**Aspect Based Sentiment Analysis Annotation Guidelines**

# Background

The motivation of this annotation task is to be able to train a model to detect and analyze the sentiments expressed toward certain GM crop related entities in African media. To achieve this, we have acquired a large corpus of articles from African media and extracted direct quotations contained in them. We will be labeling these quotations. There will be two subtasks, all described below.

* Aspect Based Sentiment Analysis
* Quotation Extraction validation

# ABSA Introduction

An aspect belongs to a category, which is defined by an entity and an attribute. An entity refers to a certain part that is important for the domain in question. An attribute defines a property of the entity and details the category of the aspect. Given a quote, the task of the annotator is to identify the following types of information:

* **Aspect Category** (**Entity** and **Attribute**) → Identify the entity E and attribute A pair E#A towards which an opinion is expressed. E and A should be chosen from the inventories of entity types (e.g., **GM\_CROP**, **ORGANIZATIONS**) and attribute labels (e.g., **PRODUCTIVITY**, **RESEARCH\_DEVELOPEMEN**) that are described in the following sections. The identified entities should be assigned one or more attribute labels based on the context of the sentence they appear in. The E#A pair defines an aspect (category).
* **Opinion Polarity** → Each identified E#A pair of a sentence has to be assigned a polarity, from a set P = {positive, negative, neutral, conflict}. Refer to the following sections for more details.

In ABSA, an **ENTITY (E)** generally refers to a concept, object, or topic that can be associated with sentiment, and an **ATTRIBUTE (A)** usually referring to a specific feature or characteristic of that entity.

## ENTITY LABELS:

**E1**: **CROPS**: Any specific mention of a crop variety, genetically modified crop, GM traits or crops in general.

**E2**: **ORGANIZATIONS**: Companies, research institutes, government bodies, groups of individuals, etc., or collections thereof.

**E3**: **AGRICULTURAL\_PRACTICES**: Methods, techniques, or processes associated with farming and agriculture.

**E4**: **TECHNOLOGY**: Any mention of the specific tools used in the genetic modification of crops, such as CRISPR, transgenics, etc., and/or associated technologies, such as pesticides. TECHNOLOGY should encompass any mentions of the technology itself, while AGRICULTURAL\_PRACTICES should encompass mentions specifically related to use (i.e. “Pesticides are bad” vs “The use of pesticides is bad”)

**E5**: **GEOGRAPHIC\_LOCATIONS**: Specific countries, regions, or locations

**E6**: **ENVIRONMENTAL\_CONDITIONS**: References to weather, climate, or other environmental conditions (earthquakes, viruses, etc.) and their impact on crop growth, distribution, or the suitability of specific geographies for crops.

**E7**: **LEGAL\_ASPECTS & POLITICS**:

* specific laws, policies, or regulations
* political debates, political party positions, or government statements. This would cover politicization of the GM crop debate and legislative influence on GM crop usage and policies.

**E8**: **ECONOMIC FACTORS**: References to economic factors such as inflation, imports/exports, unemployment, and supply and demand.

**E9**: **OTHER**: for opinions expressed about entities and their impact on GM crops, not fitting into any of the above eight entity types. (Note: Most instances should fit into one of the above categories. For those where you believe it should be labeled as OTHER, please share the example with the group.)

**!!! NOTE**: When an entity E is only implicitly referred (e.g., through pronouns) or inferred in a sentence, then the slot is assigned the value “NULL”.

For example, “*This will greatly improve maize production over the next ten years*”. “*This*” would be labeled as NULL as it is impacting crop production, but we do not know what the antecedent of “*this*” is.

## ATTRIBUTE LABELS:

**ATTRIBUTES (A)** - the qualities or specifics related to the entities that the sentiment might be about. An **ATTRIBUTE** can either be explicit (i.e, directly mentioned in the text) or implicit (i.e., implied by the context).

**A1**: **RESISTANCE** (for CROPS): Refers to the crops' ability to resist pests, diseases, and adverse weather conditions.

**A2**: **CONSUMER\_PERCEPTION** (for CROPS): Views, attitudes, or beliefs of the general public or specific consumer groups towards crops. This could include perception of taste, nutritional value, shelf life.

**A3**: **SAFETY** (for CROPS, TECHNOLOGY, AGRICULTURAL\_PRACTICES): Concerns related to the safety of consuming crops. It includes:

* human health as it relates to the consumption or exposure to crops.
* animal health as it relates to the consumption or exposure to crops, for animals that are part of the human food chain or otherwise.

**A4**: **FOOD\_SECURITY** (for CROPS, ORGANIZATION, LEGAL\_ASPECTS & POLITICS, AGRICULTURAL\_PRACTICES, TECHNOLOGY, ECONOMIC\_FACTORS ): References to the role of crops, policies, organizations, and practices in addressing food security, hunger, or malnutrition. Also refers to the status of food security in specific locations.

**A5**: **PRODUCTIVITY** (for CROPS, AGRICULTURAL\_PRACTICES, TECHNOLOGY, ORGANIZATIONS, LEGAL ASPECTS & POLITICS, ECONOMIC FACTORS): Refers to the yield or efficiency of crops or the farming practices associated with them. Also refers to the impact of various factors on the yield or efficiency of crops.

**A6**. **ECONOMIC\_IMPACT** (for CROPS, AGRICULTURAL\_PRACTICES, LEGAL\_ASPECTS & POLITICS, ENVIRONMENTAL\_CONDITIONS, TECHNOLOGY): Refers to the economic implications of growing crops, including cost of production, market price, profitability, trade policies etc. This might also include the cost of fuel or other inputs related to the cultivation, transportation, and distribution of crops.

**A7**: **RESEARCH\_DEVELOPEMENT** (for ORGANIZATIONS, LEGAL\_ASPECTS & POLITICS, TECHNOLOGY): The processes that organizations undertake to develop new genetically modified crop varieties.

**A8**: **ENVIRONMENTAL & ETHICAL\_CONCERNS** (for CROPS, ORGANIZATIONS):

* any mention of the environment or ecosystem in the context of crops.
* ethical issues surrounding crops, including but not limited to issues of biodiversity, biopiracy, or the rights of farmers and indigenous people, and the preservation of traditional agricultural practices.
* mentions of the impact of crops on genetic diversity, either within a single crop species or the wider ecosystem. This could include the risk of cross-contamination between GM and non-GM crops, or the possibility of creating "superweeds" resistant to herbicides.
* the degree to which organizations adhere to rules, regulations, and ethical guidelines.

**A9: MISCELLANEOUS** for attributes that do not fall into any of the aforementioned cases.

## **!!!** **What should be annotated:**

* Nominal phrases explicitly mentioning aspects. Notice that in the following example, the aspect term is *maize lethal necrosis disease*, not simply maize.

*The maize lethal necrosis disease has caused losses worth millions of dollars for farmers and seed companies in the affected regions in sub-Saharan Africa.*

### Remarks:

* Subjectivity indicators(i.e., words/phrases expressing opinion, evaluation etc.) are NOT considered aspect terms or parts of aspect terms and should NOT be annotated (e.g., “*good*”, “*great*”, “*overpriced*”).
* Terms that are often used as aspect terms (e.g., “*maize*”) may not always be aspect terms.
* If an aspect term appears in a sentence more than once, only the first occurrence in the sentence should be annotated.

*Based on the great demand Monsanto Company (MON) have seen in 2016, Monsanto Company (MON) know farmers are looking forward to the benefits of the Roundup Ready Crop System, including in-crop use of dicamba and glyphosate.*

* The identified aspect terms should be annotated even if they appear in quotation marks or brackets. Notice that *Monsanto Company (MON)* is a single aspect term:

*Monsanto Company (MON) are excited to enable another piece of the Roundup Ready Crop System to farmers in next year.*

* The identified aspect terms should be annotated as they appear, even if misspelled.
* Determiners (e.g., “*a*”, “*the*”, “*some*”, “*many*”, “all”) should not be included in aspect terms (unless they are parts of embedded noun phrases).

*We want the high cost of production to be addressed through provision of quality seeds that are drought tolerant and resistant to Maize Necrosis Lethal disease and armyworm.*

***!!!******NOTE:*** *"drought-resistant and resistant to Maize Necrosis Lethal disease and armyworm " can be considered a direct attribute of quality seeds*

## !!! What should NOT be annotated as aspect term

* References to the target entity (e.g., the crop the sentence is about) as a whole and mentions of other entities (e.g., “*maize*” below). No aspect term should be annotated in the following sentence:
* *Great alternative.*

Only *cassava* should be annotated in the following sentence:

* *Cassava is drought-resistant and a good alternative to maize.*
* Pronouns (e.g., “*it*”, “*they*”, “*this*”) even if they refer to an aspect. For example, “*it*” should not be annotated below:

*From Cassava, we can get ethanol, noodles, beverages, clothing materials, glucose and soap. It also has many uses that could be improved upon to earn forex for Nigeria.*

!!! NOTE: When a sentence contains more than one mention (e.g., nominal and pronominal) to the same entity then the most informative one should be annotated. (*cassava* in the above example)

## OPINION POLARITY

Based on the sentiment that is expressed in the sentence, one of the following polarities has to be assigned:

* *positive*
* *negative*
* *conflict* (both positive and negative sentiment)
* *neutral* (neither positive nor negative sentiment)
* **Positive, negative, conflict polarity**

A term should be classified as *positive*, *negative* or *conflict* if the sentence expresses a positive, negative, or both positive and negative, respectively, attitude, opinion, evaluation, emotion, or feeling etc. of an opinion holder towards the aspect term.

* **Neutral polarity**

Aspect terms should be annotated as *neutral* in the following cases:

* when a neutral sentiment, wish, or desire toward the aspect term is expressed, e.g.,

*Mandisa Nzuza, another top achiever from the same school would like to study chemical engineering or biochemistry at the University of Cape Town.*

* when factual information (no sentiment) about the aspect term is provided:

*So, cassava, rice, Maize, poultry for meat (broiler), aquaculture and cattle fattening are the 6 commodities that they are piloting with over 140,000 farmers.*

* when positive or negative polarity about the named aspect might be inferred, without being explicit.

*It took the federal government a month to submit comments on the applications submitted by Monsanto to issue two Permits to Monsanto to deploy GMO in Nigeria.*

* when expressions like “*moderate*”, “*in line*”, “*nothing out of the ordinary*”, “*not an issue*” etc. are used:

*Tanzania Leaf Tobacco Company (TLTC) provides extension services to villages, a service which is in line with sustenance of the environment and has centres of excellence in Urambo in Tabora, Lupa Tingatinga in Chunya and Likenangena in Ruvuma.*

IMPORTANT: If a sentence conveys both (a) **neutral** and (b) **negative** (or **positive**) opinions about an aspect category, then the **negative** (or **positive**) **polarities dominate over the neutral** ones.

**NOTE:** Each of these entities and attributes can be linked to sentiment in a given text. For instance, a sentence like "*The productivity of GM crops is impressive, but their environmental impact is worrying*" would be annotated to indicate positive sentiment towards the `**PRODUCTIVITY**` attribute and negative sentiment towards the `**ENVIRONMENTAL\_IMPACT**` attribute of the `**CROPS**` entity.

## ANNOTATION CONFIDENCE

For each aspect term and aspect category annotation, please provide a confidence level rating according to the following three-level scale:

1.     *Not Confident* (i.e., the annotator is not sure about an aspect term)

2.     *Average Confidence* (i.e., the annotator is confident about the aspect term but not about its polarity)

3.     *Confident*

# Quotation Extraction Validation

Our quotation extraction model will necessarily be imperfect. Both in order to understand how well it works on our dataset and to avoid labeling malformed examples, we will be notating when the quotation extraction fails, and selecting the correct quotation from the original text. For our purposes, a quotation extraction failure will be when the presented quote is:

* Clearly not just the quotation (only the quote itself should be shown, not the speaker)

*He said ‘Annotation is fun!’*

* Not presenting a complete thought or a sentence fragment

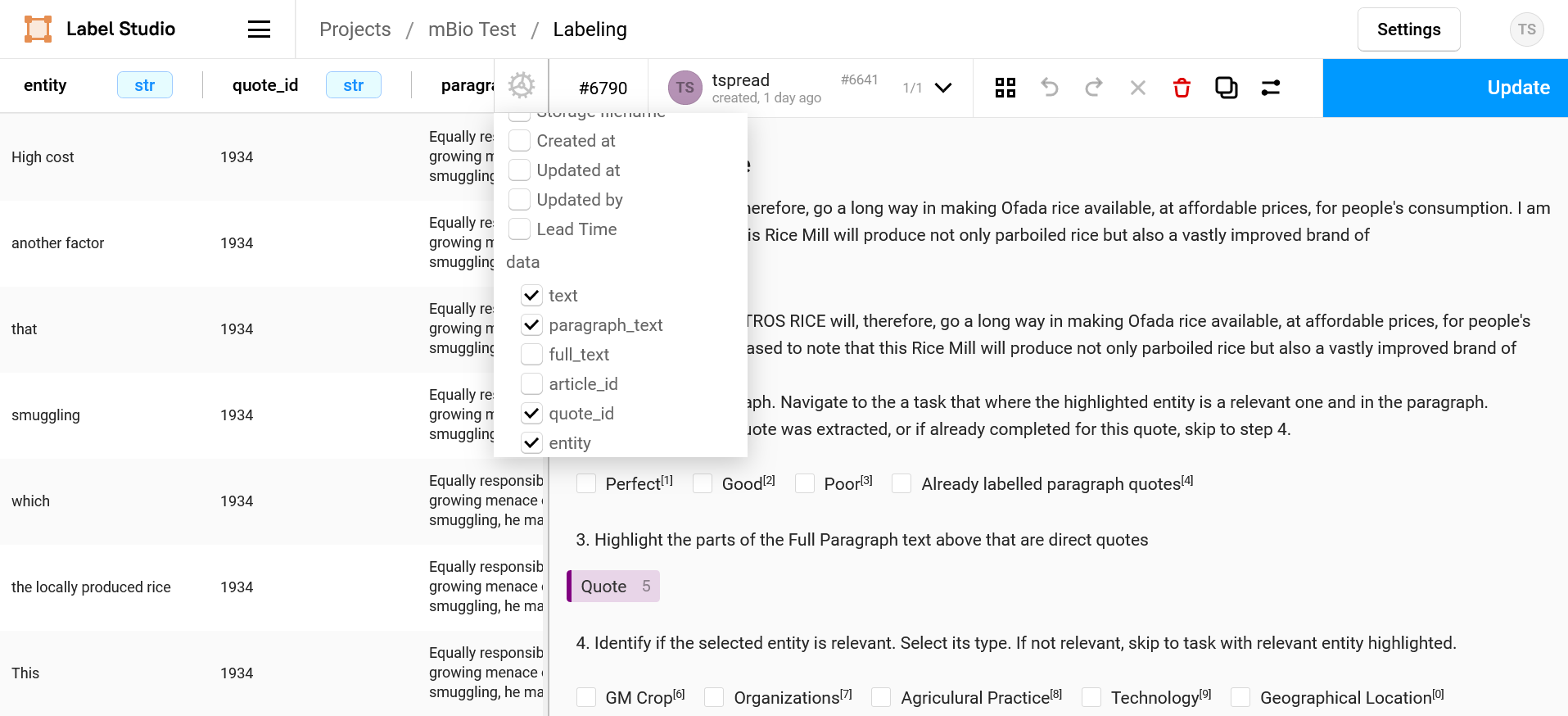
*It took the federal government a month to submit comments on the applications submitted by*

The annotator will evaluate the quality of the quote extraction by selecting one of the following options:

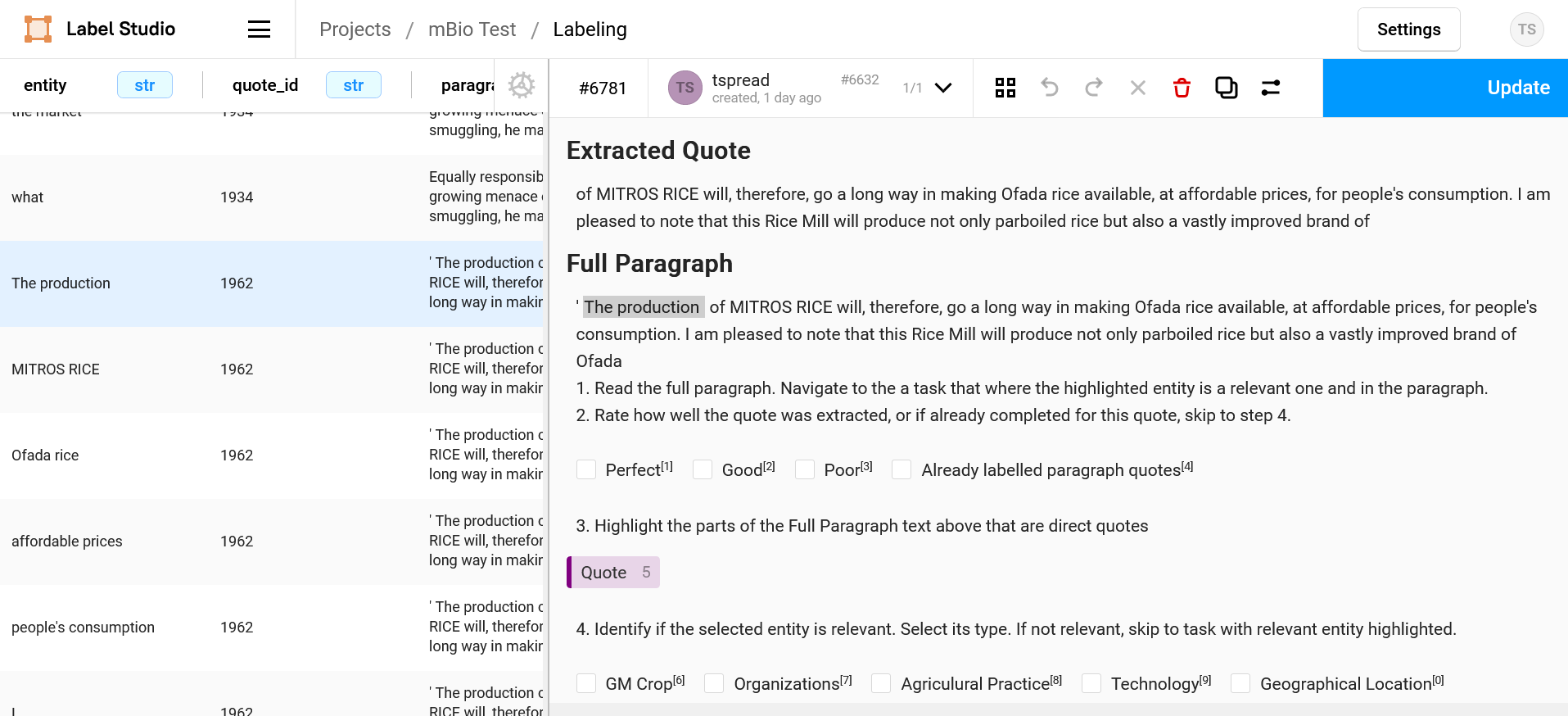
* *perfect* (the FULL quote was extracted flawlessly)
* *good* (maximum of 3-4 words were either omitted or added – please consider the length of the quote when selecting this option)
* *poor* (the extraction quality is unsatisfactory)

# Example

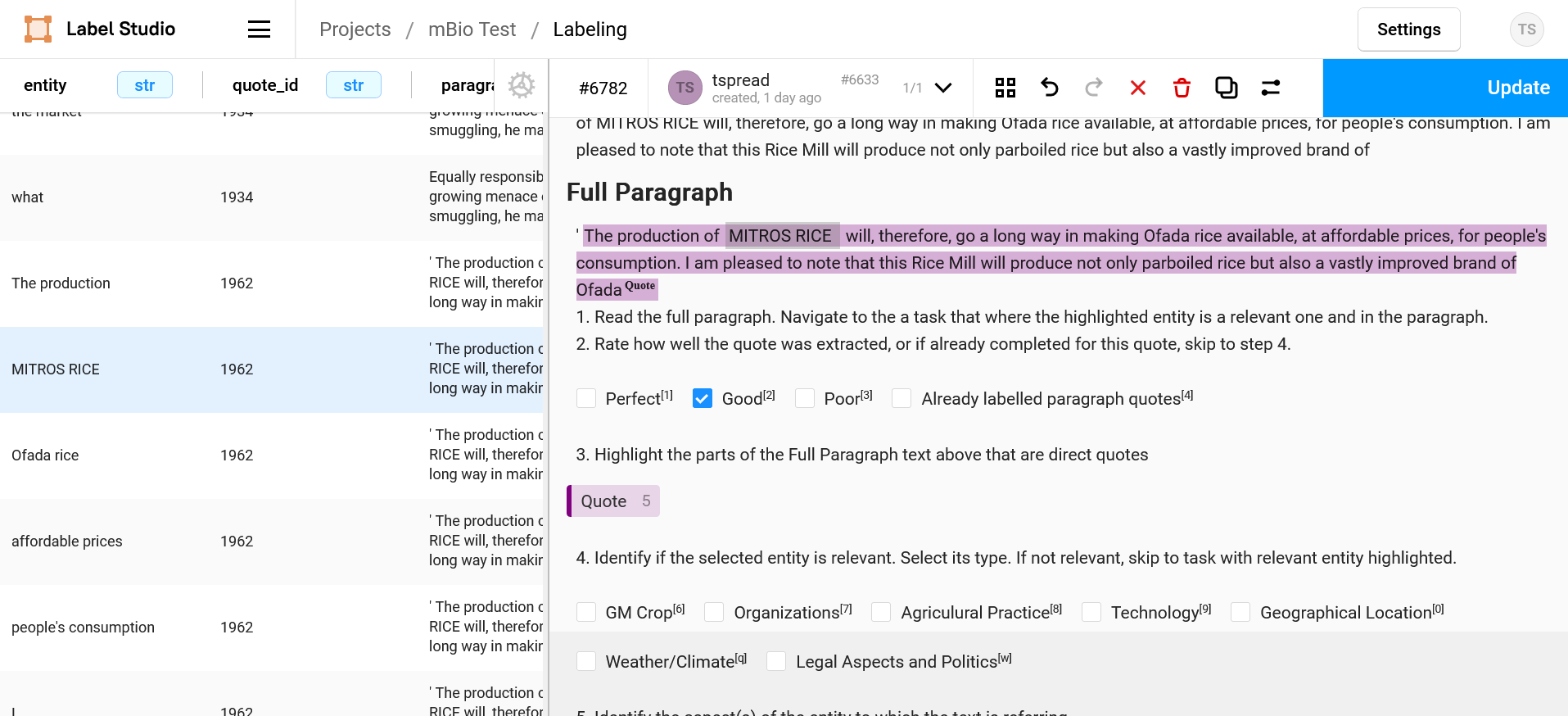
* To make it easier, display the ‘entity’ and ‘quote\_id’ columns



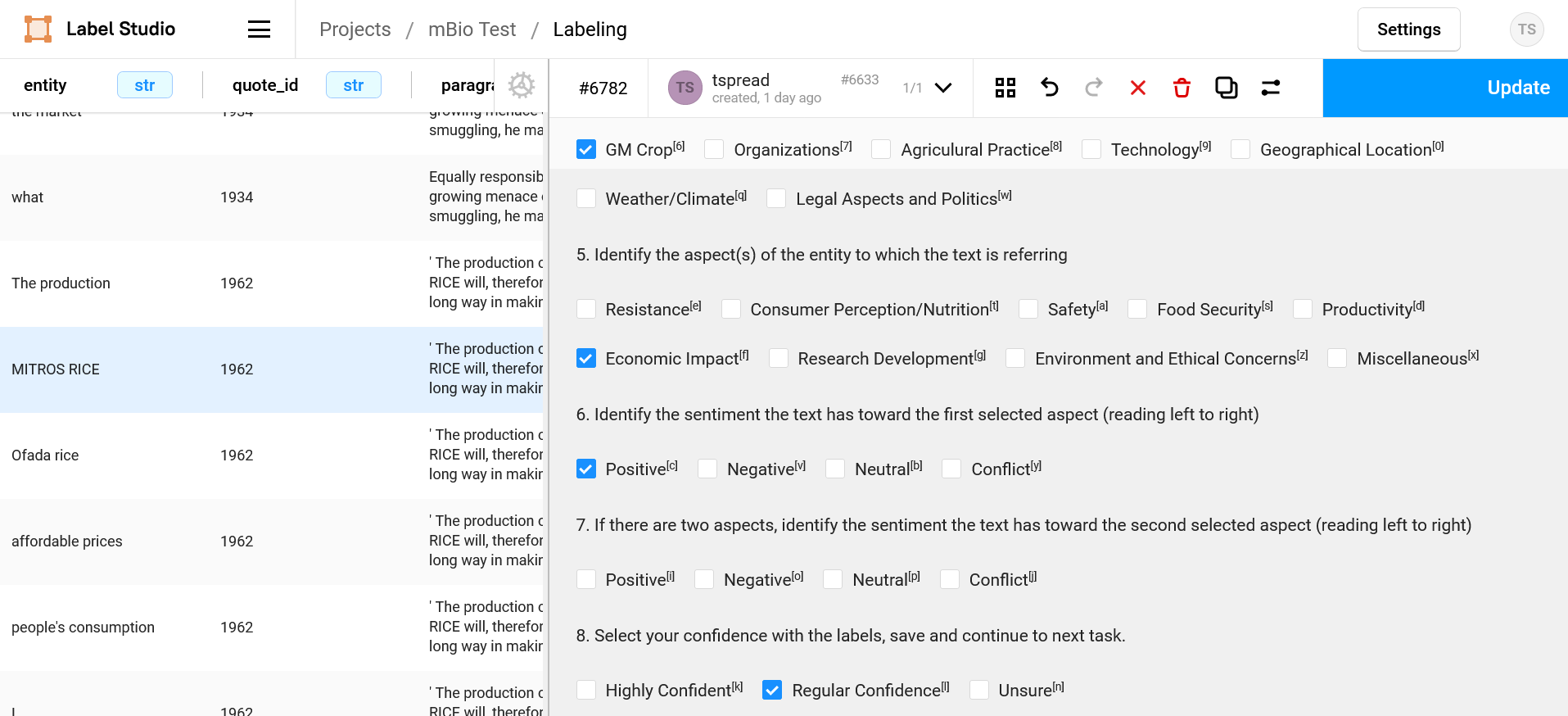
* Go to the next quote\_id. Read the quote and identify relevant entities.



In this case, they are ‘MITROS RICE’ and ‘Rice Mill’. Navigate to the task that has ‘MITROS RICE’ as the entity. Mark that the quote extraction was ‘Good’ and highlight the correct quote.



Next select the entity type of ‘MITROS RICE’ as ‘GM Crop’ because it is a crop. The aspect that is being discussed is its economic impact because they state it will help drive affordable prices. The sentiment toward that aspect of the entity is positive.



Now continue to the next entity in the quote and repeat, this time without bothering with the quote extraction questions. Once you have completed all relevant entities, continue to the next quote\_id.

# Workflow

* Download the batch from the [drive folder](https://drive.google.com/drive/folders/15C0kEtiRNHA5HvXQGQzrEjylU5-T9mBB)
* Start by creating a new project in label-studio with the name of your new batch
* Annotate!
* Once finished, export your results. Upload with the original name of the batch with “-v4-YOUR\_NAME.json” to the “Annotated Batches” folder.