

# Guidelines for Manual Event Detection and Classification in Historical Texts

Rachele Sprugnoli  
Fondazione Bruno Kessler and University of Trento  
`sprugnoli@fbk.eu`

January 2018

In our guidelines we adopt a wide definition of event by referring to the notion of eventuality<sup>1</sup> defined by Bach [2008] and re-elaborated in [Dölling et al., 2014] (see Figure 1) and, from the linguistic point of view, by taking into consideration different parts of speech and syntactic constructions. In addition, we define a set of 22 semantic classes with the aim of providing an exhaustive categorisation of events, thus overcoming the limited classifications proposed by other initiatives such as ACE and Rich ERE.

## 1 Event Annotation Guidelines

In this Section we detail the annotation guidelines designed to detect and classify event mentions in texts.

### 1.1 Event Linguistic Realization

Syntactically, the linguistic elements which may realize an event are the following:

- verbs in both finite and non-finite form;

(1) *she **expected** to be **attacked***

- past-participles in the nominal pre-modifier position that represent resultatives events. Interpreted as a state, the following example can be paraphrased as “the state of having been imprisoned”;

(2) *an **imprisoned** criminal*

---

<sup>1</sup>In this chapter we use the terms “event” and “eventuality” interchangeably.

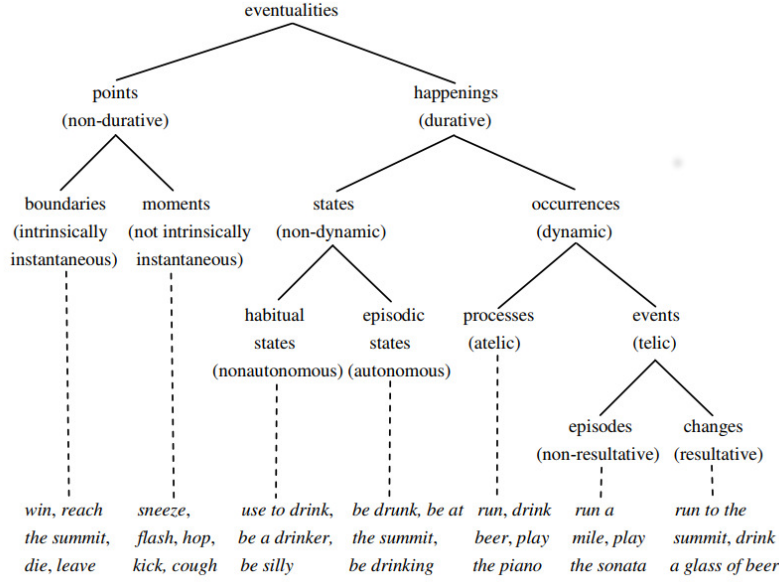


Figure 1: Taxonomy of eventualities, image taken from [Dölling et al., 2014].

- present-participles in the nominal pre-modifier position that represent in-progress events. In the following example, the modifier describes an event in progress so that it can be paraphrased as “the audience that is smiling and applauding”;

(3) *a **smiling** and **applauding** audience*

- adjectives in predicative position;

(4) *the museum itself **was damp***

- nouns which can realize eventualities in different ways:

- deverbal nouns denoting an activity or an action;

(5) *the **running** of these ferries*

- nouns which have an eventive meaning in their lexical properties even if they not derive from verbs;

(6) *delegates of Russia against the **war***

- post-copular nouns;

(7) *it **was a lie***

- nouns which normally denote objects but which are assigned an eventive reading either through the process of type-coercion Pustejovsky [1991], or through the processes of logical metonymy and coercion induced by temporal prepositions.

(8) *I am finishing this **letter** rather hurriedly*

- pronouns related to previously mentioned events.

Differently from the Rich ERE annotation, we do not annotate implied events indicated by nouns like *murderer* and *protestor* so to make a clear distinction between events and entities and avoid confusion.

The factuality status of events does not impact on the annotation: all events have to be annotated whether they are presented as a fact, a counterfact or a possibility.

(9) *here we **saw** all manner of beautiful and hideous creatures*

(10) *Professor Pais wrote a paper on this which I have not **seen***

(11) *one may **see** small towns*

## 1.2 Event Extent

Eventualities have different extents: the annotation of single-token, multi-token and discontinuous expressions is allowed as detailed below.

**Finite and non-finite verb forms:** annotate only the verbal head without auxiliaries of any form (multiple, modal, negative).

(12) *you **wish** to **know*** = 2 annotations

(13) *having been **destroyed** by the father*

(14) *we could **appreciate** to-day*

(15) *I do not **understand** Beloch*

**Phrasal verb constructions:** the main verb should be annotated together with the particle and/or the preposition forming the phrasal verb because they form a single semantic unit whose meaning cannot be understood by looking at the meaning of each single part. In case the verb and the preposition are separated having the direct object in the middle, as in (17), a discontinuous annotation should be performed.

(16) *I abjectly **stepped into** his cab* = 1 annotation

(17) *he would have **carried it off** to France* = 1 annotation

**Light verbs:** the whole predicate formed by the main verb and the following expression, usually a noun, is to be annotated even if not continuous.

(18) ***make her a visit*** = 1 annotation

(19) ***get a snap-shop*** = 1 annotation

**Copular constructions:** the literature den Dikken and O’Neill [2016] distinguishes different types of copular constructions on the basis of a taxonomy of four copular elements: (i) support copula; (ii) predicational copula; (iii) equative copula; (iv) silent copula. The first two cases are to be annotated with a multi-token span including both the copula and the whole copula complement (20, 21). As for equatives, whose linguistic status is unclear Mikkelsen [2005], only the copula should be annotated (22). No annotation is provided in case of silent copula, given that the linguistic realization of the copula is missing: in (23) only the main verb is annotated.

(20) *Our welcome to Genoa **was not cheerful***= 1 annotation

(21) *Propertius **was a contemporary** of Virgil*= 1 annotation

(22) *Dr. Jekyll **is** Mr. Hyde*

(23) *Petrarch **considered** this tomb (to be) sufficiently important to plant a laurel*

Inverse copular constructions Moro [1997] should be annotated as well, taking into consideration the reversed word order from the canonical *subject-copula-predicative expression* to *predicative expression-copula-subject*:

(24) ***an interested onlooker was** former Coroner Gustav Scholer*= 1 annotation

Please note that the verb *to be* is not the only English copula to take into account<sup>2</sup>:

(25) *I **felt strange***= 1 annotation

**Periphrastic causative constructions:** are composed of a causative verb such as *cause*, *get*, *have* or *make* combined with another verb to express causation Kemmer and Verhagen [1994]. These two verbs should be annotated separately.

(26) *urging him to **make** his brother **drive** more carefully*= 2 annotations

**Fixed expressions:** phrases, idioms, nominal expressions whose meaning cannot be understood from the individual meanings of their elements have to be annotated as a unique mention.

(27) *in order to **get rid of** him*= 1 annotation

(28) *I would not **have a leg to stand on***= 1 annotation

(29) *a hostile **air raid** this evening*= 1 annotation

**Nouns:** can be annotated within a multi-token or discontinuous expression if part of a copular construction, a light verb construction or a fixed expression. In addition, also named events such as “First World War” can have a multi-token extent. In all the other cases, the noun itself should be annotated alone, without including determiners or adjectives.

(30) *both in **peace** and **war***= 2 annotations

<sup>2</sup>A list of English copulae is available on Wikipedia: [https://en.wikipedia.org/wiki/List\\_of\\_English\\_copulae](https://en.wikipedia.org/wiki/List_of_English_copulae).

### 1.3 Semantic Classes

Each annotated event mention should be classified by assigning a value to the **CLASS** attribute. The classification we have designed is based on semantic criteria and developed by re-elaborating the semantic categories of the Historical Thesaurus of the Oxford English Dictionary (HTOED) Kay et al. [2009] . The HTOED has been defined over several decades with the aim of conceptualising and classifying the meaning of the English language: it consists of a hierarchical structure made of a primary tripartite division (*External World*, *Mental World*, and *Social World*), 37 categories and 377 sub-categories<sup>3</sup>.

In HTOED a distinction is made between categories connected with a physical existence and those having a social dimension: due to this subtle difference an event of movement can belong to the **TRAVEL AND TRAVELING**, the **SPACE** or the **MOVEMENT** category. In other words, discerning between physical and social dimensions is ambiguous. Therefore, starting from the original complex and extremely fine-grained classification, we worked to find an appropriate level of granularity by merging categories with a common conceptual core. This choice led us to create a unique class for events related to the concept of space (**SPACE-MOVEMENT**) and for those involving forces beyond scientific understanding or the laws of nature (**RELIGION-SUPERNATURAL**). In addition, we collapsed into the same class events in the area of production and trade of services and goods (**ECONOMY**), those in the public domain (**LAW-AUTHORITY**), and those involving all the types of living things and their health conditions (**LIFE-HEALTH**). Events connected to the faculties of the mind characterized by reasoning or knowledge are brought together in the **MENTAL-ABSTRACT** class, while instinctive or intuitive mental activities accompanied by a certain degree of pleasure or displeasure are joined in the **EMOTIONS-EVALUATIONS** class. Figure 2 provides a graphical representation of how our classes were defined starting from the HTOED categories.

---

<sup>3</sup><http://historicalthesaurus.arts.gla.ac.uk/>

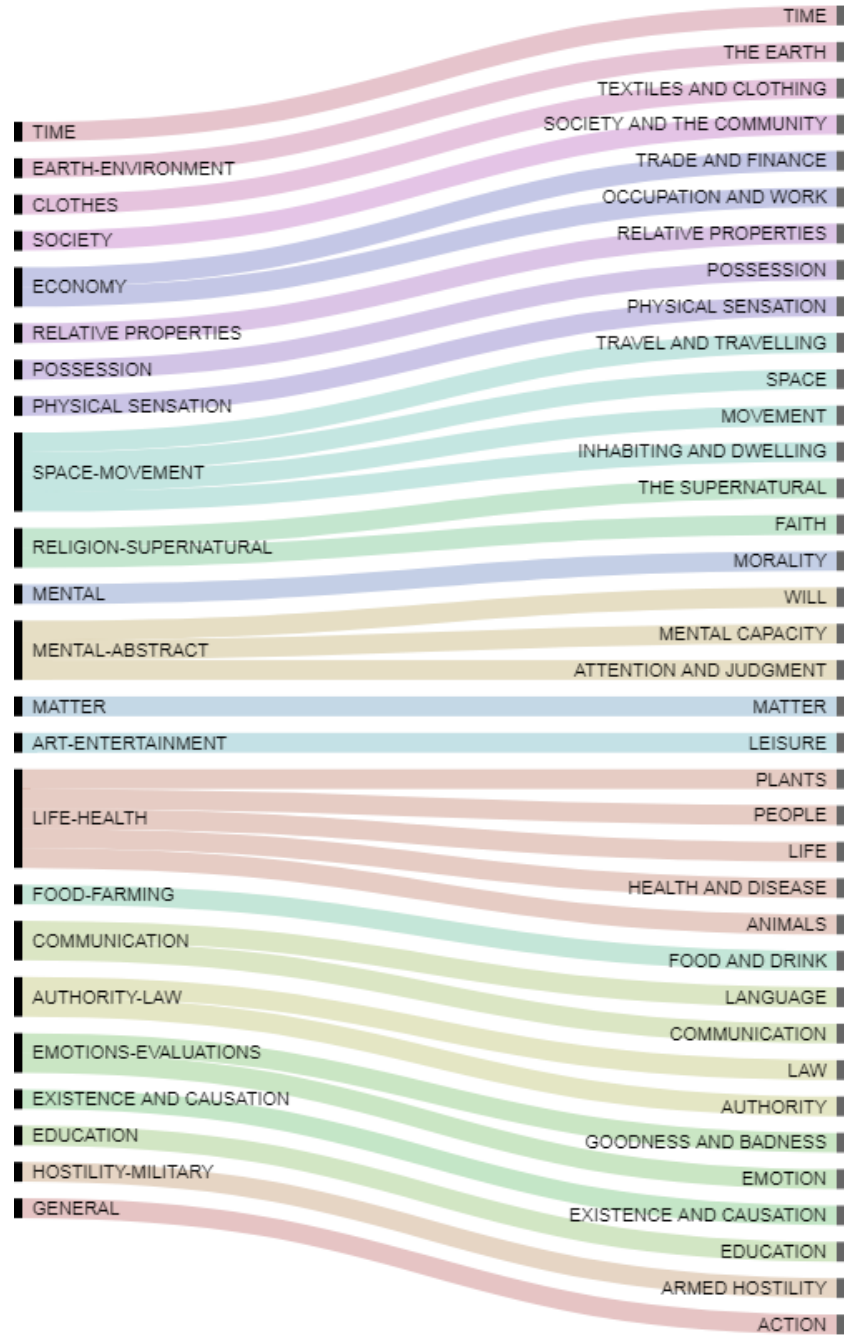


Figure 2: Mapping of our HISO classes (left) to the second-level HTOED categories (right).

### 1.3.1 Description of Semantic Classes

Classes are described below together with a set of examples. Event extension is highlighted in bold.

1. EARTH-ENVIRONMENT, eventualities related to geography (31, 32), climate/weather conditions (33), environmental issues (34).

- (31) *the streets **are like** caverns*
- (32) *the settings **are** Spanish*
- (33) *It has been **raining** for days*
- (34) ***deforestation** has denuded the mountain-side*

2. LIFE-HEALTH, eventualities related to living things, i.e. humans (35), animals and plants (36), including life (37), death (38), physical conditions, diseases and medical treatments (39).

- (35) *he **was a** Caprian paesant*
- (36) *oranges do not **grow up***
- (37) *which **lives upon** public charity*
- (38) *he will not **kill** the woman*
- (39) *in charges of contagious **diseases***

3. FOOD-FARMING, eventualities pertaining to food (40), food preparation and consumption (41), drink (42), agriculture (43) and hunting (44).

- (40) *they **are only** nuts*
- (41) *let us **breakfast** together*
- (42) *I luxuriously **sip** my coffee*
- (43) *an elderly man was **plowing** with a pair of oxen*
- (44) *this Diana **is not a** huntress*

4. CLOTHES, eventualities associated to textiles (45), clothes (46, 47) and other personal belongings (48).

- (45) *they are renowned for their skill in **weaving***
- (46) *he **took off** his hat*
- (47) *he **wore** only a shirt*
- (48) *it **is a rather heavy** portmanteau*

5. MATTER, eventualities connected to substances and materials, their properties, constitution and conditions (49). This class includes terms relating to liquids (50), solids, gases, electricity (51), light (52), colours (52), shapes.

- (49) *burdens that **seem too heavy***
- (50) *the miracle of **liquefaction***
- (51) *the power **transmitted** need not be necessarily destructive*
- (52) *their full black hair **shines** like satin*

(53) *his faces **being rather red***

6. **EXISTENCE-CAUSATION**, eventualities relating to the concepts of being as in existential clauses McNally [1998] (54), occurring (55), existing and causation (56) and their lack. It includes both creation (57) and destruction (58), damage (59), break and demolition.

(54) *in this court **are** a number of handsome sarcophagi*

(55) *these **occurrences** are fanning a spirit of revenge*

(56) *the cases **caused** me a genuine thrill*

(57) *a cloud is **occasioned** by the column of steam*

(58) *sudden **destruction** of the buried city*

(59) *the regular ambulance was **wrecked** last night*

7. **SPACE-MOVEMENT**, brings together all the eventualities pertaining to space (extensions (60), directions, presence (61)), movements (motions of entire bodies (62) and part of bodies, objects (63) but also changes of place, transfers, impacts), lack or end of movement (various stages of inactivity, such as stillness (64), stops (65), waiting), and travel (referring to ways of travelling on land (66), by water (67), by air).

(60) *mitre of gold is **covered** with precious gold*

(61) *is there any significance in the **presence** of the Mayor?*

(62) *he had **walked through** the shaded park*

(63) *the waters of two fountains **mingle** and **flow** together*

(64) *the Temple of Minerva **standing** beside twelfth-century buildings*

(65) *the little inn at which we are **stopping***

(66) *after our exciting **drive***

(67) *we **sailed** from New York six weeks ago*

8. **TIME**, eventualities associated to frequency (68), duration (69), change (70), age (71) and the spending of time (72). Aspectual terms (73) are included in this class because they denote distinct parts of the internal temporal structure of eventualities Damova and Bergler [2000].

(68) *this **is the first time** the cup will leave France*

(69) *the raid **lasted** for about half an hour*

(70) *abrupt **changes** of temperature*

(71) *I'm **only 20 years old***

(72) *she **spent** some weeks at Sydney*

(73) *for the **beginning** of our drive*

9. **GENERAL**, general eventualities denoting not specific operations upon something like doing, using (74), trying (75), helping (76), finding but also events and states relating to safety/danger (77), difficult/easiness, success/failure.

(74) *the very best tobacco **used** in the cigar factories*



- (75) *a man who is **trying** to free himself*  
 (76) *with the **help** of a band*  
 (77) *Doctor Antonio hesitated about **imperilling** her neck*
10. **RELATIVE PROPERTIES**, eventualities pertaining to measurements (78, 79), numbers (80) (except those relative to the temporal dimension that have to be annotated with the **TIME** class) and quantities (81).
- (78) *this island of Luzon **is so large***  
 (79) *he **counted** the guns which were fired*  
 (80) *because of a **reduction** in their wages*  
 (81) *these Paris delegates **are thirty-five***
11. **RELIGION-SUPERNATURAL**, eventualities relating to religions (government of organized religions (82), sects, religious ceremonies (83) and worship in general) and the supernatural (84) (occult, paranormal, supernatural manifestations, deities).
- (82) *Mr. Zeglen **is a young priest***  
 (83) *the **high mass** is **celebrated***  
 (84) *the departed **haunt** the silent town*
12. **MENTAL-ABSTRACT**, includes all mental actions and processes (85, 86) (reasoning, thinking, believing, knowing, understanding, remembering), attention and judgment (87), expressions of will (88) (necessities, inclinations, intentions, decisions and motivations). The lack of mental capacities, attention and free will is included as well in this class (89).
- (85) *Victor Emanuel **seems to have thought** that...*  
 (86) *Sir George **had knowledge** of the traditions*  
 (87) *he could **take care** of me and himself*  
 (88) *having **decided** to meet Zephine and Angela*  
 (89) *people **were fairly ignorant***
13. **EMOTIONS-EVALUATIONS**, emotional actions, states and processes or eventualities expressing the lack of emotions (90, 91) (excitement/calmness, pleasure/suffering, compassion/indifference, courage/fear, love/hate). This class includes also other aspects of subjectivity, i.e. evaluations about goodness or badness, inferiority or importance (92).
- (90) *the days brought me **enjoyment and delight***  
 (91) *the witnesses **were amazed** at the man's **calmness***  
 (92) *the second election **was not less important** that the first*
14. **POSSESSION**, includes eventualities associated to concepts such as having, not having, losing, taking, giving, allocating, acquiring, receiving, sharing (93, 94, 95, 96) and the opposition between wealthy and poverty (97).
- (93) ***having** like it four colossal bronze lions at the base*

- (94) *my mother is **keeping** it*  
 (95) *they would not **take** a large sum of money for the experience*  
 (96) *its **withdrawal** or **supply** will bring about the same results*  
 (97) *all his neighbors would testify to his **poverty***
15. **COMMUNICATION**, linguistic actions, states and processes that is eventualities connected to both the intellectual activity of speaking a language, naming things, producing speech acts (98, 99, 100) and the social activity of expressing, transmitting and receiving information in different ways through the media (101, 102) (e.g., writing, printing and publishing books, reading books and journals, using telecommunication technologies or correspondence).
- (98) *a lovely little square **called** Acquaverde*  
 (99) ***crying out**: “ecco, ecco, signora!”*  
 (100) *she **refused** to be blindfolded*  
 (101) *any one who can **write** letters as interesting as yours*  
 (102) *your mother will remember **reading** this story to me*
16. **SOCIAL**, eventualities involving the society in general or a specific community. The class includes social actions, states and processes such as the participation, or lack of participation (103), in meetings (104) and relationships of different types: intimate, between family members (105), within groups (106) and associations.
- (103) *spend a summer month in much-advertised **seclusion***  
 (104) *the **meeting** was addressed by anarchists*  
 (105) *after her **marriage** with Lord Cleverton*  
 (106) *she will be calling me soon to **join** her*
17. **HOSTILITY-MILITARY**, eventualities related to different aspect of military life (107, 108) (operations, service, use of weapons) and acts of hostility. War and peace, attack and defense are included (109, 110).
- (107) *they shall not **conscript***  
 (108) *Tully gave this story of the **shooting***  
 (109) *America stands supremely for **peace***  
 (110) *this chapel has escaped the vicissitudes of **revolutions** and **wars***
18. **AUTHORITY-LAW**, eventualities associated to political and governmental activities (111) and in general to the exercise (112) or lack of authority (power, rule) but also to criminal activities (113) and to the legal system (114, 115) (legislation, legal power, punishments). Among criminal activities, offences against the person such as murder, manslaughter and wounding are to be annotated as **LIFE-HEALTH** while offences against the property like theft and robbery fall in the **POSSESSION** class.
- (111) *favorite candidate for the next municipal **election***

- (112) *Commander Clifford **commanding** the Pampanga*  
 (113) *they had willfully **disobeyed** the law and were **locked up***  
 (114) *during the **trial** here in Buffalo*  
 (115) *the troops were already drawn up for the **execution***
19. EDUCATION, eventualities pertaining to teaching, learning but also to the administration of educational institutions (116, 117).
- (116) *he was **graduated** from Princeton University in 1906*  
 (117) *they are not learning anything*
20. ECONOMY, eventualities connected to money (118) (change of money, payments and taxation), commerce (119) (business affairs, trading operations, buying and selling), work and employment (120) (occupations but also lack of work).
- (118) *she **sold** her pearls to **raise money** to feed the poor*  
 (119) *conditions will quickly settle and **trade** revive*  
 (120) *Larcher **is a locksmith***
21. ENTERTAINMENT-ART eventualities related to entertainment (121) (night-life, hobbies), arts (122, 123) (performing art, music, visual arts), sports (124) and games in general (125).
- (121) *to expiate his **gambling** debts*  
 (122) *not content with this they **gave a dance** that same evening*  
 (123) *this **is a curious statue***  
 (124) *I **wrestled** quite a bit*  
 (125) *the number of **games** to be **played** here will be at least three*
22. PHYSICAL SENSATION, eventualities related to the perception by senses (126, 127) (touch, taste, smell, sight, hearing) but also the sleeping/waking (128) and the cleanness/dirtiness (129). The use of cigarette or drugs is included in this class as well (143).
- (126) *the wind is scarce **felt**, though you may **hear it sighing***  
 (127) *an unexpected **glimpse** into the valley*  
 (128) *cancel all speaking engagements and **take a complete rest***  
 (129) *the Neapolitan city **is even dirtier***  
 (130) *students drank, talked and **smoked***

### 1.3.2 How to Assign the Class Value

In light verb constructions and constructions with support or predication copulae, the main meaning resides not in the head verb but in the noun or copula complement: for this reason, the class attribution is based on the meaning of the noun or adjective attached to the verb.

(131) *if we **make the trip** in an automobile* = SPACE-MOVEMENT

(132) *war-horses **were monsters*** = RELIGION-SUPERNATURAL

(133) *they **were capable*** = MENTAL-ABSTRACT

(134) *the raisers **were less numerous*** = RELATIVE PROPERTIES

As for phrasal verbs and fixed expressions, the meaning of the whole linguistic unit, and not of the single parts, should be taken into consideration to assign the correct class.

(135) *how three women **get on together*** = SOCIAL

(136) ***get back** to Chiaia by five o'clock* = SPACE-MOVEMENT

Some of the previous examples show that the same token (like “were” and “get”) can be annotated with different classes according to the context, i.e. the semantics of the other parts that make up the whole event mention. In addition, other classes could be attributed to that token when it appears outside a light verb or copular construction. The verb “to be”, for example, is annotated as belonging to the EXISTENCE-CAUSATION class when in existential clauses, often in combination with the word “there” used as pronoun, while the verb “to get” can be assigned to the POSSESSION class.

(137) *there **is** a handsome modern statue* = EXISTENCE-CAUSATION

(138) *I haven't **got** it now* = POSSESSION

The verb “to fall” is another example of the importance of the context for class attribution. It is usually annotated as belonging to the SPACE-MOVEMENT class both as a single-token verb and as part of phrasal verbs, but when referring to weather, the correct class is the ENVIRONMENT one.

(139) ***falling down** with a rattling noise* = SPACE-MOVEMENT

(140) *two bombs **fell** in the Thames* = SPACE-MOVEMENT

(141) *the rain **fell** in torrents* = EARTH-ENVIRONMENT

Similarly, “to break” usually refers to the separation of objects into pieces and is annotated with the EXISTENCE-CAUSATION class; however, when referring to an injury involving the fracture of a body part, the class is LIFE-HEALTH

(142) *the concussion of this bomb **broke** glass* = EXISTENCE-CAUSATION

(143) *suffered a **broken** leg* = LIFE-HEALTH

## References

- Emmon Bach. The Algebra of Events. *Formal Semantics: The Essential Readings*, 9(1):324–333, 2008. ISSN 01650157. doi: 10.1002/9780470758335.ch13.
- Mariana Damova and Sabine Bergler. The aspectual type begin. In *Conference TALN*, 2000.
- Marcel den Dikken and Teresa O’Neill. Copular Constructions in Syntax, 2016. URL <http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-137>.
- Johannes Dölling et al. Aspectual coercion and eventuality structure. *Events, arguments, and aspects. Topics in the semantics of verbs*, pages 189–226, 2014.
- Christian Kay, Jane Roberts, Michael Samuels, and Irené Wotherspoon. *Historical Thesaurus of the Oxford English Dictionary*. Oxford University Press, 2009.
- Suzanne Kemmer and Arie Verhagen. The grammar of causatives and the conceptual structure of events. *Cognitive Linguistics*, 5(2):115–156, 1994.
- Louise McNally. Existential sentences with existential quantification. *Linguistics and Philosophy*, 21(4):353–392, 1998.
- Line Mikkelsen. *Copular clauses: Specification, predication and equation*, volume 85. John Benjamins Publishing, 2005.
- Andrea Moro. *The raising of predicates: Predicative noun phrases and the theory of clause structure*, volume 80. Cambridge University Press, 1997.
- James Pustejovsky. The generative lexicon. *Computational linguistics*, 17(4): 409–441, 1991.