ParTUT:
Manual Word Alignment Guidelines

Last update:
March, 2016
Chapter 1

General Principles

The task of the annotator is to identify correspondences in the source and target sentences. The annotator is asked to select the minimal translation unit in the source text that has an equivalent in the target text.

In line with other annotation guidelines, the general principle to select this translation unit is:

- select as many words as necessary in the source and in the target sentence to ensure a two-way equivalence
- select as few words as possible in the source and in the target sentence, while preserving two-way equivalence

**Absence of correspondence:** If a word or group of words does not have any corresponding translation in the other language, it must remain unlinked.

1.1 Phrasal correspondences

Multiple alignments are allowed, both one-to-many (or many-to-one) and many-to-many. Such alignments are used to properly handle, for example, cases of idiomatic or fixed expressions that should be linked as a whole. In particular, such links are allowed if both conditions (see below) hold:

- an exact correspondence can be identified, and such correspondence is admitted even out that specific context:

```
    media
   /   |
mezzi di comunicazione
```

- the correspondence identified is admitted in that specific context only:
Chapter 2

Detailed guidelines

Alignment links are also allowed in specific cases that are listed below:

- **Category shifts:**
  - noun $\leftrightarrow$ adjective (e.g. *Community* $\leftrightarrow$ *comunitaria*):
  - adjverb $\leftrightarrow$ adjective (e.g. *substantially* $\leftrightarrow$ *sostanziale*):
  - verb $\leftrightarrow$ noun (e.g. *contributing* $\leftrightarrow$ *contributo*)

  **NOTE:** In case of nominalization, as the transformation from verb into its deverbal noun entails the introduction of a preposition (*of* in English, *de* in French, *di* in Italian), both the deverbal noun and the preposition that introduces its argument should be aligned to the verb:

  
  
  \[
  \text{creating} \quad \text{di} \quad \text{creazione}
  \]

- **Structural shifts:**
  
  - Word order: pre vs postmodification
    When a nominal premodifier is translated using a prepositional postmodifier, besides the corresponding nouns, the omitted preposition is also linked to noun:

    
    
    \[
    \text{security} \quad \text{di} \quad \text{socpowered}
    \]

  - Passivization/Depassivization
    When a transformation occurs from a passive into an active form (or vice versa), both the corresponding verbs are aligned, together with their arguments (whenever lexically realized):
we allow joint accounts

gli account congiunti sono consentiti

you create the password

la password' creata da te

− Pro-drops
    If the subject pronoun is omitted in the source text, a link should be introduced between the lexically-realized pronoun in the target text and the verb in the source text (or vice versa):

you can
    puoi

In case of compound verbs, the pronoun should be linked to the main verb (not the auxiliary):

you have had
    hai avuto

− Auxiliaries:

* when in the source text a modal auxiliary (e.g. the English do/does) is used in order to form the verb tense and/or mood (e.g. the conditional) and such features are expressed by a single verb in the target text, the auxiliary is aligned to the corresponding verb:

created
    ha creato

* when the auxiliaries in the source and target sentence exactly correspond, they
are aligned (e.g. **ha fatto** ←→ **has done**)
* when the auxiliaries in the source and target sentence do not exactly correspond, but their corresponding verbs do, they are aligned (e.g. **è stato** ←→ **has been**)
* when more auxiliaries are attached to the main verb in one language, and just one in the other, group the auxiliaries and link them with the corresponding auxiliary:

was used

è stato usato

– Infinitive **to**:
The English infinitive marker **to**, together with the infinitive it is attached to, should be connected to the infinitive’s translation:

to reproduce

riprodurre

– Phrasal verbs:
Whenever the preposition or adverb combined with the verb is part of the meaning of the verb itself, such preposition or adverb should be aligned to the corresponding verb in the source text:

log in

accedere

– Locutions and idioms should be linked as a whole:

to bring that home

pour vous faire comprendre

– Extra determiners in source or target language should also be linked:
Semantic shifts:

- Mutation/Paraphrases:
  If a meaning is paraphrased or expressed more explicitly in the source or target sentence, use multiple links to draw the correspondence:

  highlight

  mettere in luce

  **BUT** if a phrase is paraphrased in such a way that some words in both source and target language are not expressed in the other language, do not apply any alignment at all:

  recognition as a person before the law

  reconnaissance de sa personnalité juridique

- Additions/Deletions:
  Words whose meaning is not expressed at all in the other language (either source or target language) should not be aligned:

  the respect of and observance of human rights

  le respect des droits

Other:
- Punctuation:
  - Corresponding punctuation marks are linked
  - If a punctuation mark corresponds to a word or to another type of punctuation mark, it can be linked to that word/punctuation mark anyway (e.g. ; $\leftrightarrow$ . or and $\leftrightarrow$,)
  - If a conjunction is expressed by a comma AND a conjunction in one language, and only by a comma OR a conjunction in the other language, just link the corresponding items (i.e. comma with comma, or conjunction with conjunction) and leave the omitted item unlinked:
    
    , and

    /

    e