Instructions for Annotation of Sentences with Respect to their General Interpretive Value for a Selected Phrase

Jaromir Savelka and Kevin D. Ashley

December 5, 2018

Each of the three following sheets contains an annotation task (i.e., the "independent economic value," "identifying particular," and "common business purpose" sheets).

At the top of each sheet there is a cell with a light yellow background that contains an abbreviated text of a single statutory provision. We will call this provision the source provision. In the source provision a short phrase (a couple of words) is printed in blue. This is the phrase of interest, that is, the phrase in whose meaning we are interested.

Below the source provision there is a list of sentences retrieved from the Court Listener web service. These sentences come from the top 20 documents responsive to the query in the form of the phrase of interest in double quotes (e.g., "identifying particular") for 120 Federal jurisdictions as retrieved on February 15, 2016. The sentences are grouped according to the decisions they come from and they are listed in the order in which they appear in the decisions. The annotator's task is to evaluate each sentence in terms of its usefulness for interpretation/explanation of the phrase of interest beyond what is already known from just reading the source provision.

Specifically, the task of the annotator is to decide in which of the following four categories the sentence referring to the phrase of interest belongs:

1) **high value** – This label is reserved for sentences the goal of which is to elaborate on the meaning of the phrase of interest.

2) **certain value** – An annotator should select this label if the goal of the sentence is not to elaborate on the meaning of the phrase of interest but the sentence still provides grounds to draw some (even modest or quite vague) conclusions about the meaning of the phrase of interest.

3) **potential value** – This label is appropriate if the sentence does not appear to be useful for elaboration on the meaning of the phrase of interest but the sentence provides some additional information (even quite marginal) over what is known from the source provision.

4) **no value** – This label should be selected if the sentence does not provide any additional information over what is already known from the source provision.

For the sake of clarity let us give a couple of examples using the following artificial source provision, coming from, say, New York state law, and the phrase of interest (printed in blue):

"No vehicles are allowed in the park."

The sentences that directly elaborate on the meaning of the "vehicle" belong to the "high value" category.

1. Any mechanical device used for transportation of people or goods is a vehicle.

2. A vehicle usually has wheels, engine and controls.

The sentences that assign or contrast the phrase of interest to some other phrase also belong to the "high value" category.

3. A car is a vehicle.

4. Not every vehicle is a man-made object.

The sentences that can be used to elaborate on the meaning of the "vehicle" but do not directly elaborate on the meaning themselves belong to the "certain value" category.

5. Today I took my horse for a ride in that park where no vehicles are allowed.

6. The main reason why no vehicles are allowed in that park is to secure a tranquil environment there.

The sentences that do not seem to be useful for elaboration on the meaning of the "vehicle" but at the same time provide additional information over what is known from the source provision belong to the "potential value" category.

7. The park where no vehicles are allowed was closed during the last month.

8. The courts often need to analyze the provision stating that "No vehicles are allowed in the park."

9. The Maryland law also provides that "No vehicles are allowed in the park."

If the sentence does not provide any additional information over what is already known from the source provision it belongs to the "no value" category.

10. The provision states that: "No vehicles are allowed in the park."

11. A vehicle is forbidden from entering the park.

Finally, there are three possible scenarios that fall between the cracks of the above categorization. If they occur it is necessary to reassign the sentence to a lower label than would otherwise be assigned to the sentence according to the above rules.

These scenarios include:

a) The sentence uses the phrase of interest from a statutory provision or case law from a different jurisdiction.

b) The sentence is attributed to a person who has a personal interest in interpreting the phrase of interest in a certain way.

c) The phrase of interest in the source provision and the phrase of interest in the sentence have different meanings.

If the sentence uses the phrase of interest from a statutory provision or case law from a different jurisdiction its usefulness should be discounted. If using the standard rules the sentence is assigned to the "potential value" or the "no value" category it is not necessary to do the discounting. Such is the case of the example 9. However, if the sentence is assigned to the "high value" or the "certain value" category it should be re-assigned to a category one step lower.

12. The Indiana law states that: "Vehicle is anything which serves as a means of transport." [Although, this sentence would normally belong to the "high value" category it should be assigned to the "certain value" category to take into account that it describes the state of the affairs in a different jurisdiction.]

13. [Continuation of example 12] Therefore a car is a vehicle. [ditto]

If the sentence is attributed to a person who has a personal interest in interpreting the phrase of interest in a certain way its usefulness should be discounted. If using the standard rules the sentence is assigned to the "potential value" or the "no value" category it is not necessary to do the discounting. If the sentence is assigned to the "high value" or the "certain value" category it should be re-assigned to a category one step lower.

14. The defendant claimed he did not break the rule since roller skates cannot be considered a vehicle. [Although, this sentence would normally belong to the "high value" category it should be assigned to the "certain value" category to take into account that the defendant has a personal interest in interpreting the "vehicle" in this way.]

If the phrase of interest in the source provision and the phrase of interest in the sentence have different meanings the usefulness of the sentence should be discounted. If the meanings are significantly different then the sentence should be labeled as "no value".

15. A body is a vehicle for a soul. [Although, this sentence would normally belong to the "high value" category it should be assigned to the "no value" category to take into account that it uses the "vehicle" in a significantly different meaning.]

If the meanings are different but strongly related the sentence should be re-assigned to a category one step lower if it would normally be labeled with the "high value" or the "certain value" category. If it would normally be labeled with the "potential value" or the "no value" category the usefulness of the sentence should not be discounted.

16. Any autonomous vehicle is subject to the approval of the executive committee. [Although, this sentence would normally belong to the "certain value" category it should be assigned to the "potential value" category to take into account that the "vehicle" is used in a slightly different meaning in the sentence.]

It may happen that the annotator encounters a sentence that does not clearly belong to any one of the four categories. If the annotator is deciding between the two adjacent categories (e.g., between the "potential value" and the "certain value" categories) he should assign the sentence with the one he intuitively feels as more appropriate. However, if the doubt is more serious and the annotator is not deciding merely between the two adjacent categories, the sentence should be flagged as problematic and it should be explained why the annotator thinks it does not belong to any one of the four categories.