### Manual Evaluation of Data/Text Correspondence: Instructions for Evaluators

#### Introduction

In this work, you will be shown 100 different data/text pairs (10 in the training session). Your task is to assess the fluency of the texts, and, separately, the similarities and differences between the data and the text.

The following is an example of a data/text pair:

| **Data** | **Text** |
| --- | --- |
| cityServed(Adirondack\_Regional\_Airport, Lake\_Placid,\_New\_York),  runwayLength(Adirondack\_Regional\_Airport, 2003.0),  cityServed(Adirondack\_Regional\_Airport, Saranac\_Lake,\_New\_York) | adirondack regional airport serves both lake placid and saranac lake. its 's runway length is 2003.0. |

#### Data Expressions

The data is a list of expressions of the form predicate(argument 1, argument 2), and can be interpreted as follows, using the above data expression as an example:

cityServed(Adirondack\_Regional\_Airport, Lake\_Placid,\_New\_York) – the cityServed of Adirondack\_Regional\_Airport is Lake\_Placid,\_New\_York, i.e. Adirondack\_Regional\_Airport serves the city of Lake\_Placid,\_New\_York

runwayLength(Adirondack\_Regional\_Airport, 2003.0) – the runwayLength of Adirondack\_Regional\_Airport is 2003.0

cityServed(Adirondack\_Regional\_Airport, Saranac\_Lake,\_New\_York) – the cityServed of Adirondack\_Regional\_Airport is Saranac\_Lake,\_New\_York, i.e. Adirondack\_Regional\_Airport serves the city of Saranac\_Lake,\_New\_York

Note that predicate and arguments are not strictly speaking words, but concepts. E.g. “Lake\_Placid,\_New\_York” is a single unique identifier for this particular city, and its content is therefore the city, not the city and the state (New York) both.

#### Texts

The text is supposed to express the content (meaning) of the data expression, but in standard fluent language. While texts often use different *words* than can be recognised in the concepts in the data, such words should nevertheless have the same *meaning* as the corresponding part of the data. Some of the texts you will encounter will be more fluent, others will be less fluent.

As you can see in the example above, texts can be lower case throughout, without any capital letters. While normally this would be considered wrong, for the purpose of the present evaluation it is irrelevant and we would like you to ignore it.

#### Evaluation Process

##### Assessment of fluency

First, looking just at the text, you need to rate the fluency of the text, as in the example on the right.

In this example, there are two texts to assess. Assessment fields will initially be empty and yellow, and flagged “↓ incomplete”. You need to click on the downward-pointing triangle to the right to select from a scale of 1 to 5 to express your assessment of the fluency of the text (5 being the highest rating).

First, read the instructions at the top of the table. Then, read the next text to assess, and select a rating. Once you have selected a rating, the field will turn green to indicate that you have completed it.

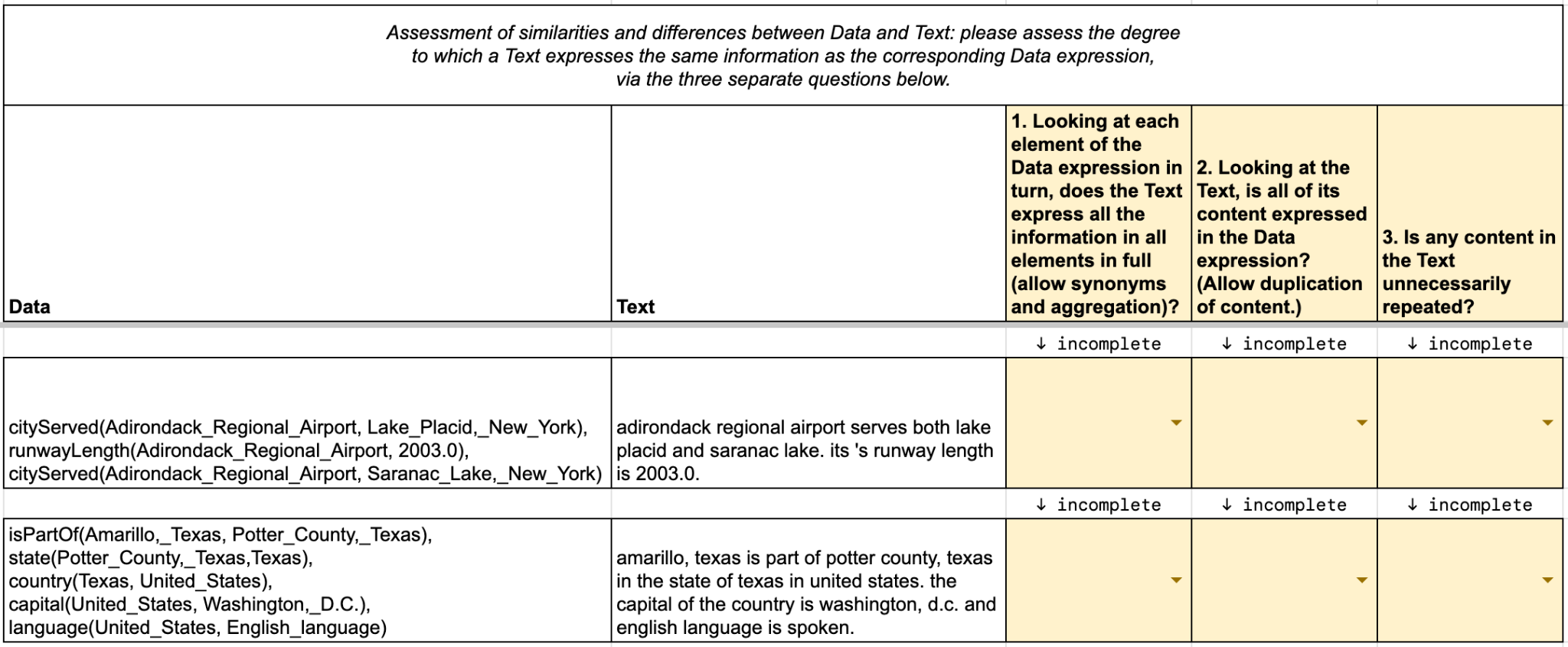
You can change your rating(s) as many times as you like.

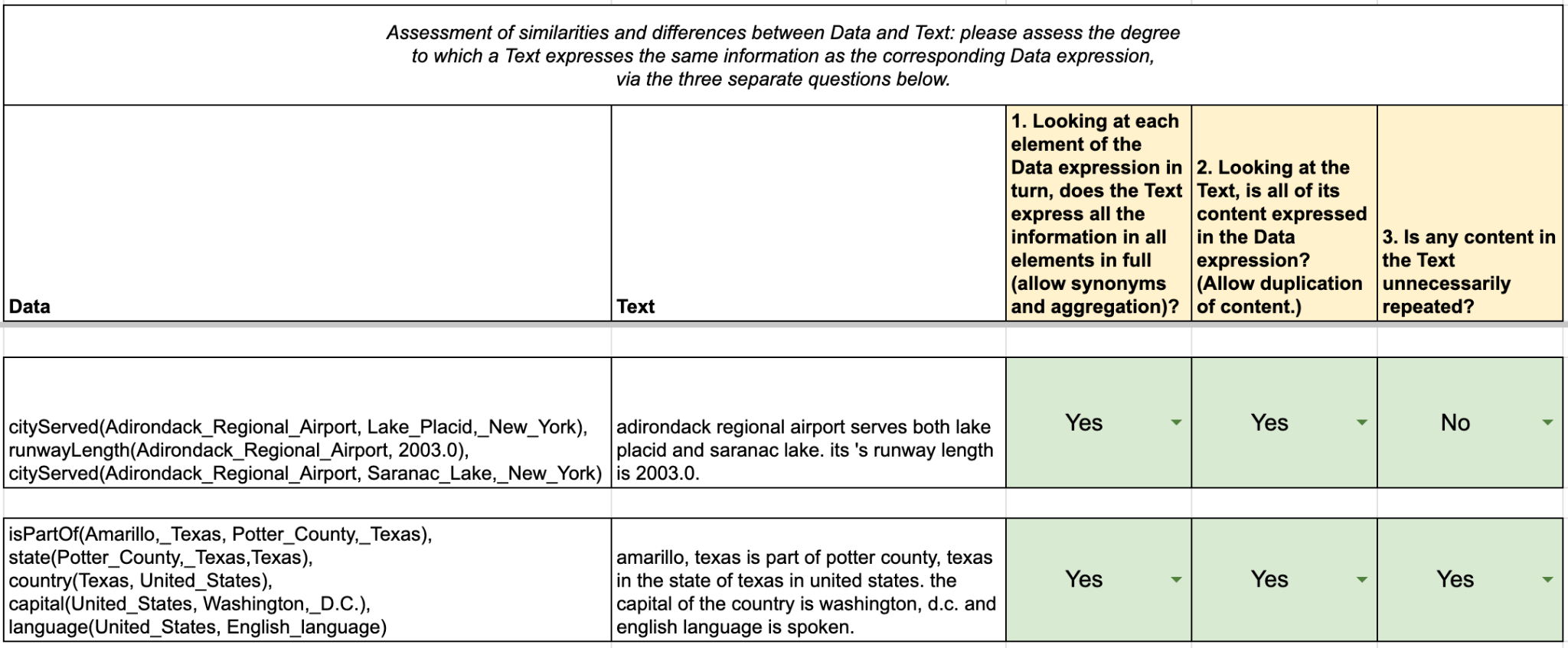
Please ignore lower-casing vs. upper-casing, formatting of numbers and layout issues in assessing the fluency of a text.

Note that there is no single ‘correct’ fluency rating – you may find a text more, or less, fluent then another person. In the above example, the evaluator has put a 3 in the rating field, because of the superfluous occurrence of “‘s”, but you could have (legitimately) chosen another rating, e.g. a 4.

##### Assessment of similarities and differences between data and text

Next, you need to assess the similarities and differences between the data and the text in three separate steps, from 1 to 3, in this order. The first screenshot below shows two data/text pairs with empty assessment fields, the second shows the same two texts with fields fully completed.





Below we consider each of the three questions in turn, and explain how the evaluator in this example arrived at the assessments in the spreadsheet above for the first data/text pair.

Note that data or text may occasionally contain factual inaccuracies. Please ignore these. In this evaluation, we are only interested in similarities/differences between data and text.

1. *Looking at each element of the Data expression in turn, does the Text express all the information in all elements in full (allow synonyms and aggregation)?*

Our task is to look at each predicate-argument expression, and each of its three components, i.e. predicate, argument 1, and argument 2, in turn, considering whether they are fully expressed by the text.

cityServed(Adirondack\_Regional\_Airport, Lake\_Placid,\_New\_York): All three components are expressed by “adirondack regional airport serves … lake placid..” Recall that Lake\_Placid,\_New\_York identifies only the city, so expressing the state is optional.

runwayLength(Adirondack\_Regional\_Airport, 2003.0): All three components are covered by “its 's runway length is 2003.0.” Here, the possessive pronoun “its” is a valid way of expressing Adirondack\_Regional\_Airport, because it has already been mentioned.

cityServed(Adirondack\_Regional\_Airport, Saranac\_Lake,\_New\_York): All three components are expressed by “adirondack regional airport serves … saranac lake.” As explained above, Saranac\_Lake,\_New\_York is a single identifier for the city of Saranac Lake, therefore expressing the state is optional.

Note that the “both … and“ construction is a valid way of avoiding repetition of “adirondack regional airport serves”. This is a form of aggregation.

1. *Looking at the Text, is all of its content expressed in the Data expression? (Allow duplication of content.)*

Here our task is to look at each part of the text in turn, considering whether each is fully expressed by the data.

“adirondack regional airport serves”: This part of the text is expressed by two occurrences of cityServed(Adirondack\_Regional\_Airport, ...)

“both lake placid and saranac lake”: This part of the text is expressed by Lake\_Placid,\_New\_York and Saranac\_Lake,\_New\_York

“its 's runway length is 2003.0.”: This last part of the text is fully expressed by runwayLength(Adirondack\_Regional\_Airport, 2003.0)

Recall that variations in the wording or formatting are fine, as long as the meaning (content) is the same.

In answering Question 2, please ignore any duplication of content. E.g. if “adirondack regional airport serves” had occurred twice in the text, that would have been fine. However, if there is duplication, there should be as many occurrences of the duplicated content on the data side as on the text side.

1. *Is any content in the Text unnecessarily repeated?*  There is no repetition at all so this is a No.

In accordance with the above reasoning, we mark both Question 1 and Question 2 Yes, and we mark Question 3 No.

You can change selected values as many times as you like. If you want to reset any field to empty, simply click on the field once, then hit the delete button (back arrow).

#### Next Steps

Now please go to the spreadsheet in the same folder as this instruction document and complete the evaluations in it.

Once you’re happy with the evaluations, please get in touch with [anya.belz@adaptcentre.ie](mailto:anya.belz@adaptcentre.ie) to let us know you’ve finished and any questions you may have.