MuLMS Corpus
Annotation Guidelines: Argumentative Zoning

Matteo Finco **Maira Indrikova** **Felix Hildebrand** **Annemarie Friedrich**

Contents

1 Overview

This document details guidelines for creating **argumentative zoning** annotations on scientific papers in the materials science domain.

1.1 Motivation

Argumentative zones are **functional sentence types**, i.e., they capture the rhetorical function of the sentence as intended by the author. For example, the function of a sentence may be to explain the research's aim, results obtained in the current study, giving background on related work, or stating a conclusion. Argumentative zoning of scientific articles has been proposed by Teufel et al. [1999] and variations thereof have been applied to a variety of domains in the past, including computational linguistics, computer science, biomedical, and chemistry.

Analysing scientific papers sentence-wise according to their argumentative zones is helpful because it allows **summarizing** papers from various perspectives, e.g., extracting the motivation, experimental results, or conclusions. This could be helpful for displaying relevant paragraphs during search, or narrowing parts of an article for information extraction in order to make the latter more precise.

In this project, we developed argumentative zoning guidelines that are **specific to the materials science domain**. The main differences versus existing annotation schemes is a more fine-grained labeling regarding the descriptions of experimental setups, differentiating for example the preparation process and the characterization of materials from results or explanations of the observed results.

1.2 Materials Science Study Types Covered

The argumentative zoning annotation guidelines have been developed aiming to the interpretation of scientific papers in the field of materials science. In particular, experimental and simulation works were taken in consideration for tailoring the definitions following in this document. In experimental papers, the authors perform physical experiments, while in simulation papers the authors perform computations to simulate materials properties that they later verify realizing experiments.

1.3 Annotation Layers

The basic unit for annotation (markables) are sentences. Sentence segmentation is performed automatically prior to annotation. In general, annotators are instructed to choose the best-fitting label for each sentence, but it is also allowed to assign more than one label if several types of information are present in the sentence. In cases where annotators feel it makes sense to split the sentence into several spans and assign labels accordingly, they may do so. **As a post-processing step to prepare for automatic modeling, we map all labels occurring in a sentence to the entire sentence, i.e., we perform a multi-label classification task on sentence-level.**

1.4 Use of Labels and Subtypes

Our annotation scheme is hierarchical, i.e. some of the labels have subtypes [Example 1]. Each span can have multiple annotations, which may also include different labels on the highest level in the taxonomy. It is also possible to use several subtypes for a label in the same span [Example 2].

1. However, researchers found that high Pt loading will result in smaller ECSA.
   BACKGROUND / PRIOR WORK
2. Finally, the working electrode was purged with N_2 again for 1 min,
   EXPERIMENTAL/ PREPARATION
   and CO stripping was carried out by conducting a CV scan between 0.05 and 0.9 V.
   EXPERIMENTAL/ CHARACTERIZATION

1.5 Labels and Subtypes

There are two sets of labels, grouped by whether their function is to represent content information or structural information.

Content information labels and subtypes are:

- MOTIVATION
- BACKGROUND
  - PRIOR WORK
- EXPERIMENTS
  - CHARACTERIZATION
  - PREPARATION
- SYNTHESIS PROCEDURE
- EXPLANATION
  - ASSUMPTION
  - HYPOTHESIS
- RESULTS
- CONCLUSION

Structure information labels are:

- HEADING
- ABSTRACT
- METADATA
- CAPTION
- FIGURE/TABLE

2 Content Information Labels

This set of labels refers to the function of a sentence according to its *information content*.

2.1 MOTIVATION

Includes the aims of the study overall [Example 1], non-scientific background (like socio-economical premises) [Example 2] and domain-specific reasons that motivate the study [Example 3]. Rhetorical questions on objectives of the study [Example 4]. Organization of the paper [Example 5].

1. In this study, we perform a systematic analysis of the underlying microstructure of platinum-cobalt catalyst dispersed on various HSC supports in terms of their surface area and pore-size distribution.
2. Increased energy consumption and forefront global issues have led to renewed focus on energy generation and storage.
3. This technique is applied to $La_{0.6}Sr_{0.4}CoO_{3-\delta}$ (LSC), one of the most promising SOFC/SOEC oxygen electrode material.
4. The following scientific questions can therefore be addressed by further film growth and IPLD analysis: i. What is the absolute value of the oxygen surface exchange resistance of LSC (RLSC, surf exch) immediately after preparation?
5. The rest of the paper is organized as follows.

2.2 BACKGROUND

Domain knowledge including technical knowledge [Example 1-2] such as the description of processes and technologies, description of textbook-level formulas [Example 3] and information on the literature's state of advancement [Example 4].

1. The substrate thickness must be less than 50 microns to gain an advantage of easy peel off, after the hot-press of MEA.
2. In fuel cell system, balance of plant (BOP) and stack development encompass 71% of the fuel cell system cost.
3. The method is based on the Kelvin equation and accounts for the formation of a condensed layer of gas molecules on the pore wall prior to capillary condensation using the Harkins-Jura equation:

$$t = \sqrt{\frac{13.99}{0.034 - \ln \frac{p}{p_0}}} \tag{1}$$

   where, t is the thickness of the adsorbed layer on the pore wall (Å), p the partial pressure of N_2, and p0 the saturation pressure of N_2 at 77 K.
4. Furthermore, to the best of our knowledge, no literature is available related to ideal decal substrates for the complete transfer of catalyst layer to the Nafion.

Do not include listing of techniques used to realize the study and mechanisms used to describe phenomena.

1. In this paper, the electrochemical performance of Pt/C catalyst electrodes with the same I/C ratio and different Pt loadings were measured by the RDE system. (EXPERIMENT)

2.3 BACKGROUND/PRIOR WORK

This label is used to mark information on prior work that is of specific relevance to the work presented in the current paper. It is distinguished from the more general BACKGROUND by referring to a specific study. The reference can be either explicit through the use of citations [Example 1,2] or implicit in the context, in the latter it possible to find explicit references in the surrounding sentences [second sentence in Example 3].

1. Among the various techniques of MEA development, studies have shown that decal technique has improved efficiency over spray and other conventional coating techniques[12,13].
2. Irmawati et al. [15] has concluded that PTFE cloth and film furnish only 90% catalyst transfer.

3. A recent study revealed that LSC surfaces might be very inhomogeneous in terms of their oxygen exchange current density, with only a few highly active, presumably Co related reaction sites in an otherwise less active Sr-terminated surface environment.[19,21] [BACKGROUND/PRIOR WORK]
   It was shown that minor surface composition changes, hardly detectable by analytical methods, had a significant impact on the oxygen exchange kinetics. [BACKGROUND/PRIOR WORK]

2.4 EXPERIMENTS

Sentences in which experimental features are presented or discussed [Example 1,2]. For more examples refer to the subtypes' examples. In simulation papers, this label is applied to sentences that report and describe the methods of execution of the simulations.

1. Specifically, we evaluate PtCo nanoparticle catalyst distributed on six different carbon black supports of varying micro-, meso-, and macro-porosity.
2. We first consider RWE_{sheet} in more detail.

2.5 EXPERIMENTS/PREPARATION

Sentences that describe the steps and devices for the preparation of samples [Example 1]. Sentences containing information about materials relevant to the preparation of the experiments reported in the current paper [Example 2, 3]. Also sample labelling is included in this section [Example 4]. Note that the label EXPERIMENTS/PREPARATION is **not** used for every sentence that contains a material, but only the ones in the experimental section. However, sentences containing insights and descriptions of the preparation procedure can be labelled as such throughout the document.
In simulation works, the preparation of the experiment includes the setting up of the simulation including the input parameters (like cut-off energy, meshing of the k-points, etc.), the boundary conditions and the adaptation of formulas to the latter.
For modeling papers, this label includes also the training steps and the description of the functionals adopted. In case that information is generally known or previous work is cited, an appropriate second label is apposed.

1. The mixture was subjected to stirring for 60 minutes, sonication for 30 minutes and stirring again for 30 minutes.
2. Other decal substrates such as aluminium foil, stainless steel, perfluoroalkoxy polymer, polytetrafluoroethylene, Kapton HN (Polyimide), and Kapton FN (80% polyimide and 20% fluorinated ethylene propylene (FEP)) etc. have also been considered with partial catalyst transfer and low transfer efficiency (Table 1).
3. Nafion®212 polymer electrolyte membrane was obtained from Fuel cell store, Nafion solution (10 wt.%) was obtained from Sainergy Fuel cell India PVT.LTD., Serpentine flow channel patterned, single cell PEMFC unit was obtained from Wonatech for Hydrogen-Oxygen test fixture.
4. The carbon black supports were classified based on their surface area and are labeled as HSC-a, HSC-b, HSC-e, HSC-f, HSC-g and MSC-a where HSC and MSC stand for high surface area and medium surface area carbons, respectively.

2.6 EXPERIMENTS/CHARACTERIZATION

This label covers sentences that report information about characterization techniques, methods and conditions used in the study [Example 2]. It includes information on how data are extracted, like in [Example 3] where the error bar amplitude (the measure) is taken from a parameter.
In simulation papers, this includes the computational domain and any other information about the performing of the simulation, since the simulation itself consist in the characterization of the model that is the object of the study. Therefore it includes descriptions of optimizations and sampling.

1. To understand these interactions, structure-property relations are evaluated using *contact angle and spectroscopic studies*.
2. Bulk proton transport resistance of the catalyst layer was measured using an electrochemical impedance spectra (EIS) methodology.[34,43] *This involved the flow of H_2 on the anode and N_2 on the cathode at 80 °C and various RH values from 20% to 122%.*
3. The error bars in the plots indicate the standard deviation of the measurements.

The label does not include background information on characterization techniques such as in the example below.

1. Normally, in order to reduce the effect of mass transfer resistance, the study of oxygen reduction for Pt/C catalysts by RDE is mainly focused on the low Pt loading range 4-36μ g cm$_{-2}$.[13] [BACKGROUND/PRIOR WORK]
   However, the actual Pt loading in cathode for PEMFC is usually up to $400\mu g\ cm^{-2}$, which is a large difference in the oxygen reduction test by RDE.[BACKGROUND/PRIOR WORK]
   Therefore, the traditional RDE test does not fully reflect the actual catalytic activity of the cathode catalyst in PEMFC. [BACKGROUND/PRIOR WORK]

2.7 EXPLANATION

Mechanisms [Example 1] and assumptions that are specific to the study itself [Example 2]. Note that this label only applies to **explanations** of the mechanisms etc. **pertaining to the work presented in the paper to be annotated**! General text-book like information will be marked as BACKGROUND.

1. When the electrode thickness is very thin, the H^+ in solution can be quickly penetrated into the electrode, the impact of mass transfer can be ignored; when the electrode thickness increases gradually, the decrease of the H^+ permeation rate will cause the shortage of H^+ in the Pt catalyst surface, which would lower the measured electrode potential. [EXPLANATION]
2. In our calculation, all Pt loadings in the electrode were considered to be electrochemically active. [EXPLANATION/ASSUMPTION]

2.8 EXPLANATION/ASSUMPTION

Approximations that are specific of the study itself. Assumptions that lead to the formulation of hypothetical scenarios.

1. The impact of scan rates can be ignored because the scan rate used in this experiment (50mV/s) was very slow.

2.9 EXPLANATION/HYPOTHESIS

Sentences in which the author hypothesize an explanation to justify empirical observation without supplying evidence for it. The label is used to mark post-hoc explanations of observations in experiments (but not final conclusions).

1. During hot-press conditions of MEA, thermal decomposition of the decal substrate generate many products and side products that in turn may poison the catalyst/electrode.

2.10 RESULTS

This label is used to mark experimental results of the current paper [Example 1], trends in data [Example 2], and content of images [Example 3, 4].

1. The hydrogen adsorption/desorption peak is at about 0.2V, which is consistent with previous literature. [15]
2. At low Pt loading region ($< 100\mu g\ cm^{-2}$, or electrode thickness <2.17 μm), we observed significant hydrogen adsorption/desorption peaks with longitudinal symmetry.
3. Figure 1a shows the N_2 adsorption pore size distribution of PtCo nanoparticle catalysts on six different carbon supports that have been investigated in this study.
4. The micropore surface area is then calculated from the difference between the total BET surface area and the external surface area
   $$S_{micro} = S_{BET} - S_{ext} \tag{2}$$
   These values are tabulated in Table I.

2.11 CONCLUSION

This label marks sentences that provide overall judgements or findings based on the results of the paper itself. This includes assumptions that are confirmed by experimental data [Example 1], or introduction of new concepts to describe the results [Example 2]. These sentences are the ones that carry the takeaway message of the paper. The label can also apply to sentences within the "highlights" section, which is, for some of the publications, placed right after the abstract and contains snippets of the conclusions of the work.

1. This result indicated that the adsorption/desorption process of hydrogen on the electrode surface is reversible, and there is no mass transfer.
2. The concept of effective Pt loading in high Pt loading electrode ($> 100\mu g\ cm^{-2}$, or electrode thickness $> 2.17\mu m$) was proposed and could be defined as the ratio of ECSA for high to low Pt loading.

3 Structure Information Labels

3.1 ABSTRACT

Sentences that constitutes the abstract.

3.2 HEADING

Headings present in the article which include the title of it and the title of each paragraph even if a paragraph is introduced only with a number.

3.3 METADATA

All the metadata present in the text. Which consists in ancillary information like editor, authors, author's email, DOI, licensing, acknowledgements, hosting institution or company etc.

3.4 CAPTION

Captions of figures and tables.

3.5 FIGURE/TABLE

This label is applied whenever the text content of a figure or a table has been transcripted into the annotation text.

4 List of Interesting Examples

1. A series of Pt/C catalyst electrodes with different Pt loadings ranging from $12\mu g\ cm^{-2}$ to $400\mu g\ cm^{-2}$ was characterized in our experiments by rotational disk electrode (RDE) system. *EXPERIMENTAL/CHARACTERIZATION, MOTIVATION*

   Being the first sentence of the paper, it is meant to carry the aims of the study, but at the same time it is important to add labels compatible with all the relevant information included.

2. Therefore, the traditional RDE test does not fully reflect the actual catalytic activity of the cathode catalyst in PEMFC. *BACKGROUND*

   Results from prior work are labelled as BACKGROUND/PRIOR WORK and not as RESULTS in order not to poison the results obtained by the experimental work of the paper.

3. In this paper, the reaction current at 0.9V is divided by the true surface area and the actual mass of the Pt to obtain the specific activity (SA). The reaction current at 0.9V is divided by the actual mass of the Pt to obtain the mass activity (MA).

   a. Formulas, and their descriptions, are sorted into BACKGROUND it they are textbook-level information or into EXPERIMENTS/CHARACTERIZATION if they are specific for the paper. In this case it is a textbook-level information.

   b. If authors adopt weird wording to introduce sentences, like "In this paper" while referring to general conventions, it is better to label manually in order to avoid the introduction on contradictory information. The reason for doing so is that during the training these words can be mistaken by the machine as referring to paper-specific sentences which are comprised in the EXPLANATION label.

4. In this context, pore size distribution shown in Fig. 1a is divided into three different regions for the specific application here as fuel cell catalysts—micropores (8 nm), mesopores (2 to 8 nm) and macropores (>8nm). *EXPERIMENTAL/CHARACTERIZATION*

5. The carbon black supports were classified based on their surface area and are labeled as HSC-a, HSC-b, HSC-e, HSC-f, HSC- g and MSC-a where HSC and MSC stand for high surface area and medium surface area carbons, respectively.

   Where do we put the description of the tested materials reported in the results section?

6. ECSA of the catalyst, shown in Table II, measured using HUPD method is a direct reflection of the PtCo particle sizes with the various HSC catalysts yielding 40–50 m^2gPt^{-1} and MSC yielding 26 m^2gPt^{-1}.

   a. In my opinion it is a result, do we additionally put exp-characterization tag as it mentions the method used? Or we label methods just in experimental section? (similar to exp-materials examples in the past)

7. The utilization of PtCo nanoparticles present inside the micro/mesoporous regions can be quantified using electrochemical CO adsorption/stripping measurements that require water as a reactant which can either be furnished by the condensed water molecules in the catalyst layer or by the bound/unbound water in the ionomer.

   $Pt-CO_{ads} + H_2O \longrightarrow CO_2{}^+ 2\,H^+ + 2\,e^-$

   Sentence within the results paragraph, motivation for characterization technique and background? Not entirely sure

8. CO-stripping measurements as a function of RH was carried out to understand the nature of PtCo distribution in the internal and external surface of the carbon primary particles.

   a. In results section, I would classify it as experimental characterization, but also motivation; shall we label motivation within the whole paper or just from introduction text? Discussed: Motivation tag will be used on a higher level, motivation for the research, but not for the method, plots, why they did smth, ... (will be updated to do not)

9. Similarly, R Pt-interior corresponds only to the fraction of Pt nanoparticles that is present in the interior volume of the carbon primary particle that does not directly contact the ionomer film.

   They developed a model, therefore, I could label it as hypothesis, but out of context it would be background information.

5 Culinary Examples

For the sake of clarity and in order to make the annotation scheme more easily understandable for readers without a materials science background, we here apply the labels definition to pancake preparation. Examples for this part have been invented but ideally resemble what could be found in a kitchen blog.

5.1 MOTIVATION

The motivation to cook pancakes might be trivial and/or superfluous. However, the aim of developing a new recipe is usually stated in blogs [Example 1]. This first example consists in a sentence containing both (motivation and aim), since the motivation to deepen the knowledge in pancake making is motivated by the offspring and the aim is to optimize the recipe.

1. I wanted to perfect my pancakes before preparing them with my boys!

5.2 BACKGROUND

In the context of cooking pancakes, the background is represented by all those sentences that describe the general appearance of pancakes and their history [Example 1]. But it also includes state of advancement in the development of the recipe [example 2].

1. A good pancake must be golden, fluffy but most importantly delicious!
2. Over the years I have tried so many recipes for homemade pancakes, and because I never found one I loved I would always go back to boxed stuff. . . until now!

5.3 BACKGROUND/PRIOR WORK

Outlining other recipes fits well within this subtype.

1. My grandma's recipe has always been a classic in my family.

5.4 EXPERIMENTS/PREPARATION

Sentences that report preparation steps [Example 1,3] and starting materials [Example 2] as well as insights on those aspects [Example 4] are labelled as such.

1. Preheat the griddle to 200°C.
2. The basic ingredients of pancakes are: baking powder, sugar, salt, flour, canola oil, eggs, milk and vanilla.
3. You may use a 1/2 cup to scoop out the batter.
4. Bigger scoop means bigger pancakes!

5.5 EXPERIMENTS/CHARACTERIZATION

An easily accessible example for this subtype is represented by the serving method as a description of the tasting of the pancake (which is meant to be a measure on the prepared samples, the pancakes themselves) [Example 1].
Otherwise we can also hypothesize that the cooker might need to probe the cooking level while preparing the pancakes, measuring the degree of cooking. [Example 2]

1. I recommend to taste these pancakes with maple syrup while they are still warm!
2. The perfect cooking is reached when the color becomes golden brown.

5.6 EXPLANATION

Explanations in pancakes preparation can consists in approximations specific for a specific preparation. [Example 1]
Also hypothesis that can be made to describe the outcome. [Example 2]

1. I assume that your pancakes will be delicious even if they don't come out in perfect circles.
2. It is possible that the pancakes burned because of the different brand of milk that I bought.

5.7 RESULTS

The results consist in the outcome of measures. In the case of pancakes, we can assume eating them to be a measuring process of the goodness and the outcome can be the expressions of the served people. [Example 1]

1. My youngest liked them so much that he tried to steal mine!

5.8 CONCLUSION

The conclusion is usually intended as the take-away message. In other words, a sentence that sums up all the results and possibly also the trials that lead to them. [Example 1]

1. A lot of pancakes after (a few burnt, a lot more eaten), we finally perfected our recipe!

References

Simone Teufel et al. *Argumentative zoning: Information extraction from scientific text*. PhD thesis, Citeseer, 1999.