NounAtlas: Filling the Gap in Nominal Semantic Role Labeling

Roberto Navigli[1]   **Marco Lo Pinto**[1]   **Pasquale Silvestri**[1]
Dennis Rotondi[3]   **Simone Ciciliano**[4]   **Alessandro Scirè**[1,2]

[1]Sapienza NLP Group, Sapienza University of Rome
[2]Babelscape, Italy
[3]University of Stuttgart [4]Free University of Bozen
{navigli, scire}@diag.uniroma1.it, dennis.rotondi@ki.uni-stuttgart.de
{lopinto.1746911, silvestri.1749274}@studenti.uniroma1.it, sciciliano@unibz.it

Abstract

Despite the significant advances made in Semantic Role Labeling (SRL), much work in this field has been carried out with a focus on verbal predicates, with the research on nominal SRL lagging behind. In many contexts, however, nominal predicates are often as informative as verbal ones, thus calling for proper treatment. In this paper we aim to fill this gap and make nominal SRL a first-class citizen. We introduce a novel approach in order to create the first large-scale, high-quality inventory of nominal predicates and organize them into semantically-coherent frames. Although it is automatically created, NounAtlas – our nominal frame inventory – is subsequently fully validated. We then put forward a technique for generating silver training data for nominal SRL and show that a state-of-the-art SRL model can achieve good performance. Interestingly, thanks to our design choices, which enable seamless integration of our predicate inventory with its verbal counterpart, i.e., VerbAtlas, we can mix verbal and nominal data and perform robust SRL on both types of predicate.

1 Introduction

In the ever-evolving field of Natural Language Processing (NLP), the pursuit of Natural Language Understanding (NLU) has garnered increasing attention over the past two decades. Among the open challenges in this domain, Semantic Role Labeling (SRL) is still far from being solved. Pioneered by Gildea and Jurafsky (2002), SRL is the task of identifying the semantic relations between predicates and their arguments (i.e. "Who did What to Whom, When, Where, and How?"). SRL has already proven to be a valuable asset, contributing to diverse tasks, such as question answering (Shen and Lapata, 2007), textual inference (Bastianelli et al., 2013), neural machine translation (Song et al., 2019), visual semantic role labeling (Chen et al., 2021; Sadhu et al., 2021), text summariza-

tion (Trandabăț, 2011) and storytelling (Peng et al., 2022).

One common oversimplification regarding SRL is the assumption that most predicates are verbs. Indeed, research has focused heavily on verbs as the driving force of meaning in sentences, also propelled by the development of verb-centric resources (Palmer et al., 2005; Schuler, 2005; Di Fabio et al., 2019). In many contexts, however, nominal predicates are at least as important as their verbal counterparts. This phenomenon is particularly evident in media content, including newspaper headlines, blog titles, short text messages, etc. Consider for instance the sentence "Astounding discovery in the laboratory yesterday night: Dr. Jones unveils groundbreaking research findings!". In this example, the noun "discovery" serves as the PREDICATE, indicating an action, with its arguments being "Astounding" (ATTRIBUTE of the discovery), "in the laboratory" (LOCATION), "yesterday night" (time), and "Dr. Jones" (AGENT). Such nominal predicates need to be specifically addressed in order for downstream applications to take full advantage of SRL.

While some resources have been provided that include nominal framesets, such as NomBank (Meyers et al., 2004) and FrameNet (Baker et al., 1998), their coverage is limited and they provide either predicate-specific roles (NomBank) or over-specific or incomplete frame scenarios (FrameNet).

Our work aims to fill this gap and equip Semantic Role Labeling with both a comprehensive inventory of nominal predicates (paired with verbal ones) and annotated datasets to perform the task.

In summary our contributions are the following:

- We introduce NounAtlas, the first large-scale inventory of predicates for nominal SRL;

- To facilitate the design of SRL approaches leveraging our resource, we create a large-scale dataset annotated with nominal predi-

cates and roles from NounAtlas;

- We conduct an extensive evaluation of several baselines trained on our dataset. Our experiments pave the way for the development of a state-of-the-art unified system for both nominal and verbal SRL.

- We release our code and resource at https://github.com/SapienzaNLP/nounatlas under a CC license.

2 Related Work

2.1 Nominal Resources

Existing resources for SRL are predominantly centered on verbs. The earliest examples of such resources include PropBank (Palmer et al., 2005) and VerbNet (Schuler, 2005). More recently, Di Fabio et al. (2019) introduced VerbAtlas, a hand-crafted lexical-semantic resource that provides a wide-coverage verbal predicate inventory offering coarse semantically-coherent frames and informative semantic role labels. The frames are created by systematically clustering all verbal synsets in WordNet (Miller, 1995) using FrameNet-style (Baker et al., 1998) scenarios, while adopting VerbNet's human-interpretable and cross-frame role labels.

Based on the frame semantic theory of Fillmore (2006), FrameNet was the first resource to be used for SRL (Gildea and Jurafsky, 2002), and one that includes nominal predicates in its frames. However, FrameNet does not grant full coverage[1], proving challenging to scale for out-of-domain SRL (Hartmann et al., 2017). Furthermore, although it is explicit and thus human-readable, FrameNet's roles are frame-specific, limiting their generalizability across frames with similar roles.

Concerning nominal predicate resources, the state of the art is currently represented by PropBank (Palmer et al., 2005) and NomBank (Meyers et al., 2004). NomBank is a project that aims to annotate a subset of noun predicates that take arguments. It utilizes a framework similar to that of the first version of PropBank, but for annotating noun predicates rather than verbal ones. Differences arise from variations in noun and verb argument structures, as well as disparities in how nouns and verbs are handled within the Penn Treebank (Marcus et al., 1993). These differences in predicate

treatment pose challenges in developing an SRL system for both verbal and nominal predicates due to the overhead of treating them in distinct ways, as seen in CoNLL-2009 (Hajič et al., 2009).

With NomBank not having been augmented or updated for a long time, nominal predicates began to be included in PropBank, starting from version 2 (Babko-Malaya et al., 2004), where they are referred to as event variables. However, PropBank is verb-oriented: all nominal units have been integrated as aliases of previously-annotated verbal rolesets, which means they are not as thoroughly annotated as their verbal counterparts. Both NomBank and PropBank are primarily limited by their use of enumerated arguments (Arg0, Arg1, etc.), with the absence of an explicit classification of the semantics of predicate roles, making the annotation hard to interpret and overloading the meaning of an argument label. For instance, the first arguments of "drinking" and "feeling" are both labeled as Arg0, despite representing distinct semantic roles (AGENT and EXPERIENCER, respectively). Also, as of today, NomBank and PropBank suffer from limited coverage of nominal predicates, including 5577 and 4295 nouns, respectively. Moreover, NomBank is built upon PropBank, with 1308 frames and 861 rolesets being shared between the two resources, and a total of 8815 distinct rolesets.

2.2 Nominal Semantic Role Labeling

As Orlando et al. (2023) highlighted in their experiments, nominal Semantic Role Labeling is still far from being solved. Here, a clear indication to focus on enriching nominal resources is the fact that the best SRL neural models (Shi and Lin, 2019; Conia et al., 2021a; Conia and Navigli, 2022; Conia et al., 2022), trained exclusively on verbal predicates, struggle to generalize to unseen nominal ones. Research on non-verbal predicates remains significantly underdeveloped, with existing efforts centered mainly around transferring knowledge from verbal predicates to their nominal counterparts in an unsupervised manner (Klein et al., 2020; Zhao and Titov, 2020), rather than creating resources that enable SRL to generalize easily between verbal and nominal predicates. The limited attention given to non-verbal predicates may be attributed to the prevailing design of current SRL datasets, which utilizes predominantly verbal predicates (Daza and Frank, 2020; Tripodi et al., 2021; Jindal et al., 2022) annotated based on the aforementioned predicate inventories.

---

[1] Only a little more than half of its frames are deemed complete, including around 5000 nominal units, see https://framenet.icsi.berkeley.edu/current_status

With our work we aim to fill the gap in nominal SRL. First, we focus on constructing NounAtlas, an inventory of over 10 000 nominal predicates, curated by means of leveraging and inheriting knowledge and semantic frames from VerbAtlas. This addresses the above-mentioned coverage issues and provides integration between nominal and verbal predicates, which can belong to the same frames. Second, we employ the resulting inventory to produce, for the first time, a large dataset for wide-coverage nominal SRL.

3 NounAtlas

Before delving into the process of constructing NounAtlas, our large-scale nominal predicate inventory, we highlight a few features in VerbAtlas that enable this process.

VerbAtlas As mentioned in Section 2, VerbAtlas uses verbal synsets from WordNet as verbal predicate proxies, organizing them into semantically-coherent frames. This includes more than 400 synset clusters, each assigned a human-readable label, such as EAT-BITE and SPEAK.

A key advantage of VerbAtlas over its alternatives is its common prototypical argument structure and cross-frame set of semantic roles à la VerbNet: for instance, the frame SPEAK features the roles of AGENT and TOPIC. Importantly, being univocally associated with exactly one frame, each synset is associated with the corresponding frame semantics and predicate argument structure, which can be exploited for SRL. For instance, in Figure 1 the verbal synsets EMBARRASS.V.01 and PRIDE.V.01 are included in the CAUSE-MENTAL-STATE frame.

3.1 Creating a Large-Scale Nominal Predicate Inventory

Our first objective is to produce NounAtlas, a large-scale inventory of nominal predicates. To achieve this goal, we leverage the VerbAtlas frames to group WordNet nominal synsets that express analogous semantics in nominal form. For instance, we utilize the EAT-BITE verbal frame, which contains synsets such as EAT.V.01, BITE.V.01, and DEVOUR.V.04, to cluster nominal synsets like EATING.N.01, BITE.N.08 and CHOMPING.N.01. We now detail the procedure for creating our nominal predicate inventory.

3.1.1 Event-characterizing synset selection

First, we define a nominal predicate as a concept which conveys an event, action or situation, allowing a clear identification of roles such as AGENT, PATIENT, THEME, etc. In our work we aim to identify nominal synsets with such properties (which we call event-characterizing synsets) as proxies for nominal predicates. As an example, we include WORK.N.01 (activity directed toward making or doing something) as an event-characterizing synset, whereas we exclude non-event synsets, such as WORKER.N.01 (a person who works at a specific occupation) or TABLE.N.02 (a piece of furniture). To do so, we sort the full list of nominal synsets in WordNet 3.0 by their number of hyponyms (i.e., descendants in the nominal taxonomy) and select the top synsets s_j in the list which are deemed to characterize events. This results in the identification of nine general event-characterizing synsets which are then reduced to four due to the removal of those subsumed by other synsets in the set: EVENT.N.01, EVENT.N.02, PROCESS.N.02 and PROCESS.N.06. We then select as our final set of event-characterizing synsets all the 10 086 nominal synsets that are descendants, i.e., direct or indirect hyponyms, of our set of four general event-characterizing synsets, including the latter.

3.1.2 WordNet-based synset-to-frame mapping

With the goal of clustering nominal synsets into frames, we bootstrap the synset-to-frame mapping process by exploiting the WordNet derivationally-related form (DRF) relations, which establish connections between senses of nominal synsets and their corresponding verbal counterparts. For instance, the nominal synset EATING.N.01 (the act of consuming food) is mapped to the verbal synset EAT.V.01 (take in solid food) through a DRF relation. Since the latter synset is contained in the EAT-BITE frame in VerbAtlas, we identify EATING.N.01 as a potential candidate for this frame.

However, multiple cases need to be considered at this stage for a given nominal synset s:

- s is linked to a single verbal synset v. This allows a straightforward inclusion in the corresponding VerbAtlas frame (**Unambiguous link**), i.e. $frame(s) := frame(v)$, where $frame$ returns the VerbAtlas frame of the input synset.
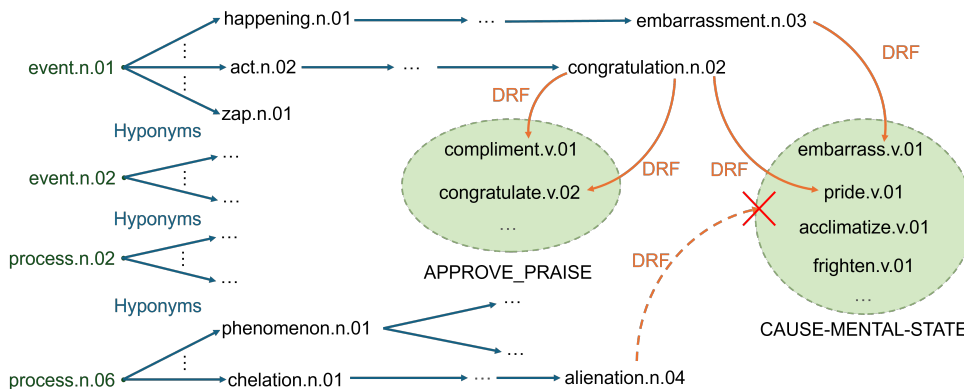
Figure 1: Visual representation of the different types of link between nominal and verbal synsets (and the corresponding VerbAtlas frames). Orange arrows represent the derivationally related forms (DRFs) connections in WordNet, whereas blue arrows represent the WordNet hyponymy relations. Best seen in colors.

- There are multiple verbal synsets V linked to s. If all the verbal synsets in V belong to the same frame, the frame mapping is again straightforward, i.e. $frame(s) := frame(v)$ for any $v \in V$. Otherwise, we manually map the nominal synset s to the most suitable frame among those containing verbal synsets linked to s (**Manually-curated links**).

- s is not associated with any verbal synset through DRFs (**Non-existing links**).

We provide a visual representation of each type of link in Figure 1. In this example, since EMBARRASSMENT.N.03 is connected to EMBARRASS.V.01 through an Unambiguous link, we can include it within the CAUSE-MENTAL-STATE frame. Conversely, CONGRATULATION.N.02 requires a manual annotation, since its DRF links lead to synsets belonging to multiple frames (i.e. APPROVE_PRAISE and CAUSE-MENTAL-STATE). Finally, ALIENATION.N.04 is an example of a Non-existing link, since it is not connected via DRF to any verbal synset from VerbAtlas.

Results and Statistics We report the statistics concerning the WordNet-based mapping of nominal synsets to VerbAtlas frames in Table 1. In Figure 2 we show the expectedly skewed distribution of verbal and nominal synsets within each frame, with some frames comprising over 100 synsets (e.g. SPEAK and HIT), and others being limited up to 3 (e.g. RESERVE and FORGET). Interestingly, the figure shows that frames containing more verbal synsets tend to also include a higher number of nominal synsets.
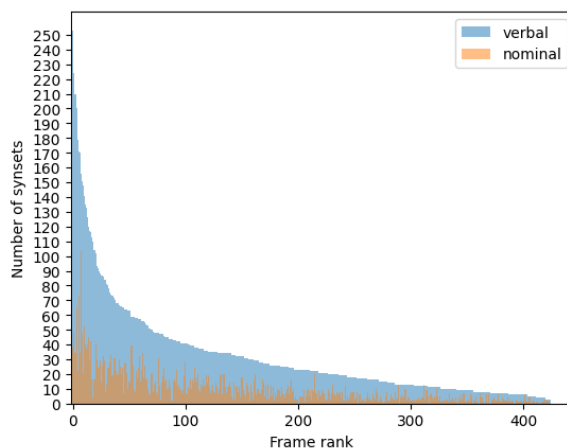
To evaluate the quality of the Unambiguous

Figure 2: Distribution of verbal (represented in blue) and nominal synsets (orange) across frames. On the x-axis, frames are ordered by their number of verbal synsets. Best seen in colors.

Mapping Type	Count
Unambiguous links	3053
Manually-curated links	1129
Non-existing links	5904

Table 1: Statistics about WordNet-based synset-to-frame mapping for the 10 086 nominal synsets.

links, we manually classified a random subset comprising 100 instances without using the DRF information and we found that 82% of the manual classifications were in agreement with the automatic ones, with another 17% upon inspection found to be equally valid. These results demonstrate that the WordNet DRF relations can be leveraged effectively to classify a substantial portion, i.e. 31%, of the overall event-characterizing nominal synsets.

16248

We then moved on to the Manually-curated links. To estimate the complexity of this task, we computed the inter-annotator agreement on a randomly-sampled subset of 100 annotations, with a resulting Cohen's κ of 0.57 (moderate agreement).

The results on both types of link can be attributed to the frequent scenario where the proposed VerbAtlas frames can provide overlapping semantics, leading to multiple valid options for a given nominal synset. For example, the nominal synset EX-AMINATION.N.01 was mapped to the ANALYZE frame by one annotator and to VERIFY by another.

3.1.3 Ranking frames for unlinked synsets

To address the challenge of Non-existing links, and thereby increase the coverage of our mapping, we design an automatic approach aimed at mapping nominal synsets to the best suited verbal frame.

First, we gather all the nominal synsets involved in Unambiguous and Manually-curated links together with all the verbal synsets from VerbAtlas.

Then, we create a dataset organized into pairs of (nominal synset, verbal synset) definitions as follows: given a nominal synset s and its mapped frame f, we pair the definition of s with the definition of the verbal synset $v \in f$ which is linked to s through a DRF relation. We label this pair as positive. Then, we augment our dataset of positive pairs by additionally associating the definition of s with the ones of K randomly-selected verbal synsets also belonging to f. Finally, we obtain $K + 1$ negative pairs by repeating the same process, but this time selecting as many random frames f_i, such that $f_i \neq f \ \forall i \in \{1, \ldots, K + 1\}$, to sample the verbal synsets from. Each definition is expanded with the corresponding synset's lemmas, which are prepended in a comma-separated format. For instance, we generate a positive pair by associating the nominal synset APPOINTMENT.N.01 ("the act of putting a person into a non-elective position"), with the verbal synset DELEGATE.V.02 ("give an assignment to a person to a post, or assign a task to a person") both linked to the frame ASSIGN-SMT-TO-SMN. We then obtain a negative pair by associating the same nominal synset with a verbal synset from a different frame e.g. with WASH.V.09 ("remove by the application of water or other liquid and soap or some other cleaning agent"), belonging to the frame WASH_CLEAN.

Ranking approach. We use the resulting dataset in a binary classification task, with a positive label

indicating that the two synsets in the pair belong to the same frame and a negative label indicating otherwise. We train a Cross-Encoder model (Reimers and Gurevych, 2019) on our dataset. Then, at inference time, given a nominal synset s, we link it to a VerbAtlas frame through the following strategy:

1. From each VerbAtlas frame, we randomly select a maximum number of available verbal synset definitions, up to 10, and pair them with the definition of s;

2. We utilize our model to score each pair;

3. We derive the final score for a frame as the average score of its 10 pairs;

4. We rank frames according to their scores, arranged from highest to lowest. A higher score indicates a greater likelihood of the nominal synset being associated with the respective frame.

Using this approach, for each nominal synset with Non-existing links we are able to construct a similarity ranking of all the VerbAtlas frames.

Experimental setup. We used mpnet-v2 as our Encoder model,[2] and we finetuned it on our dataset. We added a linear binary classification head on top. Binary cross entropy was employed as loss, while we leveraged Adam as optimizer, with an initial learning rate of 3×10^{-5}. All the experiments in our work were conducted on a computing system equipped with an NVIDIA RTX 3090. We implemented the Cross-Encoder in PyTorch and PyTorch Lightning, using the Transformers library (Wolf et al., 2020). We selected the hyperparameter values according to the best F1 score on the validation split of our dataset.

Training and evaluation datasets. With the aim of assessing the quality of our ranking approach, we created a gold standard evaluation dataset composed of pairs of definitions of (nominal, verbal) synsets. To create the training and validation splits, we selected nominal synsets involved in Unambiguous links. Specifically, for each frame, we partitioned its synsets into 80% for training and 20% for validation. To build the test set and ensure its reliability, we used the definitions of all the nominal synsets with a Manually-curated link. We

---

[2]huggingface.co/sentence-transformers/all-mpnet-base-v2. Number of parameters: 109M.

paired each nominal synset definition in the train, validation and test sets with a maximum of $2 \cdot K$ definitions of verbal synsets, generating an equal number of K positive and negative pairs. Recall from the first part of this section that, given a nominal synset s, half of the definition pairs are positive pairs built using definitions of verbal synsets sampled from the same frame that s belongs to, i.e. $frame(s)$, while the other half are built using definitions of verbal synsets sampled from random frames, each different from $frame(s)$.

We set K to 30, 3, and 3 for our training, validation, and test splits, respectively. As a result, we obtained $132\,484$ pairs for the training set, 3648 for the validation set, and 6546 for the test set.

Results and discussion. We evaluate the performance of our system for the binary classification task with the accuracy and F1 metrics, and for the ranking task using the Top-k accuracy scores. We report the results of our evaluation in Table 2.

As can be seen, our model achieves high accuracy and F1 on binary classification. Moreover, the Top-k accuracy achieved highlights the ability of our system to provide the correct frame among the top-5 ranked. This outcome paves the way for an effective manual annotation process, aimed at selecting one frame among the top-5 ranked by our system, given a nominal synset (cf. next section).

3.1.4 Manual mapping of unlinked synsets

We can now proceed to mapping unlinked synsets to frames: given a nominal synset s, we tasked an expert annotator to select the most suitable frame from among the top-5 provided in the similarity ranking of s. In special cases where the annotator was not able to choose a suitable candidate from the provided frames, they had the option to consult VerbAtlas, and select a more appropriate frame. The annotator was specifically recruited for this task. We paid them in accordance with the standard salary of their geographical location. They took an average of 3 minutes to annotate one item. The result of our entire semi-automatic mapping process was a large-scale nominal predicate inventory for SRL, which we have called NounAtlas. We note that, by construction, the VerbAtlas and NounAtlas frames share the same labels, making it possible to bring together nominal and verbal synsets into a unified nominal and verbal resource.

Results. The annotator successfully linked 5876 out of the 5904 synsets with Non-existing links
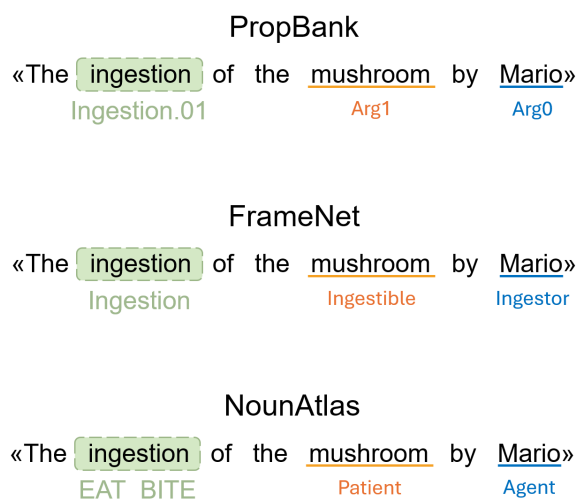

Figure 3: Example of SRL tagging with PropBank and FrameNet, compared to NounAtlas.

(99.6%). They selected 5207 frames (88.2% of the total) among the top-5 ranked by our system, with 2912 (49.3%) being the top-1. The remaining synsets were classified by directly accessing the full VerbAtlas frame inventory.

3.1.5 Qualitative Comparison

To highlight the inherent advantages of using NounAtlas, consider the annotations reported in Figure 3 of a sample sentence annotated with different lexical-semantic nominal inventories. As mentioned in Section 2, PropBank's underspecified role semantics can result in models that struggle to generalize to out-of-domain predicates. FrameNet's roles, instead, are explicit, but frame-specific, resulting in a very large set of different role labels: in the example above, the roles "Ingestor" and "Ingestible" are specific only for the frame "Ingestion". These problems are solved when using NounAtlas, which provides just a few, explicit and cross-frame semantic roles over a very large set of nominal predicates, enabling a more extensive coverage.

4 Creating a Nominal SRL Dataset

As a result of the work outlined in Section 3.1, we obtained an inventory of nominal predicates organized into frames. However, we still have to build a large dataset for nominal Semantic Role Labeling, i.e. annotated with our nominal predicates. In what follows, we achieve this goal by (i) collecting WordNet verbal sense occurrences (Section 4.1), (ii) transforming each verbal predicate annotation into a nominal one, adjusting the sen-

Binary classification		Ranking				
Accuracy	F1-Score	Acc@1	Acc@2	Acc@3	Acc@4	Acc@5
91.9	91.4	44.0	61.3	70.6	77.7	82.1

Table 2: Performance of the Cross-Encoder in the binary classification task and the top-k accuracies ($k = 1, \ldots, 5$).

tence accordingly, and obtaining the corresponding frame annotation (Section 4.2), (iii) propagating the argument role annotations (Section 4.3).

4.1 Initial Sentence Corpus

The initial step involves collecting sentences featuring sense-annotated verbal predicates. For this purpose, we use SemCor (Miller et al., 1993) – the largest corpus manually annotated with WordNet senses – to extract those sentences which contain at least one verb annotated with a synset v such that i) v is directly linked via DRF to a nominal synset s and, ii) v and s are linked through an Unambiguous link.

4.2 Predicate Nominalization

The next step consists in transforming each verbal predicate occurrence in our collected sentences into nominal form. To do so, first, we use a Large Language Model (LLM) prompted to nominalize a given verbal predicate annotated with synset v using any of the synonyms in the corresponding nominal synset $s = \{n_1, \ldots, n_m\}$, i.e. its deverbal forms. We use the following prompt:

Change the sentence by nominalizing the verb "w_i". Use exactly one of these deverbal nouns: "n_1, ..., n_m". Indicate the chosen deverbal noun with **: "$w_1 \ w_2 \ldots$ **w_i** $\ldots w_n$"

where $w_1 \ldots w_n$ is the sentence, and w_i is an inflected form of a lemma in synset v. The prompt tasks the LLM with indicating the chosen deverbal noun within the sentence (we report some examples of instantiated prompts in Appendix B). To enhance the model's understanding of the task, we provide it with additional inputs as system-level instructions, and a list of manually-crafted nominalization examples for few-shot prompting, to enable in-context learning and coherent output formatting (see Appendix B). The few-shot examples are manually crafted from selected representative verbal samples of SemCor, in order to provide the LLM with a variety of possible nominalization forms and avoid a repetitive structure of the nominalized sentences. Note that, in the case of N verbal predicates occur-

ring in the same sentence, we create N different prompts, each revolving on a different predicate to nominalize.

Second, each LLM-generated sentence is then automatically validated through the following heuristics to ensure reliability, automatically checking that: (i) a deverbal noun is clearly indicated in the output using the formatting requested in the prompt and outlined in the few-shot examples; (ii) the sentence and the chosen deverbal noun are POS-tagged to make sure that the chosen deverbal noun is indeed a noun; (iii) the lemmatized form n_j of the chosen deverbal noun is one of the candidates provided, i.e. $n_j \in s$. If all checks are successfully passed, the sentence is considered valid and we tag the nominal predicate occurrence in the generated sentence with $frame(s)$. We then proceed to the role propagation step.

During the predicate nominalization, different LLMs and prompt combinations were tested, including ChatGPT models (OpenAI, 2024) and the Gemini-Pro model (Google, 2024) (see Appendix A for a comprehensive list and individual results). In order to finalize the prompt, we evaluated the results quantitatively using the aforementioned validation pipeline, but also qualitatively by manual inspection, over a representative subset of samples, using criteria such as grammatical correctness, adherence to the original sentence, and correct usage of the nominal predicate. These tests highlighted the need to craft a prompt that guides the LLM toward the desired result, and this was the reason for our decision to include the list of admissible deverbal nouns, few-shot examples, and clear, standardized output formatting.

4.3 Verbal-to-Nominal Role Propagation

The last step is to label the roles in each sentence. Recall that, by construction, $frame(s) = frame(v)$, i.e. the frame associated with the nominal predicate n_j is the same as that of its verbal counterpart (cf. Section 3.1.2).

To label the roles of the target nominal predicate, we carry out the following steps using both the original sentence σ_v used as input for the LLM,

Training data	Overall			Predicates			Roles		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
V	62.8	54.3	58.2	13.5	57.3	21.8	66.6	56.7	61.3
N	71.6	72.1	71.9	**98.3**	**97.9**	**98.1**	75.5	76.0	75.7
N+V	**72.4**	**72.4**	**72.4**	77.7	94.0	85.0	**76.4**	**76.4**	**76.4**

Table 3: Precision, Recall and F1 results of the SRL baselines on our nominal dataset. We report the performance of our baselines trained on our nominal dataset (N), on OntoNotes (V), and on the combination of the two (V+N).

and the corresponding output sentence σ_n in which the predicate is nominalized.

First, σ_v is processed using a state-of-the-art SRL system, i.e. InVeRo-XL (Conia et al., 2021b), obtaining automatically-annotated span-based roles for the target verbal predicate $frame(v)$.

Second, we use a bi-encoder architecture[3] to encode σ_v and σ_n, and compute the following steps for each role r in σ_v:

1. We compute the span embedding E_r of role r in σ_v by mean pooling and L2 normalization of the embeddings of all the tokens in the span.

2. We slide a window of the same size as the role span over σ_n by one word until the end of the sentence. For each resulting span sp, we compute the span embedding E_{sp} as above.

3. We select the span sp in σ_n that maximizes the cosine similarity between E_r and E_{sp}.

4. We propagate the role annotation to the selected span in σ_n.

The result of this process is a large automatically annotated SRL dataset, with nominal predicates identified and tagged using VerbAtlas/NounAtlas frames, and predicate arguments tagged with the corresponding roles.

4.4 Results

Statistics. Among its sentences, SemCor contains 36 295 verbal predicate occurrences, each of which is used to prompt the LLM to generate a corresponding nominal version. Of these, 29 221 passed the validation step, resulting in the same number of valid occurrences of nominal predicates. These latter are then passed into the role propagation step, obtaining 28 055 sentences annotated with nominal frames and their roles. The drop in

---

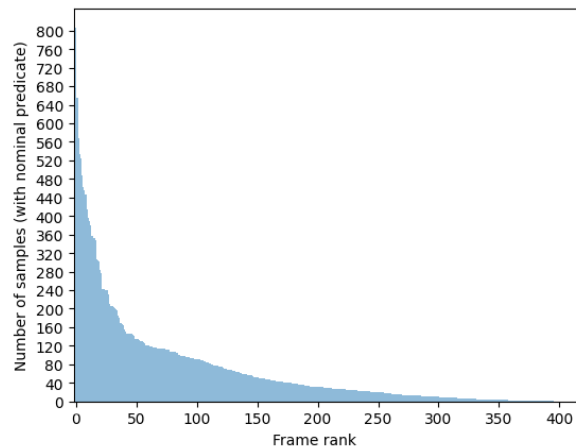[3] https://huggingface.co/sentence-transformers/all-mpnet-base-v2

Figure 4: Distribution of sentences with nominal predicate across frames (the y-axis indicates the number of sentences whose predicate belongs to the same frame).

the number of sentences in the role propagation step occurs because we consider as good sentences only the ones that have all the roles from the verbal sentence also present in the nominal counterpart. If even a single role is missing, then the sentence is discarded. The distribution of the latter sentences across frames is reported in Figure 4. As expected, a few frames have a considerable number of sentences, with these numbers decreasing exponentially. However, a reasonable number of sentences is also maintained in the long tail (e.g., 32 sentences at the median rank).

Dataset quality assessment. We set aside 500 sentences and tasked a human expert to annotate those sentences according to the following guidelines: for each sentence the expert was asked to (i) correct the sentence if it was found to be ungrammatical or nonsensical, (ii) double-check the frame assigned to the nominal predicate and correct it if needed, and (iii) correct each argument span and (iv) the corresponding role label. The annotator was specifically recruited for this task. We paid them in accordance with the standard salary of their geographical location.

	Overall			Predicates			Roles		
Training data	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
V	89.3	89.8	89.6	93.2	95.7	94.4	88.6	89.0	88.8
N+V	89.3	89.6	89.5	92.8	94.4	93.6	88.5	88.9	88.7

Table 4: Precision, Recall and F1 results of the SRL baselines on OntoNotes. We report the performance of our baselines trained on OntoNotes (V), and on the combination of our nominal dataset with OntoNotes (V+N).

An example of a manual annotation carried out by the expert is:

Original sentence: [The girl]$_{Agent}$ crawled out into the renewing warmth of the sunshine [hugging]$_{FIT}$ [her shoulders]$_{Theme}$.

Generated sentence with nominal predicate: [The girl]$_{Agent}$ crawled out into the renewing warmth of the sunshine, performing an [embrace]$_{FIT}$ of [her shoulders]$_{Theme}$.

Corrected sentence: [The girl]$_{Agent}$ crawled out into the renewing warmth of the sunshine, performing an [embrace]$_{FIT}$ [of her shoulders]$_{Theme}$.

In this example, the automatic pipeline made a mistake in capturing the span of the THEME, neglecting to include the preposition "of".

To determine the quality of the automatically annotated nominal SRL dataset, we computed the percentage of i) grammatical sentences (92.11%), ii) unchanged frame assignments (95.56%), iii) unchanged role spans (77.11%), iv) unchanged role labels (95.27%). These results show the robustness of our approach.

4.5 Baselines for Nominal SRL

We used our automatically annotated dataset to train and test a state-of-the-art SRL model in order to provide a baseline for the nominal task. We removed from our dataset the 500 sentences annotated for dataset quality assessment, and used them as our test set. We then split our silver-quality dataset into 90% training and 10% validation.

We selected the state-of-the-art SRL model introduced by Conia and Navigli (2020) as the backbone model for our baselines, closely following the training procedure outlined by the authors. We fine-tuned this model on three different datasets: (a) OntoNotes 5.0 (Weischedel et al., 2010),[4] which serves as the standard dataset for verbal SRL, (b)

our silver dataset for nominal SRL, (c) the concatenation of the two datasets. Because OntoNotes sentences are annotated with PropBank predicates, we use the PropBank-to-VerbAtlas mappings made available by Conia and Navigli (2020)[5] to convert the annotations.

Results. In Table 3, we report the performance of the aforementioned baselines on our manually-curated test set. We note that a state-of-the-art SRL approach, trained on verbal data (V), only moderately generalizes to the nominal setting, thereby confirming the findings of Orlando et al. (2023). Conversely, training the same model on our nominal SRL dataset (N) leads to a significant improvement in performance (21% on average). Interestingly, the baseline trained on the union of verbal and nominal data (V+N) achieves state-of-the-art performance on both OntoNotes (Table 4) and our gold nominal test set (Table 3).

This outcome paves the way for the development of a competitive unified system performing both nominal and verbal SRL.

5 Conclusions

With our work, we fill the gap in nominal SRL by introducing NounAtlas, a large-scale predicate inventory for this task. NounAtlas is constructed by linking nominal predicates from WordNet to VerbAtlas frames through a semi-automatic approach. We additionally create a large dataset for nominal SRL with the help of LLMs, annotated through label propogation with nominal predicates and semantic roles from NounAtlas. Finally, we benchmark multiple baselines consisting of state-of-the-art SRL approaches, trained and evaluated on our dataset. Our experiments demonstrate that a unified SRL approach, trained on both nominal and verbal data, achieves state-of-the-art performance in both settings. We release our code and resource at https://github.com/SapienzaNLP/nounatlas.

---

[4]https://huggingface.co/datasets/conll2012_ontonotesv5

[5]https://github.com/SapienzaNLP/multi-srl

Limitations

While this study provides a valuable starting point for nominal SRL, additional steps could be taken to further enhance its impact. First, our approach utilizes WordNet synsets as predicates, which makes multilingual support feasible through BabelNet (Navigli and Ponzetto, 2010). Indeed each WordNet synset corresponds to a BabelNet synset, which provides translations into tens of other languages. However, lexical coverage varies depending on language (see https://babelnet.org/statistics). Generating a multilingual nominal SRL dataset is more challenging, as one should either use available multilingual corpora annotated with BabelNet, with varying coverage (Pasini et al., 2021), or translate SemCor into other languages and propagate annotations. We leave this to future work.

Additionally, our study focused on nominal arguments, neglecting the exploration of adjectival and adverbial SRL. Furthermore, although the inventory underwent a final step of manual annotation, the training set remains automatically generated. A manual annotation of the entire dataset is expected to bring improved performance on the task. Addressing these limitations in future studies would further advance SRL into a full-fledged NLU task, offering deeper insights into the structured analysis of text.

Acknowledgements

References

Olga Babko-Malaya, Martha Palmer, Nianwen Xue, Aravind Joshi, and Seth Kulick. 2004. Proposition Bank II: Delving deeper. In *Proceedings of the Workshop Frontiers in Corpus Annotation at HLT-NAACL 2004*, pages 17–23, Boston, Massachusetts, USA. Association for Computational Linguistics.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98/COLING '98, page 86–90, USA. Association for Computational Linguistics.

Emanuele Bastianelli, Giuseppe Castellucci, Danilo Croce, and Roberto Basili. 2013. Textual inference and meaning representation in human robot interaction. In *Proceedings of the Joint Symposium on Semantic Processing. Textual Inference and Structures in Corpora*, pages 65–69, Trento, Italy.

Long Chen, Zhihong Jiang, Jun Xiao, and Wei Liu. 2021. Human-like controllable image captioning with verb-specific semantic roles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16846–16856.

Simone Conia, Andrea Bacciu, and Roberto Navigli. 2021a. Unifying cross-lingual semantic role labeling with heterogeneous linguistic resources. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 338–351, Online. Association for Computational Linguistics.

Simone Conia, Edoardo Barba, Alessandro Scirè, and Roberto Navigli. 2022. Semantic role labeling meets definition modeling: Using natural language to describe predicate-argument structures. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 4253–4270, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Simone Conia and Roberto Navigli. 2020. Bridging the gap in multilingual Semantic Role Labeling: A language-agnostic approach. In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020*, pages 1396–1410, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Simone Conia and Roberto Navigli. 2022. Probing for predicate argument structures in pretrained language models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4622–4632, Dublin, Ireland. Association for Computational Linguistics.

Simone Conia, Riccardo Orlando, Fabrizio Brignone, Francesco Cecconi, and Roberto Navigli. 2021b. InVeRo-XL: Making cross-lingual Semantic Role Labeling accessible with intelligible verbs and roles. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 319–328, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Angel Daza and Anette Frank. 2020. X-SRL: A parallel cross-lingual semantic role labeling dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3904–3914, Online. Association for Computational Linguistics.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. 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 627–637, Hong Kong, China. Association for Computational Linguistics.

Charles Fillmore. 2006. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280:20 – 32.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3):245–288.

Google. 2024. Gemini: A family of highly capable multimodal models.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 1–18, Boulder, Colorado. Association for Computational Linguistics.

Silvana Hartmann, Ilia Kuznetsov, Teresa Martin, and Iryna Gurevych. 2017. Out-of-domain FrameNet semantic role labeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 471–482, Valencia, Spain. Association for Computational Linguistics.

Ishan Jindal, Alexandre Rademaker, Michał Ulewicz, Ha Linh, Huyen Nguyen, Khoi-Nguyen Tran, Huaiyu Zhu, and Yunyao Li. 2022. Universal Proposition Bank 2.0. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1700–1711, Marseille, France. European Language Resources Association.

Ayal Klein, Jonathan Mamou, Valentina Pyatkin, Daniela Stepanov, Hangfeng He, Dan Roth, Luke Zettlemoyer, and Ido Dagan. 2020. QANom: Question-answer driven SRL for nominalizations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3069–3083, Barcelona, Spain (Online). International Committee on Computational Linguistics.

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.

Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, and Ralph Grishman. 2004. The NomBank project: An interim report. In *Proceedings of the Workshop Frontiers in Corpus Annotation at HLT-NAACL 2004*, pages 24–31, Boston, Massachusetts, USA. Association for Computational Linguistics.

George A. Miller. 1995. Wordnet: a lexical database for english. *Commun. ACM*, 38(11):39–41.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross T. Bunker. 1993. A semantic concordance. In *Human Language Technology: Proceedings of a Workshop Held at Plainsboro, New Jersey, March 21-24, 1993*.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden. Association for Computational Linguistics.

OpenAI. 2024. Gpt-4 technical report.

Riccardo Orlando, Simone Conia, and Roberto Navigli. 2023. Exploring non-verbal predicates in semantic role labeling: Challenges and opportunities. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12378–12388, Toronto, Canada. Association for Computational Linguistics.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Tommaso Pasini, Alessandro Raganato, and Roberto Navigli. 2021. Xl-wsd: An extra-large and cross-lingual evaluation framework for word sense disambiguation. In *AAAI Conference on Artificial Intelligence*.

Xiangyu Peng, Kaige Xie, Amal Alabdulkarim, Harshith Kayam, Samihan Dani, and Mark O. Riedl. 2022. Guiding neural story generation with reader models.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Arka Sadhu, Tanmay Gupta, Mark Yatskar, Ram Nevatia, and Aniruddha Kembhavi. 2021. Visual semantic role labeling for video understanding. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5585–5596.

K.K. Schuler. 2005. *Verbnet: A Broad-coverage, Comprehensive Verb Lexicon*. University of Pennsylvania.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 12–21, Prague, Czech Republic. Association for Computational Linguistics.

Peng Shi and Jimmy Lin. 2019. Simple bert models for relation extraction and semantic role labeling.

Linfeng Song, Daniel Gildea, Yue Zhang, Zhiguo Wang, and Jinsong Su. 2019. Semantic Neural Machine Translation Using AMR. *Transactions of the Association for Computational Linguistics*, 7:19–31.

Diana Trandabăț. 2011. Using semantic roles to improve summaries. In *Proceedings of the 13th European Workshop on Natural Language Generation*, pages 164–169, Nancy, France. Association for Computational Linguistics.

Rocco Tripodi, Simone Conia, and Roberto Navigli. 2021. UniteD-SRL: A unified dataset for span- and dependency-based multilingual and cross-lingual Semantic Role Labeling. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2293–2305, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ralph Weischedel, Mitch Marcus, Martha Palmer, Eduard Hovy, Robert Belvin, Sameer Pradhan, and Lance Ramshaw. 2010. Ontonotes: A large training corpus for enhanced processing. In *Handbook of Natural Language Processing and Machine Translation*. Springer, New York, NY.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Huggingface's transformers: State-of-the-art natural language processing.

Yanpeng Zhao and Ivan Titov. 2020. Unsupervised transfer of semantic role models from verbal to nominal domain.

A   LLM Tests for Predicate Nominalization

We tested several LLM and prompt combinations for the task of nominalizing a verbal predicate. The tested LLMs are: OpenAI's ChatGPT models (gpt-3.5-turbo-1106, gpt-3.5-turbo-instruct, text-davinci-003) and Google's Gemini-Pro model. We

evaluated them both quantitatively, using the validation pipeline described in Section 4.2, and qualitatively by manual inspection.

Our initial prompt strategy left ample freedom for the LLM to choose the deverbal noun, since it had only the verbal predicate to convert as input. This approach provided sentences in which the nominal predicate was not easily identifiable, preventing its linkage to an existing frame. Also, the model struggled in adhering to the desired format. Adding explicit instructions to keep the formatting consistent helped, but did not completely solve, this issue. The strategy of adding examples was more effective for the purpose of keeping the format consistent. To facilitate the identification of the resulting deverbal noun, a request to mark it in the output was added to the prompt. This solved any issue concerning the parsing of the output. At this stage, one problem still persisted: in some cases, the deverbal nouns produced by the LLM were impossible to link to an existing frame. To solve this issue, we included in the prompt a list of candidate deverbal which were directly linkable to a frame to choose from. Following this process, the resulting sentences were: i) adherent to the required format, ii) indicating and using a correct nominal predicate, i.e. a predicate linked via DRF to the verbal one, and iii) as similar as possible to the original one while preserving grammar and logic adequacy. As far as the selection of the LLM was concerned, we observed that gpt-3.5-instruct followed the instructions provided but it would often generate grammatically incorrect sentences when forced to use a specific deverbal noun. Instead, gpt-3.5-turbo and Gemini-pro produced the most natural sounding sentences. Gemini-pro was finally chosen based on a manual inspection of the results.

The validation results on a subset of 150 SemCor sentences converted using the selected prompt are shown in Table 5.

B   Prompts and Examples for LLMs

Some examples of instantiated prompts for the LLM are reported below:

Change the sentence by nominalizing the verb "produced" indicated by **. Use exactly one of these deverbal nouns: "production". Indicate the chosen deverbal noun with **: "The Fulton County Grand Jury said Friday an investigation of

Atlanta 's recent primary election **produced** " no evidence " that any irregularities took place ."

Change the sentence by nominalizing the verb "charged" indicated by **. Use exactly one of these deverbal nouns: "charge", "mission", "commission", "direction". Indicate the chosen deverbal noun with **: "The September October term jury had been **charged** by Fulton Superior Court Judge Durwood Pye to investigate reports of possible " irregularities " in the hard-fought primary which was won by Mayor-nominate Ivan Allen Jr. ."

The system directive used for the LLM is:

You are a linguist that can nominalize sentences. The sentences must be grammatically correct.

Finally, we report two of the 10 examples used for few-shot prompting. The examples are shuffled at each prompt request and mimic the structure of a realistic User-LLM chat-like conversation:

User: Change the sentence by nominalizing the verb "absorb" indicated by **. Use exactly one of these deverbal nouns: "soaking up", "absorption". Indicate the chosen deverbal noun with **: "The kidneys **absorb** sodium as a part of renal physiology ."

LLM: The **absorption** of sodium by the kidneys is a part of renal physiology.

User: Change the sentence by nominalizing the verb "mention" indicated by **. Use exactly one of these deverbal nouns: "citation". Indicate the chosen deverbal noun with **: "Vandiver likely will **mention** the $ 100 million highway bond issue approved earlier in the session as his first priority item ."

LLM: Vandiver's **citation** of the $100 million highway bond issue approved earlier in the session will likely be his first priority item.

C Guidelines for the human evaluation

In this Section, we report the instructions provided to the annotator about the task of classifying the set of nominal synsets with non-existing links (cf. Section 3.1.2) into frames.

LLM model	valid samples
GPT-3.5-turbo-1106	132/150 (88%)
GPT-3.5-turbo-instruct	130/150 (86.6%)
text-davinci-003	102/150 (68%)
Gemini-Pro	136/150 (90.6%)

Table 5: Number of samples passing validation across LLMs

We provided two Excel files.

The first file:

- Contains the nominal synsets requiring classification.
- For each synset, presents five frames automatically selected by a machine learning model.
- Annotators must choose the single most fitting frame from these five options for each synset.

The second file:

- Lists all 425 available frames within VerbAtlas along with their respective definitions.
- For each frame, lists all the included verbal synsets and their respective definitions.
- Serves as a reference for exploring frames beyond the five presented in the first file for specific synsets.

The classification process has to follow these steps for each synset:

1. Review the synset itself (column A), its lemmas (column B), and its definition (column C).

2. Examine the five proposed frames (column B) with their semantic definitions (column C). Additionally, consider the definitions explaining how synsets within the frame relate to its specific roles. For example, the frame "LOSE" might have the definition "An agent LOSES a theme to a recipient".

While selecting the appropriate frame, possible cases arise:

1. **Suitable frame in the 5 proposed:** Place the "x" in the corresponding row.

2. **No suitable frame among the 5 proposed:**

   - Consult the second file to find an appropriate frame among the 425 options.

- Replace the name of a proposed frame with your chosen frame from the second file (keep the original definition).
- Place an "x" in the corresponding row.

3. **Synset belongs to multiple frames:** Place an "x" in each relevant frame row (Minimize this occurrence; use it only for true ambiguity where a single choice is impossible. Add a note in column D explaining the ambiguity).

4. **No existing frame seems appropriate:** Don't mark any frames.

5. **Synset is not a nominal predicate:** Don't mark any frames. This is uncommon; attempt classification even for broader event concepts.