

Human Assessment

This study is carried out as part of [REDACTED], specifically the study on error annotation of system outputs generated by natural language generation (NLG) systems. The project aims to check for semantic inconsistencies in existing NLG system outputs and to assess the informativeness of our error taxonomy by performing a comparative study of these systems (traditional vs. neural as in the WebNLG shared tasks) to see which system performs better and why.

* Indicates required question

1. Email *

2. English language proficiency (as per [IELTS CEFR scale](#)): *

Mark only one oval.

- ☐ Proficient user (C1,C2)
- ☐ Independent User (B1,B2)
- ☐ Basic User (A1, A2)

3. Prior experience with error annotation (i.e. labelling and categorising text data, providing the instructions and the error types):

Mark only one oval.

- ☐ Yes
- ☐ No

Note

The following texts were generated by existing NLG systems. We will ask participants to contact [REDACTED] in the unlikely event that they encounter any offensive or inappropriate content to them personally. We will also provide the following link: [REDACTED]

4. Example of error annotation:

Instructions:

Step 1: Examine whether each element in the input triples is verbalised or not. Also, check whether the triples are in the correct order in the verbalisation such that if there is a repetition of a certain element of the triple in its verbalisation, check for this element in the triple. Otherwise, mark the element as an **omission** error type in the triple.

Step 2: Examine whether all the content phrases in verbalisation include all the elements in the triples. If the content phrase is an extra in the verbalisation without a relation in the input triples, mark it as an **addition** error type.

Step 3: Examine whether any part of the verbalisation is repeated, including close paraphrases. This is the case e.g. if an element in the triples is rendered more than once. If there is a content phrase that is repeated in this sense, mark it as a **repetition** error type.

Step 4: Examine whether all the content phrases in verbalisation include all the elements in the triples. If the content phrase in the verbalisation is an extra and replacing an element in the input triples, mark it as a **substitution** error in the verbalisation.

Input triple: Arlington,_Texas | elevationAboveTheSeaLevel | 184.0
Arlington,_Texas | areaTotal | 258.2 (square kilometres)

Verbalisation: arlington, texas, 184.0, has a total area of 258.2 square metres.

Check all that apply.

- ☐ Omission
- ☐ Addition
- ☐ Repetition
- ☐ Substitution
- ☐ Other: _____