

## DELA Project – Annotation Guidelines

The goal of these guidelines is to set the rules that direct this annotation process. The issues tagged here are issues that would occur in a translation from English into Brazilian Portuguese when no context information is given. For example, if we give the following single sentence for a translator to translate:

*“And thanks for the case.”*

The translator would not be able to translate this with 100% certainty because:

1- we don't know the gender of the person who is saying 'thanks' as Portuguese differentiate between masculine or feminine.

2 - we don't know what the word “case” is as it has a few different meanings that would fit this sentence, i.e could be some type of protective box (a case for my phone, a case for my glasses), a woman's bag, a pencil case, a folder, a suitcase, or a police case to be investigated, each one with a different translation in Portuguese. Consequently, the translation of “for” will have a different gender depending on the meaning of the word “case”.

If a translator would be evaluating the translation of the same sentence given by 3 different MT systems:

GG: *“E obrigado pelo caso.”* (masculine, police case)

MS: *“E obrigado pelo estojo.”* (masculine, pencil case)

DeepL: *“E obrigado pela caixa.”* (masculine, box)

The translator would say that the translations of all 3 systems are correct as without more context, it is impossible to know the correct translation of the sentence, which should be:

HT: *E obrigada pela capa.* (feminine, phone case)

Therefore, the issues tagged in the corpus will be issues that might arise in the translation of sentences when the full context is not given.

### The issues tagged are:

- A. Gender
- B. Number
- C. Ellipsis
- D. Reference
- E. Lexical ambiguity
- F. Terminology

Issues that are solved within the single sentence are not tagged. For example:

*Sorry, I don't know if anybody is cutting class.*

Even though the word ‘cut’ has a different translation into Portuguese when meaning ‘cutting class’, the sentence itself gives enough information for a translator to be able to identify it.

## Format

The issues are tagged per sentence and in the order they appear (one per line). Each issue type requires an explanation of why this issue could be problematic in the translation:

text = source	issue	explanation	Translation (extra explanations) and notes
leave the url, title, and corpus where the text is from	gender	give the solution, "term(s)" = gender. ex. it = feminine	Ela
	number	give the solution: ex. "term(s)" = number. ex. surrender = plural	Entreguem-se
	ellipsis	give the solution: ex. "term(s)" = reference in text. ex. I do = I think	Eu acho.
	reference	give the solution: ex. it = support group	o grupo de suporte
	lexical ambiguity	give the meanings that could be confusing vs correct meaning, wrong vs right. ex party = group of political people vs celebration	o partido vs a festa
	terminology	give the meanings that could be confusing vs correct meaning, wrong vs right. wind farm = farm vs park	fazenda eólica vs parque eólico.

## Context-Related Issues

### GENDER

Gender is tagged:

- when the gender cannot be defined within the single sentence.
- when the translation is likely to be wrong. (i.e MT defaults everything to masculine)
- even when the most used translation is gender-neutral. (some synonyms might be gender specific)

### EXAMPLE

1 - *For that reason, a lot of what you do with this air fryer is going to have to be through trial and error.*

### NOTHING TO BE TAGGED.

There is nothing else connected to air fryer in this sentence that needs to be translated into the feminine. The translation of air fryer is feminine and because the word appears in the sentence, it is solved within the sentence that anything related to 'air fryer' needs to be in the feminine.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
For that reason, a lot of what you do with this air fryer is going to have to be through trial and error.			Por esse motivo, muito do que você fizer com esta fritadeira de ar comprimido terá que ser por tentativa e erro.

2 - *It is more of an olive green than a mint green.*

NO GENDER ISSUE TO BE TAGGED.

Even though the reference 'it' relates to 'chair' there is nothing else in the sentence that needs to be translated into feminine (chair). (Note that the reference issues is still tagged)

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
It is more of an olive green than a mint green.	reference	It = chair	

3 - *I'm so **glad** that **it** comes with the extender, so I have more levels to use to continue to get smaller.*

Here the gender is marked even when the most used translation of 'glad' in portuguese is the gender neutral 'feliz', because it can also be translated as 'satisfeita', 'grata', 'encantada', 'animada'

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
I'm so glad that it comes with the extender, so I have more levels to use to continue to get smaller.	gender reference	Glad = feminine It = waist cincher	Estou tão feliz/animada que ele vem com o extensor, então tenho mais níveis para usar para continuar a diminuir.

4 - **AWESOME!**

Here "awesome" is not just an expression, the gender is connected to the word as the sentence before asks: "what do I think of it?" referring to the air fryer which is translated into feminine in PT-BR.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
AWESOME!	gender	Awesome = feminine	Maravilhosa!

5- *She waited for a few minutes longer, but nothing happened, no one followed her, so she made her way back to the motel **they** were **staying at**.*

Here gender is tagged because the verb 'staying in' in PT-BR is translated as an adjective and needs to follow the gender.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
She waited for a few minutes longer, but nothing happened, no one followed her, so she made her way back to the motel they were staying at.	gender gender	They = feminine were staying at = feminine	Ela esperou mais alguns minutos, mas nada aconteceu, ninguém a seguiu, então ela voltou para o motel em que elas estavam hospedadas.

## NUMBER

Number is tagged:

- when it is not possible to identify in the sentence whether the word needs to be translated into plural. (generally the pronoun 'you' and some verbs)

### EXAMPLES:

1 - *A third wrote: "Don't be deceived!"*

NO NUMBER ISSUE TO BE TAGGED HERE.

Even though we cannot say if it is singular or plural just with the sentence, the standard is to translate into the singular, which is the case here. (Note that gender is tagged as this order was given to a female).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
A third wrote: "Don't be deceived!"	gender	deceived = feminine	Não seja enganada!/Não se engane!

2 - *Three years after their 2012 nuptials, the Gossip Girl alum and the Pokémon: Detective Pikachu star **welcomed** James.*

NO NUMBER ISSUE TO BE TAGGED HERE.

It is possible to identify within the sentence that 'welcomed' has to be in the plural as two people are named - the gossip Girl alum and the Pokemon star. (Note that gender is tagged as alum refers to a female).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
Three years after their 2012 nuptials, the Gossip Girl alum and the Pokémon: Detective Pikachu star welcomed James.	gender	Alum feminine	= Três anos após suas núpcias em 2012, a ex-aluna de Gossip Girl e a estrela de Pokémon: Detetive Pikachu deram as boas-vindas a James.

### 3 - *I was praying for **you**.*

Here, number is tagged because it is not possible to identify within this single sentence whether 'you' should be translated into plural (which it does) or not.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
I was praying for you.	number	You = plural	Eu estava rezando por vocês.

### 4 - ***You** should be more **agreeable**.*

Here, number is tagged because it is not possible to identify within this single sentence whether 'you' should be translated into the plural which it does. In consequence, the adjective has to be translated in the plural as well. (Note that gender is also tagged as agreeable refers to females).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
You should be more agreeable.	number number gender	you = plural agreeable = plural agreeable = feminine	Vocês deveriam ser mais simpáticas/agradáveis.

## ELLIPSIS

Ellipsis is tagged:

- if it cannot be solved within the single sentence.
- when the lack of information affects the translation.

### EXAMPLES:

1 - *Also, not once did I feel a blast of hot air like I **do** when taking things out of the oven.*

NOTHING TO BE TAGGED.

The ellipsis 'do' is solved in the sentence.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
Also, not once did I feel a blast of hot air like I do when taking things out of the oven.			Além disso, nenhuma vez senti uma lufada de ar quente como quando tiro coisas do forno.

2 - *In my laughter, I bellied out a “YES, I **do**!”*

Here ellipsis is tagged because it causes lexical ambiguity that cannot be solved within the sentence:

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
In my laughter, I bellied out a "YES, I do!!"	ellipsis lexical ambiguity	do = think do = make vs think	Sim, eu faço! vs Sim, eu acho!

3 - **Several more are planned** for the rest of the year, including *The Angry Beavers Seasons 1 & 2*, *Danny Phantom Season 1*, *Aaahh!! Real Monsters Season 1*, *Catdog Season 1 Part 1* and *The Wild Thornberrys Season 2 Part*

Here ellipsis is tagged because it causes gender issues that cannot be solved within the sentence

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
Several more are planned for the rest of the year, including <i>The Angry Beavers Seasons 1 &amp; 2</i> , <i>Danny Phantom Season 1</i> , <i>Aaahh!! Real Monsters Season 1</i> , <i>Catdog Season 1 Part 1</i> and <i>The Wild Thornberrys Season 2 Part 1</i> .	ellipsis gender	several more = releases several more are planned = feminine	Several more are planned... Várias outras estão planejadas...

## REFERENCE

Reference is tagged:

- when it is not possible to identify in the sentence who the pronouns 'it' or 'they' are.
- It should not be annotated when it has merely a grammatical function".

## EXAMPLES

1 - ***It** made Carrie worry if she had learned too well from Momma that way, making someone so much the centre of her world that they could suffocate from the weight of **it**, the way she'd sometimes felt about Momma.*

NOTHING TO BE TAGGED.

Even though there is an 'it', its reference is within the sentence "making someone the centre of her world". (Note that the first mention of the pronoun 'it' is an impersonal pronoun which describes or stands for a thing or verb or any nonliving thing but not for a person - we are not interested in those types as they do not carry any reference within the context).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
It made Carrie worry if she had learned too well from Momma that way, making someone so much the centre of her world that they could suffocate from the weight of it, the way she'd sometimes felt about Momma.			Carrie ficou preocupada por ter aprendido muito bem com mamãe dessa maneira, fazendo de alguém o centro de seu mundo a ponto de sufocar com o peso, como às vezes ela se sentia por mamãe.

## 2 - *It is understandable though since **it** was shipped from China.*

Here, reference was tagged because we cannot solve what 'it' is within the single sentence. Moreover, since 'it' refers to 'ship' which is masculine in Portuguese, no gender needed to be tagged. (Note that the first mention of the pronoun 'it' is an impersonal pronoun which describes or stands for a thing or verb or any nonliving thing but not for a person - we are not interested in those types as they do not carry any reference within the context).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
It is understandable though since it was shipped from China.	reference	it = ship	É compreensível pois (o navio) foi enviado da China

## 3 - ***It** shut.*

Here reference was tagged because we cannot solve what 'it' is within the single sentence. Moreover, since 'it' refers to 'door' which is feminine in Portuguese, the gender also needed to be tagged.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
It shut.	reference gender	it = door door = feminine	Ela/A porta fechou.

## 4 - ***They** actually hurt*

Here reference was tagged because we cannot solve who 'they' is within the single sentence. Moreover, since 'they' refers to 'choices' which is feminine in Portuguese, the gender also needed to be tagged.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
They actually hurt.	reference gender	they = choices they = feminine	Elas/ As escolhas realmente doem

## LEXICAL AMBIGUITY

Lexical ambiguity is tagged:

- when the word has more than one possible translation/use in general.
- when more than one meaning can be applied to the single sentence.

### EXAMPLE

1- *The material is soft and **feels** nice on my skin.*

NOTHING TO BE TAGGED.

Even though the MTs confuse 'feel' for 'the material's feeling is soft' vs 'the touch of material on my skin is soft', this ambiguity can be completely solved within the sentence.

Note: Nunca se traduziria, ou esperaria que a tradução fosse “o material se sente macio na minha pele”. Um tradutor sabe que é “o material é macio na minha pele.”

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
The material is soft and feels nice on my skin.			O material é macio na minha pele.

2 - *Set is originally 99 at local store, idk if I can say where I **got** mine or if it's allowed to give the real retail value price.*

NOTHING TO BE TAGGED.

Even though the MTs confuse 'got' for 'where I have mine' (instead of 'where I purchased mine'), this ambiguity can be completely solved within the sentence by human translators.

Note: nunca se traduziria, ou esperaria que a tradução fosse “eu não sei se eu posso dizer onde tenho o meu ou se é permitido dar o valor real.” Um tradutor sabe que é “eu não sei se eu posso dizer onde comprei/consegui o meu ou se é permitido dar o valor real”.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
Set is originally 99 at local store, idk if I can say where I got mine or if it's allowed to give the real retail value price.			O conjunto é originalmente 99 na loja local, não sei se posso dizer onde comprei o meu ou se é permitido fornecer o preço de valor real de varejo.

3 - *Khloe Kardashian's **cheating ex** Tristan Thompson leaves ANOTHER desperate comment on her Instagram pic after rumours he wants her back*

NOTHING TO BE TAGGED.



Even though the MTs confused the word order 'cheating ex' for 'Tristan is a former cheater and doesn't cheat anymore' (instead of 'Tristan is and an ex who is a cheater'), this ambiguity can be completely solved within the sentence.

Note: nunca se traduziria, ou esperaria que a tradução fosse “O ex traidor de Khloe Kardashian, Tristan Thompson, deixa OUTRO comentário desesperado na foto dela no Instagram após rumores de que ele a quer de volta.” Um tradutor sabe que é “Tristan Thompson, o ex-namorado infiel de Kloe Kardashian....”

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
Khloe Kardashian's cheating ex Tristan Thompson leaves ANOTHER desperate comment on her Instagram pic after rumours he wants her back			Tristan Thompson, ex-namorado infiel de Khloe Kardashian, deixa OUTRO comentário desesperado em sua foto do Instagram após rumores de que ele a quer de volta

4 - I really liked this **waffle maker** after receiving **it**.

NOTHING TO BE TAGGED.

Even though the MTs confused 'waffle maker' for 'the manufacturer of waffles' instead of 'the waffle machine', this ambiguity can be completely solved within the sentence. (Note that reference was also not tagged as 'it' is solved within the sentence)

Note: nunca se traduziria, ou esperaria que a tradução fosse “Eu realmente gostei desse fabricante de waffles depois que recebi.” Um tradutor sabe que é “Eu realmente gostei dessa máquina de waffles depois que recebi”.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
I really liked this waffle maker after receiving it.			Eu realmente gostei dessa máquina de waffles depois que a recebi.

5 - There is a learning curve on figuring out how to get **it** hooked on and **be** tight enough at the same time.

Here lexical ambiguity is tagged because we don't know what 'it' is, we cannot tell if 'be' is inherent or a transitory characteristic. (Note that reference is also tagged).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
There is a learning curve on figuring out how to get it hooked on and be tight enough at the same time.	reference lexical ambiguity	it = waist cincher be: transitory VS inherent	to be tight enough: Há uma curva de aprendizado para descobrir como

			engancha e ser apertado o suficiente ao mesmo tempo.
--	--	--	--

6 - *The first day I got **it**, I meant to wear **it** for an hour to try **it** and I ended up forgetting **it was on** and wore it for 4 hours.*

Here lexical ambiguity is tagged because as we don't know what 'it' is we cannot tell if 'it was on' is 'wearing' or 'connected to electricity' even though we have the verb 'wear' - we can wear things that are electric too! (Note that reference is also tagged).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
The first day I got it, I meant to wear it for an hour to try it and I ended up forgetting it was on and wore it for 4 hours.	reference lexical ambiguity	it = waist cincher it was on = connected to electricity vs to be worn	No primeiro dia que comprei, pretendia usar por uma hora para experimentar e acabei esquecendo que tinha colocado e usei por 4 horas.

7 - *He came back in the house and said “So you think this is funny?!” up the stairway at me and **I LOST IT.***

Here lexical ambiguity is tagged because we cannot tell if 'I lost it' refers to the expression 'lose control' or if in fact an object (it) was lost. (Note that there is no reference tagged as the pronoun 'it' does not have a previous reference in the context).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
He came back in the house and said “So you think this is funny?!” up the stairway at me and I LOST IT.	lexical ambiguity	lost it = lose something vs to lose control	Ele voltou para dentro de casa e disse “Então você acha isso engraçado ?!” subir a escada para mim e eu PERDI O CONTROLE..

8 - *Period.*

Here lexical ambiguity is tagged because the word 'period' has different meanings (period of time, menstruation, punctuation mark) and there is not enough information to clarify it for sure.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
Period.	lexical ambiguity	period = era or menstruation vs full stop	Ponto Final.

9 - *This was my first **cup** so i'm not sure what it's supposed to be like but i really wanted to like this **cup**.*

Here lexical ambiguity is tagged because the word 'cup' has different meanings and translations and there is not enough information within the single sentence to clarify this issue.

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
This was my first cup so i'm not sure what it's supposed to be like but i really wanted to like this cup.	lexical ambiguity	cup = glass/mug vs menstrual collector	Este foi o meu primeiro coletor, então não tenho certeza de como deveria ser, mas eu realmente queria gostar deste coletor.

## TERMINOLOGY

Terminology is tagged:

- when the word that has different meanings happen only for that specific type of domain.
- terminology is different from lexical ambiguity.

### EXAMPLES:

1 - *Her research group has been growing and she was the **head** of it.*

NO TERMINOLOGY ISSUE TO BE TAGGED HERE.

The word 'head' is a common word with different meanings in many different situations. In this case, it is just a case of lexical ambiguity. (Note that reference was not tagged for 'it' because it refers to the 'research group' within the single sentence).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
Her research group has been growing and she was the head of it.	lexical ambiguity	head = body part vs leader	Seu grupo de pesquisa estava crescendo e ela era a chefe dele.

2 - *The center will also conduct testing (power curve, mechanical loads, noise, and power quality) at its own experimental **wind farm**.*

Here, the terminology was tagged because the word "farm" in this context has a specific translation in Portuguese for this domain (wind energy).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
The center will also conduct testing (power curve, mechanical loads, noise, and power quality) at its own experimental wind farm.	terminology	wind farm = park	O centro também conduzirá testes (curva de potência, cargas mecânicas, ruído e qualidade de energia) em seu próprio parque eólico experimental.

3 - The bones are rigid and hold their form, but do not restrict your movement.

Here, the terminology was tagged because the word 'bones' in this context has a specific translation in portuguese for this domain (corsets). (Note that gender is also tagged because since 'bones' is translated into 'barbatanas' which is feminine)

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
The bones are rigid and hold their form, but do not restrict your movement.	terminology gender	bones = skeleton vs fin rigid = feminine	As barbatanas são rígidas e mantêm sua forma, mas não restringem seus movimentos.

4 - *NECK: The soft tissues and the **strap** muscles of the neck, aside from the previously described focal hemorrhages, exhibit no abnormalities.*

Here, the terminology was tagged because the word 'strap' in this context has a specific translation in Portuguese for this domain (medicine).

SENTENCE	ISSUE	EXPLANATION	TRANSLATION
NECK: The soft tissues and the strap muscles of the neck, aside from the previously described focal hemorrhages, exhibit no abnormalities.	terminology	strap = belt vs neck muscles	PESCOÇO: os tecidos moles e os músculos Supra e Infra-Hióideos, além das hemorragias focais descritas anteriormente, não apresentam anormalidades.

**Decision Tree:** The decision tree guides the annotation of the context-related issues.

