How to Perform the Divisive Alignment

1 Tool

Firefox is the single required tool to perform the divisive alignment. No other navigator can do the job, be it IE, Chrome, Safari or Opera.

Under **Firefox**, go to the following Web site and login.

$$\boxed{\texttt{https://yawat-annot.limsi.fr/divisive}}$$

We can visit the site at the LIMSI. However, connections from outside the LIMSI are restricted due to security reasons. We must ask the IT service of the LIMSI to allow exterior visits, by providing the IP address.[1] However, one might find his/her IP address changing regularly. If this is the case, **please tell Yong the new address as early as possible.**

The Web site hosts an annotation interface called Yawat. We will explain how to use Yawat later. Figure 1 shows how it looks like.

Figure 1: A screenshot of Yawat, with three sentence pairs.

2 What is divisive alignment

In general, in Yawat, we have one French sentence (or several ones) on the left, and the translation in English on the right. We ask the following difficult question:

Which part of the French sentence(s) corresponds to which part of the English one(s)?

Divisive alignment tries to answer the question with a top-down approach. We first identify the correspondences between large groups of words, then refine them to obtain finer-grain matchings. In practice, we follow the general procedure:

1. split the French side into two segments, say into F_l (left) and F_r (right);

2. split the English side into two segments, say into E_l (left) and E_r (right);

3. establish the correspondence between the four segments. There are two possibilities:

   - F_l matches E_l, F_r matches E_r (called *direct split*);
   - F_l matches E_r, F_r matches E_l (called *inverse split*);

   and we choose the better one.

Now, suppose we have chosen the direct split, leading to two new pairs: F_l with E_l, F_r with E_r. For each new pair, we re-perform the procedure.

This is terribly abstract. In the next, we illustrate the procedure with actual examples, in the Yawat environment.

---

[1]To obtain this address, visit `https://www.iplocation.net/find-ip-address`

3 How to do divisive alignment in Yawat

We show how to do this in Yawat for the first sentence pair in Figure 1. When we log into Yawat, we encounter the pair as this:
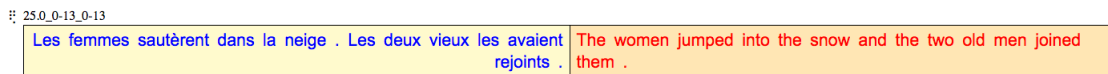
⁞ 25.0_0-13_0-13

| Les femmes sautèrent dans la neige . Les deux vieux les avaient rejoints . | The women jumped into the snow and the two old men joined them . |

Figure 2: The example sentence pair.

How to split the French side? Choose a word on the French side, **left-click on it**. In Figure 3, I have left-clicked on the period in the middle. Observe that a rectangle pops out.

⁞ 25.0_0-13_0-13

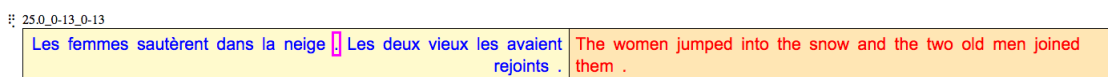| Les femmes sautèrent dans la neige . Les deux vieux les avaient rejoints . | The women jumped into the snow and the two old men joined them . |

Figure 3: Left click on the middle period on the French side.

What does this do? By clicking on one word, we **virtually** split the French side into two segments: the left side of the word, and the right side of it. In our example, *Les femmes sautèrent dans la neige .* becomes F_l, and *Les deux vieux les avaient rejoints .* becomes F_r. Note: the word on which we left-click **belongs always to its left-side part.**

How to split the English side? *After* we have *virtually* splitted the French side, we do the same on the English side: choose a word, and **left-click** on it. In Figure 4, we have left-clicked on the word *and*.

⁞ 25.0_0-13_0-13

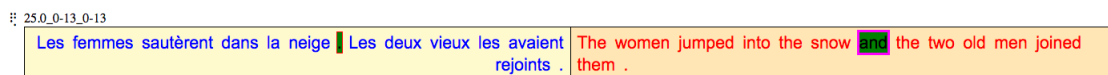| Les femmes sautèrent dans la neige . Les deux vieux les avaient rejoints . | The women jumped into the snow and the two old men joined them . |

Figure 4: Left click on the word *and* on the English side.

Notice that this time Yawat marks the two words on which we have clicked. This is a signal that we have *virtually* splitted both sides.

What we have now? Although the physical split has not taken place, we know that, once it happens, there would be four segments:

- F_l: *Les femmes sautèrent dans la neige .*
- F_r: *Les deux vieux les avaient rejoints .*
- E_l: *The women jumped into the snow and*
- E_r: *the two old men joined them .*

Clearly, the direct split is correct (that is, F_l with E_l, F_r with E_r). How to tell Yawat about this?

How to indicate direct split or inverse split? *After* we have chosen a word on both sides, we need to choose the split direction (direct or inverse). To do this, we **right-click** on either of the two chosen words. In Figure 5, I have right-clicked on the word *and*.

⁞ 25.0_0-13_0-13

| Les femmes sautèrent dans la neige . Les deux vieux les avaient rejoints . | The women jumped into the snow them . |

remove 'and' from this group
dissolve this group
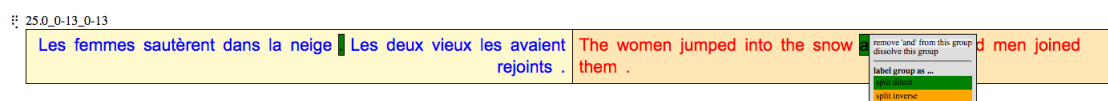label group as ...
split direct
split inverse

Figure 5: Right-click on the chosen word *and*.

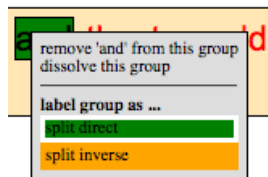Now, Yawat pops up a menu, zoomed in in Figure 6.

Figure 6: The menu, appeared after we right-clicked on the chosen word *and*.

In the menu, under the line *label the group as ...*, there are two options: *split direct* and *split inverse.* To choose direct split, we **left-click** on the option *split direct.* Figure 7 shows the result.
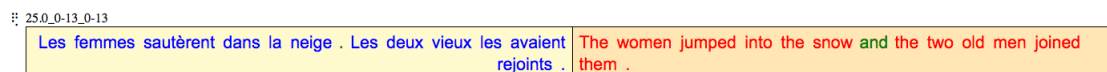

Figure 7: Finished the procedure for the example.

This ends the procedure for annotating the example. Note that both chosen words remain green. For Yawat to record our annotation, click on the **save** button on the top-right corner of the page.
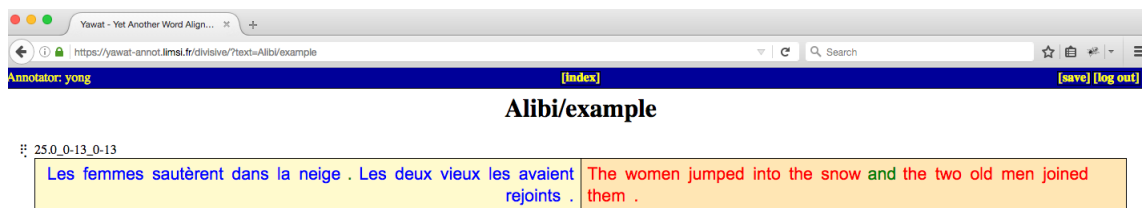

Figure 8: Don't forget to click on the *save* button regularly.

When does the actual segmentation happen?   For one pair, the annotations so far contain all information needed to perform the segmentation. However, for now, the actual segmentation is done by a program that is not (yet) a part of Yawat. So, when a page of pairs are all annotated, **tell Yong to carry out the actual segmentation.**[2] We are trying to make this step automatic.

After the actual segmentation is done, the new pairs are shown in Figure 9.
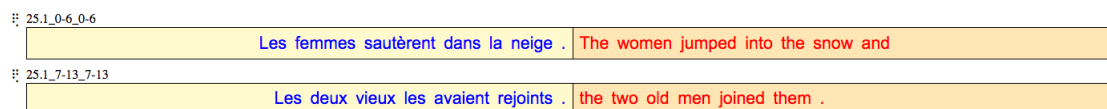

Figure 9: The two new pairs, issued from the segmentation.

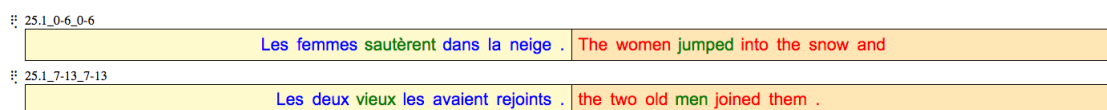The annotations for them are in Figure 10.

Figure 10: The annotations for the two new pairs.

---

[2]Before this, be sure that the annotations have been saved.

What if we want to correct the annotation? Say, we have chosen the two words on the two sides, but we find better ones and want to change the decision. In that case, **right-click** on one chosen word. The menu will pop up. Then **left-click** on the option *dissolve this group*. **Do this on both chosen words.**

However, this kind of correction is only possible **before the physical segmentation takes place**. Once the actual segmentation is done, it is difficult to step back. So please check the annotations before the actual segmentation.

When is the inverse split used? Our example is a pair of very literal translations, so mostly we use the direct split. However, eventually we will arrive at the following pair in Figure 11.

25.3_10-12_11-12
les avaient rejoints | joined them

Figure 11: A pair that needs inverse split.

We know that *les* corresponds to *them*, and *avaient rejoints* corresponds to *joined*. To indicate this, we first left-click on the words *les* and *joined*, and obtain four segments:[3]

- E_l: *les* (*les* is the only word of its left-side part)

- E_r: *avaient rejoints*

- F_l: *joined* (*joined* is the only word of its left-side part)

- F_r: *them*

Now we right-click on *joined*, and choose the *split inverse* option in the menu. The result is in Figure 12.
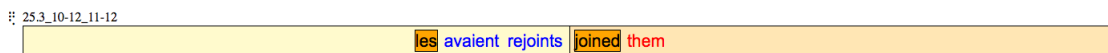
25.3_10-12_11-12
les avaient rejoints | joined them

Figure 12: Inverse split.

After the physical segmentation, we obtain the desired new pairs.

25.4_10-10_12-12
les | them
25.4_11-12_11-11
avaient rejoints | joined

Figure 13: New pairs resulted from the inverse split.

---

[3]Recall that, the word on which we left-click belongs always to its left-side part.