Type	ID	Question	Answer	Sentence in the entry fiel
PP	PP1	Is this Japanese tea?	Yes	It has the term "Powder Green Tea" on the package. Therefore, this is Japanese tea.
	PP2	Is this a carbonated drink?	Yes	We see the term 'Cider' on the package
NP	NP1	Is it an instant food?	Yes	Because we do not cook it with a traditional way such as putting leaves into a pot, I think the answer can be yes
PN	PN1	Is it coffee?	No	Because dressing is not coffee
	PN2	Is this tea?	No	It is coffee.
NN	NN1	Is this green tea?	No	Since milk tea does not contain green tea leaves
	NN2	Is this a seasoning?	No	A seasoning is something that adds a fl vor to foods
PNN	PNN1	Is this coffee?	No	Because this is herb tea and contains no caffeine, this is not coffee
	PNN2	Is this an instant food?	No	Since an instant food is a food that is easy to cook and the beans must be roasted before eating, I think the answer is no.

Table 2: Annotation Examples

Table 2 gives examples to explain how we annotated the viewpoints.

PP2. "We see the term Cider on the package" is the reason. The answer is "Yes". Therefore, it is PP. Note that we can obtain the conclusion even if the sentence in the entry fiel does not have an explicit conclusion.

PN2. The answer is No and we can interpret this sentence as "Because this is coffee, this is not tea." Therefore, "This is coffee" is the reason and the type is PN.

NN2. This is a complicated case. Although "A seasoning is something that adds a fl vor to foods" does not contain any negative expression, it is NN. The reason is this. First, since the question is "Is this a seasoning?" and the answer is "No", "It is *not* a seasoning" is the conclusion and the viewpoint is either PN or NN. Second, the sentence "A seasoning is something that adds a fl vor to foods" means "if it is a seasoning, it adds fl vor to foods" and its contraposition is "if it does not add any fl vor to foods, it is not a seasoning". Since the answer is No, this sentence is exactly her viewpoint. Therefore, "if it does not add any fl vor to foods" is the reason and it is NN. In general, if (1) the question is "Is this X", (2) the reason contains a necessary condition, and (3) the answer is "No", we conclude the viewpoint is of type NN.

PNN2. The sentence contains two reasons "An instant food is a food that is easy to cook" and "the beans must be roasted before eating", and the conclusion "I think the answer is no." The combination of the firs reason and the answer has the same pattern as NN2 and is NN. The combination of the second reason and the answer is a typical PN. Therefore, it is PNN as a whole. If we have more than one reason in a sentence, we determine the types of (sub)viewpoints and combine them to determine the (composite) type of the viewpoint.

3.4 Inter-annotator Agreement and Corpus Statistics

In this section, we report the inter-annotator agreement and summary statistics for our annotated corpus. The collected viewpoints were manually annotated by two judges. We measured inter-annotator agreement and the κ value, a standard measure for showing agreements, was 0.968, showing excellent agreements between two annotators [3]. There were only 28 difficul cases where annotators disagreed in the type of viewpoints. The high agreement shows the clarity of the annotation guideline.
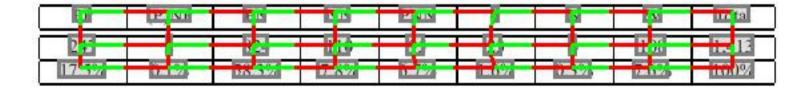

Table 3: Numbers and percentages of viewpoints of different types

Table 3 shows the numbers and percentages of viewpoints of different types. There is a large number of viewpoints labeled as PN. W stands for wrong. NP and PNP each appears only once in the corpus. As the table shows, there are few viewpoints of types NP and PNP. This is not surprising and we can justify this. Given a viewpoint $\neg r \to c$ of type NP, its contraposition is $\neg c \to r$. In many cases, there are too many things that do not satisfy c and it is difficul to give a simple phrase to describe the common attribute of those things, that can be used for r. For example if c is "it is a coffee", it is not easy to give a simple phrase to describe the common attribute of the things that is not coffee.

4. SUBJECTIVE EVALUATION

To evaluate the effectiveness of the logic-oriented annotation scheme to select generic viewpoints, we crowdsourced judging whether each viewpoint is applicable to a variety of data items. We call this a subjective evaluation of our annotation scheme.

4.1 Crowdsourced Experiments

Tasks. Remember that in order to obtain a viewpoint $r \to (q, a)$, we asked workers why they gave the answer a to the question q with the shown data item d. The subjective evaluation tasks ask other workers to judge whether the viewpoint is useful to answer the question q with another data item $d'(\neq d)$.

To generate the subjective evaluation tasks, we used the viewpoints obtained in Section 3 and the *gold standard data* that contains a correct answer for each data item. More specificall , the tasks were generated in the following way.

Step 1. For each viewpoint $r \to (q, a)$ obtained from (w, q, d, a, r), we randomly selected $d'(\neq d)$ whose answer was a in the gold standard data. The condition means that the answers to q with d and d' are the same. We did not generate any task for viewpoints whose types are "wrong" since viewpoints of the type do not make any sense.

Step 2. For each d' chosen in Step 1, we generated a subjective evaluation task (Figure 2). It has two questions; The firs question is the same as q. We need it because we want to know whether the worker agrees on the conclusion of the gold standard data. The second question is to ask whether the shown viewpoint is applicable to d' assuming that the worker agrees on the conclusion.

We generated ten such tasks with different d's for each viewpoint. Note that the more positive answers to the second question we obtain for a viewpoint, the more generic the viewpoint is. Conversely, the viewpoint that obtains only negative answers to the second question is too specifi so that it can be applied to d only.

Workers. In order to take a majority vote, we recruited nine unique workers for each task by Yahoo! crowdsourcing platform. The number of workers who performed our tasks was 1058.