Yelp Reviews Annotation Guidelines – Draft

Qishen "Justin" Su, Kelley Lynch, Yuanyuan Ma

March 20, 2016

Contents

1 Introduction

This document provides annotation guidelines for the Yelp restaurant reviews annotation task. The Yelp restaurant reviews annotation task involves (1) annotating the relation between a specific dish mentioned in a review and its described quality or characteristics, (2) annotating the relation between a dish and its ingredients or its parts, and (3) annotating the anaphors of dishes. These guidelines provide details on how to annotate Yelp restaurant reviews.

This document is organized in the following manner: Section 2 is focused on annotating extent tags, i.e. FOOD, QUALITY and ANAPHORA, while Section 3 is focusing on annotating link tags, i.e. PART_OF, OPINION, and COREFERENCE.

When examples of annotation are provided in this document, texts with different extent tags are marked in different colors: blue is for FOOD, red is for QUALITY, and yellow is for ANAPHORA. A QUALITY tag contains three types: *positive*, *negative*, and *neutral*, and in this document, a type of QUALITY will be presented in subscript. Then, each link tag is represented using a predicate argument structure given in Table 1.

Table 1: Link Tag Predicate Argument

Link Tag	Predicate Argument Structure
PART_OF	PART_OF(from=text#1, to=text#2)
OPINION	OPINION(from=text#1, to=text#2, relation=opinion_type)
COREFERENCE	COREFERENCE(from=text#1, to=text#2)

The following is an example of how the format is utilized: For example:

1. I had a **Shrimp dish** that was **out of this world**$_{positive}$ and **it** was served over **amazing**$_{positive}$ **fresh**$_{positive}$ **veggies**.
   PART_OF(from=**veggies**, to=**it**)
   COREFERENCE(from=**it**, to=**Shrimp dish**)
   OPINION(from=**out of this world**, to=**Shrimp dish** , relation="explicit-opinion")
   OPINION(from=**amazing**, to=**veggies** , relation="explicit-opinion")
   OPINION(from=**fresh**, to=**veggies** , relation="explicit-opinion")

2 Extent Tags

2.1 FOOD

The FOOD tag concerns with specific food or beverage names, i.e. corns, beans, burritos, street tacos, margaritas, etc. When annotating a food name, articles (i.e. *a* and *the*) should be excluded. For example:

2. The **chips** and **salsa** were **great**$_{positive}$.
   OPINION(from=**great**, to=**chips** , relation="explicit-opinion")
   OPINION(from=**great**, to=**salsa** , relation="explicit-opinion")

Moreover, when the toppings or sides of a dish is mentioned in the food name, the whole noun phrase should be tagged as a FOOD tag, as shown in the following example.

3. I had the **carne asada potato with cheese** and oh my **it** was **delicious**$_{positive}$.
   COREFERENCE(from=**it**, to=**carne asada potato with cheese**)
   OPINION(from=**delicious**, to=**it**, relation="explicit-opinion")

There are instances where there should be more than one FOOD extent tags for the same food name. Take the following review as an example, **agua frescas** in this case are two different instances: one is mango kind, and the other one is cantaloupe/watermelon kind. Since they are two different dishes, there should be two different FOOD tags for the phrase **agua frescas**, and their anaphora should be connected to the corresponding dish name.

4. We ordered two **agua frescas**. **One** was **mango** and **the other** was **cantaloupe/watermelon** on the waitress' recommendation. Wow! **both** were so **fresh**$_{positive}$ and **delicious**$_{positive}$.
   COREFERENCE(from=**One**, to=**agua frescas**$_1$)
   COREFERENCE(from=**the other**, to=**agua frescas**$_2$)
   PART_OF(from=**mango**, to=**One**)
   PART_OF(from=**cantaloupe/watermelon**, to=**the other**)
   COREFERENCE(from=**both**, to=**mango**)
   COREFERENCE(from=**both**, to=**cantaloupe/watermelon**)

One exception where a specific food name is not annotated is when there is no description for that dish or beverage, in terms of ingredients or quality. The following example is a complete review of a customer. In this review, a specific dish, *Beach Burger*, is mentioned. It is not annotated with a FOOD tag, because there is no description for this dish.

5. Phew, this is a GREAT place for sure. The reviews are what got me here and no doubt I WILL be back. Had the Beach Burger.

When the generic term "food" is used, it should not be annotated. In the following examples, the terms "food" and "meal" are not annotated with a FOOD tag, because they are not specific food names.

6. The food here is great!

7. We stopped for our first meal in Phoenix and loved it.

2.2 QUALITY

The quality tag marks words or phrases that indicate an explicit or implicit description of the quality of a specific food or list of foods. Quality is often indicated by by an adjective describing a particular food or an modifying an anaphora that is coreferent to a particular food. Quality should be marked as either positive, negative, or neutral. For example, in the following review, the quality descriptions indicate positive qualities of the food.

8. The **tacos** there are **SO good**$_{positive}$, and **SO cheap**$_{positive}$...its easy to justify coming back again and again and again.
   OPINION(from=**SO good**, to=**tacos**, relation= "explicit-opinion")
   OPINION(from=**SO cheap**, to=**tacos** relation="explicit-opinion")

In some occasions, when the price of a dish is directly associated with its value, the phrase that describes the price should also be tagged as quality. For example, the key words are generally "cheap", or "overpriced".

9. The **enchilada** here is **overpriced**$_{negative}$.
   OPINION(from=**overpriced**, to=**tacos**, relation= "explicit-opinion")

In the following review the quality of "Don't order" is neutral because the description of "chicken" is not indicating anything about the chicken, but instead, mentioning it to serve as a comparison. Also note that in the tag for "You will be amazed", the span is the entire sentence.

10. Be adventurous. **Don't order**$_{neutral}$ the **chicken**. Try **lengua**, or **suadero**! You will be **amazed**$_{positive}$.
    OPINION(from=**Don't order**, to=**tacos**, relation="explicit-opinion")
    OPINION(from=**You will be amazed**, to=**lengua**, relation="implicit-opinion")
    OPINION(from=**You will be amazed**, to=**suadero**, relation="implicit-opinion")

2.3 ANAPHORA

The anaphora tag is to mark names or pronouns that refers to the food items or entities mentioned in previous sentences or paragraphs. An anaphora could be a pronoun like "it", "they", "both", or another name for the food. The use of anaphora enables the link between quality and food be in the same sentence or clause, which will be discussed in "OPINION" session. Some examples of anaphoras are listed below:

4

11. I had the **carne asada potato with cheese** and oh my *it* was **delicious**$_{positive}$.

    COREFERENCE(from=*it*, to=**carne asada potato with cheese**)

    OPINION(from=**delicious**, to=*it*, relation="explicit-opinion")

12. We ordered two **agua frescas**. *One* was **mango** and *the other* was **cantaloupe/watermelon** on the waitress' recommendation. Wow! *both* were so **fresh**$_{positive}$ and **delicious**$_{positive}$.

    COREFERENCE(from=*One*, to=**agua frescas**$_1$)

    COREFERENCE(from=*the other*, to=**agua frescas**$_2$)

    PART_OF(from=**mango**, to=*One*)

    PART_OF(from=**cantaloupe/watermelon**, to=*the other*)

    COREFERENCE(from=*both*, to=**mango**)

    COREFERENCE(from=*both*, to=**cantaloupe/watermelon**)

When some quantifiers are used (i.e. "all", "everything", etc.), depending on the context of a review, they might or might be ANAPHORA. As a rule of thumb, if a quantifier describes all the dished that a customer orders, it should be an anaphora; if it generally describes the food in a restaurant or if it describes all the dishes there, it should not be tagged as an anaphora. Please consider the following two examples. The word "everything" in the first example describes the food that the customers ordered; therefore, it should be tagged as ANAPHORA. On the other hand, in the second example, while the same word is used, it describes the food in a restaurant in general; thus, it should not be tagged.

13. *Everything* that we had last night was **great**$_{positive}$! The **burrito** was **flavorful**$_{positive}$, and the **chips** were **crispy**$_{positive}$!

    COREFERENCE(from=*Everything*, to=**burrito**)

    COREFERENCE(from=*Everything*, to=**chips**)

    OPINION(from=**great**, to=*Everything*, relation="explicit-opinion")

    OPINION(from=**flavorful**, to=*burrito*, relation="explicit-opinion")

    OPINION(from=**crispy**, to=*chips*, relation="explicit-opinion")

14. my husband and I always come here, because everything is just cheap and delicious!

3 Link Tags

3.1 PART_OF

The PART_OF link tag describes a part-whole relation between an ingredient and a dish, or between a dish and a combo. The predicative argument structure PART_OF(from=text#1, to=text#2) indicates, *text#1 is part of text#2*. The following is an example of PART_OF relation between ingredients and a dish. The ingredients should be connected to the closet mention of a dish or its anaphora.

15. I just couldn't believe my eyes when I saw all the stuff they squeezed into **both sandwiches**! **Ham**, **pork sirloin**, **chorizo**, **sausage**, **egg**, **avocado**, **jalapenos**, **breaded beef**.
    PART_OF(from=**Ham**, to=**both sandwiches**)
    PART_OF(from=**pork sirloin**, to=**both sandwiches**)
    PART_OF(from=**chorizo**, to=**both sandwiches**)
    PART_OF(from=**sausage**, to=**both sandwiches**)
    PART_OF(from=**egg** to=**both sandwiches**)
    PART_OF(from=**avocado**, to=**both sandwiches**)
    PART_OF(from=**jalapenos**, to=**both sandwiches**)
    PART_OF(from=**breaded beef**, to=**both sandwiches**)

3.2 OPINION

The opinion tag links quality descriptions to the foods that they are referring to. The links are labeled with explicit or implicit. Explicit opinions are those that clearly describe aspects of the food, such as fresh, delicious, or cheesy. Implicit opinions are those where the quality is inferred from the description, such as in example 7, where the sentence You will be amazed does not directly refer to the foods lengua and suadero, but it implicitly applies that the foods are good.

16. The only **okay**$_{neutral}$ dish was their **chips** and **salsa**. **Do not order**$_{negative}$ the **enchilada suiza**. It tastes **nasty**$_{negative}$ and **do not order**$_{negative}$ their **beans** and **rice**.
    OPINION(from=**okay**, to=**chips**, relation="explicit-opinion")
    OPINION(from=**okay**, to=**salsa**, relation="explicit-opinion")
    OPINION(from=**Do not order**, to=**enchilada suiza**, relation="implicit-opinion")
    OPINION(from=**nasty**, to=**it**, relation="explicit-opinion")
    OPINION(from=**do not order**, to=**beans**, relation="implicit-opinion")
    OPINION(from=**do not order**, to=**rice**, relation="implicit-opinion")
    COREFERENCE(from=**it**, to=**enchilada suiza**)

When a review has certain connotation, such as sarcasm, generally the quality should be marked as "implicit-opinion". Please consider the following example: the reviewer used quotation marks to quote the word **burrito** to indicate sarcasm and the poor quality of the dish.

17. Well, I ordered a **"burrito"**$_{negative}$, and it was disappointing.

    OPINION(from=**"burrito"**$_{negative}$, to=**burrito**, relation="implicit-opinion")

Finally, as a rule of thumb, when an opinion is ambiguous, it should be tagged as "implicit-opinion".

3.3 COREFERENCE

The CORREFERENCE tag links the anaphora with previous mention of the same entity. It serves, in addition, as the "to" object of the OPNION tag. Examples are as follows:

18. I had the **carne asada potato with cheese** and oh my **it** was **delicious**$_{positive}$.

    COREFERENCE(from=**it**, to=**carne asada potato with cheese**)

    OPINION(from=**delicious**, to=**it**, relation="explicit-opinion")

In the following example, the ANAPHORA **both** should be linked with the closest previous mentions of the entity, which is **mango** and **cantaloupe/watermelon** instead of **agua frescas**.

19. We ordered two **agua frescas**. **One** was **mango** and **the other** was **cantaloupe/watermelon** on the waitress' recommendation. Wow! **both** were so **fresh**$_{positive}$ and **delicious**$_{positive}$.

    COREFERENCE(from=**One**, to=**agua frescas**$_1$)

    COREFERENCE(from=**the other**, to=**agua frescas**$_2$)

    PART_OF(from=**mango**, to=**One**)

    PART_OF(from=**cantaloupe/watermelon**, to=**the other**)

    COREFERENCE(from=**both**, to=**mango**)

    COREFERENCE(from=**both**, to=**cantaloupe/watermelon**)

In some cases, an ANAPHORA can be linked to a QUALITY tag as well. In that case, the ANAPHORA should still be linked to the dish that it describes with an OPINION link tag.

20. The **chips** were **good**$_{positive}$ and **so** was the **burrito**!

    COREFERENCE(from=**so**, to=**good**)

    OPINION(from=**good**, to=**chips**, relation="explicit-opinion")

    OPINION(from=**so**, to=**burrito** ,relation="explicit-opinion")