Annotation Manual Situation Entity Types

Version 1.1

Annemarie Friedrich, Kleio-Isidora Mavridou and Alexis Palmer

Department of Computational Linguistics
Saarland University

afried, apalmer@coli.uni-saarland.de

NOTE: this manual describes the approach to situation entity annotation as of April 2015. The manual is still undergoing some revision, but the core approach to annotation and the features to be annotated are stable.

Contents

1	Intr	oductio	on .		3
2	Tern	ninolog	gy		6
3	Segi	mentati	ion		9
	3.1	What	is a situation?		9
	3.2	Auton	natic Preprocessing		9
	3.3	Segme	entation Error Handling by Annotators	. 1	11
4	Ann	otation	Guidelines	-	13
	4.1	Situati	ion-Related Features	. 1	13
		4.1.1	Genericity of main referent	. 1	13
		4.1.2	Aspectual class (lexical aspect)	. 1	19
		4.1.3	Habituality	. 2	22
	4.2	Situati	ion Entity Types	. 2	28
		4.2.1	State	. 2	28
		4.2.2	Event	. 2	29
		4.2.3	General Statives	. 3	30
		4.2.4	Abstract Entities	. 3	32
		4.2.5	Speech Act Types	. 3	34
	4.3	Derive	ed Situation Entity Types	. 3	35
		4.3.1	Perfect tense	. 3	35
		4.3.2	Negation	. 3	36
		4.3.3	Future and modals	. 3	36
		4.3.4	Conditionals	. 3	37
	4.4	Summ	nary	. 3	39
A	Ann	otation	Scheme for German	4	42
	A.1	Introd	uction	. 4	42
	A.2	Gener	icity of main referent	. 4	42
	A.3		ed Situation Entity Types		44
		A.3.1	Perfect Tense	. 4	44
		A.3.2	Subjunctive	. 4	45
		A.3.3	Statal passive voice	. 4	45

A.3.4	Final clause with "damit"	45
A.3.5	Modal constructions	46

1 Introduction

Discourse modes and situation entity types

Linguistic expressions form patterns in discourse. Passages of text can be analyzed in terms of the individuals, concepts, times and *situations* that they introduce to the discourse. Within a text, one recognizes stretches that are intuitively of different types, and can be clustered by their characteristic linguistic features and interpretations. Smith (2003) posits five *discourse modes*: Narrative, Report, Description, Informative and Argument/Commentary. This list of modes may not be exhaustive, and texts of almost all genre categories have passages of different modes.

The discourse modes are characterized by (a) the type of situations (also referred to as *situation entities*) that are introduced in a text passage, and (b) the principle of text progression in the mode (temporal or atemporal). In this annotation project, we address the first of these characteristics, the *situation entity types* (*SE types*). *Situations* are expressed at the clause level. Following Smith (2003), we distinguish the following *SE types*: **Eventualities** describe particular situations such as Events (1) or States (2).

- (1) The lobster won the quadrille. (EVENT)
- (2) The Colonel owns the farm. (STATE)

The class of **General Statives** includes Generalizing sentences (2), which report regularities, and Generic sentences (4), which make statements about kinds or classes.

- (3) Mary often feeds my cats. (Generalizing)
- (4) The lion has a bushy tail. (GENERIC)

Abstract Entities are the third class of SE types, and comprise Facts (5) and Propositions (6). In addition to the types proposed by Smith (2003), we mark the type Resemblance. These situations differ from the other types in how they relate to the world: eventualities and General Statives are located spatially and temporally in the world, but Abstract Entities are not. Facts are objects of knowledge, and Propositions are objects of belief.

- (5) I know that Mary refused the offer. (FACT)
- (6) I believe that Mary refused the offer. (Proposition)
- (7) She looked as if <u>she had seen a ghost</u>. (RESEMBLANCE)

We limit the annotation of Abstract Entities to the clausal complements of certain licensing predicates, as well as clauses modified by a certain class of adverbs (see Section 4.2.4), as it is not always possible to identify sentences directly expressing Facts or Propositions on linguistic grounds (Smith, 2003, 74). In (6), believe is the licensing predicate, and Mary refused the offer is a situation that is introduced as not being *in* the world, but about the world (Smith, 2003, 74). We use the type Resemblance to mark embedded situations that are complement clauses of constructions such as look like. Some object or situation in the main clause is described as being similar to the embedded situation.

Overview of annotation project

In this annotation project, we mark the SE types of clauses with the aim of providing a large corpus of annotated text for the following purposes:

- (1) Training, development and evaluation of automatic systems classifying situation entities, as well as sub-tasks which have (partially) been studied by the NLP community, but for which no large annotated corpora are available (for example, aspectual information or the genericity of clauses and noun phrases).
- (2) Assessment of the applicability of classification of SE types as described by Smith (2003): how many situations can be classified easily, which borderline cases occur, and how do humans perform on this task?
- (3) Assessment of the applicability of the theory of Discourse Modes: as a second step, we are planning to annotate and evaluate the discourse modes of passages.

We aim to annotate data from a variety of genres and domains, and plan to start with the written part of the MASC corpus (Ide et al., 2008), the manually-annotated subcorpus of the Open American National Corpus. MASC provides texts from 20 different genres and has already been annotated with various linguistic and semantic phenomena. In addition, MASC is freely available, and we aim to distribute our corpus in an easy way. As mentioned above, situations are expressed at the clause level. Segmentation of texts into clauses is a research topic unto itself. In order to provide a common basis for annotators, we segment the texts into clauses using a freely available discourse parser, SPADE (Soricut and Marcu, 2003), and apply customized post-processing steps (described in Section 3.2).

This annotation manual is structured as follows. Section 2 is a glossary of selected key terms, including clarification of how we use some ambiguous terminology. Section 3

¹http://www.americannationalcorpus.org/MASC/Full_MASC.html

describes the automatic segmentation process, and Section 4 first describes features of situations that help to distinguish the SE types, and then presents the SE types.

Throughout this manual, short summaries of the annotation instructions are shown in boxes like this.

2 Terminology

In the literature on linguistics and in the literature on the grammar of the English language, several terms are used with different meanings in various contexts. This section aims to provide for a common understanding of terminology among the annotators involved in this project.

Utterance An utterance is a structure expressing a 'complete thought', and includes at least a subject and a predicate. For the purpose of classifying situation entities, we mainly consider elementary discourse units (see Section 3), which are usually finite clauses, to be utterances.²

Tense describes the correspondence between the form of the verb and our (non-linguistic) concept of time, which has three divisions: past, present and future (Quirk and Greenbaum, 1973, section 3.26).

Reference Time Reichenbach (1980) classifies the tenses of verbs using three time points: the speech time (the point at which an utterance is made), the event time (the time point at which an event takes place) and the reference time, relative to which the event is located in time. As an example, consider (8). In (a), the reference time is in the future ('after the point of speech'), and the event takes place before the reference time. In (b), the reference and event time are located 'before the point of speech', and the event time is the same as the reference time. In (c), the reference time is also located before the point of speech, and the event is even temporally located before the reference time.

- (8) (a) John will have made coffee (when we arrive).
 - (b) John made coffee.
 - (c) John had already made coffee.

Aspect is probably the most ambiguous and thus sometimes misleading term in this area of literature. Part of this difficulty arises from the fact that the English language does not clearly distinguish tense from aspect. While some Romance languages express tense and aspect through verb conjugation, English mostly expresses them by combining different verb forms. For the purpose of our annotation project, it is important to understand and distinguish the following concepts (we are following Filip, 1999).

²http://calleteach.wordpress.com/2010/02/10/tampa-the-basics/

- (a) Aspect is sometimes used as a cover term for the perfective-imperfective distinction. This dimension of aspect is also know as viewpoint aspect (Smith, 1991). While the perfective aspect describes a situation consisting of a single event that does not have any internal structure, the imperfective aspect describes a situation that has some internal structure and is ongoing, habitual or repeated. This distinction is independent of tense. Specifically, the perfective aspect does not imply that the event happened prior to reference time. In (9), the action of entering the room happens at the reference time, which is some point in the future (later than speech time).
 - (9) The students will be quiet when I enter the room.

The event of entering the room is a single event without structure, and hence the when-clause has perfective aspect. In our annotation effort, it is not required to make this distinction explicitly, it is, however, related to – or an underlying step of – determining the *lexical aspect*, and also the situation entity type of an utterance.

- (b) The *lexical aspect* is an attribute of a verb and its arguments. It is also called *aktionsart*, *aspectual class* or *eventuality type*. It is concerned with determining the 'real-life shape' of the event described by the verb and its arguments³, such as state (*love*), process (*walk*) or event (*realize*) (e.g. Bach, 1986). While most verbs have a dominant lexical aspect, they can easily shift depending on the context. The basic lexical aspect of *walk* is the one of a process, but in certain circumstances, it can take on other lexical aspects. Example (10) illustrates such a shift. The lexical aspect of (a) is a process, while the addition of the adverbial *a mile* causes the utterance to describe an event.
 - **(10)** (a) John walked.
 - (b) John walked a mile.

There is no unified system for classifying lexical aspect. Vendler (1957) suggests a distinction into states (desire, want, prefer), activities (run, write), accomplishments (write a letter, collapse) and achievements (win the race, start, realize). The latter three classes can be considered to contain eventive or dynamic verbs, and these three classes differ with respect to whether the action described has a natural endpoint (e.g., collapse) or not (e.g., prefer) and whether the action has a duration (run) or is punctual (realize). States are the easiest category to separate from the others, although processes also share certain properties with states. For the purpose of our annotation project, we distinguish only stative and eventive lexical aspect, as for the classification of situation entity types, we are interested in whether something is the case (= stative verb constellations) or whether something happens (= eventive verb

³http://ell.stackexchange.com/tags/aspect/info

- constellations). Following this argumentation, we consider all activities, accomplishments and achievements to be eventive. More information about how to distinguish stative and eventive verbs is given in Section 4.1.2.
- (c) There is another usage of the term *aspect* more closely related to the construction or analysis of verb forms; this is sometimes referred to as *grammatical aspect*. As mentioned earlier, English does not cleanly distinguish tense and aspect from one another, and thus it is more convenient to talk about verb forms by analysing tense and aspect simultaneously. Sometimes the marking of verb forms using the present or past participle is called *progressive* or *perfect*⁴ aspect, respectively, but this term then relates purely to the verb form. Any English verb form can convey the type of aspect described in (a), even forms unmarked for aspect such as "*Jerry jumped*.", which describes the event as a whole and thus has perfective aspect.

⁴Quirk and Greenbaum (1973) call this perfective, but seem to refer only to the verb form in doing so.

3 Segmentation

The aim of this annotation task, as previously mentioned, is to label each clause of text with the type of situation it introduces to the discourse. In other words, the granularity of our labeling is at the clause level. Because many sentences contain multiple clauses, this means that we first need to segment sentences into their component clauses, before labeling can begin. It turns out that clause segmentation is not an entirely trivial task, and even humans don't always agree on clause boundaries. This section discusses our approach to clause segmentation.

3.1 What is a situation?

We have said that each clause introduces a situation (or situation entity) to the discourse. The type of the introduced SE indirectly depends on, among other things, the internal temporal properties of the *verb constellation*, i.e., the verb and its arguments (Smith, 2003, p. 23). Adverbials and other linguistic factors may affect the interpretation of a verb constellation. We do not mark infinitives or nominal gerunds, but all other verb constellations.

3.2 Automatic Preprocessing

In order to make the annotation process both replicable and more efficient, we use an existing tool for automatically segmenting texts into clauses and add a few simple customized post-processing steps. The tool in question is the SPADE discourse parser (Soricut and Marcu, 2003). SPADE's discourse segmenter takes a sentence-split text as input and outputs a linear segmentation of the text into clauses. Sentence splitting is performed using the PTBTokenizer provided by the Stanford CoreNLP toolkit.⁵

The segmentation provided by SPADE is at times more fine-grained than is appropriate for this task. For example, (11) below shows the segmentation produced by SPADE for one sentence of a text. The first segment – *for every dollar* – doesn't introduce any situation to the discourse. In order to have a segmentation that as nearly as possible follows the expectation of one situation per segment, we do some post-processing of the SPADE output.

The aim of the post-processing step is to merge situationless segments with the appropriate neighboring segment(s). In our implementation, we target segments that meet the following two requirements: (a) the segment lacks a verb, and (b) the segment is not

⁵http://nlp.stanford.edu/software/tokenizer.shtml

identified as a complete sentence. The first step in determining the appropriate neighboring segment for merging is consideration of sentence boundaries: if, as in (11), the verbless segment occurs at the start of the sentence, we merge it with the immediately following segment. (A verbless segment at the end of the sentence would be merged with the immediately preceding segment.)

(11) SPADE segmentation:

For every dollar donated to Goodwill in 1998, we helped our "graduates" earn an estimated \$ 102.

Modified segmentation:

For every dollar donated to Goodwill in 1998, we helped our "graduates" earn an estimated \$102.

In some cases, however, the verbless segment occurs between two segments belonging to the same sentence. In these cases, we look at the parse tree (from the Charniak parser; parses are produced by SPADE during segmentation) to determine in which direction to merge the segment. We attach our verbless segment to the neighboring segment which is 'closest' in the parse (if all three segments are at the same level, we merge with the preceding segment). The procedure is as follows: first, we identify (a) the lowest common ancestor node of the preceding and the current segment and (b) the lowest common ancestor node of the current and the following segment. If one of these nodes is dominated by the other, the current segment is attached to the dominated node's segment.

Example (12) shows the segmentation produced by SPADE for the utterance "Remember what she said in my last letter?", and Figure 3.2 shows the corresponding parse. The second segment of the SPADE segmentation contains no verb. The lowest node in the parse tree dominating the first two segments of the SPADE segmentation is the VP node shown on the same line as *Remember*, and the lowest node in the parse tree dominating the last two segments is the SBAR node. As the VP node dominates the SBAR node, we attach the segment *what* to the segment which follows it, resulting in the modified segmentation shown in (12).

(12) SPADE segmentation:

Remember what she said in my last letter?

Figure 1: Constituent parse for example (12) produced by the Charniak parser.

Modified segmentation:

Remember what she said in my last letter?

3.3 Segmentation Error Handling by Annotators

The automatic segmentation procedure described above produces segments containing exactly one situation most of the time,⁶ but nevertheless sometimes fails to produce adequate segments. The segmentation provided by SPADE (designed for a slightly different task) is at times too fine-grained, and some of these cases are not caught by our simple modification step. For this reason, the annotation interface includes a way for annotators to both mark such segmentation errors and indicate how they should be corrected.

Examples (13) and (14) show such cases. In (13), the final two segments together introduce one situation to the discourse, an indirectly asked question. The last segment does not constitute a self-contained situation, as the contribution to the discourse is *not* a statement that *geography is destiny*, but the entire question *if that means geography is destiny*. Similarly, the last segment of example (14) is not a single situation either, but only a modifier of *an interesting one*.

```
(13) I wonder
if that means
geography is destiny.
```

⁶according to experiments conducted on a four-document "Mini-MASC" corpus

(14) So the shift in the image of Gates has been an interesting one for me to watch.

If a situation is split over multiple segments, we mark the main segment (containing the dominating verb) with the appropriate situation type and indicate that it is not a complete situation. We indicate for the other segment that it is not a complete situation entity as well, and that it belongs to the previous (or following) segment. In addition, we provide for cases of discontinuous situations, which, for example, occur when a relative clause interrupts the main clause, by numbering the segments and allowing the annotator to indicate the number of the segment which forms a situation together with the current segment.

There also is the reverse case, segments containing more than one situation.⁷ Example (15) shows an unclear case; the last segment (*which ... back entrance*) could be considered to contain two situations. (This particular example could also be considered to describe a single situation, as the two events are related and happen synchronously – the two phrases actually describe the same event.)

(15) Once selected, objects rode crated or wrapped in foam in armed convoys of trucks, which sped along highways and wove through Cairo traffic to reach the museum's guarded back entrance.

Segments invoking more than one situation should be annotated with the situation type related to the first situation-evoking verb in the segment. Additionally, annotators are asked to mark the segment as containing 'more than one situation' and give a comment explaining the second (or more generally, other) situation(s).

There are also segments not introducing a situation to the discourse at all, and which are not part of another situation. An example of such a situation is given in (16) below.

(16) Cheers, Kara

Segments not invoking any situation which do not belong to another situation are marked with 'no situation'.

⁷We did not observe any such cases during the development of the automatic segmentation procedure, but they are theoretically possible.

4 Annotation Guidelines

Some of the situation entity types we are interested in are easier to recognize than others. While some can be identified on the basis of surface structure and clear linguistic indicators, others depend on internal temporal (and other) properties of the verb constellation or on other, perhaps less obvious aspects of the clause. For annotation purposes, we propose a classification of introduced situations along several dimensions.

In this section, we first present situation-related features and then discuss the situation types to be used for annotation. Next, in Section 4.3 we detail the effect of negation, future tense, and conditional constructions on the final situation type. Finally, Section 4.4 presents an overview of the types and their features. It is our hope that this overview will be helpful for making decisions during the annotation process.

Figure 2 shows a screenshot of the annotation tool developed for this project. For each segment (highlighted in the text presented on the left-hand side), annotators are asked to mark features and types on the right-hand side. In addition, they are given the possibility to indicate segmentation problems and give additional comments.

4.1 Situation-Related Features

In this section we describe three features related to situations. ⁸

4.1.1 Genericity of main referent

This feature is designed to answer the question *What is this clause about?* (or *What kind of thing is this clause about?*

Usually, but not always, the *main referent* of a clause is realized as its grammatical subject. That said, it may not always be clear what the main referent of a clause is, and sentences without such a referent do exist. We appeal to the annotator's intuitions of what the clause is about in order to determine this feature.

In case the main referent does not coincide with the grammatical subject, this is to be indicated during the annotation. For example, in the first clause of (17), *Cornell* is the main referent, while the decision for clause (b) is not as clear. Without any prosodic information given, we cannot assume that the grammatical subject *he* is the main referent, and we might also have the impression that the clause rather makes a statement about the *employment options*.

⁸Throughout this section, examples sometimes are presented with only their value for the feature under discussion (e.g. examples (19)-(32)) and sometimes with the feature value and the situation entity type of the clause (e.g. examples (33)-(40)). SE types appear in small caps, feature values in normal text.

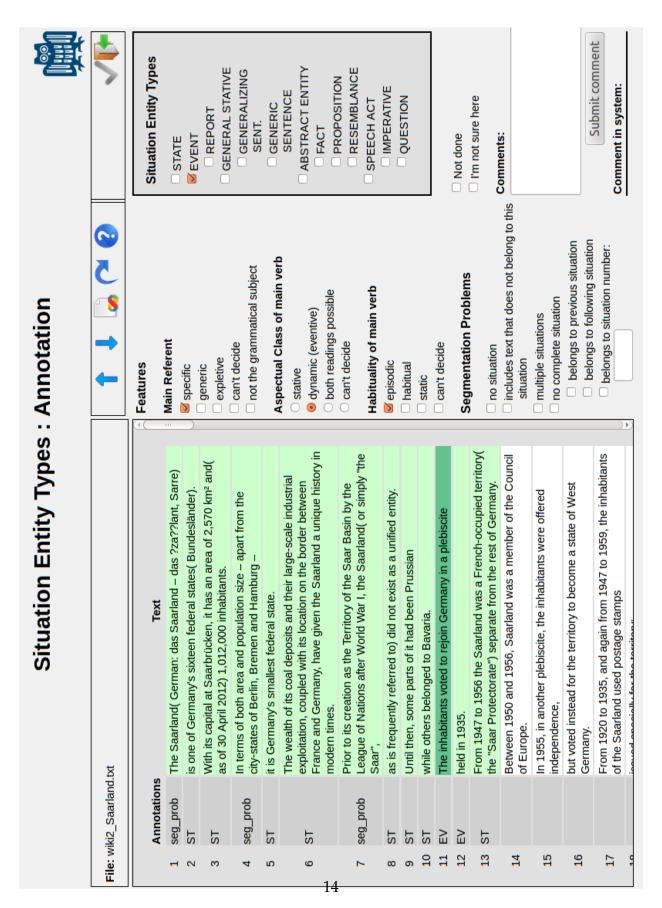


Figure 2: Situation Entity Annotation Tool: types, features and comments.

- (17) (a) Cornell found a number of employment options
 - (b) that he never dreamed existed.

The syntactic mechanism of *topicalization* puts emphasis on one of the syntactic constituents by moving it from its canonical position to the front of the sentence. In (18), the speaker of the sentence, who appears as the grammatical subject, is making a statement about the pizza rather than himself or herself.

(18) That pizza, I won't eat.

Some of the situation entity types are distinguished by whether they make a statement about some *non-generic*⁹ main referent or about a *generic* main referent. In the following, we discuss a number of examples to clarify this distinction.

Non-generic main referents are particular entities (19), particular groups of entities (21), organizations (20), particular situations (22) or particular instantiations of a concept (23).

- (19) Mary likes popcorn. (particular entity \rightarrow non-generic)
- (20) The students met at the cafeteria. (a particular group \rightarrow non-generic)
- (21) IBM was a very popular company in the 80s. (organization \rightarrow non-generic)
- (22) That she didn't answer her phone really upset me. (particular situation \rightarrow non-generic)
- (23) <u>Today's weather</u> was really nice. (particular instantiation of a concept \rightarrow non-generic)

The majority of **generic** main referents are noun phrases referring to a *kind* rather than to a particular entity, and generic mentions of concepts or notions. Definite NPs (24) and bare plural NPs (25) are the main kind-referring NP types (Smith, 2003, p. 73).

⁹At the beginning of defining these guidelines, this feature values was dubbed *specific*, and appears this way in Friedrich and Palmer (2014). The term *specific* is used in the literature (e.g. Krifka et al., 1995) to describe NPs referring to specific individuals rather than *nonspecific* NPs, which do not refer to particular entities. Our feature value does not refer to this sense of *specific/nonspecific* and should rather have been called *non-generic* instead. Both specific and nonspecific NPs can be **generic** in our scheme: *The lion is a dangerous animal* (specific, kind-referring) and *A lion roars when it is hungry* (nonspecific, non-kind-referring) and *A lion must be standing in the bush over there* (nonspecific, non-kind-referring, see (Krifka et al., 1995, p. 15)) are marked **specific=particular** because they do not make statements about classes but rather about particular individuals. We renamed the feature value to *non-generic* as of October 2014.

- (24) The lion has a bushy tail. (generic)
- (25) <u>Dinosaurs</u> are extinct. (generic)

At times, it can be difficult to decide whether a situation makes a statement about a kind or not. One test that can be applied to decide whether a NP is kind-referring is to ask whether it can be used as the subject argument of *be extinct/die out* or the object argument of *invent/exterminate*. Only kinds (not objects) can die out / be invented. If this test applies, a kind-referring NP can be assumed. While some NPs clearly make a reference to a 'well-established kind', other cases are not so clear cut, as humans tend to make up a context in which an NP describes some kind (Krifka et al., 1995, p. 11). Sentence (26) gives an example for such a case: While *lions in captivity* are not a generally well-established kind, we can think of this term as describing a class of entities rather than a particular group of lions in this context, and hence we mark it as kind-referring as well.

(26) Lions in captivity have trouble producing offspring. (generic)

However, as mentioned above (see example (21)), if plurals refer to a particular set of particular individuals, they are not considered to be kind-referring. In most cases, such plurals will be expressed by definite NPs.

Example (27) can also be considered as kind-referring, applying a *taxonomic* reading, i.e., the NP *a lion* refers to a particular subtype of *lion*.

(27) A lion has a bushy tail. (a type of lion \rightarrow generic)

Kind-readings only apply to indefinite NPs if a frequency adverbial such as *typically* or *usually* can be added to the clause felicitously. While this is possible in (27), it is not possible in (28), as we see in (29).¹⁰

- (28) A lion escaped from the zoo. (non-generic)
- (29) A lion usually has a bushy tail. #A lion usually escaped from the zoo.

There are cases, such as (30) and (31), which describe neither a particular individual nor a kind. Rather, they describe a concept which may have various components. For example, the concept *weather* comprises all kinds of weather phenomena such as storms, but also general indicators like temperature and humidity. Similarly, *security* stands for the security regulations, technical devices and humans involved in the *security concept* of some organization.

¹⁰'#' indicates an utterance that is grammatical but not semantically felicitous.

- (30) The weather is something you can't count on. (generic)
- (31) Security is an important issue in US electoral campaigns. (generic)

It is worth noting that many NPs denote particular individuals rather than concepts, even if they do not have a 'physical' referent in the world. An example is given in (32): although the noun *plot* seems to express a concept, in this context it denotes a particular plot, and is hence to be marked as non-generic.

(32) The plot of 'Pirates of the Caribbean' is sometimes confusing. (non-generic)

Gerunds may occur as the subject in English sentences. They usually describe some process or kind of process, and thus also do not have a clear physical referent. When they describe a particular process such as in (a), we mark them as non-generic individuals. When they describe a *kind* of process, however, we mark them as concepts. (The reason for this decision will become apparent in Section 4.2.3.)

- (33) (a) Knitting this scarf took me 3 days. (non-generic, EVENT)
 - (b) Knitting a scarf is generally fun. (generic, Generic Sentence)

Similarly, example (34) and (35) illustrate that the NP *many people* can refer to people in general (in the first case) or to a particular group of people.

- (34) <u>Many people</u> think (generic, Generic sentence) that foxes are cute. (generic, Proposition, Generic sentence)
- (35) Many people were on the bus today. (non-generic, State)

There are cases (a) where an NP refers to a kind in the sense that the NP serves as a placeholder for all individuals belonging to the kind, as in example (24), and (b) where an NP directly refers to the kind or 'species' rather than its members as in (36) or (37). Both cases are marked as **generic**.

- (36) A new species of foxes was discovered. (generic, Event)
- (37) One species of foxes has grey ears. (generic, Generic Sentence)

Similarly, in (38), the main referent refers to a type of sport.

(38) Baseball is a bat-and-ball game. (generic, Generic Sentence)

Quantifiers have no effect on the genericity of the main referent: example (39) is a GENERIC SENTENCE, no matter whether it describes a smaller or bigger subclass of people.

(39) Few/Most/Many/Some people like spinach. (generic, Generic Sentence)

Example (40) is a tricky case. Strictly applying our annotation scheme leads to the following conclusion: *species of foxes* is **generic**, as it refers to several kinds of foxes. The quantifier *Only twelve* then picks out twelve of these kinds, which still results in the collection of several species of foxes, hence referring to kinds, and hence receiving the label **generic**.

(40) Only twelve species of foxes actually belong to the Vulpes genus of "true fox". (generic, Generic Sentence)

Expletives Identifying the main referent when the expletive *it* is the grammatical subject can sometimes be a challenge. In example (41), we propose to consider both segments to 'be about John', and hence the main referent is non-generic. It should also be marked that the main referent is not the grammatical subject.

(41) It was <u>John</u> (non-generic, main referent is not the grammatical subject) who wrote the letter. (non-generic)

In example (42), we suggest marking the main referent in the first segment as **no-main-referent**.

(42) It seemed (no-main-referent, State) like <u>he</u> would win. (non-generic)

Existential clauses In English, sentences expressing existence often start with an unstressed *there*, and the logical subject occurs as a complement of the verb (Quirk and Greenbaum, 1973, 14.21). These sentences are 'about' the logical subject, e.g., sentence (43) is about the books. Annotators are asked to indicate that the main referent is not the grammatical subject in such cases.

(43) There are <u>two books</u> on the table. (non-generic)

If the main referent of the clause is a kind, the genericity feature for the segment is marked as **generic**. If the main referent is a particular entity, an organization, a particular group of entities, a particular situation, or a particular instantiation of a concept, the segment is marked with **non-generic**.

In case there is no clear main referent as the segment contains only an expletive *it* as the grammatical subject and there is no other identifiable main referent (about whom/which the segment makes a statement), the main referent feature can be marked as **no-main-referent**.

There are many cases where this distinction is not easily made, and annotators are asked to give these features only if sure, and otherwise mark the *NP type* feature of the segment as **can't decide**.

4.1.2 Aspectual class (lexical aspect)

As introduced in Section 2, the *aspectual class* or *lexical aspect* concerns the "real life shape" of an event (or state) that a verb, in the context of its arguments, names. It is very important to note that the aspectual class of a verb can be different from the type of situation entity introduced by the clause as a whole. The aspectual class of a verb in a particular clause is determined primarily by the verb constellation (the verb and its arguments); sometimes modifiers (such as adverbials) provide additional information relevant to lexical aspect.

In this work, we distinguish two types of aspectual class:

- (a) Verbs with **stative** lexical aspect name states of being. They are *static* and unchanging; they hold in time. They often show thoughts or opinions, senses, emotions or possession. The state may have a start or end point (e.g., if you *love* someone, you probably started doing so at some point in time, but it's still a state). Verbs classically considered as stative include: *be, like, live, hate, prefer, mean, consist of, mind, recognize, seem, prefer, dbout, need, belong, believe, cost, get, impress, know, taste, love.*
- (b) Verbs with **dynamic/eventive** lexical aspect name actions or processes that change over time or that cause changes over time. They usually have some duration, but they may also be punctual (*He knocked*). Events are *dynamic*, can happen at once or have successive stages that take time, and they take place in time. They may or may not have a natural endpoint. Verbs classically considered as eventive include: *fix*, *kill*, *collapse*, *realize*, *die*, *forget*, *walk*, *run*, *see*, *drive*, *read*, *blink*, *knock*, *squeeze*.

Only a small number of verbs have an absolute (i.e. unchanging) lexical aspect, but most verbs have a dominant lexical aspect, which may change in certain domains or contexts

(e.g. as the verb constellation changes). Many verbs can readily take either aspect (e.g. <u>stand</u>: "She stood in one place for hours" vs. "She suddenly stood and announced her decision").

Linguistic tests for lexical aspect In this section, we present some linguistic tests to distinguish stative and dynamic/eventive verb constellations. As is usually the case for linguistic tests, tests for lexical aspect are only indicators – it could be that a particular instance of a verb passes three tests for stativity and one for dynamicity. Nevertheless, these are helpful when trying to determine a verb's lexical aspect.

- (a) **Stative** verbs (e.g. *prefer, like, own*):
 - disallow progressive constructions (#I am owning a house.)
 - don't occur as imperatives (#Own a house!)
 - can't appear as the complement of the verb force (#He forced me to own a house.)
 - disallow pseudo-cleft constructions (#What Alice did was own a house.)
- (b) **Dynamic/eventive** verbs (e.g. run, realize, knock, teach):
 - allow only limited use of simple present tense (#John runs to class. vs. John likes apples.)
 - habituals (*John runs to class every day.*)
 - generics (Water runs downhill.)
 - "historical present"; often occurs in jokes & sports announcing (*The president signs the bill; A man walks into a bar; He shoots, he scores.*)
 - can answer the question "What happened?" (John ran to class. vs. #John liked apples.)
 - similarly, can be referred to anaphorically as e.g. "This happened while..."

Aspectual verbs. A special case is the aspectual verb *continue*. If it is used in the sense of *re-starting*, it should be marked as *dynamic*, as in (44). In example (45), however, *continue* describes that something on-going is still going on, and is hence to be marked as *stative*.

(44) John stopped eating his ice cream. (Event, dynamic aspectual class)
John continued eating his ice cream. (Event, dynamic aspectual class)

(45) John ate his ice cream. (Event, dynamic aspectual class)

I tried to talk to him, (Event, dynamic aspectual class)

but John continued eating his ice cream. (Event, stative aspectual class)

Special cases: There are some cases of verbs that seem to allow either an interpretation as stative or as dynamic+habitual. Some examples are *be called* (in the sense of *my name is*), *study at* (in the sense of *my university is*), *work at* (in the sense of *my employer is*). We consider these to be stative predicates, assigning properties to the subject, which have evolved from extremely frequent habitual use of the dynamic sense of the predicate.

Please mark occurrences of verbs such as *call*, *refer* etc. indicating the name of the subject as **stative**. It is possible to analyze these cases as dynamic and habitual, but we believe that the stative meaning (something has the name of ...) is more prominent. Sentence (46), says that the term 'tennis' had the property of being a name for real tennis in the 19th century (remember that states may have beginning and end times).

(46) During most of the 19th-century in fact, the term "tennis" referred to real tennis, not lawn tennis. (stative aspectual class, STATE)

Of course, this does not imply that you should mark all instances of these verbs as stative. Observe the difference of the (a) and (b) cases of examples (47) and (48).

- (47) (a) I work for IBM. (stative aspectual class, State)
 (b) I work hard in the gym every Thursday. (dynamic aspectual class, 'normal' habitual, Generalizing Sentence)
- (48) (a) They are called Pfannkuchen in Berlin. (stative aspectual class, Generic Sentence)
 - (b) He calls me a liar no matter what I say. (**dynamic** aspectual class, 'normal' habitual, Generalizing Sentence)

Aspectual class and situation entity type. The aspectual class of *building a house* is **dynamic**, and in the examples below we see this verb constellation appearing in clauses with two different SE types. Example (49) describes an Event. Clause (50), on the other hand, is a Generalizing Sentence, as it describes a pattern of events; this is a situation with a *derived* type.

- (49) John built a house. (Event, dynamic aspectual class)
- (50) John builds houses. (Generalizing Sentence, dynamic aspectual class)

Example (51) shows a clause whose SE type is a GENERIC SENTENCE, as it refers to a kind (the species *whales*), but its aspectual class is **stative**, as it introduces a property of the main referent to the discourse.

(51) Whales are huge. (Generic Sentence, stative aspectual class)

Regardless of the actual situation entity type, situations are marked with their *aspectual class*, which is either **stative** or **dynamic**. If you can see both readings, mark the aspectual class as **both**.

4.1.3 Habituality

Another dimension along which situation types can be distinguished is whether the situation is *episodic* or *habitual*. The discussion related to this linguistic feature in this section follows Carlson (2005). It is important to note that unlike *aspectual class, habituality* is an attribute of the entire situation.

Some situations describe particular events which occur once, as in (52). Carlson calls this type of sentence *episodic*.

(52) Mary ate oatmeal for breakfast this morning. (episodic)

It is possible that the truth value of episodic sentences may depend on multiple occurrences of particular events, as in (53), which does not express a regularity, but an event that consists of multiple sub-events (each student submitting his or her assignment).

(53) Each student in the class handed in a completed assignment. (episodic)

Sentences expressing regularities about the world which constitute generalizations over events and activities are called *habituals*, as in Example (54).

(54) Mary eats oatmeal for breakfast. (habitual)

Habituals are generalizations and hence have the property of tolerating exceptions. If we learn that Mary eats oatmeal for breakfast, it does not necessarily need to be true that she eats oatmeal at every breakfast. The term *habitual* as used in this annotation project covers more than what is usually considered a matter of habit. Examples (55) and (56) show cases that are also considered to be habituals as they describe regularities.

(55) Soap is used to remove dirt. (habitual)

(56) Glass breaks easily. (habitual)

For the sake of completeness, we introduce a third option to mark *static* clauses which express particular states or properties of the main referent, as in (57), (58) and (59).

- (57) Mary owns three cats. (static, stative aspectual class, State)
- (58) Mary is happy. (static, stative aspectual class, State)
- (59) Cats have fur. (static, stative aspectual class, Generic Sentence)

Linguistic tests:

- If one can add a frequency adverbial such as *typically/usually* to the clause, and the meaning of the resulting sentence differs at most slightly from the meaning of the original sentence, the sentence expresses a regularity, i.e., is habitual.
- If one can extend the sentence with a clause starting with ... and that was when... it is most likely episodic.

Habituals not only describe patterns of events, they also cover situations describing episodic¹¹ states that are taken on repeatedly, as in (60). The verb constellation I feel is stative, but this state is taken on repeatedly by the speaker.¹²

(60) I often feel as if I only get half the story. (habitual, stative aspectual class, Generalizing Sentence)

Situation entities describing states that are taken on habitually as in example (60) are *habitual*; those situation entities should be classified as General Statives. Example (61) gives an example for a General Sentence which is habitual, and example (60) is a Generalizing Sentence.

(61) Sloths often feel tired. (generic main referent, stative aspectual class, habitual \rightarrow Generic sentence)

The adverbial *always* invokes many instances of a state or an event. In example (62), *think* has a *stative* lexical aspect, and *always* invokes many past instances of thinking. It does not mean that the speaker *continuously* thought.

(62) I personally always thought... (habitual \rightarrow Generalizing sentence)

¹¹In our annotation of STATES, we generally do not distinguish between *stage*- and *individual-level* predicates according to Kratzer (1995), leaving this distinction to future work. However, an explanation is due here: only *stage-level* stative predicates (e.g. be tired, be happy, feel hungry) can occur as habituals. *Individual-level* predicates (be tall, own a farm) do not occur as habituals.

¹²Carlson (2005) make the episodic/habitual distinction only for eventive clauses. We extend their definition to include (as habitual) clauses which refer to states taken on repeatedly.

Habituality and General Statives. We introduced the *habituality* feature as we are interested in the distinction between Events (episodic frequency, happens once or a certain number of times) and General Statives (something that happens habitually/regularly). As it applies to the situation as a whole, the value for this feature must be **static** when the entire situation is marked as a State, even if it's a derived state (a negated event, a future event etc., see Section 4.3).

It is further worth noting that for Generic Sentences, whose subject refers to kinds, habitual sentences include cases where a situation occurs regularly for (possibly different) members of the class. For example, in (63), a single spider can die only once, but the sentence is habitual as it generalizes over situations in which spiders die. If the subject is non-generic, of course, the situation must repeat for the same subject. If the situation does not repeat, it is episodic or static.

(63) Spiders die in autumn after producing an egg sac. (habitual, generic main referent, Generic Sentence)

Habituals and negation. Sentences containing negation can still be *habitual*. The adverbial *never* indicates that an event does not occur repeatedly, or that a state is not taken on repeatedly. Clauses with *never* thus are considered to be **habitual**. As illustrated by example (64), the negation does not result in the entire situation being classified as a STATE (compare to Section 4.3.2).

(64) Mary never eats ice cream. (habitual \rightarrow Generalizing sentence)

One interpretation of (65) is *John does not have the habit of smoking ('John never smokes'*) This interpretation is habitual. In the case of negation, if you feel that the sentence makes a statement like 'in any relevant situation, it would be the case that the subject does not do X', please mark the sentence as habitual.

(65) John does not smoke. (habitual \rightarrow Generalizing Sentence)

Habituals and modals. There are cases of Generic Sentences and Generalizing Sentences that contain modals, and it is sometimes difficult to tell whether the sentence should be marked as static (see section on coercion) or not. In short, if you feel that a clause containing a modal describes an action that <u>actually</u> happens regularly or a state that is <u>actually</u> taken on regularly (in all relevant situations), mark it as habitual.¹³ Note

¹³In clauses without modals, you don't need to worry about whether it happens *actually*, they are defined to describe things that happen regularly (in all relevant situations), even if this might be the case only very infrequently, for example *Jane handled the mail from Antarctica* is habitual even if Jane never actually received any mail from Antarctica (Hacquard, 2009).

that this is just a rule-of-thumb to make sure you don't mark cases that merely describe general abilities / possibilities / necessities etc. as habitual, as detailed in the examples below.

The modal *can* causes the habituality feature to be *static*, see example (66). This sentence could be paraphrased by *Zebras are able to recognize one other by their stripes.*, which describes an ability, not whether they actually do this regularly. Example (67), however, is *habitual*.

- (66) Zebras <u>can</u> recognize one another by their stripes. (Generic Sentence, generic main referent, dynamic aspectual class (\rightarrow recognize), static)
- (67) Zebras recognize one another by their stripes. (Generic Sentence, generic main referent, dynamic aspectual class (\rightarrow recognize), habitual)

Examples (68) and (69) illustrate more clearly that the modal *can* makes a huge difference in meaning: (68) says that Melissa has the ability to swim but not whether she actually goes swimming, (69) says that she actually goes swimming regularly. The interpretation of (66) / (67) and (70) / (71) is analogous.

- (68) Melissa <u>can</u> swim. (State, non-generic main referent, dynamic aspectual class (\rightarrow swim), static)
- (69) Melissa swims. (Generalizing Sentence, non-generic main referent, dynamic aspectual class (\rightarrow swim), habitual)
- (70) Tennis is a racquet sport that <u>can</u> be played individually against a single opponent or between two teams of two players each. (Generic Sentence, **generic** main referent, **dynamic** aspectual class (→ recognize), **static**)
- (71) Tennis is a racquet sport that is played individually against a single opponent or between two teams of two players each. (Generic Sentence, generic main referent, dynamic aspectual class (→ recognize), habitual)

Modals, such as *can / could / may / might* usually coerce the situation to be **static**, as in (72), (73) and (74). In these cases, it is said that someone has the ability / permission to do something, rather than that something is actually done regularly.

(72) In an old version of tennis, ...
the ball <u>could</u> be hit off the wall. (**generic** main referent, **dynamic** aspectual class, **static**, Generic Sentence)

- (73) An obstetrician is a doctor who <u>can</u> provide care not only in healthy pregnancies... (**generic** main referent, **dynamic** aspectual class, **static**, GENERIC SENTENCE)
- (74) For instance, in the United Kingdom, animal treatment <u>may</u> only be performed by registered vets. (dynamic aspectual class, static, Generic Sentence)

Some sentences containing modals like *can/was able to* or *must* allow the inference that the predicate within the scope of the modal actually happened. Hacquard (2009) gives (75) as an example, and (76) as a counter-example. (These two examples are not habitual.)

- (75) Yesterday, Jane was able to swim across Lake Balaton. (She actually did it. /*She didn't do it.) (stative aspectual class, static, State) \rightarrow actuality implied.
- (76) In her twenties, Jane was able to swim across Lake Balaton. (She did it quite often. / Though she never did it.) (stative aspectual class, static, State) → no actuality implied.
- (77) is an example where we can infer that the corresponding situation occurred regularly, and are thus marked as *habitual*.
- (77) I <u>had to</u> eat an apple every day when I was a child. (Generalizing Sentence, habitual)

Often, in the present tense (as opposed to past tense as above), the implication that something happens actually is not given, and these sentences are static rather than habitual, see (78). Contrast this to how a speaker would express a habitual in the present (79).

- (78) I have to eat an apple every day. (static, STATE)
- (79) I eat an apple every day. (habitual, Generalizing Sentence)
- (80) I usually have to/must help in the kitchen. (Generalizing Sentence, habitual)
- (81) I usually can help in the kitchen. (habitual)

In (81), can must have the meaning of to be available in order to allow the habitual reading. Note that in English, the signals indicating habituality are often implicit, and cases containing modals where you feel that it cannot easily be inferred that something happened regularly should be marked as **static**.

Clauses with **epistemic** modals, which express the speaker's knowledge of / assumptions about the world as in (82), should be marked as **static**. This clause expresses the fact that the speaker believes that the subject is exercising every day, so the "top-most" layer of the situation does not occur regularly (though the believed embedded predicate might).

(82) *She looks great. She must be exercising every day.* (*static*)

Note that there are cases *with* modals that are **habitual**. The modal *would* can be a simple past tense form for habituals, as in (83).

(83) Our cook would make a carrot for Christmas every year. (Generalizing Sentence, habitual)

Will can also be used in habituals when describing rules of a game as in (84).

(84) The opponent who is unable to return the ball will not gain a point, while the opposite opponent will. (Generic Sentence, habitual)

'May' used to describe subsets of kinds / groups. We found several cases in our corpus, such as (85), where the modal *may* is used to express that the habitual predicate applies only to a subset of the kind / group referred to by the subject. We can infer from these cases that something is <u>actually</u> done regularly by at least part of the group/kind. In such cases, you can mark the clause as **habitual** (this is in line with our definition above).

- (85) Paraveterinary workers either assist vets in their work, (Generic Sentence, dynamic, habitual) or may work within their own scope of practice. (Generic Sentence, dynamic, habitual)
- (86) Many states have birthing centers, (State) where a midwife may work individually or in a group. (Generic Sentence, dynamic, habitual)

Situation entities are marked as **episodic** if they describe an event that occurred once (or a specified number of times), or as **habitual** if they describe an action <u>actually</u> carried out repeatedly/habitually/regularly or if a state is <u>actually</u> taken on repeatedly. If the situation describes a state (e.g. indicates an ability, permission, possibility, ... using modals), the habituality feature can be set to **static**.

4.2 Situation Entity Types

The previous section addressed situation-related features of clauses. In this section, we describe in detail the situation entity types (SE types) which can be assigned to clauses. We consider the effects of negation, modality, conditionals, and future tense in Section 4.3.

4.2.1 State

States introduce particular static properties of main referents marked as **non-generic** to the discourse. Smith (2003) (among others) distinguishes two types of states: *static* states, which are not associated with a start and/or end time, and *durative* states, which involve some start and/or end time. Example (87) is a static state, as we don't consider the fact that a person has blue eyes as having a start or end time (at least not during the lifetime of the person). Sentence (88), on the other hand, is an example of a durative state, as John can start knowing the answer at the earliest after the question has been posed.

- (87) She has blue eyes. (static state)
- (88) John knows the answer. (durative state)

Example (89) illustrates the frequently-occurring case of past participles used as reduced relative clauses. Such constructions allow two possible readings. In reading (a) the participle refers to an event, and the reduced relative clause is a passive construction. In reading (b) the participle functions as an adjective and hence describes a property of the noun that it post-modifies (b).¹⁴ If annotators feel that the participle functions as an adjective, they are asked to treat it as such and consequently mark the segment as stative and State.

- (89) A movie, filmed in black and white, ...
 - (a) A movie, which was filmed in black and white, ...
 - (b)- A movie, which is filmed in black and white, ...

To what extent the participle of a verb focuses on the state resulting from the action, and hence describes a property of the noun rather than the event that caused this property, needs to be decided on a case-to-case basis. Example (90) gives a few more examples which we consider as adjectival uses, and hence states.

¹⁴The simple past does not allow us to draw a conclusion regarding whether this case is an adjectival or passive use of the participle.

- (90) (a) her hair, cut short and dyed brown, ...
 - (b) Mary, named after ...
 - (c) The movie, <u>based</u> on a book by X, ...

Clauses introducing static properties of **non-generic** main referents to the discourse are marked as STATE. Note that clauses introducing properties of **generic** main referents are not marked as STATE (compare to Section 4.1.1 and Section 4.2.3). An exception to this rule are EVENTS with generic main referents which are coerced to be STATES, see Section 4.3.

4.2.2 **Event**

EVENTS introduce particular dynamic events or actions to the discourse. An important characteristic of EVENTS is that they have the effect of advancing narrative time in the discourse. Two special cases should be noted. First, the subtype Report for situation entities is introduced by verbs of speech. Second, although the subject NP of an EVENT is usually non-generic as in (91), it can also be a generic noun phrase as in (94).

- (91) Mickey painted the house. (EVENT)
- (92) "Carl is a tenacious fellow.", (STATE)
- (93) said a source close to USAir. (REPORT)
- (94) The lion disappeared from Asia. (EVENT)

Progressives with dynamic aspectual class express ongoing actions, and are labeled as Events. Example (95) shows such cases. Note that events may extend in time like states do, but nevertheless, in (b), some activity is going on, while in (e), a state is described. Progressives with stative aspectual class are hence marked as State.

- (95) (a) The girl is riding a horse. (EVENT)
 - (b) The girl, who was riding a horse, (EVENT)
 - (c) sneezed. (EVENT)
 - (d) The girl, who was sitting on a bench, (STATE)
 - (e) sneezed. (EVENT)

Clauses introducing particular dynamic events or actions to the discourse are marked as Event. The main referent of Event may be non-generic or generic.

4.2.3 General Statives

General Statives do not express particular states or events. Instead, they express regularities of events, or properties of kinds of entities. They are relatively abstract and non-dynamic in nature (Smith, 2003, p. 24). General Statives have two subtypes: Generic Sentences and Generalizing Sentences.

Generic Sentences say something about a kind or a generic concept or notion. Example (96) makes a general statement about the class of computers, not expressing a state of a particular individual or a particular event.

(96) Computers are very useful. (Generic Sentence)

Note, however, that not all sentences making some statement about a kind are Generic Sentences. Example (97) has a kind-referring main referent, but still expresses an Event. The key factor here is that the clause refers to a specific episodic happening.

(97) Computers were invented less than 100 years ago. (Event)

GENERIC SENTENCES may have eventive/dynamic verb constellations, as in (98).

(98) Lions eat meat. (GENERIC SENTENCE)

GENERIC SENTENCES say something about a **generic** main referent. Their aspectual class may be **stative** or **dynamic**. If the aspectual class is dynamic, they are always **habitual** as they express generalizations. If the aspectual class is stative, and the situation does not express a state that is taken on repeatedly, the habituality feature can be set to **static**.

GENERALIZING SENTENCES express regularities, not particular facts. Such sentences often have a frequency adverbial (*sometimes, always, never, usually*), or a frequency adverbial (*usually*) can be added felicitously. Present simple tense with a dynamic aspectual class is also often interpreted as a GENERALIZING SENTENCE, as in (99).

(99) John drives to work. (Generalizing Sentence)

Example (100) illustrates that Generalizing Sentences often have dynamic aspectual class, as they imply a series of sub-events which are dynamic (Smith, 2003, p. 78).

(100) John always fed the cats last year. (Generalizing Sentence)

GENERALIZING SENTENCES describe patterns of particular events or states; they are **habitual**. The main referent must be **non-generic**. The aspectual class may be **stative** or **dynamic** (although the dynamic case is by far more frequent).

Our definition of the two subcategories of General Statives is somewhat simplified from that in Smith (2003). We consider as Generic Sentences *all* clauses that refer to something typically holding of a class or kind. We make no distinction between whether the clause states inherent properties of the kind, describes actions carried out repeatedly by the kind, or describes something that is usually done with instances of the kind. It is the latter case on which we differ from previous work. According to Smith (2003, p. 73), kind-referring NPs may also occur in Generalizing Sentences, as in (101).

(101) Potatoes are served whole or mashed as a cooked vegetable. (Generic Sentence)¹⁵

While we can see that this example describes a pattern of *what people usually do with potatoes* rather than *an inherent property of potatoes*, we prefer the more unified notion of GENERIC SENTENCES described above.

The distinction between Generic and Generalizing Sentences is far from being a clear one, and researchers have developed many different notions of *genericity* (Carlson and Pelletier, 1995). The common denominator of Generic Sentences and Generalizing Sentences is that at least one of the two basic varieties of genericity applies (Krifka et al., 1995): (a) the clause says something about a kind in general, excluding one-time events related to the kind (as in (94)), and/or (b) the clause expresses a regularity (i.e., is **habitual**), no matter whether it refers to a kind or a particular individual. This means that if a regularity is expressed (i.e., if the feature **habitual** applies), we can mark the segment as a General Stative without specifying the subtype, and without having to decide whether the clause says something about a kind or a particular individual.

If the clause is **habitual** (expresses a regularity), but it isn't clear whether the subject NP is **generic** or not, the segment is marked with GENERAL STATIVE.

If, however, the clause expresses a state rather than a regularity, and we cannot decide whether it describes a state of a particular individual or a kind, we cannot mark the sentence as General Stative. Stative clauses that introduce properties of particular individuals are States in our annotation scheme, while stative clauses introducing properties of kinds are Generic Sentences. The aspectual class of such cases should

 $^{^{15}}$ This particular example is classified as a Generalizing Sentence by Smith (2003, p. 73).

be marked as **stative**. In such cases, the genericity depends solely on the type of the main referent. If the main referent is a generic one, it is implied that the state is taken on repeatedly, i.e., can be said to be habitual in the sense described in Section 4.1.3, and the clause is generic. If, however, the main referent is not generic, and the clause does not contain any markers of genericity, it has to be assumed to be a particular sentence describing a non-iterating state.

4.2.4 Abstract Entities

The class of *abstract entities* describes the situations and discourse entities that consist of Facts, Propositions and clauses describing Resemblances. We use these labels in a very particular sense here to refer to a small number of linguistic constructions which serve a particular discourse function. By referring to an event, rather than directly describing its occurrence, an author introduces the event as either something known to be true (Fact) or believed to be true (Proposition).

According to Smith, facts are not in the world, they are about the world. While EVENT-type situation entities have the effect of advancing narrative time, there is no such advancement when the same event is referred to in a factive or propositional construction. For the most part, abstract entities are recognizable by appearing as clausal complements of certain predicates. For instance, the clausal complement of *know* in Example (102) refers to a FACT, and the clausal complement of (104) refers to a Proposition (compare to (103)).

Please note that the licensing clause itself is not considered to be an Abstract Entity, but only its complement (see example (102)).

- (102) I know (State, licensing) that Mary refused the offer. (Fact, Event)
- (103) Mary refused the offer. (EVENT)
- (104) Mary's refusing the offer was unlikely. (Proposition)
- (105) It looked as if were were in the back of a nightclub. (RESEMBLANCE)

FACT and Proposition are directly taken from Smith (2003). In addition, we introduce the type Resemblance for situations that are licensed by constructions such as *it looked like/it seemed as if/....* The clausal complements of these constructions are neither Propositions nor Facts, but seem to be abstract entities as well. We therefore added the type Resemblance, see example (105).

¹⁶Annotators will be provided with a list of such predicates.

Multiple type annotation. In both of the constructions described above, the introduced abstract entity itself has an SE type. In these cases, annotators are asked to additionally mark the SE type of the clause that is either embedded as a complement or modified with a factive or propositional adverb. For example, (106) shows case where the clausal complement of *know* describes an EVENT.

```
(106) I know (State) that she said something. (Fact, Event)
```

We treat cases in which the ABSTRACT ENTITY is expressed by an interrogative complement clause (i.e., clauses introduced by one of the question words *who*, *what*, *how* etc.) the same way. The *content* of the Fact or Proposition may not be known to the reader in such a case. The fact that something expressed by a clausal complement is *known* by the subject of *know* makes the clausal complement a Fact (see example (107)).

```
(107) I know (State)
what she said. (Fact, Event)
```

Abstract entities are usually found as

- clausal complements licensed by certain predicates;
- clauses containing adverbs expressing degree of certainty or likelihood.

In addition to marking abstract entities as either FACT, PROPOSITION or RESEMBLANCE, please annotate the type of the embedded situation.

Examples: Licensing constructions for Facts. 17

FACTS are underlined.

learn	Sherry learned through our Future Works class
	that she could rise out of the welfare system.
be clear	It's clear
	that the balance of power is shifting again.
be sure	AOL would never have existed, if it had been founded here,
	I am sure.
regret	She regretted that she had not kept her mouth shut.
realize	He realized that she regretted it.

¹⁷We thank Ruth Kühn for collecting these.

Examples: Licensing constructions for Propositions.

Propositions are underlined.

think	I think that is why I welcomed doing a book.	
	personally always thought	
	it was a risk to pin the image of Microsoft on just one man.	
	I think it almost does not matter	
	since I think it was just that insularity that helped them	
guess	I guess she went to the supermarket.	
imagine	and I imagine it did not seem possible to people living then	
feel	Will Shuster, who felt that the fiesta had become dull and commercialized.	
believe	Lovelace believed that intuition and imagination were critical to	
	effectively applying mathematical and scientific concepts.	

4.2.5 Speech Act Types

Speech acts can be defined as utterances which perform some function beyond the communication of information (Austin, 1962; Searle, 1969). Some utterances both introduce a situation and perform a speech act; in this case we are concerned only with two types of speech acts which are purely performative: Questions and Imperatives.

QUESTIONS and IMPERATIVES in English are usually easily identifiable by their surface form. IMPERATIVE clauses (108) for the most part issue commands or requests, or offer advice. In their most common form, English imperative constructions omit the subject and begin with the verb, in its bare infinitive form. Generally the subject is understood to be the (second-person) addressee(s).

(108) *Stay calm.*

Head for the hills!

Take a deep breath.

Don't worry about it.

Mix the batter until it's smooth.

QUESTIONS in English are formed via inversion and thus usually begin with an auxiliary or modal verb (109). Usually they end with a question mark.

(109) Is John a teacher?

Has John ever taught?

Wouldn't John be a good teacher?

Do you think John would be a good teacher?

Did John teach during graduate school?

Questions can also be posed indirectly or embedded in reported speech (110); for this project, such cases are not considered to be speech acts.

(110) I wonder if it will rain today.

He asked me whether I would like to sing.

4.3 Derived Situation Entity Types

In some cases, the situation entity type of a clause (i.e. Event, Report, State, both types of General Stative, Fact, Proposition, Question, and Imperative) changes based on the addition of some linguistic indication of uncertainty about the status of the situation described, or by the addition of linguistic information (specifically, the English perfect) focusing on the post-state of some event. We refer to these as **derived situation entity types**.

The linguistic features which can trigger this change in situation entity type are negation, modality, future tense, conditionality, and sometimes subjectivity. When these features appear in a clause that would otherwise be marked as Event, it causes coercion to the category of State. For the other situation entity types, the coercion does not occur.

4.3.1 Perfect tense

Clauses in present perfect and past perfect are considered to be states, as they focus on the circumstances of an action's being completed at the time of reference (Smith, 2003, p. 71). This perspective is also supported by Katz (2003). Contrasting examples are given in (111).

- (111) (a) She left at that moment. (EVENT)
 - (a) At that moment, she was already leaving. (EVENT)
 - (c) She has left. (STATE)
 - (d) She had left. (State)

Please note that in general, *all* States should receive the habituality value of **static**, however, in the case of Events that were coerced to States via the perfect, and that describe events that actually took place (no other coercion may apply), you can mark the clause as both **episodic** and **static**. If for example a modal is included as in (112), the clause can only be marked **static**.

(112) She might have left.

This definition leads to a particular case of STATE having a **generic** main referent. Consider example (113).

(113) Coffee has replaced the ubiquitous tea. (generic main referent, State)

Please not that clauses with generic main referent can only be marked as STATES if no additional coercion (as listed in the sections below) occurs, i.e. if the event has in fact taken place, and the only coercion is perfect. In the case of a dynamic aspectual class, you can mark the clause as both static and episodic, if you feel that it really also tells that an event has taken place. For all other cases please see the sections below.

4.3.2 Negation

An earlier, to date unpublished annotation project (Shore and Palmer) explicitly introduced the notion of considering negated Events to be States. Negated events do not in fact take place, but rather such clauses (e.g. (114)) introduce the possibility of an event. Negated eventive clauses are considered to be States when predicating over a specific entity, but as Generic Sentences when predicating over a class of entities. When negating generalizing sentences (as in (115)) or generic sentences (as in (117)), the derived situation entity type stays the same.

- (114) John did <u>not</u> win the lottery. (negated Event \rightarrow State)
- (115) John does <u>not</u> drive to work. (Generalizing Sentence)
- (116) Our cat <u>never</u> hunts birds in our backyard. (Generalizing Sentence)
- (117) Whales are not small. (Generic Sentence)

4.3.3 Future and modals

In addition to negated events, events in future tense denoted by constructions such as *will* or *shall* and modal constructions using *can*, *could*, *would* or modal adverbs expressing uncertainty are marked as STATES, as they are considered to describe the 'possibility' of an EVENT.

Again, future clauses with generic main referents are not marked as STATES (see (120) and (121)).

- (118) John will move to his own place next week. (State (specific event in the future))
- (119) John could move to his own place. (possible event \rightarrow State)

¹⁸This coercion is sometimes indicated in Smith (2003), but it is not explicitly addressed in that text.

- (120) Whales will not be extinct in 100 years. (Generic Sentence)
- (121) John <u>will</u> drive to work from next week on. (Generalizing Sentence: a repeated action in the future)
- (122) It is now clear (state)
 that the whole of Eritrea <u>could</u> become a battlefield. (State)
 (from Smith (2003))

In addition, adverbials such as *probably*, *likely*, *certainly* can transform a situation entity into a State, as they show the speaker's belief regarding the likelihood of the situation.

(123) John probably argued with his parents yesterday. (State)

Again, this does not apply to general statives:

(124) Mickey probably paints houses. (Generalizing Sentence)

However, note that modals like *may*, *can*, *could* and *must* express ability, possibility or necessity rather than the fact that something is done regularly, and are marked as STATES, see (125).

(125) Mickey can/may/could/must paint houses for the rest of his life. (static, State)

GENERIC SENTENCES are marked as *static*, but they are not coerced to another situation entity type, as shown in (126).

(126) Kangaroos can/may/could/must jump all the time. (static, Generic Sentence)

4.3.4 Conditionals

Examples (127) and (128) introduce yet another case of derived situation entity types. In (127), the condition clause and the consequence clause introduce possible EVENTS, and can hence be seen as STATES. The counterfactual conditional introduces an EVENT which did not take place (John winning the lottery) and a STATE which isn't the case (John not living with his parents) to the discourse. Both of them can be seen as STATES, as they describe possibilities of EVENTS or STATES.

(127) If you finish your homework by 3pm, (State) we can go swimming. (State)

- (128) If John had won the lottery, (STATE) he would not be living with his parents. (STATE)
- (129) If whales were smaller, (Generic Sentence),
 I would have one in my aquarium. (State)
- (130) If our cat chased mice more often, (Generalizing Sentence)
 I would go crazy. (State)
- (131) I don't know (STATE) whether Saarland is in Germany. (STATE)
- (132) I asked him (EVENT) if he would sing. (STATE, example taken from Smith (2003), p. 14, text 6)

The consequence clause of example (129) describes a possible State, and is hence a State. The condition clause (*If whales were smaller*) also has the situation type of a GENERIC SENTENCE.

Similarly, (130) shows a conditional clause with the basic type of a Generalizing Sentence. The conditional clause makes a general statement about a specific individual, and is hence also a Generalizing Sentence.

Consider examples (131) and (132): Following the argumentation from above, we consider possible Events or States to be States, and sentences making statements about kinds as Generic Sentences, no matter whether they describe an actual action or state of the kind or just a possible action or state. Example (132) shows that Smith (2003) applied the same idea.

Subjectivity Consider example (133). Although this could be seen as a Proposition, we consider this sentence to be a state. The sentence is clearly subjective, but situation entities are classified as the speaker sees them (and the speaker of (133) considers her proposition that John should move in with his girlfriend a state), not what they objectively are (following Moens and Steedman, 1988). However, if licensed by a predicate such as *believe* in (134), the same clause is considered as a Proposition.

- (133) John should move in with his girlfriend. (State)
- (134) I believe that John should move in with his girlfriend. (Proposition)

4.4 Summary

This section intends to give a helpful overview for the decision process during annotation. Utterances differ with respect to how easy it is to determine their SE type, and in some cases it is easier to find values for some features than it is to arrive at a final determination of SE type. For this reason, we recommend first considering all "easy" cases (described below), then proceeding to the difficult cases, taking a features-first, labels-second approach.

Step 1: Identify Speech Acts

These types include questions and imperatives, and (at least in English) they are easily recognizable by their surface linguistic form. For more detail, see Section 4.2.5. If you can label a clause as a QUESTION or IMPERATIVE, there's no need to look further.

- (135) Why can't we teach our children to read, write and reckon? (QUESTION)
- (136) Make your due-process system visible. (Imperative)

Step 2: Identify Abstract Entities

FACTS, PROPOSITIONS and RESEMBLANCES – the three SE types within the category of Abstract Entities – are also relatively easy to identify based on their surface form. There are three different types of constructions to look for.

Embedded complement clause. First, these SE types often occur as complement (i.e. obligatory argument) clauses to certain types of predicates, as in (137) (Peterson, 1997). Annotators will be provided with a list of factive and propositional predicates: look for full clauses following such predicates, with or without the complementizer *that*.

(137) Simone knew that the position would be offered to her. (FACT) Simone believed she was a good candidate. (Proposition) Simone looked as if she had seen a ghost. (RESEMBLANCE)

Remember to mark the SE type of the embedded or modified clause.

Step 3: Determine the feature values

- (a) **Genericity of main referent** Does the clause say something about
 - a particular individual/organization/situation/process...? → specific

- a kind/class? → generic
- some kind of general concept? → **generic**
- (b) **Aspectual class** Does the verb constellation describe
 - a event or a series of events? → **dynamic** verb constellation
 - some kind of state? → **stative** verb constellation
- (c) **Habituality** Does the clause describe
 - a one-time event? \rightarrow **episodic**
 - a regularity of events/states that happen/are taken on repeatedly?
 - \rightarrow habitual/regularity
 - the state of a specific individual, valid always or for a certain duration?
 - \rightarrow static

Step 4: Identify the SE type

Using the features identified in Step 3 and Table 1 or Figure 3, you can determine the situation types Event, State, General Stative and its two subtypes, General Sentence and Generalizing Sentence. If all feature values can be determined unambiguously, the situation type should be clearly identifiable as well. When looking for a valid feature combination for one of the SE types, the table is to be read from top to bottom. General Stative is only to be assigned without a subtype if the subject NP type cannot be determined.

situation type	main referent	aspectual class	habituality
EVENT	specific	dynamic	episodic
	generic		
State*	specific	stative	static
GENERIC SENTENCE	generic	dynamic	habitual
		stative	static or habitual
GENERALIZING	specific	dynamic	habitual
SENTENCE		stative	
GENERAL STATIVE	specific	dynamic	habitual
	generic		

Table 1: Situation types and their features. *Other feature combinations are possible for State, see Section 4.3 (derived SE types).

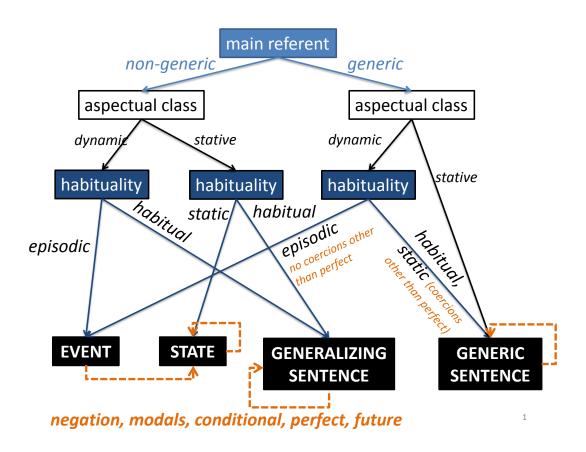


Figure 3: Decision tree: SE types and features

A Annotation Scheme for German

A.1 Introduction

The guidelines regarding the annotation of the situation-related features (genericity of main referent, aspectual class, and habituality) apply for the German annoatation scheme, as well. In the following sections we present cases that appear in German, and which do not occur in the English scheme. In Section A.2, we describe how to mark the genericity of the main referent in German, especially in constructions in which the main referent is not the grammatical subject of the clause. In Section A.3, we discuss the German-specific derived situation entity types. Please note that the same rules that apply for the derived types in English are to be followed for the German as well. The derived situation entity types that we describe in Section A.3 are simply additional types for the German part. Moreover, we introduce a new subtype of Event, namely the Event-Perf-State to mark clauses containing perfect tense, that cannot be described neither as State nor as Event only.

A.2 Genericity of main referent

As we have already seen for English, the question to be answered in order to find the main referent of a clause is what the sentence is about. We have seen in the English part that the main referent is usually realized as the grammatical subject of a clause. However, this is not always true and such cases are more frequent in German. One such example is the impersonal passive, which can be formed in German (in contrast to English) for intransitive verbs. During the formation of the passive clause, the pronoun "es" serves as a placeholder for the subject at the beginning of the clause or disappears when the word order is inverted. Since there is no grammatical subject, annotators have to think about whether the main referent of the clause is a specific person/group/organization, as in (138) and (140) where the implied main referent is a specific group of people, or a kind or abstract individual/entity, as in (139), where the implied main referent are all people who sew in general.

- (138) Jetzt ist Pause, (non-generic, State) es wird wieder geredet. (non-generic, Event)
- (139) Früher gab es <u>keine Nähmaschinen</u>, (generic, Generic Sentence) heute wird anders genäht. (generic, Generic Sentence)
- (140) Abends wird im Gemeinschaftsraum gespielt, getrunken und gefeiert. (non-generic, Generalizing Sentence)

Please note that, when a clause containing an impersonal passive starts with the pronoun "es", like in example (138), it should **not** be marked as expletive. "Es" has different usages and appears very often as an impersonal pronoun in German. Thus, annotators have to keep in mind what the clause is about. In examples (141) and (142), the pronoun functions as a placeholder and can be omitted, the main referents are "ein Bild" and "der Flieder" respectively.

- (141) Es hängt ein Bild an der Wand. (non-generic)
- (142) Es blühte <u>der Flieder</u>. (non-generic)

There is a group of impersonal verbs that express the perceptions of a person, which are usually expressed with stative verbs, and require an argument in dative, as in (143), or accusative, as in (144). In both cases, the argument in dative or accusative is considered to be the main referent of the clause.

- (143) Es gruselt mir vor dir. (non-generic)
- (144) Mich friert es. (non-generic)

Similarly, when the impersonal pronoun "es" appears with verbs that express existence, the main referent of the of the clause is the object or person the verb is referring to. This is analogous to the existential clauses in English.

- (145) Derzeit gibt es viel zu tun. (non-generic)
- (146) Es war einmal ein Drache. (non-generic)

Annotators are asked to annotate the pronoun "es" as **no-main-referent** in clauses where no main referent can be identified, for example with verbs expressing natural phenomena and especially those describing the weather, where the pronoun is obligatory and is the grammatical subject of the verb.

- (147) Es grünt und blüht in Wald und Flur. (no-main-referent, State)
- (148) Es regnet. (no-main-referent, State)
- (149) Es ist kalt. (no-main-referent, State)

Another interesting case for German is the indefinite pronoun "man". Like the English words people, they, one etc., "man" can refer to a specific group of people, like in example (150), or to abstract individuals or entities, like in example (151).

- (150) Jetzt ist Pause, man redet wieder. (non-generic, Event)
- (151) Heute näht man anders. (generic, Generic Sentence)

A.3 Derived Situation Entity Types

A.3.1 Perfect Tense

In English, we consider clauses in present perfect and past perfect to be states, as they focus on the circumstances of an action's being completed at the time of reference. In German, however, this rule does not apply since it is not strictly defined when to use the present perfect (*Perfekt*) or the simple past (*Präteritum*). These two tenses can be used interchangeably (especially in spoken language) to describe events that were completed in the past, as in (152), (153) and (154).

- (152) Schiller hat "die Räuber" im Jahre 1781 geschrieben. (EVENT)
- (153) Gestern sind wir alle ins Kino gegangen. (EVENT)
- (154) Ich möchte allen danken, die so intensiv an diesem Thema mitgewirkt haben. (EVENT)

Like in English, however, clauses in present perfect and past perfect can also be considered to be states in German, as they focus more on the outcome or the state reached after a process rather than the process itself. In example (155), the interpretation of the clause is that there is now snow around, in example (156) that the speaker is not hungry and in example (157) that the person was aswoon when she was found.

- (155) Guck mal, es hat die ganze Nacht geschneit. (State) (In contrast: Es schneite die ganze Nacht. (EVENT)
- (156) Ich habe schon gegessen. (State)
- (157) Sie hatte das Bewusstsein verloren, (State) als ihre Retter sie endlich befreiten. (Event)

In contrast to English, it is not clear in German where the boundary lies between State and Event in clauses containing perfect tense. In many cases, it is a matter of underspecification that makes it difficult to decide whether the clauses express State or Event and this is the reason we have introduced a new subtype of Event, the Event-Perf-State. When the clause is underspecified and annotators feel that a clause is not primarily a State neither primarily an Event, they are asked to mark them with the Event-Perf-State subtype.

- (158) Das scheint ein wichtiges Thema zu sein, und ich sage Ihnen das als jemand, der sich selber damit beschäftigt hat. (Event, Event-Perf-State)
- (159) Ich habe die Stelle! Sie haben mir den Job gegeben. (Event, Event-Perf-State)
- (160) Bitte, setzen Sie sich, ich habe schon zwei Bier bestellt. (Event, Event-Perf-State)

A.3.2 Subjunctive

There are two different types to express the subjunctive mood (*Konjunktiv*) in German: The Subjunctive I (based on the present tense) and the Subjunctive II (based on the past tense). Contrary to the indicative, the subjunctive is used to express doubt, possibility, speculations, conditionality and other "unreal" conditions, and in this case Events are marked as States. General Statives in subjunctive mood are not coerced to States.

- (161) Hätten wir das Geld, (State) gingen wir morgen schon im Urlaub.(conditionality \rightarrow State)
- (162) Ein plötzliches Zerreissen eines Tisches und sauberes Zerspringen eines Brotmessers habe er beobachtet. (doubt \rightarrow State)
- (163) Sie sagte, (Report) sie tue das regelmäßig um Menschen in Not zu helfen. (Generalizing Sentence)

A.3.3 Statal passive voice

The statal passive, in contrast to the processual passive, focuses on the result or the "state" reached after a process. Clauses in statal passive are therefore marked as STATES.

- (164) Die Tür ist geöffnet. (State)
- (165) Der Brief ist geschrieben. (State) (In contrast: Der Brief ist geschrieben worden. (Event)
- (166) Sein Herz war für Melanie Nowara entflammt. (State)

A.3.4 Final clause with "damit"

Final clauses are used to describe a purpose, an intention or a goal. Clauses starting with "damit" that contain Events are coerced to States, as they are considered to describe the possibility of an Event.

- (167) Es wird Druck ausgeübt, (EVENT) damit die Pharmaindustrie neue Impfstoffe entwickelt. (State)
- (168) Erinnere mich nochmal, (Imperative) damit ich pünktlich komme. (State)
- (169) Damit sie pausenlos arbeiten, (Generalizing Sentence) werden sie ständig überwacht. (Generalizing Sentence)

A.3.5 Modal constructions

Modality, as we have already seen in the English scheme, can trigger a change in the type of the situation entity. It is not only expressed through the modal verbs (in German: müssen, sollen, dürfen, können, mögen, wollen) but also through various grammatical constructions. When these constructions appear in clauses that would be normally marked as Events, the situation entity type is coerced to State. Two common structures are described in this section.

habenl sein + zu + *infinitive*. Phrases containing this construction are used to denote necessity or possibility and, thus, are marked as STATES.

- (170) Sie haben es zu unterlassen, vertrauliche Informationen weiterzugeben. (necessity \rightarrow State)
- (171) Sie hat etwas zu verschenken. (possibility \rightarrow State)
- (172) Diesem Kommentar ist nichts hinzuzufügen. (possibility \rightarrow State)
- (173) Die Wohnung ist beim Auszug zu renovieren. (necessity \rightarrow State)

sich lassen + *infinitive*. This construction is one of the alternatives for the passive voice. It can be analyzed with the semantics of the verb "can".

- (174) Den Beteiligten muss klar sein, (State) dass sich dieser Konflikt ohne Kompromisse lösen lässt. (possibility \rightarrow State)
- (175) Der Umweltskandal ließ sich verbergen. (State)

As we have seen before, this does not apply to general statives.

- (176) Geschickt lassen sich jetzt aus Gegnern Freunde machen. (Generic Sentence)
- (177) Mit dieser Software lassen sich täglich Millionen von Tweets auswerten. (Generalizing Sentence)

Note that it should not be confused with another *sich-lassen* construction, which requires an animated subject and has the meaning of "allow", as in (178) or "prompt", as in (179).

- (178) Anna ließ sich nicht unter Druck setzen. (= Anna ließ nicht zu, dass ...). (Event)
- (179) Er ließ sich die Haare schneiden. (= Er veranlasste, dass ...). (EVENT)

References

- John Austin. How to Do Things With Words. Harvard University Press, 1962.
- Emmon Bach. The algebra of events. *Linguistics and philosophy*, 9(1):5–16, 1986.
- Gregory N Carlson. Generics, habituals and iteratives. *The Encyclopedia of Language and Linguistics*, 2005.
- Gregory N Carlson and Francis Jeffry Pelletier. *The generic book*. University of Chicago Press, 1995.
- Hana Filip. Aspect, eventuality types and nominal reference. Taylor & Francis, 1999.
- Annemarie Friedrich and Alexis Palmer. Situation entity annotation. *LAW VIII*, page 149, 2014.
- Valentine Hacquard. On the interaction of aspect and modal auxiliaries. *Linguistics and Philosophy*, 32(3):279–315, 2009.
- Nancy Ide, Collin Baker, Christiane Fellbaum, and Charles Fillmore. Masc: The manually annotated sub-corpus of american english. 2008.
- Graham Katz. On the stativity of the english perfect. *Perfect explorations*, pages 205–234, 2003.
- Angelika Kratzer. Stage-level and individual-level predicates. *The generic book*, 125:175, 1995.
- Manfred Krifka, Francis Jeffry Pelletier, Gregory Carlson, Alice ter Meulen, Gennaro Chierchia, and Godehard Link. Genericity: an introduction. *The generic book*, pages 1–124, 1995.
- Marc Moens and Mark Steedman. Temporal ontology and temporal reference. *Computational linguistics*, 14(2):15–28, 1988.
- Philip L. Peterson. Fact Proposition Event. Kluwer Academic Publishers, 1997.
- Randolph Quirk and Sidney Greenbaum. *A concise grammar of contemporary English*. Harcourt Brace Jovanovich New York, 1973.
- Hans Reichenbach. *Elements of symbolic logic*. Dover Publications, 1980.
- John Searle. Speech Acts. Cambridge University Press, 1969.

Todd Shore and Alexis Palmer. Situation entity segmentation and annotation guidelines, version 1.0.

Carlota S Smith. *The Parameter of Aspect*, volume 43. Kluwer Academic Press, 1991.

Carlota S Smith. *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press, 2003.

Radu Soricut and Daniel Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 149–156. Association for Computational Linguistics, 2003.

Zeno Vendler. Verbs and times. The philosophical review, 66(2):143–160, 1957.