

SUPPLEMENT

X-SRL: A Parallel Cross-Lingual Semantic Role Labeling Dataset

A.1 Annotation Guidelines for Translators

In this annotation task, you will be provided with pairs of sentences taken from the business section of the Wall Street Journal newspaper. Each pair consists of the original English source and its corresponding translation in a given target language (German, Spanish or French). Every target sentence was produced via automatic translation software. We are interested in three main aspects of the translations:

A.1.1 Validation of Automatically Translated Text - Overall Quality

We are interested in how much the automatic translation preserves the general meaning that the source intended to communicate. Considering the target sentence as a whole, mark on a scale from 1-5 (worst to best) how well the target translation captures the original meaning:

Quality Mark on a scale from 1-5 how grammatical the sentence is:

- 1 = Ungrammatical or meaningless.
- 2 = Grammatical but does not preserve the source meaning.
- 3 = The translation has most of the source meaning but has key errors.
- 4 = This is a translation with small distortions.
- 5 = This translation could have been done by a (non-professional) human.

A.1.2 Key Predicates

We are also interested in the preservation of source events (verbs) in the target sentence. We are aware that sometimes the best translation of a sentence does not contain a direct translation of each verb. For our specific purposes, however, we would like to have sentence pairs that contain them in both the source and the target. We will provide a list of the main verbs of the source and you should indicate if

there is a one-to-one corresponding verb on the target side. We only aim to keep verbal predicates, therefore target corresponding nouns should be ignored even if they are a correct translation.

A.1.2 Key Arguments

We also want to measure how many keywords (arguments) from the source are still preserved in the target sentence. We will provide a list of such words taken from the source and you should indicate if there is a corresponding word on the target side. Note that in this case, rather than a one-to-one correspondence, we want to find semantically related sub-phrases inside the sentence (please see the “Difficult Examples” and “Important notes” sections for more details). We provide you with:

A.1.3 Annotation Elements

1. **Source sentence:** the original English sentence. Example:

```
Heavy selling of Standard & Poor's 500-stock index futures
in Chicago relentlessly beat stocks downward.
```

2. **Target sentence:** the automatic translation of the source (in English, German or French). Example:

```
Starke Verkäufe von Standard & Poor's 500-Aktienindex-Futures
in Chicago schlagen Aktien unerbittlich nach unten.
```

3. **Indexed Source:** the same English source but indicating, on the left side of each word, in which position it is located inside the source sentence. This will help you to identify the specific words we are interested in. Example:

```
1_Heavy 2_selling 3_of 4_Standard 5_ 6_Poor 7_'s 8_500
9_- 10_stock 11_index 12_futures 13_in 14_Chicago 15_religiously
16_beat 17_stocks 18_downward 19_.
```

4. **Indexed Target:** the same translated sentence but indicating, on the left side of each word, in which position it is located inside the target sentence. Example:

```
1_Starke 2_Verkäufe 3_von 4_Standard 5_ 6_Poor 7_'s 8_500
9_- 10_Aktienindex-Futures 11_in 12_Chicago 13_schlagen
14_Aktien 15_unerbittlich 16_nach 17_unten 18_.
```

5. **Source Predicates:** the relevant events (verbs) that occur in the source sentence. Example:

```
16_beat
```

6. **Source Arguments (keywords):** these are words that bear most of the semantic meaning of the source, therefore we would like to find them as well in the target. Example:

2_selling
15_relentlessly
17_stocks
18_downward

7. **Propose an Alternative Translation:** if you think that the whole translation should be rephrased, please provide it in this space.

A.1.4 Difficult Examples

1. When an Argument (keyword) is a **preposition or a syntactic marker**, you should read the complete sub-phrase that is associated to it and try to find the translated sub-phrase on the target side: we are interested in matching the same semantic meaning, even if the keywords do not correspond. Example:
As the market plunged 90 points, it barely managed **to** stay this side of chaos.
Als der Markt 90 Punkte einbrach, gelang es ihr kaum , **auf** dieser Seite des Chaos zu bleiben.
11_to → 13_auf
2. When in doubt, it is preferable that you fill-in more than one word in the EQUIVALENT IN TARGET column, and mark it as a **special case**. The special cases that you find will be analyzed on a further steps by other annotators. For example:
18_downward → 16_nach 17_unten
15_surrendered → 2_gaben 21_auf
20_unable → 27_nicht 28_in 29_der 30_Lage
3. Sometimes the English side has a preposition but the target refers directly to a main noun, therefore you should pair it with the noun. Example:
Martha **gave** a book *to Jonas* .
Martha **gab** *Jonas* ein Buch .
5_to → 3_Jonas
4. Sometimes the proper names in the text are not marked properly (even in the source). Please provide always the complete proper name (even when they are many words). Example, for “The New York Stock Exchange” you should answer like the following:
7_Exchange → 4_New 5_Yorker 6_Börse

A.1.5 Important Notes

1. Please remember to always put the answers in the form of `index_word` (i.e. `16_beat` instead of just `beat`).
2. If the translation quality is 2 or 1 then you can ignore the predicates and arguments and you don't have to do the translation fix (such examples will be dropped).
3. If there is no equivalent predicate in the target, you should indicate this by putting the token `<NONE>` on the target predicate cell
4. Avoid spending too much time thinking about "the best possible translation", if you think the sentence should be improved but you can't think of it right away, please fill-in the token `<FIX-LATER>` and skip it (for now). Later go back and improve it if you find the time.
5. If you find concepts whose translation you don't know how to handle, add the Sentence-ID to your log so we can comment on it later.
6. If you found that a predicate or an argument could be present by fixing a mistaken word (just a one-word fix), please fill-in `<FIX-ARG>` and on the immediate cell to the right write the word that you consider most appropriate.
7. Please ALWAYS keep the original word indices, even after you provided different translations for the sentences or you fix word-segmentation issues (e.g. two words have a single index). Mark this cases on the log and they will be revisited later to be fixed.