

# MRK Alignment Scoring Guidelines

Christopher Handy & Marieke Meelen

Version	Date	Changes	Contributor(s)
1.0	2 January 2020	First complete version	Chris & Marieke
2.0	27 March 2020	Added table of contents, getting started instructions and section on (new) input instructions & options	Marieke
2.1	14 April 2020	Added content for section §5 on CSV upload	Marieke
2.2	4 May 2020	Updated content on CSV upload	Marieke

## Table of content

§1 Introduction

§2 Getting Started

§3 Identifying & Entering Alignments

§4 Scoring alignments

§5 Upload CSV with alignments and scores

## §1 Introduction

The aim of the Alignment Input & Scoring Tool is to facilitate the creation of training data for an automatic crosslinguistic alignment engine.

This document provides descriptions and examples of manually aligned passages in Classical Tibetan and Classical Chinese texts, including instructions for identification, input and scoring of alignments. The alignment engine developed for the Open Philology project will automatically detect bilingual aligned passages within the MRK or other text collections using training data created with this manual alignment input & scoring tool; it is important to create training data in order to increase the accuracy of the automated system. These training data consist of manually aligned passages from the MRK with a score indicating the strength of each alignment.

Note that in principle any other witnesses in any language can be added to the database, for now, all MRK Witnesses from the Derge (Tibetan) and Taisho (Chinese) collections are there and Sanskrit will follow as soon as reliable digital witnesses are available.

## §2 Getting Started

To add alignments and scores to the database, you need the following:

- (reasonably) up-to-date version of the [Mozilla Firefox browser](#)
- know which Witnesses you are working with (e.g. 'Derge 75' and 'Taisho 310-31')
- open the following link with Firefox: <https://engine.openphilology.eu/optools/>
- obtain username and password from Christopher Handy ([c.a.handy@hum.leidenuniv.nl](mailto:c.a.handy@hum.leidenuniv.nl))

Note that further documentation & code can be found here:

<https://github.com/handyc/openphilology.eu>

## §3 Identifying and Entering Alignments and Scores

### 3.1 What are cross-linguistic alignments?

Cross-linguistic alignments can in principle be identified on various levels, e.g. words, sentences, paragraphs or entire witnesses. For the purposes of creating useful training data for the Open Philology model, we consider the following parameters:

Minimum length: sentence (or clause, if preferred from a philological perspective)

Maximum length: “paragraph”

The OP source texts generally do not identify individual paragraphs. For NLP purposes, it is useful to keep alignments as short as possible, but not shorter than the sentence/clause level. “Paragraphs” are therefore indicated as minimally coherent sets of sentences. If an individual paragraph contains several sentences that could be aligned separately, these sentences should be manually marked as aligned sentences. Passages of text for which alignment can really not be determined at the sentence or clause level should be manually aligned as paragraphs. At this time, the software does not make any formal distinction between a sentence and a paragraph.

### 3.2 How to enter alignments and scores?

Manual alignments are added to the database using the OPTools utility

<https://engine.openphilology.eu/optools/> - see image of homepage below:

The screenshot displays the OPTools utility interface. At the top, there are two witness text boxes, labeled 'Witness 1: D45' and 'Witness 2: D45'. Both boxes contain the same text in a Tibetan script. Below the text boxes, there are two large empty rectangular areas, each with the text '[THIS SPACE INTENTIONALLY LEFT BLANK]'. At the bottom of the interface, there are several control elements: a 'Witness' dropdown menu set to 'D45', a 'Provisional Subsegment' input field, a 'Score' dropdown menu set to 'N/A', and a 'Submit' button. To the right of these controls, there is a section titled 'Alignments'.

The current implementation of this utility is a Django app. To access the tool, please obtain a login and password from Chris Handy ([c.a.handy@hum.leidenuniv.nl](mailto:c.a.handy@hum.leidenuniv.nl)).

This utility connects directly to the project database, so that any manual alignments entered using the tool are instantly part of the project database.

There are two ways to submit and alignments and scores with the tool:

Option 1. By selecting and saving alignments using the main window (see §3.3 below)

Option 2. By uploading a CSV file (see section §5 below)

### 3.3 Selecting & Scoring alignments with the tool ('Option 1')

Option 1 for submitting alignments and scores uses the tool with the following 5 steps:

1. Choose two witnesses from the drop-down menus
2. Highlight the text to be aligned within each witness by selecting it
3. Check if the correct parts are selected in the selection windows
4. Score the alignment with the dropdown list of scores
5. Submit the alignment to the database with the submit button

#### Step 1 – Choose two witnesses from the drop-down menus

Witness 1: D75

Witness 2: T11n0310\_30

@##/\_/'phags pa dkon mchog brtsegs pa chen po'i chos kyi rnam grangs le'u stong phrag brgya pa las le'u sum cu rtsa gcig pa ste/\_gang gA'i mchog gis zhus pa lung bstan pa/\_bam po phyed pa/\_rgya gar skad du/\_Ar+ya gang+go t+ta ra pa ri pr-i ts+tshA nA ma ma hA yA na sU tra/\_bod skad du/\_'phags pa gang gA'i mchog gis zhus pa zhes bya ba theg pa chen po'i mdo/\_sangs rgyas dang byang chub sems dpa' thams cad la phyag 'tshal lo/\_/'di skad bdag gis thos pa dus gcig na/\_bcom ldan 'das mnyan yod na rgyal bu rgyal byed kyi tshal mgon med zas sbyin gyi kun dga' ra ba na bzhugs te/\_de'i tsho grong khyer chen po mnyan yod na dge bsnyen ma gang gA'i mchog ces bya ba zhis gnas so/\_/de nas dge bsnyen ma gang gA'i mchog mnyan yod nas byung nas rgyal bu rgyal byed kyi tshal mgon med zas sbyin gyi kun dga' ra ba ga la ba der song ste phyin nas bcom ldan 'das kyi zhabs la mgo bos phyag 'tshal te phyogs gcig tu 'dug go/\_/phyogs gcig tu 'dug pa dang /\_ bcom ldan 'das kyi mkhyen bzhi du dge bsnyen ma gang gA'i mchog la gang gA'i mchog khyod da gzod gang nas 'ongs zhes smras so/\_/gang gA'i mchog gis gsol pa/\_bcom ldan 'das gang zhig gis sprul pa la smras te/\_khyod da gzod gang nas 'ongs zhes de skad bgyis na ji skad lan 'debs par 'gyur lags/\_bcom ldan 'das kyi bka' stsal pa/\_gang gA'i mchog sprul pa la ni 'greng ba med/\_/'dug pa med/\_nyal ba med/\_ong ba med/\_gro ba med/\_chi 'pho med/\_skye ba med na 'di nas 'ongs so zhes de ci zhig ston par 'gyur/\_gang gA'i

如是我聞：一時佛在王舍城耆闍崛山中，與大比丘眾千二百五十人，菩薩摩訶薩十千人俱。時王舍城有長者女名為妙慧，年始八歲，面貌端正容色姝好，諸相具足見者歡喜，曾於過去無量諸佛親近供養種諸善根。時彼女人詣如來所，頂禮佛足右遶三匝，長跪合掌而說偈言：「無上等正覺，為世大明燈；菩薩之所行，唯願聽我問。」佛告妙慧：「今恣汝問。當為解說，令斷疑網。」爾時妙慧即於佛前以偈問曰：「云何得端正，大富尊貴身？復以何因緣，眷屬難沮壞？云何見己身，而受於化生，千葉蓮花上，面奉諸世尊？云何能證得，自在勝神通。」

[THIS SPACE INTENTIONALLY LEFT BLANK]

[THIS SPACE INTENTIONALLY LEFT BLANK]

Witness D75 ( ) clear1  
Witness T11n0310\_30 ( ) clear2  
Provisional Subsequent §  
Score N/A  
Submit

Alignments

Witness 1: D75

Witness 2: T11n0310\_30

@##/\_/'phags pa dkon mchog brtsegs pa chen po'i chos kyi rnam grangs le'u stong phrag brgya pa las le'u sum cu rtsa gcig pa ste/\_gang gA'i mchog gis zhus pa lung bstan pa/\_bam po phyed pa/\_rgya gar skad du/\_Ar+ya gang+go t+ta ra pa ri pr-i ts+tshA nA ma ma hA yA na sU tra/\_bod skad du/\_'phags pa gang gA'i mchog gis zhus pa zhes bya ba theg pa chen po'i mdo/\_sangs rgyas dang byang chub sems dpa' thams cad la phyag 'tshal lo/\_/'di skad bdag gis thos pa dus gcig na/\_bcom ldan 'das mnyan yod na rgyal bu rgyal byed kyi tshal mgon med zas sbyin gyi kun dga' ra ba na bzhugs te/\_de'i tsho grong khyer chen po mnyan yod na dge bsnyen ma gang gA'i mchog ces bya ba zhis gnas so/\_/de nas dge bsnyen ma gang gA'i mchog mnyan yod nas byung nas rgyal bu rgyal byed kyi tshal mgon med zas sbyin gyi kun dga' ra ba ga la ba der song ste phyin nas bcom ldan 'das kyi zhabs la mgo bos phyag 'tshal te phyogs gcig tu 'dug go/\_/phyogs gcig tu 'dug pa dang /\_ bcom ldan 'das kyi mkhyen bzhi du dge bsnyen ma gang gA'i mchog la gang gA'i mchog khyod da gzod gang nas 'ongs zhes smras so/\_/gang gA'i mchog gis gsol pa/\_bcom ldan 'das gang zhig gis sprul pa la smras te/\_khyod da gzod gang nas 'ongs zhes de skad bgyis na ji skad lan 'debs par 'gyur lags/\_bcom ldan 'das kyi bka' stsal pa/\_gang gA'i mchog sprul pa la ni 'greng ba med/\_/'dug pa med/\_nyal ba med/\_ong ba med/\_gro ba med/\_chi 'pho med/\_skye ba med na 'di nas 'ongs so zhes de ci zhig ston par 'gyur/\_gang gA'i

如是我聞：一時佛在王舍城耆闍崛山中，與大比丘眾千二百五十人，菩薩摩訶薩十千人俱。時王舍城有長者女名為妙慧，年始八歲，面貌端正容色姝好，諸相具足見者歡喜，曾於過去無量諸佛親近供養種諸善根。時彼女人詣如來所，頂禮佛足右遶三匝，長跪合掌而說偈言：「無上等正覺，為世大明燈；菩薩之所行，唯願聽我問。」佛告妙慧：「今恣汝問。當為解說，令斷疑網。」爾時妙慧即於佛前以偈問曰：「云何得端正，大富尊貴身？復以何因緣，眷屬難沮壞？云何見己身，而受於化生，千葉蓮花上，面奉諸世尊？云何能證得，自在勝神通。」

[THIS SPACE INTENTIONALLY LEFT BLANK]

[THIS SPACE INTENTIONALLY LEFT BLANK]

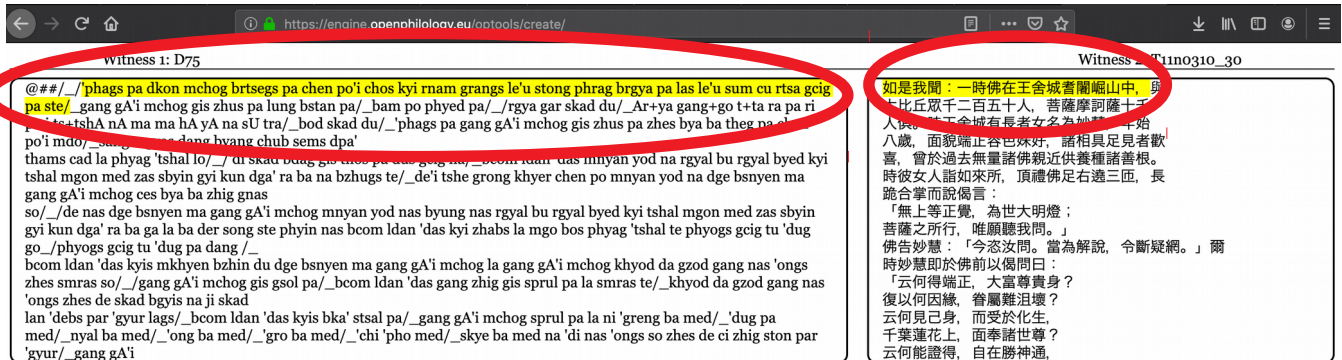
Witness D75 ( ) clear1  
Witness T11n0310\_30 ( ) clear2  
Provisional Subsequent §  
Score N/A  
Submit

Alignments

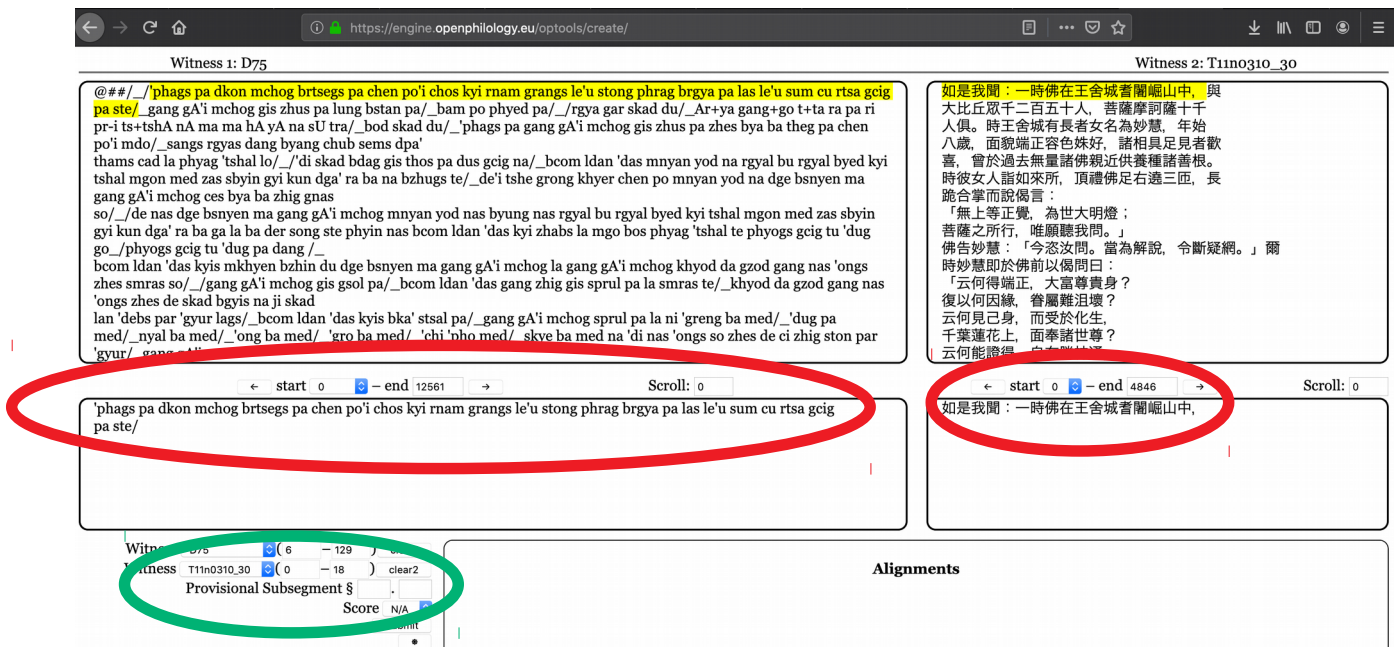
## Step 2 - Highlight the text to be aligned within each witness by selecting it

A passage in one witness with no corresponding passage in the other witness (e.g., a phrase in a Tibetan witness does not appear in the equivalent Chinese witness) should be left blank where it does not appear. There are thus three types of allowed alignments that can be submitted to the database:

- text selected on left, text selected on right
- text selected on left, blank on right OR



- text selected on right, blank on left



Step 3 - Check if the correct parts are selected in the selection windows. If your selection is not correct, simply go back to Step 2 and make a new selection. Your old selection highlighted in yellow will disappear once you click again in the text field of the witness. Your new selection will be highlighted instead and it will be displayed in the box with the red circle so you can check again if necessary.

## Adding provisional subsegment labels (optional).

If you want to add a provisional subsegment label, you can enter a number + letter in the boxes circled in green. Adding provisional subsegment labels is entirely optional, but it may benefit philologists in the longer run when they start editing their alignments in the philological editing environment (currently being developed by Bruno Laine).

## Step 4 - Score the alignment with the dropdown list of scores

The screenshot shows the OpenPhilology engine interface. Two witness texts are displayed side-by-side for alignment. The left witness (D75) is in Tibetan, and the right witness (T11n0310\_30) is in Chinese. A dropdown menu is open over the alignment area, showing a score of '3' selected. The interface includes scroll bars and witness identifiers.

When a Tibetan or Chinese clause/sentence does not correspond to anything, you can keep one of the languages blank. In the selection check box you will see the message 'THIS SPACE INTENTIONALLY LEFT BLANK'. These alignments with one of the two boxes blank cannot be scored and the score is just left as 'not applicable' (n/a). Note that when you are scoring alignments that are manually inserting, your score can never be '1' (see guidelines on scoring in section §4 below). Scores can be added at the time of the alignment submission, or at a later time (see below).

## Step 5 - Submit the alignment to the database with the submit button

The screenshot shows the OpenPhilology engine interface. The alignment is now submitted, and the 'Submit' button is highlighted with a red circle. The interface includes scroll bars and witness identifiers.

Check if the selections, provisional subsegments (optional) and score (optional – can be done later, see below) are correct and then submit the alignment & score to the database clicking the 'Submit' button.

Scores can be added at the time of the alignment submission, or at a later time. Alignments already in the system for any particular choice of witness pairs are visible in a list at the lower right portion of the screen and highlighted in green in the text boxes.

When clicking on an alignment in the bottom right 'Alignment box', that alignment will be highlighted in pink and displayed in the check boxes. The score for each alignment can be added/modified or deleted using the 'modify' and 'delete' buttons on its item in the list indicated with the red circle.

## §4 Scoring Alignments

Our alignment scoring system adapts best practices from Machine Translator Evaluation in which human annotators evaluate automatically translated sentences produced by a machine. Since ‘Fluency’ is not a relevant parameter when it comes to judging alignments, we focus on the ‘Adequacy’ parameter only. ‘Adequacy’ of a translation, or in our case ‘alignment’, refers to how well content is preserved between two passages. The following brief descriptions for each of the scores and included examples are intended to help guide philologists in scoring their manually aligned witnesses. When in doubt between two scores, use the higher score. Brief descriptions of each score:

- 1 no match
- 2 content is somewhat similar, major differences in word choice
- 3 content is the same, significant differences in word choice
- 4 content is the same, one or two differences in word choice
- 5 meaning is exactly the same with no differences

Cases for which there is not a Tibetan/Chinese equivalent to a sentence/paragraph of a witness should be indicated with the “intentionally left blank” selection in the alignment tool (the default selection before text is highlighted). In such cases, no score is needed and a score of ‘not applicable’ (N/A) should be selected from the drop-down list in the alignment input tool. A complete manual alignment of two witnesses should therefore account for every phrase in both witnesses (there should be no cases that have no match between the Tibetan and Chinese). It is however possible for the computer to select two sentences that have no matching content, so we reserve the score of ‘1’ for these cases. Human users of the software should not ever select a score of ‘1’ for any purpose.

### Score 1 – No match

This score can only apply to alignments produced automatically by the machine: there is absolutely no match in content or wording between the Tibetan and Chinese. Note that this score concerns only the content of the core proposition. For example, if the passage in question is a dialogue between two characters, and the only common feature shared by the Tibetan and Chinese phrases is “X said/asked”, while the internal content of the question/statement remains markedly different, the alignment can still be scored as 1.

- |   |  |  |
|---|--|--|
| 1 | Tibetan  | Chinese  |
|   | de nas bcom ldan 'das la tshe dang ldan pa kun dga'<br>bos 'di skad ces gsol to ll bcom ldan 'das chos kyi<br>rnam grangs 'di'i ming ci lags l 'di ji ltar bzung bar bgyi<br>l | 是人已曾無量佛所親近、供<br>養，種諸善根。」   |
|   | Then the Venerable Ānanda asked the Blessed One:<br>“Blessed One, what is the name of this exposition of<br>the teaching? How should we hold it in mind?”                      | “This person previously brought<br>herself close to innumerable<br>buddhas, worshipped them and<br>planted roots of virtue.” |
| 2 | Tibetan  | Chinese  |
|   | des pas bltas nas smras pa/  | 爾時菩薩以無染眼觀彼諸<br>女，告之言曰：   |

## Score 2 - content is somewhat similar, major differences in word choice

Score 2 is distinguishable from score 1 because examples with score 2 have at least some overlap in content. For score 2, however, there can still be major differences in word choice.

3	Tibetan	Chinese
	'di ltar grong khyer mnyan yod na l dge bsnyen ma 'di lta bu gnas la l de de bzhin gshegs pa dang lhan cig smra yang l de'i lus ngal bar yang mi 'gyur ba dang l	「此優婆夷甚為希有，能 與如來共相訓對，得無所 畏。
	“In the town of Śrāvastī lives such a female lay disciple who, even though she spoke together with the Tathāgata, did not experience physical exhaustion.*	“This female lay disciple is exceedingly wonderful; she is able to engage in question and answer together with the Tathāgata, without being afraid.*
4	Tibetan	Chinese
	/mthong nas 'di snyam du sems te/ ci grong khyer gyi dbul po 'di yi dam la brtan nam mi brtan bdag gis brtag go snyam mo/	便自念言：「今此善順於 諸梵行曾不懈息，將不為 求帝釋處耶？或貪王位及 欲樂耶？」
5	Tibetan	Chinese
	/de nas lha'i dbang po brgya byin gyis dus gzhan zhig na mi gzhan zhig sprul to/	爾時天帝復自親試持俱胝 金，
6	Tibetan	Chinese
	/smras pa/ de bas na gnas brtan nyid la yongs su dris khyod kyis ji ltar bzod pa bzhin du lan thob shig	文殊師利言：「須菩提！ 於意云何？若有醫人將護 病者，不與辛酸苦澁等 藥。而彼醫人於彼病者， 為與其差？為與死耶？」



### Score 3 - content is the same, significant differences in word choice

Examples with score 3 have the same content, but there can be significant differences in word choice or in the way the content is presented (i.e. in a different order).

7	Tibetan	Chinese
	<p>de nas bcom ldan 'das kyi zhal gyi sgo nas 'od zer kha dog sna tshogs kha dog du ma 'di lta ste l sngon po dang l ser po dang l dmar po dang l dkar po dang l btsod ka dang l shel dang l dngul gyi kha dog lta bu dag byung nas l de dag gis 'jig rten gyi khams mtha' yas mu med pa dag snang bar byas te l 'od kyis khyab par byas nas l tshangs pa'i 'jig rten gyi bar du mngon bar byung ste l slar log nas bcom ldan 'das kyi zhal gyi sgor nub bo ll</p> <p>Many variously colored rays of light—such as blue, yellow, red, white, vermillion, crystal and silver*—came out of the Blessed One's mouth, and illuminated the endless, boundless worlds of the universe. The light penetrated the worlds, ascended to the world of Brahmā, and returned to disappear back into the Blessed One's mouth.</p>	<p>從其面門放種種光，青、黃、赤、白、紅、頗梨色。其光普照無量國界，上至梵世，還從如來頂上而入。</p> <p>All kinds of rays of light radiated from his mouth—blue, yellow, red, white, purple and crystal colored.* Those rays of light shone throughout immeasurable realms, ascending to the world of Brahmā, returning and entering from the crown of the Tathāgata's head.</p>
8	Tibetan	Chinese
	<p>/rgyal pos smras pa/ bu mo de dag ni nga dang mtshungs pa ma yin no/</p>	<p>王語女言：「彼非我類，我云何迎？」</p>
9	Tibetan	Chinese
	<p>bcom ldan 'das 'di ltar nyon mongs pa rnam kyi rang bzhin gang lags pa de nyid sangs rgyas kyi yul gyi rang bzhin lags te/ de'i slad du de bzhin gshogs pas chos thams cad mnyam pa nyid du mngon par dzogs par sangs rgyas so zhes ston lags so/</p>	<p>以煩惱性即佛界性故，說如來住平等性。」</p>

## Score 4 - content is the same, one or two differences in word choice

Examples with score 4 also have the same content, but they can be differentiated from those with score 3 because only one or two words are different. It is expected that many of the manually aligned sentences/paragraphs will fall into this category.

10	Tibetan	Chinese
	de'i tshe grong khyer chen po mnyan yod na l dge bsnyen ma gang ga'i mchog ces bya ba zhig gnas so ll	時舍衛城有優婆夷, 名恒 河上。
	At that time, in the great town of Śrāvastī there lived a female lay disciple named Gaṅgottarā.	At that time in Śrāvastī there was a female lay disciple called Gaṅgottarā.
11	Tibetan	Chinese
	'di skad bdag gis thos pa dus gcig na l bcom ldan 'das yul barga na chu srin byis pa gsoḍ lta bu'i ri 'jigs su rung ba'i nags ri dags kyi nags na	聞如是：一時佛遊於披 祇國妙華山中恐懼樹間鹿 所聚處，
	Thus I heard at one time: the Blessed One dwelt in the state of Bhaggas, in the mountain that looked like a child-killing water-monster ( <i>susumāragiri</i> ), in the Bhesakaḷā Forest, at the Deer's Grove,	Thus I heard: at one time, the Buddha travelled in the state of Baqi, in the Marvelous Flower Mountain, in the Bhesakaḷā Grove, at the Deer's Haunts,
12	Tibetan	Chinese
	gzhan yang de bzhin gshegs pa zhal phyir phyogs pa mthong ba dang/	復次夢見如來背坐；
		Furthermore, he sees the tathāgata facing backwards;
13	Tibetan	Chinese
	bcom ldan 'das kyis bka' stsal pa l gang ga'i mchog sprul pa ni ngan song du mi 'gro l mtho ris su mi 'gro l yongs su mya ngan las 'das par mi 'gro na l ci gang ga'i mchog kyang ngan song du mi 'gro l mtho ris su mi 'gro l yongs su mya ngan las 'das par mi 'gro 'am l	世尊告曰：「是幻化人不 往惡趣，不生天上，不證 涅槃。恒河上，汝亦爾 耶。」
	The Blessed One said: “Gaṅgottarā, if an artificial creation does not go to evil destinies, does not go to heaven, does not go to parinirvāṇa, do not you, Gaṅgottarā, also not go to evil destinies, not go to heaven, not go to parinirvāṇa?”	The Blessed One said: “This illusorily artificial person does not go to evil destinies, is not born in heaven, does not attain nirvāṇa. Gaṅgottarā, are you also like this?”

14	Tibetan  /bcom ldan 'das yul khyad par ma mchis pa ni sang rgyas kyi yul lags te/	Chinese  世尊！非境界是佛境界。
15	Tibetan  bcom ldan 'das kyis bka' stsal pa ll cī khyod mya ngan las 'da's pa'i dbyings la gzhol bar myi byed dam ll  The Blessed One said: "Are you not applying yourself toward the realm of nirvāṇa?"	Chinese  佛言：「汝豈不趣涅槃界 耶。」  The Buddha said: "Is it not, certainly, the case that you are headed toward the realm of nirvāṇa?"
16	Tibetan  sngon ma mthong ba'i dge slong chos smra ba dag mthong ba dang/	Chinese  又見未曾見法師；
17	Tibetan  maud gal gyi bus smras pa/ bu mo de ni mi shes so/	Chinese  目連答言：「不能知也。」

## Score 5 - meaning is exactly the same with no differences

Score 5 should be given to perfect alignments, i.e. passages that match completely as in the case of an ideal translation, accounting for every single word with the same style and content. In theory this could also apply to texts that are not translated from each other directly, though it is likely to occur less often in those.

18	Tibetan	Chinese
	gang ga'i mchog gi's gsol pa' l gal te de ltar na l ci ltar byang cub kyī ched du dge ba'i rtsa ba stsogs par bgyid lags ll	恒河上言：「若如是者，云 何爲菩提故，積集善根。」
	Gaṅgottarā asked: "If that is the way it is, how did you collect roots of virtue for the sake of awakening?"	Gaṅgottarā said: "If that is the way it is, how did you collect roots of virtue for the sake of awakening?"
19	Tibetan	Chinese
	tha na dgra la yang khro bar mi byed pa dang /	一者，於惡友所不起瞋 心；
20	Tibetan	Chinese
	bcom ldan 'das gang zhig gis sprul pa la smras te l khyod da gzod gang nas 'ongs shes de skad bgyis na l ji skad lan 'debs par 'gyur lags l	「世尊，若問化人『汝從何 來』，如是問者，當云何 答。」
	"Blessed One, if someone were to say to an artificial creation,* 'Where have you come from just now?' how should he answer?"	"Blessed One, if one were to ask an artificially created person,* 'Where have you come from?' how should he answer such a question?"*

## §5 CSV upload

Alignments and scores can also be added to the database by uploading a CSV file ('Option 2'). This sections explains how this is done.

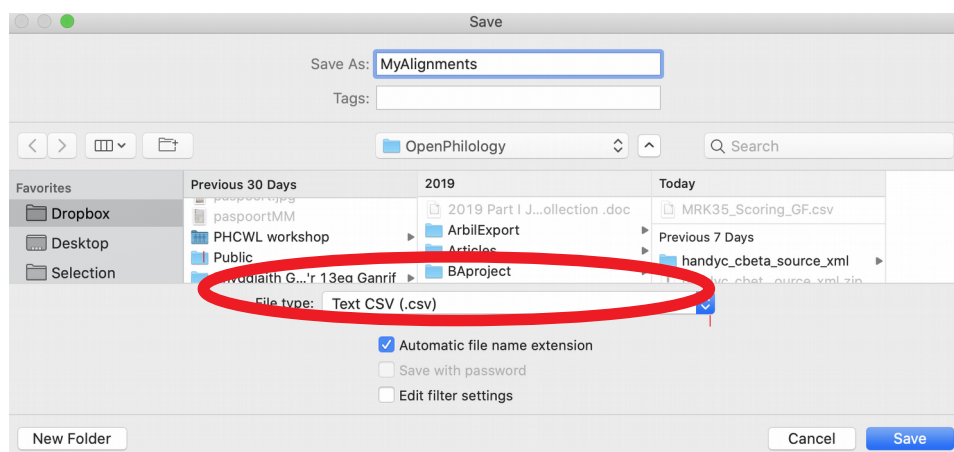
### Step 1 – Prepare a CSV file with alignments (and optional scores)

Use any software application you prefer (e.g. MS Excel, LibreOffice/OpenOffice Calc, Mac Numbers, etc.) to create a CSV file with the following specifications:

- UTF-8 encoding
- Text from Witness 1 in column A & Text from Witness 2 in column B (note that either A or B could be blank if there is not Tibetan/Chinese corresponding to a sentence)
- Score (n/a, 1, 2, 3, 4, 5) in column C (optional)
- Subsegment label, e.g. '1.a' in column D (optional)
- Other notes (e.g. 'not found in Chinese') in column E (optional)

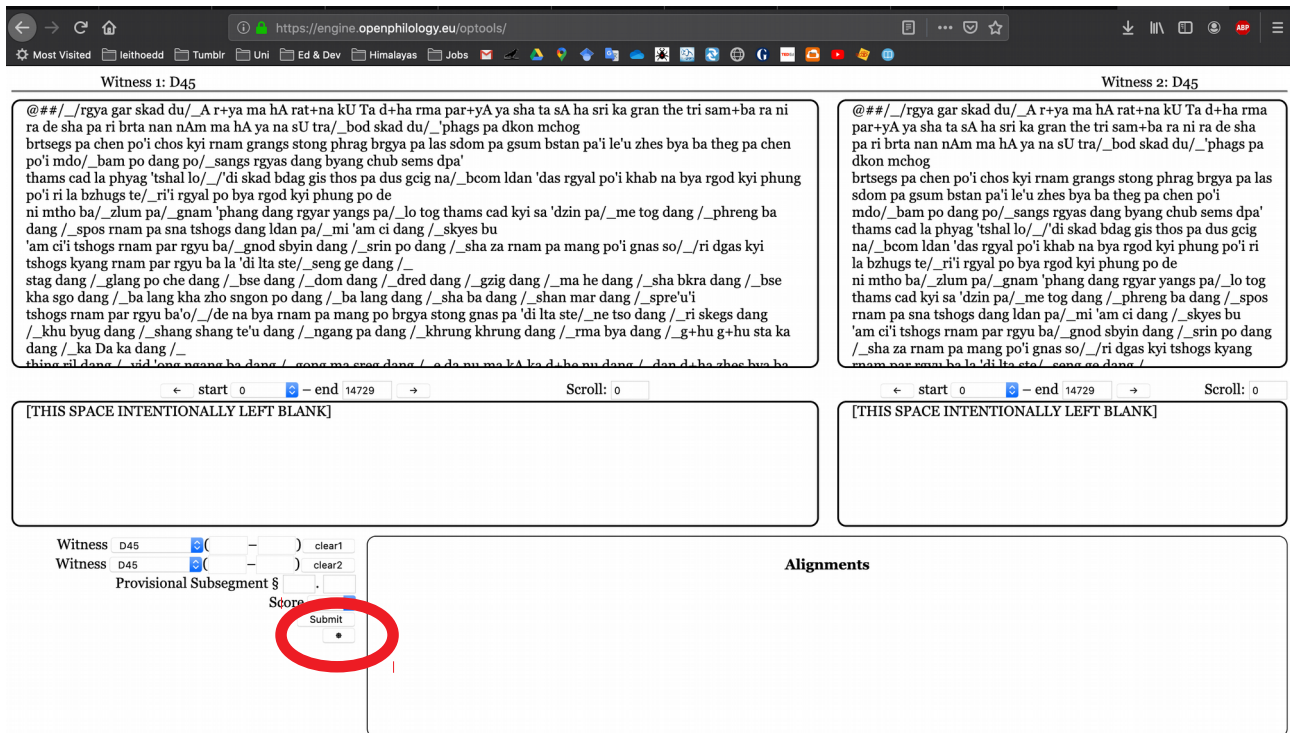
	A	B	C	D
1	//phags pa dkon mchog brtsegs pa chen po'i chos	大寶積經善德天子會第三十五 大唐三藏菩提流志奉詔譯		1.1.a
2	/di skad bdag gis thos pa dus gcig na/ bcom ldan	如是我聞： 一時佛在舍衛國祇樹給孤獨園，與大比丘眾		2.1.b
3	/yang de'i tshes 'jam dpal gzhon nur gyur pa 'khor d	是時文殊師利菩薩摩訶薩與善德天子俱在會中。		3.2.a
4	/de nas bcom ldan 'das kyis 'jam dpal gzhon nur gy	爾時世尊告文殊師利：「汝當為此諸天大眾及諸菩薩，演說		3.3.a
5	/gsol pa/ bcom ldan 'das sangs rgyas kyi yul ni gar	文殊師利白佛言：「唯然世尊！若善男子善女人欲知佛境界		3.3.b
6	/bcom ldan 'das yul khyad par ma mchis pa ni san	世尊！非境界是佛境界。		3.4.a
7	/bcom ldan 'das de bzhin gshegs pas byang chub	以是義故，如佛所得阿耨多羅三藐三菩提，為何境界耶？」		2.4.b
8	bka' stsal pa/ 'jam dpal lta bar gyur pa mnyam pa r	佛言：「空境界，諸見平等故。無相境界，一切相平等故。」		4.4.c
9	/gsol pa/ bcom ldan 'das 'dus ma bgyis kyi yul gan	文殊師利言：「世尊！何等是無為境界？」		5.4.d
10	bka' stsal pa/ 'jam dpal sems med pa ni 'dus ma by	佛言：「無念是無為境界。」		5.4.e
11	/gsol pa/ bcom ldan 'das sems ma mchis pa ni san	文殊師利言：「世尊！若無為等是佛境界，為無念者依何而		3.5.a
12	/bka' stsal pa/ 'jam dpal sangs rgyas kyi yul gang	佛言：「文殊師利！佛境界當於何求？」		5.6.a
13	gsol pa bcom ldan 'das sangs rgyas kyi yul ni sem	曰：「於一切眾生煩惱中求。何以故？眾生煩惱性不可得，		3.6.b
14	/bka' stsal pa/ 'jam dpal sangs rgyas kyi yul la 'ju	佛言：「文殊師利！佛境界有增減耶？」		1.6.c
15	gsol pa/ bcom ldan 'das sangs rgyas kyi yul ni 'jug	曰：「無增減也。」		1.7.a
16	/bka' stsal pa/ 'jam dpal gal te sangs rgyas kyi yu	佛言：「云何了知一切眾生煩惱本性？」		1.7.b
17	gsol pa/ bcom ldan 'das ji ltar sangs rgyas kyi yul 'j	曰：「如佛境界無有增減，煩惱本性亦無增減。」		1.8.a
18	/bka' stsal pa/ 'jam dpal nyon mongs pa rnams kyi	佛言：「云何名為煩惱本性？」		4.9.a
19	gsol pa/ bcom ldan 'das sangs rgyas kyi yul gyi rar	曰：「煩惱本性是佛界本性。」		3.10.a
20	bcom ldan 'das gal te nyon mongs pa rnams kyi ra	世尊！若煩惱性異佛境界，則不說佛住一切法平等性中。		2.11.a
21	bcom ldan 'das 'di ltar nyon mongs pa rnams kyi ra	以煩惱性即佛界性故，說如來住平等性。」		2.11.b

When using spreadsheet software like Excel, Calc, Numbers etc. make sure you save the file as .csv (this will automatically render the right format for uploading).

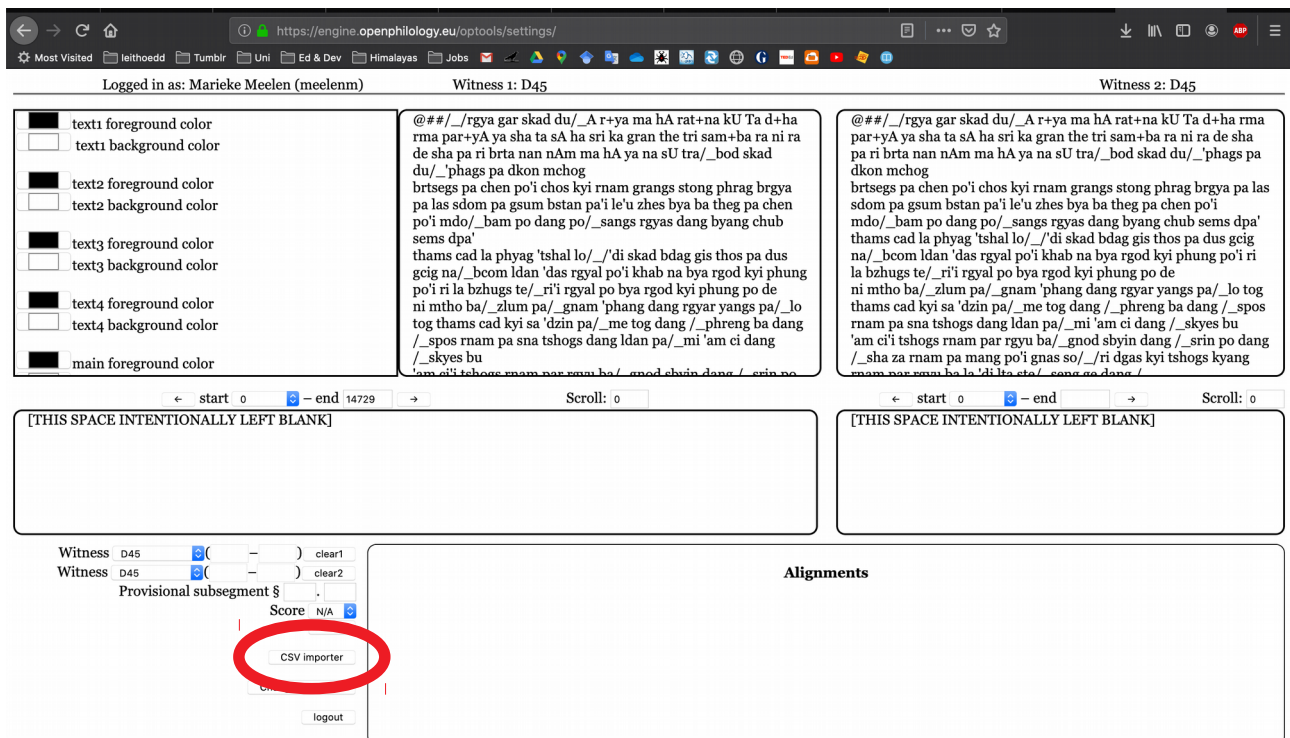


## Step 2 – Navigate to the CSV upload function in the tool

Click on the settings/dharma wheel in the home screen of the Alignment Input & Scoring Tool.



This will make the 'CSV importer' button appear – click on that to access the part of the tool that allows for CSV uploads.



### Step 3 – Select Witnesses and Browse to upload your CSV file

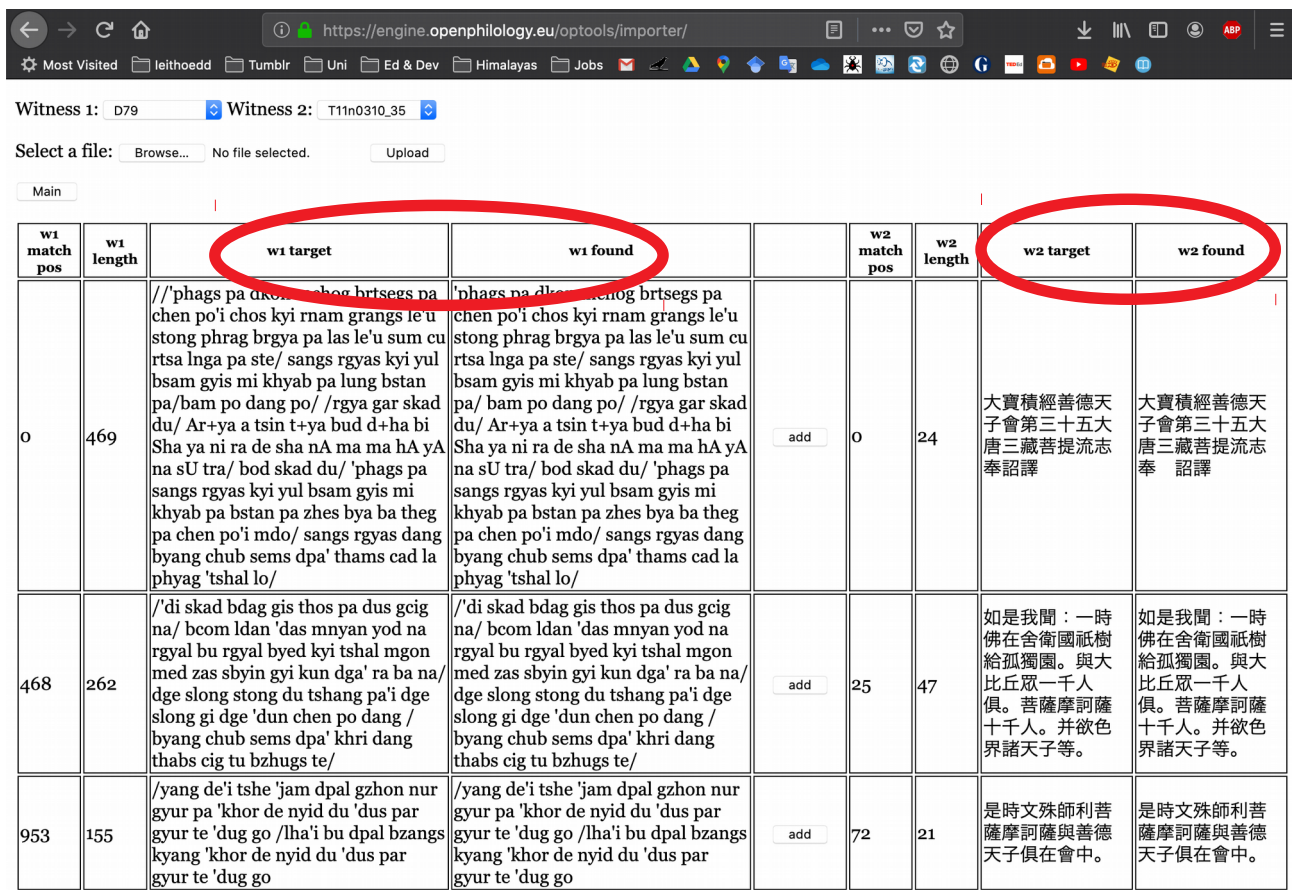
Choose your Witness 1 corresponding to your Column A in the CSV file from the dropdown list and Witness 2 for Column B. Then upload your CSV file choosing the file from your computer with the Browse button.



Please note that uploading the file may take a while (depending on the size of your file) and the cursor does not change shape to a 'wait cursor' in the process.

### Step 4 – Check and add alignments and scores to the database

Once the CSV file is uploaded, you will see a table with 'Witness 1 target', 'Witness 2 found', a score and an 'Add' button and the same for Witness 2 (ignore the other columns).



The column 'w1 target' shows the input from your CSV file and the 'w2 found' shows the whichever part of that witness in database forms the closest match with what you have in your CSV file. The same goes for the 'target' and 'found' columns of the second witness.

Check if the matches in the ‘found’ columns are close to what you intended. They don’t have to match perfectly: sometimes there may be some slight differences in punctuation for instance or a word may be added or missing. Even if some punctuation or a word or two differs, this will still be useful training data. So click the ‘Add’ button to add the alignment and score to the database.

w1 match pos	w1 length	w1 target	w1 found		w2 match pos	w2 length	w2 target	w2 found
0	469	//'phags pa dkon mchog brtsegs pa chen po'i chos kyi rnam grangs le'u stong phrag brgya pa las le'u sum cu rtsa lnga pa ste/ sangs rgyas kyi yul bsam gyis mi khyab pa lung bstan pa/bam po dang po/ /rgya gar skad du/ Ar+ya a tsin t+ya bud d+ha bi Sha ya ni ra de sha nA ma ma hA yA na sU tra/ bod skad du/ 'phags pa sangs rgyas kyi yul bsam gyis mi khyab pa bstan pa zhes bya ba theg pa chen po'i mdo/ sangs rgyas dang byang chub sems dpa' thams cad la phyag 'tshal lo/	'phags pa dkon mchog brtsegs pa chen po'i chos kyi rnam grangs le'u stong phrag brgya pa las le'u sum cu rtsa lnga pa ste/ sangs rgyas kyi yul bsam gyis mi khyab pa lung bstan pa/ bam po dang po/ /rgya gar skad du/ Ar+ya a tsin t+ya bud d+ha bi Sha ya ni ra de sha nA ma ma hA yA na sU tra/ bod skad du/ 'phags pa sangs rgyas kyi yul bsam gyis mi khyab pa bstan pa zhes bya ba theg pa chen po'i mdo/ sangs rgyas dang byang chub sems dpa' thams cad la phyag 'tshal lo/	add	0	24	大寶積經善德天子會第三十五大唐三藏菩提流志奉詔譯	大寶積經善德天子會第三十五大唐三藏菩提流志奉詔譯
468	262	//'di skad bdag gis thos pa dus geig na/ bcom ldan 'das mnyan yod na rgyal bu rgyal byed kyi tshal mgon med zas sbyin gyi kun dga' ra ba na/ dge slong stong du tshang pa'i dge	//'di skad bdag gis thos pa dus geig na/ bcom ldan 'das mnyan yod na rgyal bu rgyal byed kyi tshal mgon med zas sbyin gyi kun dga' ra ba na/ dge slong stong du tshang pa'i dge	add	25	47	如是我聞：一時佛在舍衛國祇樹給孤獨園。與大比丘眾一千人俱。菩薩摩訶薩	如是我聞：一時佛在舍衛國祇樹給孤獨園。與大比丘眾一千人俱。菩薩摩訶薩

Once you click ‘Add’, the button will turn blue to indicate this alignment & score has been added to the database so you can keep track of your work.

If there are blank cells in one of the witnesses, still add the alignment (especially if you have added subsegment labels, this may be useful for you when you are editing your texts in the longer run.

8201	0			add	1770	53	於是首耶阿須倫女日行王女識乾執樂第一夫人。於冥夜靜時。於須賴所止地之分界。行詣於彼。說溫暖甜辭。與須賴言：	於是首耶阿須倫女日行王女識乾執樂第一夫人。於冥夜靜時。於須賴所止地之分界。行詣於彼。說溫暖甜辭。與須賴言：
9375	15	des pa bzhengs/	des pa bzhengs/	add	1832	6	「起。仁者！」	「起。仁者！」
9391	57	bdag cag gnyis ni khyod kyi bsnyen bkur bgyid pa lags so/	bdag cag gnyis ni khyod kyi bsnyen bkur bgyid pa lags so/	add	1838	7	我等故來相事。	我等故來相事。
9450	379	des pa bdag cag gnyis kyi gzugs phun sum tshogs pa la gzigs/ sha tshugs dang / kha dog dang / dbyibs dang / lus byugs pa dang / rgyan dang / chas dang / lhab lhub dag la gzigs/ na tshod dang po'i lang tsho la bab pa la gzigs/ des pa khyod kyi bsnyen bkur bgyid pa 'di 'dra ba bdag nyid kyi bsod nams phun sum tshogs pa la gzigs/ na tshod dang po'i lang tsho la bab pa la gzigs/	des pa bdag cag gnyis kyi gzugs phun sum tshogs pa la gzigs sha tshugs dang / kha dog dang / dbyibs dang / lus byugs pa dang / rgyan dang / chas dang / lhab lhub dag la gzigs/ na tshod dang po'i lang tsho la bab pa la gzigs/	add	1846	31	觀。須賴！我等形容之嚴好。塗栴檀香瓔珞被服鮮明適在盛時視此。	觀。須賴！我等形容之嚴好。塗栴檀香瓔珞被服鮮明適在盛時視此。
9675	102	des pa khyod kyi bsnyen bkur bgyid pa 'di 'dra ba bdag nyid kyi bsod nams phun sum tshogs pa la gzigs/	des pa khyod kyi bsnyen bkur bgyid pa 'di 'dra ba bdag nyid kyi bsod nams phun sum tshogs pa la gzigs/	add	1877	15	須賴！以汝之福故。得我等執事。	須賴！以汝之福故。得我等執事。

If an entire clause is missing in the ‘found’ column, compare both ‘found’ columns (indicate with green circles) and decide whether these could also work as alignments (even though they differ from what you had in your CSV). If they are fine, add them too. If not, skip them.

After checking and adding the first ten alignments, click on the ‘next’ button at the bottom of the page to check and add the following ten alignments of your text.

Once you’ve checked/add all alignments you’re done! If you want to edit modify alignments or scores, or add subsegment labels, use the main tool window as described in Section §3.3. Thank you for adding & scoring alignments!