

EN.601.769 Assignment 2: Synthesis Paper

Anonymous ACL-IJCNLP submission

1 A&R Factors

Abend and Rappoport (2017) described a plethora of factors to classify existing Semantic Representations of Text (SRTs). The described factors include: **connection to lexical items** (can SRTs be mapped to the original text?) and **syntax** (is there correspondence between syntactic and semantic representations?), **cross-sentence annotations** (do SRTs support cross-sentence relations and larger discourse contexts?), **logical structure** (for symbolic SRTs, what is the underlying formal language?), **inference capabilities** (how well suited the SRTs are for downstream inference tasks, say recognizing textual entailment?), **universality** (how well does each semantic scheme generalize across languages?) and **annotation effort** (is the annotation fully manual? do we need skilled experts or crowd workers? how long does the annotation take?).

2 Proposed Factors

2.1 Options Considered

To suggest two new factors, I took a practical standpoint, considering what a potential end-user, interested in extracting events from text, needs to be aware of, before locking in their application or model with a particular SRT. The options I considered were: **representations of event predicates** (what is the vocabulary of event predicates?), **initial layer of annotation** (are annotations done over raw texts? or dependency trees? or another existing SRT?), **completeness of annotations** (is the goal to annotate everything in a given corpus? or provided a certain number of “exemplars” per each label, akin to (Baker et al., 1998)?), **mappings to other well-known resources** (can we map from a given SRT to, say, FrameNet? PropBank? VerbNet?) and **corpus domain** (what domain(s) does the gold annotated data come from?).

2.2 Final Choice of Factors

I chose **representations of event predicates** and **corpus domain** as my two factors.

Representations of event predicates specify what counts as an event predicate and how is it represented. Is it a lexical unit with an associated verb sense, as in PropBank, or a separate label that generalizes multiple lexical units, as in VerbNet and FrameNet? I chose this particular factor for a practical reason: for my final project, I am interested in Causal Discovery over arbitrary frequent event sets. The existing CD tools expect reasonably sized vocabularies of events, and minimally the size of the event vocabulary is at least as big as the size of the event predicate vocabulary $|V_{Pred}|$. While for all the representations considered in this report this number is on the order of all unique English verbs ($|V_{Pred}| \sim 10^4$), there exist more parsimonious representations such as FrameNet ($|V_{Pred}| = 1221$) and VerbNet ($|V_{Pred}| \sim 10^2$).

Corpus domain, broadly, specifies the sampling distribution of all the existing gold annotated data¹ for a given SRT. More specifically, it controls the genre and the style of texts, the distribution of tokens (which tokens are considered frequent? rare?) and various document-level statistics (what’s the average length of a passage? a document?). Why and where is the corpus domain important? Indeed, all the schemes considered here claim to be capable of providing general-purpose representations. However, the end-user still needs to be aware if, say, their downstream task is inference in the domain of science fiction literature, and all the gold annotated data for an SRT of choice is newswire texts or Reddit posts.

¹However, for the purposes of this report, I will only consider the corpora mentioned in the respective papers, where each SRT was introduced.

2.3 Classification of Schemes

AMR To represent events (or concepts in the paper’s terminology), AMR (Banarescu et al., 2013) uses **PropBank** (Kingsbury and Palmer, 2002; Palmer et al., 2005) **framesets**. A single concept might be realized into different lexical forms. For instance, the concept “desire-01” might be realized as a verb (“desire”), a noun (also “desire”), or an adjective (“desirous”). The concepts need not necessarily be verbal: some adjectives are mapped into verbal predicates (e.g. “be aware (of X)” → “realize-01”) while for others the adjectives themselves are used as predicate names (e.g. “be-efficient (at X)” → “efficient-01”). The AMR bank is constructed manually over exemplar sentences from domains such as **tutorials, newswire, web data, novels, and broadcast conversations**.

UCCA The equivalent of an event in UCCA (Abend and Rappoport, 2013) is called a *scene*. Each scene (further subcategorized as a Process or a State) is represented as a DAG where nodes are lexical units and edges are categories (e.g. Process, State, Participant, Adverbial). Therefore, event predicates are simply **English words** in UCCA. Predicate arguments are not further sub-typed beyond the Participant category. The annotated corpus is mostly based on **English Wikipedia articles about celebrities**. This domain is chosen for its diversity and accessibility to annotators.

GMB Semantic representations of GMB (Basile et al., 2012) are based on Boxer (Bos, 2008) which represents events as neo-Davidsonian-style predicates, where event predicates are simply **English words** and event arguments are expressed with an inventory of VerbNet (Kipper et al., 2008) thematic roles. GMB is annotated over open-domain text corpora such as **newswire** (from Voice of America), **country descriptions** (from the CIA Factbook), **a collection of texts from Open American National Corpus**, and **Aesop’s fables**.

UDS Predicate extraction in UDS (White et al., 2016) is handled via PredPatt. **PredPatt predicates** are built recursively from a Universal Dependency parse via 1) predicate root identification, where certain tokens such as governors of *nsubj/dobj*, or dependents of *acl/ccomp* are designated as predicate roots, 2) predicate phrase extraction, where the subtree of a root token is expanded until a stop criterion is met. As an example, a PredPatt predicate could be simply a verb

(“finds”) or a verb with a negation and an adverbial modifier (“did not sleep quietly”). The paper had preliminary annotations over exemplar sentences from PropBank and SemLink. A larger UDS1.0 dataset, collected later by White et al. (2019), provides an annotation layer over English Web Treebank (Bies et al., 2012) and covers **weblogs, newsgroups, email, reviews, and QA** genres.

3 QA-SRL

Here, I will discuss QA-SRL (He et al., 2015) through the lens of the A&R factors (Section 1) and the factors proposed in the report (Section 2).

3.1 Existing Factors

QA-SRL deals with *argument annotation*: given a sentence s and a pre-identified verbal predicate v , the annotators must produce a set of wh-questions that contain v and whose answers are phrases in s . Possible predicates are chosen to be verbs from s . Answer spans are effectively subsequences of the words from s and need not be contiguous. As such, they are **connected lexically** (but **not syntactically**) to a sentence s (and its syntactic representation). The annotations are done **on a sentence level** and **no explicit logical structure is assumed**. The **inference capabilities** of QA-SRL are also **limited** since QA-SRL deals with identifying arguments of an existing event, and not the whole event. Both question generation and answer identification experiments are performed in English, but the proposed QA-driven approach could be easily extended to other languages and is thus **universal**. Finally, the annotation took **6-9 minutes per sentence**, depending on the domain of the text. The annotators were **crowd workers**. Screening and training them took less than 2 hours on average.

3.2 Proposed Factors

The corpus domains covered in QA-SRL experiments are **newswire** texts (subsampling from the training data of CoNLL-2009 shared task) and **Wikipedia**. The data comes in the form of 3000 exemplar sentences with nearly 8000 verbs. Since the goal of QA-SRL is to identify arguments, the predicates have been pre-identified as simply **all the English verbs from the exemplar sentences**² using either gold or predicted POS tags, depending on the corpus domain.

²The annotators had the ability to discard the irrelevant verb predicates; around 2-3% of predicates were discarded.

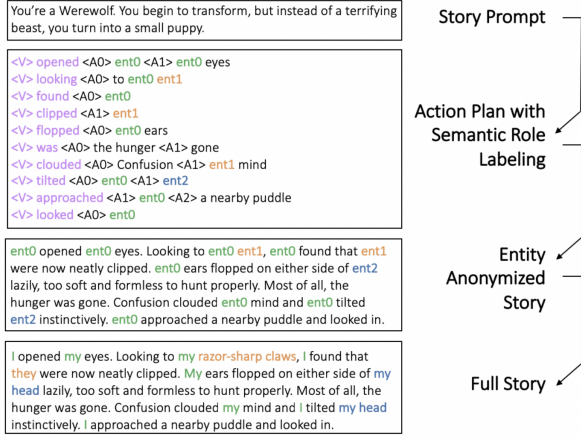


Figure 1: Story generation pipeline of Fan et al. (2019). PropBank framesets and roles are used in Action Plans.

4 Story Generation with Semantic Plans

4.1 Introduction

The task of Story Generation is often decomposed into a planning and a realization stage (Yao et al., 2019; Fan et al., 2019; Rashkin et al., 2020). In the first stage, given a short prompt, a plan (also called an outline, or a skeleton) is generated. In the second stage, a model (typically, a conditional language model) realizes the plan into a coherent natural language story. This decomposition allows for more explainable generation and facilitates interactive writing (Goldfarb-Tarrant et al., 2019).

Existing work on story planning demonstrates a variety of representations. Chen et al. (2019) use external summarization models. Yao et al. (2019) extract most important words of each story sentence based on word frequency. Fan et al. (2019) identify PropBank predicates and their arguments with external pretrained models. They then employ coreference resolution to identify mention chains across the extracted predicates and rename (“anonymize”) those to *ent0*, *ent1*, ... to make story plans more general. An example of their PropBank-based story plan and a realized story is shown on Figure 1.

In this section, I argue for a more semantically rich story plan representation based on **FrameNet** frames as well as **WikiData** entities (qnodes) and relations (pnodes). This proposed representation is similar to Ou et al. (2021) but elaborates more on the representation of arguments and relations.

4.2 Predicate Representation

To represent events succinctly while generalizing over multiple lexical forms, I propose using FrameNet frames as event predicates. As shown by (Ou et al., 2021), such a representation for story plans is able to give enough guidance for the realization model while not constraining the generation too much. It is also more **universal** than the language-specific VerbNet, GMB, and UCCA and can potentially improve the **sampling complexity** of a PropBank-based planning model of Fan et al. (2019), owing to smaller event predicate vocabulary and fewer parameters to learn as a result.

Practically, when mining plans from stories, it might be reasonable to keep the most salient frames only, similar to how Yao et al. (2019) only keeps one keyword per sentence. To that end, PredPatt can be used to first identify spans that represent event predicates, and then frame extraction may be constrained to these spans only. Alternatively, one can identify the most event salient spans (Liu et al., 2018) and similarly constrain frame extraction to event salient spans only.

4.3 Argument Representation

The anonymization process of Fan et al. (2019) is reasonable but discards all the typological information about entities. After anonymization, it is unclear from the plan whether an entity was a human, an animal, a commodity, or a location. I propose storing entity types as part of the story plan. Entity types may be obtained by linking the entities to WikiData, and traversing the “subclass of” relation until a desired entity typing level is reached. Using this approach, e.g. “poodle” (Q38904) can be typed as “dog” (Q144) → “pet” (Q39201) → “domesticated animal” (Q622852) → “animal” (Q729).

4.4 Representing Entity Relations

Suppose we want the entities in a story to be related in a particular way. For example, we may want to say that *ent0* is a person who is employed at the *ent1* university. Or say that *ent0* is a person who is employed at a particular university, say Q193727³. This information might be generated automatically by the planning model, or provided manually by a human. Similar as before, we can employ WikiData and represent such a relation as a typed edge: *ent0* $\xrightarrow{\text{P108}}$ *ent1*, where P108 is a WikiData property representing employment.

³<https://www.wikidata.org/wiki/Q193727>

References

- Omri Abend and Ari Rappoport. 2013. [Universal Conceptual Cognitive Annotation \(UCCA\)](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 228–238, Sofia, Bulgaria. Association for Computational Linguistics.
- Omri Abend and Ari Rappoport. 2017. [The state of the art in semantic representation](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 77–89, Vancouver, Canada. Association for Computational Linguistics.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. [The Berkeley FrameNet project](#). In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada. 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.
- Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. [Developing a large semantically annotated corpus](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3196–3200, Istanbul, Turkey. European Language Resources Association (ELRA).
- Ann Bies, Justin Mott, Colin Warner, and Seth Kulick. 2012. English web treebank. *Linguistic Data Consortium, Philadelphia, PA*.
- Johan Bos. 2008. [Wide-coverage semantic analysis with Boxer](#). In *Semantics in Text Processing. STEP 2008 Conference Proceedings*, pages 277–286. College Publications.
- Gang Chen, Yang Liu, Huanbo Luan, Meng Zhang, Qun Liu, and Maosong Sun. 2019. Learning to predict explainable plots for neural story generation. *arXiv preprint arXiv:1912.02395*.
- Angela Fan, Mike Lewis, and Yann Dauphin. 2019. Strategies for structuring story generation. *arXiv preprint arXiv:1902.01109*.
- Seraphina Goldfarb-Tarrant, Haining Feng, and Nanyun Peng. 2019. Plan, write, and revise: an interactive system for open-domain story generation. *arXiv preprint arXiv:1904.02357*.
- Luheng He, Mike Lewis, and Luke Zettlemoyer. 2015. [Question-answer driven semantic role labeling: Using natural language to annotate natural language](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 643–653, Lisbon, Portugal. Association for Computational Linguistics.
- Paul Kingsbury and Martha Palmer. 2002. [From Tree-Bank to PropBank](#). In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC’02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).
- Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation*, 42(1):21–40.
- Zhengzhong Liu, Chenyan Xiong, Teruko Mitamura, and Eduard Hovy. 2018. Automatic event salience identification. *arXiv preprint arXiv:1809.00647*.
- Jiefu Ou, Nathaniel Weir, Anton Belyy, Felix Yu, and Benjamin Van Durme. 2021. Infillmore: Neural frame lexicalization for narrative text infilling. *arXiv preprint arXiv:2103.04941*.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. [The Proposition Bank: An annotated corpus of semantic roles](#). *Computational Linguistics*, 31(1):71–106.
- Hannah Rashkin, Asli Celikyilmaz, Yejin Choi, and Jianfeng Gao. 2020. Plotmachines: Outline-conditioned generation with dynamic plot state tracking. *arXiv preprint arXiv:2004.14967*.
- Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. [Universal compositional semantics on Universal Dependencies](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723, Austin, Texas. Association for Computational Linguistics.
- Aaron Steven White, Elias Stengel-Eskin, Sidharth Vashishtha, Venkata Govindarajan, Dee Ann Reisinger, Tim Vieira, Keisuke Sakaguchi, Sheng Zhang, Francis Ferraro, Rachel Rudinger, et al. 2019. The universal compositional semantics dataset and decomp toolkit. *arXiv preprint arXiv:1909.13851*.
- Lili Yao, Nanyun Peng, Ralph Weischedel, Kevin Knight, Dongyan Zhao, and Rui Yan. 2019. Plan-and-write: Towards better automatic storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7378–7385.