Conference Paper Review

**Paper ID**: 002

**Title**: Comparing Annotation Guidelines and Annotators: A Preliminary Investigation into the Lexical Complexities Assigned in Two Complex Word Identification Datasets

**Author(s)**: Kai North

**Reviewer's Recommendations:**

1. **Writing**
   * Minor improvement suggested
2. **Novelty**
   * Original
3. **Suitability**
   * Very related
4. **Reviewer's Expertise**
   * Knowledgeable
5. **Recommendation**
   * Absolute accept

**Reviewer's detailed comments: strength, weakness, suggestion**

1. **Comments for the Authors:**
   1. **Strengths**

The paper's objective to compare the effects of annotation guidelines on the quality of complex word identification datasets is made well clear in the abstract. Likewise, the paper's significance has been well stated by giving examples of some application areas like text simplification and lexical simplification.

The related works have been explained very well. It includes the current progress and the changes in methodology due to the introduction of Lexical Complexity Prediction.

The two different annotation tasks and the two datasets and processes involved have been explained and differentiated well. I also liked the structure of introducing the two well-defined hypothesis and validating them through the results later. The results are well presented, as well.

The conclusion covers all the key contributions of the paper. Likewise, the impact and significance of the research have been well stated in the future recommendations section.

* 1. **Weaknesses and Suggestions**

I found a few grammatical errors in the paper. For example, the first sentence in the abstract, no space after a period in section 4.1 second sentence, spelling error in "instance" in section 6.3, etc.

Although the disadvantages of annotators' choice have been clearly stated in section 3.1, the advantages are missing. Likewise, removing people not doing their tasks properly has been proposed to eliminate financial incentives in the annotation. However, it is not mentioned clearly how such people are selected.

A suggestion is to make the research question proposed measurable in terms of quantitative or qualitative measurements, if possible.

The formula for Keyness contains the term (Ok – Ek)2. However, the explanation mentions that "the observed frequency values of each complex word: 'O' were then subtracted from their expected values: 'E'". Although the results will be the same, the sentence can be made consistent with the formula.

A concern is that only a sample of words was used to validate the hypothesis, which might not be accurate.

One suggestion is to shorten the title if possible. Although the title is well descriptive, I felt that the title is too long.

1. **Comments for the Program Committee (will be kept confidential and NOT released to the authors)**

The paper presents the effects of the annotation guidelines on the quality of complex word datasets quite well. It has possible impact in creation of similar datasets in the future, which can be used in various application areas.