AutoAspect: Automatic Annotation of Tense and Aspect for Uniform Meaning Representations

Daniel Chen
University of Colorado at Boulder
Boulder, CO, USA
daniel.chen-1@colorado.edu

Martha Palmer
University of Colorado at Boulder
Boulder, CO, USA
martha.palmer@colorado.edu

Meagan Vigus
University of New Mexico
Albuquerque, NM, USA
mvigus@unm.edu

Abstract

We present AutoAspect, a novel, rule-based annotation tool for labeling tense and aspect. The pilot version annotates English data. The aspect labels are designed specifically for Uniform Meaning Representations (UMR), an annotation schema that aims to encode crosslingual semantic information. The annotation tool combines syntactic and semantic cues to assign aspects on a sentence-by-sentence basis, following a sequence of rules that each output a UMR aspect. Identified events proceed through the sequence until they are assigned an aspect. We achieve a recall of 76.17% for identifying UMR events and an accuracy of 62.57% on all identified events, with high precision values for 2 of the aspect labels.

1 Introduction

As the field of Natural Language Processing advances, there are increasing demands for more sophisticated applications and richer representations. Abstract Meaning Representations (AMR; Banarescu et al. 2013), and their more recent crosslingual incarnation as Uniform Meaning Representations (UMR; Van Gysel et al. 2021), are a response to that demand. AMR/UMRs provide an abstract, directed acyclic graph representation of a complete sentence, focusing on the underlying "who" did "what" to "whom" elements of the events being described. The more information that can be associated with those events, in terms of whether they have been completed, or whether they have achieved their intended results, the better.

The increased richness of UMR Tense, Aspect and Modality annotations, as described below, can more clearly identify the completion and achievement of events in a cross-lingual context, providing a firmer baseline for comparing typologically distinct languages. Automating such a complex semantic processing task provides valuable qualitative and temporal crosslingual features that applications like translation models and virtual assistants can utilize to more accurately capture the semantic nuances of events. Given the substantial amounts of English AMR annotation, the question immediately arises of how to efficiently add these new annotation features to pre-existing English AMRs.

This paper describes an implementation of an automatic system that relies on VerbNet, a rich lexical resource, as the basis for categorizing event descriptions according to the Aspect guidelines discussed below. Our initial results are quite promising, and there are obvious next steps to take.

2 Background Information

Previous automatic annotation models operate on different definitions of aspect. In Friedrich et al. (2016), clauses are annotated for situation entity types, which capture some of the same semantic distinctions as the UMR aspect annotation scheme, including state and habitual. Friedrich et al. (2016) also include modal distinctions in their annotation, such as questions and imperatives. Unlike Friedrich et al. (2016), the UMR aspect annotation distinguishes between different types of dynamic (non-stative) clauses (Activity, Endeavor, and Performance); these are all annotated as Event in Friedrich et al. (2016).

Friedrich and Gateva (2017) annotate a binary telicity distinction: telic vs. atelic. The UMR aspect annotation annotates a three-way distinction for non-stative events, which takes both the qualitative and temporal dimensions of event semantics into account.

The UMR aspect annotation consists of a feature assigned to events that indicates their internal qualitative and temporal structure (Van Gysel et al., 2019, 2021). Every node classified as an "event" in UMR receives an aspect annotation. UMR defines "event" based on the typological prototype of a verb as described in Croft (2001), and Croft (in press). UMR events exhibit either the prototypical information-packaging of a verb, predication (as opposed to modification and reference), or the prototypical semantic class of a verb, a process (as opposed to a property or an entity).

The aspectual distinctions made in UMR are based on Croft (2012)'s two-dimensional analysis of aspect and build on the aspect annotations from Donatelli et al. (2018, 2019). Croft (2012) analyzes aspectual structure as having both a temporal dimension and a qualitative dimension; the temporal dimension measures out the event's unfolding over time and the qualitative dimension measures out the change that occurs (or does not occur) during the event.

The UMR aspectual distinctions do not have a direct correspondence with either specific verbs or specific constructions in a language; instead, they annotate the aspectual structure of an event in its context. In order to ensure the maximum cross-linguistic comparability of annotation values, UMR uses lattices of compatible annotation values for certain annotation categories, including aspect (Van Gysel et al., 2019).

In addition to the lattices, UMR also considers a certain level of specificity in annotation as the base level, corresponding to distinctions that tend to be straightforward to annotate in a majority of languages. For aspect, there are four base-level annotation values: STATE, ACTIVITY, ENDEAVOR, and PERFORMANCE. In addition, there is a HABITUAL value. The annotation of these categories manually relies on a series of decisions that distinguish event types.

First, event nominals are annotated as PROCESS, one of the more coarse-grained categories on the lattice. Processes in reference (i.e., event/action nominals) lack grammatical clues as to their aspectual structure in English and many other languages; therefore, they are simply annotated as PROCESS.

Next, the HABITUAL value applies to all events that are repeated on a regular basis, regardless of the internal aspectual structure of each individual (repeated) event. This means that further aspectual distinctions (including between PROCESS and STATE) are collapsed for habitual events.

The rest of the aspectual values make a distinction between stative and non-stative events. STATE is the base-level value that captures all stative events; non-stative events are divided into ACTIVITY, ENDEAVOR, and PERFORMANCE.

UMR defines stative events as those in which no change occurs on the qualitative dimension during the event (Vendler, 1967; Croft, 2012). In addition to prototypical stative events, UMR extends the STATE value to "nonverbal predication" (*He **is** a teacher. / There **is** a sandwich in the kitchen.*), events modalized by ability modals (*This car can **go** 100mph.*), and modal complement-taking predicates (*She **wants** to eat at noon.*).

In addition, UMR uses the STATE value for a class of events termed "inactive actions" in Croft (2012). Inactive actions are semantically intermediate between states and processes; this class includes verbs of sensation, perception, cognition, emotion, and position. These types of events can variably be construed as either states or processes, even within the same language. Since the construal of an inactive action can be hard to ascertain in context, UMR annotates all inactive actions with the STATE value.

The rest of the base level aspect annotations (ACTIVITY, ENDEAVOR, PERFORMANCE) work to characterize non-stative events (PROCESSES in UMR). These events are defined as those that involve change on the qualitative dimension.

The distinction between the ACTIVITY annotation value and the ENDEAVOR and PERFORMANCE values is based on whether the event is still ongoing at Document Creation Time (DCT), or whether it has ended. Events annotated as ACTIVITY indicate that the event may be ongoing at DCT.

Both ENDEAVOR and PERFORMANCE characterize non-stative events that have ended prior to DCT. The ENDEAVOR and PERFORMANCE values differ from each other in whether the event has been terminated or completed. Events annotated as ENDEAVOR signal that the event has terminated, without reaching completion. This means that the event has not reached a distinct result state on the qualitative dimension. The PERFORMANCE label indicates that an event has been completed, reaching a distinct result state. The UMR PERFORMANCE value corresponds to Vendler's achievements and accomplishments.

UMR also incudes a number of more fine-grained aspectual types on its aspect lattice (see Van Gysel et al. 2019). Many of these fine-grained aspect values make distinctions that cross-cut the base-level annotation values. These include whether an event is punctual or durative, which cross-cuts the ENDEAVOR/PERFORMANCE distinction. For durative ENDEAVORS and PERFORMANCES, in addition to all ACTIVITIES, there is also the cross-cutting distinction of incremental vs. nonincremental change (Dowty, 1991; Croft, 2012).

The base-level UMR aspectual values were selected because they capture the most salient pieces of aspectual structure and they can be consistently annotated, even in languages with minimal grammatical aspect marking like English. The presence or absence of change on the qualitative dimension is captured by the PROCESS vs. STATE distinction. Boundedness on the temporal dimension is captured for processes by the distinction between ACTIVITY and ENDEAVOR/PERFORMANCE. Finally, boundedness on the qualitative dimension is captured by the distinction between ENDEAVOR and PERFORMANCE.

3 Rule-Based Classification

The rule-based classifier[1] follows the sequential manual annotation steps as closely as possible, immediately exiting the sequence as soon as an annotation label has been assigned by a numbered step. The annotation loop processes text by sentence, such that every identified event in a sentence receives an annotation.

3.1 Step 1a: Syntactic Split

Step 1a syntactically separates annotation branches into *verbs* and *event nominals*. Since all event nominals are assigned the same aspect by Step 1b, the *event nominals* branch does not pass through the sequence of rules and is analyzed separately. The *verbs* branch is explored by allowing any tokens that the spaCy English Web Core Large NLP parser (Honnibal et al., 2021) marks as having a part-of-speech (POS) corresponding to any of the Penn Treebank VERB tags[2] to continue on to the sequence of rule-based decisions (Steps 2a-8).

[1]All code can be found at https://github.com/dchensta/AutoAspect.

[2]VB (base form), VBD (past tense), VBG (gerund or present participle), VBN (past participle), VBP (non-3rd person singular present, and VBZ (3rd person singular present).

We experimented with various parsers for extracting verbal events, which included running the ClearTAC parser (Myers and Palmer, 2019) and iterating through the list of events SemParse (Gung, 2020; Gung and Palmer, 2021) generates for the input sentence. At this stage, we are able to extract more verbal events using the spaCy NLP parser.

3.2 Step 1b: Event Nominals Branch

While SemParse misses some verbal events, it is the sole parser that can identify nominal tokens that correspond to VerbNet frames. With this functionality, a derived nominal like *explosion* shares the same VerbNet ID as its parent verb *explode*. It can thus be recognized by SemParse as an event nominal. Thus, to follow the *event nominals* annotation branch, we run SemParse separately and only extract spans of text in the sentence that constitute event nominals.

We restrict identification of event nominals to spans of text that do not contain any Penn Treebank VERB tags. Thus, the span *historic visit* is identified as an event nominal because both tokens have a non-verbal POS: adjective and singular noun, respectively. But the span *to lay the groundwork* does not get identified as an event nominal because *lay* has a verbal POS and will thus be handled in the verbal annotation branch. All event nominals receive the PROCESS label from Step 1b, and thus exit the annotation search.

The *event nominals* branch is explored only after the *verbs* annotation branch terminates. The final list of events and corresponding aspects collected for each sentence combines the outputs of the *verbs* and *event nominals* annotation branches.

3.3 Steps 2-3: Verbs Branch

The remaining steps all take place in the *verbs* annotation branch. Step 2a handles non-verbal predication, which analyzes all English copula forms that are followed by predicate nominals, adjectives, and locationals. It assigns all copula forms the aspect STATE, with the condition that the copula does not function as a helping verb, i.e. no verb form directly follows it. Note that UMR annotates the nominal or adjectival predicate rather than the copula, so the sentence *"He is friendly."* results in a UMR aspect annotation for the event *friendly* rather than the verbal token *is*. We compare the aspect results accordingly, not requiring the token that receives the label from AutoAspect to correspond exactly to the predicate token(s) representing

the UMR event.

Step 2b annotates for STATE as well, and it does so based on VerbNet class, requiring a semantic processing tool. We run SemParse and extract the VerbNet senses corresponding to the verbal tokens in the sentence. For each VerbNet sense *y* in the pre-determined list of VerbNet class IDs that are labeled STATE, we match the SemParse sense *x* to *y* by backing off to the most coarse-grained class. For example, the SemParse VerbNet class ID for the lemma *want* is *x* = *want-32.1-1-1*, but the specified class in the pre-determined list is *y* = *want-32.1*. We run a regular expressions match to ensure that the class ID *y* is contained within the span of the more fine-grained class ID *x*, confirming that *x* and *y* belong to the same class. Any verb event that does not receive a STATE label from Steps 2a and 2b receives the umbrella label PROCESS and continues on to Step 4.

3.4 Steps 4-8: Verb Branch

Steps 4-8 subcategorize the umbrella label PROCESS carried over from Step 2b into the labels ACTIVITY, PERFORMANCE, and ENDEAVOR.

Steps 4 and 5 continue to classify based on VerbNet class. Step 4 assigns all verbs that receive a participial[3] Penn Treebank label from spaCy the label ACTIVITY. Verbs with inceptive and continuative auxiliary verbs[4] like *started* and *continued* also receive the ACTIVITY label. Step 5 assigns all completive auxilaries[5] — like *finished* — the label PERFORMANCE and all terminative auxiliaries[6] — like *stopped* — the label ENDEAVOR.

Step 6 assigns verbs that occur in clauses with container adverbials — marked by the preposition *in* — the label PERFORMANCE. Step 6 also assigns verbs that occur in clauses with durative adverbials — marked by the preposition *for* — the label ENDEAVOR. We use the spaCy dependency parser to check that those prepositions occur in the same clause as the verb being analyzed.

Step 7 annotates verbs that occur with non-result paths as ENDEAVOR, marked by prepositions like *around*, *along*, and *past*. A non-result path is defined for the tool as the occurrence of a preposition immediately after the verb it is a dependent of, e.g. in the sentence *"He **walked along** the river."*, where *along* occurs immediately after the main verb *walked*.

Finally, Step 8 assigns any verbal event that has not yet received an annotation the label PERFORMANCE.

4 Results

Many of the annotation steps are themselves complex semantic processing tasks. Therefore, the performance of this rule-based model depends heavily on the limits of the classifier components being utilized. Given the lack of substantial training data that would be required for a machine learning-driven implementation, we focus on specific examples that reveal the linguistic blind spots of our model.

4.1 Annotation Results

The following results were generated from running AutoAspect on four gold standard news articles from the LDC REFLEX English core set of newswire and web text documents (Strassel and Tracey, 2016) that were annotated by our team. Table 1 shows results for the critical subtask of event identification, which yielded a recall of 76.17%, a reasonably high recall for a pilot study with limited gold data. For this specific task, recall is the only appropriate measure for statistical analysis, because precision penalizes the model for annotating an event that is not present in the gold data. Given the variability in using syntactic and semantic cues in the annotation manual, it is more appropriate to see how much of the human-identified labels it is capturing.

For example, AutoAspect analyzes the initial split by analyzing Penn TreeBank verbs separately from Penn TreeBank nominals. In the phrase *investigation of bombing campaign*, it is important to note when AutoAspect only identifies the *investigation* event and fails to identify the *campaign* event. However, a sentence like *"I think the court would be highly politicized."* does not assign an aspect to the verb *think*, but to the event *be_politicized*. Since AutoAspect will analyze every token labeled with a Penn TreeBank VERB part-of-speech, it would be disingenuous to penalize the model for labeling an event that is subsumed by a UMR event purposefully abstracting away from strictly syntactic cues.

[3] VBG (present participle or gerund), VBN (past participle)
[4] *begin-55.1*, *continue-55.3*, and *sustain-55.6*
[5] *complete-55.2*
[6] *stop-55.4*

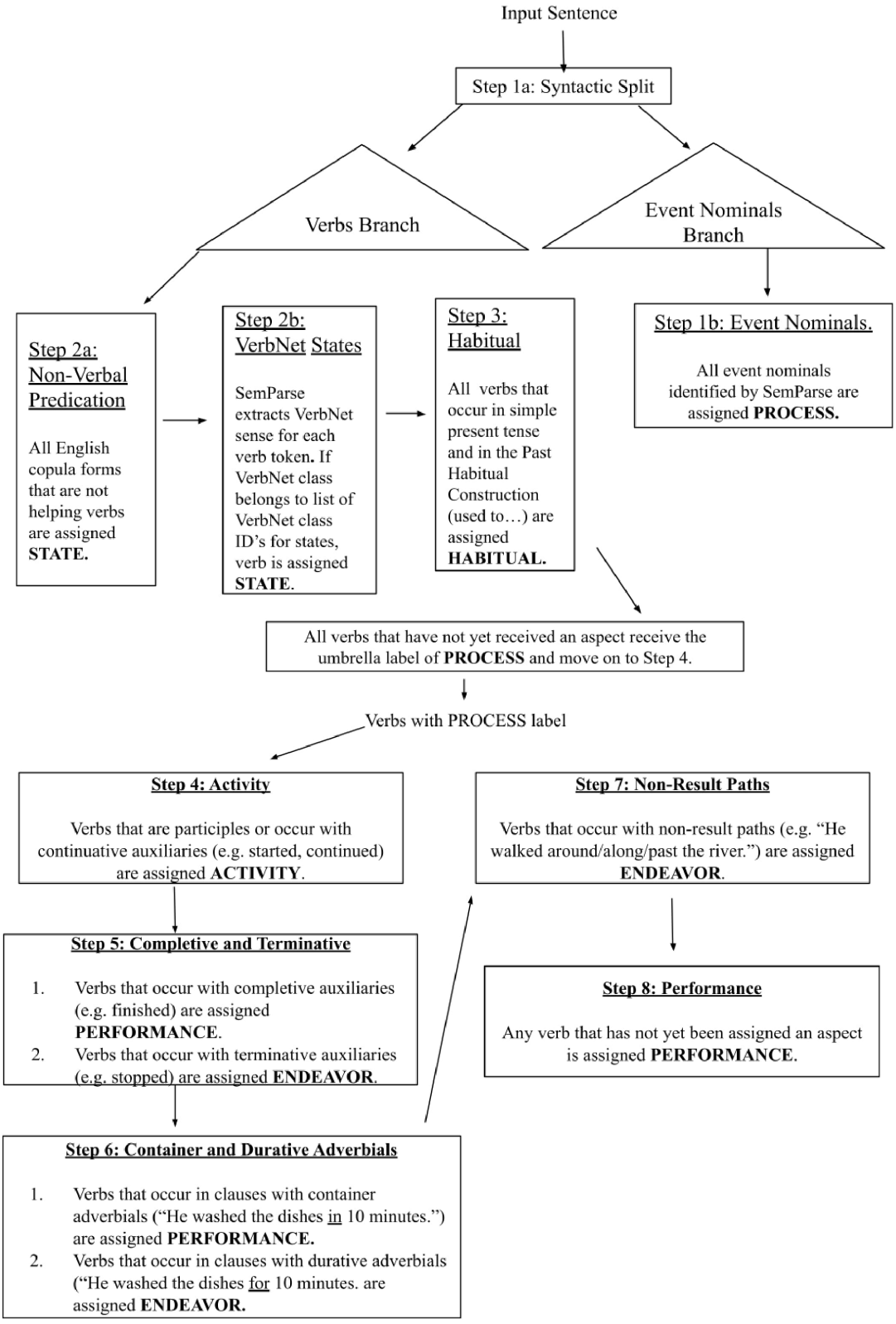

Figure 1: Flow Chart of Automatic Classification

Task	FN	TP	Recall	Acc
Event Nominal ID	56	179	76.17	
ID-ed Events Performance		112		62.57

Table 1: Recall for event identification subtask and performance of model on all gold events that were identified by the model, across all aspect labels.

Of the 235 gold events spread throughout the 4 gold files, the model failed to identify 56, leaving 179 events that were identified. Of those 179, the model correctly labeled 112 gold events out of the 179 it could identify, yielding an accuracy of 62.57%.

Label Error	FP	TP	# Gold	Precision
HABITUAL	27	3	5	10
STATE	7	43	71	86
ACTIVITY	12	6	9	33.33
PROCESS	1	4	56	N/A[7]
PERFORMANCE	15	56	94	78.87
ENDEAVOR	5	0	0	0

[7]Given that almost all of the 56 events the model failed to detect were event nominals, the precision calculation for PROCESS is vacuous.

Table 2: Counts and precision of each type of mislabeled annotation. For each aspect label, FP and TP counts are from the 179 events that AutoAspect successfully detected, number of gold events is from the 235 total gold events.

In Table 2, we show the counts for each type of incorrect label that AutoAspect assigned to an event. This allows us to identify both the linguistic errors made by the model and which errors it makes more frequently. Further development of the tool can thus identify the error-triggering linguistic inputs and improve how a specific rule-based component processes those linguistic inputs.

Table 2 also depicts high precision for STATE and PERFORMANCE, indicating that SemParse is successfully matching VerbNet states to verb tokens and that gold PERFORMANCE tokens are generally able to avoid triggering Steps 2a-7. We did obtain poor precision for HABITUAL and ACTIVITY, showing that the model is too accepting of false positive inputs in Steps 3 and 4.

4.2 Linguistic Analysis

The following linguistic inputs consistently led to errors in the output of AutoAspect.

Sentence	UMR Events
A. Clinton's ***signature*** would be an important symbolic ***gesture***.	sign, be_important, gesture
B. He told survivors of the ***genocide*** there the United States would support the International Court.	tell, genocide, support
C. ***Pleasure.***	be_pleasure
D. Despite American ***objections***, the International Criminal Court enjoys wide support.	object, enjoy
E. ...it will sell weapons to Iran, contrary to the earlier made ***agreement***.	sell, agreement
F. Carla Delponte briefly contemplated an investigation of NATO's bombing ***campaign***.	contemplate, investigate, campaign

Table 3: Missed event nominal corresponding to a gold UMR event in the sentence is bolded and italicized.

Sentence	UMR Events
G. Part of the purpose of this **historic visit** is to lay the groundwork...	visit, be_part, lay
H. It is his **decision**.	decide
I. But the **opposition** here in the United States is intense.	opposition, intense
J. Carla Delponte briefly contemplated an **investigation of NATO's bombing campaign**.	contemplate, investigate, campaign

Table 4: Accurately identified span containing a UMR event nominal is bolded.

1. **Less Explicitly Deverbal Event Nominals:** Failure to detect event nominals made up 45.5% of errors. Table 3 depicts event nominals that SemParse did not detect and Table 4 depicts event nominals that SemParse did detect. Sentence C is notable because it involves a nominal found in a dialogic omission of the main verb. SemParse still fails to identify *pleasure* as an event in the sentence *"It's been a pleasure."*, citing it as an attributive argument of the main verb, the event *seem-109-1-1*.

   These examples indicate that abstracting away from syntactic cues like having a main verb remains a difficult NLP task. SemParse is trained on Unified PropBank corpora, mapping nominal and adjectival predicates to VerbNet roles. Since VerbNet roles are syntactically defined for verbs, mappings exist for linking sentences like *"John has a fear of spiders"* to *"John fears spiders"* and *"John is afraid of spiders"*(Gung, 2020). Thus, an event like *campaign* in Sentence E will not be identified because SemParse currently only

Sentence	UMR Event	AutoAspect	Gold Aspect
K. ...the Pardon Commission, after it has made its decision, **sends** it to the President...	send	HABITUAL	HABITUAL
L. He will spend the next several days at the medical center there before he **returns** home with his wife Sherry.	return	HABITUAL	PERFORMANCE
M. Marsha, thank you very much for **speaking** with us.	speak	ACTIVITY	PERFORMANCE
N. ...they were afraid of the spy mania **rising** in Russia...	rise	ACTIVITY	ACTIVITY

Table 5: Annotation of present tense verb forms.

identifies event nominals/adjectivals that function as arguments of the main verb of the sentence, *contemplated*. A human annotator can identify another argument structure where the noun phrase headed by *investigation* has its own argument roles that could be identified as event nominals, but SemParse identification requires clear sentential structure as input and tends to be more limited to the main verb. Even then, event nominal identification is not guaranteed, given that *survivors* and *genocide* in Sentence B go undetected, despite being the direct object of the main verb *told*.

One possibility is that SemParse handles more explicitly deverbal nominals better. In Table 3, less explicitly deverbal nominals like *signature* and *gesture* are undetected. Table 4 shows that the deriving suffix *-tion* appears to make for more readibly detectable event nominals in *decision*, *opposition*, and *investigation*, all core arguments of their main verb. In F, like *gesture*, the nominal *visit* shares an identical form with its verbal lemma, but SemParse identifies *visit* and not *gesture*. Notably, the event nominal in F also does not occur as part of the core arguments of the main verb *is*, and was successfully identified as its own nominal phrase.

But SemParse also missed *objections* in D and *agreement* in E, both nominals that have a transparently derivative suffix that attaches to the verb lemma. Both of those event nominals occur in adjunct clauses that are not core arguments of the main verb and themselves do not contain a verb.

2. **Dialogic Sentences:** In addition to Sentence C, UMR annotates dialogic sentences like *"One last question."* as a singular event that labels the adjective: *be_last*. The current syntactic split of verbs and nominals does not allow AutoAspect to label predicative nominals and adjectives that lack a main verb. Additionally, multi-sentence coreference is common in dialogue, as in the sentences *"Is this case likely to strain US-Russian relations? I'm afraid it might."*, where an event from the previous clause (*strain*) becomes elided in a successive clause.

3. **Present Tense Verbs:** Table 5 depicts mislabeled and correctly labeled present tense verbs. Mislabeling present tense verbs as HABITUAL was the most common error made by the model aside from the main subtask of event identification, making up 22% of errors. Mislabeling verbs in the present participle form as ACTIVITY was another consistent error, occurring at Step 4. The most common gold label for these erroneous HABITUAL and ACTIVITY labels was PERFORMANCE. For example, in Sentence L, *returns* is mislabeled as HABITUAL. The future tense verb *will spend* in the first clause changes the aspect for the successive clauses of L, i.e. the clause *"...before he **returns** home with his wife Sherry"*.

   The prevalence of gold PERFORMANCE labels indicates that Steps 3 and 4 are prematurely assigning an aspect and not letting certain present participles continue throughout the sequence and make it all the way to Step 8. However, AutoAspect also correctly labeled some present tense forms, as seen in Sentences K and N. Experimenting with using tense and aspect annotations[8] from the ClearTAC parser resulted in even more false positives for HABITUAL and ACTIVITY. Reducing the number of false positives for HABITUAL and ACTIVITY necessitates building a semantic parser that can discern between sentences like L with multi-tense contexts and sentences like M with dialogic contexts.

[8]Tenses: present, past, future; Aspects: progressive, perfect, perfect progressive

4. **Container and Durative Adverbials:** The AutoAspect decision-making for container and durative adverbials in Step 6 — as well as the non-resultative paths in Step 7 — currently only checks if specific prepositions like *in* and *for* are found as dependents of the main verb. Thus, the sentence *"They also said this court did not give the lawyers for the defense due procedure."* incorrectly assigns the verb *give* the aspect ENDEAVOR, because the prepositional phrase *for the defense* is incorrectly parsed by the spaCy dependency parser as being a dependent of the verb *give*.

   Future work can incorporate semantic processing of prepositional phrases to help AutoAspect refine its analysis of prepositional dependencies. One such system is SNACS (Schneider et al., 2018), which outputs disambiguated supersenses like TIME and DURATIVE that could match the semantic properties of the container and durative adverbials.

5. **Stativity:** 15 STATIVE verbs such as *expected* and *contemplated* were mislabeled as PERFORMANCE. This implies that the pre-defined list of VerbNet class IDs that correspond to *state* events does not yet fully cover the range of stative verbs, and/or that SemParse isn't able to find all the VerbNet classes. One solution is to pursue development of a stativity annotator, such as SitEnt (Friedrich et al., 2016), which houses training data that annotates verbs as *stative*, *dynamic*, or neither.

   The overall task of classifying SitEnt types achieved an accuracy of 76% using a CRF model that utilizes hand-crafted feature sets. A key difference is that the SitEnt features are accessed simultaneously by the model, while AutoAspect follows the sequential UMR annotation steps. Future research directions could incorporate developing a feature-based model to achieve an accuracy comparable to SitEnt, which also was scored on larger corpora (Brown and MASC) than the 4 gold documents AutoAspect scored.

5 Conclusion

The linguistic blind spots discovered from analyzing the results highlight many areas for future development for the AutoAspect annotator. It is clear that the primarily syntax-driven rules are unable to capture semantic properties like stativity, disambiguation of prepositional phrases, and genre of text. Feeding in input sentence-by-sentence also prevents AutoAspect from properly analyzing multi-sentence events. The tool currently has a recall of 76.17% for identifying events and a 62.57% accuracy for identified events. As more gold UMR data is processed, further linguistic blind spots can be identified for development past the pilot version.

Acknowledgments

We gratefully acknowledge the support of NSF 1764048 RI: Medium: Collaborative Research: Developing a Uniform Meaning Representation for Natural Language Processing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or the U.S. government. We thank James Gung and Ghazaleh Kazeminejad for their valuable assistance in teaching us how to use the SemParse tool. We thank Skatje Myers for assisting us with the ClearTAC tool.

References

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.

William Croft. 2001. *Radical construction grammar*. Oxford University Press, Oxford.

William Croft. 2012. *Verbs: Aspect and causal structure*. Oxford University Press, Oxford.

William Croft. in press. *Morphosyntax: Constructions of the World's Languages*.

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.

Lucia Donatelli, Nathan Schneider, William Croft, and Michael Regan. 2019. Tense and aspect semantics for sentential amr. *Proceedings of the Society for Computation in Linguistics*, 2(1):346–348.

David Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67:547–619.

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, 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.

James Gung. 2020. *Abstraction, Sense Distinctions and Syntax in Neural Semantic Role Labeling*. Ph.D. thesis, University of Colorado at Boulder.

James Gung and Martha Palmer. 2021. Predicate representations and polysemy in verbnet semantic parsing. In *14th International Conference on Computational Semantics (IWCS)*, online.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2021. spaCy: Industrial-strength Natural Language Processing in Python.

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.

Nathan Schneider, Jena D Hwang, Vivek Srikumar, Jakob Prange, Austin Blodgett, Sarah R Moeller, Aviram Stern, Adi Bitan, and Omri Abend. 2018. Comprehensive supersense disambiguation of english prepositions and possessives. *arXiv preprint arXiv:1805.04905*.

Stephanie Strassel and Jennifer Tracey. 2016. Lorelei language packs: Data, tools, and resources for technology development in low resource languages. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3273–3280.

Jens E. L. Van Gysel, Meagan Vigus, Pavlina Kalm, Sook-kyung Lee, Michael Regan, and William Croft. 2019. Cross-linguistic semantic annotation: Reconciling the language-specific and the universal. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 1–14, Florence, Italy.

Jens EL Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Jan Huang, Chu-Ren 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. *Künstliche Intelligenz*, pages 1–18.

Zeno Vendler. 1967. *Linguistics in philosophy*, chapter Verbs and times. Cornell University Press, Ithaca.

A Supplementary Material

A.1 List of VerbNet classes corresponding to the aspect STATE

- want-32.1
- long-32.2
- try-61.1
- intend-61.2
- wish-62
- allow-64.1
- let-64.2
- admit-64.3
- forbid-64.4
- tingle-40.8.2
- pain-40.8.1
- stimulus_subject-30.4
- keep-15.2
- support-15.3
- contain-15.4
- being_dressed-41.3.3
- simple_dressing-1.3.1
- function-105.2.1
- lodge-46
- exist-47.1
- bulge-47.5.3
- meander-47.7
- contiguous_location-47.8
- terminus-47.9
- put_spatial-9.2-1
- cling-22.5
- entity_specific_modes_being-47.2
- light_emission-43.1
- smell_emission-43.3
- sound_emission-43.2
- sound_existence-47.4
- substance_emission-43.4-1
- swarm-47.5.1-1
- animal_sounds-38
- carve-21.2-1
- modes_of_being_with_motion-47.3
- snooze-40.4
- body_internal_states-40.6
- spatial_configuration-47.6
- peer-30.3
- see-30.1