

# Adverbs Revisited: Enhancing WordNet Coverage of Adverbs with a Supersense Taxonomy

Jooyoung Lee<sup>1,2</sup>, Jader Martins Camboim de Sá<sup>1,3</sup>, Cédric Pruski<sup>1</sup>

<sup>1</sup>Luxembourg Institute of Science and Technology  
5 avenue des Hauts-Fourneaux, L-4362, Esch-sur-Alzette, Luxembourg,

<sup>2</sup>Department of Linguistics - Brown University  
Providence, RI 02912, United States,

<sup>3</sup>FSTM - University of Luxembourg  
2 place de l'Université, L-4365, Esch-sur-Alzette, Luxembourg,

Correspondence: [first.second@list.lu](mailto:first.second@list.lu)

## Abstract

WordNet offers rich supersense hierarchies for nouns and verbs, yet adverbs remain underdeveloped, lacking a systematic semantic classification. We introduce a linguistically grounded supersense typology for adverbs, empirically validated through annotation, that captures major semantic domains including manner, temporal, frequency, degree, domain, speaker-oriented, and subject-oriented functions. Results from a pilot annotation study demonstrate that these categories provide broad coverage of adverbs in natural text and can be reliably assigned by human annotators. Incorporating this typology extends WordNet's coverage, aligns it more closely with linguistic theory, and facilitates downstream NLP applications such as word sense disambiguation, event extraction, sentiment analysis, and discourse modeling. We present the proposed supersense categories, annotation outcomes, and directions for future work.

**Keywords:** WordNet, Adverbs, Taxonomy

## 1. Introduction

As a primary lexical class, adverbs perform a range of semantic functions, from answering fundamental questions about an event, such as how it was performed (manner), when it occurred (temporal), or to what extent a property holds (degree), to expressing speaker attitude, discourse stance, and logical relations between propositions. Despite this semantic richness, adverbs have long occupied an ambiguous and often marginalized position in linguistic classification, frequently described as a "residual" or "wastebasket" category (Huddleston and Pullum, 2002; Nikolaev et al., 2023). Words are often assigned to this category not because they share definable grammatical properties, but because they fail to conform to the morphological and syntactic criteria of nouns, verbs, adjectives, prepositions, or conjunctions. This stands in contrast to the detailed formal and semantic analyses developed for nouns and verbs. (Conlon and Evens, 1994).

The heterogeneity of adverbs arises from the multiple dimensions along which they vary. Syntactically, adverbs can attach at different heights, within the VP, clause, or discourse level, modifying events, participants, or propositions. Semantically, they range from event-oriented (*quickly*, *carefully*) to epistemic and evaluative (*probably*, *fortunately*) to domain-restrictive (*legally*, *technically*) meanings. Morphologically, English adverbs derive through diverse pathways, productively with *-ly*, through lexicalization (*often*, *maybe*), or borrowing (*a priori*, *per se*), each contributing to internal diversity. Col-

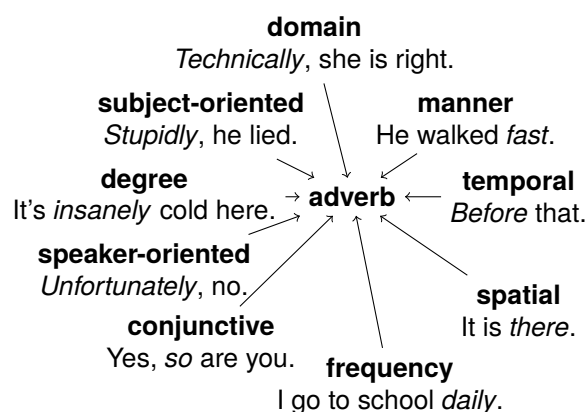


Figure 1: Taxonomy for adverb supersenses.

lapsing such functionally distinct elements under a single label produces an uninformative and theoretically incoherent grouping that obscures meaningful semantic distinctions.

This conceptual heterogeneity has far-reaching implications beyond linguistic theory. In computational linguistics and large lexical databases, adverbs often receive limited or inconsistent representation, reflecting their ambiguous grammatical status. Major resources such as WordNet, for instance, tend to prioritize manner adverbs while overlooking domain, epistemic, or evaluative meanings. This imbalance illustrates how theoretical uncertainty surrounding adverbs has translated into gaps in computational modeling and lexical coverage, motivating the need for a more systematic and semantically grounded classification.

Distinguishing these meanings is not merely a matter of theoretical precision. Adverbs frequently serve as indicators of how speakers assess states of affairs, making them highly informative for computational tasks such as opinion mining, stance detection, and sentiment analysis (Pang and Lee, 2008). Their semantic function extends beyond simple modification to include attitudinal and intensificational meanings that shape pragmatic interpretation. Without an explicit taxonomy capturing distinctions such as manner, evaluative, and intensifier uses, NLP systems cannot systematically model these meanings.

This paper addresses this gap by proposing a linguistically motivated and empirically validated typology of adverbs and demonstrating its effectiveness through computational classification. We present a taxonomy encompassing event-related, domain-related, subject-oriented, and speaker-oriented adverbs (see Creffig:taxonomy), and implement an annotation pipeline to facilitate the systematic expansion and annotation of WordNet adverbs.

## 2. Background

In most syntactic and computational frameworks, adverbs have been treated as adjuncts, optional modifiers situated outside the core predicate–argument structure. Because early parsing systems were primarily designed to capture argument relations, adverbial phrases were typically attached in a flat, non-hierarchical fashion, with little representation of their semantic contribution. This simplified treatment, inherited by subsequent treebanks and dependency parsers, has had enduring effects: adverbs are encoded as syntactically peripheral and semantically underspecified elements (Ernst, 2002; McConnell-Ginet, 1982).

These structural simplifications have shaped how adverbs are represented in modern computational resources and tasks. Large lexical databases such as WordNet (Miller, 1995; Fellbaum and Miller, 1998), syntactic annotation schemes like the Penn Treebank (Marcus et al., 1993), and semantic frameworks including semantic role labeling all encode adverbs in simplified or undifferentiated ways, reflecting a broader trend of structural and semantic underrepresentation across NLP.

### 2.1. WordNet

WordNet is among the most influential lexical–semantic resources in computational linguistics, Natural Language Processing (NLP), and cognitive science. Its architecture organizes words into synsets grouped by supersenses, capturing broad conceptual domains across parts of speech. It covers 26 noun supersenses and 15 verb su-

persenses, allowing for systematic semantic generalization. On the other hand, all adverbs are placed under a single supersense, “adv.all”, which collapses their internal diversity into a monolithic category. This flattening of the semantic space not only underrepresents the heterogeneity of adverbial meaning but also constrains the usefulness of WordNet for adverb-related research.

The WordNet documentation explicitly states that “there are only few adverbs in WordNet . . . as the majority of English adverbs are straightforwardly derived from adjectives via morphological affixation” (Princeton University, 2025). This assumption simplifies adverb formation to a purely morphological process and conflates morphological derivation with semantic equivalence, failing to recognize the interpretive autonomy of adverbs. For instance, while the adverb *stupidly* in *He danced stupidly* concerns the manner in which John danced, the counterpart in *Stupidly, he danced* concerns the speaker’s evaluation on John’s decision to dance and does not have any inference about *how* he danced. Such semantic contrast cannot be predicted from adjectival meaning alone. The assumption that adverb formation is semantically straightforward is therefore untenable.

### 2.2. Penn Treebank

The Penn Treebank (PTB), a foundational syntactic resource for NLP, serves as the basis for countless parsing models and downstream datasets. However, its treatment of adverbs mirrors the structural simplifications seen in WordNet. The tagset distinguishes only four adverb categories: RB for general adverbs such as *quickly*, *often*, and *perhaps*; RBR for comparative adverbs such as *faster* and *earlier*; RBS for superlative adverbs such as *best* and *fastest*; and WRB for wh-adverbs such as *when*, *where*, and *why*. These tags primarily capture morphological and comparative information, making no reference to syntactic scope or semantic function. As a consequence, adverbs with vastly different meanings and syntactic behaviors such as *quickly* (manner), *probably* (epistemic), *fortunately* (speaker-oriented), and *technically* (domain) are all collapsed into the identical RB tag, treating them as a single class.

This coarse-grained tagging is not merely a semantic omission; it creates a problem for syntactic parsing itself. The primary challenge a parser faces with an RB tag is attachment ambiguity. It has no information to determine if the adverb should attach low to the VP (e.g., *He spoke clearly* in a manner sense) or high to the entire Sentence (e.g., *Clearly, he is joking* in a speaker-oriented sense). By lumping all these adverbs together, the tagset discards the single most important clue for resolving this ambiguity. Since an adverb’s interpretive role is in-

separable from its syntactic scope, a semantically grounded taxonomy is not merely desirable but necessary for improving both syntactic and semantic accuracy.

### 2.3. PropBank

The Proposition Bank (PropBank; [Palmer et al. 2005](#)) extends the Penn Treebank by adding a layer of semantic role labeling, enriching syntactic parses with predicate–argument structure. Its design represents a significant step forward: PropBank introduced a detailed inventory of adjunct types that partially disaggregate adverbial meaning, including labels such as ARGM-TMP (temporal), ARGM-LOC (locative), ARGM-CAU (causal), ARGM-PRP (purpose), ARGM-MNR (manner), and ARGM-DIS (discourse connective). These distinctions capture coarse semantic relations that the PTB’s RB tag entirely ignores, allowing for more interpretable role structures and finer-grained feature extraction for semantic parsing models.

However, despite these improvements, PropBank’s treatment of adverbs still relies on a catch-all category, ARGM-ADV, which functions as a repository for any adverb that does not fit the existing labels. As [Nikolaev et al. \(2023\)](#) observe, this design reproduces the “wastebasket”: multiple distinct semantic types including epistemic, evaluative, subject-oriented, and domain adverbs—are merged into one ambiguous slot. A classic example like *stupidly* in *Stupidly, he answered* (meaning it was stupid of him) is neither a simple manner nor a speaker evaluation. With no dedicated category, it is inevitably forced into one of these ill-fitting boxes, most likely ARGM-MNR, thus demonstrating that PropBank’s schema is blind to this core linguistic distinction.

### 2.4. NLP tasks

The simplified representation of adverbs in foundational resources such as WordNet, the Penn Treebank, and PropBank has cascading effects across NLP tasks, causing a systematic loss of semantic precision. By treating adverbs as a single or weakly differentiated category, these resources obscure distinctions in semantic scope, orientation, and function, differences that are crucial for interpreting event structure and speaker stance. This problem is especially clear in Word Sense Disambiguation, where the sense inventories themselves encode the bias. For instance, SenseBERT ([Levine et al., 2020](#)) doesn’t learn polysemy for adverbs and similarly WiC models ([Pilehvar and Camacho-Collados, 2019](#)) are not trained for adverb differentiation. Adverb are often merged into one entry, leaving systems unable to identify the correct sense

because the relevant contrast is missing. Such limitations affect machine translation and sentiment analysis as well, where adverbs of stance, degree, and certainty are often mistranslated or misinterpreted. A semantically grounded taxonomy of adverbs would provide explicit cues for scope and orientation, improving semantic role identification.

This problem extends directly to Semantic Role Labeling (SRL), where the semantic layer in PropBank inherits the Treebank’s structural simplifications. Although PropBank improves on purely syntactic annotation by introducing a range of adjunct roles, its framework still conflates functionally distinct adverbs under general labels such as ARGM-ADV or ARGM-MNR. This design erases the distinction between adverbs that modify the event, those that evaluate the agent’s participation, and those that comment on the proposition or discourse. As a result, SRL models cannot exploit adverb semantics to constrain predicate–argument interpretation or infer pragmatic scope. A semantically grounded taxonomy of adverbs, which encodes the level of modification and orientation, would therefore provide crucial cues for both syntactic attachment and argument interpretation, enhancing the performance of SRL, WSD, and related tasks that depend on fine-grained semantic distinctions.

## 3. Related Work

Developing semi-automatic methods for WordNet expansion remains a long-standing challenge ([Ciarmita and Johnson, 2003](#); [Curran, 2005](#)), since automatic approaches have yet to meet the quality standards expected of lexicographic resources ([Rundell, 2024](#)).

[Tsvetkov et al. \(2014\)](#) address the absence of a semantic hierarchy for adjectives in English WordNet, a limitation closely parallel to that of adverbs. While WordNet provides a rich taxonomic structure for nouns and verbs, adjectives are organized only into flat, unstructured clusters. To overcome this limitation, the authors introduce a coarse-grained taxonomy of 13 adjective supersenses, adapted from GermaNet, thereby imposing a hierarchical organization onto the adjective lexicon. Acknowledging that fully manual annotation of all 18,156 adjective synsets would be infeasible, they develop a semi-automatic classification approach.

[Maziarz et al. \(2016\)](#) tackle the underrepresentation of adverbs in plWordNet, arguing that simply treating them as derivatives of adjectives is insufficient. Their methodology is twofold: first, they define a formal set of semantic relations for adverbs, including hyponymy, gradation, and antonymy, which is largely adapted from their adjective model. Second, they implement a semi-automatic procedure to bootstrap the adverb

lexicon from the existing adjective network. This procedure generates adverbial counterparts for adjective lexical units and groups them into synsets that mirror the structure of the adjective synsets, providing a starting point for manual verification.

While previous work has investigated the expansion of supersenses for adjectives and the extension of adverb coverage in the Polish WordNet, to the best of our knowledge, no prior study has developed a taxonomy or systematically expanded the coverage of adverbs in the English WordNet.

#### 4. Taxonomy

The typology proposed here builds on the tripartite classification of manner, subject-oriented, and speaker-oriented adverbs (Jackendoff, 1972), which first recognized that adverbs differ in the level of meaning they target: the event, the agent, or the speaker, respectively.

Building on Jackendoff's tradition, our typology expands this inventory to include classes that systematically differ in interpretive scope, such as frequency, temporal, spatial, degree, domain, focus, and conjunctive adverbs. We ground this expanded classification in the rigorous syntax-semantics mapping proposed by Cinque (1999) and the scope-based approach of Ernst (2002). Both works argue that these distinct semantic functions are not random but correspond to a rigid hierarchy of syntactic positions, finding that is critical for resolving the attachment ambiguity that plagues computational parsers.

This combined theoretical grounding provides a principled framework for our classification. It allows us to move beyond a flat list and organize adverbs by their level of semantic scope: from event-internal modifiers (manner, spatial, temporal, frequency) and agent-evaluative modifiers (subject-oriented), to proposition-level modifiers (speaker-oriented, domain) and discourse-level operators (conjunctive, focus).

- **Manner** adverbs describe how an event or action is performed. They can typically be paraphrased using the expression *in a [X] manner*. For example, *He danced stupidly* may be rephrased as *He danced in a stupid manner*.
- **Subject-oriented** adverbs attribute a property or attitude to the subject of the sentence in relation to the event. These adverbs often allow a paraphrase of the form *It was [X] of [SUBJECT] to [VERB]*, as in *Stupidly, he walked into traffic* → *It was stupid of him to walk into traffic*.
- **Speaker-oriented** adverbs express the speaker's stance, evaluation, or attitude toward the proposition or the act of speaking.

They can frequently be paraphrased with constructions such as *I [believe/say/judge] that... or It is [unfortunate/fortunate/evident] that...* For instance, *Presumably, he missed the deadline* corresponds to *I presume that he missed the deadline*, while *Unfortunately, the project failed* can be restated as *It is unfortunate that the project failed*, and *Frankly, I disagree* as *I say this frankly*.

- **Frequency** adverbs indicate how often an event occurs and typically answer the question *How often?*, as in *She frequently visits her grandmother*.
- **Temporal** adverbs specify when an event happens or its duration, answering questions like *When?* or *For how long?* (e.g., *She arrived yesterday*; *He stayed briefly*).
- **Spatial** (or locative) adverbs refer to where the event takes place and can be identified by the question *Where?*, as in *He stood outside*.
- **Degree** adverbs describe the intensity or extent of another element, typically answering *To what extent?* or *How much?*, as in *She is very happy*.
- **Domain** adverbs restrict the interpretation of the proposition to a particular semantic or disciplinary domain, often paraphrasable as *In a [X] sense* or *From a [X] perspective*; for example, *Politically, the policy is controversial* → *In a political sense, the policy is controversial*.
- **Focus** adverbs highlight or limit the scope of the proposition to a specific constituent, marking inclusion, exclusion, or emphasis. They answer questions such as *What part of the clause is being emphasized or limited?* or *Does it indicate exclusivity, inclusion, or emphasis?* Examples include: *Only John passed the exam* (exclusivity); *Even Mary noticed the mistake* (unexpected inclusion); *She also attended the meeting* (addition); and *He mainly spoke about politics* (restriction).
- **Conjunctive** adverbs signal opposition, correction, connection, or contrast between propositions or discourse segments, relating the current clause to an alternative or prior context. They are often paraphrasable by *in contrast*, *on the other hand*, or *however*. Typical examples include *However, she decided to stay* (contrast with a preceding statement), *Conversely, rural areas saw a decline* (opposition), *Instead, he chose to wait* (correction), and *Nevertheless, they continued the project* (concession).

## 5. Methodology

The expansion of adverb supersenses in Open English WordNet (OEWN) follows a semi-automatic annotation pipeline, in which each computational stage is manually validated by professional linguists. The process first involves categorizing existing adverbs in OEWN to determine which semantic senses warrant inclusion in the extended taxonomy. New adverbial senses are then identified through large-scale corpus analysis, where adverbs are extracted, their contextual meanings distinguished, and representative examples selected. These examples are subsequently integrated into OEWN in accordance with its structural and formatting conventions. An overview of this workflow is presented in Figure 2.

### 5.1. Adverb Extraction

The first step in Figure 2, consists in collecting a large, representative corpus and extracting all adverbial forms. Adverb candidates are identified through a combination of dictionary lookup and morphological detection, primarily targeting common adverbial suffixes (e.g., *-ly*). Surface forms are preserved throughout the process to maintain alignment with corpus usage. Ambiguous forms that can function as multiple parts of speech (e.g., *fast*, *hard*) are excluded at this stage to avoid introducing noise into the adverb inventory.

### 5.2. Lexical Sense Differentiation

For each extracted adverb, Figure 2 second step, distinct lexical usages are identified using cosine distance of the contextualized embeddings of each usage. The model compares contextual embeddings across occurrences of the same adverb, computing pairwise cosine distances to cluster similar contexts. New sense discovery is entirely data-driven, with the number of senses emerging from the clustering process rather than being predefined. Following the OEWN guidelines, a cluster should be composed of at least 100 elements to be considered a valid sense (McCrae et al., 2021). Each resulting sense is subsequently assigned a dictionary definition by annotators, who also classify the adverb according to its semantic category (e.g., manner, degree, temporal, frequency).

### 5.3. Example Sentence Selection

To provide representative examples for each sense, sentences are ranked according to the contextual embedding uncertainty estimated by the differentiation model. The sentence with the lowest entropy value is selected as the prototypical example. Sentences are further filtered by length, retaining only

those containing between 3 and 20 tokens to ensure naturalness and readability. Annotators then manually verify each selected example to confirm that it accurately illustrates the intended sense.

### 5.4. Synset Formation

Synset construction is performed by aligning newly identified adverb senses with existing or newly created adverb synsets. Sense alignment is based on cosine similarity between sense embeddings, applying a similarity threshold to determine equivalence. Only adverbs belonging to the same semantic category are eligible for grouping. Existing OEWN resources are consulted to validate and anchor new synsets within the current lexical network. When multiple senses show near-equivalence or redundancy, they are manually reviewed and merged as appropriate.

### 5.5. Validation and Integration

All newly generated entries are formatted in YAML, following the OEWN schema. The final output includes the adverb lemma, sense identifier, gloss, and verified example sentence. Inter-annotator agreement is computed for the taxonomy for the initial round, then each annotator works on independent examples. Finally, at the synset construction, a manual post-processing step ensures consistency by removing or merging near-duplicate synsets before integration into OEWN.

## 6. Experiments

To improve the adverbs coverage in OEWN (McCrae et al., 2021) the open version derived from (Miller, 1994), we start by training two annotators in the taxonomy mentioned in Section 4 and detailed in the Appendix. These annotators have English as their second language and a background in linguistics. After some short rounds to train and clear annotators doubts, we run a full annotation round. We observed a substantial annotator agreement (Cohen's kappa = 0.67; n = 229).

After the initial round, the annotators independently annotate examples from WordNet following the guidelines, in Table 1 we demonstrate examples of the supersenses annotated for existing examples in OEWN. In Figure 3 we plot the distribution of categories for adverbs after this annotation.

To extend existing usages for adverbs and discover new ones., we began by compiling a large-scale corpora. For each adverb, we extracted millions of usage instances from OpenSubtitles (Community, 2025) and Fineweb-edu (Lozhkov et al., 2024). To ensure linguistic diversity, we included data from multiple domains and registers. Each

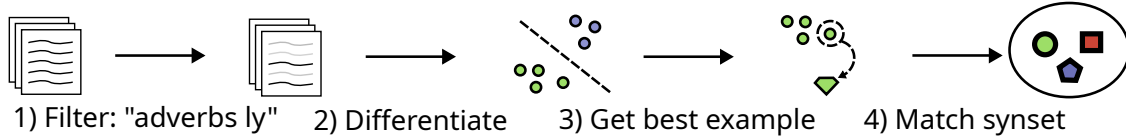


Figure 2: Pipeline for assisting the annotation of adverbs.

| Category         | Synset ID  | Adverb          | Usage  |
|------------------|------------|-----------------|--|
| conjunctive      | 00043413-r | thus            | it is late and thus we must go                       |
| degree           | 00006640-r | significantly   | our budget will be significantly affected [...]      |
| domain           | 00131423-r | semantically    | semantically empty messages                          |
| focus            | 00009062-r | alone           | [...] rests on the prosecution alone                 |
| frequency        | 00256795-r | biannually      | we hold our big sale biannually                      |
| manner           | 00248938-r | horrifyingly    | he laughed horrifyingly                              |
| spatial          | 00259792-r | inland          | the town is five miles inland                        |
| speaker_oriented | 00201575-r | hopefully       | hopefully the weather will be fine on Sunday         |
| subject_oriented | 00473325-r | superstitiously | superstitiously he refused to travel on Friday [...] |
| temporal         | 00061170-r | previously      | he was previously president of a bank                |

Table 1: Examples of adverb categories, their WordNet synset IDs, and usage examples.

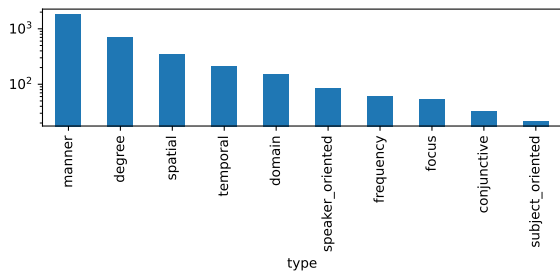


Figure 3: Counts of types for adverbs in WordNet.

instance was stored along with its surrounding sentence or paragraph context to support later sense clustering and disambiguation.

To capture the potential polysemy of each adverb, we organized its usage instances into clusters representing distinct contextual senses. This step was performed using XL-Lexeme (Cassotti et al., 2023), available in HuggingFace<sup>1</sup>. XL-Lexeme, a multilingual sense differentiation with empirical good generalization, computes contextual embeddings for each adverb occurrence and then we group them using a Agglomerative Clustering with a distance threshold of 0.4. Each resulting cluster corresponds to a potential sense of the adverb, defined by semantically coherent usage contexts.

After the supersense classification, we used the ChatGPT-5-mini API to generate sense definitions for each cluster (prompt provided in Section A). Each definition was produced by prompting the

<sup>1</sup><https://huggingface.co/pierluigic/xl-lexeme>

model with representative examples from the cluster and requesting a concise, WordNet-style definition. We selected the examples with the lowest entropy, as described in Equation (1), when the adverb is masked and feed to the XLM-RoBERTa model.

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i) \quad (1)$$

Where  $P(x_i)$  is the probability of the  $i$ -th token in the model's vocabulary being the correct token for the mask. This process allowed us to automatically select high-quality examples and derive lexicographically interpretable sense entries for each adverb.

All automatically generated senses underwent manual evaluation by trained annotators. The evaluation involved verifying the accuracy of the supersense label, checking for redundancy across clusters, and refining the sense glosses for lexical and stylistic consistency. To this task we assigned two annotators where English is their second language. In the next section we report inter-annotator agreement. Errors were corrected, redundant clusters were merged, and unclear definitions were rewritten following WordNet conventions. The final resource includes 418 new adverbs or usages not present in WordNet and constitutes a validated extension of adverb supersenses aligned with the WordNet ontology.

We discarded existing senses based on the adverb category and in Figure 4 we present the statistics of the new obtained category and senses.

The resulting resource, including the adverb

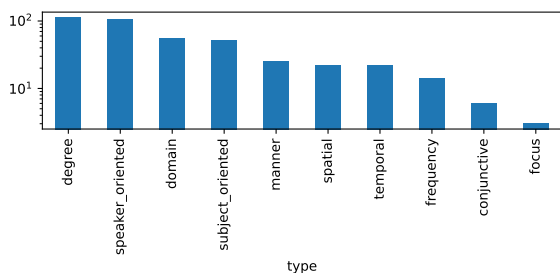


Figure 4: Counts of types for adverbs not present in WordNet.

sense inventory and annotation guidelines, will be released under an open license and made available via Github.

## 7. Discussion

The primary objective of this work was to expand the coverage of adverbs in the OEWN through a novel semi-automatic pipeline to assist professional annotators. The methodology, combining large-scale computational analysis with essential manual validation, proved effective, resulting in the creation of 418 new adverb synsets previously absent from WordNet. This outcome validates the hybrid *human-in-the-loop* approach, demonstrating that modern NLP tools can significantly accelerate and scale lexicographic efforts.

A key strength of our pipeline is the integration of large models (LLMs) for complex annotation tasks. This allowed our trained annotators to focus on validation and refinement rather than from scratch sense discovery. The selection of large, diverse corpora (Community, 2025; Lozhkov et al., 2024) ensured that the identified senses reflect contemporary and varied language use.

The validation phase yielded a substantial inter-annotator agreement (Cohen’s Kappa =  $0.67 \pm .095$ ,  $n = 229$ ) for the semantic category classification. This result underscores the inherent ambiguity and complexity of adverb classification, even for trained annotators working with detailed guidelines. It also confirms that the task is feasible and that our guidelines provide a solid basis for consistent annotation. The manual verification by two trained L2 English annotators was crucial for correcting errors, merging redundant clusters, and ensuring the final definitions met OEWN’s lexicographical standards. The comparison between the original adverb distribution in Figure 3 and our new additions in Figure 4 highlights that our method successfully identified senses in categories that were previously under-represented.

## 8. Conclusion

This paper presented a semi-automatic pipeline for the identification, differentiation, and integration of new adverb synsets into OEWN. By combining contextualized embeddings for sense clustering, LLMs for bootstrapped annotation, and expert manual validation, we added 418 new high-quality adverbial senses and categorized 3,493 existing senses in OEWN. This work makes two key contributions: first, it delivers a substantial and concrete expansion of OEWN, an essential lexical resource for the NLP community; second, it introduces a replicable methodology that effectively balances the scalability of computational models with the precision of human annotation.

The resulting adverb inventory and accompanying annotation guidelines, which will be released publicly, constitute a valuable resource for future research in lexical semantics, computational lexicography, and natural language understanding. Future work should extend this pipeline to address POS-ambiguous adverbs excluded from the present study and to evaluate the transferability of this methodology to other parts of speech whose sense inventories remain underspecified or outdated.

## Limitations

Despite the pipeline’s success, several limitations warrant acknowledgment. First, the methodology excluded ambiguous forms that can function across multiple parts of speech (e.g., fast, hard). While this restriction was necessary to minimize noise during extraction and clustering, it leaves both our resource and OEWN with limited coverage of morphosyntactically ambiguous adverbs, which are among the most frequent and semantically complex in English. Addressing these cases remains an important direction for future work. Second, the quality of the automated components depends on model-specific factors—particularly the XL-Lexeme embedding model and the clustering threshold (set to 0.4). Alternative models or parameters could yield different levels of sense granularity, and the need for manual post-merging indicates that clustering occasionally produced redundant or overly fine-grained senses. Finally, although two annotators carried out the validation phase, incorporating a larger and more diverse annotation pool, including L1 English-speaking lexicographers, would likely improve consistency and further enhance the robustness of the resulting resource.

## Acknowledgements

The research reported in this publication was supported by the Luxembourg National Re-

search Fund (FNR), project D4H grant number PRIDE21/16758026

## 9. Bibliographical References

- Pierluigi Cassotti, Lucia Siciliani, Marco DeGemmis, Giovanni Semeraro, and Pierpaolo Basile. 2023. [XL-LEXEME: WiC pretrained model for cross-lingual LEXical sEMantic changE](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1577–1585, Toronto, Canada. Association for Computational Linguistics.
- Massimiliano Ciaramita and Mark Johnson. 2003. [Supersense tagging of unknown nouns in WordNet](#). In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 168–175.
- Guglielmo Cinque. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. Oxford University Press.
- Sumali Pin-Ngern Conlon and Martha Evens. 1994. An adverbial lexicon for natural language processing systems. *International Journal of Lexicography*, 7(3):197–221.
- James Curran. 2005. [Supersense tagging of unknown nouns using semantic similarity](#). In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 26–33, Ann Arbor, Michigan. Association for Computational Linguistics.
- Thomas Ernst. 2002. *The syntax of adjuncts*, volume 96. Cambridge University Press.
- Christiane D. Fellbaum and George A. Miller. 1998. [Combining local context and wordnet similarity for word sense identification](#).
- Rodney Huddleston and Geoffrey K. Pullum. 2002. [The cambridge grammar of the english language by rodney huddleston](#).
- Ray S Jackendoff. 1972. Semantic interpretation in generative grammar.
- Yoav Levine, Barak Lenz, Or Dagan, Ori Ram, Dan Padnos, Or Sharir, Shai Shalev-Shwartz, Amnon Shashua, and Yoav Shoham. 2020. [SenseBERT: Driving some sense into BERT](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4656–4667, Online. Association for Computational Linguistics.
- Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of english: the penn treebank. *Comput. Linguist.*, 19(2):313–330.
- Marek Maziarz, Stan Szpakowicz, and Michal Kalinski. 2016. [Adverbs in plWordNet: Theory and implementation](#). In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 209–217, Bucharest, Romania. Global Wordnet Association.
- Sally McConnell-Ginet. 1982. Adverbs and logical form: A linguistically realistic theory. *Language*, pages 144–184.
- John P. McCrae, Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka, and Luis Morgado Da Costa. 2021. [The GlobalWordNet formats: Updates for 2020](#). In *Proceedings of the 11th Global Wordnet Conference*, pages 91–99, University of South Africa (UNISA). Global Wordnet Association.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Dmitry Nikolaev, Collin F Baker, Miriam RL Petruck, and Sebastian Padó. 2023. Adverbs, surprisingly. *arXiv preprint arXiv:2305.19650*.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.
- Bo Pang and Lillian Lee. 2008. [Opinion mining and sentiment analysis](#). *Found. Trends Inf. Retr.*, 2:1–135.
- Mohammad Taher Pilehvar and Jose Camacho-Collados. 2019. [WiC: the word-in-context dataset for evaluating context-sensitive meaning representations](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1267–1273, Minneapolis, Minnesota. Association for Computational Linguistics.
- Princeton University. 2025. Wordnet: A lexical database for english. <https://wordnet.princeton.edu/front>. Accessed: 2025-10-24.
- Michael Rundell. 2024. Automating the creation of dictionaries: are we nearly there? *Humanising Language Teaching*, 26(1).

Yulia Tsvetkov, Nathan Schneider, Dirk Hovy, Archana Bhatia, Manaal Faruqui, and Chris Dyer. 2014. [Augmenting English adjective senses with supersenses](#). In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4359–4365, Reykjavik, Iceland. European Language Resources Association (ELRA).

## 10. Language Resource References

OpenSubtitles Community. 2025. [OpenSubtitles](#). OpenSubtitles Website. Dataset updated on 4th August 2025; licensed under CC0.

Lozhkov, Anton and Ben Allal, Loubna and von Werra, Leandro and Wolf, Thomas. 2024. [FineWeb-Edu](#). Hugging Face Website. Dataset updated on 4th August 2025; licensed under ODC-BY.

John P. McCrae, Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka, and Luis Morgado Da Costa. 2021. [The GlobalWordNet formats: Updates for 2020](#). In *Proceedings of the 11th Global Wordnet Conference*, pages 91–99, University of South Africa (UNISA). Global Wordnet Association.

Miller, George A. 1994. [WordNet: A Lexical Database for English](#). ACM New York, NY, USA.

### A. Prompt for Sense Modeling

Below we present the template for querying ChatGPT-5-mini for sense definitions.

```
Give a definition for the adverb
"<word>" in the sentence: "<sen-
tence>". Using wordnet-style, re-
spond only with the definition.
```