

**Our proposed approach: Parse Tree Fragmentation**

- Identify well-formed syntactic structures for the parts that make sense
- **Parse tree fragmentation** is the process of breaking up the tree
- **Fragments** are reasonable isolated parts of parse trees

---

**Developing a Fragmentation Corpus**

1) **Pseudo Gold Fragmentation (PGold)**

Given an ungrammatical sentence and its error corrections:

- ESL sentence: *I am very good at swimming.
- Teacher corrections: *I am very good swimming.

Replacing error Unnecessary error Missing error

---

2) **REference Fragmentation (REF)**

Given an ungrammatical sentence and a grammatical version of the same sentence:

- Automatically find alignments between two trees
- Because we don’t necessarily know what the error is without some detailed human correction annotations
- Assign fragments to aligned nodes

---

**Experiments**

**Extrinsic Evaluation: Fluency Judgment**

Binary classification: a sentence has virtually no error or many errors
Regression: Predict number of errors in ESL dataset or HTER in MT dataset

**Our feature set**: number, avg, size, min size, max size of fragments

---

**Conclusion**

- Introducing the new task of **parse tree fragmentation**
- Extracting gold fragments using existing corpora for other NLP applications
- Proposing two practical fragmentation methods