The Materials Science Procedural Text Corpus: Annotating Materials Synthesis Procedures with Shallow Semantic Structures

Sheshera Mysore[1*] **Zach Jensen**[2*] **Edward Kim**[2] **Kevin Huang**[2]
Haw-Shiuan Chang[1] **Emma Strubell**[1] **Jeffrey Flanigan**[1]
Andrew McCallum[1] **Elsa Olivetti**[2]

[1]College of Information and Computer Sciences
University of Massachusetts Amherst
{smysore, hschang, strubell, jflanigan, mccallum}@cs.umass.edu

[2]Department of Materials Science and Engineering
Massachusetts Institute of Technology
{zjensen, edwardk, kjhuang, elsao}@mit.edu

Abstract

Materials science literature contains millions of materials synthesis procedures described in unstructured natural language text. Large-scale analysis of these synthesis procedures would facilitate deeper scientific understanding of materials synthesis and enable automated synthesis planning. Such analysis requires extracting structured representations of synthesis procedures from the raw text as a first step. To facilitate the training and evaluation of synthesis extraction models, we introduce a dataset of 230 synthesis procedures annotated by domain experts with labeled graphs that express the semantics of the synthesis sentences. The nodes in this graph are synthesis operations and their typed arguments, and labeled edges specify relations between the nodes. We describe this new resource in detail and highlight some specific challenges to annotating scientific text with shallow semantic structure. We make the corpus available to the community to promote further research and development of scientific information extraction systems.

1 Introduction

Systematically reducing the time and effort required to synthesize novel materials remains one of the grand challenges in materials science. Massive knowledge bases which tabulate known chemical reactions for *organic* chemistry (Lawson et al., 2014) have accelerated data-driven synthesis planning and related analyses (Segler et al., 2018; Coley et al., 2017). Automated synthesis planning for organic molecules has recently achieved human-level

In a typical procedure for the synthesis of β-MnO$_2$ nanowires, **2.5 mL** of **50 wt% Mn(NO$_3$)$_2$ solution** was **diluted** to **25.0 ml**, and **ozone** was fed into the bottom of the **solution** for **30 min** under vigorous **stirring**. With the indraught of **ozone**, **black solid** **appeared** gradually and the **clear solution** **turned** into **black slurry** finally. Then the **suspension** was **transferred** into an **autoclave** of **48.0 ml**, **sealed** and **maintained** at **200ºC** for **8 h**. After this, the **autoclave** was **cooled** to **room temperature** naturally. The resulting **solid products** were **washed** with **water**, and **dried** at **120ºC** for **8 h**. The obtained products were collected for the following characterization.

Figure 1: Example synthesis procedure text from a materials journal article (Dong et al., 2009). **Bold red** indicates the operations (predicates) involved in the synthesis; **bold black** indicates arguments; underlines demarcate entity boundaries.

planning performance using massive organic reaction knowledge bases as training data (Segler et al., 2018). There are, however, currently no comprehensive knowledge bases which systematically document the methods by which *inorganic* materials are synthesized (Kim et al., 2017a,b). Despite efforts to standardize the reporting of chemical and materials science data (Murray-Rust and Rzepa, 1999), inorganic materials synthesis procedures continue to reside as natural language descriptions in the text of journal articles. Figure 1 presents an example of such a synthesis procedure. To achieve similar success for inorganic synthesis as has been achieved for organic materials, we must develop new techniques for automatically extracting structured representations of materials synthesis procedures from the unstructured narrative in scientific papers (Kim et al., 2017b).

To facilitate the development and evaluation of
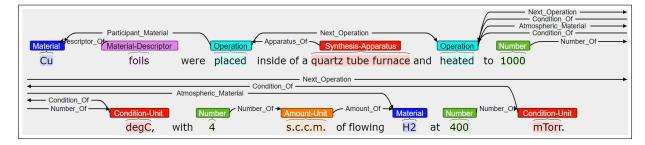
---

*Equal contribution

Figure 2: An example annotated sentence. Shallow semantic structures generally consist of verbal predicates and arguments of these predicates as nodes and labeled edges between predicate and argument nodes, example. Heated(*Condition_of*: degC, *Atmospheric_Material*: H2, *Condition_of*: mTorr). We also label relations between argument entities and non-predicate entities, for example. *Descriptor_of* (Cu, foils) and relations between predicates, for example. *Next_Operation*(placed, heated).

machine learning models for automatic extraction of materials syntheses from text, in this work we present a new dataset of synthesis procedures annotated with semantic structure by domain experts in materials science. We annotate each step in a synthesis with a structured frame-semantic representation, with all the steps in a synthesis making up a Directed Acyclic Graph (DAG). The types of nodes in the graph include synthesis operations (i.e. predicates), and the materials, conditions, apparatus and other entities participating in each synthesis step. Labeled edges represent relationships between entities, for example *Condition_of* or *Next_Operation*. Our dataset consists of 230 synthesis procedures annotated with these structures. An example sentence level annotation is given in Fig. 2. We make the corpus available to the community to promote further research and development of scientific information extraction systems for procedural text.[1]

2 Description of the Annotated Dataset

Here we describe the manner in which synthesis procedures were chosen for annotation (§2.1), present a description of the structures we annotate (§2.2), summarize key statistics of the dataset (§2.3), highlight specific annotation decisions (§2.4) and present inter-annotator agreements (§2.5). All annotations were performed by three materials scientists using the BRAT[2] annotation tool (Stenetorp et al., 2012).

2.1 Selecting Synthesis Procedures for Annotation

The 230 synthesis procedures annotated were selected from our database of 2.5 million publications

describing materials synthesis. The database was built from agreements with major scientific publication companies. Synthesis procedure text were obtained by first parsing the HTML text of the full publications, then automatically identifying candidate synthesis paragraphs using a trained classifier. This paragraph classifier was trained on a set of manually labeled paragraph examples and has a F1 score of 90.2 on a held out test set.[3] The paragraphs selected by the classifier were manually verified as containing complete, valid materials synthesis procedures by domain experts. While a given synthesis procedure is most often a single paragraph, there are cases where it spans multiple paragraphs, we consider all the paragraphs to be a single synthesis procedure. All the semantic structures were then manually annotated in these selected synthesis procedures. Fig 1 depicts an example paragraph containing a synthesis procedure. In selecting a synthesis procedure for annotation, a small number of valid synthesis procedures ($\sim 20\%$) are ignored; this is done for the synthesis procedures which are not amenable to annotation from a sentence-level frame-semantic view of synthesis steps, or ones in which most entities in the synthesis do not agree with our definitions of operations and argument entities (see §2.4 for further discussion).

2.2 Structures Annotated

An annotated graph consists of nodes denoting the participants of synthesis steps and edges denoting relationships between the participants in the synthesis. Operation nodes define the main structure of the graph and the arguments for each operation include different materials, conditions and appara-

---

Entity type	Count
Material	4843
Number	4095
Operation	3786
Amount-Unit	1659
Condition-Unit	1621
Material-Descriptor	1430
Condition-Misc	535
Synthesis-Apparatus	490
Nonrecipe-Material	475
Brand	348

(a)

Operation-argument relations	*Recipe-target, Solvent-material, Atmospheric-material, Recipe-precursor, Participant-material, Apparatus-of Condition-of*
Non-operation entity relations	*Descriptor-of, Number-of, Amount-of, Apparatus-attr-of, Brand-of, Coref-of,*
Operation-Operation relations	*Next-operation*

(b)

Table 1: Entity types and relation labels annotated in our dataset. The table (a) depicts the 10 most frequent of the 21 entity types defined in our dataset, and the table (b) highlights the 14 relation labels among entities possible in our dataset.

tus. For annotating the text describing a synthesis procedure, we define a set of span-level labels that identify the operations and typed arguments in the text, i.e. the nodes of the graph. We also define a set of relationships between text spans, which label the edges of the synthesis graph. We detail these two kinds of labels next.

Span-level Labels: Each span is a sequence of tokens or characters which form one entity mention (for example. "quartz tube furnace"). Entity mentions are associated with *entity types* which specify a category/kind for the entity mention. Our dataset defines a total of 21 entity types, with the least frequent entity type occuring 32 times. The 10 most frequent entity types defined for our dataset are listed in Table 1a. We describe a notable subset of the entity types in more detail below alongside examples of their occurrence in text. In examples, the text underlined is the span to be annotated.

Material: Materials that are used in the synthesis of the target. For example: Cr_2O_3, Strontium carbonate, $BaTiO_3$, Li_2CO_3, Water, Ethanol.

Nonrecipe-Material: Chemically specified materials that are not used in the synthesis of the synthesis target. For example: "$BaTiO_3$ powder (Ba/Ti=0.999)", "Li_2CO_3 was used as the Li source", "Si/Al ratio was 5".

Operation: Discrete actions physically performed by the researcher or discrete process steps taken to synthesize the target.

Material-Descriptor: Describes a material's structure, shape, form, type, role, etc. and must be directly or nearly adjacent to the material it describes. Does not include

amounts, concentrations, or purities of materials. For example: $CaCu_3Ti_4O_{12}$ compound, Copper ion, GaAs nanowires, Anatase TiO_2, Deionized water.

Meta: A canonical name to specify a particular overall synthesis method used for synthesis. For example: "Graphite oxide was prepared by oxidation of graphite powder according to the modified Hummers' method". "Bi2S3 nanorods with orthorhombic structure were prepared through the hydrothermal method". "Graphene oxide (GO) was prepared from graphite powder by the Staudenmaier method."

Amount-Unit: These describe absolute amounts, concentrations, purities, ratios, flow rates and so on. For example: mg, mL, M, %.
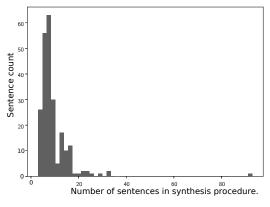
Condition-Unit: These describe the units of measurement for intangible conditions under which operations are performed. For ex: °C, K, sec, RPM, mW.
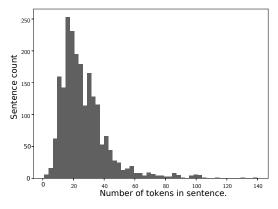
Condition-Misc: Qualitative descriptions of conditions. For example: Room temperature, Dropwise, Naturally, Vacuum.

Synthesis-Apparatus: Equipment used to perform an operation involved in the synthesis.

Characterization-Apparatus: Equipment used to characterize a materials properties.

Relation Labels: We define a set of relationships between entity mentions, which label the edges of the synthesis graph. A subset of these relations describe direct relationships between operations and their arguments, others describe re-

58

(a) Sentences per synthesis document.

(b) Tokens per sentence.

Figure 3: Sentences count statistics of the corpus; On average a synthesis procedure contains 9 sentences, each of which contain 26 tokens on average.

lationships between argument mentions, and the *Next-Operation* relation describes relationships between operations so as to step towards annotating full recipe graphs. The different relation labels we define are tabulated in Table 1b, a subset of these are defined below:

Recipe-target: Indicates a material assigned to an operation which is the target of the synthesis procedure.

Participant-material: A material that is part of a particular synthesis step.

Recipe-precursor: Indicates a material which is the source of an element for the target material used in a specific synthesis operation.

Apparatus-of: Denotes an apparatus to be used in a synthesis operation.

Condition-of: Denotes a reaction condition for a synthesis operation.

Coref-of: Intended to capture coreferent mentions of entities presented by abbreviations, text in parenthesis and so on. For example: "Air (O_2/N_2 mixture gases)" and "He were supplied to the porous support side ...". "Air" is coreferent with O_2 and N_2.

Amount-of: Links a number entity to the corresponding unit of measurement.

Next-operation: A relation intended to denote the true synthesis order of the synthesis steps; the relation is also intended to implicitly denote the flow of intermediate materials in the synthesis. However, in this first release of the data, as a placeholder for future annotations, *Next-Operation* is used simply used to indicate the next operation in text order rather than

in true synthesis order.

We refer readers to our annotation guidelines for definitions of the complete set of entity type and relation labels in the dataset.

2.3 Dataset Statistics

Some key statistics of the dataset such as number of documents, tokens, entities and unique operations are listed in Table 2 and Fig. 3. In reporting these statistics we perform tokenization and sentence segmentation using the ChemDataExtractor package (Swain and Cole, 2016).[4]

2.4 Annotation Decisions

Next we highlight specific points of contention in creating the current set of annotations.

What constitutes an operation?: While our definition of the `Operation` entity type specifies only actions performed by a lab researcher to be valid operations, there are cases where our annotations diverge from this definition. This happens in the following cases:

- Cases where an operation isn't explicitly performed by the researcher. For example: "After this, the autoclave was `cooled` to room temperature naturally".
- Cases with nested verb structures. For example: "white precipitate which was `harvested` by `centrifugation` ...".

In the current set of annotations, we allow experts to decide when a particular candidate operation

---

[4]`https://pypi.org/project/ChemDataExtractor/1.2.2/`

59

Item	Count
Synthesis procedures	230
Sentences	2113
Tokens	56510
Entity mentions	20849
Entities	4883
Unique operations	409
Entity types (Table 1a)	21
Relation types (Table 1b)	14
Avg. sentence length (Fig. 3b)	26
Avg. sentences/synthesis proc (Fig. 3a)	9

Table 2: Various dataset statistics. Additional details provided in referred figures. To determine unique operations, `Operation` entity mentions are lemmatized with the WordNet lemmatizer and the unique lemmas are counted.

should be considered valid and when it can be omitted. As our inter-annotator agreements will demonstrate, experts tend to agree often on what should be considered an operation. The question of what constitutes an operation is analogous to the notion of what constitutes an "event" in the broader NLP literature as highlighted by Mostafazadeh et al. (2016).

Argument state and argument re-use: Annotation of semantic structures often allow for argument spans to have multiple parents (Surdeanu et al., 2008; Banarescu et al., 2013; Oepen et al., 2015). For example in Figure 2, the material "Cu" could be considered an argument of the operations "placed" and "heated". Allowing for arguments to have multiple parents however runs into complications when the operation causes the state of a material to change (incidentally, this is not the case in the example we highlight above). When a materials state changes due to a specific operation, considering the same text span to be the argument of a different operation would not be chemically valid. For example, in the sentence:

> After that, the mixed solution was aged at 60 degC for 48 h, followed by heating at 900 degC for 2 h with a heating rate of 5 degC min-1 in an N2 atmosphere.

"solution" is labeled as *Participant-material* for "aged", but it isnt considered a *Participant-material* to "heating" since aging caused it to be a different material. Similarly, in:

> 1.6632 g lithium acetate was dissolved into 26 mL of ethanol-water mixture

(12:1 in volume) and slowly dropped into the above suspension.

"lithium acetate" is only labeled as *Participant-material* for "dissolved" and not for "dropped" whose sole argument is "suspension". This clearly highlights an instance of a material entirely absent from the text being the true argument of an operation. Therefore the current set of annotations does not allow for arguments to have multiple parents. Further, the tracking of state itself is also complicated by the difficulty in being able to write down precise states at a meaningful level of granularity for all possible materials, this is further complicated by the ambiguity presented by under-specified materials in synthesis text, for example in the sentence:

> With the indraught of ozone, black solid appeared gradually and the clear solution turned into black slurry finally.

Most of the entities, "black solid", "clear solution" and "black slurry" are chemically under-specified, with precise specification even unnecessary for describing the synthesis procedure.

Relations across sentences: Often, in synthesis procedures, a given synthesis step is described across multiple sentences. In these cases it would be meaningful to allow for relationships between operation-argument entities which are in different sentences. For the sake of simplicity and to stick more closely to a sentence level shallow semantic annotation, our current iteration of the annotations has avoided this annotation, however a very small number of instances of cross-sentence relations do exist ($< 1\%$ of all relations in the dataset). Examples of this type are as follows:

> First, sulfuric acid and nitric acid were mixed well by stirring 15 min in an ice bath, and then graphite powder was dispersed into the solution. After 15 min, potassium chlorate was added into the system - very slowly to prevent strong reaction during the oxidation process.

> Oxygen with 20 sccm flow rate and argon with 40 sccm flow rate were used as the sputtering gas. Growth temperature was 400 degC and the RF power was 90 W.

Here "min" and "dispersed" are related by a *Condition_Of* relation. Similarly, "degC" and "W",

Entity type	Fleiss' Kappa
Material	0.916
Number	0.971
Operation	0.859
Amount-Unit	0.967
Condition-Unit	0.985
Material-Descriptor	0.638
Condition-Misc	0.784
Synthesis-Apparatus	0.860
Nonrecipe-Material	0.371
Brand	0.862

(a)

Annotation	Fleiss' Kappa
Span-level labels	0.861
Relation labels	0.941

(b)

Table 3: Annotator agreements in our dataset. The table (a) depicts the percent agreements on 10 most frequent of the 21 entity types defined in our dataset, and the table (b) denotes overall agreements on the different annotations in our dataset.

both are annotated with *Condition_Of* relations to "used". Annotations of this kind were created when annotators deemed such an annotation absolutely necessary. Synthesis procedures which required annotation primarily of cross-sentence relations were ignored.

2.5 Inter-annotator Agreement

Next we report a host of inter-annotator agreements for the different levels of semantic annotation in our dataset. The agreements we report are based on a collection of 5 synthesis procedures which were annotated separately by all three expert annotators. All the numbers we report are Fleiss' Kappa scores for the 3 expert annotators.

Span-level Labels: Agreements on span level labels correspond to the agreement on entity type labels assigned to individual tokens. We observe the overall agreement on the token level labels to be 0.861. A break down of this agreement by the entity type is presented in Table 3a. As this indicates, there seems to be high agreement on labels which have clear definitions; namely. Number, Amount-Unit. Labels which by definition are a lot more ambiguous, however, have a lower agreement. The two entity types Material-Descriptor and Nonrecipe-Material see the lowest agree-

ments. We believe these to be inherently more subjective entity types. In the case of Material-Descriptor it is often that some annotators may consider the descriptor and the adjacent material to be Material in its own right, for example: "Deionized Water" may be considered as a material in its own right or "deionized" may be considered to be a descriptor. In the case of Nonrecipe-Material, a similarly harder decision needs to be made by the annotator, since these are materials which aren't involved in the synthesis but are still mentioned in the text for completeness information. Often it is up to the interpretation of the annotator to decide whether a material is indeed involved in the synthesis leading to the low agreement on this entity type.

Relation Labels: Agreement on relation labels were computed for the set of cases where a pair of annotators agreed on the token level annotations, this happens 66% of the time in our repeated annotations. For a pair of entities, if both annotators indicate the same relation type the annotators are considered to be in agreement. For relation labels we observe a agreement score of 0.941. Since we only consider cases where the token labels are in agreement, we believe that it is likely that when annotators agree on the token level annotations they also tend to agree on the relation level labels.

3 Related Work

Shallow semantic parsing in NLP: Prior work in the NLP community has defined and annotated semantic structures for text. These structured representations often seek to generalize about sentence level predicate-argument structure, abstracting away from the surface nuances of natural language and representing its semantics (Abend and Rappoport, 2017). A large body of work has created these resources for non-scientific text, as done in PropBank (Palmer et al., 2005; Surdeanu et al., 2008), FrameNet (Fillmore and Baker, 2010), AMR (Banarescu et al., 2013), semantic dependencies (Oepen et al., 2015) and ACE event schemas (Doddington et al., 2004). The GENIA project has defined event structures for biomedical data (Kim et al., 2003) while Garg et al. (2016) extended the AMR framework to biomedical text. Closer still to the work presented here, Mori et al. (2014) have annotated cooking recipes with sentence and discourse level semantic relations. There has also been an interest in labeling scientific wetlab proto-

col text, with semantic structures and to facilitate training supervised models for the extraction of these structures (Kulkarni et al., 2018). Kulkarni et al. make use of an altered version of the EXACT2 ontology, created for the annotation of biomedical procedural text (Soldatova et al., 2014). The dataset presented here can be viewed to fit within the theme of sentence level semantics for procedural text, specifically tailored to materials science synthesis.

Materials Science & Chemistry: Prior work in the materials science community have shown that manual extraction and subsequent text mining can be an effective approach to analysis of synthesis routes for specific compounds and classes of materials (Raccuglia et al., 2016; Ghadbeigi et al., 2015); these approaches however have been limited by scale due to the manual extraction step. There has also been strong a consensus that comprehensively extracting the knowledge contained within written inorganic materials syntheses is a key step towards reducing the overall discovery and development time for novel materials (Butler et al., 2018). We believe that the dataset we release fills an important gap in the existing work on extraction of inorganic materials synthesis procedures, by allowing exploration into extraction at a scale not attempted before. Parallel with this work, work by Kim et al. (2018) and Tamari et al. (2019) adopt the dataset released here to aid extraction of structured representations from synthesis procedures and with Kim et al. presenting early experiments in synthesis planning from extracted synthesis.

The focus of existing datasets and resources in the materials science community, has been on materials structures and properties knowledge bases (Jain et al., 2013), rather than reactions and synthesis. In pursuit of more scalable methods for materials synthesis data extraction, Young et al. (2018) have made use of automated methods for extracting specific categories of materials synthesis parameters, while Mysore et al. (2017) and Kim et al. (2017a) have both presented preliminary methodological explorations for automated extraction of elements of a synthesis graph from materials science literature. However, these lines of work have not presented general purpose annotated data with which to train information extraction models for extraction of structured synthesis representations at scale, the focus of this work.

4   Conclusion and Future Directions

In this work we present a shallow semantic parsing dataset consisting of 230 synthesis procedures. The dataset was annotated by domain experts in materials science. We also highlight specific difficulties in the annotation process and present agreement metrics on the different levels of our annotation. We believe the dataset will enable the development of robust supervised entity tagging models and is suitable for evaluating models trained to extract shallow semantic structures. This is evidenced by the adoption of the dataset by work contemporaneous with this work (Kim et al., 2018; Tamari et al., 2019).

Future work in the development of this dataset could involve methods for the scaling up of the annotation process, perhaps by adapting the guidelines to enable annotation by non-experts at some stages of the annotation process. Further, we also plan to quantitatively establish the limits of our annotation schema for the kinds of information it isn't able to capture. We also plan to add additional layers of annotation, including: co-reference relations between synthesis steps, states of argument entities, and linking annotated entities to entries in materials science knowledge bases such as The Materials Project.[5]

5   Acknowledgements

References

Omri Abend and Ari Rappoport. 2017. The state of the art in semantic representation. In *ACL*.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*.

---

[5] https://materialsproject.org/

Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. 2018. Machine learning for molecular and materials science. *Nature*.

Connor W Coley, Regina Barzilay, Tommi S Jaakkola, William H Green, and Klavs F Jensen. 2017. Prediction of organic reaction outcomes using machine learning. *ACS central science*.

George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie Strassel, and Ralph M Weischedel. 2004. The automatic content extraction (ace) program-tasks, data, and evaluation. In *LREC*.

Yuming Dong, Hongxiao Yang, Kun He, Shaoqing Song, and Aimin Zhang. 2009. β-mno 2 nanowires: a novel ozonation catalyst for water treatment. *Applied Catalysis B: Environmental*.

Charles J Fillmore and Collin Baker. 2010. A frames approach to semantic analysis. In *The Oxford handbook of linguistic analysis*.

Sahil Garg, Aram Galstyan, Ulf Hermjakob, and Daniel Marcu. 2016. Extracting biomolecular interactions using semantic parsing of biomedical text. In *Thirtieth AAAI Conference on Artificial Intelligence*.

Leila Ghadbeigi, Jaye K Harada, Bethany R Lettiere, and Taylor D Sparks. 2015. Performance and resource considerations of li-ion battery electrode materials. *Energy & Environmental Science*.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. 2013. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *Apl Materials*.

Edward Kim, Kevin Huang, Adam Saunders, Andrew McCallum, Gerbrand Ceder, and Elsa Olivetti. 2017a. Materials synthesis insights from scientific literature via text extraction and machine learning. *Chemistry of Materials*.

Edward Kim, Kevin Huang, Alex Tomala, Sara Matthews, Emma Strubell, Adam Saunders, Andrew McCallum, and Elsa Olivetti. 2017b. Machine-learned and codified synthesis parameters of oxide materials. *Scientific Data*.

Edward Kim, Zach Jensen, Alexander van Grootel, Kevin Huang, Matthew Staib, Sheshera Mysore, Haw-Shiuan Chang, Emma Strubell, Andrew McCallum, Stefanie Jegelka, et al. 2018. Inorganic materials synthesis planning with literature-trained neural networks. *arXiv preprint arXiv:1901.00032*.

J-D Kim, Tomoko Ohta, Yuka Tateisi, and Junichi Tsujii. 2003. Genia corpusa semantically annotated corpus for bio-textmining. *Bioinformatics*.

Chaitanya Kulkarni, Wei Xu, Alan Ritter, and Raghu Machiraju. 2018. An annotated corpus for machine reading of instructions in wet lab protocols. *NAACL*.

Alexander J Lawson, Jürgen Swienty-Busch, Thibault Géoui, and David Evans. 2014. The making of reaxystowards unobstructed access to relevant chemistry information. In *The Future of the History of Chemical Information*.

Shinsuke Mori, Hirokuni Maeta, Yoko Yamakata, and Tetsuro Sasada. 2014. Flow graph corpus from recipe texts. In *LREC*, pages 2370–2377.

Nasrin Mostafazadeh, Alyson Grealish, Nathanael Chambers, James Allen, and Lucy Vanderwende. 2016. Caters: Causal and temporal relation scheme for semantic annotation of event structures. In *Proceedings of the Fourth Workshop on Events*, pages 51–61.

Peter Murray-Rust and Henry S Rzepa. 1999. Chemical markup, xml, and the worldwide web. 1. basic principles. *Journal of Chemical Information and Computer Sciences*.

Sheshera Mysore, Edward Kim, Emma Strubell, Ao Liu, Haw-Shiuan Chang, Srikrishna Kompella, Kevin Huang, Andrew McCallum, and Elsa Olivetti. 2017. Automatically extracting action graphs from materials science synthesis procedures. *Workshop on Machine Learning for Molecules and Materials at NIPS*.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinková, Dan Flickinger, Jan Hajic, and Zdenka Urešová. 2015. Semeval 2015 task 18: Broad-coverage semantic dependency parsing. *SemEval-2015*, page 915.

Martha Palmer, Paul Kingsbury, and Daniel Gildea. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*.

Paul Raccuglia, Katherine C Elbert, Philip DF Adler, Casey Falk, Malia B Wenny, Aurelio Mollo, Matthias Zeller, Sorelle A Friedler, Joshua Schrier, and Alexander J Norquist. 2016. Machine-learning-assisted materials discovery using failed experiments. *Nature*.

Marwin HS Segler, Mike Preuss, and Mark P Waller. 2018. Planning chemical syntheses with deep neural networks and symbolic ai. *Nature*.

Larisa N Soldatova, Daniel Nadis, Ross D King, Piyali S Basu, Emma Haddi, Véronique Baumlé, Nigel J Saunders, Wolfgang Marwan, and Brian B Rudkin. 2014. Exact2: the semantics of biomedical protocols. *BMC bioinformatics*.

Pontus Stenetorp, Sampo Pyysalo, Goran Topi, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. BRAT: A web-based tool for NLP-assisted text annotation.

Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The conll-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pages 159–177. Association for Computational Linguistics.

Matthew C Swain and Jacqueline M Cole. 2016. Chemdataextractor: a toolkit for automated extraction of chemical information from the scientific literature. *Journal of chemical information and modeling*, 56(10):1894–1904.

Ronen Tamari, Hiroyuki Shindo, Dafna Shahaf, and Yuji Matsumoto. 2019. Playing by the book: An interactive game approach for action graph extraction from text. In *Workshop on Extracting Structured Knowledge from Scientific Publications at NAACL 2019*.

Steven R Young, Artem Maksov, Maxim Ziatdinov, Ye Cao, Matthew Burch, Janakiraman Balachandran, Linglong Li, Suhas Somnath, Robert M Patton, Sergei V Kalinin, et al. 2018. Data mining for better material synthesis: The case of pulsed laser deposition of complex oxides. *Journal of Applied Physics*.