A Kind Introduction to Lexical and Grammatical Aspect, with a Survey of Computational Approaches

Annemarie Friedrich[1] **Nianwen Xue**[2] **Alexis Palmer**[3]
[1]Bosch Center for Artificial Intelligence, Renningen, Germany
[2]Brandeis University [3]Dept. of Linguistics, University of Colorado Boulder
annemarie.friedrich@de.bosch.com xuen@brandeis.edu
alexis.palmer@colorado.edu

Abstract

Aspectual meaning refers to how the internal temporal structure of situations is presented. This includes whether a situation is described as a state or as an event, whether the situation is finished or ongoing, and whether it is viewed as a whole or with a focus on a particular phase. This survey gives an overview of computational approaches to modeling lexical and grammatical aspect along with intuitive explanations of the necessary linguistic concepts and terminology. In particular, we describe the concepts of stativity, telicity, habituality, perfective and imperfective, as well as influential inventories of eventuality and situation types. Aspect is a crucial component of semantics, especially for precise reporting of the temporal structure of situations, and future NLP approaches need to be able to handle and evaluate it systematically.

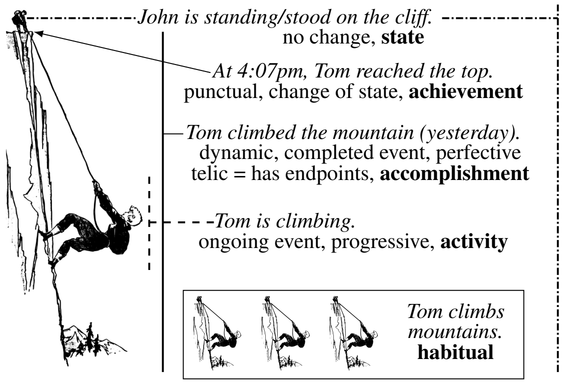

Figure 1: **Aspect** is like the camera lens of language, the device by which we focus on particular phases of a situation (Vendler, 1957; Smith, 1997).

1 Introduction

Lexical and grammatical aspect play essential roles in semantic interpretation (Smith, 2003), and yet even state-of-the-art natural language understanding (NLU) systems do not address these linguistic phenomena systematically (Metheniti, 2022). Consider this example: an NLU-based personal assistant, noticing the boarding time of a flight, tells a passenger (who is still shopping at the airport) "You miss flights" (i.e. on a regular basis) instead of "You are missing the flight" (now!). The traveler might misinterpret this utterance as chit-chat and indeed miss the flight. Aspectual encoding errors impair fluid and correct communication.

While there has been a notable amount of work on modeling lexical and grammatical aspect in the computational linguistics community (Moens and Steedman, 1988; Siegel and McKeown, 2000b; Friedrich et al., 2016; Kober et al., 2020), this area is still a niche in natural language processing (NLP). In this paper, we survey the state of research in this area and argue that a good computational understanding of lexical and grammatical aspect is paramount for capturing temporal information signaled by linguistic encoding. We thereby add to the on-going discussion in the NLP community about what is required to achieve true natural language understanding (Bender and Koller, 2020; Dunietz et al., 2020; Trott et al., 2020).

When describing a situation, as exemplified in Figure 1, producers of language have options for how to depict the situation and which subparts of it to highlight. Just as a cinematographer uses focus to highlight certain elements of a scene and thus bring them to the attention of the viewer, different aspectual choices focus on different subparts of the situation at hand, leaving the rest as background (analogy due to Smith, 2003). For example, many situations can be construed as either *foregrounded* events ("Tom climbed the mountain"), moving narrative time forward, or as states ("John stood on the cliff"), which often function as the *background* in a narrative or explanation. Aspectual choices may trigger hard inferences, softer implicatures, or simply expectations, having a non-negligible impact on the reader or listener. One aim of this survey is to clarify the sometimes complex linguistic work on aspect, giving NLP researchers and practitioners a

Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pages 599–622
May 2-6, 2023

Dataset	Reference	Language	Size	Annotated Categories
Asp-Ambig	Friedrich and Palmer (2014a)	EN	7875 clauses	stativity
SitEnt	Friedrich et al. (2016)	EN	40,000 clauses	situation entities, stativity, habituality
Tense-Europarl	Loáiciga and Grisot (2016)	EN/FR	435 verb phrases	boundedness (in sentence context)
MASC-telicity	Friedrich and Gateva (2017)	EN	1863 clauses	telicity
Captions	Alikhani and Stone (2019)	EN	2600 captions	stative, durative, telic/atelic, punctual
SdeWac-Aspect	Egg et al. (2019)	DE	4200 clauses	stativity, durative vs. punctual, boundedness
DIASPORA	Kober et al. (2020)	EN	927 utterances	stative, telic, atelic
UDS	Gantt et al. (2022)	EN	16,624 sentences; granularity varies	stativity, some telicity, durativity, other event properties

Table 1: Available **Datasets** labeled with aspectual information.

conceptual toolkit, plus a glossary of aspect-related terminology (Appendix A). The second aim is to survey computational work on modeling aspect.

Aspectual categories are *semantic* notions that are conveyed through a variety of mechanisms (Section 2). Fundamental aspectual distinctions (Section 3) may be part of the lexical meaning of a verb and are also influenced by context. These distinctions drive some influential inventories of eventuality types (Section 3.3). Grammatical aspect (Section 4) is marked morphologically in some (e.g., Slavic) languages, or via the choice or absence of tense-aspect features in languages such as English. Section 5 explains the idea of situation types, aspectual distinctions at the clause level.

As a consequence of the often distributed or even implicit encoding of aspect, computational models need to consider a variety of lexical and grammatical features as well as discourse and pragmatic information. While some NLP systems trained on end-to-end datasets may get some of these cases right, we are not aware of a system treating such problems in a principled manner. A recent study finds that transformer-based embeddings are useful for classifying clausal aspect (Metheniti, 2022), but it remains unclear just how these models learn or incorporate aspectual information. This survey focuses on modeling aspectual information as it is annotated in the datasets listed in Table 1.

In Section 6, we discuss potential ways of moving the field forward. We argue that explicit modeling or evaluation of aspectual categories should become a routine perspective in future NLP: getting aspect right has high stakes for applications such as temporal question answering, machine translation, or computer-aided language learning.

2 Where Does Aspect Live?

Comrie (1976) notes that there is a terminological and conceptual confusion around tense and aspect. This is at least partially because the expression of temporal location is intertwined morphologically with aspect in many languages of the world (Smith, 1997), so the traditional grammatical terminologies of individual languages often introduce tense/aspect combinations simply as tenses. Tense is *deictic* and locates the time of a situation relative to the utterance or relative to another situation. Aspect, in contrast, refers to situation-internal time.

Aspectual distinctions are *covert* categories (Whorf, 1945; Smith, 1997), expressing particular temporal semantic concepts in a way that does not have a one-to-one mapping to lexical, grammatical or syntactic categories (Fillmore, 1969). For example, the English Progressive[1] often indicates an ongoing situation without focusing on its endpoints ("I am eating an apple"), but it also signals future events ("I am going to Paris"). The grammatical form of such tense-aspect combinations can often be automatically detected with high accuracy (Ramm et al., 2017; Myers and Palmer, 2019). This type of aspect, encoded in the tense system of a language, is different from *semantic* concepts which refer to how the internal structure of a situation is presented; this survey focuses on the latter.

So where does *covert, semantic* aspect live? Aspectual meaning is compositional, composed by the verb's inherent meaning, its arguments, its tense, any morphological aspect markers, some prepositional phrases, and the adverbs of the sentence (Verkuyl, 1972; Mourelatos, 1978; Smith, 1997). Eventuality and situation type inventories make use of several dimensions of aspect to distinguish between different types of temporal structures. *Lexical aspect* refers to information contained in the meaning of verbs (or verb senses), with the consequence that their values often can be determined

[1]Following Comrie (1976), we use initial capitals for the names of language-particular categories and lower case for language-independent semantic distinctions.

only when observing how the verb interacts with its clausal context. A second dimension, *grammatical* or *viewpoint* aspect, refers to whether there is a focus on part of the situation or whether the situation is viewed as a whole or even as repeating. Grammatical aspect and situation types are both semantic categorizations at the sentence/clause level (e.g., Vendler, 1957; Verkuyl, 2005; Smith, 2003).

3 Inherent Lexical Aspect

Here we explain two fundamental notions related to inherent *lexical aspect*: *stativity* refers to whether a situation is viewed as an event or state, and *telicity* refers to whether an endpoint is visible in a clause.

3.1 Stativity

The most fundamental distinction made in the hierarchies of eventuality types is that between *states* ("love," "own") and *events* ("run," "buy"). In contrast to *dynamic* predicates, *stative* verbs entail no change (Filip, 2012). States obtain in time but they do not take time; events occur, happen, or take place (Smith, 1997). Events may have a beginning and/or an end (Comrie, 1976), and they are often *durative*, i.e., conceived as lasting for a certain period of time. Some events are dynamic situations describing a change of state such as "John reached the goal"; these are perceived as *punctual*. For punctual, single-stage events which cause no change of state, Smith (1997) adds the situation type *semelfactives*, for verbs like "knock," "flash," or "blink." While dynamic in nature, semelfactives return to their initial state at their end and often occur as *iteratives* (Filip, 2012). The stativity distinction pertains to verb senses, not verb types. For example, the verb type "make" has both a dynamic (1a) and a stative (1b) sense.

(1) (a) She is **making** a cake. (*dynamic*)
(b) She **makes** a great host. (*stative*)

In theory, the analysis of stativity could be applied to inventories of verb senses such as WordNet (Fellbaum, 2010). In practice, most computational works focus on classifying verbs in context.

Early computational approaches. The dynamic-stative distinction is at the heart of early approaches to computational modeling of tense and aspect. In the PUNDIT system for temporal information processing (Passonneau, 1988a), the lexical aspect of verbs in tensed clauses can be read from the system's output: transition events use `become`, processes use `do`; other cases signal states. In the context of lexicon induction, Klavans and Chodorow (1992a) suggest representing event structure of a verb as its *degree of stativity*, which is estimated by the proportion of occurrences in a corpus that are in the Progressive. Brent (1991a) also presents a program for identifying stativity using syntactic indicators: verbs that occur frequently with the Progressive or that combine with rate adverbs (e.g. "quickly", "slowly") are usually dynamic. Both works evaluate by manually inspecting system output.

Datasets. Siegel (1999) presents a manually annotated dataset of 739 training and 739 test clauses taken from medical discharge summaries, covering 222 different verb types. With the aim of reproducing and extending their work, Friedrich and Palmer (2014a) have three annotators label 7,875 clauses from MASC (Ide et al., 2008), marking the clause's main verb as *stative* or *dynamic*.If annotators see both readings, they may select both. The dataset has later been extended to cover 30,333 clauses from MASC and 10,607 clauses from Wikipedia (Friedrich et al., 2016). Kober et al. (2020) introduce the DIASPORA dataset, in which 927 utterances from a corpus of human-human phone conversations (Brennan et al., 2013) are labeled with whether the first verb phrase of each utterance is stative or dynamic. The dataset spans 69 different verb types. Chen and Palmer (2022) produce 292 contrast sets for stativity in English and outline strategies for converting between stative and dynamic expressions of the same situation. For more datasets that annotate stativity as one feature among others, see Section 3.3.

Modeling.[2] Siegel (1999, 1997) and Siegel and McKeown (2000b) propose a machine-learning approach to classifying stativity that describes each verb occurrence exclusively using corpus-based statistics of the corresponding verb type. The verb-type based indicators are normalized counts that reflect, for instance, how often the verb co-occurs with the past tense, in the perfect, or in negated form (for a full list, see Appendix D.1). Using these features, they train logistic regression models, decision trees, and genetic programming algorithms.

Friedrich and Palmer (2014a) compare these corpus-based *linguistic indicators* to instance-based syntactic-semantic features representing the clausal context, e.g., the part-of-speech tag of the

[2]Appendix B summarizes computational systems and approaches to modeling aspect.

verb, tense, voice, and WordNet information (Fellbaum, 2010) for the verb's arguments. Using the LCS database (Dorr, 2001) and a procedure described by Dorr and Olsen (1997), they also construct three seed sets with verb types that either occur exclusively as *dynamic*, only as *stative*, or can take both aspects. Based on these seed sets and a pre-trained syntactic distributional model (Thater et al., 2010), similarity values are computed and used as additional features. Kober et al. (2020) input sums of non-contextualized word embeddings (Mikolov et al., 2013) for the clause's main verb as well as selected context words to a logistic regression classifier. Overall, they find that using the local context in the form of a word window outperforms the verb-type only classifier, but that feeding in the entire sentence hurts. Metheniti (2022) also classify stativity (called "duration" in their work) by fine-tuning various transformer models and classification layers. They find BERT (Devlin et al., 2019) models work better than RoBERTA (Liu et al., 2019), AlBERT (Lan et al., 2019) and XLNet (Yang et al., 2019).

3.2 Telicity

The second important distinction related to inherent lexical aspect is that of *telicity*. The term *telic*, derived from Greek *télos* (goal) was introduced by Garey (1957). In his definition, telic verb senses have a built-in goal: when that goal is reached, a change of state occurs and the event is complete (Smith, 1997). Telicity is also sometimes referred to as *boundedness* (e.g., by Loáiciga and Grisot, 2016). As illustrated by (2), telicity is a feature of the entire clause (Verkuyl, 2005).

(2) (a) He was swimming in the lake. *(atelic)*
(b) He was swimming across the lake. *(telic)*

When a telic verb is used in the *imperfective* as in (3a), the arrival or nonarrival at the goal is hidden (Garey, 1957). If the same verb is applied in the *perfective* as in (3b), it means that the goal has been reached at the time of reference. Hence, if (3a) is true at some particular point in time, it cannot be the case that (3b) is true at the same point in time.

(3) (a) John was recovering. *(telic, imperfective)*
(b) John has recovered. *(telic, perfective)*

In contrast, *atelic* verbs do not have to wait for a goal for their realization; they are realized as soon as they begin. If an atelic verb is used in imperfective form as in (4a), we can infer that the sentence in perfective form (4b) is also true.

(4) (a) Sue was singing. *(atelic, imperfective)*
(b) Sue has sung. *(atelic, perfective)*

Datasets. Siegel and McKeown (2000b) describe a small dataset annotated for *completedness*, i.e, telicity. Their training and test sets each consist of approximately 300 clauses taken from 10 novels and covering 204 different dynamic verbs. Friedrich and Gateva (2017) manually annotate 1863 clauses taken from MASC (Ide et al., 2008) for telicity. The Captions dataset (Alikhani and Stone, 2019) marks telicity in several image caption corpora. They find that the proportion of telic verbs ranges from 6% to 59% across corpora, with atelic descriptions prevailing in almost all corpora. In the DIASPORA dataset (Kober et al., 2020), in each utterance, the first VP is annotated with its *predicational* aspect (stative, telic, or atelic).

Modeling. Siegel and McKeown (1996, 2000b) propose a machine-learning approach to classifying telicity of verb occurrences using their above-described corpus-based linguistic indicators. For example, verbs frequently occurring in the Progressive are likely atelic. Leveraging these corpus-based *linguistic indicators*, Friedrich and Gateva (2017) integrate syntactic-semantic contextual features into their logistic regression model. They also leverage additional silver standard training data in the form of projected annotations from the English-Czech InterCorp (Čermák and Rosen, 2012; Rosen and Vavřín, 2012). Their approach, using the machine-readable Czech dictionary Vallex (Žabokrtský and Lopatková, 2007), is based on the assumption that most perfective Czech verbs are translated using telic verb constructions, and that most imperfective verbs correspond to atelic constructions. Loáiciga and Grisot (2016) create an automatic classifier similar to that of Friedrich and Gateva (2017) to classify *boundedness* of French verbs, i.e., whether the endpoint of an event has occurred or not. They show that this is useful for picking the correct tense in French translations of the English Simple Past. Several more recent studies have shown that distributional and neural models can be trained to predict telicity as annotated in available datasets (Kober et al., 2020; Metheniti et al., 2021; Metheniti, 2022). BERT-style models perform well on existing telicity datasets (with larger models outperforming smaller models), yet it is still unclear how or whether they actually capture aspect. Metheniti (2022) observe that models are always highly confident in their predictions, re-

Type	*dynamic*	*durative*	*telic*	Examples
state	-	+	-	know the answer, love Mary
activity	+	+	-	laugh, stroll in the park, swim
accomplishment	+	+	+	build a house, walk to school
achievement	+	-	+	win a race, recognize
semelfactive	+	-	-	tap, knock, hiccup, tap, wink

Table 2: Eventuality types (Vendler, 1957; Smith, 1997).

gardless of their accuracy (a common phenomenon for neural models). We are not aware of any probing or explainability studies in this area yet, and how to obtain better calibrated models is an open research question.

3.3 Eventuality Types

We now explain some influential inventories of *eventuality types*, which are ontologies of the temporal structures of events and/or states. Table 2 shows the set of situation types introduced by Vendler (1957): *state*, *activity*, *accomplishment*, and *achievement*. There are other similar schemata (Kenny, 1963; Mourelatos, 1978, more details in Appendix C.1); we use Vendler's terminology here.

States are inherently stative and atelic. They are *durative* as they usually extend in time, though this time period may be very short as in (5).

(5) He was very quiet for two seconds.

The other three Vendler classes are all dynamic. They differ in whether they have a built-in endpoint, and in whether a clearly defined process leads up to this endpoint. *Activities* as in (6), consisting entirely of a process, use atelic verbs. *Accomplishments* as in (7) consist of a process that leads up to a built-in terminal point. Similarly, *achievements* as in (8) have an endpoint including a change of state, but the verb meaning does not include a process leading up to this point. Smith (1997) adds the *semelfactive type* (see Section 3.1).

(6) Mary was laughing. *(activity)*

(7) Mary wrote a letter. *(accomplishment)*

(8) He arrived at the station. *(achievement)*

Moens and Steedman (1988) work with aspectual profiles of sentences (for details see Appendix C.1) that are classified by making reference to a so-called *nucleus*. For example, "John built a house" consists of the *preparatory phase* (the house is being built), a *culmination point* (the moment at which it is completed), and a *consequent state* (the house is complete). The process of **aspectual coercion** shifts aspectual verb types based on their arguments or other aspectual operators such as adverbials. When used with a predicate whose lexical entry corresponds to a culminated process, the English Progressive strips off the culmination point and makes visible only the preparatory process as in "John was running a mile."[3]

Datasets. Zarcone and Lenci (2008) create a dataset of 3129 occurrences of 28 Italian verbs manually annotated for Vendler-style event types. Keelan (2012) works on an eight-way classification task for categories based on Leech's (1971) classes (see Appendix C.2). More recently, in the context of the Richer Event Description (RED) annotation scheme (Ikuta et al., 2014; O'Gorman et al., 2016; O'Gorman et al., 2021), the annotation of events with finer-grained Vendler-style situation types has been proposed (Croft et al., 2016). Falk and Martin (2016) select 167 frequent French verbs from a lexical resource (François et al., 2007), and label the corresponding 1199 entries ("readings") with eight aspectual classes similar to those of Mourelatos (1978) as listed in Appendix C.1. Their classification task is finer-grained than labeling verb types, but coarser-grained than clause-level labeling. Egg et al. (2019) annotate 4200 clauses from the German SdeWac (Faaß and Eckart, 2013) with the features stative, durative vs. punctual, and boundedness. With the aim of improving zero-shot image to verb recognition, Zellers and Choi (2017) crowdsource a dataset of 1710 verb templates (such as "put up") for 1203 different verbs annotated with the four Vendler categories. The Caption dataset (Alikhani and Stone, 2019) annotates image captions with the features stative, durative, punctual, telic, and atelic.

Modeling. Zarcone and Lenci (2008) train a maximum entropy classifier that uses adverbial, morphological, and syntactic features, as well as features capturing argument structure. Keelan (2012) uses an SVM (Cortes and Vapnik, 1995) with features similar to the linguistic indicators of Siegel and McKeown (2000b). Zellers and Choi (2017) use GloVe embeddings (Pennington et al., 2014) to represent the input verb. They concatenate this embedding with a phrase embedding of the verb's dictionary definition computed using a recurrent neural network, and add a linear layer on top. Hermes et al. (2015) induce Vendler classes for German verb types, using Siegel-style distributional

[3] Thus offering an elegant solution to the *imperfective paradox* (Dowty, 1979; Lascarides, 1991).

features extracted from dependency-parsed corpus data (3000 sentences per verb type) and SVMs for type-level classification. Hermes et al. (2018) compare the above framework to shallow distributional vectors considering only co-occurring word types for each verb. Finally, Gantt et al. (2022) propose a generative model for event types (and other information) fitted from English data annotated with a range of aspectual and other properties. The resulting 4-class event type inventory closely resembles Vendler classes in the distinctions it draws.

The main obstacle to systematically comparing the works mentioned in this section is that their authors make differing choices for the granularity of both the annotation scheme and the computational modeling, and conclusions drawn from the studies depend on these choices. For better comparability, we need better benchmarking.

4 Grammatical Aspect / Viewpoint

Phenomena treated as *grammatical aspect* or *viewpoint* (Smith, 1997) take different views (the entire situation vs. a part of it), or signal recurrence.

4.1 Perfective vs. Imperfective

The *perfective* viewpoint presents situations as complete with both an initial and a final endpoint, while the *imperfective* viewpoint makes only certain parts of the situation visible to the receiver (Smith, 1997). The situation in (9a) is presented imperfectively, focusing on the middle phase of John's eating. In contrast, (9b) is viewed in the perfective. Here, the interpretation is that (b) happens during the interval at which (a) is true.

(9) (a) John was eating a sandwich *(imperfective)*
(b) when Susan entered. *(perfective)*

The perfective/imperfective distinction is observed cross-linguistically, although some languages make even finer distinctions. For example, Chinese has two perfective aspect markers and two imperfective aspect markers. The two perfective markers *le* (10a) and *guo* (10b) differ in the present relevance, with the former indicating completion of a situation while the latter emphasizes the experience of having been through a situation. The two imperfective aspect markers *zhèngzài* and *zhe* both indicate a situation is on-going at the reference time, but the latter emphasizes the resulting state of a situation (Chao, 1968; Li and Thompson, 1989; Ljungqvist, 2007).

(10) (a) tā jìn le fángjiān "*He has entered the room (and is still in the room)"*
(b) tā jìn guò fángjiān "*He entered the room (at some point but is no longer in the room)"*
(c) tā zhèngzài jìn fángjiān "*He is entering the room"*
(d) fángjiān de mén kái zhe "*The door of the room is open."*

While also referring to endpoints, the concept of viewpoint differs from telicity: telicity describes types of situations independent of which phase is focused (e.g., "eating a sandwich" in (9a) is an (inherently-telic) accomplishment); viewpoint adds the focus (in this case on the phase during which the eating happens and the endpoint has not yet been reached; more details in Appendix C.3). It is the perfective-imperfective distinction that has traditionally been referred to as *aspect* in Slavic linguistics (Filip 1999). Computational work leverages parallel corpora to map the distinction, which is partially *overt* (i.e., explicit in the morphosyntax) in Slavic languages, to English text (Stambolieva, 2011; Friedrich and Gateva, 2017). Despite overall high accuracy, the models still struggle to reach high scores for the minority class *atelic*. In addition, such annotation projection approaches are not easily scalable, as they require strong knowledge of the languages involved.

4.2 Habituality

Habituals such as (11b) are sentences that "express regularities about the world which constitute generalizations over events and activities" (Carlson, 2005); on a sentence-level, they can be regarded as "derived statives" (Smith, 1997). In contrast, the term *episodic* refers to particular events. Habituals allow exceptions, e.g., (11b) is still true if Mary sometimes takes the train.

(11) (a) Mary cycled to work. *(episodic)*
(b) Mary cycles to work. *(habitual)*

Habitual sentences may also use stative predicates as in (12), generalizing over situations in which some state applies (Smith, 2005).

(12) Sloths sometimes rest on trees. *(habitual)*

Habituals are not restricted to what one would usually consider a matter of habit (Carlson, 2005); they can also have inanimate subjects as in (13).

(13) Glass breaks easily. *(habitual)*

Habituality is not to be confused with *iterativity*, which states that a situation occurs repeatedly, but not regularly, as in "the light flashed." Borderline cases are discussed in Appendix C.4.

Data and Modeling. Mathew and Katz (2009) randomly select 1052 sentences for 57 verbs from the Penn TreeBank (Marcus et al., 1993) and manually mark them as habitual or episodic. They train both a decision tree and a Naive Bayes classifier on syntactic features extracted from gold parses. Friedrich and Pinkal (2015a) argue that in order to be able to apply such a model on free text, a third category, which they call *static*, needs to be taken into consideration, which covers lexically stative, negated, and modalized clauses. They experiment with Random Forest classifiers (Breiman, 2001) using syntactic-semantic features and linguistic indicators (see Section 3.1). Govindarajan et al. (2019) label UD-EWT (Silveira et al., 2014) with real-valued properties indicating e.g., the "degree" of habituality. Their multi-layer perceptron uses features from verb databases such as LCS and GLoVe embeddings. Results for predicting habituality are overall promising, although there is a need for experiments with careful consideration of and controls over the verb types involved.

Other recent related work (Williams, 2012; Williams and Katz, 2012; Vempala et al., 2018) extracts typical durations (in terms of actual time measures) for verb lemmas from Twitter. They distinguish episodic and habitual uses of the verbs, using the method of Mathew and Katz (2009), and collect typical durations (e.g., "seconds" or "weeks") for episodic and habitual uses separately for each verb. The problem has further been studied in the context of acquiring common sense knowledge (e.g., Zhou et al., 2019; Yang et al., 2020).

4.3 Grammatical Aspect in Recent Syntactic-Semantic Frameworks

The **Universal Dependencies** (UD) guidelines[4] define six aspectual *features* for verbs: Hab, Imp, Iter, Perf, Prog, and Prosp (prospective, for relative future). The categories are assumed to be language-specific and close to the respective morphologies, and the exact definition of each feature is left to the individual languages. UD parsers frequently treat the identification of these features as a tagging task (e.g., Kondratyuk and Straka, 2019).

In the context of a pilot study on integrating aspect into Abstract Meaning Representations (Banarescu et al., 2013, **AMR**), Donatelli et al. (2018) propose to indicate whether a clause is habitual or ongoing (in addition to marking stativity and telicity). Tense and aspect annotation has been incorporated into Dialogue-AMR and used to annotate a corpus of human-robot interactions (Bonial et al., 2019, 2020). The Uniform Meaning Representations (**UMR**) framework (Van Gysel et al., 2021) uses a lattice for typologically-motivated annotation of aspect. The lattice begins by making the distinction between three categories: HABITUAL, which covers recurring states and events; PROCESS, which covers all non-recurring dynamic situations; and IMPERFECTIVE, designated for non-recurring states and atelic processes. The PROCESS category is further divided into ongoing ACTIVITIES, ENDEAVORS which have begun but not finished, and PERFORMANCES, which are completed dynamic processes. See Appendix D.2 for the full lattice.

Chen et al. (2021) build a rule-based system leveraging syntactic and semantic cues to annotate English sentences according to the UMR aspect lattice. Due to the recency of these frameworks, so far there is only very limited computational work.

5 Situation Entities

In this section, we review work on automatically classifying *situation entity* (SE) types (see inventory in Appendix D.3). SEs are "semantic concepts organized according to their internal temporal properties" (Smith, 2003). They are introduced to the discourse by a clause's verb constellation, i.e., the clause's main verb and its arguments and modifiers. Deciding on the type of an SE thus involves the combination of lexical and syntactic factors. In contrast to eventuality types (Section 3.3), situation entities capture the linguistic nature of the clause (Friedrich and Palmer, 2014b). EVENTs include all dynamic verb constructions referring to particular non-recurring situations, no matter whether there is a pre-defined endpoint or not. They may occur in the perfective or imperfective. GENERALIZING SENTENCEs correspond to habituals as described in Section 4.2, with the exception of GENERIC SENTENCEs that comprise all sentences making statements about kinds (Krifka et al., 1995).

Datasets. Palmer et al. (2007) present 6065 clauses taken from the Brown corpus (Francis and Kučera, 1979) and MUC-6 (Grishman and Sundheim, 1996), manually annotated with SE types. The SitEnt dataset (Friedrich et al., 2016; Friedrich, 2017) consists of 40,000 clauses from 13 genres annotated for SE types with substantial agreement (Friedrich and Palmer, 2014b). Govindarajan et al.

[4] https://universaldependencies.org/u/feat/Aspect.html

(2019) crowdsource UDW-G, a dataset of 37,146 arguments and 33,114 predicates in 16,222 sentences with continuous judgments on a scale from 1 to 5 with distinctions essentially following the schema of Friedrich et al. (2016).

Modeling. Palmer et al. (2007) use a maximum entropy sequence tagger with part-of-speech and CCG-based (Steedman, 2000) syntactic features to predict labels for each sentence separately. In an oracle experiment, they show that including the label of the previous sentence(s) as a feature improves performance. Friedrich et al. (2016) present the first true sequence labeling approach to SE types classification using conditional random fields (Lafferty et al., 2001) with distributional features in the form of Brown cluster IDs (Brown et al., 1992) and syntactic-semantic features. Kober et al. (2020) propose to classify predicational aspect with distributional semantics, using non-contextualized word embeddings and sums of the word vectors of the target words and their context words. Several works (Palmer et al., 2007; Palmer and Friedrich, 2014; Friedrich et al., 2016) find that sequence information mostly helps when training in-domain and has the biggest impact on identifying GENERIC SENTENCES, which often require discourse context for classification, even by human annotators (Friedrich and Pinkal, 2015b).

SE types have also been modeled using a variety of neural approaches. Becker et al. (2017) employ a GRU-based (Cho et al., 2014) RNN with an attention mechanism. Dai and Huang (2018) dynamically build context-aware clause representations, informed by their paragraph-wide contexts. They propose a hierarchical recurrent neural network that reads entire paragraphs at once, learning interdependencies for clauses. Their model uses word2vec (Mikolov et al., 2013) embeddings for words, and first computes contextualized word embeddings using a BiLSTM over the entire paragraph. Clause embeddings are formed by max pooling over the word embeddings of each clause, and then clause embeddings are further contextualized via another BiLSTM. For further improving SE classification, they add a CRF layer on top, and finally predict SE labels via softmax. Recently, BERT (Devlin et al., 2019) and GPT-2 (Radford et al., 2019) have been employed for the classification task by Rezaee et al. (2021), who also compare to using ParBERT (Cohan et al., 2019). While BERT only considers one clause at a time, ParBERT reads several sentences at once, separated by [SEP] tokens, and then uses the embeddings of the [SEP] tokens to predict a label for the corresponding clause. Comparing a variety of neural models based on non-contextualized word embeddings on their situation-entity related regression tasks, Govindarajan et al. (2019) reach levels similar to human agreement.

The computational problem of identifying SE types has recently been studied extensively. Overall, SE patterns are specific to the domain, genre or *discourse mode* (Smith, 2003). Future research could use SE-style analysis for NLP tasks like temporal processing or information extraction.

6 How Can We Move Forward?

In this section, we discuss potential steps that could lead to more successful modeling of aspectual information, as well as how to leverage that information in NLP applications.

6.1 Dataset Construction

To date, there is no consensus or standard across languages regarding where aspect lives. Ongoing efforts in the UMR community aim to develop typologically-informed aspectual representations that work across languages, but so far very few UMR datasets exist. Whether achieving a standard for aspectual representation is a necessary step or not, in order to make more rapid progress on the computational modeling of aspect, first the various models need to be **benchmarked** on the same tasks and datasets. **Dataset construction** in this area is challenging. Annotators not only need to have extensive linguistic training but also often domain expertise. Formulating the problem as a crowdsourcing task is a possible direction (Govindarajan et al., 2019), but obtaining consistently high agreement is (as for most tasks) difficult in this setup. **Multilingual** datasets and models are still under-researched, yet they are of particular relevance due to applications in computer-aided language learning or the evaluation of machine translation (see examples in Appendix E).

6.2 Modeling

More research is also needed on whether recent **embeddings** or text encoders capture or abstract away from aspectual features. Recent experimental findings on the SitEnt dataset show that, when training data within the same domain and genre is available, performance of pre-trained transformers is in the range of human agreement (Metheniti, 2022; Dai and Huang, 2018). The current state-of-the-art model based on BERT-GPT achieves a macro-average F1 of 79.1 compared to an estimate of human performance (78.6). Yet, most verbs are either not ambiguous or have a strongly predominant majority class that the models memorize (Friedrich and Palmer, 2014a; Kober et al., 2020). We expect much insight to come from further studying verb types that behave differently depending on the context. Moreover, results on cross-genre classification of SE types are yet inconclusive. Dai and Huang (2018) report high cross-genre scores for BERT, ParBERT, and BERT+GPT, but out of these models, there is no clear winner across all genres. Overall, when only using out-of-genre training data, performance of the BERT+GPT model still drops to an F1 of 70.7.

Investigating aspectual information in the context of **document-level models** (e.g., Beltagy et al., 2020) is another opportunity: cues for aspectual interpretion may occur earlier in the discourse than the sentence being interpreted (see App. E).

Linguistic indicators (Siegel and McKeown, 2000a) have been shown to be useful for predicting aspectual features in the absence of training data. Another research direction should hence look at how aspectual information can be **induced from raw data** or parallel corpora on a larger scale and for a broader set of distinctions. Future work should also investigate the interaction with other elements of semantics such as argument structure.

In sum, systems for predicting aspectual information encounter several challenges: (a) many verb types are ambiguous, (b) datasets for the task show significant class imbalance, and (c) for many verb types, no training data at all is available. Initial experiments with careful controls for different groups of verbs were highly promising (Friedrich and Palmer, 2014a; Kober et al., 2020), but available annotated data is sparse, especially for ambiguous verb types. More research is required to disentangle the effects of potentially lexically biased datasets, class imbalance, and sparsity.

6.3 Applications

Automatic classifiers for aspectual distinctions clearly have **applications** in the **digital humanities** and quantitative linguistics, where research questions may directly target the use of aspect. Recent highly fluent (but not necessarily always accurate) language generation models such as GPT-3 (Brown et al., 2020) and ChatGPT[5] open up a new level of writing assistance or language learning systems. We believe that a fruitful direction for future research is to make such systems explainable, also in the sense that they are able to give linguistically founded qualitative feedback to the user about why an aspectual form is correct or wrong in a particular context. Benchmarking on datasets explicitly annotated for aspectual information will let us estimate the degree to which text generated by the models is in line with the author's intentions. In Appendix F, we provide some chat logs of conversations with ChatGPT that illustrate how far it is from being an adequate tool for teaching the linguistic concepts described in this survey.

However, we argue that paying attention to aspect is necessary to arrive at fluent human-level NLP systems as motivated by our "miss the flight" example in Section 1. In Appendix E, we discuss two potential application areas: temporal processing and machine translation. In a nutshell, some works suggest that embedding spaces still struggle to capture the nature of tense and aspect, with consequences for commonsense reasoning driven by temporal properties, and can hence lead to wrong conclusions. Similarly, machine translation systems may (and still do) go wrong when cues for the correct temporal form occur much earlier in the document then the sentence being translated. More research is needed on how to **incorporate features** explicitly or **guide systems** implicitly towards the right output with regard to aspectual information, and also on how to convincingly evaluate generated output with regard to aspectual plausibility.

7 Conclusion

This paper provides an overview of the linguistic concepts and terminology associated with aspect, and also surveys existing datasets and prior work in computational modeling of aspect. Because the expression of aspect varies across languages, the proper incorporation of aspect into downstream NLP tasks requires language-dependent research rooted in the language-independent categories surveyed here. The semantic concepts expressed by lexical and grammatical aspect play essential roles in semantic interpretation, and their proper coding and evaluation must not be neglected if we are to one day develop human-level NLU systems.

[5] https://openai.com/blog/chatgpt

Limitations

This survey reviews linguistic and computational work on lexical and grammatical aspect. While we believe that we covered the vast majority of relevant computational works, we of course only scratch the surface of the body of theoretical linguistic work and need to simplify at times.

As the term *aspect* is highly ambiguous and used in NLP primarily within sentiment analysis, it is difficult to search for relevant papers. Our search hence relied mainly on following citation networks.

Ethical Considerations

We did not identify any potential ethical issues with this survey.

Acknowledgements

We thank the anonymous reviewers for their extremely valuable comments. This survey incorporates insights from countless extremely valuable discussions with many researchers and students. In particular, we thank Manfred Pinkal, Josef van Genabith, Bonnie Webber, Mark Steedman, Mike Lewis, Omri Abend, Melissa Peate Sorensen, Michaela Regneri, Andrea Horbach, Stefan Thater, Vera Demberg, Ines Rehbein, Caroline Sporleder, Fatemeh Torabi Asr, Asad Sayeed, Alessandra Zarcone, Hanna Kermes, Damyana Gateva, Anna Nedoluzhko, Michal Novák, Martha Palmer, and Laura Michaelis.

References

Malihe Alikhani and Matthew Stone. 2019. "caption" as a coherence relation: Evidence and implications. In *Proceedings of the Second Workshop on Shortcomings in Vision and Language*, pages 58–67, Minneapolis, Minnesota. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Maria Becker, Michael Staniek, Vivi Nastase, Alexis Palmer, and Anette Frank. 2017. Classifying semantic clause types: Modeling context and genre characteristics with recurrent neural networks and attention. In *Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)*, pages 230–240, Vancouver, Canada. Association for Computational Linguistics.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *CoRR*, abs/2004.05150.

Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.

Steven Bethard. 2013. ClearTK-TimeML: A minimalist approach to TempEval 2013. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 10–14, Atlanta, Georgia, USA. Association for Computational Linguistics.

Steven Bethard and James H. Martin. 2006. Identification of event mentions and their semantic class. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 146–154, Sydney, Australia. Association for Computational Linguistics.

Steven Bethard, Guergana Savova, Wei-Te Chen, Leon Derczynski, James Pustejovsky, and Marc Verhagen. 2016. SemEval-2016 task 12: Clinical TempEval. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1052–1062, San Diego, California. Association for Computational Linguistics.

Steven Bethard, Guergana Savova, Martha Palmer, and James Pustejovsky. 2017. SemEval-2017 task 12: Clinical TempEval. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 565–572, Vancouver, Canada. Association for Computational Linguistics.

Claire Bonial, Lucia Donatelli, Mitchell Abrams, Stephanie M. Lukin, Stephen Tratz, Matthew Marge, Ron Artstein, David Traum, and Clare Voss. 2020. Dialogue-AMR: Abstract Meaning Representation for dialogue. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 684–695, Marseille, France. European Language Resources Association.

Claire Bonial, Lucia Donatelli, Stephanie M. Lukin, Stephen Tratz, Ron Artstein, David Traum, and Clare Voss. 2019. Augmenting Abstract Meaning Representation for human-robot dialogue. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 199–210, Florence, Italy. Association for Computational Linguistics.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages

632–642, Lisbon, Portugal. Association for Computational Linguistics.

Leo Breiman. 2001. Random forests. *Machine learning*, 45(1):5–32.

Susan Brennan, Katharina Schuhmann, and Karla Batres. 2013. Entrainment on the move and in the lab: The walking around corpus. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 35. Cognitive Science Society.

Michael R. Brent. 1991a. Automatic semantic classification of verbs from their syntactic contexts: An implemented classifier for stativity. In *Fifth Conference of the European Chapter of the Association for Computational Linguistics*, Berlin, Germany. Association for Computational Linguistics.

Michael R Brent. 1991b. Automatic semantic classification of verbs from their syntactic contexts: an implemented classifier for stativity. In *Proceedings of the fifth conference on European chapter of the Association for Computational Linguistics (ACL)*, pages 222–226, Berlin, Germany.

Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. 1992. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *cs.CL*.

Gregory N Carlson. 2005. Generics, habituals and iteratives. In Keith Brown, editor, *The Encyclopedia of Language and Linguistics*. Elsevier Ltd.

Franti Čermák and Alexandr Rosen. 2012. The case of intercorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, 17(3):411–427.

Nathanael Chambers. 2013. NavyTime: Event and time ordering from raw text. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 73–77, Atlanta, Georgia, USA. Association for Computational Linguistics.

Yuen Ren Chao. 1968. *A grammar of spoken Chinese*. Univ of California Press.

Daniel Chen and Alexis Palmer. 2022. Contrast sets for stativity of English verbs in context. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4028–4036, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Daniel Chen, Martha Palmer, and Meagan Vigus. 2021. AutoAspect: Automatic annotation of tense and aspect for uniform meaning representations. In *Proceedings of The Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 36–45, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder–decoder approaches. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 103–111, Doha, Qatar. Association for Computational Linguistics.

Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, and Dan Weld. 2019. Pretrained language models for sequential sentence classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3693–3699, Hong Kong, China. Association for Computational Linguistics.

Bernard Comrie. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*, volume 2 of *Cambridge Textbooks in Linguistics*. Cambridge University Press.

Corinna Cortes and Vladimir Vapnik. 1995. Support-Vector Networks. *Machine Learning*, 20(3):273–297.

Francisco Costa and António Branco. 2012. Aspectual type and temporal relation classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 266–275, Avignon, France. Association for Computational Linguistics.

William Croft, Pavlína Pešková, and Michael Regan. 2016. Annotation of causal and aspectual structure of events in RED: a preliminary report. In *Proceedings of the Fourth Workshop on Events*, pages 8–17, San Diego, California. Association for Computational Linguistics.

Zeyu Dai and Ruihong Huang. 2018. Building context-aware clause representations for situation entity type classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3305–3315, Brussels, Belgium. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages

4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Lucia Donatelli, Michael Regan, William Croft, and Nathan Schneider. 2018. Annotation of tense and aspect semantics for sentential AMR. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018)*, pages 96–108, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Bonnie J Dorr. 2001. Lcs verb database, online software database of lexical conceptual structures and documentation, university of maryland.

Bonnie J. Dorr and Mari Broman Olsen. 1997. Deriving verbal and compositonal lexical aspect for NLP applications. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 151–158, Madrid, Spain. Association for Computational Linguistics.

David R Dowty. 1979. *Word meaning and Montague grammar – The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Studies in Linguistics and Philosophy. Springer.

Jesse Dunietz, Greg Burnham, Akash Bharadwaj, Owen Rambow, Jennifer Chu-Carroll, and Dave Ferrucci. 2020. To test machine comprehension, start by defining comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7839–7859, Online. Association for Computational Linguistics.

Markus Egg, Helena Prepens, and Will Roberts. 2019. Annotation and automatic classification of aspectual categories. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3335–3341, Florence, Italy. Association for Computational Linguistics.

Gertrud Faaß and Kerstin Eckart. 2013. Sdewac–a corpus of parsable sentences from the web. In *Language processing and knowledge in the Web*, pages 61–68. Springer.

Ingrid Falk and Fabienne Martin. 2016. Automatic identification of aspectual classes across verbal readings. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 12–22, Berlin, Germany. Association for Computational Linguistics.

Christiane Fellbaum. 2010. Wordnet. In *Theory and applications of ontology: computer applications*, pages 231–243. Springer.

Hana Filip. 2012. Lexical aspect. *The Oxford handbook of tense and aspect*, pages 721–751.

Hana Filip and Gregory N Carlson. 1997. Sui generis genericity. *University of Pennsylvania Working Papers in Linguistics*, 4(2):7.

Charles J Fillmore. 1969. Types of lexical information. In *Studies in syntax and semantics*, pages 109–137. Springer.

W Nelson Francis and Henry Kučera. 1979. Brown corpus manual. *Brown University*. Http://clu.uni.no/icame/brown/bcm.html.

Jacques François, Denis Le Pesant, and Danielle Leeman. 2007. Présentation de la classification des verbes français de jean dubois et françoise dubois-charlier. *Langue française*, 1:3–19.

Annemarie Friedrich and Damyana Gateva. 2017. Classification of telicity using cross-linguistic annotation projection. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2559–2565, Copenhagen, Denmark. Association for Computational Linguistics.

Annemarie Friedrich and Alexis Palmer. 2014a. Automatic prediction of aspectual class of verbs in context. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 517–523, Baltimore, Maryland. Association for Computational Linguistics.

Annemarie Friedrich and Alexis Palmer. 2014b. Situation entity annotation. In *Proceedings of LAW VIII - The 8th Linguistic Annotation Workshop*, pages 149–158, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

Annemarie Friedrich, Alexis Palmer, and Manfred Pinkal. 2016. Situation entity types: automatic classification of clause-level aspect. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1757–1768, Berlin, Germany. Association for Computational Linguistics.

Annemarie Friedrich and Manfred Pinkal. 2015a. Automatic recognition of habituals: a three-way classification of clausal aspect. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2471–2481, Lisbon, Portugal. Association for Computational Linguistics.

Annemarie Friedrich and Manfred Pinkal. 2015b. Discourse-sensitive automatic identification of generic expressions. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1272–1281, Beijing, China. Association for Computational Linguistics.

Annemarie Silke Friedrich. 2017. *States, events, and generics: computational modeling of situation entity types*. Ph.D. thesis, Saarland University.

William Gantt, Lelia Glass, and Aaron Steven White. 2022. Decomposing and recomposing event structure. *Transactions of the Association for Computational Linguistics*, 10:17–34.

Howard B Garey. 1957. Verbal aspect in french. *Language*, 33(2):91–110.

Venkata Govindarajan, Benjamin Van Durme, and Aaron Steven White. 2019. Decomposing generalization: Models of generic, habitual, and episodic statements. *Transactions of the Association for Computational Linguistics*, 7:501–517.

Ralph Grishman and Beth Sundheim. 1996. Message understanding conference-6: A brief history. In *Proceedings of the 16th conference on Computational linguistics-Volume 1*, pages 466–471. Association for Computational Linguistics.

Valentine Hacquard. 2009. On the interaction of aspect and modal auxiliaries. *Linguistics and Philosophy*, 32(3):279–315.

Jürgen Hermes, Michael Richter, and Claes Neuefeind. 2015. Automatic Induction of German Aspectual Verb Classes in a Distributional Framework. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology (GSCL)*, pages 122–129, Duisburg-Essen, Germany.

Jürgen Hermes, Michael Richter, and Claes Neuefeind. 2018. Supervised classification of aspectual verb classes in german-subcategorization-frame-based vs window-based approach: A comparison. In *ICAART (2)*, pages 653–662.

Nancy Ide, Collin Baker, Christiane Fellbaum, Charles Fillmore, and Rebecca Passonneau. 2008. MASC: the manually annotated sub-corpus of American English. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Rei Ikuta, Will Styler, Mariah Hamang, Tim O'Gorman, and Martha Palmer. 2014. Challenges of adding causation to richer event descriptions. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 12–20, Baltimore, Maryland, USA. Association for Computational Linguistics.

Hyuckchul Jung and Amanda Stent. 2013. ATT1: Temporal annotation using big windows and rich syntactic and semantic features. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 20–24, Atlanta, Georgia, USA. Association for Computational Linguistics.

Richard Keelan. 2012. *Lexical Aspectual Classification*. University of Ottawa (Canada).

Anthony Kenny. 1963. *Action, emotion and will*. Routledge.

Judith L. Klavans and Martin Chodorow. 1992a. Degrees of stativity: The lexical representation of verb aspect. In *COLING 1992 Volume 4: The 14th International Conference on Computational Linguistics*.

Judith L Klavans and Martin Chodorow. 1992b. Degrees of stativity: the lexical representation of verb aspect. In *Proceedings of the 14th conference on Computational linguistics (Coling)*, volume 4, pages 1126–1131, Nantes, France.

Thomas Kober, Malihe Alikhani, Matthew Stone, and Mark Steedman. 2020. Aspectuality across genre: A distributional semantics approach. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4546–4562, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Thomas Kober, Julie Weeds, Lorenzo Bertolini, and David Weir. 2021. Data augmentation for hypernymy detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1034–1048, Online. Association for Computational Linguistics.

Dan Kondratyuk and Milan Straka. 2019. 75 languages, 1 model: Parsing Universal Dependencies universally. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2779–2795, Hong Kong, China. Association for Computational Linguistics.

Manfred Krifka, Francis Jeffrey Pelletier, Gregory N. Carlson, Alice ter Meulen, Godehard Link, and Gennaro Chierchia. 1995. Genericity: An Introduction. In Gregory N. Carlson and Francis Jeffry Pelletier, editors, *The Generic Book*, Studies in Communication, Media, and Public Opinion, pages 1–124. University Of Chicago Press.

John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the eighteenth International Conference on Machine Learning (ICML)*, volume 1, pages 282–289, Williamstown, Massachussetts.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. ALBERT: A lite BERT for self-supervised learning of language representations. *CoRR*, abs/1909.11942.

Alex Lascarides. 1991. The Progressive and the Imperfective Paradox. *Synthese*, 87(3):401–447.

Geoffrey N Leech. 1971. *Meaning and the English verb*. Pearson Education.

Charles N Li and Sandra A Thompson. 1989. *Mandarin Chinese: A functional reference grammar*, volume 3. Univ of California Press.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Marita Ljungqvist. 2007. Le, guo and zhe in mandarin chinese: a relevance-theoretic account. *Journal of East Asian Linguistics*, 16(3):193–235.

Hector Llorens, Estela Saquete, and Borja Navarro-Colorado. 2010. TimeML events recognition and classification: Learning CRF models with semantic roles. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 725–733, Beijing, China. Coling 2010 Organizing Committee.

Sharid Loáiciga and Cristina Grisot. 2016. Predicting and using a pragmatic component of lexical aspect of simple past verbal tenses for improving English-to-French machine translation. In *Linguistic Issues in Language Technology, Volume 13, 2016*. CSLI Publications.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Thomas A. Mathew and E. Graham Katz. 2009. Supervised Categorization of Habitual and Episodic Sentences. In *Sixth Midwest Computational Linguistics Colloquium*, Bloomington, Indiana: Indiana University.

Eleni Metheniti. 2022. About time: Do transformers learn temporal verbal aspect? In *Cognitive Modeling and Computational Linguistics (CMCL)*, Dublin,Ireland.

Eleni Metheniti, Tim van de Cruys, and Nabil Hathout. 2021. Prédire l'aspect linguistique en anglais au moyen de transformers (classifying linguistic aspect in English with transformers). In *Actes de la 28e Conférence sur le Traitement Automatique des Langues Naturelles. Volume 1 : conférence principale*, pages 209–218, Lille, France. ATALA.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Marc Moens and Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics*, 14(2):15–28.

Alexander PD Mourelatos. 1978. Events, processes, and states. *Linguistics and philosophy*, 2(3):415–434.

Skatje Myers and Martha Palmer. 2019. ClearTAC: Verb tense, aspect, and form classification using neural nets. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 136–140, Florence, Italy. Association for Computational Linguistics.

Tim O'Gorman, Kristin Wright-Bettner, and Martha Palmer. 2021. The Richer Event Description Corpus for Event-Event Relations. *Computational Analysis of Storylines: Making Sense of Events (Studies in Natural Language Processing)*.

Tim O'Gorman, Kristin Wright-Bettner, and Martha Palmer. 2016. Richer event description: Integrating event coreference with temporal, causal and bridging annotation. In *Proceedings of the 2nd Workshop on Computing News Storylines (CNS 2016)*, pages 47–56.

Alexis Palmer and Annemarie Friedrich. 2014. Genre distinctions and discourse modes: Text types differ in their situation type distributions. In *Proceedings of the Symposium on Frontiers and Connections between Argumentation Mining and Natural Language Processing*, Bertinoro, Italy.

Alexis Palmer, Elias Ponvert, Jason Baldridge, and Carlota Smith. 2007. A sequencing model for situation entity classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 896–903, Prague, Czech Republic. Association for Computational Linguistics.

Rebecca J. Passonneau. 1988a. A computational model of the semantics of tense and aspect. *Computational Linguistics*, 14(2):44–60.

Rebecca J Passonneau. 1988b. A computational model of the semantics of tense and aspect. *Computational Linguistics*, 14(2):44–60.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Adam Poliak, Aparajita Haldar, Rachel Rudinger, J. Edward Hu, Ellie Pavlick, Aaron Steven White, and Benjamin Van Durme. 2018. Collecting diverse natural language inference problems for sentence representation evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 67–81, Brussels, Belgium. Association for Computational Linguistics.

James Pustejovsky, José M Castano, Robert Ingria, Roser Sauri, Robert J Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R Radev. 2003. Timeml: Robust specification of event and temporal expressions in text. *New directions in question answering*, 3:28–34.

James Pustejovsky, Kiyong Lee, Harry Bunt, and Laurent Romary. 2010. ISO-TimeML: An international standard for semantic annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Anita Ramm, Sharid Loáiciga, Annemarie Friedrich, and Alexander Fraser. 2017. Annotating tense, mood and voice for English, French and German. In *Proceedings of ACL 2017, System Demonstrations*, pages 1–6, Vancouver, Canada. Association for Computational Linguistics.

Roi Reichart and Ari Rappoport. 2010. Tense sense disambiguation: A new syntactic polysemy task. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 325–334, Cambridge, MA. Association for Computational Linguistics.

Mehdi Rezaee, Kasra Darvish, Gaoussou Youssouf Kebe, and Francis Ferraro. 2021. Discriminative and generative transformer-based models for situation entity classification. *CoRR*, abs/2109.07434.

Alexandr Rosen and Martin Vavřín. 2012. Building a multilingual parallel corpus for human users. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2447–2452, Istanbul, Turkey. European Language Resources Association (ELRA).

Roser Saurí, Robert Knippen, Marc Verhagen, and James Pustejovsky. 2005. Evita: A robust event recognizer for QA systems. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 700–707, Vancouver, British Columbia, Canada. Association for Computational Linguistics.

Eric V. Siegel. 1997. Learning methods for combining linguistic indicators to classify verbs. In *Second Conference on Empirical Methods in Natural Language Processing*.

Eric V. Siegel. 1999. Corpus-based linguistic indicators for aspectual classification. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 112–119, College Park, Maryland, USA. Association for Computational Linguistics.

Eric V Siegel and Kathleen R McKeown. 1996. Gathering statistics to aspectually classify sentences with a genetic algorithm. In *Proceedings of the Second International Conference on New Methods in Language Processing*, Ankara, Turkey.

Eric V Siegel and Kathleen R McKeown. 2000a. Learning methods to combine linguistic indicators: Improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–628.

Eric V. Siegel and Kathleen R. McKeown. 2000b. Learning methods to combine linguistic indicators:improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–627.

Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer, and Christopher D. Manning. 2014. A gold standard dependency corpus for English. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*.

Carlota S Smith. 1997. *The parameter of aspect*, volume 43 of *Studies in Linguistics and Philosophy*. Springer Science & Business Media.

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

Carlota S Smith. 2005. Aspectual entities and tense in discourse. In *Aspectual inquiries*, pages 223–237. Springer.

Maria Stambolieva. 2011. Parallel corpora in aspectual studies of non-aspect languages. In *Proceedings of The Second Workshop on Annotation and Exploitation of Parallel Corpora*, pages 39–42, Hissar, Bulgaria. Association for Computational Linguistics.

Mark Steedman. 2000. *The syntactic process*, volume 24 of *Language, Speech, and Communication*. MIT Press.

Tianxiang Sun, Xiangyang Liu, Xipeng Qiu, and Xuanjing Huang. 2021. Paradigm shift in natural language processing. *arXiv preprint arXiv:2109.12575*.

Stefan Thater, Hagen Fürstenau, and Manfred Pinkal. 2010. Contextualizing semantic representations using syntactically enriched vector models. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 948–957, Uppsala, Sweden. Association for Computational Linguistics.

Sean Trott, Tiago Timponi Torrent, Nancy Chang, and Nathan Schneider. 2020. (Re)construing meaning in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5170–5184, Online. Association for Computational Linguistics.

Naushadand UzZaman, Hectorand Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 1–9, Atlanta, Georgia.

Jens Van Gysel, Meagan Vigus, Jin Zhao, and Nianwen Xue. 2022. Developing a uniform meaning representation for natural language processing. In *LREC 2022 tutorial*.

Jens E. L. Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James H. Martin, Stephan Oepen, Martha

Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. Designing a uniform meaning representation for natural language processing. *KI - Künstliche Intelligenz*, 35(0):343–360.

Eva Vanmassenhove, Jinhua Du, and Andy Way. 2017. Investigating 'aspect'in nmt and smt: Translating the english simple past and present perfect. *Computational Linguistics in the Netherlands Journal*, 7:109–128.

Siddharth Vashishtha, Adam Poliak, Yash Kumar Lal, Benjamin Van Durme, and Aaron Steven White. 2020. Temporal reasoning in natural language inference. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4070–4078, Online. Association for Computational Linguistics.

Alakananda Vempala, Eduardo Blanco, and Alexis Palmer. 2018. Determining event durations: Models and error analysis. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 164–168, New Orleans, Louisiana. Association for Computational Linguistics.

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

Henk J Verkuyl. 1972. *On the Compositional Nature of the Aspects.*, volume 15 D of *Foundations of Language, Supplementary Series*. D. Reidel Publishing Company, Dordrecht-Holland.

Henk J Verkuyl. 2005. Aspectual composition: Surveying the ingredients. In *Perspectives on aspect*, pages 19–39. Springer.

Benjamin Lee Whorf. 1945. Grammatical categories. *Language*, 21(1):1–11.

Jennifer Williams. 2012. Extracting Fine-grained Durations for Verbs from Twitter. In *Proceedings of the ACL 2012 Student Research Workshop*, pages 49–54, Jeju Island, Korea.

Jennifer Williams and Graham Katz. 2012. Extracting and modeling durations for habits and events from Twitter. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 223–227, Jeju Island, Korea. Association for Computational Linguistics.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.

Zonglin Yang, Xinya Du, Alexander Rush, and Claire Cardie. 2020. Improving event duration prediction via time-aware pre-training. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3370–3378, Online. Association for Computational Linguistics.

Zdeněk Žabokrtskỳ and Markéta Lopatková. 2007. Valency information in vallex 2.0. *The Prague Bulletin of Mathematical Linguistics*, 87:41–60.

Alessandra Zarcone and Alessandro Lenci. 2008. Computational models for event type classification in context. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Rowan Zellers and Yejin Choi. 2017. Zero-shot activity recognition with verb attribute induction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 946–958, Copenhagen, Denmark. Association for Computational Linguistics.

Yuchen Zhang and Nianwen Xue. 2014. Automatic inference of the tense of Chinese events using implicit linguistic information. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1902–1911, Doha, Qatar. Association for Computational Linguistics.

Ben Zhou, Daniel Khashabi, Qiang Ning, and Dan Roth. 2019. "going on a vacation" takes longer than "going for a walk": A study of temporal commonsense understanding. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3363–3369, Hong Kong, China. Association for Computational Linguistics.

APPENDIX

A Glossary

This glossary intends to provide a concise alphabetically ordered overview of the linguistic terminology used in this paper.

accomplishment eventive / dynamic situation type according to (Vendler, 1957; Smith, 1997), consisting of a preparatory phase and an endpoint, e.g., "build a house."

achievement eventive / dynamic situation type according to (Vendler, 1957; Smith, 1997) consisting only of the *punctual* event that changes a state, e.g., "win the race."

activity eventive / dynamic situation type according to (Vendler, 1957; Smith, 1997), describes a process that does not have a pre-defined endpoint, e.g., "swim."

atelic describes situations or verbs that do not have a built-in endpoint, e.g., "draw."

bounded see *telic*.

covert describes linguistic categories that are not directly inferrable from the lexicon or from grammatical/syntactic structure (Whorf, 1945). They usually consist of *semantic* concepts that require taking into account the context or pragmatic factors.

durative describes states that extend in time.

dynamic describes a type of situation / verb that has a part to its meaning that applies a force or changes a state.

episodic describes particular events that "happen" or "have happened" (opposite of *habituals*).

eventive see *dynamic*.

habitual clause that expresses a situation that repeats regularly and expresses a characteristic, e.g., "John drives to work" or "Bishops move diagonally" (Krifka et al., 1995).

imperfective describes a perspective on a situation focusing one or none of the potential endpoints, but not both

iteratives clauses (usually with *semelfactive* verbs) that signal that the event expressed by the verb is executed a number of times, e.g., "the light flashed."

overt describes linguistic categories that are directly inferrable from the lexicon or from grammatical/syntactic structures.

state a situation that is not changing, e.g., "John owns a house."

stative describes verbs or situations that express a state.

perfective describes constructions that show a situation with its endpoint(s), e.g., "John traveled to the US (last week)."

progressive describes constructions that focus on a particular phase of a situation that is currently ongoing, e.g., "John is traveling to the US (at the moment)."

punctual event that occurs at a single point in time, e.g., "hiccup." – *dynamic* by definition.

semelfactive *punctual* situation or event type (Smith, 1997).

telic describes situations or verbs that have a built-in endpoint, e.g., "capture."

B Overview of Modeling Approaches

Table 3 gives a concise overview of existing modeling approaches.

C Further Linguistic Background

C.1 Eventuality Types

In Section 3.3, we explain a taxonomy of eventuality types according to Vendler (1957). As shown in Figure 2, Kenny (1963) adds *performances*, which are "actions that tend towards a goal." Mourelatos (1978) criticizes that these earlier analyses focus too much on predicates that require human agency, and suggests the terminology in Figure 2.

As explained in Section 3.3, Moens and Steedman (1988) introduce an inventory of eventuality types similar to those of Vendler (1957), but referring to the characteristics of whether an event is *atomic*, *extended* (*durative*) and whether it has a *consequent state*. The definitions of the types along with examples are given in Table 4.

Type of Model	Reference(s)	Categories Targeted / Task
Rule-based	Passonneau (1988b) (PUNDIT)	stativity
	Chen et al. (2021)	UMR aspect features
Progressive as indicator	Klavans and Chodorow (1992b)	stativity
Linguistic indicators	Brent (1991b)	stativity
	Siegel and McKeown (1996, 2000b)	stativity, boundedness
Naive Bayes + Decision Tree	Mathew and Katz (2009)	habituality
Random Forest Classifiers	Friedrich and Pinkal (2015a)	habituality
Logistic regression	Friedrich and Palmer (2014a)	lexical aspect
	Loáiciga and Grisot (2016)	boundedness
	Friedrich and Gateva (2017)	telicity
	Zarcone and Lenci (2008)	Vendler types
Distributional models	Kober et al. (2020)	telicity
SVM	Keelan (2012)	Leech's classes
	Hermes et al. (2015, 2018)	Vendler classes
Max Ent Sequence Tagger	Palmer et al. (2007)	situation entity types
CRF	Friedrich and Pinkal (2015b)	genericity
	Friedrich et al. (2016)	situation entities
word2vec + GRU	Becker et al. (2017)	situation entity types
GloVe + GRU	Zellers and Choi (2017)	verb-level Vendler classes
Paragraph-level neural model	Dai and Huang (2018)	situation entity types
ELMO + MLP	Govindarajan et al. (2019)	habituality (continuous features, UDS-G)
BERT + classifiers	Metheniti et al. (2021); Metheniti (2022)	telicity
BERT + tagger	e.g., Kondratyuk and Straka (2019)	UD aspectual features
BERT, GPT, ParBERT	Rezaee et al. (2021)	situation entity types
Generative model	Gantt et al. (2022)	event types

Table 3: **Computational systems and approaches** to modeling aspect.

Vendler (1957)	*dynamic*	*durative*	*telic*	Moens and Steedman (1988)	*atomic event*	*extended event*	*consequent state*	**Examples**
State	-	+	-	state	NA	NA	NA	know the answer, love Mary, understand
Activity	+	+	-	process	-	+	-	laugh, stroll in the park, run, swim
Accomplishment	+	+	+	culminated process	-	+	+	build a house, walk to school
Achievement	+	-	+	culmination	+	-	+	win a race, reach the top, recognize
Semelfactive	+	-	-	point event	+	-	-	tap, knock, hiccup, wink

Table 4: **Eventuality** types (Vendler, 1957; Smith, 1997) / **Temporal Ontology** (Moens and Steedman, 1988).

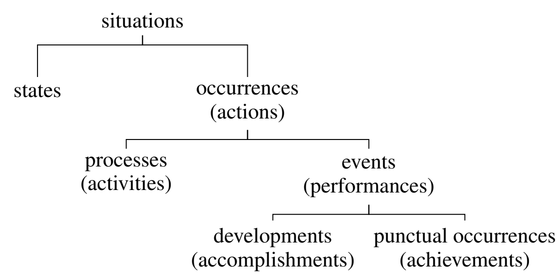

Figure 2: Classification of aspectual oppositions according to Mourelatos (1978), terms used by Vendler (1957) and Kenny (1963) in parentheses.

C.2 Leech's Classes

In addition to categories corresponding to those of Figure 2, Leech (1971) distinguishes between *Attitude* states ("hate," "hope") and *Relationship* states ("own," "resemble"). There are also two categories for *Perception* and *Cognition* verbs such as "hear," "see," or "feel," which are generally hard to classify along the stative-dynamic dimension (Comrie, 1976).

C.3 Details on Viewpoint Aspect

In some cases, viewpoint aspect (Section 4) and situation type interact. The perfective viewpoint is naturally available for the situation types (Section 3.3) activity, accomplishment, semelfactive and achievement; the imperfective viewpoint is available for states, activities and accomplishments (Smith, 1997). The usage of the perfective viewpoint with stative predicates indicates aspectual coercion as introduced in Section 3.3. For example, in Chinese the stative predicate "gāo" means to be tall. When used with the perfective marker "le," an ingressive meaning is implied, i.e., "tā gāo-le" translates as "he became tall."

C.4 Habituals: Borderline cases

Habitual sentences (Section 4.2) describe situations that are characteristic of an extended period of time (Comrie, 1976), a decision that is of conceptual rather than of linguistic nature. In fact, Filip and Carlson (1997) even argue that *sentential genericity*, which corresponds to habituality, is independent from tense and aspect, and that habitual sentences such as (14) can occur in the perfective.

(14) In the eighties, John went to work by bus.

The interaction of habituality and modality is by no means trivial (Hacquard, 2009) as illustrated by (15a), and negated sentences are another unclear case (15b-c).

(15) (a) I had to swim every day. *(habitual?)*
(b) John smokes. *(habitual)*
(c) John does not smoke. *(habitual?)*

The concept of habituality does not include *dispositional sentences* such as "John can swim," which denote abilities or preferences.

D Further Computational Background

D.1 Linguistic Indicators

Table 5 reports the full set of linguistic indicator features as proposed by Siegel and McKeown (2000b) and related works.

Feature	Example
frequency	-
past	*said*
perfect	*had won*
progressive	*is winning*
negated	*not/never*
particle	*up/in/...*
no subject	-
continuous adverb	*continually, endlessly*
evaluation adverb	*better, horribly*
manner adverb	*furiously, patiently*
temporal adverb	*again, finally*
in-PP	*in an hour*
for-PP	*for an hour*

Table 5: **Linguistic indicators** computed over large syntactically parsed text corpora (Siegel and McKeown, 2000b).

D.2 Aspect in UMR

As briefly described in Section 4.3, the UMR (Uniform Meaning Representation) (Van Gysel et al., 2021) framework approaches annotation of aspect from a typological perspective. Aspect is represented as an attribute for events in UMR, and since cross-linguistically languages mark aspect in different ways, both grammatically and lexically, the UMR aspect categories are arranged in a lattice. This allows UMRs to be annotated at the level of granularity that is most appropriate for a particular language.

The resulting lattice (Van Gysel et al., 2022) appears in Figure 3. The top-level of the lattice represents broad distinctions between aspect categories across languages, while the bottom level represents the finer distinctions.

SE type	Example
Eventualities	
STATE	The colonel owns the farm.
EVENT	John won the race.
REPORT	"...", said Obama.
General Statives	
GENERIC SENT.	The lion has a bushy tail.
GENERALIZING SENTENCE	Mary often fed the cat last year.
Abstract Entities	
FACT	(I know) that she refused the offer.
PROPOSITION	(I believe) that she refused the offer.
QUESTION	Who wants to come?
IMPERATIVE	Hand me the pen!

Table 6: Situation entity types (Smith, 2003).

D.3 Situation entity inventory

Table 6 provides the complete inventory of situation entity types. Section 5 offers a partial inventory, focusing on those categories for which aspect is a key determining property.

D.4 Event Classes in TimeBank

TimeML (Pustejovsky et al., 2003, 2010) events are "situations that occur or happen," but also include "states or circumstances in which something obtains or holds true." Thus, the usage of the term *event* denotes a *situation* in the sense of Smith (1997). In TimeML, tensed verbs, stative adjectives and event nominals that describe situations temporally located in the text can be marked with the categories shown in Table 7.

TimeML event classification corresponds to a word-sense level task. Saurí et al. (2005), in their event recognition system, simply assign the class that was most frequently observed for each verb type in the training data to events and reach an accuracy of 82.3% on TimeBank 1.2. The top-performing systems (Jung and Stent, 2013; Bethard, 2013; Chambers, 2013) of the 2013 TempEval challenge (UzZaman et al., 2013) use corpus-based features, WordNet synsets, parse paths and fea-

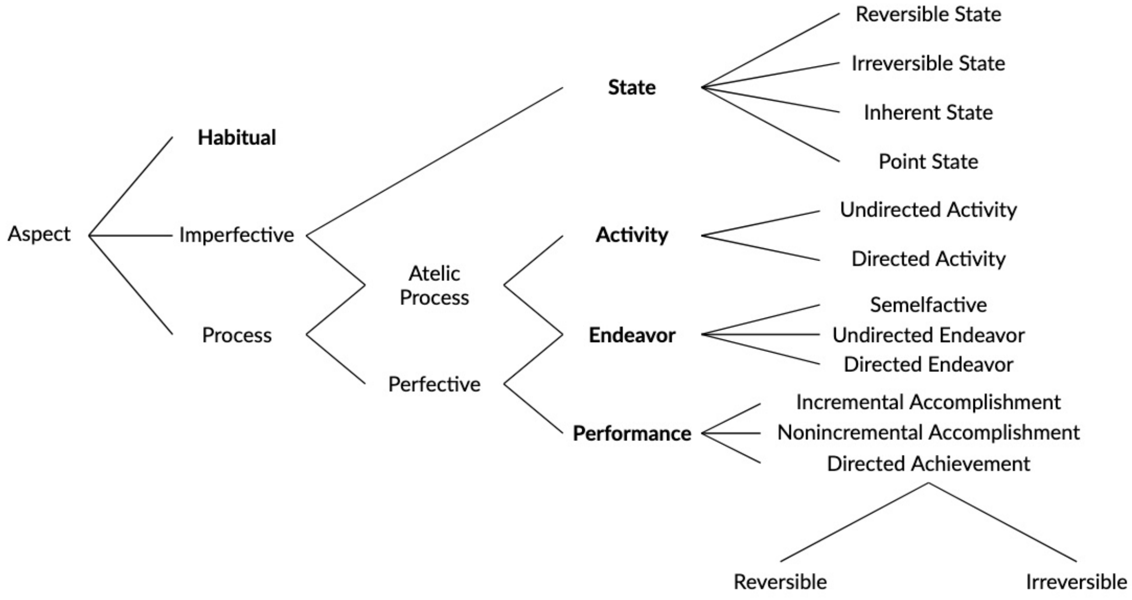

Figure 3: UMR annotation lattice for aspect (Van Gysel et al., 2022).

tures from typed dependencies to classify events as a joint task with determining the event's span. Bethard and Martin (2006) phrase the recognition of EVENTs and their semantic class as a chunking task using syntactic-semantic features such as part-of-speech, morphological information, word clusters and WordNet hypernyms. Llorens et al. (2010) extend this idea by using a conditional random field (Lafferty et al., 2001) enhanced with semantic role information. Costa and Branco (2012) explore the usefulness of a wider range of explicitly aspectual features, including linguistic indicators, for temporal relation classification in Portuguese.

More recent TempEval challenges (Bethard et al., 2016, 2017) did not offer a task for classifying event classes.

Event Class	Explanation and Examples
OCCURRENCE	situations that happen: *die, crash, merge*
STATE	circumstances in which sth. holds: *like, own, the kidnapped girl, on board*
I_ACTION	intensional actions: *try, persuade*
I_STATE	intensional states: *love, believe, enjoy*
ASPECTUAL	predicates that pick out a phase of the event: *begin, start, continue*
I_REPORTING	capture attribution: *said*
PERECPTION	physical perception of another event: *see, hear, feel*

Table 7: TimeML event classes (Pustejovsky et al., 2003).

E (Potential) Applications

Despite its importance for understanding, few NLP tasks explicitly incorporate aspectual information. Here we discuss the potential uses in two application areas; many other areas could also benefit, including argumentation mining, computer-aided language learning, and information extraction.

Temporal Processing. Together with tense, aspect is essential to the linguistic system encoding temporal information (Smith, 1997, 2003). Yet there is little work systematically leveraging aspectual information in temporal relation extraction systems, possibly because TimeML event classes (Pustejovsky et al., 2003, 2010, Appendix D.4) are not strictly rooted in linguistic theory. We believe that recognizing the mode of discourse and the type of temporal progression (which is different e.g., in a narrative, an encyclopedia entry, or a news article) is key to recognizing temporal structure in text.

With the help of templates, Vashishtha et al. (2020) re-cast existing temporal relation and temporal duration datasets as natural language inference (NLI) datasets. Existing models perform well on classifying English sentence-internal relations this way, indicating that the number of tense pair patterns is somewhat limited. However, using large-pretrained transformer based models with a paradigm shift (Sun et al., 2021) is unlikely to be the solution to temporal reasoning, as shown by Kober et al. (2021). They create a dataset consisting of 11,138 pairs of short sentences labeled with a binary annotation scheme (*entailment* vs. *non-entailment*), addressing tense and the perfect vs. progressive distinction in English. For example, “John is visiting London” entails “John has arrived in London,” but “John will visit London” does not. They evaluate a range of models including non-contextualized and contextualized embeddings pre-trained on SNLI (Bowman et al., 2015) and DNC (Poliak et al., 2018) and find that none of these models outperforms a majority class baseline on the new dataset. This suggests that embedding spaces struggle to capture the more latent nature of tense and aspect.

In our view, a first step towards leveraging the knowledge about tense and aspect provided by linguistic theory could be to systematically study how recent language models and NLP systems succeed or fail with regard to these categories, as pioneered in the pre-neural age for example by Zhang and Xue (2014); Reichart and Rappoport (2010).

Machine Translation. While recent MT systems often perform well, they do not systematically treat aspectual notions (Vanmassenhove et al., 2017), but rather rely on common translations for a particular domain. Consider the following example. In contrast to English, the German simple present may or may not indicate habituality. When translating to English, a choice must be made based on the context. The translation[6] in (16b) wrongly indicates habitual viewpoint, while the context suggests an ongoing event (the correct translation would be “is riding”).

(16) (a) Tim ist schon weg. Er <u>fährt</u> gerade mit dem Fahrrad zur Arbeit. *(ongoing event)*
(b) Tim is already gone. He just <u>rides</u> his bike to work. *(habitual, wrong)*

The cue need not be in the immediately preceding sentence, but could occur much earlier in the discourse. In (17), “fährt” should be translated as “is riding,” but the cue is in the first sentence.

(17) (a) Tim ist schon weg. Er arbeitet an der Universität. Er <u>fährt</u> gerade mit dem Fahrrad zur Arbeit. *(ongoing event)*
(b) Tim is already gone. He works at the university. He just <u>rides</u> his bike to work. *(habitual, wrong[7])*

Aspect is a crucial component of semantic interpretation, and special care is needed to make sure translations correctly capture it.

F ChatGPT Logs

Accessed January 26 and 27, 2023.

In this section we provide logs from our interactions with ChatGPT on some of the main topics covered by this survey. Our aim is to investigate the extent to which a system like ChatGPT could serve as a tool for teaching about aspect; in other words, what does ChatGPT know about aspect?

Note that some of ChatGPT’s analyses are surprisingly accurate, however, it also presents several completely wrong analyses and only produces correct analyses if guided to them by the user during the generation process.

Each subsection contains one or more segments from a chat log about a particular aspectual notion. We use the formatting shown below:

- *Prompt*
- Text generated by ChatGPT (with interesting parts highlighted in **bold**)
- *[Comments]*

F.1 Talking about Stativity

Consider the following text. "Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, and ‘what is the use of a book,’ thought Alice, ‘without pictures or conversation?’" Which verb phrases refer to states, which verb phrases refer to events in this passage?

In this passage, "was beginning to get very tired" and "having nothing to do" are verb phrases that refer to states. "She had peeped" and **"had no pictures or conversations in it" are verb phrases that refer to events**.

[The verb phrases "have nothing to do" and "had no pictures or conversations in it" are states; "had peeped" is an event; "was beginning" is a special case at it signals the beginning of a state

[6] Google Translate, October 5, 2022

[7] Google Translate, October 19, 2022

(inchoative aspect). ChatGPT's response is simplistic and in some cases completely wrong.]

Why is "had no pictures or conversations in it" an event?

The phrase "had no pictures or conversations in it" is an event because it describes something that happened in the past, specifically that the book did not contain pictures or conversations. The verb "had" is in the past tense which indicates that **the action of the book not having pictures or conversations** is completed, and it happened in the past. An event can be a completed past action or something that is happening currently or in the future.

[...]

Could you interpret the phrase as a state as well?

It could be interpreted as a state in the sense that the book currently lacks pictures or conversations, however, in the context of the passage **it is clear that the author is describing a past event** where Alice had looked into the book and found that it did not contain pictures or conversations, **as opposed to the book being in a state of lacking pictures or conversations**. The past tense of the verb "had" and the context of Alice peeping into the book suggest it's an event that happened in the past. Additionally, the phrase "'what is the use of a book,' thought Alice, 'without pictures or conversation?'" also suggest that it's an event that happened in the past. So, it can be interpreted as a state, but in this context, it is more likely an event.

[*Even in this context, the verb phrase describes a state (a property of the book).*]

I agree that "peeped into the book" is an event. But what about "the book had no pictures in it"?

In the sentence "the book had no pictures in it", the phrase "had no pictures in it" describes the state of the book, specifically that it did not contain any pictures. The verb "had" is in the past tense, which indicates that this is a statement about the book's state at a specific point in the past. The phrase describes a state that existed at a certain point in the past. Even though "had" is in past tense, it is used to describe a state, not an event. Additionally, the phrase is not describing something that just happened, but a characteristic of the book, which implies a state.

[*This explanation is correct. However, the above conversation clearly illustrates that ChatGPT generates text explaining the aspectual nature of the various verb phrases without having a deeper understanding of their semantics and hence without being consistent in its relatively random interpretations.*]

Telicity is a property of an event type expressed by verbs, i.e., referring to events that have an endpoint. Atelic verbs refer to event types that do not have a built-in endpoint. Can you give me examples for each case?

Sure, here are some examples of telic verbs and the events they describe:

"She finished the book." (finish)

"He graduated from college." (graduate)

"They completed the project." (complete)

And here are some examples of atelic verbs and the events they describe:

"She is reading a book." (read)

"He is studying for the exam." (study)

"They are working on the project." (work)

[*These examples are correctly classified except for "read a book."*]

Why is the event type "read a book" atelic?

The event type "read a book" is considered atelic because it does not have a built-in endpoint. In other words, when someone says "She is reading a book," it does not necessarily imply that the act of reading will come to a definite conclusion. Reading a book can be an ongoing activity that may or may not be completed. **It does not have an inherent endpoint.**

[*Here, the analysis confuses between progressive aspect and analysing telicity at the event type (predicate) level. (Admittedly, this is a very fine-grained notion that humans also need to think carefully about.) In "She was reading a book", the event type "read a book" is telic, while the sentence in the progressive presents the ongoing situation of reading a book, hiding the end points.*]

Wouldn't the endpoint mean she finished reading the entire book?

Yes, in the sentence "She finished the book" the endpoint would mean she finished reading the entire book. The verb "finish" implies that an action has come to completion, so when it is used in the context of reading a book, it implies that the book has been read in its entirety. However, it's worth noting that some verbs can have multiple meanings and context can affect the interpretation of a sentence.

In contrast, when someone says "She is reading a book" it doesn't necessarily imply that the act of reading will come to a definite conclusion or that she will finish reading the entire book, she could stop reading at any time. And that's why the event is atelic.

I think the event is still telic, but the sentence focuses on a phase of it.

You are correct that the event of reading a book, in general, is telic, as it does have a natural endpoint, which is when the person finishes reading the book. However, when we use the verb "read" in the present progressive form "is reading," it describes the action as being ongoing and does not imply a definite conclusion. It describes the event as a phase or a state of being rather than a completed action.

So in the sentence "She is reading a book," the emphasis is on the current state of reading and not on the endpoint of finishing the book. The sentence "She finished the book" implies that the action of reading is completed and the verb "finish" implies that an action has come to completion.

So, in summary, while the overall event of reading a book is telic, in certain sentence constructions, such as "She is reading a book," the focus is on the ongoing state of the event, making it atelic.

[*The last part of the analysis is correct, except that the event of reading a book is always telic, and the progressive focuses on part of the situation. This does not make the event type atelic. The explanations given in this section focus more on the progressive vs. non-progressive distinction, not on the telic vs. atelic distinction.*]

F.2 Eventuality Types

(Prompted with the first two paragraphs of Section 3.3 of this paper including examples.)

Can you please analyse: "Tom held a meeting"

The sentence **"Tom held a meeting" is an example of an achievement** according to Vendler's (1957) classification of situation types. The sentence describes an event that has an endpoint, the meeting, and it is also a change of state. The verb "held" is also telic, meaning it has a specific endpoint and it does not include a process leading up to this point, which fits with the definition of an achievement. The phrase "held a meeting" implies that there was a start and an end to the meeting and it implies that it has been completed.

Can you please analyse: "Tom was holding a meeting"?

The sentence **"Tom was holding a meeting" is an example of an activity** according to Vendler's (1957) classification of situation types. The sentence describes an ongoing process of holding a meeting without any specific endpoint. The verb "was holding" is also atelic, meaning it does not have a specific endpoint, which fits with the definition of an activity. The phrase "holding a meeting" implies that there is a meeting in progress, but it does not specify when or if the meeting is completed.

[...]

[*The predicate "hold a meeting" is in fact an accomplishment, as it is dynamic, durative, and telic. The change in grammatical aspect from perfect to progressive does not change the underlying eventuality type. Again, the system seems to be misled by the progressive aspect.*]

Can you please analyse: "Anna was swimming across the lake"?

The sentence "Anna was swimming across the lake" is an example of an activity according to Vendler's (1957) classification of situation types. The sentence describes an ongoing process of swimming without any specific endpoint. The verb "was swimming" is also atelic, meaning it does not have a specific endpoint, which fits with the definition of an activity. **The phrase "across the lake" indicates that Anna is swimming over a significant distance, covering a large geographical area,** but the sentence does not specify when or if the journey is completed.

[*The default interpretation for this sentence is as an accomplishment, with the goal of reaching the other side of the lake. There is a second plausible, though less likely, interpretation in which Anna is engaged in the activity of swimming in a location that is across the lake from the speaker. ChatGPT proposes a third, highly-unlikely interpretation, akin to describing something as happening "all across the world."*]

Can you please analyse "Tim swam across the pool three times".

The sentence **"Tim swam across the pool three times" is an example of an accomplishment** according to Vendler's (1957) classification of situation types. The sentence describes a process (swimming) that leads to a built-in terminal point, the completion of swimming across the pool. The verb "swam" is telic, meaning it has a specific endpoint, which fits with the definition of an accomplishment. Additionally, the phrase "across the pool" indicates that Tim is swimming over a significant distance, covering a large geographical area and the sentence implies that the journey has been completed. The number "three times" indicates that the process was repeated a specific number of times, and it gives the sense that the accomplishment has been reached multiple times.

[*When substituting "pool" for "lake" and adding the frequency adverb "three times", GPT's interpretation of the event type of "swim across X" is correct, again hinting at the randomness of the generated explanations.*]