

Annotation Guidelines of Translation Techniques for English-Chinese

Yuming Zhai
LIMSI, CNRS
Université Paris-Sud
Université Paris-Saclay
Orsay, France
zhai@limsi.fr

Lufei Liu
Université Paris Diderot
Paris, France
luffyliu1108@gmail.com

Xinyi Zhong
Université Sorbonne Nouvelle
Paris, France
zhongxinyisophie@gmail.com

Contents

1	Introduction	3
1.1	Task description	3
1.2	Translation techniques	3
2	General instructions	3
2.1	How to deal with minor misspelling?	3
2.2	How to deal with Chinese segmentation errors?	3
2.3	Three conditions of annotation	4
2.4	External resources	4
3	Decision helper	4
4	Annotation tool	5
5	Annotation conventions	6
5.1	How to decide the segment boundary?	6
5.2	How to align punctuation?	8
5.3	Mutually exclusive categories	8
5.4	How to annotate unaligned segments?	8
5.5	How to deal with linguistic anaphora?	9
6	Annotation in practice	9
6.1	Annotation of aligned segments	9
6.1.1	Literal	9
6.1.2	Equivalence	11
6.1.3	Transposition	11
6.1.4	Modulation	13
6.1.5	Mod+Trans	14
6.1.6	Particularization	15
6.1.7	Generalization	15
6.1.8	Figurative translation	16
6.1.9	Lexical shift	17
6.1.10	Translation error	17
6.1.11	Uncertain	17
6.2	Annotation of unaligned segments	17
6.2.1	Unaligned - Explication	17
6.2.2	Unaligned - Reduction	18
6.2.3	Unaligned - No Type	19

1 Introduction

1.1 Task description

This document is a guide for annotating translation techniques in a parallel corpus. For example, in this work, our compiled English-Chinese parallel corpus contains 2 200 pairs of sentences, covering eleven different domains: *Art*, *Educational Materials*, *Literature*, *Law*, *Microblog*, *News*, *Official document*, *Spoken*, *Science*, *Subtitle* and *Scientific article*. Each sentence pair contains on average 24 English tokens and 38 Chinese characters. The translation direction is from English to Chinese, except for the domain of *Scientific article*.

The Chinese corpus has been tokenized by the tool *THULAC* (Li and Sun, 2009)¹ and English corpus by *Stanford Tokenizer*². The parallel corpus has been further automatically aligned at word level by the tool *TsinghuaAligner* (Liu and Sun, 2015)³.

1.2 Translation techniques

Consider this example (all examples in this guide are shown with tokenization):

I know what people think when they see this .

→ 我知道人们想什么当他们看到这个的时候。

For this pair of sentences, we could conduct word alignments as follows, where all the source segments have been translated literally.

I → 我

know → 知道

what → 什么

people → 人们

think → 想

when → 当 ... 的时候

they → 他们

see → 看到

this → 这个

. → 。

Translation techniques constitute an important subject of study for translators and linguists (Vinay and Darbelnet, 1958; Newmark, 1981; Newmark, 1988; Chuquet and Paillard, 1989; Molina and Hurtado Albir, 2002; Gibová, 2012), which distinguish literal translation from other translation techniques on word or phrase level. Based on the above cited work in translation techniques, by annotating and analyzing our English-Chinese parallel corpus, we have proposed a typology of translation techniques (see figure 1) in order to have a global view of these categories. The table 1 and 2 provide a recapitulation of definition and important rules for each translation technique.

2 General instructions

2.1 How to deal with minor misspelling?

In order to guarantee the corpus quality and to generate a clean data set for developing our automatic classifier, we correct minor spelling errors in the corpus, for example 作法 → 做法.

During the annotation, please note down the sentence ID and the misspelled pair.

2.2 How to deal with Chinese segmentation errors?

The word segmentation has first been done automatically with the tool *THULAC* (Li and Sun, 2009). The remaining errors should be corrected manually before or during the annotation. If you see segmentation errors, please **skip the sentence first**, and tell us to correct the segmentation. (Because if we align all the words and then correct the segmentation, the alignment indices will be influenced.) For example:

¹<http://thulac.thunlp.org/>

²<https://nlp.stanford.edu/software/tokenizer.html>

³<http://nlp.csai.tsinghua.edu.cn/~ly/systems/TsinghuaAligner/TsinghuaAligner.html>

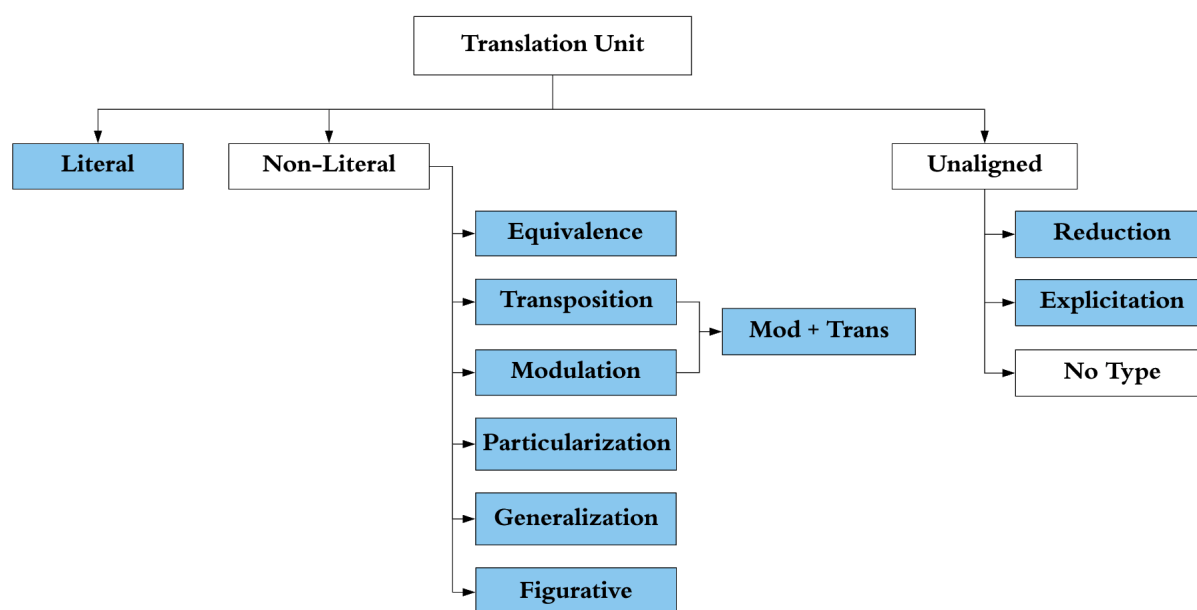


Figure 1: Typology of translation techniques

就是, 只是: should be tokenized into "就 是" (就: emphasise something), "只 是"
 们: it is a suffix and should be tokenized

2.3 Three conditions of annotation

- Raw text without any manual annotation: you should conduct segmentation of translation units, correct existing automatic word alignments, attribute categories of translation technique.
- Text which has already been annotated once: in order to guarantee the quality of our work, you need to verify the annotation. You can modify the alignment and category attribution if there is a disagreement. You can also modify the boundary of translation units.
- Text which has been annotated twice, and it was you who conducted the first pass: please try to reach consensus with the other annotator on the remaining differences between you two. We will provide you with the differences between the two versions to accelerate the reviewing.

2.4 External resources

Annotators are encouraged to use (online) authoritative resources, such as *Cambridge Dictionary*, *Longman Dictionary*, *TheFreeDictionary*, etc. to consult word senses, part of speech information, meaning of multiword expressions, terminology translation and so on. When annotators have doubt about word senses, they should refer to an English monolingual dictionary, instead of a bilingual dictionary.

For example, online dictionaries give all possible parts of speech of words with examples, which helps to choose the translation technique category (see figure 2).

3 Decision helper

In order to facilitate the annotation task, please see figure 3 to help you to make decisions on the most confusing categories. Please note that this table recapitulates the most distinguishing aspects for each category, and doesn't include all the definitions and rules presented below.

During the annotation, there exist different categories concerning idioms⁴ and fixed expressions⁵ in

⁴Idiom: a phrase or an expression that has a figurative, or sometimes literal, meaning. The figurative meaning is based on the whole rather than on the individual words in it. Idiomatic expressions are strongly cultural and have different meanings derived from the cultures they come from. For example: *a piece of cake*, *every cloud has silver lining*.

⁵Fixed expression: a standard form of expression that has taken on a more specific meaning than the expression itself. It is used as a part of a sentence, and is the standard way of expressing a concept or idea. Unlike idioms, they are generally transparent in meaning. For example: *as a matter of fact*, *all of a sudden*, *to whom it may concern*.

Translation Technique	Definition and important rules
Aligned segments	
Literal (6.1.1)	<p>Word-for-word translation: <i>a bronze ring</i> → 一个青铜戒指</p> <p>Borrowing words using transliteration: <i>a cup of coffee</i> → 一杯咖啡</p> <p>Possible literal translation of idioms: <i>ivory tower</i> → 象牙塔</p> <p>Corresponding expression when absolute literal translation does not make sense: <i>I give you my word.</i> → 我向你保证。("I promise you.")</p>
Equivalence (6.1.2)	<p>Non-literal translation of proverbs, idioms, or fixed expressions: <i>A friend in need is a friend indeed.</i> → 患难见真情。("Misfortune tests the sincerity of friends.")</p> <p>No change in meaning and point of view, a word-for-word translation makes sense but the translator has produced a different translation: <i>protect all locations at all times</i> → 日夜("day and night") 保护所有的地点</p>
Transposition (6.1.3)	<p>Change grammatical categories without changing the meaning: <i>She was careful not to question him, fearful that he might leave them.</i> → 她也小心地("carefully") 从不问起, 生怕("to fear") 他走了。</p>
Modulation (6.1.4)	<p>Change the point of view, can be encountered both in lexis and syntactic structures: <i>I like the dreams of the future better than the history of the past.</i> → 我不("don't") 缅怀("recall") 过去的历史, 而("but") 致力于("devote myself to") 未来的梦想。</p> <p>Slight meaning change in lexical level according to the context: <i>he had rudely bellowed across the supper table to her</i> → 他隔着餐桌对她大声("loudly") 吼叫</p>
Mod+Trans (6.1.5)	<p>Combine the transformations in <i>Modulation</i> and <i>Transposition</i>: <i>One by one the other elders now timidly rise with innocuous requests.</i> → 其他的长老一个接一个怯生生地站起来, 提出了("put forward") 一些不关痛痒的要求</p>
Particularization (6.1.6)	<p>The source segment could be translated into several target segments with more specific meaning, and the translator has chosen one of them according to the context: <i>"Yes, put you to bed"</i> → "是的, 服侍("serve") 你上床睡觉"</p> <p>Specify the meaning of a segment in context: <i>On his best days, Gomes is a very nice, solid bench player.</i> → 当他打得好的时候("play well"), 戈麦斯是很优秀、很得力的板凳球员。</p> <p>Translate a pronoun by the thing(s) it references: <i>He then requested her to stay where she was</i> → 他先让苔丝("Tess") 在外面等着</p>
Generalization (6.1.7)	<p>Several source words or expressions could be translated into a more general target word or expression, and the translator used the latter to translate: <i>a research that will be embraced by millions of bleary-eyed Britons</i> → 一项即将被广大("numerous") 睡眼惺忪的英国人所知道的研究</p> <p>The translation of an idiom by a non-fixed expression: <i>Every man has a fool in his sleeve.</i> → 人人都有糊涂的时候。("Every man is a fool sometimes.")</p> <p>The removal of a metaphorical image: <i>But should clouds gather over the Atlantic, or tempers rise in the Middle East [...]</i> → 如果大西洋风云再起, 中东战火重燃("war resumes") 的话 [...]</p>
Figurative translation (6.1.8)	<p>Introduce an idiom to translate a non-fixed expression, or a metaphorical expression to translate non-metaphor: <i>He gave the required information, in words as suitable as he could find.</i> → 他字斟句酌地("weigh one's words") 作了回答。</p> <p>Use personification to translate: <i>For Joanne, new opportunities are opening.</i> → 对乔安娜而言, 新的机遇现已向她招手。("are waving to her")</p>
Lexical shift (6.1.9)	<p>Change of verbal tense, verbal modality or of determiner, changes between singular and plural forms, and other minor changes alike. <i>My cousin is launching a photography exhibit in the USA this week.</i> → 我的表哥本周在美国推出一个摄影展。</p>
Translation error (6.1.10)	<p>Obvious translation errors. <i>database connection method</i> → 数据库访问方式</p>
Uncertain (6.1.11)	<p>Not sure about which category to assign, need more discussion.</p>

Table 1: Definition and important rules for aligned segments

the source or target language, the table 3 recapitulates them. Below is their definition.

4 Annotation tool

We use the web application Yawat (Germann, 2008) for our annotation, if you don't know how to use this tool, please read the section 7.

Translation Technique	Definition and important rules
Unaligned segments	
Explicitation (6.2.1)	<p>Introduce clarifications that are implicit in the source text: <i>the building blocks of the universe</i> → 宇宙形成("form")的最("most")基本单位</p> <p>Add Chinese-specific words: <i>the knife</i> → 这把刀 (Chinese measure word)</p> <p><i>I will bring it to China.</i> → 我可以把它带到中国来。(necessary addition due to syntactic order change in translation)</p>
Reduction (6.2.2)	<p>Deliberately remove certain words in translation:</p> <p>Removal of preposition: <i>A spokesman from the Ministry of National Defense</i> → 国防部发言人</p> <p>Removal of copula: <i>Peter is six years old.</i> → 彼得六岁。</p> <p>Removal of anticipatory « it »: <i>It was a pleasant surprise to learn of her marriage.</i> → 得知她结婚是件令人惊喜的事。</p>
No Type (6.2.3)	<p>Function words necessary in English but not in Chinese: <i>The tragedy of the world is that those who are imaginative have but slight experience.</i> → 世界的悲剧就在于有想象力的人缺乏经验。</p> <p>Segments not translated but which do not impact the meaning: <i>The present state, application and development of coal mine hydraulic drill rig are described in this paper.</i> → 介绍了煤矿用液压钻车现状, 使用情况及发展。</p> <p>Target segments added without reason, which do not correspond to any source segment.</p>

Table 2: Definition and important rules for unaligned segments

where

adverb, conjunction • UK  /weə/ US  /wer/

★ A1 to, at, or in what place

去哪里; 在哪里

Where does he live?

他住在哪里?

"I put it on your desk." "Where? I can't see it."

“我把它放在你书桌上了。”“在哪里? 我怎么没看到?”

Figure 2: Annotators are encouraged to consult authoritative dictionaries

	word-for-word	non-literal translation of proverbs, idioms, fixed expressions	change point of view	change syntactic structure	change meaning	change grammatical category	more specific meaning	more general meaning
Literal	x							
Equivalence		x						
Transposition						x		
Modulation			perhaps	perhaps	perhaps			
Mod+Trans			perhaps	perhaps	perhaps	x		
Particularization					x		x	
Generalization					x			x
Lexical shift								

Figure 3: A table to help annotators to make decisions on the most confusing categories

5 Annotation conventions

5.1 How to decide the segment boundary?

In principle

Translation phenomenon	Category attributed
literal translation of an idiom	<i>Literal</i> (see subsection 6.1.1)
idiom → equivalent idiom	<i>Equivalence</i> (see subsection 6.1.2)
idiom → non-fixed expression	<i>Generalization</i> (see subsection 6.1.7)
non-idiom → idiom	<i>Figurative</i> (see subsection 6.1.8)
non-fixed expression → fixed expression	<i>Equivalence</i> (see subsection 6.1.2)

Table 3: Different categories concerning idioms and fixed expressions

The segment boundary is not provided to annotators and it should be fixed by respecting the given tokenization, while excluding the part not involved:

I'll get my sleeve back → 然后我把袖子挽起来

experience of reading → 阅读体验 (include « of » in the boundary)

For simple literal lexical translation We annotate the smallest semantic unit as we can, for example given this pair:

this is not a Hollywood special effect → 这不是好莱坞特效

We should segment and align like this:

this → 这

not → 不

is → 是

Hollywood → 好莱坞

a special effect → 特效

Articles and prepositions

- In the following two examples, we annotate the article together with the noun, because the English article sometimes corresponds to an empty position in Chinese:

1. *have a screen and a wireless radio* → 有显示屏和无线电装置

2. *explore the answer* → 探寻答案

- Here we align « to » with 《给》, both of which are followed by the indirect object of the action « show ».

I'll show it to you → 我可以展示给你们

- Align « want to » with 《想》: regroup the preposition « to » with the verb « want ».

I want to start out by asking you → 我想请大家

- In this example, « based on » is aligned to 《基于》. Adding the preposition « on » here changes the meaning of the verb, because « base » alone means « to situate at a specified place as the centre of operations ». Then « the content » is aligned to 《内容》, because « the » corresponds to an empty position in Chinese.

based on the content inside the image → 基于图片的内容.

- Some English prepositions are translated by Chinese discontinuous segments:

in → 在...中/里; 在...的时候

for → 对...来说

as → 在...时

from → 从...中

since → 从...开始

if → 如果...的话

dive in → 钻到...里

related to → 跟...有关

paring ... down to → 分解到

so...that → 如此...甚至

stitched → (被)...拼合 (passive voice)

For non-literal translations Sometimes it is necessary to enlarge the boundary of the segments, in order to clarify the meaning, even though there are words which could be annotated as *Literal* inside this segment. For example :

1. *you get to choose which one you want to use* → 任君选择
(*Equivalence*, non-literal translation with no change in meaning, grammatical categories and point of view, subsection 6.1.2)
2. *so many of* → 不少
(*Modulation*, affirmative form → negative form, subsection 6.1.4)
3. *we damaged about 50 of the magnets* → 大概有 50 个磁铁受损
(*Modulation*, active voice → passive voice)
4. *this question was so compelling* → 这样的疑问不断萦绕在我的脑海中
(*Transposition*, adjective → verb, subsection 6.1.3)
5. *immense celebration* → 欢呼雀跃
(*Transposition*, noun → verb)
6. *become linked together* → 交织在一起
(*Transposition*, adjective → verb)

Composition of categories In the above cases, the literal part is considered as neutral, when combined with another category, the latter becomes the translation technique for the entire segment.

Thus, in our work, the translation unit could be a word, a phrase or a short sentence (do not include the final punctuation):

Nice one → 好的 (*Equivalence*, subsection 6.1.2)

What happen is this . → 它是这样发生的 。 (*Equivalence*)

5.2 How to align punctuation?

See table 4 for a recapitulation.

Punctuation	Alignment
if the punctuation doesn't change	
Final punctuation (period, exclamation mark, question mark), comma, colon, semicolon, quotation marks, angle quotes, brackets, ellipsis, etc.	Align them out of the segments.
Apostrophe, dash, hyphen	Respect the given tokenization. If you disagree with them, please contact us.
if the punctuation changes, or change between a conjunction word and punctuation	
For example, the translator replaces a double dash by a comma.	Annotate as <i>Lexical shift</i> (6.1.9).

Table 4: How to align punctuation

5.3 Mutually exclusive categories

There exist difficult borderline examples, but we do not allow multiple categorization in this task, which means that a pair of segments receive always one category listed in our typology (see figure 1).

For borderline examples, after discussion, annotators should agree on a category which better reflects the technique used by the translator.

5.4 How to annotate unaligned segments?

For the categories of *Unaligned - Explicitation* and *Unaligned - Reduction* (see definition in table 2), please annotate separately each span of reduced or added segment (separated by aligned segments), do not make a whole group.

For example in figure 4, please annotate these reduced instances separately: « *the terms of the* », « *by* ».

241	the ruling judge prager said under the terms of the 1998 agreement signed by the 46 us states , tobacco companies are barred from running cigarette ads in popular teenager magazines and from targeting youths .
	主审法官普拉格说,根据一九九八年美国四十六个州签署的协议,烟草公司不准在青少年普遍阅读的刊物上刊登香烟广告,不准以青少年为拉拢对象。

Figure 4: How to annotate unaligned reduced instances

5.5 How to deal with linguistic anaphora?

Linguistic anaphora is different from rhetorical anaphora which consists of repeating words at the beginning of phrases in order to make emphasis.

Linguistic anaphora can be the repetition of noun subject or words of other parts of speech. In case of using anaphora in translation, we annotate all the repeated words, which are not necessarily the same, with the English source word together. The annotation type is decided by the translation technique.

Here are some examples :

1. The noun subject is repeated and annotated as *Literal* (see figure 5). «两国» ("the two countries") refers to « Pakistan and India » which are not directly translated by their names.

124	pakistan and india have massed about 1 million troops along their common border and have been exchanging bellicose statements . they have also been exchanging fires in the endlessly disputed kashmir in the past week .
	巴基斯坦和印度已在两国边界集结约一百万大军,两国交相发表好战声明,过去一周并在争执不断的克什米尔交火。

Figure 5: Repetition of subject, annotate as literal

2. In figure 6, the verb and adjective are repeated and annotated as *Particularization*. « many » is translated into «一次又一次»和«无数次»;«made» is translated into «发»和«下».

2	And in this agony of mind , I made many vows and resolutions
	在这种惶恐不安的心情下,我一次又一次地发誓,下了无数次决心

Figure 6: Repetition of verb and adjective, annotate as *Particularization*

3. The determiner is repeated by using the same word, annotate as *Literal*.

I've loved another with all my heart and soul .

→我用我的心和我的灵魂爱过另一个人。

6 Annotation in practice

6.1 Annotation of aligned segments

Literal translations

6.1.1 Literal

Definition Word-for-word translation (including insertion or deletion of articles), or possible literal translation of some idioms:

the sunlight is very warm → 阳光很温暖

I've loved another with all my heart and soul . → 我用我的心和我的灵魂爱过另一个人

(According to bilingual dictionaries, « with » can be translated by «用» according to the context, we consider this as a literal translation.)

The book is a collection of his reminiscences about the actress .

→ 这本书辑录了他对那位女演员的回忆。

Rule 1 Borrowing English words using transliteration:

a cup of coffee → 一杯咖啡

Rule 2 When an absolute word-for-word translation does not make sense in the target language, the corresponding expression is deemed to be literal:

in other words (?用 其它 话) → 换句话说

I give you my word . (?我 给你 我的 话) → 我 向你 保证 。

Rule 3 The change of numbers between the spelled out form and the Arabic form is annotated as *Literal*:

nearly 4 million tons of food must be imported → 必须 要 进口 接近 四 百万 吨 的 粮食

Rule 4 When the Chinese structural particle 《的》 marks the possession:

1. *Some people come into **our** life and quickly go* .

→ 一些 人 闯进 我们 的 生活 , 又 匆匆 走 了 。

(regroup 《的》 and the personal pronoun before it, annotate as *Literal*)

2. *in any deal , you need to know your opponent 's breaking point*

→ 在 任何 交易 中 你 都 需要 知道 对手 的 忍耐 度

(when 《的》 is after a noun, align 《的》 with «'s »)

3. *His father emptied sacks of stale rye bread into the vat* .

→ 他 (的) 父亲 把 一 袋 袋 发霉 的 黑 面包 倒 进 大 桶 里 。

(Sometimes the 《的》 after a personal pronoun could be omitted in oral conversation or informal text, in this case, annotate as *Literal*.)

Rule 5 When the Chinese particle 《的》 is added after an attributive adjective, annotate 《的》 and the adjective together as *Literal*:

It 's a beautiful 3D graphics . → 它 是 漂亮 的 3D 图画

Rule 6 The translation of proper noun is annotated as *Literal*:

*Saltanov will discuss Russian president **Putin** 's proposal* .

→ 沙福诺夫 将 讨论 俄罗斯 总统 普京 的 提议 。

Rule 7 Fixed modulation is annotated as *Literal*:

a life jacket → 一 件 救生 衣

(The means 《救生》 ("save life") substituted for the result « life », from a linguistic point of view, the category should be *Modulation*. However, since there is no other possible translation for « life jacket », and it's a recorded pair in bilingual dictionaries, we annotate this case by *Literal*.)

For each translation technique, we show some counterexamples and borderline examples:

Counterexamples:

the arms reduction treaty → 裁减 武器 条约

(The translation is word-for-word, but 《裁减》 ("reduction") in Chinese is a verb. The category should be *Transposition*.)

Borderline examples:

in the past week → 上 周

(This is a borderline example with *Equivalence*, but should be annotated as *Literal*.)

Non-literal translations

6.1.2 Equivalence

Definition There is no change of point of view like in *Modulation* (subsection 6.1.4). A word-for-word translation makes sense but the translator has expressed differently. However, if there exist changes of grammatical categories, the pair should be annotated with *Transposition* (see examples in subsection 6.1.3).

She decides to string along with others as she has nothing else to do .

→ 她反正也无事可做，所以就跟着大家去了。

(« as » should be translated into 《由于》 ("since") but in this sentence, it's translated into 《反正》 ("anyway") to emphasize the reason.)

Rule 1 Non-literal translation of proverbs, idioms or fixed expressions:

A friend in need is a friend indeed . → 患难见真情。

Like father , like son → 有其父必有其子

Rule 2 Cultural equivalence :

In the figure 7, the segment « great to » is translated into 《荣幸》 ("it's a great honour to ..."). A word-for-word translation makes sense here but it's necessary to show the honoured side in the Chinese context.

10	There 's been a great reaction on the doorstep all over London , London LibDems are working so hard , great to have such support !
	伦敦 市民 很 欢迎 我们 敲门 到访 , 感谢 自民党 团队 的 努力 , 很 荣幸 得到 那么 多 支持

Figure 7: Cultural equivalence

swear to God → 对天发誓

Rule 3 Change measure into the one used in the target culture:

I bought 5 half kilo of banana . → 我买了五斤香蕉。

Rule 4 Translate a non-fixed expression by a fixed expression:

Rule 5 Translate an abbreviation into a full version or vice versa:

SAFE → 国家外汇管理局

(The full version of « SAFE » is « State Administration of Foreign Exchange ».)

Rule 6 Translate the title of a book:

THE COMPLETE FORTUNE-TELLER → 《算命大全》

Counterexamples:

in other words → 换句话说

(The word-for-word translation does not make sense in Chinese so the category is *Literal*.)

Borderline examples:

along the Kashmir demarcation of control separating India from Pakistani territories

→ 沿着分隔印度与巴基斯坦辖区的克什米尔控制分界线

(Borderline with *Particularization*. According to Cambridge Dictionary, « territories » corresponds to 《领土，领域，活动范围》，not 《辖区》. Annotate as *Equivalence* for now.)

6.1.3 Transposition

Definition Translating words or expressions by using other grammatical categories than the ones used in the source language, without altering the meaning of the utterance.

The change of grammatical categories could occur on a complete syntagm, for example:

- " *He who asserts must prove* " **is rational** → “谁主张谁举证”有其合理性
(« is rational » is translated into 《有 合理性》 ("has rationality"))
- *Those who **are experienced** have feeble imagination .*
→ 有 经验 的人 缺乏 想象力 。
(« are experienced » is translated into 《有 经验》 ("have experience"))

verb -> noun

*London LibDems are **working** so **hard*** → 感谢 自民党 团队 的 努力

noun -> verb

*It was a pleasant surprise to learn of her **marriage** .*

→ 得知 她 结婚 是 件 令人 惊喜 的 事 。

adj -> verb

*I 'd like to talk about the oldest **nutritional** method on Earth .*

→ 我想 讲 一下 地球 上 最 原始 的 获得 营养 的 方式 。

adj -> noun

In figure 8, the English adjective « russian » is translated by a noun 《俄罗斯》 (the name of the country). Because in Chinese grammar, the noun is used as an adjective here (Fang, 2008).

20

returning together with shuttleworth to earth are the **russian** spacecraft commander gidzenko and the italian engineer vittori **who** entered space with him .

与 夏特沃斯 一同 返回 地球 的 , 是 这次 和 他 一起 进入 太空 的 **俄罗斯** 太空船 指挥官 吉曾柯 与 义大利 工程师 维托利 。

Figure 8: In Chinese, a noun can be used as an adjective

prep -> verb

*patients **over** the age of 40* → 超过 四十 岁 的 病人

When the Chinese particle 《的》 is used to replace a preposition or a relative pronoun in order to describe the subject, the annotation type should be *Transposition*.

prep -> particle

*people **of** Iran* → 伊朗 的 人们

*The Irish representative team is currently training on Saipan **in** the pacific ocean .*

→ 爱尔兰 代表队 目前 在 太平洋 的 塞班岛 集训 。

*The air force **from** the two nations will provide support .*

→ 两国 的 空军 将 给 以 支援 。

relative pronoun -> particle

*Every citizen **who** is able to work .* → 一切 有 劳动 能力 的 公民 。

determiner -> pronoun

*It was a pleasant surprise to learn of **her** marriage .*

→ 得知 她 结婚 是 件 令人 惊喜 的 事 。

Counterexamples:

*Now boarding all passengers for flight 624 **to** Newyork .*

→ 现在 所有 乘坐 624 航班 飞往 纽约 的 旅客 请 登机 。

(There is not only the change of grammatical category but also the meaning: 《飞往》 ("fly to"). So it should be annotated as *Mod+Trans*.)

Borderline examples:

*There is no need to see the little detailed words **on** it .*

→ 不用 去 看 里面 具体 的 小字 。

(Here, « on » is a preposition and 《里面》 ("inside") is a noun. The meaning and the part of speech are both changed. Although it is borderline with *Lexical shift*, it's annotated as *Transposition* now.)

By the standards of history, his execution in 399 B.C. was singularly humane.

→ 按照历史的标准来看，公元前399年处死他的方式是异常人道的。

(Borderline with *Equivalence* (in this specific context). The point of view does not change.)

6.1.4 Modulation

Definition This translation technique consists of a change in the point of view that enables the translator to convey the same phenomenon in two languages in a different way.

- can be encountered both in lexis and in syntactic structures
- reveal a specific way of seeing things for the speakers of the target language
- circumvent translation difficulties
- the translator desires to achieve expression naturalness in the target language
- could bring meaning changes between source and target text

According to Vinay and Darbelnet (1958), there exist two types of modulation:

- metonymical modulations (the cause substituted for the effect, the container for the content, the part for the whole, etc. and vice versa)
- grammatical modulations (change between affirmative form and negative form; between injunction and interrogation; between passive voice and active voice; the subject becomes the object, etc.)

Examples:

Metonymical modulations:

1. *If you want, you can lay color in the mold, and get rid of the paint shop.*
→ 如果你愿意，你可以在模具里直接加颜色，就省得给车喷漆了。
(Use the concrete action 《给车喷漆》 ("paint the car") to substitute for the abstract « the paint shop ».)
2. *I'm thinking of becoming a suicide bomber.*
→ 我正在想如何成为一个自杀炸弹。
(Use the object 《炸弹》 ("bomb") to substitute for the « bomber ».)
3. Geographical modulation:
In English « Downing street » is often used to refer to « the office of the UK prime minister ».
By metonymy, English speaker often refer to 《中国政府》 ("Chinese government") by « Beijing ».

Change the point of view:

4. *Louis Hicks, what's the matter? Work a little slow?*
→ 路易斯希克斯，怎么了？没事做吗？
5. *Just needs a little something more.* → 还少了一点东西。

Change between passive voice and active voice:

6. *nearly 4 million tons of food must be imported* → 必须要进口接近四百万吨的粮食

Affirmative form to negative form, the negation of the opposite:

7. *I like the dreams of the future better than the history of the past.*
→ 我不缅怀过去的历史，而致力于未来的梦想。
(《不.....而》 means « not...but ». Since the affirmative form « like...better than » is changed into negative form « not...but », the sentence's order is reversed: « the history of the past » is now at the beginning of the sentence.)
8. *It's difficult.* → 这不简单。

The subject becomes the object:

9. *What do you hope to gain from those colors?*
→ 那颜色能给你什么希望？

Obligatory syntactic change but no change in meaning:

10. **There have been increased** contacts between the United States and India .

→ 美印间的接触 **增加** 。

(The English structure « There be + noun » can't be translated in a word-for-word way in Chinese, so the adjective « increased » is transformed to the intransitive verb 《增加》 ("increase"). Thus, the noun phrase « contacts between the United States and India » becomes the subject in Chinese.)

Circumvent translation difficulties, achieve expression naturalness:

11. **Maybe** we should postpone till the weekend . → **不如** 我们推迟到周末吧 。

Slight meaning change in lexical level according to the context:

12. British airways **posts** first annual loss in 15 years .

→ 英航 **出现** 十五年来首次年度亏损 。

13. **Stop by** tomorrow and drop off your ID badge and all that .

→ 明天 **去** 店里交上你的工作证还有别的 。

(《去》 means « go to ». In this sentence, « stop by » is translated into « go to » which is more natural in the context.)

14. I want to tell **you** three things today . → 我今天想告诉大家三件事 。

(In this specific context, the speaker is talking to an audience, so « you » is translated into 《大家》 ("everybody").)

15. she remembered how frequently Gerald had **rudely** bellowed across the supper table to Suellen

→ 她想起杰拉尔德时常隔着餐桌对苏伦 **大声** 吼叫

Counterexamples:

Borderline examples:

My plan is to live here until I 'm 30 , and then enter a hermitage .

→ 我的计划是在那里一直住到我三十岁，然后 **过** 起隐居生活 。

(Borderline with *Equivalence*, annotated as *Modulation* now. 《隐居生活》 ("live in seclusion") substitutes « hermitage » which is the place to live in seclusion.)

6.1.5 Mod+Trans

This category combines the transformations in *Modulation* and *Transposition*, for example:

In the figure 9, the noun phrase « the attack on the national parliament in New Delhi last december » is transformed into a verbal phrase 《去年十二月新德里国会遭攻击》. As a result, the preposition « on » is semantically aligned to the verb 《遭》 ("encounter"). This is a case of transposition made necessary by the syntactic changes.

13

after the sept. 11 terrorist attacks , and the attack on the national parliament in new delhi last december that led to military tensions between india and pakistan , there have been increased contacts between the united states and india .
继九一一恐怖攻击事件，以及去年十二月新德里国会遭攻击致使印度与巴基斯坦呈现军事紧张情势之后，美印间的接触增加。

Figure 9: Necessary transposition caused by a syntactic change

1. Tobacco companies are barred from running cigarette ads in popular teenager magazines and from **targeting** youth .

→ 烟草公司不准在青少年普遍阅读的刊物上刊登香烟广告，不准以青少年为拉拢对象 。

(The verb « target » in English becomes a noun 《拉拢对象》 ("target") in Chinese, and the verbal phrase « targeting youth » is transformed into a noun phrase 《以青少年为拉拢对象》 ("taking

youth as target"), so the whole part is annotated as *Mod+Trans*.)

2. When a preposition is translated into a verb to define the action, annotate the preposition with the verb as *Mod+Trans*.

For example, in the figure 10, the preposition « with » is translated into 《提出了》 ("put forward").

10

One by one the other Elders now timidly rise with innocuous requests , which Rahm receives warmly .
其他的 长老 一个 接 一个 怯生生 地 站起来 , 提出了一些 不关痛痒 的 要求 , 拉姆 全 热情 地 接纳 了 。

Figure 10: Translate a preposition into a verb

Counterexamples:

What I 'd love is for my highly paid publicist to take care of it .

→ 我想要我高收费的公关经理搞定这事。

(There is only *Modulation* used in this translation. The point of view is changed from « highly paid » to 《高收费》 ("highly charged").)

Borderline examples:

6.1.6 Particularization

Definition The source word could be translated into several target words with more specific meaning, and the translator has chosen one of them according to the context.

" Yes , *put* you to bed , " she added lightly → " 是的 , 服侍你上床睡觉 , " 她小声补充说

Longer boundary Annotators could enlarge the boundary to help disambiguate the meaning of the segment:

1. *On his best days , Gomes is a very nice , solid bench player .*

→ 当他打得好的时候 , 戈麦斯是很优秀、很得力的板凳球员。

Rule 1 Specify the meaning of a word in context is annotated as *Particularization*:

drag yellow diamonds to adjust arrowhead or tail → 拖动黄色菱形可以调整箭头或箭尾。

Rule 2 Translate a pronoun by the thing(s) it references is annotated as *Particularization*:

Who is which depending on whether the customer 's account is in credit or is overdrawn .

→ 究竟谁是债务人谁是债权人 , 要看储户是有结余还是透支。

Rule 3 Use different words to give colour to translation is annotated as *Particularization*:

Pakistan and India have massed about 1 million troops along their common border .

→ 巴基斯坦和印度已在两国边界集结约一百万大军。

Counterexamples:

the plan includes a lay-off of 2,000 employees → 这项计划括及裁员两千人

(This is an instance of *Generalization*.)

Borderline examples:

6.1.7 Generalization

Definition This translation technique is the opposite of *Particularization*. Several source words or expressions could be translated into a more general target word or expression, and the translator used the

latter to translate. Or some semantic properties have been lost via the generalized translation.

1. *a research that will be embraced by millions of bleary-eyed Britons*
→ 一项即将被广大睡眠惺忪的英国人所知道的研究
(《广大》("vast") is much more general than « millions of ».)
2. *the plan includes a lay-off of 2,000 employees* → 这项计划括及裁员两千人
(« employees » is translated into 《人》("people") which is a more general word.)

Rule 1 The translation of an idiom by a non-fixed expression is annotated as *Generalization*:

Every man has a fool in his sleeve .

→ 人人都有糊涂的时候。

(Literally, the translation means « everyone has a confused moment » which is not an idiom like in English.)

be a far cry from → 和.....完全不同

Rule 2 The removal of a metaphorical image is annotated as *Generalization*:

- *Don 't make me go through all of this and not make it .*

→ 别让我的辛苦白费了。

(« Go through all of this and not make it » is translated into 《辛苦》("hard work") 《白费》("waste"), so there is no more metaphorical image like in the source sentence.)

- *But should clouds gather over the Atlantic, or tempers rise in the Middle East, the price could jump again.*

→ 如果大西洋风云再起，中东战火重燃的话，无疑，油价会再度上涨。

(« tempers rise » is translated into 《战火》("flames of war") 《重燃》("burn again"), so there is no more metaphorical image like in the source sentence.)

Rule 3 Use pronoun to translate the thing(s) that it references:

the couple broke up shortly before this interview → 他们在接受这次采访的不久前分手了

Counterexamples:

He is trying to fight off the advances of his rival .

→ 他正试图击溃对手的进攻。

(This is an instance of *Particularization*.)

Borderline examples:

All relevant departments will co-operate to publicise this MPF system which affects the general public .

→ 有关部门会齐心协力，推广这项影响广大市民的计划。

(It's a borderline with *Modulation*, annotated as *Generalization* for now. (MPF : Mandatory Provident Fund System.))

6.1.8 Figurative translation

Rule 1 Introduce an idiom to translate a non-fixed expression, or a metaphorical expression to translate non-metaphor:

pressured the people a little bit about it → 刨根问底 ("inquire into the root of the matter")

He gave the required information, in words as suitable as he could find.

→ 他字斟句酌地作了回答。

Rule 2 Use personification to translate:

For Joanne, new opportunities are opening . → 对乔安娜而言，新的机遇现已向她招手。

(《现已向她招手》 means « is waving to her ».)

Counterexamples:

Borderline examples:

6.1.9 Lexical shift

Definition The translation is not literal, but there is no change in meaning. They are minor lexical level changes, which do not involve any translation technique.

change verbal tense or verbal modality (without changing the verb):

*He also **indicated** that the United States will hold negotiations with Cuba*

→ 他还表示，美国将与古巴举行谈判

differences between plural and singular form:

*include the following additional **responsibilities*** → 包括下列新增的职责

remove the passive voice:

*The game **was** temporarily **suspended** for 20 minutes.*

→ 比赛暂停了二十分钟。

(« was suspended » should be translated into 《被暂停》，but in Chinese, it's not always necessary to translate 《被》 ("be"). In this case, the pair of verb is annotated as *Lexical shift*.)

Counterexamples:

*The book is a **collection** of his reminiscences about the actress .*

→ 这本书辑录了他对那位女演员的回忆。

The meaning of the word doesn't change but the part of speech is altered, so it is a *Transposition*.)

6.1.10 Translation error

Definition This category concerns obvious translation errors.

1. *a funny video **about** a newly married couple .* → 一个有趣的视频对新婚夫妇。
2. *database **connection** method* → 数据库访问方式

6.1.11 Uncertain

Please attribute this category when you do not know which label to assign: because it is a difficult borderline example, or because there is something not clear in this annotation guide, and you want more discussion about this pair of segments.

6.2 Annotation of unaligned segments

Definition Certain segments are left unaligned, there exist these three cases:

6.2.1 Unaligned - Explicitation

Definition Translations that could not be aligned to any source segment.

Rule 1 Resumptive anaphora (Charolles, 2002): add a phrase or sentence summarizing the preceding information (which could be present in the previous sentence), to help understanding the present sentence.

and it does n't matter how much information we 're looking at , how big these collections are or how big the images are

→ 不管所见到的数据有多少、图像集有多大以及图像本身有多大，*Seadragon* 都拥有这样的处理能力。

(The last phrase is an example of resumptive anaphora, which refers to information present in the previous sentence (in the corpus of TED Talks).)

Rule 2 Introduce in the target language clarifications that remain implicit in the source language but emerge from the situation; Personal interpretation or explanation of the translator:

- *the building blocks of the universe* → 宇宙形成的最基本单位
- *most of them are* → 大部分都是
(to emphasise)
- *hold negotiations with Cuba on resuming direct mail service*
→ 与古巴就重新恢复直接通邮举行谈判
- *London LibDems are working so hard, great to have such support!*
→ 感谢 ("Thanks to") 自民党团队的努力，很荣幸得到那么多支持！

Rule 3 Add Chinese-specific words:

- Chinese measure words:
the knife → 这把刀
a pharmaceutical → 一种药物作用
- Verbs added in special text genre, for example in newspaper article:
AFP, Washington → 法新社华盛顿电
- Add obligatory word for the date:
October 13 → 十月十三日

Rule 4 Add logical connectives:

Get 'boring' done as early as possible today and you will be prepared for anything.
→ 今天只要尽快搞定那些无聊的东西，你就能应付各种事情了。

Rule 5 Necessary addition due to syntactic order change in translation. In Chinese, we often use preposition like 《把》，《将》，《使》 etc. to advance the object to the position before the verb.:

- *I will bring it to China.* → 我可以把它带到中国来。
(By adding 《把》，« bring it » is translated into 《把》《它》("it")《带到》("bring to").)
- *Strong political commitment has significantly reduced the burden of tuberculosis borne by India and China.*
→ 坚定的政治承诺使得印度和中国的肺结核疾病负担显著缩减。
(By adding 《使得》，the transitive verb « has reduced » becomes an intransitive verb 《缩减》("decrease") and is moved to the end.)

Rule 6 When the relative pronoun is absent in the source side while 《的》 exists in the target side, annotate 《的》 as *Explicitation*.

The US federal government has grossly violated the promise it made in the trade negotiation (which was) held last November in Doha.

→ 美国联邦政府违背了去年十一月在卡塔尔首都多哈举行的贸易谈判上所作的承诺。

6.2.2 Unaligned - Reduction

Definition Remove certain words that could have been translated. See rules below for each specific case.

1. *we started this over 15 years ago* → 我们于15年前开始这项工作
2. *one by one the other elders now timidly rise*
→ 其他的长老一个接一个怯生生地站起来
3. *let me show you* → 我演示给你们看

4. *Irish soccer team captain Keane announces he **will** retire after the World Cup .*

→ 爱尔兰足球队队长基恩宣布世界杯之后退休。

Rule 1 Removal of preposition:

*A spokesman **from** the Ministry of National Defense* → 国防部发言人

(In the translation, there is no preposition between 《发言人》 ("spokesman") and 《国防部》 ("the Ministry of National Defense").)

Rule 2 Removal of determiner:

*Bush proposed in **his** speech to step up the humanitarian assistance to Cuba*

→ 布什在演说中提议，增加对古巴的人道援助

(There is no determiner before 《演说》 ("speech").)

Rule 3 Removal of noun:

*Declaration on the Protection of Right of **Persons** belonging to National , Ethnic , Religious and Linguistics Minorities .*

→ 保护民族，种族，语言，宗教上属于少数人的权利宣言。

(The noun « Persons » is not repeated in the translation.)

Rule 4 Removal of pronoun:

*When she gets there , **she** finds out that there are no doctors .*

→ 但是当她到那里时却发现并没有医生。

(The subject « she » is not repeated in the translation.)

Rule 5 Removal of copula:

*Peter **is** six years old .* → 彼得六岁。

(The copula « is » is not translated.)

*Its glimmer **was** yet dim in the plain below .*

→ 暗淡的微光，在平原依稀可见。

Rule 6 Removal of anticipatory « it » : In English, « it » is used to introduce or ‘anticipate’ the subject or object of a sentence, especially when the subject or object of the sentence is a clause. It’s not translated in Chinese.

***It** was a pleasant surprise to learn of her marriage .*

→ 得知她结婚是件令人惊喜的事。

6.2.3 Unaligned - No Type

Rule 1 Function words necessary in English but not in Chinese:

*The tragedy of the world is **that** those who are imaginative have but slight experience .*

→ 世界的悲剧就在于有想象力的人缺乏经验。

the two gentlemen leaned forward and looked at each other

→ 这两位绅士身体前倾，互相你看看我、我看看你

(« and » is often replaced by a comma)

Rule 2 Segments not translated but which do not impact the meaning:

*The present state , application and development of coal mine hydraulic drill rig are described **in this** paper .*

→ 介绍了煤矿用液压钻车现状，使用情况及发展。

***you know** , we do n’t know necessarily what it ’ll look like*

→ 我们并不一定知道它看起来是怎样

Rule 3 Target segments added without reason, which do not correspond to any source segment:

then , after about 30 seconds , it reshuffles , and you have a new set of letters and new possibilities to try .

→ 大约 30 秒后 Siftables 会自动根据正确的单词拼写顺序改变原来显示的字母，组成正确的单词，你还可以尝试用其他的字母拼出新的单词。

(The content is added by the translator but not expressed by the speaker.)

Rule 4 《鼓掌》("applause"), 《笑声》("laughter"): specific in the subtitles of TED Talks.

7 Tutorial for using Yawat

In order to conduct all the operations available in Yawat, please use Firefox (version 67 or before)⁶ as navigator instead of Chrome or Safari.

Url : <https://pakin.limsi.fr/cgi-bin/cgi/yawat.cgi>

Every annotator has an account created by the administrator. After logging in, according to the current task, please find the corresponding sub-corpus in different directories (e.g. pass1, pass2, pass3). Then click the corpus name to begin the annotation. (We also prepare a small test corpus for you to be familiar with the functions in Yawat.)

Here are some instructions about using the tool:

1. You can change the layout to show pairs of sentences in two lines or in two columns, by clicking on the button in the top right corner.
2. The black color is by default which means the category *Literal*. (But this does not mean that the boundary and alignment are all correct. Annotators should verify the alignment of all words.)
At source side: blue means *unaligned and no type*.
At target side: red means *unaligned and no type*.
3. Automatic word alignments have been imported, annotators should correct them when necessary. For example, in figure 11, there was just *open* and *ouvre* aligned automatically, and we want to add *up* to *open*.
First, left click on either *open* or *ouvre*, the pair will be in a state of "being chosen", then left click *up*, we can see a new discontinuous boundary is created. Finally right click on any of these three words, the new alignment will be confirmed.

but that 's not really what this is about . right ? and then I open my hand up .
ce n' est pas vraiment le sujet . là , j' ouvre ma main .

Figure 11: How to correct alignment in Yawat

4. If we want to remove a word from a segment, right click on it, and click *remove ... from this group*. If it is a lexical pair like the example in figure 12, when removing one word from one side, the link between them is broken in consequence. The corresponding word will change into the *unaligned and no type* state a moment later, you can hover your mouse on it to confirm it.

the knife , which you can examine ; my hand , which you could examine .
le couteau , que vous pouvez examiner ; ma main , que vous pouvez examiner .
1850
I 'm just going to hold the knife in my fist like this .
remove 'pouvez' from this group dissolve this group label group as ... literal equivalence

Figure 12: How to remove a word from a segment

5. To facilitate the alignment of a pair of long segments, for example in figure 13, first left click *nothing* and *il*, then right click either of them to create this pair.

⁶If you need to use higher version of Firefox for other applications, you can continue to use Yawat by downloading the Firefox ESR version here: <https://www.mozilla.org/en-US/firefox/organizations/>.

Next, left click this pair (by clicking either word inside) and add the rest of the segment by left clicking the other words, and finally right click on any word of them to confirm this long segment: *nothing goes up or down my sleeve* → *il n'y a rien dans mes manches*.

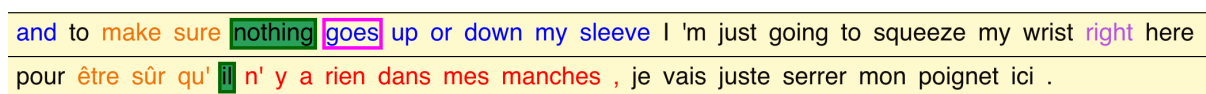


Figure 13: How to align long segments in Yawat

6. In order to annotate unaligned segments, *i.e.* *Explicitation* and *Reduction*, left click words to fix the boundary then right click the segment to choose the category from the menu (figure 14).

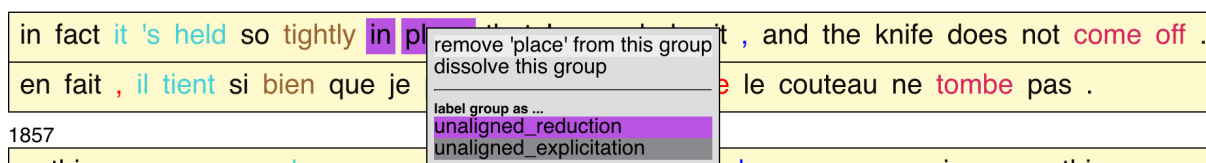


Figure 14: How to annotate unaligned segments

7. Once the boundary and the alignment are fixed, we can attribute a category. Right click any word in the pair, and choose a category from the menu (figure 15).

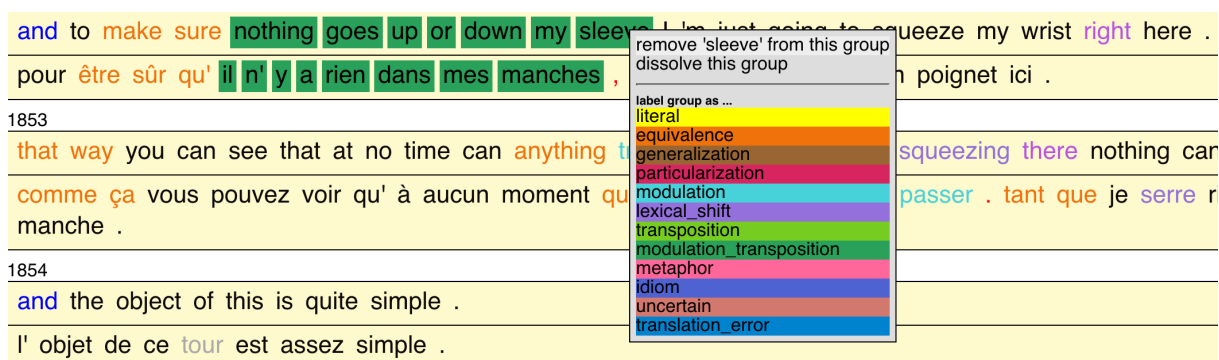


Figure 15: Choose category from the menu

8. Hover the mouse on words to check the boundary and the alignment of segments.
 9. Make sure to save your annotations before logging out (click the **[save]** button in the top right corner, it appears only after changes).
- Please contact us if you have any trouble in using this interface.

References

- Michel Charolles. 2002. *La référence et les expressions référentielles en français*. Ophrys.
- Hélène Chuquet and Michel Paillard. 1989. *Approche linguistique des problèmes de traduction anglais-français*. Ophrys.
- Yuqing Fang. 2008. 实用汉语语法(第二次修订本) (*Practical Chinese Grammar*). 北京语言大学出版社.
- Ulrich Germann. 2008. Yawat: Yet Another Word Alignment Tool. In *ACL 2008, Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, June 15-20, 2008, Columbus, Ohio, USA, Demo Papers*, pages 20–23. The Association for Computer Linguistics.
- Kludia Gibová. 2012. *Translation Procedures in the Non-literary and Literary Text Compared*. Books on Demand.

- Zhongguo Li and Maosong Sun. 2009. Punctuation As Implicit Annotations for Chinese Word Segmentation. *Computational Linguistics*, 35(4):505–512, December.
- Yang Liu and Maosong Sun. 2015. Contrastive unsupervised word alignment with non-local features. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI’15, pages 2295–2301. AAAI Press.
- Lucía Molina and Amparo Hurtado Albir. 2002. Translation Techniques Revisited: A Dynamic and Functionalist Approach. *Meta*, 47(4):498–512.
- Peter Newmark. 1981. *Approaches to Translation (Language Teaching Methodology Series)*. Oxford: Pergamon Press.
- Peter Newmark. 1988. *A textbook of translation*, volume 66. Prentice Hall New York.
- Jean-Paul Vinay and Jean Darbelnet. 1958. *Stylistique comparée du français et de l’anglais: méthode de traduction*. Bibliothèque de stylistique comparée. Didier.