

Article

Extracting Composition Expression Patterns from Materials Science Patent Documents Using SEP-Tags

Toshihiko Sakai ^{1,*} , Nobuhiko Chiwata ²  and Tsunenori Mine ³ ¹ Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan² Digital Transformation Strategy Office, Proterial, Ltd., Tokyo 135-0061, Japan; nobuhiko.chiwata.qj@proterial.com³ Department of Advanced Information Technology, Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan; mine@ait.kyushu-u.ac.jp

* Correspondence: toshihiko.sakai.a46@s.kyushu-u.ac.jp

Abstract

Extracting composition expressions from materials science patent documents is essential for patent document searches. Composition expressions describing a single unit of elements and quantities (e.g., “Al: 0.02% or more and 0.08% or less”) tend to appear clustered together. In such cases, researchers in the field of materials science who conduct patent searches have found that boundary-indicating phrases are effective for searching. However, there are no concrete examples that have implemented this approach, and the validity of this approach has not been evaluated to date. In this paper, we propose a Separator Tag (SEP-tag) framework as an explicit boundary for composition expressions in named entity recognition labels. This allows the named entity recognition model to simultaneously perform entity recognition and pattern boundary learning within a single end-to-end process. Furthermore, we propose a four-axis evaluation framework that extends the conventional single-entity F1 score to evaluate named entity recognition models using SEP-tags. (1) Entity F1 score excluding structural tags, (2) Exact match rate for correct spans, (3) Predicted span pattern extraction F1 score, (4) Pattern extraction F1 score. We conducted evaluations using RoBERTa-base and BERT-base-Japanese on materials science patent datasets in English (10,166 sentences) and Japanese (975 sentences). Experimental results show that training the model with SEP-tags improved the span exact match rate on the English dataset by approximately 59.72 percentage points (from 15.95% to 75.67%), and reduced false positives in pattern extraction to 1/117 (F1 score: 0.0784 → 0.8503). In the Japanese dataset, false positives were reduced to 1/123 (F1 score: 0.0361 → 0.4877). For both languages, the entity F1 score was equivalent to that of the model without SEP-tags (English: $|\Delta F1| < 0.002$, Japanese: $|\Delta F1| < 0.001$), with no significant difference found for any of the 13 labels. These results demonstrate that explicit structural boundary tokens are highly effective for extracting composition expression patterns in domain-specific named entity recognition.



Academic Editor: Fabrizio Messina

Received: 26 April 2026

Revised: 16 June 2026

Accepted: 25 June 2026

Published: 3 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: named entity recognition; materials science; composition expression; patent documents; pattern extraction

1. Introduction

Materials informatics is emerging as a key field for accelerating materials discovery by extracting structured knowledge from the growing body of scientific and patent literature [1–3]. In particular, in the field of materials science, composition expressions

described by combinations of entity tokens such as element names, composition values, units, and constraints play a central role in describing the properties of new alloys and compounds [4,5]. Named entity recognition is a technique used to extract such entities from text [6–10]. In this study, we use named entity recognition to extract information on composition expressions from patent documents in the field of materials science. Composition expressions tend to be described in clusters, and they can be efficiently identified by utilizing words that mark the boundaries of these clusters. It is well-established among researchers in the field of materials science that words such as “comprising” and “wherein” are commonly used to mark these boundaries and typically appear immediately before the composition expressions [11]. Nevertheless, extracting complete composition expressions from patent documents remains challenging [12,13]. Although transformer-based models perform well at recognizing individual entities, they often fail to identify the correct boundaries of composition expressions consisting of multiple entities [14–16]. There are several types of composition expressions. In this study, we will refer to them as composition expression patterns.

Typical composition expressions in patents in the field of materials science include the following: “wherein the steel comprises Al: 0.02% to 0.08%, Si: 0.1% to 0.5%,”. A standard named entity recognition model correctly labels {Al, 0.02%, 0.08%} and {Si, 0.1%, 0.5%} as the element, lower limit, and upper limit, respectively. However, without explicit boundary information, it cannot group {Al, 0.02%, 0.08%} as one composition expression pattern and {Si, 0.1%, 0.5%} as another (See Section 4). This confusion regarding the boundary leads to a large number of false positives when attempting to enumerate composition patterns. Our research goal is to rapidly narrow down pattern candidates from a large volume of patent documents at low training and inference cost. The reason for this is that in practical patent searches, if search systems cannot return results quickly, search efficiency decreases. Therefore, we aim to extract composition expression patterns using a method that minimizes both training and inference costs as much as possible.

To address this issue, we propose and define the SEP-tag as a separator. The SEP-tag is a new label class introduced into named entity recognition labeling that allows for the explicit identification of the start and end boundaries of composition expression patterns. Specifically, following the conventions of BIO encoding, the start and end of a composition expression pattern are distinguished by a single SEP-tag. By treating the boundary SEP-tag as a single entity during training, the model can simultaneously learn token-level entity recognition and pattern-level segmentation in a single end-to-end process. Since the SEP-tag is a structural entity with properties different from conventional entities, its effectiveness cannot be fully evaluated using standard entity-level F1 scores. Therefore, this paper defines and evaluates a four-axis evaluation framework. This framework enables multifaceted evaluation of not only entity-level performance but also the extraction of composition expression patterns using SEP-tags. The first axis (Axis 1) evaluates the entity-level F1 score excluding structural tags. This is a general entity-level evaluation. The second axis (Axis 2) evaluates the correct span exact match rate, which measures the proportion of predicted entity sequences that exactly match the correct spans in true SEP-tag spans. The third axis (Axis 3) evaluates the F1 score for the extraction of composition expression patterns from predicted spans, which measures the precision and recall of the composition expression patterns derived from the predicted SEP-tag boundaries. Finally, the fourth axis (Axis 4) evaluates composition expression pattern extraction F1 score based solely on the boundaries predicted by models with and without SEP-tags, respectively.

The main contributions of this paper are as follows:

- We propose the SEP-tag which denotes the boundaries of composition expression patterns, to efficiently extract such patterns from materials science patent documents.

This method introduces explicit structural boundary tokens into named entity recognition training.

- We propose a four-axis evaluation framework that comprehensively assesses both entity recognition and pattern-level extraction quality. We demonstrate on English (10,166 sentences) and Japanese (975 sentences) materials science patent datasets that SEP-tag training achieves significant improvements in composition expression pattern extraction F1 score while preserving entity F1 score.
- We confirmed that phrases assigned SEP-tags are lexically distinct from composition expression pattern components. We also confirmed that there is no segmentation caused by SEP-tags in either the English or Japanese data.
- Based on the correlation between SEP-tag prediction quality and extraction accuracy of composition expression patterns, we confirmed that accurate prediction of SEP-tags directly leads to improved composition expression pattern extraction accuracy.

This paper is organized as follows. Section 2 provides an overview of related work. Section 3 describes our proposed method, the SEP-tag framework and evaluation design. Section 4 describes the experimental setup and results. Section 5 provides analysis and discussion. Section 6 concludes the paper.

2. Related Work

2.1. Named Entity Recognition Approaches

The BIO (Beginning–Inside–Outside) encoding scheme [17] is the de facto standard for sequence labeling in named entity recognition. Extensions such as BIOES (adding End and Single tags) have been shown to improve performance on fine-grained boundaries [18]. Span-based approaches [19,20] enumerate candidate spans and perform adjacency scoring, thereby avoiding the strict constraints of left-to-right sequential decoding. All of these methods focus on the boundaries of individual entities or spans. Therefore, they do not address the problem of grouping multiple entities into a single pattern.

LLM-Based Approaches

More recently, large language model (LLM)-based approaches have reframed named entity recognition as an open generative task, departing from fixed BIO tag sets entirely. Wang et al. [21] proposed a method for representing boundaries using special markers such as @@...## when having LLMs perform named entity recognition. Zhou et al. [22] proposed UniversalNER, which distills the entity recognition capability of ChatGPT into smaller models through mission-focused instruction tuning, achieving state-of-the-art performance across 43 datasets and 9 domains without domain-specific supervision.

The SEP-tag framework proposed in this study differs from these approaches in two key ways. First, while both BIO/BIOES and span-based methods focus on identifying individual entities or spans, SEP-tags address the problem of grouping co-occurring entities into a single composition expression pattern. This grouping problem cannot be solved by BIO/BIOES or span-based methods. Second, while span-based methods require dedicated span enumeration and scoring mechanisms, SEP-tag can be directly applied to BERT-based models simply by extending the label set. SEP-tag introduces label categories dedicated to structural boundaries without utilizing an LLM-based decoder structure. Unlike LLM-based methods, it does not reformulate NER as open-ended text generation. Since our method can be applied directly to BERT-based models, it is well-suited for specialized domains with limited labeled data.

2.2. Named Entity Recognition in Materials Science

2.2.1. Early and Pre-Trained Models

Named entity recognition has been widely applied to scientific literature to extract structured information regarding materials, processes, and properties [2]. Kim et al. [23] conducted early research on the rule-based extraction of composition-related entities from patent documents. Weston et al. [1] proposed an LSTM-based model to extract mentions of inorganic materials, sample descriptors, phase labels, material properties and applications, as well as the synthesis and characterization methods used. Subsequent research explored domain-adaptive pre-training. Gupta et al. [24] released MatSciBERT, a BERT model trained on materials science papers, which achieved state-of-the-art performance on multiple downstream tasks, including named entity recognition. Trewartha et al. [25] similarly proposed MatBERT, pre-trained on materials science literature, and demonstrated its effectiveness for downstream named entity recognition tasks. Mavracic et al. [26] proposed the automatic extraction of material properties using ontologies.

2.2.2. Recent Advances

More recently, Huang et al. [27] reformulated materials science named entity recognition as a machine reading comprehension task, achieving state-of-the-art performance on five public benchmarks by treating entity types as natural-language questions. Foppiano et al. [28] showed that even LLMs such as GPT-4 fail to match the zero-shot accuracy of fine-tuned BERT for materials science named entity recognition. Similarly, Potu et al. [29] compared in-context learning using LLMs with domain-specific language models such as MatSciBERT [24] and DeBERTa [30] in ontology-constrained named entity recognition. Their results showed that domain-specific language models outperform LLMs.

Despite advances ranging from domain-specific pre-training to LLM-based formulations, the majority of existing systems focus on recognizing individual entity tokens (such as elements, property values and units) rather than extracting grouped composition expression patterns. This distinction is crucial in patent analysis. Patents may describe several types of compositions, and correctly grouping related tokens into a single composition expression pattern is essential for improving search accuracy and identifying related patents.

2.3. Structural and Relational Annotation in Named Entity Recognition

Several studies have introduced auxiliary annotation layers to encode structural relationships between entities. Luan et al. [31] proposed a multitask framework that jointly trains named entity recognition and relation extraction in scientific papers, enabling the utilization of structural dependencies between entity spans. Wadden et al. [32] extended this work to nested entities and coreference, demonstrating that structural signals improve entity recognition accuracy. The most closely related prior study is the use of sentence-level bracket tokens in dialogue state tracking [33], which is a method that inserts special tokens into the input tokens to delimit slot value spans. A similar approach has also been attempted for relation extraction. Soares et al. [34] performed relation extraction by inserting [E1] . . . [/E1] into the input, demonstrating the origin of the idea of inserting brackets into the input. Zhou et al. [35] demonstrated improved performance for sentence-level relation extraction by incorporating entity representations using typed markers. Sainz et al. [36] demonstrated that explicitly encoding annotation guidelines into LLM fine-tuning substantially improves zero-shot information extraction, underscoring the broader principle that precise structural annotation design directly benefits extraction performance.

These studies collectively demonstrate that various forms of explicit structural annotation, including boundary markers, typed entity markers, and annotation guidelines, consistently improve information extraction quality. In contrast to relation extraction, which labels directed links between pre-identified entity pairs, the SEP-tag approach proposed in this study encodes composition pattern boundaries directly into the named entity recognition label sequence. This allows boundary information to be learned using an end-to-end sequence labeling model. Unlike boundary marker techniques, which insert special tokens into the input sequence, SEP-tags are label classes assigned to tokens that already exist within the text. Consequently, the input sequence remains unchanged, and structural boundary information is encoded solely in the label space. This study is the first to formalize the expertise of materials science experts regarding composition pattern boundaries as a named entity recognition label.

3. Proposed Method

3.1. Task Definition and Composition Expression Pattern Structure

Let $\mathbf{x} = (x_1, x_2, \dots, x_T)$ be a token sequence from a patent sentence. The goal of composition expression pattern extraction is to identify a continuous subsequence of tokens that collectively describes a single sequence of composition entities (such as elements, composition ranges and units) that constitutes a composition expression pattern. We define a composition expression pattern region as a token span enclosed by SEP-tags. In the annotation scheme, a composition expression pattern region takes the following form:

$$\text{SEP} \underbrace{e_1 \ e_2 \ \dots \ e_k}_{\text{content entities}} \text{SEP} \quad (1)$$

where, e_i is the label of an entity (e.g., atom, fig_LL). At least one SEP-tag is required at the start and end boundaries. A SEP-tag and the sequence of entities enclosed between SEP-tags are extracted as a single composition expression pattern.

Figure 1 shows an annotated example sentence. A single sentence contains multiple composition expression pattern regions (one region per composition expression pattern). Table 1 shows all 16 entity labels used in both datasets. The 13 content labels represent the semantic roles of tokens within composition expression patterns, while the structural labels (SEP, limitation, selection) indicate auxiliary roles for extracting composition expression pattern boundaries and content labels. Frequency statistics for entities in the English and Japanese datasets are shown in Appendix C.

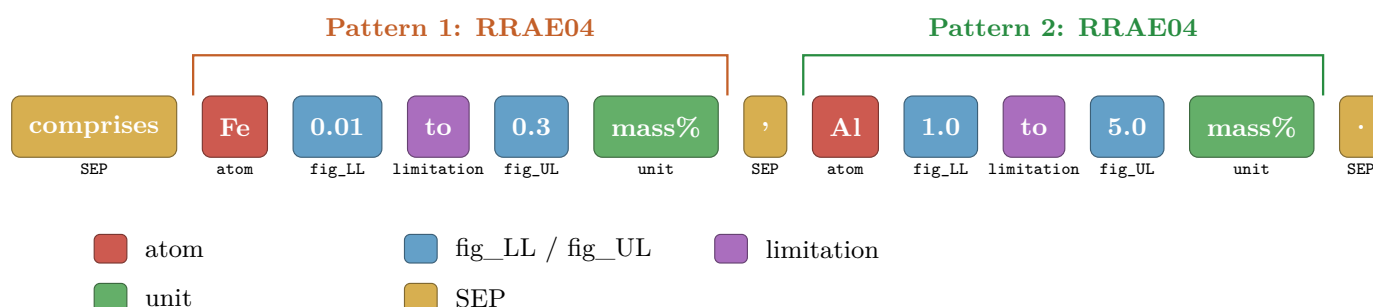


Figure 1. SEP-tag annotation for the sentence fragment “comprises Fe 0.01 to 0.3 mass%, Al 1.0 to 5.0 mass%.”. Gold tokens (SEP) are separator words in the patent text. They enclose Pattern 1 and Pattern 2. Each color corresponds to one entity label type (see legend).

Table 1. Entity label schema (16 types shared across English and Japanese datasets). Structural labels (SEP, limitation, selection) are excluded from entity-level F1 score (Axis 1).

Label	Type	Description and Example
atom	content	Chemical element or material name (<i>Fe, aluminum, carbon</i>)
fig_LL	content	Lower bound numeric value (<i>0.01, 1.0</i>)
fig_UL	content	Upper bound numeric value (<i>0.3, 5.0</i>)
fig	content	Single numeric value (no bound distinction) (<i>5</i>)
unit	content	Unit of measurement (<i>mass%, wt%, %</i>)
unit_def	content	Unit definition prefix (<i>in mass %, by weight</i>)
use	content	Intended use or product name (<i>steel, alloy</i>)
substance	content	Compound or substance name (<i>oxide, carbide</i>)
variable	content	Variable or parameter symbol (<i>x, n</i>)
formula	content	Chemical formula (<i>Fe₃C, Al₂O₃</i>) or formula (<i>(Mo + W) ≥ 4.3%</i>)
balance	content	Remainder/balance expression (<i>balance, remainder</i>)
f_no	content	Formula or figure reference number (<i>Formula (1)</i>)
sum	content	Sum or total expression (<i>total, in sum</i>)
limitation	structural	Constraint or comparison keyword (<i>to, or less, at most, ≤</i>)
selection	structural	Selection expression (<i>one or more of, selected from</i>)
SEP	structural	Composition expression pattern boundary separator (<i>wherein, comprising</i>)

Italic text indicates example tokens.

3.2. SEP-Tag Named Entity Recognition Model

We fine-tune a pre-trained transformer model for token classification with the augmented label set $\mathcal{L} = \mathcal{L}_{\text{entity}} \cup \mathcal{L}_{\text{SEP}}$, where $\mathcal{L}_{\text{SEP}} = \{\text{B-SEP}, \text{I-SEP}\}$. Given a tokenized input $\mathbf{x}$, the model produces a contextualized representation $\mathbf{H} = (\mathbf{h}_1, \dots, \mathbf{h}_T) \in \mathbb{R}^{T \times d}$ from the transformer encoder. A linear classification head maps each token representation to a label distribution:

$$P(y_t | \mathbf{x}) = \text{softmax}(\mathbf{W}\mathbf{h}_t + b), \quad \mathbf{W} \in \mathbb{R}^{|\mathcal{L}| \times d} \quad (2)$$

The model is trained end-to-end using a cross-entropy loss over all tokens. For the English dataset we use RoBERTa-base [37] (roberta-base), and for the Japanese dataset we use BERT-base-Japanese [38] (cl-tohoku/bert-base-japanese-whole-word-masking) (<https://huggingface.co/tohoku-nlp/bert-base-japanese-whole-word-masking>, accessed on 19 April 2026). Both models use standard WordPiece/BPE subword tokenization with a maximum sequence length of 512 tokens. We deliberately restrict our comparison to standard BERT-derived models to isolate the effect of SEP-tag training from confounding factors such as model architecture or pre-training corpus differences. Evaluating the impact of SEP-tags on larger or more recent models (e.g., RoBERTa-large, LLMs) remains a direction for future work.

The baseline without SEP-tags is trained under the same settings with $\mathcal{L}_{\text{SEP}} = \emptyset$, meaning tokens annotated with SEP-tags are treated as 0 during training. Hereinafter, the model trained with SEP-tag labels is referred to as the “model with SEP-tags”, and the model trained with SEP-tag labels replaced by O is referred to as the “model without SEP-tags”.

3.3. Extraction of Composition Expression Patterns from Predicted Labels

Given a sequence of predicted labels $\hat{\mathbf{y}}$, we extract SEP-tags and the composition expression patterns enclosed by them. Specifically, we identify SEP-tags by scanning the

token sequence from left to right and collect content entity and structural entity tokens up to the next SEP-tag. Tokens with the 0 label within the composition expression pattern region are ignored. This is because they do not affect the identity of the composition expression pattern.

3.4. Four-Axis Evaluation Framework

In this study, we define a four-axis evaluation framework to evaluate the extraction of composition expression patterns using SEP-tags. Standard entity-level F1 score typically measures only the recognition of individual entity tokens. Since SEP-tags are structural entities with properties different from conventional entities, standard entity-level F1 score cannot fully evaluate their effectiveness. Beyond standard entity-level scoring, three additional evaluation aspects are necessary. First, we assess whether entity labels within composition expression patterns are correctly recognized, independently of boundary prediction (using true SEP-tag spans as correct boundaries). Second, we assess whether the model with SEP-tags can identify the correct composition expression patterns from its own predicted boundaries. Third, we compare both models on pattern extraction accuracy without any boundary oracle, which most closely reflects real-world extraction conditions. Therefore, in this study, we conduct a four-axis evaluation based on these evaluation criteria. Table 2 shows the four evaluation axes.

Table 2. Summary of the four-axis evaluation framework.

Axis	Name	Comparison Unit	Model	Overview
1	Entity-level F1 score	Token (entity span)	Both	Standard named entity recognition accuracy. Structural labels excluded.
2	Correct span exact match rate	Entity sequence	Both	Ground-truth SEP-tag spans provided. Isolates entity labeling.
3	Predicted span pattern extraction F1 score	Entity sequence	The model with SEP-tags	Self-predicted SEP-tag boundaries. End-to-end span plus entity.
4	Pattern extraction F1 score	Pattern ID	Both	No boundary hint. Most practical evaluation.

3.4.1. Axis 1: Entity-Level F1 Score

We calculate the standard seqeval [39] entity-level F1 score, excluding all auxiliary structural labels (SEP, limitation, selection). Note that in our previous research [11], we found that auxiliary structural labels contribute to improving the accuracy of content entity labels for elements and composition values we wish to extract. In this study, we also exclude auxiliary structural labels from the overall F1 score calculation. This allows us to evaluate whether training using SEP-tags degrades entity-level recognition.

3.4.2. Axis 2: Correct Span Exact Match Rate

We extract entity sequences from each span by using the correct SEP-tags boundaries as spans. We then verify whether the predicted entity sequences exactly match the correct sequences. In other words, Axis 2 assesses the recognition performance of entity labels within composition expression patterns, excluding SEP-tags. This allows us to evaluate the recognition capability of composition expression patterns independently of boundary prediction.

3.4.3. Axis 3: Predicted Span Pattern Extraction F1 Score

We extract composition expression patterns from the predicted SEP-tags and compare them with those extracted from the true SEP-tags to calculate the precision, recall, and F1 score for the multiset of composition expression pattern sequences. Tokens with the 0 label are ignored within SEP-tags. This measures the model's ability to position and identify correct composition expression patterns using its own predicted boundaries.

3.4.4. Axis 4: Pattern Extraction F1 Score

We compare both models fairly against the same ground-truth composition expression patterns. The model with SEP-tags uses the predicted SEP-tags as the extraction window, while the model without SEP-tags uses continuous entity groups (any sequence of tokens other than 0) as the extraction window. In other words, Axis 4 purely evaluates the performance of composition expression pattern extraction. This evaluation is the most representative of real-world performance.

4. Experiments

4.1. Datasets

4.1.1. English Dataset

We constructed a dataset of 10,166 English-language materials science patent sentences extracted from the USPTO (United States Patent and Trademark Office), focusing on US registered patents classified under CPC C22 issued between January 2000 and December 2004. This corpus consists of 1000 manually annotated sentences and 9166 pseudo-labeled sentences. The pseudo-labeled sentences were labeled using a RoBERTa-large model trained on the 1000 manually annotated sentences. For the 1000 sentences, 900 were used as training data and 100 as validation data for fine-tuning, achieving an F1 score of 0.9333 on the validation data. We randomly sampled 100 of the 9166 pseudo-labeled sentences and manually verified them, achieving a micro F1 score of 0.9772 (Precision: 1.0000, Recall: 0.9553). The dataset contains 16 types of entity labels (13 content labels + 3 structural labels including SEP-tags), and includes 1074 complete composition expression pattern instances corresponding to 33 of the 50 predefined composition pattern templates (66.0% coverage). Predefined composition expression patterns will be discussed later. All four axes are evaluated using 5-fold cross-validation.

4.1.2. Japanese Dataset

Although patent specifications contain a variety of descriptions, this study focuses on the “claims”. The claims are a critical component that sets forth the requirements for defining an invention and clearly delineates the scope of the invention to be protected. Furthermore, the specifications of patent applications cover a wide range of technical fields. Therefore, utilizing the International Patent Classification (IPC) categories related to “alloys”, we extracted 15,053 registered patents belonging to IPC class C22C (Alloys) that were published between January 2000 and August 2021. All of these patents include the term “composition” in their claims. This determination was based on expert advice, as it was judged that sentences containing “composition” represent expressions indicating the lower and upper limits of elemental and compound content. Of the 15,053 registered patents mentioned above, we use sentences from 975 patents that were annotated entirely by hand. We use the same entity label schema as the English dataset. Similarly, we evaluate the model using 5-fold cross-validation. For the Japanese dataset, a single expert in materials science performed the annotation based on the defined entity labels and composition expression patterns. For the English dataset, annotation was performed by one expert in materials science and one researcher. Regarding the entity labels and composition expression patterns defined by the materials science expert, the two individuals cross-checked each other’s work to ensure consistency in their understanding before proceeding with the annotation. In such cases, the opinion of the materials science expert was adopted.

4.1.3. Definition of Composition Expression Patterns

Table 3 shows typical composition expression patterns defined by domain experts in materials science, which encode the structure of general composition expressions. A

complete list of all composition expression patterns in English and Japanese is provided in Appendices A and B, respectively.

Figure 2 shows the frequency distribution of 33 observed composition expression pattern templates in the English dataset (stacked plots of the manually annotated and pseudo-labeled portions). The distribution shows a strong skew, with the top five composition expression patterns accounting for 611 out of 1074 cases (56.89%), while 17 composition expression patterns have fewer than 10 instances.

Table 3. Definitions of typical composition expression patterns common to both English (En) and Japanese (Ja). atom: element; fig_LL/fig_UL: lower/upper bound; unit: unit; limitation: constraint keyword. En/Ja counts indicate total instances in each corpus.

Token Sequence	Relation	En ID	Ja ID	Count (En/Ja)
atom → fig_LL → limitation → fig_UL → unit	Compo_RA	RRAE04	RRA008	95/1710
atom → fig_LL → unit → limitation	Compo_OS	ROSE01	ROS002	133/50
atom → fig_LL → limitation → fig_UL	Compo_RA	RRAE01	RRA001	125/38
atom → fig_LL → unit → limitation → fig_UL → unit	Compo_RA	RRAE02	RRA002	47/68
fig_UL → unit → limitation → atom	Compo_OS	ROSE09	ROS007	44/35

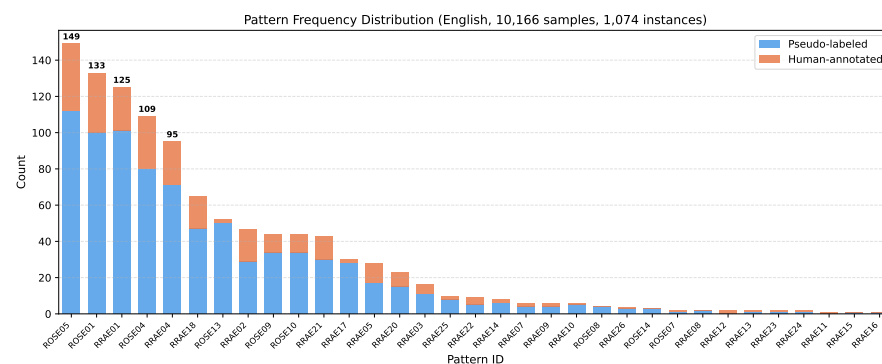


Figure 2. Frequency distribution of 33 defined composition expression patterns in the English dataset (10,166 sentences, 1074 total instances). Blue: pseudo-labeled portion; orange: manually annotated portion. Composition expression patterns are sorted in descending order by total count.

Figure 3 shows the corresponding distribution for the Japanese dataset (18 of 27 patterns were observed, totaling 3359 instances). The Japanese data distribution is even more concentrated: the top two composition expression patterns (RRA008 and ROS004) alone account for 2678 out of 3359 instances (79.72%).

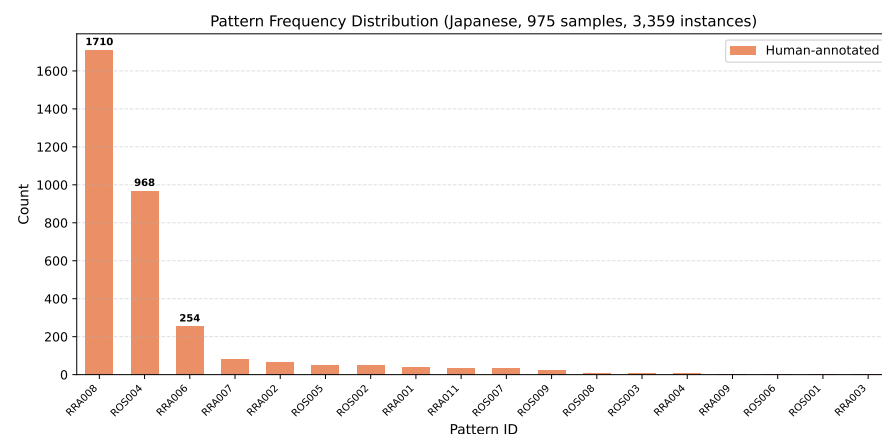


Figure 3. Frequency distribution of 18 defined composition expression patterns in the Japanese dataset (975 sentences, 3359 total instances). All data are manually annotated. Composition expression patterns are sorted in descending order by total count.

4.2. Training Settings

Table 4 summarizes the training settings used for both models with and without SEP-tags.

Table 4. Training settings.

Model/Hyperparameter	Value
English base model	roberta-base
Japanese base model	cl-tohoku/bert-base-japanese-whole-word-masking
Maximum sequence length	512 tokens
Batch size	8
Learning rate	4×10^{-5} (AdamW [40])
LR scheduler	Linear warmup (10%) + linear decay
Maximum epochs	20
Early stopping patience	5 epochs

4.3. Main Results

Table 5 shows the results for each entity in Axis 1 for both the English and Japanese datasets (English: left table, Japanese: right table). For English, the micro F1 score on a 5-fold cross-validation is 0.9201 (without SEP-tags) and 0.9216 (with SEP-tags). For Japanese, the values are 0.8992 (without SEP-tags) and 0.8984 (with SEP-tags), confirming that entity recognition performance is maintained in both languages.

Table 5. Axis 1: Entity-level F1 score (structural labels excluded, 5-fold CV). Left: English (10,166 sentences), Right: Japanese (975 sentences).

(Left)				(Right)			
Model	Precision	Recall	F1	Model	Precision	Recall	F1
without SEP-tags	0.9095	0.9310	0.9201	without SEP-tags	<u>0.8875</u>	0.9111	<u>0.8992</u>
with SEP-tags	<u>0.9110</u>	<u>0.9324</u>	<u>0.9216</u>	with SEP-tags	0.8853	<u>0.9119</u>	0.8984
Diff.	+0.0015	+0.0014	+0.0015	Diff.	−0.0023	+0.0007	−0.0008

Underline indicates the higher value between the model without SEP-tags and the model with SEP-tags.

Table 6 shows the results for both the English and Japanese datasets in Axes 2–4 (English and Japanese columns shown side-by-side). In Axis 2 (correct span exact match rate), we confirmed that the exact match rate was significantly higher for the model with SEP-tags in both English and Japanese. Since the model with SEP-tags learns SEP-tags, it is believed that it also learns the named entity structure within composition expression patterns, leading to an increase in the number of matched composition expression patterns.

Axis 3 (predicted span pattern extraction F1 score) evaluates the ability of the model with SEP-tags to extract entity sequences within the SEP-tags it predicted, independently of boundary detection based on SEP-tags. The model with SEP-tags achieved an F1 score of 0.8807 (precision: 0.8680, recall: 0.8937) in English, and an F1 score of 0.6803 (precision: 0.7830, recall: 0.6014) in Japanese. The low recall rate for Japanese may be attributed to the difference in the size of the training data (975 cases vs. 10,166 cases) and the morphological complexity of Japanese composition expressions. Next, in Axis 4 (pattern extraction F1 score), the baseline without SEP-tags achieved a high recall rate (0.8354) in the English dataset. The model without SEP-tags generates consecutive entity groups from tokens other than 0, resulting in 56,739 candidate composition expression patterns across all CV folds. While it generates nearly all true composition expression patterns, resulting in a high recall rate, the high number of false positives limits the precision to 0.0412 and the F1 score to 0.0784. In contrast, the model with SEP-tags uses predicted SEP-tags to limit

candidate composition expression patterns to 2878 across all five CV test folds, reducing false positives by a factor of 117 while maintaining a recall of 0.8630. We observed a similar trend in the Japanese dataset as well.

Table 6. Axes 2, 3, and 4 (5-fold CV). Axis 2: Correct span exact match rate. Axis 3: Predicted span pattern extraction F1 score. Axis 4: Pattern extraction F1 score. Column headers abbreviate “the model with SEP-tags” as w/ SEP-tags and “the model without SEP-tags” as w/o SEP-tags.

Axes	Metric	English (10,166)			Japanese (975)		
		w/o SEP-Tags	w/ SEP-Tags	Diff.	w/o SEP-Tags	w/ SEP-Tags	Diff.
2	Exact match rate	0.1595	<u>0.7567</u>	+0.5972	0.1754	<u>0.8655</u>	+0.6901
3	Precision	—	0.8680	—	—	0.7830	—
3	Recall	—	0.8937	—	—	0.6014	—
3	F1 score	—	0.8807	—	—	0.6803	—
4	Precision	0.0412	<u>0.8381</u>	+0.7969	0.0185	<u>0.5613</u>	+0.5428
4	Recall	0.8354	<u>0.8630</u>	+0.0276	<u>0.7826</u>	0.4312	−0.3514
4	F1 score	0.0784	<u>0.8503</u>	+0.7719	0.0361	<u>0.4877</u>	+0.4516
4	False positives	54,404	466	$\times \frac{1}{117}$	11,485	93	$\times \frac{1}{123}$

Underline indicates the higher value between the model without SEP-tags and the model with SEP-tags.

4.4. Statistical Significance Tests

This section discusses statistical significance tests for each evaluation. Axis 1: To verify whether the performance difference at the entity level is statistically significant, we apply a label-specific Wilcoxon signed-rank test with Benjamini–Hochberg (BH) correction ($\alpha = 0.05$) to the 5-fold cross-validation results of the 10,166-sentences English dataset and the 975-sentences Japanese dataset. Table 7 shows the results for both languages. No labels showed a statistically significant difference in either language (minimum raw $p = 0.0754$, 0 out of 13 labels were significant after BH correction).

Table 7. Axis 1: Per-label entity F1 score significance test (Wilcoxon signed-rank with BH correction, $\alpha = 0.05$; 5-fold CV). Positive mean diff indicates with SEP-tags > without SEP-tags. Top-4 labels by raw p -value shown per language. All others raw $p > 0.3$.

Label	English (10,166)		Japanese (975)	
	Mean Diff	p -Value	Mean Diff	p -Value
atom	−0.0126	0.0955	—	—
balance	—	—	−0.0179	0.1797
fig	+0.0595	0.1683	—	—
fig_UL	—	—	+0.0089	0.0754
substance	—	—	−0.0306	0.2619
sum	+0.0564	0.1688	—	—
unit	—	—	−0.0073	0.2845
variable	+0.0148	0.1048	—	—
(all BH-corrected $p > 0.43$, 0/13 significant after BH correction)				

To complement the analysis by label, we also applied the Wilcoxon signed-rank test to the overall micro-average F1 score calculated for each sample (excluding structural labels and performing sequeval entity-level evaluation). Note that the sample-level average F1 score shown in Table 8 is calculated differently from the aggregated micro-F1 score in Table 5. While the aggregated micro-F1 score is calculated by summing the TP, FP, and FN across all samples, the sample-level average is the mean of the F1 scores for each individual sample, and tends to be lower due to the influence of samples with fewer entities. In English, a slight improvement of 0.0023 points was found to be statistically significant in the model with SEP-tags ($p = 0.0147$), while in Japanese, there was no significant difference

($p = 0.3308$). Both results are consistent with the analysis by label. Training with SEP-tags preserves entity recognition in both languages. For Axis 2 (correct span exact match rate), the McNemar test showed $\chi^2 = 0.1552$, $p = 0.6936$ (English), and $\chi^2 = 0.0000$, $p = 1.0000$ (Japanese), confirming that there is no significant difference in the quality of matches at the span level. Note that the Japanese result ($\chi^2 = 0.0000$, $p = 1.0000$) reflects the extremely small number of discordant span pairs: no span was detected only by the model without SEP-tags, and only one span was detected only by the model with SEP-tags, out of 48 ground-truth SEP-tag spans. The McNemar test lacks statistical power in this setting and does not contradict the large absolute improvement in exact match rate (+0.6901).

Table 8. Axis 1: Overall micro-average F1 score Wilcoxon signed-rank test (two-sided, $\alpha = 0.05$). Per-sample entity-level F1 score (structural labels excluded, evaluated using sequeval).

Dataset	<i>n</i>	Without SEP-Tags	With SEP-Tags	Diff.	<i>p</i> -Value
English	10,166	0.8190	<u>0.8212</u>	+0.0023	0.0147
Japanese	975	<u>0.8652</u>	0.8625	−0.0027	0.3308

Underline indicates the higher value between the model without SEP-tags and the model with SEP-tags. Bold *p*-value indicates statistical significance ($p < 0.05$).

Regarding Axis 4 (pattern extraction F1 score), Fisher’s exact test on the 10,166 English examples using 5-fold cross-validation confirms that the improvement in precision is highly significant ($p < 10^{-10}$, without SEP-tags: 0.0412, with SEP-tags: 0.8381). Similar results were obtained with the Wilcoxon signed-rank test at the sample level ($p < 10^{-10}$, average improvement +0.5374, see Section 5.3). For the Japanese data, similar results were also obtained. In Japanese, the improvement in the precision is highly significant ($p < 10^{-10}$, without SEP-tags: 0.0185, with SEP-tags: 0.5613). Similar results were obtained using the Wilcoxon signed-rank test on a per-sample basis ($p < 10^{-10}$, average improvement +0.3069).

5. Analysis

5.1. Why SEP-Tags Do Not Lower Entity F1 Score?

Training to add SEP-tags as entities results in SEP-tags being added as a new class for classification. Nevertheless, the results for both English and Japanese show that the change in entity F1 score is negligible ($|\Delta F1| < 0.002$ for English, $|\Delta F1| < 0.001$ for Japanese) (see Table 5). In this study, we believe there are two factors contributing to this. First, SEP-tags are semantically distinct from content entities. Typically, they are function words or punctuation marks (e.g., “wherein,” “comprising,” or commas in Japanese) and do not appear as entity tokens. Therefore, the model learns to separate SEP-tags with minimal interference. Second, the Transformer’s self-attention mechanism [41] can utilize SEP-tags as contextual anchors. The presence of SEP-tags provides additional structural context and has the potential to improve entity recognition capabilities within bounded regions. This is suggested by the fact that the recall of the model with SEP-tags is slightly higher on the English dataset (model with SEP-tags: 0.9324 vs. model without SEP-tags: 0.9310). A similar trend in recall is observed in Japanese (see Table 5).

5.2. Results for Composition Expression Patterns in Axis 4

Table 9 shows the results for Axis 4, broken down by composition expression pattern, for the dataset of 10,166 English sentences. For all composition expression pattern definitions with ten or more correct examples, the model with SEP-tags consistently outperforms the model without SEP-tags. While the F1 score for the model without SEP-tags is close to zero for many composition expression patterns, the model with SEP-tags achieved F1 scores ranging from 0.714 to 0.887. Significant improvements were also observed in the “OTHER” category (2319 composition expression patterns that did not match the defined

composition expression patterns). (F1 score by the model without SEP-tags = 0.080 vs. F1 score of the model with SEP-tags = 0.857) The “OTHER” category targeted those with one or more entity types. These results indicate that the SEP-tag generalizes beyond the defined composition expression patterns.

Table 9. Per-pattern Axis 4 results on the English dataset (10,166 sentences). Support: number of ground-truth instances, $\Delta F1$ = with SEP-tags – without SEP-tags. Patterns with support ≥ 10 are shown.

Pattern	Support	Without SEP-Tags			With SEP-Tags			$\Delta F1$
		Precision	Recall	F1 Score	Precision	Recall	F1 Score	
RRAE19	116	0.014	0.155	0.025	<u>0.895</u>	<u>0.879</u>	<u>0.887</u>	+0.862
RTOE01	106	0.038	0.292	0.067	<u>0.791</u>	<u>0.821</u>	<u>0.806</u>	+0.738
RRAE04	40	0.000	0.000	0.000	<u>0.791</u>	<u>0.850</u>	<u>0.819</u>	+0.819
ROSE11	29	0.002	0.034	0.004	<u>0.844</u>	<u>0.931</u>	<u>0.885</u>	+0.881
ROSE02	21	0.000	0.000	0.000	<u>0.714</u>	<u>0.714</u>	<u>0.714</u>	+0.714
RRAE21	20	0.010	0.050	0.017	<u>1.000</u>	0.650	<u>0.788</u>	+0.771
ROSE03	20	0.200	0.050	0.080	<u>0.737</u>	<u>0.700</u>	<u>0.718</u>	+0.638
RRAE18	19	0.004	0.053	0.008	<u>0.778</u>	<u>0.737</u>	<u>0.757</u>	+0.749
ROSE04	14	0.000	0.000	0.000	<u>0.800</u>	<u>0.857</u>	<u>0.828</u>	+0.828
RRAE06	14	0.000	0.000	0.000	<u>0.833</u>	<u>0.714</u>	<u>0.769</u>	+0.769
RRAE02	10	0.067	0.100	0.080	<u>0.750</u>	<u>0.900</u>	<u>0.818</u>	+0.738
OTHER	2319	0.042	<u>0.928</u>	0.080	<u>0.841</u>	<u>0.873</u>	<u>0.857</u>	+0.776

Underline indicates the higher value between the model without SEP-tags and the model with SEP-tags.

5.3. Why Does the Precision of the Model Without SEP-Tags Decline in Axis 4?

The model without SEP-tags achieves an entity-level F1 score of 0.9201 in the English dataset (see Table 5). However, in Axis 4, the precision at the composition expression pattern level drops to 0.0412 (see Table 6). This gap clearly demonstrates that a separator such as the SEP-tag is necessary to extract composition expression patterns. One reason is that, while individual token labeling achieves high accuracy without SEP-tags, it is unable to organize those tokens into meaningful compositional groups as composition expression patterns. In the English 5-fold CV evaluation (10,166 sentences), the model without SEP-tags generates 56,739 consecutive entity groups from token spans other than O. Of these, 54,404 are false positives. Since there are only 2795 correct composition expression patterns, the majority of these false positives result from the erroneous segmentation of long text regions into multiple overlapping composition expression pattern candidates. In contrast, the model with SEP-tags is constrained by the SEP-tags and generates only 2878 candidate groups across all 5-fold CVs, of which 83.8% match the correct composition expression patterns (FP = 466 out of 2878 predictions). To confirm that this difference is statistically significant, we performed Fisher’s exact test on the aggregated TP/FP counts (H_1 : Precision (without SEP-tags) < Precision (with SEP-tags)) and a sample-level Wilcoxon signed-rank test for Axis 4 (H_1 : $F1_{\text{with SEP-tags}} > F1_{\text{without SEP-tags}}$).

Fisher’s exact test indicates an odds ratio of 8.0×10^{-3} , $p < 10^{-10}$ (Precision: without SEP-tags = 0.0412, with SEP-tags = 0.8381). The sample-level Wilcoxon test (for $n = 2407$ samples with correct composition expression patterns) shows $p < 10^{-10}$ and an average F1 score improvement of +0.5374. Both tests confirm that the improvement in precision and F1 score for the model with SEP-tags in Axis 4 is highly statistically significant.

Error Analysis

To better understand the causes of false positives, we performed a qualitative analysis of all 466 FP from the model with SEP-tags in Axis 4 on the 10,166 sentences English 5-fold

CV. As a result, we identified three types of errors. (1) Prediction of empty composition expression patterns (188 cases, 40.3%): Cases where the model predicts an SEP-tag but no entity exists within it, resulting in an SEP-tag being predicted in an area where there is no composition. (2) Predictions outside composition expression patterns (OTHER, 195 cases, 41.9%): Sequences that do not belong to defined composition expression patterns, with `unit_def` alone accounting for the majority. (3) Predictions within composition expression patterns with no correct answer (83 cases, 17.8%): Cases where the predicted composition expression pattern ID is correct, but there is no corresponding correct answer within the same sample.

In the model without SEP-tags (FP = 54,404 cases), the dominant error type is a single entity prediction. `atom` alone accounts for approximately 13,084 cases (24.0%), `use` alone accounts for 9135 cases (16.8%), and `substance` alone accounts for 6494 cases (11.9%). The reason for this is that the model without SEP-tags can accurately label individual entities. However, since it lacks a mechanism to group them into multi-element compositional units, isolated entities are extracted as individual candidate composition expression patterns. Consequently, almost all of them do not match predefined composition expression patterns. These findings confirm that the SEP-tag directly resolves the segmentation problem that causes a decrease in precision rate under settings without SEP-tags.

As future improvements, we discuss approaches to address these three types of errors. First, regarding (1) Prediction of empty composition expression patterns, one possible post-processing method is to exclude predictions where no entity exists between SEP-tags from the list of candidate composition expression patterns. Applying this post-processing is expected to reduce false positives of this error type. Next, we describe approaches for (2) Predictions outside composition expression patterns. Among the composition expression patterns falling under (2), there is a possibility that they include useful composition expression patterns that could not be discovered during the manual annotation stage. Therefore, as a future direction, we expect to expand the definition of composition expression patterns by manually inspecting the composition expression patterns extracted under error type (2). Finally, regarding (3) Predictions within composition expression patterns with no correct answer, it is possible to reduce false positives by improving SEP-tag prediction accuracy and by adding or revising annotations to supplement the corresponding ground-truth patterns, as further discussed in the Section 5.5.

5.4. Analysis of False Negatives in the Japanese Dataset

In Evaluation 4 of the Japanese dataset, the recall dropped from 0.7826 to 0.4312 after introducing SEP-tags. Since the model with SEP considers only the range enclosed by the predicted SEP-tags as pattern candidates, the true patterns of sentences where SEP-tag prediction failed are not generated as candidates and result in false negatives. In the Japanese dataset, one factor contributing to conservative SEP-tag prediction is the size of the training data. The Japanese dataset consists of 975 sentences, which is approximately one-tenth the size of the English dataset (10,166 sentences). Another factor is the diversity of words assigned SEP-tags (see Section 5.5). The number of distinct strings assigned SEP-tags was 165 in the English dataset and 299 in the Japanese dataset. In the Japanese dataset, the problem is considered difficult because boundaries must be predicted for a wide variety of tokens. Furthermore, since the number of training sentences per fold is limited, the model tends to choose the O label for uncertain boundaries, resulting in an increase in false negatives. Therefore, expanding the Japanese dataset is a key approach to improving recall, and this will be a focus for future work.

The Correlation Between SEP-Tags and the Accuracy of Composition Expression Pattern Extraction

To quantify the relationship between SEP-tag prediction accuracy and the quality of extracted composition expression patterns, we calculated the F1 score for SEP-tags and the F1 score for composition expression patterns based on Axis 4 for 2407 samples, each of which contained at least one correct composition expression pattern. As a result, Pearson $r = 0.543$ and Spearman $\rho = 0.555$, indicating a moderate positive correlation for both metrics. Additionally, Table 10 shows the F1 scores for extracting composition expression patterns within each F1 score range when detecting SEP-tags. In samples where the F1 score for SEP-tag ≥ 0.8 (2214 out of 2407), the average F1 score for Axis 4 reaches 0.899, but in samples where the F1 score for SEP-tag $\in [0.6, 0.8)$, the average F1 score for Axis 4 drops to 0.476. This confirms that the 40.3% false positive rate identified in the error analysis in Section 5.3, which consisted of predictions of empty composition expression patterns, was primarily due to samples where SEP-tag prediction failed, and that further improvements in SEP-tag prediction accuracy directly lead to improved composition expression pattern extraction accuracy.

Table 10. Axis 4 of the model with SEP-tags segmented by SEP-tag F1 scores per sample. Axis 4: Average F1 score (English 5-fold CV, 2407 samples with correct composition expression patterns).

SEP-Tags F1 Score Range	Samples	Axis 4 Average F1 Score
[0.8, 1.0]	2214	0.899
[0.6, 0.8)	146	0.476
[0.4, 0.6)	39	0.252
[0.2, 0.4)	8	0.125

5.5. Validation of the Non-Crossing Hypothesis

A key assumption underlying the SEP-tag approach is that SEP-tags do not span composition expression patterns. In other words, since SEP-tags do not appear within composition expression patterns, a single composition expression pattern is never described across multiple SEP-tags. In this study, we tested this hypothesis using three main procedures.

5.5.1. Method 1: Semantic Categorization of Strings with SEP-Tags

Analysis of SEP-tagged strings in annotated materials science patent documents (1000 English and 975 Japanese) revealed that they can be classified into three main semantic categories. (i) Composition Enumeration Starters (e.g., English: comprising, wherein; Japanese: karanaru, woganyuushi), which are strings indicating the start of a composition enumeration. (ii) Composition Enumeration Separators (e.g., English: comma, semicolon, and; Japanese: comma, oyobi), which are strings that separate individual composition expression patterns. (iii) Clause/Sentence Terminator (e.g., English: period, Japanese: dearu), a character sequence that terminates a sentence or clause. All three categories indicate boundaries rather than compositional content. Table 11 shows that 98.9% (English) and 95.9% (Japanese) of the SEP-tags in the annotation data belong to these three categories. The remaining categories included case particles and relative clauses, none of which were components of composition expression patterns and did not belong to any of the three semantic categories.

Table 11. The semantic categories of SEP-tags and the number of instances in annotation data. Ja: Japanese, En: English.

Category	En Example	Ja Example	En Coverage	Ja Coverage
Composition Enumeration Starters	comprising, wherein	karanaru, woganyuushi	30.2%	17.0%
Composition Enumeration Separators	',' and	oyobi, narabini	53.3%	66.3%
Clause/Sentence Terminator	'.'	dearu	15.4%	12.5%
Other	-	-	1.1%	4.1%

5.5.2. Method 2: Lexical Distinctness of SEP-Tag Strings from Content Entities

We directly verify that strings labeled with SEP-tags are lexically distinct from content entities (such as atom and fig). The results of the lexical overlap analysis are summarized in Table 12. The number of strings assigned SEP-tags was 165 types in the English data and 299 types in the Japanese data. Of these, 97.0% of the SEP-tags in the English data and 99.0% of the SEP-tags in the Japanese data do not appear as content entity tokens. In other words, these are strings assigned only by SEP-tags. On the other hand, the number of strings assigned to both SEP-tags and content entity tokens was 5 in the English data and 3 in the Japanese data. All of these few duplicate forms are considered to be context-dependent bracket numbers (e.g., (1), (i)) or annotation errors, and it has been confirmed that they do not represent true compositional content.

Table 12. Verification of strings assigned SEP-tags. Number of unique SEP-tag types that do not appear as content entity tokens.

Dataset	Unique SEP Types	Non-Overlapping	Overlapping
English (1000 human)	165	160 (97.0%)	5
Japanese (975)	299	296 (99.0%)	3

5.5.3. Method 3: SEP-Tag Non-Crossing Verification

At the annotation level, we examined all SEP boundary spans in both datasets (40,232 in English and 6146 in Japanese) and confirmed that no composition expression patterns spanned across SEP boundaries in either language.

Based on these three pieces of evidence, we consider this sufficient to validate the hypothesis that SEP-tags do not span composition expression patterns.

5.6. Cross-Lingual Consistency

In this study, we evaluated the proposed SEP-tag method across two different languages: English (10,166 sentences, roberta-base) and Japanese (975 sentences, BERT-base-Japanese). Despite differences in language, tokenization strategy, and dataset size, the results showed a remarkably consistent pattern of results across all four evaluation metrics.

5.6.1. Entity F1 Score (Axis 1)

The change in entity-level F1 score when adding SEP-tags is minimal in both languages. In English, $\Delta F1 = +0.0015$, and in Japanese, $\Delta F1 = -0.0008$; statistical significance tests confirmed that neither difference is significant. Therefore, adding SEP-tags as new entities does not negatively impact entity recognition performance.

5.6.2. Reduction of False Positives (Axis 4)

The model with SEP-tags significantly reduces false positives in both languages. In English, the ratio is 1/117 (FP: 54,404 $\rightarrow$ 466), and in Japanese, it is 1/123 (FP: 11,485 $\rightarrow$ 93).

These results confirm that SEP-tags significantly reduce false positives in both languages, supporting the robustness of the SEP-tag approach.

5.6.3. Differences Between Languages

The results of Axis 2 indicate that there is a difference in the correct span exact match rate between the two languages. In the model with SEP-tags, the 5-fold cross-validation for Japanese achieved an exact match rate of 0.8655, while the 5-fold cross-validation for English achieved an exact match rate of 0.7567. Both were evaluated using 5-fold cross-validation with identical hyperparameters. This gap may partially reflect the small size and homogeneity of the Japanese dataset (975 cases, 171 ground-truth SEP-tag spans). Furthermore, in the Axis 3 F1 score (predicted span pattern extraction), the Japanese micro F1 score (0.6803) is lower than the English micro F1 score (0.8807), suggesting that boundary prediction is more difficult in Japanese. This may be due to the high ambiguity in Japanese, where punctuation and particles have multiple roles. In any case, the qualitative conclusions are the same for both languages, which supports the generalizability of the proposed approach.

5.7. Comparison of Computational Costs

In this section, we conclude by comparing the proposed method with other methods in terms of training and inference costs and constraints. Table 13 shows a comparison of computational costs across different methods for extracting composition expression patterns. The training cost referred to here pertains solely to the task-specific computational overhead, excluding the encoder component of BERT. The method using BIO/BIOES and SEP-tags involves only fine-tuning the BERT model. Since it involves token-by-token learning for label prediction in the task head, the complexity is $O(n)$, resulting in a low training cost. Furthermore, since inference requires only $O(n)$ computations, the inference cost is also low. However, pattern extraction is difficult with the BIO/BIOES model. In the case of LLMs, no training costs are required since only the prompt needs to be designed. However, inference costs are high because API costs are incurred during inference. Furthermore, LLMs face challenges regarding privacy issues and reproducibility when the model is modified. Additionally, Foppiano et al. [28] have shown that even LLMs such as GPT-4 cannot match the zero-shot accuracy of fine-tuned BERT for materials science NER, indicating challenges in terms of accuracy as well. In span-based NER, all combinations of start and end positions are enumerated as span candidates, so the training cost for span enumeration is $O(n^2)$. The same applies during inference. Therefore, the SEP-tag approach proposed in this study requires only an initial annotation cost, but training and inference costs can be kept low.

Table 13. Comparison of computational costs by method for extracting composition expression patterns.

Method	Training Cost	Inference Cost	Constraints
BIO/BIOES	Low ($O(n)$)	Low ($O(n)$)	Cannot be patterned
SEP-tags	Low ($O(n)$)	Low ($O(n)$)	Initial annotation only
LLM	None	High (API) [†]	Privacy and reproducibility
Span-based	Medium ($O(n^2)$)	Medium ($O(n^2)$)	Span enumeration required

[†] LLM inference cost refers to API billing cost, not local computational complexity.

6. Conclusions

In this study, we propose the SEP-tag, a separator tag for extracting composition expression patterns in materials science patent documents. By introducing entities that explicitly mark structural boundaries as named entity recognition labels, the model can learn both entity recognition and composition expression pattern segmentation through

a single end-to-end fine-tuning process. We comprehensively evaluated performance at both the entity level and the composition expression pattern level using the proposed four-axis evaluation framework. Experiments on English (10,166 sentences) and Japanese (975 sentences) datasets showed that training with the SEP-tag improved the correct span exact match rate by 59.72 percentage points on the English dataset (Axis 2), reduced false positives in composition expression pattern extraction to 1/117 in English and 1/123 in Japanese (Axis 4), and maintained the entity F1 score across all 13 labels in both languages without any statistically significant decline (Axis 1).

Analysis by composition expression pattern confirmed that the improvement generalized across all observed composition expression pattern templates. Improvements were also observed in the “OTHER” category (2319 cases outside predefined patterns) (F1 score: 0.080 $\rightarrow$ 0.857). Error analysis revealed that the dominant type of remaining false positives was empty composition expression pattern predictions (40.3% of SEP-tags false positives), and sample-level analysis showed a moderate correlation between SEP-tag prediction accuracy and composition expression pattern extraction accuracy (Pearson $r = 0.543$).

The following issues remain to be addressed:

- Adding rule-based post-processing to the without SEP-tag baseline. We believe that this baseline without SEP-tags provides a fair comparison with the model that uses SEP-tags, as it employs the same architecture and training data, except for the difference in the presence or absence of SEP-tags. On the other hand, adding rule-based post-processing is expected to further enhance the competitiveness of the baseline.
- Improving recall by expanding the Japanese dataset.
- Direct improvement of SEP-tags prediction quality. As SEP-tags F1 score increases, the accuracy of composition expression pattern extraction improves consistently, and for samples with SEP-tags F1 score ≥ 0.8 , the average F1 score has already reached 0.899 in Axis 4.
- Evaluation using independent test data consisting solely of gold labels. We have confirmed that the errors in the pseudo-labels generated by the RoBERTa-large training model are limited and sufficient in quality. However, since the test fold also contains pseudo-labeled data, we cannot dismiss the possibility of performance overestimation. In the future, we expect to further enhance the reliability of the evaluation by conducting assessments using independent test data composed solely of gold labels.
- Extraction of composition expression patterns through the introduction of SEP-tags into other domains. This approach may be applied to domains where structural patterns consisting of multiple entities exist and where the vocabulary defining the boundaries of these patterns is well-defined. For example, in domains such as academic papers and legal documents, similar structural pattern extraction can be expected if domain experts define the vocabulary that indicates these boundaries as SEP-tags.

Author Contributions: Conceptualization, T.S.; software, T.S.; validation, T.S.; investigation, T.S.; data curation, T.S. and N.C.; writing—original draft, T.S.; writing—review and editing, T.S. and T.M.; supervision, T.M.; project administration, T.S.; funding acquisition, T.M. All authors have read and agreed to the published version of the manuscript.

Funding: This study is supported in part by KAKENHI JP26K03045 and JP26K00542.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this article are not readily available because the data were annotated by employees of a co-author company and contain proprietary information that cannot be disclosed externally.

Acknowledgments: The authors thank Kimiyuki Arashiba for assistance with annotation of the English dataset.

Conflicts of Interest: Author Nobuhiko Chiwata was employed by the company Proterial, Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. English Pattern Templates (Complete List)

Table A1 lists the 33 observed composition expression pattern templates in the English dataset. “Count” is the number of SEP-segment-bounded instances observed in the full English corpus (all 10,166 sentences including pseudo-labeled; 1074 total). Note that this differs from the “Support” column in Table 9, which counts ground-truth instances extracted from 5-fold CV test fold labels only.

Table A1. English composition expression patterns. Sorted by descending instance count.

Pattern ID	Relation	Token Sequence	Count
ROSE05	Compo_OS	atom → limitation → fig_UL	149
ROSE01	Compo_OS	atom → fig_LL → unit → limitation	133
RRAE01	Compo_RA	atom → fig_LL → limitation → fig_UL	125
ROSE04	Compo_OS	atom → limitation → fig_LL → unit	109
RRAE04	Compo_RA	atom → fig_LL → limitation → fig_UL → unit	95
RRAE18	Compo_RA	fig_LL → unit → limitation → fig_UL → unit → atom	65
ROSE13	Compo_OS	limitation → fig_UL → atom	52
RRAE02	Compo_RA	atom → fig_LL → unit → limitation → fig_UL → unit	47
ROSE09	Compo_OS	fig_UL → unit → limitation → atom	44
ROSE10	Compo_OS	fig_LL → unit → limitation → atom	44
RRAE21	Compo_RA	limitation → fig_LL → limitation → fig_UL → unit → atom	43
RRAE17	Compo_RA	fig_LL → limitation → fig_UL → atom	30
RRAE05	Compo_RA	atom → limitation → fig_LL → unit → limitation → fig_UL → unit	28
RRAE20	Compo_RA	limitation → fig_LL → unit → limitation → fig_UL → unit → atom	23
RRAE03	Compo_RA	atom → fig_LL → unit → limitation → fig_UL → unit → limitation	16
RRAE25	Compo_RA	fig_LL → unit → limitation → fig_UL → unit → udef → atom	10
RRAE22	Compo_RA	fig_LL → unit → limitation → fig_UL → unit → limitation → atom	9
RRAE14	Compo_RA	fig_LL → unit → limitation → atom → limitation → fig_UL → unit	8
RRAE07	Compo_RA	atom → fig_LL → unit → limitation → limitation → fig_UL → unit	6
RRAE09	Compo_RA	atom → fig_UL → unit → limitation → limitation → fig_LL → unit	6
RRAE10	Compo_RA	atom → limitation → fig_LL → unit → limitation → fig_UL → unit → udef	6
ROSE08	Compo_OS	atom → limitation → fig_UL → unit → udef	4
RRAE26	Compo_RA	limitation → fig_LL → unit → limitation → fig_UL → unit → udef → atom	4
ROSE14	Compo_OS	fig_UL → limitation → atom	3
ROSE07	Compo_OS	atom → fig_UL → unit → limitation → udef	2
RRAE08	Compo_RA	atom → limitation → fig_LL → unit → fig_UL → unit → limitation	2
RRAE12	Compo_RA	atom → fig_LL → unit → limitation → fig_UL → unit → udef	2
RRAE13	Compo_RA	atom → fig_LL → limitation → fig_UL → unit → udef	2
RRAE23	Compo_RA	fig_LL → unit → limitation → limitation → fig_UL → unit → atom	2
RRAE24	Compo_RA	fig_UL → unit → limitation → limitation → fig_LL → unit → atom	2
RRAE11	Compo_RA	atom → fig_LL → unit → limitation → limitation → fig_UL → unit → udef	1
RRAE15	Compo_RA	fig_LL → limitation → atom → limitation → fig_UL → unit	1
RRAE16	Compo_RA	fig_LL → limitation → atom → limitation → fig_UL	1

Appendix B. Japanese Pattern Templates (Complete List)

Table A2 lists the 18 observed compositional pattern templates in the Japanese dataset. “Count” is the total number of observed instances in the Japanese corpus (3359 total).

Table A2. Japanese composition expression patterns. Sorted by descending instance count.

Pattern ID	Relation	Token Sequence	Count
RRA008	Compo_RA	atom → fig_LL → limitation → fig_UL → unit	1710
ROS004	Compo_OS	atom → fig_UL → unit → limitation	968
RRA006	Compo_RA	atom → fig_LL → unit → limitation → fig_UL → unit → limitation	254
RRA007	Compo_RA	fig_LL → limitation → fig_UL → unit → atom	82
RRA002	Compo_RA	atom → fig_LL → unit → limitation → fig_UL → unit	68

Table A2. Cont.

Pattern ID	Relation	Token Sequence	Count
ROS005	Compo_OS	atom → limitation → fig_UL → unit	52
ROS002	Compo_OS	atom → fig_LL → unit → limitation	50
RRA001	Compo_RA	atom → fig_LL → limitation → fig_UL	38
RRA011	Compo_RA	atom → fig_LL → limitation → fig_UL → unit → limitation	36
ROS007	Compo_OS	fig_UL → unit → limitation → atom	35
ROS009	Compo_OS	limitation → fig_UL → unit → atom	24
ROS008	Compo_OS	fig_LL → unit → limitation → atom	10
ROS003	Compo_OS	atom → fig_UL → limitation	10
RRA004	Compo_RA	fig_LL → unit → limitation → fig_UL → unit → atom	9
RRA009	Compo_RA	fig_LL → limitation → fig_UL → unit	4
ROS006	Compo_OS	atom → limitation → fig_LL → unit	4
ROS001	Compo_OS	atom → fig_LL → limitation	3
RRA003	Compo_RA	fig_LL → limitation → fig_UL → atom	2

Appendix C. Entity Label Frequency Statistics

Tables A3 and A4 report the number of entity instances per label type in the English and Japanese corpora, respectively. Counts are aggregated at the entity level (each span is counted once, regardless of BIO prefix).

Table A3. Entity label instance counts in the English corpus (10,166 sentences, 132,923 total entity instances). Sorted by descending count within each group.

Label	Count	Ratio (%)
atom	19,933	15.0
unit	12,557	9.4
fig_UL	11,548	8.7
fig_LL	9450	7.1
use	9015	6.8
substance	6776	5.1
unit_def	1573	1.2
variable	1346	1.0
formula	1205	0.9
balance	827	0.6
fig	628	0.5
f_no	562	0.4
sum	230	0.2
limitation	14,563	11.0
selection	2478	1.9
SEP	40,232	30.3

Table A4. Entity label instance counts in the Japanese corpus (975 sentences, 33,515 total entity instances). Sorted by descending count within each group.

Label	Count	Ratio (%)
atom	6122	18.3
unit	4691	14.0
fig_UL	4170	12.4
fig_LL	2939	8.8
use	992	3.0
variable	594	1.8
unit_def	453	1.4
balance	382	1.1
substance	382	1.1
formula	365	1.1
f_no	285	0.9
sum	97	0.3
fig	19	0.1
limitation	4923	14.7
selection	955	2.8
SEP	6146	18.3

References

- Weston, L.; Tshitoyan, V.; Dagdelen, J.; Kononova, O.; Trewartha, A.; Persson, K.A.; Ceder, G.; Jain, A. Named Entity Recognition and Normalization Applied to Large-Scale Information Extraction from the Materials Science Literature. *J. Chem. Inf. Model.* **2019**, *59*, 3692–3702. [\[CrossRef\]](#) [\[PubMed\]](#)
- Song, Y.; Miret, S.; Liu, B. MatSci-NLP: Evaluating Scientific Language Models on Materials Science Language Tasks Using Text-to-Schema Modeling. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, ON, Canada, 9–14 July 2023; Volume 1, pp. 3621–3639. [\[CrossRef\]](#)
- Jiang, X.; Wang, W.; Tian, S.; Wang, H.; Lookman, T.; Su, Y. Applications of natural language processing and large language models in materials discovery. *npj Comput. Mater.* **2025**, *11*, 79. [\[CrossRef\]](#)
- Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.V.; Persson, K.A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019**, *571*, 95–98. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kononova, O.; He, T.; Huo, H.; Trewartha, A.; Olivetti, E.A.; Ceder, G. Opportunities and challenges of text mining in materials research. *iScience* **2021**, *24*, 102155. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, J.; Sun, A.; Han, J.; Li, C. A survey on deep learning for named entity recognition. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 50–70. [\[CrossRef\]](#)
- Yadav, V.; Bethard, S. A survey on recent advances in named entity recognition from deep learning models. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 2145–2158.
- Jehangir, B.; Radhakrishnan, S.; Agarwal, R. A survey on Named Entity Recognition—Datasets, tools, and methodologies. *Nat. Lang. Process. J.* **2023**, *3*, 100017. [\[CrossRef\]](#)
- Hu, Z.; Hou, W.; Liu, X. Deep learning for named entity recognition: A survey. *Neural Comput. Appl.* **2024**, *36*, 8995–9022. [\[CrossRef\]](#)
- Keraghel, I.; Morbieu, S.; Nadif, M. Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study. *arXiv* **2024**, arXiv:2401.10825.
- Sakai, T.; Chiwata, N.; Mine, T. Named Entity Recognition with Clue-Word Tags From Patent Documents in Materials Science. *IEEE Access* **2026**, *14*, 38332–38346. [\[CrossRef\]](#)
- Vaucher, A.C.; Zipoli, F.; Geluykens, J.; Nair, V.H.; Schwaller, P.; Laino, T. Automated extraction of chemical synthesis actions from experimental procedures. *Nat. Commun.* **2020**, *11*, 3601. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jiang, L.; Goetz, S.M. Natural language processing in the patent domain: A survey. *Artif. Intell. Rev.* **2025**, *58*, 214. [\[CrossRef\]](#)
- Wang, Y.; Yu, B.; Zhu, H.; Liu, T.; Yu, N.; Sun, L. Discontinuous Named Entity Recognition as Maximal Clique Discovery. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online, 1–6 August 2021; Volume 1, pp. 764–774. [\[CrossRef\]](#)
- Wang, J.; Shou, L.; Chen, K.; Chen, G. Pyramid: A Layered Model for Nested Named Entity Recognition. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 5918–5928. [\[CrossRef\]](#)
- Cabral, R.C.; Han, S.C.; Alhassan, A.; Batista-Navarro, R.; Nenadic, G.; Poon, J. TriG-NER: Triplet-Grid Framework for Discontinuous Named Entity Recognition. In *WWW'25: Proceedings of the ACM on Web Conference 2025, Sydney, Australia, 28 April–2 May 2025*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 2824–2837. [\[CrossRef\]](#)
- Ramshaw, L.A.; Marcus, M.P. Text Chunking Using Transformation-Based Learning. In *Proceedings of the Third Workshop on Very Large Corpora*; Massachusetts Institute of Technology: Cambridge, MA, USA, 1995; pp. 82–94.
- Ratinov, L.; Roth, D. Design Challenges and Misconceptions in Named Entity Recognition. In Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL), Boulder, CO, USA, 4–5 June 2009; pp. 147–155. [\[CrossRef\]](#)
- Lee, K.; He, L.; Lewis, M.; Zettlemoyer, L. End-to-end Neural Coreference Resolution. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP), Copenhagen, Denmark, 7–11 September 2017; pp. 188–197. [\[CrossRef\]](#)
- Fu, J.; Huang, X.; Liu, P. SpanNER: Named Entity Re-/Recognition as Span Prediction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL-IJCNLP), Online, 1–6 August 2021; pp. 7183–7195. [\[CrossRef\]](#)
- Wang, S.; Sun, X.; Li, X.; Ouyang, R.; Wu, F.; Zhang, T.; Li, J.; Wang, G.; Guo, C. GPT-NER: Named Entity Recognition via Large Language Models. In Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025, Albuquerque, NM, USA, 29 April–4 May 2025; pp. 4257–4275. [\[CrossRef\]](#)
- Zhou, W.; Zhang, S.; Gu, Y.; Chen, M.; Poon, H. UniversalNER: Targeted Distillation from Large Language Models for Open Named Entity Recognition. *arXiv* **2024**, arXiv:2308.03279.
- Kim, E.; Huang, K.; Tomala, A.; Matthews, S.; Strubell, E.; Saunders, A.; McCallum, A.; Olivetti, E. Machine-learned and codified synthesis parameters of oxide materials from scientific literature. *Sci. Data* **2017**, *4*, 170127. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gupta, T.; Zaki, M.; Krishnan, N.M.A.; Mausam, M. MatSciBERT: A materials domain language model for text mining and information extraction. *npj Comput. Mater.* **2022**, *8*, 102. [\[CrossRef\]](#)

25. Trewartha, A.; Walker, N.; Huo, H.; Lee, S.; Cruse, K.; Dagdelen, J.; Dunn, A.; Persson, K.A.; Ceder, G.; Jain, A. Quantifying the advantage of domain-specific pre-training on named entity recognition tasks in materials science. *Patterns* **2022**, *3*, 100488. [CrossRef] [PubMed]
26. Mavračić, J.; Court, C.J.; Isazawa, T.; Elliott, S.R.; Cole, J.M. ChemDataExtractor 2.0: Autopopulated Ontologies for Materials Science. *J. Chem. Inf. Model.* **2021**, *61*, 4280–4289. [CrossRef] [PubMed]
27. Huang, Z.; He, L.; Yang, Y.; Li, A.; Zhang, Z.; Wu, S.; Wang, Y.; He, Y.; Liu, X. Application of machine reading comprehension techniques for named entity recognition in materials science. *J. Cheminform.* **2024**, *16*, 76. [CrossRef] [PubMed]
28. Foppiano, L.; Lambard, G.; Amagasa, T.; Ishii, M. Mining experimental data from materials science literature with large language models: An evaluation study. *Sci. Technol. Adv. Mater. Methods* **2024**, *4*, 2356506. [CrossRef]
29. Potu, S.T.; Niranjana Murthy, R.; Thomas, A.; Mishra, L.; Prange, N.; Durmaz, A.R. Ontology-conformal recognition of materials entities using language models. *Sci. Rep.* **2025**, *15*, 18597. [CrossRef] [PubMed]
30. He, P.; Liu, X.; Gao, J.; Chen, W. DeBERTa: Decoding-enhanced BERT with Disentangled Attention. *arXiv* **2021**, arXiv:2006.03654.
31. Luan, Y.; He, L.; Ostendorf, M.; Hajishirzi, H. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 31 October–4 November 2018; pp. 3219–3232. [CrossRef]
32. Wadden, D.; Wennberg, U.; Luan, Y.; Hajishirzi, H. Entity, Relation, and Event Extraction with Contextualized Span Representations. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 5784–5789. [CrossRef]
33. Hosseini-Asl, E.; McCann, B.; Wu, C.S.; Yavuz, S.; Socher, R. A Simple Language Model for Task-Oriented Dialogue. In *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020*; Curran Associates Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 20179–20191.
34. Soares, L.B.; FitzGerald, N.; Ling, J.; Kwiatkowski, T. Matching the Blanks: Distributional Similarity for Relation Learning. *CoRR* **2019**, abs/1906.03158.
35. Zhou, W.; Chen, M. An Improved Baseline for Sentence-level Relation Extraction. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, Online, 20–23 November 2022*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2022; Volume 2, pp. 161–168. [CrossRef]
36. Sainz, O.; García-Ferrero, I.; Agerri, R.; Lacalle, O.; Rigau, G.; Agirre, E. GoLLIE: Annotation Guidelines improve Zero-Shot Information-Extraction. In Proceedings of the International Conference on Learning Representations 2024, Vienna, Austria, 7–11 May 2024; Volume 2024, pp. 47083–47107.
37. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv* **2019**, arXiv:1907.11692.
38. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4171–4186. [CrossRef]
39. Nakayama, H. Seqeval: A Python Framework for Sequence Labeling Evaluation. 2018. Available online: <https://github.com/chakki-works/seqeval> (accessed on 19 April 2026).
40. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. *arXiv* **2017**, arXiv:1711.05101.
41. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, N.; Polosukhin, I. Attention Is All You Need. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: New York, NY, USA, 2017; Volume 30, pp. 5998–6008.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.