Glove et al., 2019), as SAT can automatically segment raw comprehension can be shown in Appendix B. SAT is an extension of Seg-
ification of alignment data, we developed an online tool manual annotation procedure is challenging and time-
Figure 3: Alignment word
now, a product developed by a Japanese company (may offer a solution). In 2005 the company, Nippon Basic Co., Ltd. called Glove
“to solve this issue.”
“to solve this issue.”
“to solve this issue.”

THE IDEA UNIT
In Applied Linguistics, the Idea Unit (IU) is a “chunk of information which is viewed by the speaker/writer cohesively as it is given surface form” (Kroll, 1977). The IU can be used to assess students’ listening comprehension and written recall via segmentation and alignment (Figure 1). We expand upon our previous work (Geczhe et al., 2019) and release an updated Idea Unit Annotation Guideline (Figure 2).
Our tests show that the new annotation guidelines improve the inter-annotator agreement of the idea unit from 0.547 to 0.785 of Cohen’s κ (Cohen, 1960).

CORPUS: L2WS 2021
We release an Idea Unit gold standard corpus L2WS 2021 (L2 Written Summary). The corpus is comprised of 40 summaries written by 40 university students as part of a course assignment. All the summaries refer to a source text that describes a new device that can purify water without electricity. This source text is included in the corpus. The students were asked to read the source text (391 words) and summarize its main ideas and key details in approximately 80 words. All the students speak Japanese as a first language. The data is manually annotated according to the IU annotation guidelines released with this paper. An additional dataset comprised of 80 summaries, L2WS 2020, was also collected. However, this dataset cannot be shared with the public due to a lack of consent for sharing from the part of the students. L2WS 2020 was used exclusively for developing and testing the automatic segmentation algorithm IUExtract.

AUTOMATIC SEGMENTATION ALGORITHM: IUExtract
IUExtract is an automatic rule-based segmentation algorithm released as a python package. We developed the algorithm by deriving the annotation guidelines into a rule-based segmentation algorithm. We tested this algorithm against the L2WS 2020 test set and L2WS 2021 corpus. The algorithm was evaluated in terms of Precision, Recall, F1 score and Perfect IU ratio. The formulas for Precision, Recall, and F1 score are the following:

\[
\text{Precision} = \frac{|\text{AutoBoundaries} \cap \text{GoldBoundaries}|}{|\text{AutoBoundaries}|} \\
\text{Recall} = \frac{|\text{AutoBoundaries} \cap \text{GoldBoundaries}|}{|\text{GoldBoundaries}|} \\
F_1 = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

where AutoBoundaries is the set of automatically extracted boundaries by the algorithm and GoldBoundaries is the set of manually annotated segment boundaries.

Table 2: Evaluation results for the segmentation algorithm. Average IU length, variance, Precision, Recall and F1 score are all micro-averaged.

<table>
<thead>
<tr>
<th>IUExtract</th>
<th>L2WS 2020</th>
<th>L2WS 2021</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUExtract Gold</td>
<td>#IUs</td>
<td>#Docs</td>
</tr>
<tr>
<td>IUExtract Gold</td>
<td>1264</td>
<td>1174</td>
</tr>
<tr>
<td>IUExtract</td>
<td>76</td>
<td>63</td>
</tr>
<tr>
<td>#IUs</td>
<td>6.449</td>
<td>7.118</td>
</tr>
<tr>
<td>#Docs</td>
<td>10.59</td>
<td>10.27</td>
</tr>
<tr>
<td>#Avg Tokens</td>
<td>0.800</td>
<td>–</td>
</tr>
<tr>
<td>#Avg IUs</td>
<td>0.806</td>
<td>–</td>
</tr>
<tr>
<td>F1 Score</td>
<td>0.835</td>
<td>–</td>
</tr>
</tbody>
</table>

Table 3: A screenshot of the alignment section of SAT.

ALIGNMENT COLLECTION PLATFORM: SAT
We tested the alignment algorithm proposed in our previous work (Geczhe et al., 2019) with the more recent word-embedding models. We tested GloVe embeddings (Pennington et al., 2014), SpaCy’s Word2Vec implementation (Honnibal and Johnson, 2015) and Sentence-BERT (Reimers and Gurevych, 2019).
The results are insufficient for effective alignment, with the best model, SBERT, sporting only 0.375 in maximum precision and 0.415 in maximum recall.
We developed a Segmentation and Alignment Tool – SAT to facilitate the collection of new alignment gold standard data. SAT is a website that can be used by annotators to link Idea Units across texts in a graphical manner.

AKNOWLEDGEMENTS
This work was supported by the Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (B) [No. 20308192; PI: Yasuyo Sawaki].

REFERENCES

Figure 1: An example of Idea Unit segmentation and alignment.