A Comprehensive Survey of Semantic Role Labeling

Junjiao Li[1,2], Zhengjie Gao[1,2,*]

[1] Geely University of China; China
[2] Intelligent Computing & Applications Lab; China

Abstract. Semantic Role Labeling (SRL) is a core Natural Language Processing (NLP) task focused on identifying predicate-argument structures and assigning semantic labels like agent or goal, which is crucial for natural language understanding (NLU) in applications such as question answering and information extraction. The field has evolved from early reliance on syntactic information to advanced neural network architectures, including Transformer models and pre-trained language models (PLMs). Key approaches include span-based and dependency-based SRL, with a growing trend towards incorporating higher-order graph structures for richer interactions. A novel paradigm, Definition-based SRL (DSRL), redefines the task as natural language generation, where models explain semantic relationships in human-readable definitions. Despite progress, challenges persist in handling nominal and non-verbal predicates, cross-lingual and low-resource scenarios due to data scarcity, and annotation inconsistencies. Future work aims for end-to-end and document-level SRL, integrating broader contextual information. Large Language Models (LLMs) are significantly transforming SRL by enabling data generation and serving as core architectural components for tasks like DSRL, promising enhanced semantic understanding and zero-shot capabilities. However, challenges remain regarding computational cost and potential for erroneous outputs.

CCS Concepts: Computing methodologies → Artificial intelligence → Natural language processing

Keywords: Semantic Role Labeling, Natural Language Processing, Natural Language Understanding, Predicate-Argument Structure, Neural Networks

1. Introduction

Semantic Role Labeling (SRL), often referred to as shallow semantic parsing or slot-filling, represents a foundational task within Natural Language Processing (NLP) [1]. This process involves the identification of the predicate-argument structure inherent in a sentence, subsequently assigning specific labels to words or phrases that denote their semantic function, such

* Corresponding Author: Zhengjie Gao (gaozhengjie@guc.edu.cn)

as that of an agent, a goal, or a result. The overarching objective of SRL is to unveil the underlying meaning of a sentence by systematically classifying how various arguments relate to the central predicate [2].

1.1. Definition and Significance of Semantic Role Labeling

The importance of SRL stems from its pivotal role in natural language understanding (NLU). It empowers computational systems to transcend mere word recognition, enabling them to comprehend the nuanced ways in which words function within diverse sentence constructions. This profound level of understanding is indispensable for a broad spectrum of downstream NLP applications. These applications include, but are not limited to, question answering, information extraction, automatic text summarization, text data mining, and speech recognition [2].

The fundamental components of SRL include:

- **Predicates:** These are the pivotal elements that convey the action or state within a sentence. While traditional SRL primarily focused on verbs, modern advancements recognize that predicates can also manifest as nouns, adjectives, or even adverbs [3]. The core function of SRL is to discern the relationships between different sentence constituents and this central predicate.
- **Arguments:** These refer to the words or phrases that are semantically connected to the predicate. For example, in the sentence "Mary sold the book to John," "Mary," "the book," and "John" are identified as the arguments [4].
- **Semantic Roles:** These are the descriptive labels assigned to arguments, precisely specifying their relationship to the predicate. Illustrative examples include "agent" for "Mary," "theme" for "the book," and "recipient" for "John". These roles effectively articulate the fundamental inquiries of "who, what, when, where, how, and why" regarding the actions or states described in the text[3].

The conceptual groundwork for semantic role labeling was established by Charles J. Fillmore in 1968, a contribution that subsequently inspired the development of the FrameNet project. FrameNet emerged as a pioneering computational lexicon, offering a systematic and detailed description of predicates and their corresponding semantic roles. Following this, the PropBank corpus was introduced, enriching the Penn Treebank (a collection of Wall Street Journal texts) with meticulously hand-annotated semantic roles. PropBank has since become a cornerstone training dataset for numerous automatic SRL systems.

1.2. Evolution and Formulations of SRL

Early research in NLP regarded SRL as an intrinsic component of natural language understanding, with initial algorithms consistently integrating syntactic information for feature modeling. Over time, SRL has been approached through two primary formulations in practice:

- **Span-based SRL:** This approach involves identifying contiguous segments of text, known as spans, and subsequently assigning semantic roles to them. The objective is to delineate and categorize these spans based on their semantic connection to a specified predicate[4].
- **Dependency-based SRL:** This formulation assigns semantic roles by meticulously analyzing the syntactic dependency relations that exist between words within a sentence. It concentrates on the structural connections between words to infer their semantic functions[5].

The progression of SRL methodologies reveals a compelling pattern concerning the utilization of syntactic information. Initially, SRL models exhibited a substantial reliance on explicit syntactic parses. However, with the rise of neural network models, there was a discernible shift towards approaches that were less dependent on explicit syntactic structures, prioritizing simplicity and robustness. More recently, research has seen a re-integration of syntactic insights, but in more refined and computationally efficient ways. This includes, for example, the use of supertags [6] or the re-framing of SRL as a dependency parsing task. This trajectory suggests that while syntactic structure is undeniably valuable for SRL, the method of its incorporation is paramount. Excessive reliance on potentially noisy, full syntactic parses can introduce errors, leading researchers to explore lighter-weight syntactic features[7] or implicit syntactic modeling within end-to-end neural architectures. The continued development and coexistence of both span-based and dependency-based formulations, alongside efforts to convert between them (as exemplified by research in "Semantic Role Labeling as Syntactic Dependency Parsing"), highlight an ongoing quest for the optimal representation of underlying linguistic structure. The emergence of "syntax-agnostic SRL" as a research direction[8] further indicates a potential trade-off between model simplicity and robustness versus the depth of linguistic understanding and performance, implying that the ideal balance remains an active area of investigation. This dynamic suggests that the field is advancing towards a more sophisticated comprehension of the interface between syntax and semantics. Future SRL systems are likely to incorporate adaptive mechanisms for syntactic integration, potentially learning which syntactic cues are most salient for different predicate types or languages[9], rather than uniformly applying a single parsing strategy. This adaptability could lead to more robust and generalizable SRL across diverse linguistic phenomena.

1.3. Applications of SRL in Natural Language Processing

SRL functions as a critical intermediate representation, significantly enhancing a broad spectrum of downstream NLP applications.

- **Question Answering (QA):** By identifying the "who, what, when, where, and how" of actions, SRL substantially improves a system's capacity to accurately respond to complex queries[10].
- **Information Extraction (IE):** SRL facilitates the structured extraction of pivotal entities and their interrelationships from unstructured text, a process vital for constructing knowledge bases and organized data[10].
- **Automatic Text Summarization:** A profound understanding of sentence meaning, enabled by SRL, allows for the generation of more coherent and informative text summaries.
- **Text Data Mining:** SRL augments the ability to uncover meaningful patterns and relationships within extensive volumes of text data.
- **Speech Recognition:** By furnishing semantic context, SRL contributes to a more precise interpretation of spoken language.
- **Natural Language Inference (NLI):** SRL can assist in discerning logical relationships and entailments between sentences, a core component of NLI.
- **Machine Translation (MT):** A precise comprehension of semantic roles enables more accurate and contextually appropriate meaning transfer across different languages[10].
- **Event Argument Extraction (EAE):** Given the structural commonalities in argument identification, SRL annotations prove to be a valuable resource for EAE. This effectively frames event argument extraction as a role querying problem[11].
- **Opinion Role Labeling (ORL):** Information derived from SRL can significantly boost ORL performance by providing semantic-aware word representations. This is particularly effective as predicate agents and patients frequently correspond directly to opinion holders and targets, making SRL a valuable source of prior semantic knowledge[12].
- **Out-of-Distribution (OOD) Detection:** SRL can guide OOD detection by facilitating the extraction of fine-grained local features from arguments. This process aids in filtering out anomalous phrases and enhances the learning of global-local features through self-

supervised prediction of SRL-extracted roles, leading to more accurate identification of subtle OOD patterns[13].

2. Methodologies and Architectural Advancements in Semantic Role Labeling

2.1. Neural Network Architectures for SRL

The landscape of SRL has been significantly shaped by advancements in neural network architectures. Early neural SRL systems frequently employed Bi-directional Long Short-Term Memory (BiLSTM) networks to encode contextual information within sentences. These recurrent architectures were effective in capturing sequential dependencies in text.

The advent of Transformer architectures and large-scale pre-trained language models (PLMs), such as BERT, RoBERTa, XLM-RoBERTa, and BART-large, marked a pivotal leap in SRL performance[14]. These models demonstrate exceptional capabilities in generating deep, contextualized word representations, which are indispensable for capturing intricate semantic relationships. For example, the study on "Joint End-to-End Semantic Proto-role Labeling" leveraged a BERT-based multi-task framework, showcasing its effectiveness in simultaneously predicting predicates, arguments, and proto-role attributes. Similarly, "Making Cross-Lingual Semantic Role Labeling Accessible with Intelligible Verbs and Roles" utilized XLM-RoBERTa for robust multilingual context encoding, underscoring its strong cross-lingual capabilities.

Modern SRL models often integrate PLMs with other specialized layers to optimize performance for specific tasks. Examples include combining a BERT encoder with a BiLSTM layer for tasks such as Emotional SRL (SRL4E), or employing RoBERTa in conjunction with GELU-activated linear layers to derive predicate-argument representations in models designed for learning from compatible label sequences. Beyond sequential processing, Graph Neural Networks (GNNs) are increasingly adopted to encode complex constituent or dependency structures within sentences. These networks provide rich relational information, which has been shown to enhance SRL systems, with their effectiveness demonstrated across various standard benchmarks[15].

2.2. Advanced Modeling Techniques

2.2.1. Multi-task Learning and Joint Modeling

A notable challenge in traditional SRL arises from the existence of multiple, often disjoint, label sets, such as PropBank and VerbNet. Modeling these sets independently can lead to inconsistencies in the output structure, thereby violating underlying semantic constraints. To mitigate this, researchers have proposed joint Conditional Random Field (CRF) models. These models conceptualize the integration of different label sets as a single, unified sequence labeling task, rather than treating them as separate problems. Resources like SEMLINK, which provides explicit mappings between PropBank and VerbNet labels, play a crucial role by enforcing compatibility constraints during the decoding process, which significantly reduces label inconsistencies[16].

The consistent efforts across multiple studies to integrate disparate label sets (PropBank, VerbNet, FrameNet)[17], to develop unified evaluation frameworks (SRL4E), and to build universal encoders (as seen in InVeRo-XL and "Unifying Cross-Lingual Semantic Role Labeling") point towards a profound underlying trend in NLP[18]: a strong impetus towards more unified, comprehensive, and interoperable semantic representations. This drive is motivated by the desire to create models that are not only more robust and generalizable[19] but also less constrained by specific, often arbitrary, annotation schemes or language-specific idiosyncrasies[3]. The recognition of "discrepancies" between SRL and related tasks like Event Argument Extraction (EAE) or Opinion Role Labeling (ORL) further emphasizes this need for bridging semantic gaps, suggesting that a shared semantic foundation can improve performance across diverse tasks. This overarching trend signifies a future where SRL systems are not merely isolated taggers for specific semantic frameworks but integral components of a broader, more integrated semantic understanding pipeline[20]. The demonstrated success of joint modeling and unified frameworks[4] indicates that explicitly modeling semantic relationships across different linguistic resources and tasks leads to superior performance and enhanced generalizability[21], paving the way for more holistic NLU systems. For instance, the paper "Learning Semantic Role Labeling from Compatible Label Sequences" demonstrated that a joint CRF approach not only improved overall performance by approximately 1.5 percentage points over multi-task learning but also completely eliminated label conflicts across test sets ($\rho = 0$), showcasing the power of unified modeling.

2.2.2. Higher-Order Modeling

Traditional, or "first-order," SRL models primarily focus on localized predicate-argument relationships, often treating each argument's role prediction independently. This approach frequently fails to capture the richer, global interactions and joint constraints that exist between multiple arguments or even between multiple predicates within a sentence[22].To overcome this limitation, researchers have introduced higher-order graph structures into SRL models.

These structures explicitly model complex interactions, such as "sibling" relationships (between arguments sharing a common predicate), "common parent" relationships (where arguments or predicates share a syntactic parent), and "grandparent" relationships. Innovative decoding methods, like approximate higher-order decoding combined with variational inference, are employed to manage the increased computational complexity while maintaining high performance[23].

The evolution from first-order to higher-order modeling in SRL reflects a deeper and more sophisticated understanding of linguistic structure. Sentences are not simply linear sequences of words or isolated predicate-argument pairs; they are intricate networks of interconnected elements with complex, often non-local, dependencies. The explicit move towards higher-order models, the development of graph reasoning networks, and even the re-formulation of SRL as a dependency parsing task (which inherently models global structural relationships) collectively indicate a profound recognition: capturing these complex, global interactions is absolutely crucial for achieving robust and accurate Natural Language Understanding. This represents a fundamental shift in focus from merely achieving "local accuracy" in role assignment to ensuring "global coherence" across the entire sentence structure. This trend suggests that future SRL systems will increasingly integrate more sophisticated graph-based or advanced attention mechanisms to effectively capture these high-order relationships. This progression could naturally lead towards full discourse-level or document-level SRL, where the semantic roles of arguments are understood not just within their immediate sentence but also in relation to information spanning multiple sentences and broader discourse contexts. Such advancements are vital for applications requiring deep contextual understanding, like long-form question answering or narrative comprehension. The paper "High-order Semantic Role Labeling" achieved new state-of-the-art results on the CoNLL-2009 multilingual benchmark across seven languages. This significant improvement was directly attributed to the effective modeling of these higher-order interactions, demonstrating their crucial role in capturing the nuances of complex sentences.

2.2.3. Mixture Models for Syntactic-Semantic Correlations

A novel SRL method has been introduced that explicitly models the intricate relationship between semantic label distributions and Shortest Syntactic Dependency Path (SSDP) patterns[5]. This approach, detailed in "Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models" , aims to achieve balanced annotation performance for both short-distance and long-distance semantic dependencies[5] without relying on language-specific hyperparameters. The core of this method involves using a mixture model to categorize SSDP jump patterns into multiple distinct clusters, each corresponding to an independent semantic label distribution. The final semantic label distribution is then derived by a weighted average of the semantic distributions from each cluster, utilizing mixture weights. The model employs a shared Transformer encoder[24], responsible for the joint encoding and optimization of both syntactic and semantic dependencies, enabling a holistic understanding of sentence structure.

2.3. SRL as a Natural Language Generation (NLG) Task

Traditional SRL methods, which rely on discrete label classification, can be rigid and may not fully capture the nuanced and continuous spectrum of semantic relationships between predicates and arguments. This approach often treats semantic roles as mere symbolic labels, overlooking their rich, meaningful definitions[10]. A novel paradigm redefines SRL as a natural language generation problem, termed Definition-based Semantic Role Labeling (DSRL). In this approach, given a sentence and its predicate, the model is tasked with generating a natural language definition that describes both the predicate's meaning and its semantic relationship with its arguments. This generated definition explicitly includes the predicate's meaning, the identification of arguments, and natural language descriptions of their semantic roles. DSRL is typically implemented using powerful pre-trained language models, such as BART-large, within a sequence-to-sequence generation framework. The input representation explicitly marks the predicate's position, and the output integrates the predicate's meaning and argument roles into the original sentence using special symbols (e.g., brackets and braces), creating a self-explanatory structured output.

The paradigm shift from classification to generation in SRL (DSRL) represents a fundamental change in how the task is conceptualized and executed. This move suggests that SRL is not merely about assigning pre-defined, symbolic labels but about explaining the semantic relationships in a human-readable and interpretable manner. This aligns perfectly with the broader trend in NLP towards explainable AI (XAI) and more intuitive model outputs. By generating definitions, the model is implicitly compelled to learn a deeper, more compositional, and nuanced understanding of semantic roles, moving beyond rote memorization of labels. This generative approach also naturally addresses the challenge of incorporating the "meaningful definitions" of semantic roles that traditional symbolic methods often ignored[10]. This generative paradigm could lead to more flexible and robust SRL systems, particularly for handling unseen or low-frequency predicates, as it leverages the model's ability to compose natural language descriptions rather than simply selecting from a fixed, finite set of labels. It also opens exciting new avenues for more interactive and interpretable SRL systems, where users can not only see the assigned roles but also understand the underlying semantic rationale in plain language. This could be particularly beneficial for debugging models, educating users, and facilitating human-AI collaboration in complex NLU tasks. The paper "Semantic Role Labeling Meets Definition Modeling" demonstrated that this generative approach achieved competitive performance compared to existing state-of-the-art methods across various benchmarks. Furthermore, it showed excellent generalization capabilities across different semantic resources and in joint training scenarios.

2.4. Leveraging Syntactic Information

2.4.1. SRL as Dependency Parsing

A key empirical analysis revealed a strong correlation between syntactic and semantic structures: over 98% of PropBank-style SRL annotations for both English and Chinese data can be mapped to three common syntactic dependency patterns. These patterns include: D-mode (predicate directly governs argument), C-mode (predicate and argument share a common syntactic parent), and R-mode (argument directly governs predicate, often seen in relative clauses)[25]. Motivated by this observation, a novel approach transforms the traditional span-based SRL task into a syntactic dependency parsing task[26]. This involves designing "joint labels" that simultaneously encode both syntactic relations and the three identified semantic patterns. A "back-and-forth conversion" algorithm is then employed, allowing for highly accurate recovery of the original SRL annotations from the parsed dependency tree structures.

The discovery that a vast majority of semantic roles align with specific, recurring syntactic dependency patterns is a profound observation regarding the intrinsic connection between syntax and semantics. This suggests that syntax is not merely a superficial layer but provides a fundamental, structural scaffold upon which semantic relationships are built. By casting SRL as a dependency parsing problem, researchers are effectively leveraging well-established and efficient syntactic parsing techniques to solve a semantic task. This approach can reduce the overall complexity of the SRL problem by exploiting the inherent structural regularities of language, rather than treating semantic role assignment as an entirely independent problem. This re-formulation could lead to more efficient, accurate, and linguistically informed SRL systems[27], particularly for languages where robust syntactic parsers are readily available. It also reinforces the critical importance of interdisciplinary research at the intersection of syntax and semantics, suggesting that a deeper theoretical understanding of their interplay can yield significant practical advancements in NLU. The paper "Semantic Role Labeling as Syntactic

Dependency Parsing" demonstrated impressive reconstruction accuracy (99% F1 for English and over 97% for Chinese) and competitive F1 scores on benchmark datasets, validating the effectiveness of this re-formulation.

2.4.2. Supertags for Syntax-Aware SRL

While full syntactic parsing provides rich structural information, its complexity can lead to increased computational cost and susceptibility to parsing errors, which can propagate to SRL. Conversely, completely ignoring syntactic information can limit the performance of SRL models in certain contexts. To strike a balance, the study "Syntax-aware Neural Semantic Role Labeling with Supertags" introduces "supertags" as an intermediate, lightweight solution. Supertags are fine-grained lexical tags that encode crucial local syntactic information, such as the relationship between a lexical item and its parent or dependent words, without requiring a complete, explicit syntactic parse tree. These supertags are predicted by a BiLSTM network and then integrated as embeddings into the main SRL model, alongside word and part-of-speech embeddings[14].

The adoption of supertags represents a pragmatic and intelligent solution to the "syntax dilemma" in SRL. It acknowledges the undeniable value of syntactic information but seeks to integrate it in a manner that is less brittle, more computationally efficient, and robust to potential errors from full parsers. This reflects a broader pattern in NLP research: identifying and extracting the minimal yet most informative linguistic signals from complex structures that can be effectively learned and leveraged by neural models. It is about achieving a high signal-to-noise ratio in syntactic feature integration. This approach suggests a future where SRL models might move away from external, pre-computed full parses. Instead, they could learn to implicitly or explicitly extract only the most beneficial, high-signal syntactic cues for semantic role assignment, potentially through multi-task learning or specialized architectural components that co-learn syntactic and semantic representations. This could lead to more streamlined and efficient end-to-end SRL systems. This method achieved significant performance improvements on the CoNLL 2009 English and Spanish datasets[28] , demonstrating that supertags effectively bridge the gap between full syntactic reliance and complete syntax-agnosticism.

3. Addressing Key Challenges and Expanding Scope in SRL

3.1. Nominal and Non-Verbal Predicate Labeling

Historically, SRL research predominantly focused on verbal predicates, leaving nominal

(NSRL), adjectival, and adverbial predicates significantly under-researched. This imbalance is problematic because non-verbal predicates are highly prevalent and informative in real-world texts, such as news headlines, blogs, and social media posts[28]. Current SRL systems often struggle with these non-verbal forms and exhibit limited knowledge transfer capabilities between different predicate types. For instance, the CN-22 system demonstrated a drastic F1 score drop from 84.7 for verbs to 16.4 for nouns and 5.4 for adjectives on OntoNotes.

Solutions to this challenge include:

- **Development of New Resources:** A key solution involves constructing large-scale, dedicated noun predicate framework libraries, such as NounAtlas. The objective is to fill existing resource gaps and facilitate the integration of noun and verb predicates into a unified semantic framework, thereby improving overall SRL performance[28]. NounAtlas achieved this by extracting event- or state-describing nouns from WordNet, mapping them to existing VerbAtlas frames, and leveraging Large Language Models (LLMs) like ChatGPT for predicate nominalization and semantic role propagation to generate a substantial dataset of approximately 28,000 sentences.
- **Dataset Expansion and Challenge Sets:** Researchers are creating extended datasets (e.g., PB-Examples, PB-Unseen) to broaden the coverage of non-verbal predicates. Additionally, manually annotated challenge sets like Challenge-SRL are being developed to rigorously evaluate the generalization ability of SRL systems across diverse predicate types and unseen predicates[28].
- **Leveraging Linguistic Resources:** Exploring and integrating existing linguistic resources like FrameNet and VerbAtlas has shown promise in enhancing non-verbal SRL. Consistent semantic role definitions across different frames in these resources can significantly aid systems in generalizing to new patterns.
- **Word Sense Disambiguation (WSD) Integration:** Combining WSD with SRL has proven effective in improving predicate sense disambiguation, particularly in low-resource scenarios. WSD systems can, in fact, outperform SRL systems in this specific subtask, highlighting a beneficial synergy.

The initial focus on verbs in SRL was a pragmatic simplification. The subsequent research into nominal, adjectival, and adverbial predicates reveals that the concept of "predication" in natural language is far more expansive and complex than initially assumed. This expansion is not merely an academic exercise; it is critical for developing SRL systems that are truly applicable to real-world text, where non-verbal predicates are ubiquitous. The consistent challenge of limited knowledge transfer between predicate types and the need for entirely new datasets underscore that the semantic structures associated with non-verbal predicates are sufficiently distinct to pose significant technical hurdles. A crucial factor in addressing the data scarcity for these new predicate types has been the innovative use of Large Language Models (LLMs) for semi-automated data generation (as seen with NounAtlas). This demonstrates LLMs' potential to accelerate resource creation in under-resourced areas of NLP. Future SRL research must increasingly adopt a holistic view of predication, accounting for the full spectrum of predicate types across different parts of speech. This will likely necessitate the development of more flexible and abstract semantic representations that can generalize across various lexical categories, possibly by emphasizing proto-roles or definition-based approaches that are less tied to specific grammatical forms. The success of LLMs in data generation for NSRL also points to a future where LLMs could play a central role in rapidly expanding annotated resources for other complex or low-resource SRL subtasks.

3.2.Cross-Lingual and Low-Resource SRL

Cross-lingual SRL faces significant hurdles due to the scarcity of annotated training data in most languages, especially low-resource ones. Traditional methods often rely on noisy external tools like word alignment, machine translation, or pre-processing tools, which introduce errors and offer limited support for languages with scarce resources. Furthermore, inherent linguistic differences, such as varying word order across languages, can significantly impact annotation performance and model transferability.

Solutions to these problems include:

- **Alignment-Free Models:** A key advancement involves developing models, often based on LSTMs, that incorporate semantic role compressors and multilingual word embeddings. These models reduce the reliance on external, noisy tools and only require annotations in a source language along with access to raw parallel corpora for cross-lingual transfer[29].
- **Unified Cross-Lingual Models:** Creating comprehensive models with universal sentence and predicate-argument encoders, coupled with language-specific decoders, allows for effective SRL across diverse linguistic formalisms and multiple languages. Examples include InVeRo-XL and the unified model proposed in "Unifying Cross-Lingual Semantic Role Labeling"[29].
- **Zero-Shot Learning:** Advanced approaches, such as those in "Zero-shot Cross-lingual Conversational Semantic Role Labeling", employ hierarchical encoders and specifically

designed pre-training objectives. These objectives (e.g., Latent Space Alignment, Hard Parallel Sentence Identification, Speaker Role Identification, Utterance Order Rearrangement, Semantic Argument Identification) implicitly learn language-agnostic, conversation-structure-aware, and semantically rich representations, enabling SRL in target languages without any direct annotations[9].

- **Transfer Learning:** Leveraging the power of pre-trained multilingual models (e.g., XLM-RoBERTa) and applying transfer learning techniques has proven effective in adapting SRL systems to new, low-resource languages, as demonstrated in Transformer-based Swedish SRL.
- **New Annotated Datasets:** A crucial step in addressing resource scarcity is the creation and release of new, high-quality human-annotated datasets for low-resource languages, as seen with the Chinese and German datasets introduced in "Alignment-free Cross-lingual Semantic Role Labeling".

The consistent and prominent focus on cross-lingual and zero-shot SRL across the research landscape underscores NLP's overarching global ambition. The reality is that while language is universal, annotated data is not, creating a significant bottleneck for widespread adoption of NLP technologies. The persistent challenges of data scarcity and the limitations of traditional transfer methods (e.g., machine translation-based approaches) highlight this fundamental resource imbalance. The development of "alignment-free" and "zero-shot" methods, along with the creation of "universal" encoders, represents a strategic and necessary shift towards methodologies that are inherently more robust to data limitations across diverse languages. However, the observed variations in performance across languages (e.g., French performing better than Chinese in "Alignment-free Cross-lingual SRL”) indicate that deep linguistic typology (e.g., fundamental differences in word order or grammatical structures) still poses a significant challenge. This suggests that a truly "universal" model remains a goal, not a fully realized state, and that language-specific nuances cannot be entirely overlooked. This area will undoubtedly remain a critical research frontier. Success in low-resource and cross-lingual SRL will be transformative, democratizing access to advanced NLP technologies for a much wider range of global languages. Future research will likely continue to push the boundaries of zero-shot and few-shot learning, potentially leveraging the immense pre-training capabilities of Large Language Models (LLMs) to create more inherently language-agnostic SRL systems. However, it will also need to incorporate more sophisticated typological awareness in model design or annotation strategies to effectively handle the remaining challenges posed by deep linguistic diversity.

3.3. Integration with Related Semantic Tasks

SRL's capacity to provide a structured "who-did-what-to-whom" understanding renders it a versatile and foundational component for enhancing various other NLP tasks.

- **Emotional Semantic Role Labeling (SRL4E):** This specialized task aims to identify emotion triggers (CUE), experiencers, targets, and stimuli within text. A unified evaluation framework (SRL4E) has been developed to integrate heterogeneous emotion datasets under a consistent annotation format, enabling more reliable evaluation of pre-trained language models and facilitating cross-lingual transfer in emotion analysis[30].
- **Opinion Role Labeling (ORL):** SRL information can significantly enhance the performance of ORL, which focuses on identifying opinion holders and targets for a given opinion trigger. This is achieved by learning semantic-aware word representations from SRL, as predicate agents and patients frequently correspond to opinion holders and targets, making SRL a valuable source of prior semantic knowledge[12].
- **Word Sense Disambiguation (WSD):** Combining WSD with SRL can lead to substantial improvements in predicate sense disambiguation, particularly in low-resource settings. Research indicates that WSD systems can even outperform SRL systems in this specific subtask, highlighting a beneficial synergy between the two.
- **Event Argument Extraction (EAE):** SRL annotations can be effectively transferred to EAE tasks. This is achieved by conceptualizing EAE as a "role querying" problem and employing template-based slot querying strategies to mitigate discrepancies in role labels and handle distant arguments between the two tasks[11].
- **Out-of-Distribution (OOD) Detection:** SRL can guide OOD detection by enabling the extraction of fine-grained local features from arguments within a sentence. This process helps in filtering out noise (e.g., anomalous phrases) and enhances global-local feature learning through self-supervised prediction of SRL-extracted roles, leading to more accurate identification of subtle OOD patterns[13].

The consistent and successful application of SRL as a foundational or enhancing component for a diverse array of NLP tasks (emotion analysis, opinion mining, event extraction, OOD detection) powerfully demonstrates its versatility and fundamental nature within Natural Language Understanding. SRL provides a structured, canonical representation of "who did what to whom," which is inherently transferable and highly valuable across various semantic tasks. This indicates that SRL is not merely an isolated, end-goal task but rather a crucial intermediate representation that can be leveraged to build more sophisticated and robust NLU systems. The acknowledgment of "discrepancies" between SRL and these downstream tasks (e.g., ORL's use of non-standard predicates) further highlights the importance of tailored adaptation and specialized transfer learning techniques to maximize this synergy. SRL will undoubtedly continue to serve as a core component in multi-task learning frameworks, effectively acting as a "semantic backbone" for a wide range of downstream applications. Future research in this area will likely focus on developing more sophisticated and adaptive methods for leveraging and fine-tuning SRL models for the precise requirements of specific tasks, potentially through advanced shared representations, meta-learning approaches, or even by designing new tasks that explicitly build upon SRL's structured output.

3.4. Annotation Consistency and Data Heterogeneity

A persistent and significant challenge in SRL research stems from the inconsistencies in annotation strategies across different datasets. Existing SRL datasets often exhibit heterogeneity in terms of size, domain, specific labels used, and emotion categories, which severely hinders unified evaluation, effective model development, and cross-dataset generalization. This issue is particularly pronounced in the context of Korean SRL datasets, where variations between resources like Korean PropBank and NIKL SRL dataset pose integration difficulties[27].

Solutions to improve annotation consistency and address data heterogeneity include:

- **Unified Evaluation Frameworks:** One approach to mitigate heterogeneity is the creation of unified evaluation frameworks, such as SRL4E. These frameworks integrate multiple heterogeneous datasets under a common, consistent labeling scheme, enabling more standardized comparisons and facilitating broader model development.
- **Linguistically-Informed Annotation Strategies:** To improve consistency and accuracy, researchers are developing annotation strategies that are deeply grounded in linguistic theories and explicitly account for language-specific properties. For instance, in Korean, this involves precise definitions of arguments and modifiers, and careful consideration of unique features like Subject-Object-Verb (SOV) word order and the crucial role of postpositions[27].
- **Conversion Methods:** Providing robust methods to convert existing linguistic resources (e.g., the Sejong Verb Dictionary) into standard SRL dataset formats (e.g., CoNLL-style)

is vital. This increases the availability of consistent, high-quality data, which is essential for training and evaluating models effectively[27].

Despite significant advancements in SRL modeling techniques and architectures, the underlying quality, consistency, and standardization of annotated data remain a critical and often underestimated challenge. The recurring theme of heterogeneity across datasets and the explicit need for linguistically-informed annotation strategies (as highlighted by the Korean SRL example) indicate that data creation is far from a solved problem. This situation embodies a classic "garbage in, garbage out" principle: even the most sophisticated neural models will struggle to achieve optimal performance and generalization if the training data is inconsistent, ambiguous, or poorly defined. The continuous effort to unify datasets and develop rigorous annotation guidelines underscores the fundamental nature of this persistent problem in NLP. Future progress in SRL, particularly for less-resourced languages, specialized domains, or emerging linguistic phenomena (like conversational SRL), will heavily depend on the development of more robust, scalable, and linguistically sound annotation methodologies. This might involve exploring advanced techniques such as active learning, human-in-the-loop annotation systems, or leveraging Large Language Models (LLMs) for efficient semi-automated annotation with subsequent expert human validation. The emphasis will shift towards ensuring that the data itself is a reliable and consistent representation of linguistic reality.

4. Datasets and Evaluation Benchmarks

4.1.Overview of Major SRL Datasets

The landscape of SRL research is heavily reliant on a diverse set of annotated datasets, each contributing to different aspects of the task.

- **PropBank:** A cornerstone resource, PropBank augments the Penn Treebank (comprising Wall Street Journal texts) with manually created semantic role annotations. It serves as a foundational training dataset for a multitude of automatic SRL systems. PropBank-style annotations are prominently featured in various CoNLL shared tasks, including CoNLL-2005, CoNLL-2009, and CoNLL-2012[21].
- **FrameNet:** This computational lexicon offers a systematic description of predicates (referred to as "frames") and their associated roles (known as "frame elements")[22]. It is utilized in research exploring non-verbal predicates and in definition-based SRL approaches.

- **VerbNet:** Another significant lexical resource, VerbNet provides detailed information on argument patterns for various verbs. It is frequently used in conjunction with PropBank, with SEMLINK serving as a crucial resource that provides mappings and compatibility constraints between their respective label sets.
- **CoNLL Shared Tasks:** These annual shared tasks have played a pivotal role in standardizing SRL evaluation and fostering comparative research.
- **CoNLL-2005:** A key benchmark for span-based SRL[14].
- **CoNLL-2009:** A prominent multilingual benchmark, covering 7 languages including English, German, Spanish, Chinese, Czech, and Catalan, primarily focusing on dependency-based SRL[10].
- **CoNLL-2012:** Another widely used benchmark, primarily for span-based SRL, often built upon the OntoNotes 5.0 corpus[21].
- **Specialized Datasets:**
  - **NounAtlas:** The first large-scale noun predicate framework library, comprising approximately 28,000 sentences, specifically designed to address the gap in nominal SRL.
  - **PB-Examples & PB-Unseen:** Extended PropBank-based datasets developed to expand coverage of non-verbal predicates. PB-Unseen is particularly challenging as it contains over 4000 predicate frames not present in OntoNotes.
  - **Challenge-SRL:** A manually annotated test set of 288 sentences, specifically created to evaluate model generalization across different and unseen predicate types.
  - **SPR1 & SPR2:** Datasets for Semantic Proto-role Labeling (SPRL), derived from the Wall Street Journal and English Web Treebank respectively, annotated with binary proto-role attributes[14].
  - **SRL4E Unified Datasets:** A framework integrating 6 heterogeneous emotion datasets (e.g., Blogs, Elections, EmoTweet) to provide a consistent annotation format for emotional SRL.
  - **UPB (Universal Proposition Bank):** Used extensively in cross-lingual SRL studies, covering languages such as German, French, and Portuguese, and serving as a benchmark for zero-shot transfer across 23 target languages[15].

- **DuConv, PersonaChat, CMU-DoG:** Conversational datasets utilized for developing and evaluating cross-lingual conversational SRL systems.
- **MPQA 2.0:** A standard benchmark dataset for fine-grained opinion mining and Opinion Role Labeling (ORL)[12].
- **Sejong Verb Dictionary:** This resource serves as a source for developing linguistically-informed Korean SRL datasets through a conversion process to a CoNLL-style format[27].

Table 1. Overview of Major SRL Datasets and Benchmarks

Dataset Name	Primary Focus/SRL Type	Language(s)	Annotation Style	Source/ Origin	Key Characteristics/Size
PropBank	Verbal SRL	English	Span-based, BIO	Penn Treebank (WSJ)	Foundational, widely used for training SRL systems
FrameNet	Frame Semantics	English	Span-based	Lexical resource	Describes predicates (frames) and roles (frame elements)[22]
VerbNet	Verbal Argument Patterns	English	-	Lexical resource	Provides argument patterns for verbs, often used with PropBank via SEMLINK[8]
CoNLL-2005	General SRL	English	Span-based	Benchmark	Standard benchmark for span-based SRL[21]
CoNLL-2009	Multilingual SRL	7 languages (EN, DE, ES, ZH, CS, CA, JP)	Dependency-based	Benchmark	Prominent multilingual benchmark across 7 languages
CoNLL-2012	General SRL	6 languages (EN, ZH, CS, DE, ES, CA)	Span-based	OntoNotes 5.0	Widely used benchmark, often built on OntoNotes[22]
NounAtlas	Nominal SRL (NSRL)	English	-	WordNet, VerbAtlas, SemCor	First large-scale noun predicate framework, ~28,000 sentences[22] 4
PB-Examples	Non-Verbal Predicates	English	-	PropBank extended	Covers verb, noun, adjective predicates, 7481 framesets
PB-Unseen	Non-Verbal Predicates	English	-	PropBank extended	Challenging test set, >4000 unseen predicate frames

Dataset Name	Primary Focus/SRL Type	Language(s)	Annotation Style	Source/ Origin	Key Characteristics/Size
Challenge-SRL	Non-Verbal Predicates	English	Manual Annotation	Custom	Manually annotated, 288 sentences, for generalization testing
SPR1	Semantic Proto-roles	English	Binary Attributes	Wall Street Journal	4912 sentences, 18 attributes [22]
SPR2	Semantic Proto-roles	English	Binary Attributes	English Web Treebank	2758 sentences, 14 attributes [8]
SRL4E	Emotional SRL	Multilingual	BIO	6 integrated emotion datasets	Unified framework for emotional SRL[8]
UPB	Cross-lingual SRL	Multilingual	-	Universal Proposition Bank	Used for zero-shot transfer across 23 target languages[8]
DuConv	Conversational SRL	Chinese	-	-	Training/dev/in-domain test for CSRL
PersonaChat, CMU-DoG	Conversational SRL	English	-	Custom	Used for cross-lingual CSRL evaluation
MPQA 2.0	Opinion Role Labeling	English	-	-	Standard benchmark for fine-grained opinion mining [21]
Sejong Verb Dictionary	Korean SRL (source)	Korean	-	Lexical resource	Source for linguistically-informed Korean SRL datasets[21]

4.2. Standard Evaluation Metrics

The performance of SRL systems is rigorously evaluated using a set of standard metrics to ensure comparability and track progress across different models and approaches.

- **F1 Score:** This is the most widely adopted metric in SRL, representing the harmonic mean of precision and recall. It is extensively used to evaluate overall argument annotation performance, as well as specific subtasks such as emotion classification and predicate sense disambiguation[10].
    - **Strict F1:** A more stringent variant that demands correct prediction of the predicate, all arguments, and their associated attributes (e.g., in Semantic Proto-role Labeling).

  - **Relaxed F1:** A less strict variant that permits incorrect predicates or arguments, evaluating only the correctly predicted parts. This metric is valuable for analyzing partial correctness and identifying areas for improvement.
- **LAS (Labeled Attachment Score):** Primarily employed in dependency-based SRL, LAS quantifies the accuracy of semantic dependency recovery by assessing whether both the head of the argument and its semantic label are correctly predicted.
- **Precision and Recall:** These metrics are often reported individually alongside the F1 score, particularly when evaluating annotation strategies or specific aspects of model performance, providing a more granular view of system strengths and weaknesses[27].
- **ρ (rho):** This specific metric quantifies the proportion of predictions that violate SEMLINK constraints. It is used to assess the consistency of labels when integrating different semantic resources like PropBank and VerbNet, ensuring that the combined output adheres to established semantic relationships.
- **Accuracy:** This metric is employed for specific subtasks, such as evaluating the performance of automatic mapping processes (e.g., Top-1 and Top-5 accuracy for NounAtlas mapping).

5.Future Directions and Open Problems

5.1. Towards End-to-End and Document-Level SRL

Despite significant progress, many current SRL models still operate under simplifying assumptions. They often rely on pre-identified or manually annotated predicates, rather than automatically detecting them. Furthermore, most models primarily process sentences in isolation, failing to fully consider broader contextual information, such as cross-sentence dependencies or discourse-level relationships. This limitation significantly impacts their performance when dealing with long texts or complex narratives where semantic roles can span multiple sentences[2].

Future work in this area is directed towards:

- **End-to-End SRL Parsing:** A critical future direction is to develop truly end-to-end SRL systems that eliminate the reliance on manually annotated predicates. This involves models capable of automatically identifying predicate locations and their specific attributes within a text, achieving full automation from raw text input to semantic role output.

- **Document-Level Context Modeling:** Future research will increasingly focus on incorporating broader contextual information. This includes modeling the influence of preceding and succeeding text, as well as discourse-level phenomena, on semantic role assignment. This is a natural and necessary progression from sentence-level SRL, acknowledging that semantic roles often extend beyond the strict boundaries of individual sentences[2].
- **Long Text Processing:** To address the efficiency limitations imposed by the maximum sequence length constraints of many pre-trained language models (PLMs), future work will explore advanced mechanisms such as sliding windows, hierarchical attention mechanisms, or novel long-context Transformer architectures. These techniques aim to improve the efficiency and effectiveness of SRL for processing very long texts.

The persistent push towards end-to-end and document-level SRL signifies a fundamental shift in the field's ambition: moving from localized, fragmented sentence-level understanding to a more holistic, comprehensive interpretation of entire texts and discourses. Relying on "gold" predicates or processing sentences in isolation are simplifying assumptions that do not reflect the complex, interconnected nature of real-world language use. This indicates that SRL is maturing, pushing beyond its foundational single-sentence scope to tackle the inherent complexities of discourse, narrative, and multi-sentence comprehension. While not explicitly detailed in all provided materials, the very existence of "Semantic Role Labeling Graph Reasoning Network" and the emphasis on "High-order Semantic Role Labeling" implicitly support this trajectory by seeking to capture broader structural and relational understanding beyond simple pairwise dependencies. Future SRL systems are likely to be integrated into larger, more ambitious NLP pipelines focused on discourse parsing, knowledge graph construction, or complex reasoning tasks, where understanding cross-sentence semantic relationships and coreference is paramount. This will necessitate the development of novel architectural designs capable of handling long-range dependencies, complex anaphoric relations, and the dynamic evolution of semantic roles across a document.

5.2 Enhancing Performance for Low-Resource Languages

Despite notable advancements in cross-lingual SRL, a significant and persistent challenge remains the performance disparity in low-resource languages, particularly non-Indo-European languages. This is primarily due to the severe scarcity of annotated training data and the unique linguistic characteristics (e.g., complex morphology, free word order) that differ substantially from high-resource languages. Furthermore, specific linguistic phenomena, such as the handling of zero postpositions or special particles in Korean, can pose unique annotation and modeling challenges that current universal methods may not fully address.

Future work in this domain will focus on:

- **Data Augmentation and Self-Supervised Learning:** A key focus will be on combining more parallel corpora with advanced self-supervised learning techniques. The goal is to leverage vast amounts of unannotated text to generate or learn robust representations that can effectively enhance performance in low-resource settings without requiring expensive manual annotations.
- **Addressing Linguistic Specificities:** Future research must delve deeper into addressing specific linguistic challenges unique to various low-resource languages. This includes developing more nuanced models or annotation strategies that can correctly handle phenomena like zero postpositions or complex morphological inflections, which the current conversion methods might not fully capture.
- **Expansion to Broader Semantic Ontologies:** Expanding SRL models to integrate with a wider range of semantic ontologies, such as FrameNet and Abstract Meaning Representation (AMR), is crucial. This could improve cross-lingual generalization by providing richer, more abstract semantic representations that are less tied to specific lexical or syntactic forms.

The recurring emphasis on low-resource languages highlights a persistent and critical inequality in NLP research and application, where high-resource languages (predominantly English) disproportionately benefit from data-hungry models. The recognition of specific linguistic challenges (like Korean postpositions) underscores that a simplistic "one-size-fits-all" universal model might not be sufficient, and that typological awareness remains crucial for truly robust cross-lingual transfer. The strategic push for more parallel corpora and self-supervised learning indicates a clear direction to generate or leverage data more efficiently, reducing the reliance on costly manual annotation. This area will remain a vital research frontier, as success here is key to democratizing NLP technologies and making SRL accessible and effective for a wider range of global languages. This might involve the development of more sophisticated meta-learning, domain adaptation, or transfer learning techniques that can generalize effectively from minimal examples or exploit structural and semantic similarities across different language families. Ultimately, advancements in this area will broaden the global impact and utility of SRL.

5.3. The Role of SRL in Large Language Models (LLMs)

The research consistently highlights the increasing interaction between SRL and Large Language Models (LLMs). Studies like "NounAtlas" demonstrate the utility of LLMs (e.g., ChatGPT) for generating new training data, while "Semantic Role Labeling Meets Definition Modeling" showcases LLMs (e.g., BART-large) as core architectural components for redefining SRL as a natural language generation task. Furthermore, broader literature explicitly discusses the evolving role of SRL in the age of LLMs and its potential impact on the wider NLP landscape.

Current interactions between SRL and LLMs include:

- **LLMs for Data Generation:** LLMs are proving to be powerful tools for semi-automated data creation, assisting in tasks like converting verb predicates to noun predicates to expand SRL datasets, as seen with NounAtlas. This significantly reduces the manual annotation burden.
- **LLMs as Core SRL Models:** Pre-trained language models (PLMs) such as BERT, RoBERTa, XLM-RoBERTa, and BART-large already form the backbone of many state-of-the-art SRL systems, providing highly contextualized and semantically rich embeddings.
- **LLMs for Definition-based SRL:** LLMs' advanced natural language generation capabilities enable the transformation of SRL from a classification task into an NLG task, where models generate human-readable descriptions of predicate-argument structures.

The rise of LLMs represents a profound paradigm shift across the entire NLP landscape, and SRL is not merely an application benefiting from this; it is being fundamentally transformed by it. Initially, LLMs might have been viewed simply as "better encoders" for existing SRL architectures. However, their use for data generation (e.g., NounAtlas) and, more strikingly, for redefining the SRL task itself (e.g., DSRL as NLG) indicates a deeper, symbiotic relationship. LLMs are not just consuming SRL; they are actively transforming how SRL is conceived, implemented, and what it can ultimately achieve. This moves beyond incremental performance improvements to fundamental changes in methodology and potential. The ability of LLMs to "reason" with and generate natural language definitions of semantic roles (as in DSRL) suggests that they are internalizing semantic concepts in a more abstract and compositional way than traditional, discrete-label models could. This implies a potential for more robust generalization and a deeper, human-like understanding of meaning.

The future role and impact of LLMs on SRL are anticipated to be significant:

- **Enhanced Semantic Understanding:** LLMs, with their vast pre-trained knowledge and emergent capabilities, hold the potential to learn and internalize complex semantic role patterns more effectively and robustly than previous models.
- **Zero-Shot/Few-Shot SRL:** The strong generalization capabilities of LLMs could enable high-performance SRL in new domains or for new languages with minimal or even no labeled training data, significantly reducing annotation costs.
- **Interpretability and Explainability:** Generative SRL (DSRL) powered by LLMs could provide more interpretable outputs, explaining semantic roles in natural language, thereby enhancing trust and understanding of model decisions.
- **Integration with Knowledge Graphs:** LLMs could facilitate the extraction of structured semantic information (via SRL) at scale, enabling the automated population and expansion of knowledge graphs, which are crucial for advanced AI applications.
- **Challenges:** While LLMs offer immense potential, their application in SRL also presents challenges, including their substantial computational cost, the potential for generating "hallucinated" or incorrect outputs, and the ongoing need to ensure that their internal "understanding" of semantic roles aligns precisely with established linguistic theories and human intuition.

SRL research in the LLM era will likely focus on several key areas:

1. **Prompt Engineering for SRL:** Developing optimal strategies for prompting LLMs to perform SRL tasks effectively, including few-shot and zero-shot learning scenarios.
2. **Evaluating LLM's Implicit SRL Capabilities:** Systematically assessing how well LLMs inherently understand predicate-argument structures and semantic roles without explicit, fine-grained SRL training. This could reveal emergent semantic understanding.
3. **Hybrid Models:** Designing novel hybrid architectures that combine the strengths of LLMs (e.g., broad knowledge, contextual understanding, generation capabilities) with the precision and structural constraints of traditional SRL techniques.
4. **SRL for LLM Interpretability:** Utilizing SRL as a tool to analyze and explain the internal workings, decision-making processes, or generated outputs of LLMs, contributing to the broader field of XAI.
5. **Data Curation with LLMs:** Leveraging LLMs for highly efficient, high-quality annotation and validation of SRL datasets, which could dramatically reduce annotation costs and accelerate research in under-resourced areas.

6. Conclusion

Semantic Role Labeling has undergone a significant evolution from its foundational concepts, progressing through sophisticated neural architectures, incorporating higher-order modeling, and extending its capabilities to cross-lingual and low-resource scenarios. The field has notably expanded beyond its initial verb-centric focus to embrace a broader range of predicate types, including nominal and non-verbal forms. Furthermore, SRL is increasingly integrated with other critical semantic tasks, serving as a foundational layer for more complex Natural Language Understanding applications.

Despite these advancements, several challenges persist. These include the enduring issue of data scarcity, particularly for low-resource languages, which impedes the development of robust models. Annotation inconsistencies and data heterogeneity across various datasets continue to pose obstacles to unified evaluation and model generalization. Additionally, achieving truly end-to-end SRL and extending semantic understanding to the document level, beyond individual sentences, remains an active area of research.

The future of SRL is poised for transformative changes, largely driven by the rapid advancements in Large Language Models (LLMs). LLMs are not merely enhancing existing SRL methodologies but are fundamentally reshaping how SRL is conceived and implemented, particularly through their capabilities in data generation and natural language-based semantic representation. This symbiotic relationship promises to push SRL towards more holistic, interpretable, and globally accessible semantic understanding. As a core component of Natural Language Understanding, SRL remains a dynamic and crucial field at the heart of advancing artificial intelligence's ability to comprehend human language.

References

[1] D. Gildea and D. Jurafsky, "Automatic labeling of semantic roles," Computational linguistics, vol. 28, no. 3, pp. 245-288, 2002.

[2] H. Chen et al., "Semantic Role Labeling: A Systematical Survey," arXiv preprint arXiv:2502.08660, 2025.

[3] S. Conia, R. Orlando, F. Brignone, F. Cecconi, and R. Navigli, "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, 2021: Association for Computational Linguistics, pp. 319-328.

[4] S. Kurita, H. Ouchi, K. Inui, and S. Sekine, "Iterative Span Selection: Self-Emergence of Resolving Orders in Semantic Role Labeling," in Proceedings of the 29th International Conference on Computational Linguistics, 2022, pp. 5383-5397.

[5] T. Shi, I. Malioutov, and O. Irsoy, "Semantic role labeling as syntactic dependency parsing," arXiv preprint arXiv:2010.11170, 2020.

[6] J. Kasai, D. Friedman, R. Frank, D. Radev, and O. Rambow, "Syntax-aware neural semantic role labeling with supertags," arXiv preprint arXiv:1903.05260, 2019.

[7] J. Chen, X. He, and Y. Miyao, "Modeling syntactic-semantic dependency correlations in semantic role labeling using mixture models," in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 7959-7969.

[8] N. Wang et al., "An mrc framework for semantic role labeling," arXiv preprint arXiv:2109.06660, 2021.

[9] H. Wu, H. Tan, K. Xu, S. Liu, L. Wu, and L. Song, "Zero-shot cross-lingual conversational semantic role labeling," arXiv preprint arXiv:2204.04914, 2022.

[10] L. Zhang, I. Jindal, and Y. Li, "Label definitions improve semantic role labeling," in Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2022, pp. 5613-5620.

[11] Z. Zhang, E. Strubell, and E. Hovy, "Transfer learning from semantic role labeling to event argument extraction with template-based slot querying," in Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 2627-2647.

[12] M. Zhang, P. Liang, and G. Fu, "Enhancing opinion role labeling with semantic-aware word representations from semantic role labeling," 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), 2019, pp. 641-646.

[13] J. Zou et al., "Semantic role labeling guided out-of-distribution detection," arXiv preprint arXiv:2305.18026, 2023.

[14] E. Spaulding, G. Kazantsev, and M. Dredze, "Joint end-to-end semantic proto-role labeling," in Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2023, pp. 723-736.

[15] R. Devianti and Y. Miyao, "Transferability of Syntax-Aware Graph Neural Networks in Zero-Shot Cross-Lingual Semantic Role Labeling," in Findings of the Association for Computational Linguistics: EMNLP 2024, 2024, pp. 20-42.

[16] T. Li, G. Kazeminejad, S. W. Brown, M. Palmer, and V. Srikumar, "Learning semantic role labeling from compatible label sequences," arXiv preprint arXiv:2305.14600, 2023.

[17] C. Guan, Y. Cheng, and H. Zhao, "Semantic role labeling with associated memory network," arXiv preprint arXiv:1908.02367, 2019.

[18] S. Conia, A. Bacciu, and R. Navigli, "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, 2021, pp. 338-351.

[19] S. Conia, E. Barba, A. Scirè, and R. Navigli, "Semantic role labeling meets definition modeling: Using natural language to describe predicate-argument structures," arXiv preprint arXiv:2212.01094, 2022.

[20] C. Zheng and P. Kordjamshidi, "SRLGRN: Semantic role labeling graph reasoning network," arXiv preprint arXiv:2010.03604, 2020.

[21] S. Zhou, Q. Xia, Z. Li, Y. Zhang, Y. Hong, and M. Zhang, "Fast and accurate end-to-end span-based semantic role labeling as word-based graph parsing," arXiv preprint arXiv:2112.02970, 2021.

[22] C. Ai and K. Tu, "Frame semantic role labeling using arbitrary-order conditional random fields," in Proceedings of the AAAI Conference on Artificial Intelligence, 2024, vol. 38, no. 16, pp. 17638-17646.

[23] Z. Li, H. Zhao, R. Wang, and K. Parnow, "High-order semantic role labeling," arXiv preprint arXiv:2010.04641, 2020.

[24] D. Dannélls, R. Johansson, and L. Y. Buhr, "Transformer-based Swedish Semantic Role Labeling through Transfer Learning," in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024, pp. 16762-16769.

[25] R. Orlando, S. Conia, and R. Navigli, "Exploring non-verbal predicates in semantic role labeling: Challenges and opportunities," arXiv preprint arXiv:2307.01870, 2023.

[26] Y. Zhang, Q. Xia, S. Zhou, Y. Jiang, G. Fu, and M. Zhang, "Semantic role labeling as dependency parsing: Exploring latent tree structures inside arguments," arXiv preprint arXiv:2110.06865, 2021.

[27] Y. Chen, K. Lim, and J. Park, "A Linguistically-Informed Annotation Strategy for Korean Semantic Role Labeling," in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024, pp. 733-738.

[28] R. Navigli, M. Pinto, P. Silvestri, D. Rotondi, S. Ciciliano, and A. Scirè, "NounAtlas: Filling the Gap in Nominal Semantic Role Labeling," in Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2024, pp. 16245-16258.

[29] R. Cai and M. Lapata, "Alignment-free cross-lingual semantic role labeling," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 3883-3894.

[30] C. Campagnano, S. Conia, and R. Navigli, "SRL4E–Semantic Role Labeling for Emotions: A unified evaluation framework," in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2022, pp. 4586-4601.