

Chinese Guidelines for BTSR

You will receive an excel file with two columns: 'check' and 'contexts'. In each row of the 'contexts' column, you will find a test example in the following form:

Source word lemma::[position of the source word in source sentence] |&| [the source sentence in which the source word is marked in red] ||| Target word lemma::[position of the target word in test sentence] |&| [the target sentence where the target word is marked in red]

The source and target sentences contain a translation word pair (marked in red). Notice that the sentences are not translations of each other. The word lemma is the canonical form of the word which may be inflected in different ways in the sentence. These inflections (eg. number inflections, case, person) do not change the word sense (word sense is a meaning or usage of a word). Since a word can be ambiguous, your job is to check if the two words in the source and target contexts are expressing the same word sense, and if so, please leave the corresponding row empty in the 'check' column.

What counts as correct?

(1). Some translation pairs are more 'stable' in terms of meanings, in the case that they will always be translated the same in different contexts. For example:

數字::14 |&| 一種 (電子) 器件 , 其 輸出 數據 是 表示 輸入 數據 的 數字 之 和 。
||| number::8 |&| The second problem is , what does the **number** of one in 73 million mean ?

(2). Some translation pairs are more ambiguous. You will need contexts to determine whether the two words in context are of the equivalent word sense. For example:

欺騙::4 |&| 我 不 喜歡 被 欺騙 ||| lie::9 |&| You lied to Big Baby and you 've been **lying** ever since .

What counts as wrong?

(1). The two words express different word senses in the given contexts (This is the most important error that I would like to you pick up). It could be that the two words never express the same word sense in any context. Or it could be that the two words could be translations of each other, but not in the given contexts. Please annotate 'Wrong'

eg.

沮喪::5 |&| 爲什麼 他們 總讓 我 很 沮喪 ? ||| down::12 |&| Mama say that happiness is from magic rays of sunshine that come **down** when you feelin ' blue .

(2). Sometimes the two words' word senses seem to be related. A general rule of thumb is that if they express abstract vs concrete distinction, or if they are of different parts of speech (parts of speech are syntactic categories of a word. The main parts of speech are nouns, verbs, adjectives

and adverbs), then the pair should be marked wrong. If in doubt, please put a question mark. The following is an example where the two words in contexts are related but not in the same word sense.

降臨::2 |&| 夜幕 已經 降臨 ||| fall::10 |&| it remains only to cherish , ere yet the blow fall , the lofty thoughts that ennoble his little day ;

(3). If you happen to spot any ungrammatical or strange sentences, mark these as 'en error' if it's the English sentence that is strange or 'ch error' if it's the Chinese sentence.