

Annotation Guidelines of Translation Techniques for English-French

Yuming Zhai, Gabriel Illouz, Anne Vilnat

LIMSI, CNRS

Université Paris-Sud, Université Paris-Saclay

Orsay, France

zhai@limsi.fr

Contents

1	Introduction	2
1.1	Task description	2
1.2	Translation techniques	2
2	General instructions	4
2.1	How to deal with minor misspelling and tokenization errors?	4
2.2	Three conditions of annotation	4
2.3	External resources	4
3	Decision helper	5
4	Annotation tool	5
5	Annotation conventions	5
5.1	How to decide the segment boundary?	5
5.2	How to align punctuation?	7
5.3	Mutually exclusive categories	7
5.4	How to annotate unaligned segments?	7
5.5	How to deal with linguistic anaphora?	7
6	Annotation in practice	8
6.1	Annotation of aligned segments	8
6.1.1	Literal	8
6.1.2	Equivalence	9
6.1.3	Transposition	10
6.1.4	Modulation	12
6.1.5	Mod+Trans	13
6.1.6	Particularization	14
6.1.7	Generalization	15
6.1.8	Figurative translation	15
6.1.9	Lexical shift	16
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	18
7	Tutorial for using Yawat	19

1 Introduction

1.1 Task description

This document is a guide for annotating translation techniques in a parallel corpus. For example in our work, the corpus is composed of transcriptions and human translations of TED Talks¹.

The total corpus contains 19 talks and 2 436 lines of parallel sentences (the sentence alignment is already done). These topics are covered: technology, psychology, culture, science, biology, etc. Each sentence pair contains on average 21 English tokens on source side and 22 French tokens on target side.

1.2 Translation techniques

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

that image reminded me of something . → *cette image m' a rappelé quelque chose .*

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

that → *cette*

image → *image*

reminded of → *a rappelé*

me → *m'*

something → *quelque chose*

. → *.*

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-French parallel corpus, we have proposed a typology of translation techniques (see figure 1) in order to have a global view of these categories. The tables 1 and 2 provide a recapitulation of definition and important rules for each translation technique.

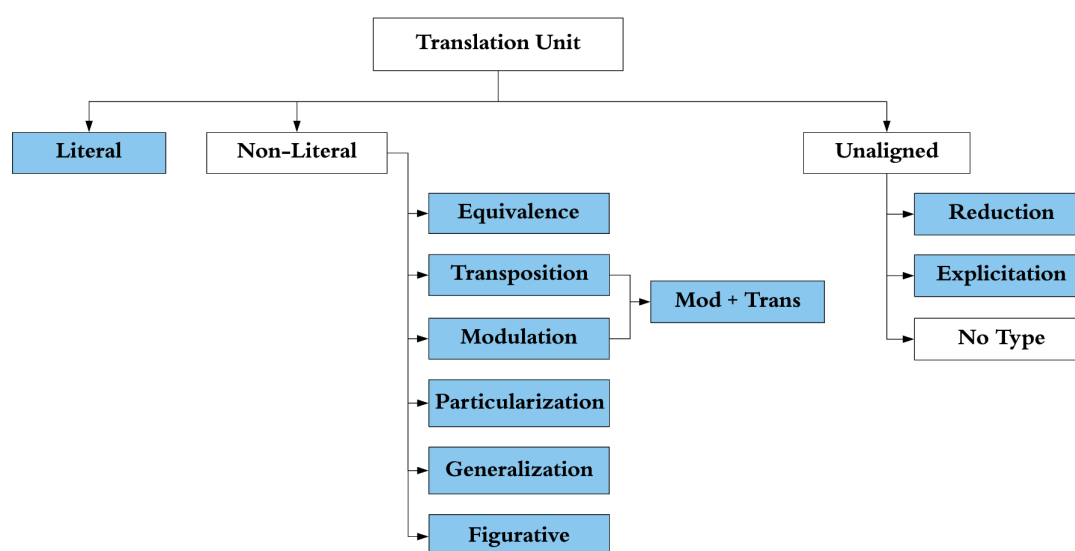


Figure 1: Typology of translation techniques

¹<https://www.ted.com/>

Translation Technique	Definition and important rules
Aligned segments	
Literal (6.1.1)	<p>Word-for-word translation;</p> <ul style="list-style-type: none"> - possible literal translation of idioms; - concerns lexical units in multiword form; - corresponding expression when absolute literal translation does not make sense. <p><i>certain kinds of → certains types de</i></p>
Equivalence (6.1.2)	<p>A word-for-word translation makes sense but the translator has expressed differently;</p> <ul style="list-style-type: none"> - non-literal translation of proverbs, idioms, or fixed expressions; - no change in meaning and point of view. <p><i>sense each other → se reconnaître entre eux</i></p>
Transposition (6.1.3)	<p>Change grammatical categories without changing the meaning.</p> <p><i>and , if anything , at a far greater rate . → et peut-être bien plus rapidement .</i></p>
Modulation (6.1.4)	<p>Change the point of view;</p> <ul style="list-style-type: none"> - can occur both on lexical level and in syntactic structures; - metonymical and grammatical modulation; - could bring slight meaning changes. <p><i>the statistics you hear about → les statistiques qui nous sont communiquées</i></p>
Mod+Trans (6.1.5)	<p>Combine the transformations of <i>Modulation</i> and <i>Transposition</i>.</p> <p><i>this is a completely unsustainable pattern . → il est absolument impossible de continuer sur cette tendance .</i></p>
Particularization (6.1.6)	<p>The translation has a more concrete or particular meaning;</p> <ul style="list-style-type: none"> - specify the meaning of a word in context; - translate a pronoun by the thing(s) it references. <p><i>they have a screen and a wireless radio → ils sont équipés d' un écran et d' une radio sans fil</i></p>
Generalization (6.1.7)	<p>The translator used a target word or segment whose meaning is more general than that of the source word or segment;</p> <ul style="list-style-type: none"> - the translation of an idiom by a non-fixed expression; - the removal of a metaphorical image; - use pronoun to translate the thing(s) that it references. <p><i>as we sit here today in Monterey → alors que nous sommes à Monterey aujourd'hui</i></p>
Figurative translation (6.1.8)	<ul style="list-style-type: none"> - introduce an idiom to translate a non-fixed expression, or a metaphorical expression to translate non-metaphor; <p><i>at any given moment → à un instant " t "</i></p> <ul style="list-style-type: none"> - keep the same metaphorical image by using a non-literal translation; <p><i>the Sun begins to bathe the slopes of the landscape → le soleil qui inonde les flancs de ce paysage</i></p>
Lexical shift (6.1.9)	<p>Change verbal tense or verbal modality, preposition, determiner, subject, position adverb, between singular and plural forms.</p> <p><i>when you do a web search for images → quand on fait une recherche web sur des images</i></p>
Translation error (6.1.10)	<p>Obvious translation errors.</p> <p><i>is not going to be remembered for its wars → ne sera pas reconnu pour ses guerres</i></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

Translation Technique	Definition and important rules
Unaligned segments	
Explicitation (6.2.1)	<ul style="list-style-type: none"> - resumptive anaphora; - introduce clarifications that is implicit in source text. <p><i>[...] live amongst those who have not forgotten the old ways, who still feel their past in the wind</i> → <i>[...] vivre parmi ceux qui n'ont pas oublié les anciennes coutumes, qui ressentent encore leur passé souffler dans le vent</i></p>
Reduction (6.2.2)	<p>Deliberately remove certain words with concrete meaning that could be translated.</p> <p><i>look carefully at the area of the eastern Pacific</i> → <i>regardez le secteur oriental de l' océan pacifique</i></p>
No Type (6.2.3)	<ul style="list-style-type: none"> - function words only necessary in one language; - segments not translated but they do not impact the meaning; - segments giving repeated information in context; - translated segments which do not correspond to any source segment. <p><i>minus 271 degrees , colder than</i> → <i>moins 271 degrés , ce qui est plus froid que</i></p>

Table 2: Definition and important rules for unaligned segments

2 General instructions

2.1 How to deal with minor misspelling and tokenization errors?

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 *ca* → *ça*, *a quel point* → *à quel point*, *l' endroit ou* → *l' endroit où*, etc.

During the annotation, please note down the sentence ID and the misspelled pair. The same for the word tokenization errors that you have found (e.g. *lorsqu'on* → *lorsqu' on*).

2.2 Three conditions of annotation

Annotators can watch corresponding videos of TED Talks before annotating to better understand the context of each sub-corpus, the links are provided in another file.

During the annotation, there will be three possible configurations:

- 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 phrase boundary.
- 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.

2.3 External resources

Annotators are encouraged to use language resources such as *Cambridge Dictionary*, *Larousse*, *Le Robert*, *TLFi*, etc. to consult word senses, widely used literal translations, the translation of multiword expressions and so on.

For example, in *Linguee*², we can see the relation between « *faint* » and « *tomber dans les pommes* », this can also help to decide the boundary:

²<https://www.linguee.com/>

if you *faint* easily → si vous *tombez dans les pommes* facilement (Category *Figurative* (subsection 6.1.8))

3 Decision helper

In order to facilitate the annotation task, please see figure 2 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.

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

During the annotation, there exist different categories concerning idioms and fixed expressions in the source or target language, the table 3 recapitulates them. Below is their definition.

Idiom: 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*.

Fixed 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 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

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.

5 Annotation conventions

5.1 How to decide the segment boundary?

In principle The segment boundary is not provided to annotators and it should be fixed by respecting the given tokenization, while excluding the part not involved (follow the bold part in this annotation guide):

have generated **sufficient** interest → ont suscité **suffisamment d'** intérêt

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

there is a measurable effect → il y a un effet mesurable

We should segment and align like this:

there is → il y a

a → un

measurable → mesurable

effect → effet

Negation

Don't believe everything she says . → Ne crois **pas** tout ce qu' elle te dira .

since it **did n't happen** here → comme ça **ne se passe pas** ici

we **did n't have** polio in this country yesterday → nous **n' avons pas** la polio dans ce pays hier

are **n't you afraid** → vous **n' avez pas peur de**

Articles and prepositions

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

1. it 's really **text** → c' est vraiment **du texte**

2. in other words , **sugar pills** have a measurable effect

→ en d' autres termes , **des pilules de sucre** ont des effets mesurables

3. you can just say a few names and **people** will understand

→ il suffit de citer quelques noms et **les gens** comprennent

- Here we align to you with vous, which are both indirect objects of the verb show and montrer.

I 'll show it **to you** → je vais **vous** le montrer

- Regroup the preposition with the verb that triggers its appearance:

we do n't **want to** encourage people to eat → on ne **veut pas** encourager les gens à manger

- In this example, lift off is aligned to soulevant. Adding the preposition off here changes the meaning of the verb, because lift alone means lever. Then the table is aligned to de la table, because de corresponds to an empty position in English (the notion of from is implicitly implied in English).

and he can bring new characters into the scene , just by **lifting** the Siftables **off the table** that have that character shown on them .

→ il peut amener de nouveaux personnages dans la scène , simplement en **soulevant de la table** les Siftables présentant ce personnage .

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. and the great indicator of that , of course , is **language loss**

→ et l' indicateur le plus fiable est bien sûr l' **extinction du langage** (Particularization, subsection 6.1.6)

2. **spend a large sum of money** → **dépenser massivement** (Transposition)

(We keep this group to clarify the very general meaning of « massivement ».)

3. **stamp a letter into it** → **avec une lettre en creux** (Transposition)

4. les Buddhists **still pursue** the breath of the Dharma

→ les Bouddhistes **continuent à rechercher** le souffle du Dharma (Transposition)

(still + verb → continuer à + verb is a pattern, which should be annotated as a group.)

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. For example:

stamp a letter into it → **avec une lettre en creux**

(*Transposition*, where *a letter* and *une lettre* is a pair of literal translation)

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 . → *Pas mal* . (*Equivalence*, subsection 6.1.2)

How long have you been here ? → *Quand êtes -vous arrivés ici?* (*Modulation*)

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	
For example, the translation 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 3, please annotate the four instances separately: *facing*, *from you*, *is just going to*, *out*.

in a moment when my hand moves from facing you to being away from you , this finger right here , my index finger is just going to shift from where it is , to a position pointing out like this .
au moment quand ma main s' écarte de vous pour s' éloigner , ce doigt juste ici , mon index , passe de là où il est à une position pointant comme ceci .

Figure 3: How to annotate unaligned reduced segments

5.5 How to deal with linguistic anaphora?

In our work we do not manually resolve the linguistic anaphora, for example, we leave the « *ma* » unaligned in the following pair:

it is also my great love and fascination → *c' est aussi mon grand amour et ma fascination*

Another two examples are shown in figure 4 and 5, see the figure caption for how to annotate.

In figure 6, we annotate the repeated preposition « *de* » after the conjunction word « *et* » as *Explicitation* (see subsection 6.2.1).

1	you know , one of the intense pleasures of travel and one of the delights of ethnographic research is the opportunity to live amongst those who have not forgotten the old ways , who still feel their past in the wind , touch it in stones polished by rain , taste it in the bitter leaves of plants .
	vous savez , un des plaisirs intenses du voyage et un des délices de la recherche ethnographique est la possibilité de vivre parmi ceux qui n' ont pas oublié les anciennes coutumes , qui ressentent encore leur passé souffler dans le vent , qui le touchent dans les pierres polies par la pluie , le dégustent dans les feuilles amères des plantes .

Figure 4: On target side, leave the second *qui* unaligned

11	now , together the myriad cultures of the world make up a web of spiritual life and cultural life that envelops the planet , and is as important to the well-being of the planet as indeed is the biological web of life that you know as a biosphere .
	aujourd'hui , les innombrables cultures dans le monde constituent un tissu de vie spirituelle et culturelle qui enveloppe la planète , et qui est aussi important pour le bien-être de la planète que l' est également le tissu biologique de la vie que vous connaissez en tant que biosphère .

Figure 5: On source side, leave the second *life* unaligned

561	are n't you afraid you 're going to keep writing for your whole life and you 're never again going to create a book that anybody in the world cares about at all , ever again ? "
	vous n' avez pas peur de passer votre vie à écrire et de ne jamais plus faire de livre qui intéresse qui que ce soit dans le monde , plus jamais ? "

Figure 6: Annotate the repeated preposition after the conjunction as *Explicitation*

6 Annotation in practice

We have already annotated 17 924 English tokens, below we will show the percentage of literally translated tokens, and the percentage of each other category among the non-literal cases.

6.1 Annotation of aligned segments

Literal translations

6.1.1 Literal

Percentage: 73.80%.

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

certain kinds of → *certain types de*

we all share the same → *nous partageons tous les mêmes*

facts are stubborn → *les faits sont têtus* (literal translation of an idiom, align word by word)

Rule 1 The literal translation also concerns lexical units but in multiword form:

1. *But we wonder whether it has gone too far .*
→ *Nous nous demandons cependant si cela n'est pas allé trop loin .*
2. *there are* → *il y a*
3. *this is* → *voici*
4. *hatpin* → *épingle à chapeau*
5. *the largest* → *le plus grand*
6. *ago* → *il y a*
7. *magic trick* → *tour de magie*

8. **NASA geospatial image** → **image géospatiale de la NASA**

(annotate as a group, useful to teach learners the composition rules in French when adding the preposition **de**)

9. **all of you** → **vous tous**

(annotate as a group: all of + pron → pron + tous)

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:

look forward to (?regarder en avant)³ → **avoir hâte de**

in other words (?en d'autres mots) → **en d'autres termes**

I give you my word . (?Je vous donne mon mot .) → **Je vous donne ma parole** .

experience a hallucination (?expérimenter une hallucination) → **avoir une hallucination**

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

go from **six and a half** to nine billion people → passer de **6,5** à 9 milliards d' êtres humains

Rule 4 Translating by using calque or anglicism is annotated as *Literal*:

the myths of the Inuit elders still **resonate with** meaning → les mythes des anciens Inuit **résonnent encore de sens**

(« résonner de sens » is not a natural expression here)

but one of them is **distinctly** worse than the other → mais l' une d' entre elles est **distinctement** pire que l' autre

Rule 5 Fixed modulation is annotated as *Literal*:

a **life jacket** → un **gilet de sauvetage**

(The means « sauvetage » 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:

that looks **kind of neat** → c' est pas mal

(The translation is not word-for-word, and the translation uses a negative form. The category should be *Modulation*.)

Borderline examples:

Sort of. → En **quelque sorte** .

(Borderline with *Equivalence*, but it should be annotated as *Literal*. If it's translated by « Peut-être. » or « Presque. », then it should be annotated as *Equivalence*.)

Non-literal translations

6.1.2 Equivalence

Percentage: 17.79% (of non-literal translations)

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).

if you 'll pardon the pun → **si vous me passez ce calembour**

sense each other → **se reconnaître entre eux**

³The question mark means this translation is word-for-word but actually is incorrect.

more than that → *plus encore*

right now he 's over Ohio → *là* il survole l' Ohio

are n't you afraid you 're never going to *be able to top that* ?

→ vous n' avez pas peur de ne jamais *réussir à faire mieux* ?

Rule 1 Non-literal translation of proverbs, idioms, or fixed expressions is annotated as *Equivalence*:

Birds of a feather flock together . → *Qui se ressemble s' assemble* .

like a bull in a china shop → *comme un chien dans un jeu de quilles*

on the brink of → *à deux doigts de*

Rule 2 Equivalence in context is annotated as *Equivalence*:

and we started talking about music , from Bach to Beethoven and Brahms , Bruckner , all the B 's , from Bartók , all the way up to Esa-Pekka Salonen .

→ *et nous avons commencé à parler de musique , de Bach à Beethoven , de Brahms , Bruckner , tous les B , de Bartók , jusqu' à Esa-Pekka Salonen .*

(need world knowledge, chronological sense)

Rule 3 Changing the measure into the one used in the target culture is annotated as *Equivalence*:

about a mile and a half deep → *vers 2500 m de profondeur*

is 20 inches → *est de 50 centimètres*

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

now they are metaphorically in the womb of the great mother

→ *ils sont maintenant dans le ventre de la grande mère , métaphoriquement parlant*

(The word "métaphoriquement" alone would be a literal translation, but "-ment + parlant" is a fixed expression, but it doesn't have a figurative meaning.)

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

UN → *Organisation des Nations unies*

IPCC → *Groupe d' experts intergouvernemental sur l' évolution du climat*

Counterexamples:

magic trick → *tour de magie*

(Word-for-word translation, the category is *Literal*.)

at no time → *à aucun moment*

(We can not say *à aucun temps* in French. This is a literal translation (see Rule 2 of *Literal*).)

Borderline examples:

which is → *soit*

(Borderline with *Literal*. Since it is not the most literal translation, it is annotated as *Equivalence*.)

that 's something the world needs right now → *c' est quelque chose dont le monde a besoin maintenant*

(Borderline with *Literal*, annotated as *Equivalence*.)

6.1.3 Transposition

Percentage: 14.81%

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:

people are suspicious → *les gens se méfient*

or locally on a term in a syntagm, which globally doesn't change the category, for example:

I said it as a joke → *J' ai dit ça pour plaisanter*

Below are some typical examples found in our corpus:

adv -> conjunction

there are **only** three fingers down here → il **n'** y a **que** trois doigts ici

it takes 142 pages **just** to print this genetic code

→ ça prendrait 142 pages **rien que** pour imprimer ce code génétique

verb -> prep

" what is life ? " is something that **I think** many biologists have been trying to understand

→ " qu' est -ce que la vie ? " est **selon moi** ce que beaucoup de biologistes ont cherché à comprendre

verb -> noun

unless **something changes** , they 're already dead

→ à moins qu' **un changement ait lieu** , elles sont déjà mortes

these two morphologically unrelated plants that **when combined** in this way

→ ces deux plantes sans aucun lien morphologique qui **lorsque mises en synergie** de cette façon

noun -> verb

and **that is the idea** that the world in which we live

→ et **cela veut dire** que le monde dans lequel nous vivons

the **computer vision** algorithms have registered these images

→ les algorithmes de **vision informatisée** ont enregistré ces images

noun -> adv

and , if anything , at a far **greater rate** . → et peut-être bien **plus rapidement** .

adj -> adv

have generated **sufficient** interest → ont suscité **suffisamment d'** intérêt

adj -> verb

even those of us **sympathetic with** the plight of indigenous people

→ même ceux d' entre nous qui **compatissons avec** les difficultés du peuple indigène

adj -> noun

we would say to be **friendly gestures** , → on pourrait qualifier de **témoignage d' amitié** ,

prep -> verb

patients **over** the age of 40 → les malades **ayant dépassé l'** âge de 40 ans

how we understand a lot of the world **around us** .

→ notre façon de comprendre une grande partie du monde **qui nous entoure** .

these were calling cards **from** the devil → c' était des cartes **venant du** diable

The « **chassé-croisé** » is a double transposition involving both a change of grammatical category and a syntactic permutation of the elements that constitute the meaning.

1. *all that we can be as an **astonishingly inquisitive** species .*

→ tout ce que nous pouvons être en tant qu' espèce **dotée d' une curiosité stupéfiante** .

2. *he **strode into** the house → il **entra à grands pas** dans la maison*

3. *so they **spearred** the five missionaries **to death***

→ ils ont donc **abattu** les cinq missionnaires **à coups de lance**

Counterexamples:

where they get mixed . → où elles se mélangent .

(The passive voice (« get + past participle » expresses a passive meaning) has been changed to active voice, the category should be *Modulation*.)

Borderline examples:

all of these peoples teach us that there are other ways of being

→ tous ces peuples nous enseignent qu' il y a d' autres façons d' être

(Here, "all" is a pronoun and "tous" is an adjective, but it is borderline with *Literal*. It is annotated as *Literal* now.)

they go , “ well , he ’s certainly not dumb enough to stab himself through the skin to entertain us for a few minutes .

→ ils font : “ il n’ est pas assez bête pour se planter des choses à travers la peau pour nous amuser quelques minutes .”

(Without preposition, "quelques temps" is not a direct object as in English. It is annotated as *Transposition* now.)

6.1.4 Modulation

Percentage: 16.03%

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

- can be encountered both in lexis as well as 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 Chuquet and Paillard (1989), there are mainly two types of modulation:

- metonymical modulations (the cause substituted for the effect, the container for the content, the part for the whole, etc.)
- 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. *Buy Coca-Cola by the carton* . → *Achetez Coca-Cola en gros* . (the abstract « en gros » substituted for the concrete « by the carton »)
2. *global warming pollution* → *pollution à effet de serre* (the means « à effet de serre » substituted for the result « global warming »)
3. *He shut the door in my face* . → *Il me claqua la porte au nez* . (the part « nez » substituted for the whole « face »)
4. Geographical modulation:
In French, *Holland* is often used to refer to the *Netherlands*, while the former is just the name of a region in the Netherlands.
By metonymy, we often refer to *United Kingdom* by *England* or *Great Britain*.

Change the point of view:

5. *How long have you been here ?* → *Quand êtes -vous arrivés ici ?*
6. *their hunters could smell animal urine at 40 paces and tell you what species left it behind* .
→ *leurs chasseurs pouvaient sentir l’ urine animale à 40 pas et vous dire de quelle espèce elle provenait* .
7. *what could be more lonely than to be enveloped in silence*
→ *comment ne pas se sentir seuls , enveloppés dans le silence*
8. *and that scar has stayed with him for his entire life*
→ *et que , toute sa vie , il a souffert de ce traumatisme*
9. *the beta-carbolines found within that liana* → *les bêta-carbolines dont est composée la liane*

Passive voice and active voice:

10. *where they get mixed* . → *où elles se mélangent* .
11. *the statistics you hear about* → *les statistiques qui nous sont communiquées*

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

12. *It's difficult* . → *Ce n'est pas facile* .

The subject becomes the object:

13. *He was knee-deep in water* . → *L'eau lui arrivait jusqu'aux genoux* .

14. *I had a really astonishing assignment* → *on m'avait confié une mission* étonnante

Syntax changes (no meaning change):

15. *something has happened* → *il m'est arrivé* quelque chose

16. *and you can go backwards, you can go forwards; you can not stay where you are* .
→ *vous pouvez aller en arrière ou en avant; vous ne pouvez pas rester où vous êtes* .

17. *those are two very different entities* → *ces deux entités sont très différentes*

Circumvent translation difficulties, achieve expression naturalness:

18. *unless you think of it in the terms that I do* , → *à moins que vous ne regardiez la chose comme moi* :

19. *which has caused me to have to recalibrate my whole relationship with this work*
→ *qui a transformé ma relation à ce travail*

Slight meaning change in lexical level:

20. *at no time can anything travel* , → *à aucun moment quoi que ce soit ne peut passer* .

21. *allowing people to more fully engage with their abilities* → *permettent aux gens de développer pleinement leur potentiel*

22. *I do n't expect that's going to change* → *je ne crois pas que cela va changer*

23. *remember the central revelation of anthropology*
→ *se rappeler de la révélation essentielle de l'anthropologie*

Counterexamples:

and saw the front page was like that → *et voyait la page d'accueil comme ça*
(The point of view does not change, the category should be *Equivalence*.)

Borderline examples:

we're looking at → *on a devant les yeux*
(Borderline with *Mod+Trans*, annotated as *Modulation* now.)

6.1.5 Mod+Trans

Percentage: 4%

This category combines the transformations of *Modulation* and *Transposition*, because the translation includes both changes in grammatical category and in syntactic structure or point of view. This often results in many-to-many alignment:

1. *this is a people who cognitively do not distinguish the color blue from the color green*
→ *c'est un peuple dont l'état des connaissances ne permet pas de faire la distinction entre la couleur bleue et verte*
2. *this is a completely unsustainable pattern* .
→ *il est absolument impossible de continuer sur cette tendance* .
3. *you can learn a great deal about deception* → *vous pouvez apprendre pas mal de choses sur l'illusion*
4. *it's going to be a stretch to do it for 9* . → *ça sera d'autant plus difficile de le faire pour 9* .
5. *so, our pace of digitizing life has been increasing at an exponential pace*
→ *bien que le rythme de notre numérisation de la vie se soit accru de manière exponentielle*
6. *an enzyme found naturally in the human gut* → *une enzyme se trouvant de façon naturelle dans l'intestin de l'homme*

(*found* → *se trouvant*: passive voice to active voice)

7. *The houses were all dark* . → *Pas une maison n' avait de lumière* .

8. *are n't you afraid that you 're going to work your whole life at this craft and nothing 's ever going to come of it and you 're going to die on a scrap heap of broken dreams with your mouth filled with bitter ash of failure ?*

→ *tu n' as pas peur de passer ta vie à écrire sans résultat et de mourir sur un tas de rêves brisés avec , dans la bouche , le goût amer des cendres de l' échec ?*

9. *you are taken by the spirit* → *l' esprit prend le contrôle*

Counterexamples:

that was done entirely computationally → *qui a été faite entièrement par calcul informatique*
(There is only *Transposition* used in this translation.)

Borderline examples:

brilliant , Juilliard-trained musician → *brillant musicien issu de Juilliard*
(Annotated as *Mod+Trans*.)

6.1.6 Particularization

Percentage: 7.71%

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.

they have a screen and a wireless radio → *ils sont équipés d' un écran et d' une radio sans fil*

this plant had in it some very powerful tryptamines

→ *cette plante contenait de très puissantes tryptamines*

the great indicator of that → *l' indicateur le plus fiable est*

because of someone 's perception of it → *grâce à la perception de quelqu' un*

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

1. *and the great indicator of that , of course , is language loss*

→ *et l' indicateur le plus fiable est bien sûr l' extinction du langage*

2. *live amongst those who have not forgotten the old ways*

→ *vivre parmi ceux qui n' ont pas oublié les anciennes coutumes*

3. *I 'll get my sleeve back* . → *je retrousse ma manche* .

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

if you 're queasy → *si vous ne supportez pas la vue du sang*

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

if you want to know how sea level rises from land-base ice melting this is where it reaches the sea .

→ *si vous voulez voir comment le niveau de la mer monte à cause de la fonte des glaces terrestres voilà l' endroit où la rivière se jette dans la mer* .

Counterexamples:

the weight of the marshmallow causes the entire structure to buckle and to collapse → *le poids du marshmallow fait que toute la structure se déforme et s' écroule*

(This is an instance of *Generalization*.)

Borderline examples:

you get to choose which one you want to use → *on peut choisir celle qu' on veut utiliser*

(It is not a very obvious example, it's annotated as *Particularization* now.)

6.1.7 Generalization

Percentage: 3.85%

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. *the most optimistic scenario in the realm of cultural diversity*
→ *le scénario le plus optimiste au sein de la diversité culturelle*
2. *and because you are possessed , you are taken by the spirit*
→ *et étant possédés , l' esprit prend le contrôle*
(remove the causality)
3. *as we sit here today in Monterey* → *alors que nous sommes à Monterey aujourd'hui*
4. *it 's held so tightly* → *il tient si bien*
(the notion of "tightly" is lost)

Rule 1 The translation of an idiom by a non-fixed expression is annotated as *Generalization*:
we use that great euphemism , " trial and error " , which is exposed to be meaningless .
→ *nous employons cet euphémisme , procéder par tâtonnements , qui est dénué de sens .*

Rule 2 The removal of a metaphorical image is annotated as *Generalization*:
ancient Tairona civilization which once carpeted the Caribbean coastal plain of Colombia
→ *anciennes civilisations tyranniques qui occupaient jadis la plaine côtière des Caraïbes de Colombie*

Rule 3 Use pronoun to translate the thing(s) that it references:
people in the back or people on video → *les gens du fond ou ceux qui regarderont la vidéo*
if you 're starting with digital information in the computer , that digital information has to be really accurate
→ *si vous partez des données numérisées sur ordinateur , il faut qu' elles soient extrêmement précises*

Counterexamples:

it does n't matter how much information we 're looking at → *ce quelle que soit la quantité d' informations que l' on visionne*

(This is an instance of *Particularization*.)

Borderline examples:

I 'm going to grab hold of my wrist → *je vais tenir mon poignet*

(It is not a very obvious example, it's annotated as *Generalization* now.)

let 's pretend right here we have a machine → *supposons que nous ayons ici une machine*

(Borderline with *Equivalence*.)

what 's interesting is the metaphor that defines the relationship between the individual and the natural world .

→ *la métaphore est intéressante , définissant la relation entre l' individu et le monde naturel .*

(Borderline with *Modulation*, syntactic change without changing the meaning.)

6.1.8 Figurative translation

Percentage: 0.57%

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

the number of pixels on your screen at any given moment

→ *le nombre de pixels sur votre écran à un instant " t "*

if you faint easily → *si vous tombez dans les pommes facilement*

and one woman almost passed out → *et une dame a presque tourné de l'oeil*

Rule 2 Keep the same metaphorical image by using a non-literal translation:

the Sun begins to bathe the slopes of the landscape → *le soleil qui inonde les flancs de ce paysage*

(Here we don't annotate it as *Équivalence*, because the non-literal translation not only keeps the same meaning, but also keeps the metaphorical image. We use the category *Figurative* to quantify this phenomenon.)

Counterexamples:

Borderline examples:

we can do a lot of things → *on peut faire un tas de choses*

(Borderline with *Équivalence*.)

6.1.9 Lexical shift

Percentage: 10.11%

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)

give you an update on how that machine worked

→ *vous mettre au courant de quelle façon cette machine fonctionne*

which might seem like an odd thing → *qui peut paraître bizarre*

change preposition

when you do a web search for images → *quand on fait une recherche web sur des images*

change determiner

if they believe enough there is a measurable effect in the body

→ *s' ils y croient assez fort , il y a un effet mesurable dans leur corps*

change subject

you can move them → *on peut les déplacer*

it's clear that we can make food → *il est clair que l' on peut faire de la nourriture*

change position adverb

there's a hole there → *il y a un trou ici*

change between singular and plural

sugar pills have a measurable effect in certain kinds of studies

→ *des pilules de sucre ont des effets mesurables dans certains types d' études*

(the boundary has been enlarged to keep the segment as a whole group)

change punctuation

I went through all the types of batteries that get made — for cars ,

→ *J' ai étudié tous les types de batteries qu' on fabrique , pour les voitures ,*

Counterexamples:

I've argued that we → *je pense que nous*

(The verb changes, not only the verbal tense, so it is a *Modulation (meaning changes in lexical level)*.)
at no time can anything travel , → *à aucun moment quoi que ce soit ne peut passer* .

(Same as above, this is a case of *Modulation (meaning changes in lexical level)*.)

6.1.10 Translation error

Percentage: 0.80%

Definition This category concerns obvious translation errors.

1. *the world in which we live does not exist in some absolute sense , but is just one **model** of reality*
→ *le monde dans lequel nous vivons n' existe pas dans un sens absolu , mais est uniquement un **exemple** de réalité*
2. *to this day , they remain ruled by a ritual **priesthood** but the training for the priesthood is rather extraordinary .*
→ *à ce jour , ils sont dirigés par un **clergé** rituel cependant , leur formation au clergé est plutôt extraordinaire .*
3. *there 's no one home anymore to **experience** a hallucination*
→ *il n' y a plus personne à la maison pour **faire** une hallucination*
4. *we found two **instances** of natural death* → *nous avons découvert deux **circonstances** de mort naturelle*
5. *is not going to be **remembered** for its wars* → *ne sera pas **reconnu** pour ses guerres*

6.1.11 Uncertain

Percentage: 3.05%

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.

- *this young man 's father had been **ascribed** to the Panchen Lama .*
→ *le père de ce jeune homme avait été **affecté** au Panchen Lama .*

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. (As a result, the annotation percentage is 0% at source side.)

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.

Rule 2 Introduce in the target language clarifications that remain implicit in the source language but emerge from the situation:

[...] live amongst those who have not forgotten the old ways, who still feel their past in the wind
→ *[...] vivre parmi ceux qui n'ont pas oublié les anciennes coutumes, qui ressentent encore leur passé souffler dans le vent*

Rule 3 Specify that the sentence comes from which source (e.g. video, another speaker, etc.).
the sun is rising . → ***vidéo** : le soleil se lève .*

Rule 4 Adding obligatory words to keep the sentence grammatically correct:

writing books is my profession → *écrire des livres* , *c' est mon métier*

Rule 5 Annotate an added preposition or a repeated preposition after the conjunction as *Explicitation*:

are n't you afraid you 're going to keep writing for your whole life and you 're never [...]

→ *vous n' avez pas peur de passer votre vie à écrire et de ne jamais plus [...]*

6.2.2 Unaligned - Reduction

Percentage: 6.83%

Definition Deliberately remove certain words with concrete meaning that could be translated:

1. *and you 'll suddenly discover what it would be like* → *et vous découvrirez ce que ce serait*
2. *they say we , who are the younger brothers , are the ones responsible for*
→ *ils disent , nous , les jeunes frères , sommes responsables de*
3. *look carefully at the area of the eastern Pacific* → *regardez le secteur oriental de l' océan pacifique*
4. *we 've never needed progress in science more than we need it right now*
→ *nous n' avons jamais eu autant besoin de la science que maintenant*
5. *We do want to help you .* → *Nous voulons t' aider .*

6.2.3 Unaligned - No Type

Percentage: 14.47% (on source side).

Rule 1 Function words necessary in one language but not in the other:

1. *the last example I have time to* → *le dernier exemple que j' ai le temps de*
2. *a sequence that you can assemble* → *une séquence que l' on peut assembler*
3. *minus 271 degrees , colder than* → *moins 271 degrés , ce qui est plus froid que*

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

1. *he 's going to land in a couple of hours , he 's going to rent a car , and he 's going to come to Long Beach*
→ *il va atterrir dans quelques heures , louer une voiture et arriver à Long Beach*
(The stylistic choice in English has been omitted in French)
2. *and everyone in this room has to get into it .* → *chacun de nous ici doit y monter .*

Rule 3 Target segments which do not correspond to any source segment:

Rule 4 Reformulation of the speaker: in figure 7, we don't annotate the reformulation which are incomplete segments.

1981
and neither of these stories is very inspiring or great -- but one of them is this distinct ...
et aucune de ces expériences n' était agréable ou motivante -- mais l' une d' entre elles peut se distinguer ...
1982
but one of them is distinctly worse than the other .
mais l' une d' entre elles est distinctement pire que l' autre .

Figure 7: Don't annotate the incomplete segment of reformulation

7 Tutorial for using Yawat

In order to conduct all the operations available in Yawat, please use Firefox (version 67 or before)⁴ 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 this 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 8, 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.

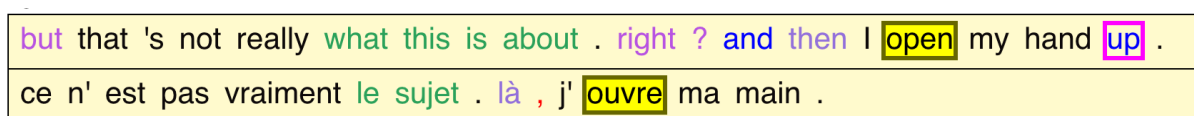


Figure 8: How to correct alignment in Yawat

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 9, 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.

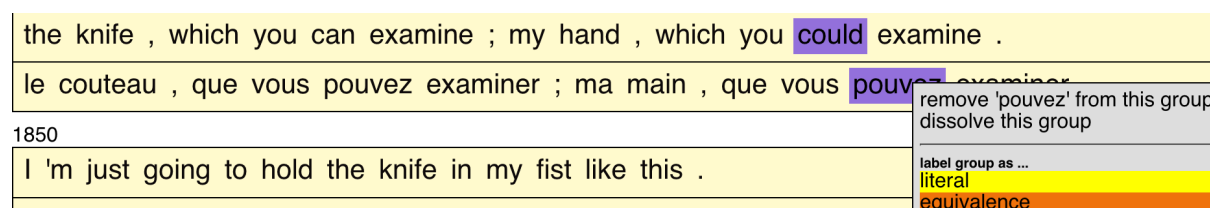


Figure 9: How to remove a word from a segment

5. To facilitate the alignment of a pair of long segments, for example in figure 10, first left click *nothing* and *il*, then right click either of them to create this pair.
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*.
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 11).

⁴If 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/>.

and to make sure nothing goes up or down my sleeve I 'm just going to squeeze my wrist right here .
pour être sûr qu' il n' y a rien dans mes manches , je vais juste serrer mon poignet ici .

Figure 10: How to align long segments in Yawat

in fact it 's held so tightly in place , and the knife does not come off .
en fait , il tient si bien que je ne peux pas le couteau ne tombe pas .

1857

nothing goes up or down my sleeve and you can examine everything

Figure 11: How to annotate unaligned segments

- 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 12).

and to make sure nothing goes up or down my sleeve I 'm just going to squeeze my wrist right here .
pour être sûr qu' il n' y a rien dans mes manches , je vais juste serrer mon poignet ici .

1853

that way you can see that at no time can anything touch the blade squeezing there nothing can

comme ça vous pouvez voir qu' à aucun moment qu' on ne peut pas passer . tant que je serre r

1854

and the object of this is quite simple .

l' objet de ce tour est assez simple .

Figure 12: Choose category from the menu

- Hover the mouse on words to check the boundary and the alignment of segments.
 - 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.
- 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.
- 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 // Pergamon Institute of English. State University of New York Press.
- Peter Newmark. 1988. *A Textbook of Translation*. English language teaching. Prentice-Hall International.
- Jean-Paul Vinay and Jean Darbelnet. 1958. *Stylistique comparée du français et de l'anglais: méthode de traduction*. Bibliothèque de stylistique comparée. Didier.