingly, we tag all the pieces separately. Thus, in 31 we treat each age separately, rather than having a single AGE over the oder-coordination. In 32, we proceed the same way but notice that here we also label the initial bare cardinal number *zwei* as DUR since in context this is the entity that the speaker is trying to produce.

(31) ich war ja erst [13 AGE] oder [14 Jahre alt AGE] .

(32) die durchgemacht haben [zwei DUR] [drei Jahre DUR] oder [vier Jahre DUR] schon dort

## 2.10 Mistokenized material

When a **genitive** ’s has been split off of a noun, we include it in the noun’s tag:

(33) [Bremen s GPE] [Bürgermeister TITLE] [Carsten Sieling PER]

If a **compound** has been spelled with blanks, still treat it as if it were a single word. That is, if the non-canonical spelling exposes any words that by themselves would be NEs, don't annotate them as such. Focus only on the larger expression's status as an NE or not.

(34) [UBS - Banker TITLE] [#RaoulWeil PER] # don't annotate UBS as ORG here!

(35) [UBS-Banker TITLE] [#RaoulWeil PER]

(36) Spanien - Reise # no annotation here!

Of course, we always ignore uses of NEs in compounds no matter how the compounds are spelled: That is, there is no annotation on these items:

(37) Italienreise

(38) Italien-Beauftragter

(39) Sepsis-Vorkommen

In social media data we often find hashtags or other strings that combine complex NEs into one token. For instance, the expression #schlossschönbrunn in 40 would be rendered in standard spelling as *Schloss Schönbrunn*. We nevertheless tag the run-together variant.

(40) Gelungener Auftakt des #bocspm in #wien im [#schlossschönbrunn FAC].

## 2.11 Syntactic coordination

In some cases the conjuncts are full phrases that can be tagged as instances of the NE type in question, as in 41 and 42. When the coordination conjoins elements that are not both full NE instances, we still label two instances of the NE type, as shown in (43).

(41) Hast du [zehn Minuten DUR] oder [zwanzig Minuten DUR] gebraucht?

(42) Neunzehnhundertdreißig oder so - vielleicht auch [neunzehnhundertvierzig DATE] oder [neunzehnhundertfünfzig DATE] .

(43) Ich war [zehn DUR] oder [zwanzig Minuten DUR] da.



Consistent with the above, for coordinated person names we err on the side of separate labeling of multiple instances. That is, in 46 we label the first name as a separate instance even though that person may well share the last name of the person in the second conjunct.

(44) [Josef PER] und [Karl Mittermaier PER]

(45) [Josef Bierwieser PER] und [Korbinian Mittermaier PER]

When titles are coordinated we also tag them separately.

(46) [Herr TITLE] und [Frau TITLE] [Müller PER]

With quantitative NE types, it is often possible to find coordination-like expressions that express a span or range of values, as in 47–49. We label the components of the sequence as separate NEs.

(47) [zehn QUANT] bis [zwanzig Kilogramm QUANT]

(48) [zehn DUR] bis [zwanzig Minuten DUR]

(49) [zehn CARDINAL] bis [fünfzehn CARDINAL] Personen

(50) Der Verein lädt zu einem zu einem märchenhaften Nachmittag Uhr für Kinder von [6 AGE] bis [8 Jahre AGE] in den Teufelsturm Menden ein.

## 3 Labels

### 3.1 AGENTS

#### 3.1.1 ORG Organizations

**Central cases** Names of companies, government agencies, political parties, educational institutions, sport teams, hospitals, museums, libraries, etc. Government and political facilities such as White House and Pentagon are marked with **ORG**.

(51) Der [FC Barcelona ORG] gibt die Verpflichtung des französischen Weltmeisters Antoine Griezmann bekannt.

(52) [Atlético ORG] zahlte ihm jährlich über 20 Millionen Euro netto, bei [Barça ORG] soll er angeblich knapp drunter liegen.

- (53) Der [BGV ORG] koordiniert anlässlich der Behördenwahlen in der Gemeinde Zell die bürgerlichen Kandidaturen.
- (54) Die [FDP ORG] liebäugelt jetzt mit Jamaika.
- (55) [Verdi ORG] ruft die Beschäftigten der [Münchner Verkehrsgesellschaft ORG] am Dienstag von 3:30 Uhr bis 14:30 Uhr zum Warnstreik auf
- (56) Die Zuständigkeit des [BVerfG ORG] für Verfassungsbeschwerden ergibt sich aus Art. 93 I Nr. 4a GG, §§ 13 Nr. 8a, 90 ff. BVerfGG.

**ORG vs FAC** Organization names can be **FAC** depending on context. Especially, hotels, restaurants, museums, hospitals, libraries, churches and temples, commercial facilities, stock exchanges, etc. are often marked with **FAC**.

- (57) Dortmunder Filiale macht dicht – [Vapiano ORG] schließt erstes Restaurant in Deutschland
- (58) Wir waren zu Mittag bei [Vapiano FAC].

**Teams** City names referring to a sport team are marked with **ORG**. This also applies to country names used to refer to sports (or other ) teams.

- (59) [Hamburg ORG] besiegte [Schalke 04 ORG] heute nachmittag.
- (60) Am Abend spielt [Deutschland ORG] gegen [Estland ORG].

A very tricky case consists of nominalizations of adjectives that belong to the NRP category. As simple adjectives they are easy to label (61). The person-referring nominalization is more ambiguous. In 62, the noun *Kölner* denotes citizens of a GPE and receives the label NRP. In sports contexts, the noun may also be used to refer to the team (63) but this is not always the case: references to individual players occur as well (64).

- (61) Die neuen [Kölner NRP] Preise für Schränke größeren Rauminhalts lagen um sechs bis zwölf Prozent niedriger als bisher.
- (62) Immer weniger [Kölner NRP] greifen zur Tageszeitung
- (63) „Wir müssen in diesem Heimspiel Akzente setzen sowie kompakt, aggressiv und konsequent gegen [die Kölner ORG] verteidigen“, sagt Schalkes U-17-Trainer Frank Fahrenhorst.
- (64) Der [Kölner NRP] foulte den Gladbacher, der sich mit einem Wischer in Richtung Vogts Kopf sowie einem leichten Tritt revanchierte.

**Media and media products** See section 3.1.2 for a discussion on how to distinguish ORG and MEDIUM uses.

**Administrative units** Certain administrative units have official names but are more frequently referred to in alternative ways. We accept variants such as 65 and 66, even though the official name for the specific entity would be *Polizeiinspektion Amberg*.

(65) Die [Polizei Amberg ORG] warnt vor Trickbetrügern.

(66) Die [Amberger Polizei ORG] warnt vor Trickbetrügern.

(67) Die Polizei in [Amberg GPE] warnt vor Trickbetrügern.

References to organizations that we can infer as unique references are still not annotated: e.g. the Kulturstadt in question probably is the „Kulturstadt Bremen“ but we don’t label it.

(68) [Max PER] wohnt in [Bremen GPE]. Er arbeitet auf dem Kulturstadtamt.

**Complex organization names** The boundary between complex instances of ORG versus co-occurrences of ORG with NRP and/or GPE is sometimes tricky. The general advice is to follow as much as possible the referential practice of the entity in question. Alternatively, we can look at the name used in Wikipedia as an external reference, if available. Accordingly, we choose candidate 69 from the two candidates below because the institution uses the domain <https://www.deutscheoperberlin.de>.

(69) ✓ [Deutsche Oper Berlin ORG]

(70) ‡ [Deutsche Oper ORG] [Berlin GPE]

For the case below, we choose 72 because the Wikipedia entry is simply „Charité“ and the internet domain <https://www.charite.de>.

(71) ‡ [Berliner Charité ORG]

(72) ✓ [Berliner NRP] [Charité ORG]

The problem also arises in cases of FAC sometimes. Again, we proceed by the same logic: since the wikipedia entry and the internet domain use the full NP „Kölner Dom“ as the name, we choose option (74).

(73) ‡der [Kölner NRP] [Dom FAC]

(74) ✓[Kölner Dom FAC]

A final note concerns elaborations of names in publicity materials. For instance, the hospital Charité frequently uses the phrase “Charité – Universitätsmedizin Berlin” in its materials. Nevertheless, we don’t consider *Universitätsmedizin Berlin* to be a part of the name or a name of its own, for that matter. Accordingly, we label this case as follows:

(75) [Charité ORG] – Universitätsmedizin [Berlin GPE]

Similarly, the FAZ newspaper often uses the phrase “Frankfurter Allgemeine Zeitung - Zeitung für Deutschland”. Again the ORG’s name consists only of the first NP “Frankfurter Allgemeine Zeitung” as shown by the ORG’s web domain and its wikipedia entry. In the self-descriptive phrase, we only label the GPE term “Deutschland”.

(76) [Frankfurter Allgemeine Zeitung ORG] - Zeitung für [Deutschland GPE]

### 3.1.2 MED MEDIUM

When media channels or platforms are referred to in the context of media use, broadcast or reporting we use MEDIUM.

(77) Journalist @drevoigt startete unmittelbar nach Spielende ein Live-Video auf [Facebook MEDIUM]

(78) Im [Tagesspiegel MED] war ein Bericht über den Fall.

(79) „Wetten, dass?“ lief im [ZDF MEDIUM].

(80) „Nazis raus“ trendet auf [Twitter MEDIUM].

(81) Die [ARD ORG] sucht eine neue Intendantin.

(82) [Twitter ORG] bekommt einen neuen CEO.

Note that the relevant expressions can often also be used as ORG.

(83) Der [Tagesspiegel ORG] sucht eine neue Chefredakteurin.

With newspapers and magazines, there is the possibility that expressions may be used to refer to an artifact rather than the organization producing the artifact. In those cases, we do not annotate the expression.

(84) Ich habe den Fisch in den Tagesspiegel eingewickelt.

### 3.1.3 PER Persons

Proper names of people including

- (85) full names (Peter Altmaier)
  - (86) first names occurring without last names (Hans)
  - (87) last names occurring without a first name (Merkel)
  - (88) individual or family names ([Königin TITLE] [Elisabeth PER])
  - (89) fictional names (Bertie Wooster)
  - (90) unique nicknames (Heckuva job, [Brownie PER]!)
- Generational markers such as Jr. and IV, and royal titles such as Queen and Sir are included in the PER label.
  - Honorific titles other than the above such as Mr., Mrs., Miss, Ms, and Dr. are covered by the category TITLE.

### 3.1.4 CREAT

Names of creatures other than humans including animals and fictional non-human characters.

- (91) [Raupe Nimmersatt CREAT]
- (92) [Biene Maja CREAT] und ihr Freund Willi sind jetzt Grundlage für ein Labyrinth in einem Maisfeld
- (93) Also, [Fritzi CREAT] beißt nicht.

Note that in some cases, the names used may be able to refer to humans in other contexts (cf. 93). Also note that for some cases such as 91 and 92 where the name exclusively or often occurs in an appositive construction with a common noun describing the species, we treat the whole NP as an instance of CREAT.

Names of gods and goddesses or names of human-like fairies are marked with NE-PER, NOT NE-CHA. E.g. Hobbits in the „lord of the rings” would be treated as persons, not creatures. The Barbie doll would be treated as NE-PER, too. By contrast, „Biene Maja” would be an instance of CREAT.

### 3.1.5 TITLE

Titles are names of public offices. In addition, we treat professional / role descriptors as titles.

Titles are not agents but in their relevant use they co-occur with /depend on persons. Note that while most common titles occur before the person's name, some do follow it (cf. 101–102).

- (94) [Kanzlerin TITLE] Merkel hat in den vergangenen drei Wochen mehrmals öffentlich gezittert.
- (95) [Präsident TITLE] Trump wirft Twitter Zensur vor.
- (96) [Premierministerin TITLE] May tritt zurück.
- (97) [First Lady TITLE] Melania Trump besuchte in Ghana frühere „Sklavenfestung“
- (98) [Ratspräsident TITLE] Donald Tusk sprach sich dagegen aus.
- (99) [Stürmer TITLE] Olivier Giroud traf in der 76. Minute.
- (100) [Sozialrechtler TITLE] Richter ist dagegen der Meinung, dass dieser Weg wohl keinen Versicherungsschutz genießen dürfte.
- (101) Frau Ulla Jelpke, [MdB TITLE], wird im Rahmen des Bundestagswahlkampfes zum zweiten mal nach ihrem Besuch 2009 am Samstag, den 7. September, nach Radevormwald kommen wird.
- (102) John Powell, [PhD TITLE], berichtet, dass die Studie zeigt, dass ein breites Spektrum junger Sportler dem Risiko erheblicher Kopfverletzungen ausgesetzt ist

We also annotate honorifics and academic titles. When they occur in series as in (104–105) we use only one combined label:

- (103) [Herr TITLE] Mai
- (104) [Dr. TITLE] Markus Söder
- (105) [Frau Dr. TITLE] Merkel

**Predicative uses** We do NOT annotate uses of job or office titles like the ones below, where the expressions are not used as titles before personal names:

- (106) Thomas spielt immer Stürmer.
- (107) Wenn ich groß bin, möchte ich Bundeskanzlerin werden.

**Solo referential uses** We also do NOT annotate them when they occur by themselves as autonomous referring expressions:

- (108) Der Sozialrechtler verwies auf Klagen gegen Österreich aus der Vergangenheit.
- (109) Der Stürmer kam 1998 nach Österreich zum DSG Klopeinersee.
- (110) Die Kanzlerin fliegt morgen nach Brüssel.

### 3.1.6 NRP NationalityReligionPolitical affiliation

Adjectival forms of GPE and non-GPE place names (such as American), named religions, heritage, and political affiliation. Regardless of context, adjectival forms such as „French” are NRPs while nominal forms such as „France” are GPEs.

- (111) die [europäischen NRP] Fluglinien
- (112) die [Muslime NRP]
- (113) [Frankreichs GPE] [sozialistische NRP] Regierung
- (114) eine [katholische NRP] Sonntagszeitung
- (115) ein [amerikanisches NRP] Unternehmen
- (116) zwei [SPDler NRP]
- (117) [BMW ORG] ist ein [bayerisches NRP] Unternehmen.
- (118) Die [Dresdner NRP] Band tritt morgen in [Berlin GPE] auf.

We also use the label NRP for words referring to political or other ideologies, as in (119–120).

- (119) Der Bericht kritisierte zaghaft eine [neoliberale NRP] Agenda.
- (120) Mehrere starke [maoistische NRP] Parteien sind bis heute in Nepal aktiv.

Note the contrast:

- (121) SPD, Katholische Kirche  $\Rightarrow$  ORG
- (122) SPD-ler, Katholik  $\Rightarrow$  NRP

### 3.1.7 GPE geopolitical entities

Names of geographical administrative entities including countries, villages, cities, states, provinces, prefectures, and other forms of municipalities. Island names are also marked with GPE if they are administrative units. **When you do not know whether a geographical entity is an administrative unit or not, give it the benefit of the doubt and mark it as GPE.**

(123) in [Dresden GPE]

(124) [Frankreich GPE]

(125) bei [München GPE]

(126) [Island GPE]

**City quarters** Named city quarters can also be marked as GPE. Note that a co-occurring adjective referring to the larger city is to be included in the GPE-tag as seen in (129).

(127) Ich wohne im [Jungbusch GPE]

(128) Peter wohnt in der [Weststadt GPE] .

(129) Peter wohnt in der [Heidelberger Weststadt GPE] .<sup>2</sup>

(130) Susanne wohnt in [Schwabing GPE] .

Places that have no administrative status should be treated as LOC:

(131) Peter wohnt in Amberg auf dem [Kreuzberg LOC]

If a phrase just descriptively references a sub-location of a GPE but is not a proper city quarter, treat it as LOC. (See also section 3.2.1.)

(132) Der HRK genauso wie sein Endspielgegner aus der [Heidelberger Nordstadt LOC] machten das beste draus.

(133) Der HRK genauso wie sein Endspielgegner aus dem [Heidelberger Norden LOC] machten das beste draus.

(134) Peter wohnt im [Norden Heidelbergs LOC].

---

<sup>2</sup>Weststadt is the name of a city quarter in the case of Heidelberg.



German states and cities as GPEs are sometimes referred to by appositive constructions such as „Land Hessen etc”. We still only label the entity’s name but do not include the descriptive word „Land” etc.

- (135) Das Land [Baden Württemberg GPE] fördert E-Mobilität.
- (136) Die Stadt [Heidelberg GPE] bietet eine großzügige Gleitzeit-Regelung.
- (137) Der Staat [Israel GPE] ist eine parlamentarische Demokratie nach westlichem Vorbild

**Uses to refer to teams: ORG** Note that uses of potential GPE-terms in context where organizations (e.g. sports teams) are meant should be annotated as ORG:

- (138) Am ersten Spieltag spielt [Bremen ORG] gegen [Hamburg ORG].

Be careful to allow for GPE uses mixed with ORG uses:

- (139) Am ersten Spieltag der Nations League spielt [Italien ORG] in [Frankreich GPE].

## 3.2 SPATIAL INFO

### 3.2.1 LOC Locations

Names of locations other than GPEs including

- (140) celestial bodies (Mars, Venus)
- (141) stars (Beteigeuze, Polarstern)
- (142) continents (Afrika)
- (143) mountains (Kilimandscharo, Zugspitze)
- (144) oceans (Atlantik)
- (145) coasts (Goldküste)
- (146) rivers (Donau, Mississippi)
- (147) lakes (Bodensee, Gardasee)
- (148) borders, etc.

Named regions, areas, and neighborhood such as „Mitteldeutschland” and „Nordeuropa” are included in this category. In addition we also include here phrasal types that combine place names and cardinal directions to pick out subregions

(149) im reichen [Dortmunder Norden LOC]

(150) im reichen [Norden von Dortmund LOC]

Further, we include hyphenated combinations of place name + cardinal directions:

(151) Die Dortmunder Polizei hat in der vergangenen Nacht (14.-15. Juni) erneut einen Schwerpunkteinsatz zur Bekämpfung der örtlichen Straßenkriminalität in [Dortmund-Nord LOC] durchgeführt und zahlreiche polizeiliche Maßnahmen getroffen.

However, we do not label formations such as Innenstadt, Altstadt, Nordstadt, Weststadt when they are not official names of city quarters and do not co-occur with the full location name:

(152) Der SPD-Ortsverein lud ihn darin zu einem Rundgang durch die *Nordstadt* ein

When they do occur with an adjective designating the town, we label Altstadt etc as LOC, if we know that the subpart is not itself an administrative subdivision of the larger city:

(153) Aber wer denkt an die [Amberger Weststadt LOC]?

If it is its own administrative subdivision, or if we’re not sure, we are to label these cases as GPE:

(154) Das war ein grandioser Abend im Herzen der [Heidelberger Altstadt GPE]

### 3.2.2 ADD Addresses

Addresses, phone numbers, and email addresses are treated under this category.

(155) [Hochschulstr. 36, 01069 Dresden, Südvorstadt-Ost ADD]

(156) [089/382-3432 ADD]

(157) [info@doodad.com ADD]

Note that in Twitter data person or organization names that are specially formatted in cosntructs such as at-mentions (158) or hash tags (159), are not treated as a separate class but as either PER or ORG, depending on what is being referred to.

(158)    [@Mainz05 ORG] siegt gegen Schalke

(159)    [#Mainz05 ORG] verliert gegen Schalke.

The simple expression „[place name], [higher-level municipality’s name]” such as „Boston, MA” is NOT marked with ADD when it is used to refer to a GPE.

(160)    Der Entwickler Joshua Nichols, aka nichoj, wohnt in [Boston, MA GPE].

### 3.2.3 URL URLs

Strings specifying Uniform Resource Locators are tagged with their own tag.

(161)    ftp://max:muster@ftp.example.com

(162)    http://de.wikipedia.org

(163)    http://de.wikipedia.org/wiki/Uniform\_Resource\_Locator

(164)    file:///foo/bar.txt

## 3.3 HUMAN CREATIONS

### 3.3.1 FAC Facilities

Names of man-made structures, including buildings, airports, stations, infrastructures (bridges and streets), monuments, oil fields, golf courses, hospitals, zoos, shopping centers, etc.

(165)    [Allianzarena FAC]

(166)    in der [Schloßstraße FAC]

(167)    Der [Berliner NRP] [Alexanderplatz FAC] ist weltbekannt.

(168)    der [Eiffelturm FAC]

(169)    das Minarett der [Al-Nur-Moschee FAC]

- (170) Das Festival wird organisiert von Bewohnern des [St. -Rembert-Stiftes FAC]

Facility names can be ORG depending on context:

- (171) Die [Stadtbibliothek Heidelberg ORG] meldet, dass ..
- (172) Der Imam der [Al-Nur-Moschee ORG] lädt zum Tag der Offenen Tür .
- (173) Das [St. Rembert Stift ORG] gibt jährlich 2 Mio Euro für Medikamente aus.

Other uses may occur as part of e.g. book or movie titles

- (174) [Berlin Alexanderplatz ART] ist ein Roman von Alfred Döblin.

### 3.3.2 LAN Language

Names of languages

- (175) ein [rumänisches LAN] Wörterbuch
- (176) auf [Italienisch LAN] verfasst
- (177) die [deutsche B-LAN] [Sprache I-LAN]
- Many forms are also nationality terms (NRP) and need to be distinguished in that function.
- (178) zwei [italienische NRP] Staatsbürger

NB: Programming languages such as Python, Perl or Java are not covered by this label. Treat them as instances of PRODUCT.

### 3.3.3 LAW

This category covers laws, treaties, and sections and chapters thereof.

- (179) der [Nordatlantikvertrag LAW]
- (180) [Hartz IV LAW]
- (181) [Art 1 GG LAW]
- (182) das [Grundgesetz LAW]

Note that we do not distinguish between LAWs that are actually on the books and bills being proposed to become LAWs: both are LAWs for our purposes.

### 3.3.4 ART Works of Art

Titles of books, songs, films, plays and other creations such as awards, stock price indexes, and social security systems including health insurance systems or pension plans.

Newspaper headlines are marked with ART only when they are referential. Headlines used as „headlines” should not be marked.

Series names, as in the „Harry Potter series”, are also marked.

(183) [Die Blechtrommel ART]

(184) [Stille Nacht ART]

(185) [Oscar ART] (movie award)

TV broadcasts are often to be treated as ART, including news broadcasts and talkshows.

(186) Ich schaue die [Tagesthemen ART].

(187) Bei „[Hart aber fair ART]” ging es um Mieten als soziale Frage - und die Sendung lief ab wie ein Wettrennen.

TV shows that are competitions we treat as EVT:

(188) Hier finden Sie alle News, Videos und Fotos zur aktuellen Staffel von [DSDS EVT]

(189) Mein Onkel ist bei [Wetten, dass? EVT] angetreten?

### 3.3.5 PRODUCT

Name of any product including non-commercial vehicles (automobiles, rockets, aircraft, ships). Although references including manufacturer and product names are marked as two different entities, products using only their manufacturer’s name (as in “I bought a Ford”) are marked with PRODUCT.

(190) [VW ORG] [Passat PRODUCT]

(191) zwei [VWs PRODUCT]

Unlike what is specified in the OntoNotes definition, financial products and services are also included in this category.

(192) Bei Suchmaschinen-Werbung im Teildienst [AdSense for Search PRODUCT] seien andere Anbieter unerlaubterweise behindert worden .

## 3.4 TIMES DATES EVENTS

### 3.4.1 EVT event

Named events and phenomena including natural disasters, hurricanes, revolutions, battles, wars, demonstrations, concerts, sports events, etc.

- (193) das [Tief Felix EVT]
- (194) der [Vietnamkrieg EVT]
- (195) die [Französische Revolution EVT]
- (196) die [Fußballweltmeisterschaft 2014 EVT]
- (197) die [Bundestagswahl 2013 EVT]

Note that entities of this type may include expressions that by themselves usually belong to a different NE-type. E.g. 193 contains the person name *Felix*. Similarly, 195 includes *Französische*, which is, however, not tagged as NRP. In 196 and 197 years are mentioned but they are not tagged as DATE.

### 3.4.2 PROJ project

Intended for names of projects, systems, policies, plans, movements.

- (198) [Green New Deal PROJ]
- (199) [#FridaysForFuture PROJ]

Note that references to doctrines (e.g. *Keynesianismus*), religions, belief systems, political ideologies are cases of NRP.

### 3.4.3 FREQ

Frequency expressions generally indicate a repeated , multiple situation.

Following the CGEL grammar, we assume that frequency adjuncts can be subdivided into two main types, bounding and non-bounding. Bounding expressions answer the question *How many times?*, non-bounding expressions answer the question *How often?*

- (200) Sie gab die Vorlesung lectured [fünf Mal FREQ]. [bounding]
- (201) Sie unterrichtet das Fach [regelmäßig / jeden Tag FREQ]. [ non-bounding]

Note that for the non-bounding type of **FREQ** expression there exist many more different forms than for the bounding type.

- (202) dauernd, oft, häufig, nie, gelegentlich, sporadisch, selten, regelmäßig, manchmal, hie und da, wieder und wieder, immer wieder, ab und an, meistens, in der Regel, etc.

Note that for the **CGEL** grammar constructions as in 203 and 204 are cases of non-bounding frequency expressions. Given their semantics and their formal properties, these could also be treated under the separate label **RATE** (cf. 205). Still, we choose to label 203 and 204 as instances **FREQ** because that identifies them as specifically temporal rate expressions.

- (203) Wir treffen uns [zweimal in der Woche **FREQ**].  
 (204) Wir telefonieren [ein Mal pro Woche **FREQ**].  
 (205) Der Wagen verbraucht [4 l pro 100 km **RATE**]

#### 3.4.4 **SORD**

Serial order expressions pick out one event from a series of repeated events. They indicate an event's temporal order relative to others in the series. The most common forms are exemplified by 206 and 207. Instead of referencing specific occasions, the event instance may be identifiable only relative to others, as in 208 and 209. 209 also shows that **FREQ** expressions may also be adverbs rather than PPs or NPs. Even vaguer forms of referencing the series are possible, as shown by 210 and 211. Note finally, that we also treat a deictic reference to a specific occasion, as in 212, under the heading **SORD**.

- (206) Ich besuchte ihn letztes Jahr [zum ersten / zweiten / ... Mal **SORD**].  
 (207) Ich sah ihn letzte Woche [das erste / zweite / ... Mal **SORD**].  
 (208) [Das nächste/letzte Mal **SORD**] begegneten sie mir in der S-Bahn.  
 (209) Diese Seite wurde [zuletzt **SORD**] am 14. März 2019 um 17:04 Uhr geändert.  
 (210) [Zum wiederholten Mal **SORD**] ist der schwedische Cellist Ludwig Frankmar im Rahmen der „Reihe“-Konzerte in Trossingen aufgetreten  
 (211) Tom traf Kim letzte Woche [wieder / erneut **SORD**].  
 (212) [Dieses Mal **SORD**] bist Du zu weit gegangen!

### 3.4.5 TIME

Prototypically, this label covers times of day: these may be clock times or known, if potentially vague, intervals such as *Morgen* (cf. 217).

- (213) Ich muss bis [3 Uhr TIME] zurück sein.
- (214) Um [15:00 Uhr TIME]
- (215) Rufen Sie nach [1 Uhr TIME] noch mal an.
- (216) Es ist [Viertel nach zwei TIME]
- (217) Am [Morgen TIME] stand ich auf und machte Tee.
- (218) Am [Morgen TIME] stand ich auf und machte Tee.
- (219) In der [ersten Stunde TIME] haben wir Mathe.
- (220) Wir werden gegen [9 TIME] daheim sein.
- (221) Wir werden [bald TIME] dort sein.

Note above that we annotate **minimal time expressions**, even if they are used in combination with prepositions to identify other times that are relevant in the discourse. E.g. we label *1 Uhr* in the third example above, (215), rather than *nach 1 Uhr*, which is an interval that does not even include 1 o'clock.

We also use TIME for adverbs such as *jetzt* and *derzeit* or phrases like *im Moment*, which pick out intervals deictically. In addition we use the label TIME for adverbs such as *danach*, *vorher*, *nachher* which pick out intervals relative to anaphorically given landmarks.

When TIME and DATE occur as separate phrases, we tag them separately, as in (222). They may also occur together in compounds, as in (223). In these cases, we use the label of the head, i.e. TIME.

- (222) Dann brach diese Euphorie schon [Freitag DATE] [Abend TIME] wieder in sich zusammen.
- (223) Am [Freitagabend TIME] beginnt diese Party bei drückender Hitze  
.

We do use TIME also in contexts of habitual or generic descriptions where no specific time is referred to.



- (224) Auf Sylt gibt's [abends TIME] Frühstück - und [morgens TIME] Steak

Expressions of time that are defined only relative to an event such as *nach Spielende* in 225 do not get annotated

- (225) Enttäuschte Gesichter bei den HSG-Mädels als der Wimpel nach Spielende an die Gegnerinnen aus Mössingen übergeben wurde.

### 3.4.6 DATE

DATE is used for a date or period of 24 hours or more, including definite article or contractions of prepositions and definite article. These can be specific calendar dates (cf. 226) or periods that are picked out relative to a deictic temporal center (e.g. 228, 229). When years are abbreviated to their last two digits, as in 232, then include the apostrophe marking the abbreviation in the tag. Finally, simple deictic expressions such as *heute*, *gestern*, *morgen*, *heuer* are annotated as DATE as well (cf. 227). We also treat the related adjectives *heutig*, *morgig*, *gestrig*, *vorgestrig*, etc. as instances of DATE.

- (226) Sie kommt am [31.12.2001 DATE] wieder.
- (227) Ich mache [heute DATE] die Vertretung. Morgen ist der [5. Mai DATE]
- (228) Jo ist [diese Woche DATE] im Urlaub.
- (229) Ich bin [kommende Woche DATE] weg.
- (230) Wir sehen uns [nächsten Monat DATE] wieder.
- (231) Ich bin [1982 DATE] geboren.
- (232) Klaus ist ['45 DATE] geboren.
- (233) Er war [letztes Jahr DATE] noch besser in Schuss.
- (234) In den [80ern DATE] waren die Mieten hier günstig.
- (235) Im [20. Jahrhundert DATE] versuchten viele Organisationen, aus den Theorien von Karl Marx und Friedrich Engels neue Staatsorganisationen herzustellen.
- (236) Ich war am [Wochenende DATE] beim Segeln.

Note that we do NOT include the **definite article** (or contractions of preposition and article) as part of DATEs. However, we do include demonstratives as well as relative time markers such as *nächst/letzt/kommend* as part of DATEs.

There are further kinds of named dates, usually holidays

(237) An [Sylvester DATE] fahren wir nach Köln.

(238) Lindner muss die FDP an [Dreikönig DATE] bei Laune halten

We also label expressions with **seasons** (broadly construed) as DATE:

(239) Zweiter Pastoralbesuch nach Deutschland von Papst Johannes Paul II. im [Frühjahr 1987 DATE]

(240) Im [Herbst DATE] kommt die neue Übungsleiterin

(241) Im [Advent DATE] findet hier der Christkindlmarkt statt.

(242) Auch [2019 DATE] findet unsere traditionelle Freizeit im [Karneval DATE] für Erwachsene statt.

(243) Im [Fasching DATE] hat Grippe Hochsaison

Note that we annotate date expressions even if they are not uniquely referring in context. For instance, we would annotate a habitual case like this (where the habituality can be read off the sentence-initial *FREQ jedes Jahr* ‘each year’)

(244) Jedes Jahr besucht Hans am [1. Mai DATE] seine Eltern.

Even outside of habitual contexts, an expression also may not be used in a way that a specific instance can be picked out. However, partial information may be relevant enough in some cases.

(245) Ich bin an einem [Sonntag DATE] geboren.

(246) Franz ist im [Mai DATE] geboren.

However, we do not label potential expressions when they are **used predictively**, as in cases 247–249.

(247) Das wird ein toller Fasching!

(248) Das war das schönste Weihnachten meines Lebens!

(249) Champagner gibt Dir das Gefühl, dass jeder Tag ein Sonntag ist.

Neither do we label cases where the expressions are used in indefinite form to **refer to a type**:

- (250) Danach ist uns so ein Weihnachten nicht mehr gelungen.
- (251) So einen Mai möchte ich nie wieder erleben!
- (252) Einen schöneren Geburtstag hätte ich mir nicht wünschen können.

Finally, note that the adverbs *gestern*, *heute*, *morgen* are treated as DATES.

### 3.4.7 DUR Durations

Periods that are given as combinations of cardinal numbers and temporal units of measurement are treated as DUR. Note that the adjective *lang* in its temporal measure sense is not included in the labeled span (cf. 253).

- (253) Ich wartete [3 Stunden DUR] lang im Regen.  
'I waited for three hours in the rain.'
- (254) Wir blieben [2 Tage DUR].  
'We stayed for two days.'

We also accept expressions with **bare temporal units**:

- (255) Die Leute haben [Tage DUR] gewartet, in der Hoffnung, dass sie Busse rausbringen.

### 3.4.8 AGE

Unlike in Ontonotes, AGE is its own category rather than being included in DATE. Also, while AGE is in its typical realization form a cardinal number, we treat AGE here in section 3.4 with the time-related labels rather than the general quantity expressions in section (3.5).

We annotate as Age various expressions, whether they refer to a person of a specific age or to just the age. Importantly, a simple cardinal number must be labeled as AGE when it is used to specify AGE (cf. 263).

- (256) ein [Hundertjähriger AGE]
- (257) Mein Neffe ist [10 Jahre alt AGE]
- (258) Meine Nichte ist [11 Jahre AGE]
- (259) Opa ist [siebzig AGE].

- (260) Die meisten Lehrer sind [älter als 45 Jahre AGE].
- (261) Peter kam [im Alter von 10 Jahren AGE] nach Österreich.
- (262) Wer geht [mit 67 Jahren AGE] in Rente?
- (263) Susanne, [34 AGE], wohnhaft in Dachau, ...
- (264) Kinder [unter 12 Jahren AGE] brauchen keine Fahrkarte.

Typically, AGE will concern persons but we tag expressions also when they apply to other entities such as e.g. *whisky* or *wine* in (265–266).

- (265) Wir haben [10 Jahre alten AGE] Whisky getrunken.
- (266) Nicht weich genug für einen Riesling Kabinett – und vor allem nicht für einen [2005er AGE].

Note that we also treat as AGE certain idiomatic references to the birth year from which age can be calculated:

- (267) Mein Vater ist [Jahrgang 34 AGE].
- (268) Paul McCartney ist [ein 42er Jahrgang AGE]

Finally, note that we do not tag simple adjectives such as *alt* and *jung* as AGE expressions. Neither do we label compound forms such as *uralt*, *blutjung*.

### 3.5 QUANTITATIVE

NB: approximators should not be included in the spans for the subtypes of QUANTITATIVE. That is, for instance, „circa” is not to be annotated, regardless of the specific NE type at hand.

- (269) circa [zwei Millionen CARDINAL] Teilnehmer
- (270) circa [drei Stunden DUR]
- (271) circa [4 Kilogramm QUANT]
- (272) circa [5 Euro pro Tag RATE]
- (273) ...

While we do annotate phrases consisting of bare measure nouns (in plural form), as in (274–275), we do not tag cases where we would have to do morphological analysis of words to pick out the QUANT morpheme(s), as in (276–277).

- (274) Das dauert wieder [Stunden DUR]
- (275) [Kilometer QUANT] lange Staus durch Straßenbau auf der B 111
- (276) milliardenschwere Investitionen # no annotation
- (277) Kilometerlange Staus sind die Folge. # no annotation

### 3.5.1 MON Money

Monetary units may be written explicitly or implicitly without mentioning the unit of currency. We also accept colloquial references to currency units.

- (278) Ich habe [50 Euro MON] verloren.
- (279) Hans bekommt [\$300 MON].
- (280) Es dauerte vielleicht 25 Minuten, bis meine [400 Dollar MON] sich in [60 MON] verwandelt hatten.
- (281) Hier könnt ihr Maronen für [10 Kröten MON] kaufen!

### 3.5.2 CARDINAL

Any natural number that is used to ‘count’ the number of entities of a specified type.

- (282) [100 CARDINAL] Personen
- (283) [vierzehn CARDINAL] Tore
- (284) [1 Million CARDINAL] Einwohner
- (285) [eine Million CARDINAL] Bewohner
- (286) fast [eine halbe Million CARDINAL] Teilnehmer

Approximators, whether simple or complex, are not included with cardinals:

- (287) weit mehr als [10000 CARDINAL] Demonstranten kamen.

(288) fast [10000 CARDINAL] Demonstranten kamen.

(289) über [10000 CARDINAL] Demonstranten kamen.

Combinations with units of measurement belong to QUANT, however.

(290) Kim wiegt [65 kg QUANT].

### 3.5.3 ORDINAL

All ordinal numbers including adverbials.

(291) der [erste ORDINAL] Punkt ist ...

(292) [Zweitens ORDINAL] hat ...

We treat **numbers written in words** the same as when written in numerals:

(293) Kim war der [1.te ORDINAL] Sieger des Turniers

(294) Petra war die [2.te ORDINAL] Frau auf dem Mond.

### 3.5.4 PERC percent

Percent symbol or the word „percent” should be included.

(295) [60 % PERC] mehr

(296) [100 Prozent PERC] der Einnahmen

### 3.5.5 FRAC fractions

(297) [Vier Fünftel FRAC] der Internetnutzer im Alter von 16-24 Jahren in der EU27 nutzten im Verlauf des Jahres 2010 das Internet zu diesem Zweck

(298) [Zwei Drittel FRAC] der Deutschen haben schiefes Becken

(299) [zwei von drei FRAC] Deutschen haben schiefes Becken

### 3.5.6 QUANT quantity

Measurements including length, distance, area, weight, heat, velocity, temperature, byte size, etc.

Units of measurement must be explicitly mentioned.

- (300) [5 km QUANT]
- (301) [2 g QUANT] Mehl
- (302) wiegt bis zu [10 kg QUANT]
- (303) [100 Meter QUANT]

### 3.5.7 RATE

A rate is the ratio between two related quantities in different units.

- (304) Dieser Roller fährt bis zu [100 km/h RATE]
- (305) Sie zahlen [7 Euro die Stunde RATE] zuzüglich Material
- (306) Du erhältst von uns [4000,- Euro im Monat RATE] + Zulagen + 30 Tage Urlaub.

As noted earlier under the section for FREQ, we treat cases where repetitions of an event over an interval are counted as cases of FREQ rather than as cases of RATE.

- (307) Ich faste [zweimal im Monat FREQ]

### 3.5.8 SCORE

For sports scores we use a separate category. It applies to all types of scores, regardless of their specific structure.

- (308) Und der nach einer weiteren Tennis-Gala im Halbfinale gegen seinen Dauerrivalen mit [7:6 (3), 1:6, 6:3, 6:4 SCORE] die Huldigungen der Zuschauer entgegennahm.
- (309) Schalke schlägt Hoffenheim [ 2:1 SCORE]
- (310) Gegen Brose Bamberg setzte sich das Team von Trainer Mladen Drijencic am Mittwochabend mit [100:89 (52:44) SCORE] durch

- (311) Die Mannschaft von RWK I turnte teils Kürübungen und teils auf Sicherheit und erhielt [99,9 Punkte SCORE].

Note that we do not use this label for lexical items classifying the score in terms of tie, win or defeat. Thus, in 313 there is no label on the noun *Unentschieden*.

- (312) Hoffenheim und Schalke spielten [2:1 SCORE].

- (313) Hoffenheim und Schalke spielten Unentschieden.

## 3.6 OTHER

### 3.6.1 MISC miscellaneous

The remainder category for NE concepts that do not fall into any of the explicitly identified categories.

- We do not label names of diseases , even if they are Latinate or Greek in origin and technical sounding.
- We do not label sayings, proverbs etc. E.g. Mein Vater sagte immer „Eine Schwalbe macht noch keinen Sommer“.

## 4 Notes

A document worth consulting when needing to extend or supplement these instructions is the Nosta-D guidelines <https://www.aclweb.org/anthology/L14-1251/>. Its documentation of subclasses is very helpful.