Annotation Guidelines

for English-Dutch Translation Quality Assessment

version 1.1

LT3 Technical Report – LT3 13-02[1]

Joke Daems, Lieve Macken, Sonia Vandepitte

joke.daems@ugent.be, lieve.macken@ugent.be, sonia.vandepitte@ugent.be

LT3- Language and Translation Technology Team

Faculty of Applied Language Studies

University College Ghent

September 13, 2013

Introduction

Assessing translation quality is a very complex task, and depending on the goal of the assessment, a different approach is needed. Though existing translation quality assessment (TQA) metrics are suitable for a general assessment or assessment within a specific company, they lack the granularity needed to compare different methods of translation and their respective translation problems. The translation situation is not always taken into account, the assessments are often subjective and they are often limited to the sentence level. In an attempt to solve these and other problems, we propose a two-step TQA approach with a fine-grained categorization of translation problems and user-defined error weights.

Categorization

The categories are divided into two main groups: adequacy and acceptability. Though some subcategories are suggested in this report, categories can easily be added or deleted to better suit certain assessment needs. Categories contain *possible* translation problems, but depending on the text type they will be considered to be errors or not.

Error weights

Rather than letting annotators judge a text as containing 'minor', 'major', or 'critical' errors, error weights are defined by the user. As such, the error weights can be adapted to the translation situation: for technical texts, for example, 'terminology' will receive a high error weight. The error weights can also be equal to zero, which can be useful for researchers or teachers interested in studying the translation characteristics, rather than errors. Hyperonymy, for example, is not always an error, but it can be interested to highlight cases of hyperonymy in order to identify differences between translators or translation methods (human translation vs. post-editing of machine translation).

Two-step

In a first step, annotators only receive the target text and they have to annotate the text for acceptability. By only giving them the target text, they can't be influenced by the source text in their acceptability assessment and they can also judge the general coherence of the text.

In the second step, annotators receive the source sentences alternated with their translations and they have to annotate the text for adequacy.

For optimal results, we recommend informing the annotators as thorough as possible on the goal of the translation and the assessment, so they know which problems to highlight. Provide them with the same information the translators had during the task (Did they receive certain

pictures or background information? Was there terminology involved?). If terminology is crucial, also provide the annotators with the glossary the translators were asked to adhere to, so they can judge the translation for adequacy.

<u>The technical report</u>

This technical report contains a possible classification of translation problems for the translation of texts from English to Dutch and guidelines on how to annotate these problems. It was applied in a pilot study on the translation of newspaper articles, which explains the content of the annotation guidelines. Annotating for acceptability and annotating for adequacy are seen as two different tasks, which is why the guidelines contain two sections with a similar introduction. The guidelines were intended for the annotators, which explains the tone.

For this study, we used the brat rapid annotation tool[2], and the guidelines also contain instructions on how to use the proposed categorization within this tool.

---

[2] http://brat.nlplab.org

Annotating acceptability

The texts that you are about to annotate are the results of a translation task from English to Dutch. To be able to judge the quality of the translations, they will be marked for two important aspects: adequacy and acceptability. Adequacy is concerned with the relationship between source text and target text, whereas acceptability is concerned with the target text and language. The goal of the current assignment is to annotate translations for acceptability. Adequacy will be dealt with separately. In order to allow you to focus on the requirements of the target text and language, without being distracted by the source text, you only receive the Dutch translation, without the English reference text.

Acceptability can be described as respecting the norms of the target language and culture. A good translation should read as a native Dutch text. This includes respecting the conventions of the language (grammar, lexicon, spelling) as well as respecting the conventions of the text structure (paragraph content and coherence) and the text type (a newspaper article requires a different style than a manual, for example).

As a reviser, it is your task to make sure the translation reaches publishable and acceptable quality. You do this by marking anything that does not follow the conventions of the Dutch language or that does not respect the demands of the text structure or the text type. To facilitate the task, you can use the brat rapid annotation tool. In this tool, the different categories (grammar & syntax; lexicon; spelling, typos & punctuation; style & register; coherence) are predefined, along with their most important subcategories.

In the following paragraphs, the basics of the tool are explained, followed by the annotation rules and an overview of all the (sub)categories.

Using the brat-tool

Hover over the word 'brat' at the top right hand corner to be able to select 'log in' and use your username and password to log in.

You are now ready to start annotating! Just double-click a word or click and drag to select smaller/larger pieces of text and the tool will give you an overview of the possible categories.
The categories and subcategories are listed below 'entity type'. You can expand or collapse the categories for a better overview, but make sure to always select a subcategory, not the main category!
Just click 'ok' when you're done or 'cancel' when you've selected a piece of text that you didn't want to select. To change an annotation, double-click the label above the word. You can change the category, subcategory and notes, you can decide to delete the annotation or you can move the annotation. To move an annotation, first select 'move' and then select the text span where you want the annotation to move to.

You can select a word or span more than once, so it is possible to assign different problem categories to the same word.

A special type of annotation: the empty span

Sometimes, you'll have to make an annotation that concerns the entire sentence instead of a single word or a short span of text. Rather than selecting the entire sentence in this case, you should insert an empty span at the beginning of the sentence.

An empty span is created by holding down the shift-key while you click.

Remember to only use it at the very beginning of a sentence (before the first word of the sentence) and to only use it for annotations that impact the entire sentence. The guidelines contain information on the cases in which you are allowed/required to use an empty span.

Linking spans

When you encounter a problem that concerns more than one word, but these words are split up by other words, you'll have to make sure that those non-relevant words are not contained in your span. You do this by selecting the two parts of the sentence that contain the problem separately (so you make two different annotations of the same category) and then you link them together with an arrow. You do this by clicking the first annotation and dragging your mouse pointer to the second annotation. You'll see an arrow appear with the words 'belongs_to'. The

guidelines contain information on when you are allowed or required to insert a link between spans.

General annotation rules

1) Always select a subcategory, not a main category.
2) Always specify why you made a certain annotation in the 'notes' section. You are allowed to use either English or Dutch for your comments.
3) If an item contains more than one problem, highlight the item as many times as there are problems, once for each problem (for example, a word that contains both a compound error and a capitalization error).
4) Only select 'other' if there is no other category that describes the problem better.
5) If you are not sure about your annotation, end your comment in the notes section with a double question mark '??' to indicate this insecurity.
6) If the same problem occurs more than once, select each occurence seperately.

When in doubt...

If you are not sure whether or not something is an error, don't be afraid to consult external sources. You are perfectly allowed to use a dictionary or a search engine to look things up. Some useful sites that you could consult are:

http://www.vandale.be/
http://taaladvies.net/
http://woordenlijst.org
http://www.vrt.be/taal/

You can refer to external sources in the 'notes' section to support your decisions. It's also allowed to look back to previous texts, to check how you annotated the same problem in a different translation.

Categorization

A detailed explanation of each subcategory can be found below the overview on this page. The information consists of the category name and colour in the brat-tool, followed by the full name, a definition, important remarks, guidelines for annotation and examples. The words that should be annotated are underlined (empty spans are indicated with an asterisk - *) and the information after the arrow sign is an example of a possible annotation note.

Grammar & syntax	Lexicon	Spelling & typos	Style & register	Coherence
Article	Wrong preposition	Capitalization	Register	Conjunction
Comparative/ superlative	Wrong collocation	Single-word spelling mistake	Untranslated	Missing info
Singular/plural	Word non-existent	Compound	Repetition	Logical problem
Verb form		Punctuation	Disfluent sentence/ construction	Paragraph
Article-noun agreement		Typo	Short sentences	Inconsistency
Noun-adjective agreement			Long sentence	Other
Subject-verb agreement			Text type	
Reference			Other	
Missing constituent/ preposition				
Superfluous word/ constituent				
Word order				
Structure				
Other				

<u>GramSyn (red) Grammar & syntax:</u>

Definition   'This does not follow the grammatical or structural[3] rules of the Dutch language.'

article

Definition   'There is an article that should not be there, or there is no article where there should be one.'

Be careful!   If you encounter a singular noun that either needs to receive an article or needs to be replaced by the plural version in order for it to be correct, also select this category.

- e.g.: ze kunnen schijnbaar terminale <u>ziekte</u> genezen

Be careful! (2)   If the article is incorrect for the noun, select 'article-noun agreement'.

Annotation   If the article is missing, select the noun it belongs to; if the article is superfluous, select the article itself. Specify the type of error in the notes-section.

- e.g.: <u>VN-milieuprogramma</u> waarschuwt → missing article, should be 'het VN-milieuprogramma waarschuwt'

comparative/superlative

Definition   'The structure or form of the comparative or superlative is incorrect.'

Annotation   Select the entire comparative/superlative.

- e.g.: <u>praktischte</u> → meest praktische
- e.g.: een <u>meer belangrijk</u> verschil is dat… → een belangrijker verschil

singular/plural

Definition   'A word that has no plural or singular has been given this form or an incorrect plural has been given.'

Annotation   Select the word.

- e.g.: Ik heb gisteren een <u>hersen</u> gegeten → 'hersenen' is always plural
- e.g.: Ze luistert graag <u>muzieken</u> → 'muziek' has no plural

verb form

Definition   'A grammatically incorrect tense or verb form.'

Annotation   If a whole verb form is incorrect, select the entire constituent (if the verbs are split up by non-verbs, select the two parts and use an arrow to link the verbs). Specify the type of error – tense or form - in the notes-section and only select the word(s) where the issue occurs.

---

[3] Though syntax/structure is often seen as synonymous to grammar or as a subset of grammar, we chose to add (near-)synonyms in these annotations rules, so that they were easy to understand by people with different linguistic backgrounds.

Be careful! Only use this category if the actual verb form is incorrect or does not exist. If the form itself is correct, but the spelling is wrong (for example 'zij', 'verhuizde') select 'spelling & typos – single-word spelling mistake). If the verb exist, but is not the one intended in this context, select 'Lexicon – wrong collocation' rather than 'verb form' (for example 'wijdt' for 'wijt' or 'opgeheven' for 'opgegeven').

- e.g.: zodat vorderingen van schijnbare genezing <u>te zien</u> in een geschikte context → verb form, should be 'gezien worden'
- e.g.: <u>Ontgind</u> → verb form, schould be 'ontgonnen'
- e.g.: ik <u>wordt</u> → verb form, should be 'ik word'

article-noun agreement

Definition 'Mismatch between article and noun.'

Annotation Select both the article and the noun and link them together with an arrow.

- e.g.: <u>de hoofd</u> → wrong article

noun-adjective agreement

Definition 'Mismatch between noun and adjective.'

Annotation Select the adjective and the noun separately and link the annotations together with an arrow.

- e.g. een <u>slimme meisje</u> → een slim meisje

subject-verb agreement

Definition 'The subject and verb differ in number.'

Annotation Select the subject and the verb separately and link the annotations together with an arrow.

- e.g.: onze <u>planeet</u>...<u>ondergaan</u> → 'planeet' is singular, verb should be singular as well

reference

Definition 'Mismatch between referring expression and referent'

Annotation Select the referent and the referring expression separately and link the annotations together with an arrow. In case of a compound, select the entire compound (even if the compound is incorrectly split up).

- e.g.: Ze heeft een <u>contact verbod</u> aangevraagd <u>die</u> hem verplicht om... → een contactverbod dat

superfluous – superfluous word or constituent

Definition 'The sentence is grammatically incorrect because there is a superfluous constituent or word. This can be a constituent that already appeared earlier in the sentence or a word that has been written twice in a row.'

Annotation   Select the superfluous word or constituent. If the word appears twice and either one of them can be deleted to make the sentence grammatically correct, select the first occurrence of the word.

- e.g.: Ze heeft het <u>verteld</u> aan de politie verteld.
- e.g.: <u>de</u> de hond

missing - missing constituent or preposition

Definition   'The sentence is grammatically incorrect because a necessary part of the structure is missing, this can be a preposition or an entire constituent such as an obligatory direct object.'

Be careful!   Missing articles should be highlighted as 'article', not as 'missing constituent or preposition'.

Be careful! (2)   Only select 'missing constituent' if a whole constituent or preposition is missing. If part of a verb form is missing (but not the whole verb), select 'grammar & syntax – verb form', not 'missing constituent'.

Annotation   If a preposition is missing, select the noun phrase following the missing preposition, even if the preposition is part of a verb with a fixed preposition. If an essential subject or object is missing, select the verb. If an entire constituent that does not belong to either a noun or a verb is missing, insert an empty span at the beginning of the sentence and specify the missing element.

- e.g.: Hij begon <u>het verzenden</u> van berichten → 'hij begon met het verzenden'
- e.g.: Ze abonneerde zich <u>het nieuwe tijdschrift</u> →'op het nieuwe tijdschrift'

word order

Definition   'This word order is grammatically incorrect.'

Be careful!   If the word order is grammatically correct, but another word order would be better (more natural), select 'style & register – disfluent', not 'grammar & syntax – word order'.

Annotation   If there are only two words or parts of the sentence that should switch places, select these parts separately and link them together with an arrow. If something more complex is wrong with the word order, insert an empty span at the beginning of the sentence and specify the incorrect and correct word order in the 'notes' section.

- e.g.: Ten derde <u>klimaatverandering kan</u> de temperatuur verhogen→ Inversion after 'ten derde'

structure - other structural problems

Definition   'There is something wrong with the structure of the entire sentence that is not just the cause of an incorrect verb form, lack of agreement or word order.'

Annotation   Insert an empty span at the beginning of the sentence and specify the problem in the 'notes' section.

- e.g.: *Wouden gaan verloren; wegen en dammen worden aangelegd; de populatie neemt toe: natuurlijke habitatten worden weggedaan voor landbouw; mijnbouw; en vervuiling van zeewater bevorderen omstandigheden waaronder nieuwe en oude pathogenen kunnen bloeien → different structures in the same sentence

gram_other – other grammatical / syntactical problems

Choose this category when you encounter a grammatical / syntactical error that does not belong to any of the abovementioned categories. Always explain why you choose this category. It is possible to link two annotations together with an arrow if they are separated by non-relevant words.

<u>Lexic (yellow) Lexical problems</u>

Definition   'This is a lexical problem/error when looking at the Dutch language.'

wrong preposition

Definition   'This expression requires a different preposition.'

Annotation   Select the preposition.

- e.g.: de ziekte komt voor <u>in</u> vliegende honden → bij vliegende honden
- e.g.: Mevrouw Nestler <u>van</u> Santa Fe → uit Santa Fe
- e.g.: ze was begonnen <u>om</u> berichten te verzenden → met berichten te verzenden

wrong collocation

Definition   'The word(s) exist(s) in Dutch, but is/are used in the wrong or in a strange way: uncommon combinations of words, errors in fixed expressions,…'

Be careful!   If the combination of words is lexically correct, but not logical in the context, select 'coherence – logical problem', not 'lexicon – wrong collocation'.

Annotation   Select the word or constituent that is used in a wrong way. If there is a problem with a fixed expression, select the entire expression (prepositions included!). If an inappropriate verb or adjective is used with a noun, select the entire verb or adjective. If the expression or collocation is split up by non-relevant words, select the two parts of the collocation separately and link them together with an arrow.

- e.g.: de vergadering <u>gaat door</u> om 5u → should be: vindt plaats
- e.g.: veranderingen in het milieu zullen een stijging van besmettelijke ziektes <u>tot stand brengen</u> → should be: met zich meebrengen
- e.g.: <u>in diepe nesten zitten</u> → should be: diep in de nesten zitten
- e.g.: De temperaturen zorgen ervoor dat pathogenen kunnen <u>bloeien</u>. → 'bloeien' is something that can't be said about 'pathogenen'

word non-existent

Definition   'This word does not exist in Dutch'

Be careful!   If the non-existent word is the result of incorrect grammar, select the appropriate category in the 'grammar & syntax' section, not 'Lexicon – word non-existent'.

- e.g.: ontgind → non-existent verb form of 'ontgonnen', belongs to grammar & syntax – verb form

Annotation   Select the word. If the word is a verb that is split up by non-relevant words, select the two parts separately and connect them with an arrow.

- e.g.: <u>Plantaties</u>→ should be: plantages
- e.g.: <u>Bosverlies</u> → should be: ontginning van bossen

<u>Spel_Typ (blue) Spelling, typos and punctuation</u>

Definition   'This does not follow the rules of spelling and punctuation of the Dutch language or this is a typo.'

Be careful!   If there is more than one type of spelling mistake in one word, select the word twice: once for each type. For example, when a compound is split up and spelled with a capital letter, you select the word once for 'capitalization' and once for 'compound'.

capitalization

Definition   'This does not follow the Dutch language rules of capitalization.'

Annotation   Select the <u>whole</u> word.

- e.g. <u>afrika</u> → Afrika

spelling_mistake - single-word spelling mistake

Definition   'General spelling mistake: wrong spelling of a single word.'

Annotation   Choose this category when you encounter an error other than 'capitalization' in a word that is not part of a compound. Always explain why you choose this category. Select the entire word.

- e.g.: <u>advokaat</u> → advocaat
- e.g.: <u>financiëel</u> → financieel

compound

Definition   'This compound is misspelled: there is a space between two elements of a one-word compound, there is a space between the first part of the compound and the hyphen, there is a superfluous hyphen,… A compound can be a noun as well as an adjective, a verb or a preposition.'

Annotation   Select the entire compound, unless the second part of the compound does not follow directly after the first, in which case you only select the first part.

- e.g.: <u>anti-semitische</u> → antisemitische
- e.g.: <u>groei -</u> en leerlijnen → no space between first part of a compound and hyphen.

punctuation

Definition   'The punctuation of the sentence is wrong or missing.'

Be careful!   If a punctuation mark within a word or compound is used in the wrong way, select 'single-word spelling mistake' or 'compound', not 'punctuation'. e.g.: de jaren <u>'90</u> → single-word spelling mistake

- e.g.: <u>brood-rooster</u> → compound

Be careful! (2)   However, if the punctuation mark is a quotation mark, you should select 'punctuation' and select the entire word, not just the punctuation mark.

- e.g.: '<u>Wonder'middeltjes</u> → quotation marks should be around the entire word.

Annotation Select the punctuation mark. If the punctuation mark is missing, insert an empty span at the beginning of the sentence. If there are quotation marks at the beginning or end of a sentence/constituent, but not on the other side, select the quotation mark that is there, and specify the problem. If both single and double quotation marks are used, select the first quotation mark and specify the problem. If the wrong quotation marks are used on both sides, select the opening quotation mark.

- e.g.: full stop after title, semicolon between two words rather than two main clauses,…
- e.g.: Hij zei: <u>"</u>Ik heb het niet gedaan. → missing quotation mark at end of quote.
- e.g.: Ze heeft het probleem <u>'</u>grondig" bestudeerd. → Both single and double quotation marks are used (better use single quotation marks for irony)
- e.g.: Ze heeft het probleem <u>"</u>grondig" bestudeerd. →use single quotation marks to express irony.

typo

Definition 'This error was probably caused by typing too fast: letters have been switched, there is a letter missing or too many, there's a superfluous or missing space,…'

Be careful! Do not interpret spelling mistakes as typos. Only select 'typo' if the error cannot be classified elsewhere, otherwise select 'spelling, typos & punctuation – single-word spelling mistake' or the appropriate category.

Annotation Select the entire word, not just the few letters that need to be changed. If a word has been typed twice, select the first time it occurs. If there is a space before a punctuation mark, select the punctuation mark.

- e.g.: <u>dennken</u> → denken
- e.g.: De vis zwom in het water <u>.</u> → no space before a punctuation mark

Style_reg (pink) Style and register

register

Definition 'The words have the same meaning, but the chosen word/expression is too formal/informal/... for the text.'

Annotation Select the word(s) or expression from the inappropriate register. If they are split up by non-relevant words, select the two parts separately and connect them with an arrow.

- e.g.: <u>dat wijf</u> → too informal / vulgar, should be 'die vrouw' in this text

untranslated

Definition 'A word / fragment of which a Dutch translation exists is left untranslated.'

Annotation Select all the untranslated words.

- e.g.: De <u>UN</u> heeft ons gewaarschuwd voor de opwarming van de aarde → 'UN' has a Dutch translation: 'VN'.

repetition

Definition 'The same or a very similar word/expression is used too often or is too close to the previous occurrence of the word/expression, it's better to use a synonym.'

Annotation Select the first occurrence of this word/expression and specify why/where it has been used too often in the notes. If the expression or constituent is split up by non-relevant words, select the two parts separately and connect them with an arrow.

disfluent - disfluent sentence/construction

Definition 'The sentence / constituent is not grammatically incorrect, but it is nonetheless very difficult to read, it could be translated in a much more idiomatic way.'

Annotation Insert an empty span at the beginning of the sentence and specify the problem.

short sentence

Definition 'This segment contains too many short sentences, which affects the readability.'

Annotation Insert an empty span at the beginning of the first short sentence.

long_sentence

Definition 'This sentence is too long, it would benefit the readability if this sentence were split up.'

Annotation Insert an empty span at the beginning of the sentence.

text type

Definition 'This is not necessarily a problem, but the text type allows or requires different constructions or deletions.'

Annotation If the issue is contained in one word or constituent, select that word/constituent. If the constituent is split up by non-relevant words, select the two parts of the

constituent separately and connect them with an arrow. If the issue affects an entire sentence, insert an empty span at the beginning of the sentence where the issue occurs.

- e.g.: *Tijdens een boswandeling hebben de scouts een schat gevonden → no articles in the title of a newspaper article, 'scouts vinden schat tijdens boswandeling' is better
- e.g.: '<u>Meneer</u> Letterman zei dat hij…' → 'Letterman zei dat hij…': usually only the last name is used in a Dutch newspaper article

style_other – other style & register problems

Choose this category when you encounter a style/ register problem that does not belong to any of the abovementioned categories. Always explain why you choose this category. It is possible to link two annotations together with an arrow if they are separated by non-relevant words.

Coher (orange) Coherence

Definition   'There is something wrong with the coherence of the text: confusing relationships, lack of logical structure, undeveloped paragraphs,…'

conjunction

Definition   'The conjunction or linking word expresses a strange relationship or there seems to be a missing relationship.'

Annotation   If there is a conjunction/linking word, select this word. If the conjunction or linking word is missing, insert an empty span at the beginning of the sentence and specify the problem in the notes.

missing info

Definition   'Information that is needed to easily understand the text is missing, thus reducing readability. This includes cases of implicitation.'

Annotation   Select the constituent that is affected by the missing info. If the constituent is split up by non-relevant words, select the two parts of the constituent separately and link them together with an arrow. If the entire sentence is affected, insert an empty span at the beginning of the sentence and specify the problem in the notes.

logical problem

Definition   'This does not follow the logical structure of the text. The idea contradicts something that has been previously stated, or the information as such is illogical/confusing when looking at the rest of the text.'

Annotation   If the logical problem is situated in one word or constituent, select this word/constituent. If the constituent is split up by non-relevant words, select the two parts of the constituent separately and link them together with an arrow. If the problem affects the entire sentence, insert an empty span at the beginning of the sentence and specify the problem in the notes.

paragraph

Definition   'The paragraph is not well-developed, contains more than one idea or the information belongs to a previous paragraph.'

Annotation   Insert an empty span at the beginning of the paragraph and specify the problem(s) in the notes.

inconsistency

Definition   'Terms or notations are used inconsistently throughout the text.'

Annotation   Select the first inconsistent occurrence of the word and specify the variant + amount of times they occur. If it is a problem contained in one word, use the following construct: "(word1): (number of occurrences)x, (word2): (number of

occurrences)x, (comment)". If it is a different type of problem, specify the problem in the notes-section.

Be careful!   If the capitalization of words is inconsistent throughout the text, insert an empty span at the beginning of the very first sentence of the text and specify the problem in the notes-section.

- e.g.: Deze ziektes zijn dodelijk voor de mens. (…) Andere besmettelijke <u>ziekten</u> kunnen genezen worden door…

  → ziektes: 3x, ziekten: 4x, a different plural is used throughout the text

coh_other – other coherence problems

Choose this category when you encounter a coherence problem that does not belong to any of the abovementioned categories. Always explain why you choose this category. It is possible to link two annotations together with an arrow if they are separated by non-relevant words.

Annotating adequacy

The texts that you are about to annotate are the results of a translation task from English to Dutch. To be able to judge the quality of the translations, they will be marked for two important aspects: adequacy and acceptability. Adequacy is concerned with the relationship between source text and target text, whereas acceptability is concerned with the target text and language. The goal of the current assignment is to annotate translations for adequacy. Acceptability will be dealt with separately. The texts are presented to you sentence per sentence. First, the original English sentence is given, followed by the Dutch translation.

Adequacy can be described as making sure the target text contains the same information as the source text. This means that all misinterpretations, contradictions, meaning shifts, additions or deletions are potential errors. Depending on the text type, a translator could be allowed to delete some details or to add a little extra information, but these are general possible translation problems. In a technical manual or medical document, for example, strict adherence to the source text information is required.

As a reviser, it is your task to make sure the translation reaches publishable quality based on information adequacy. You do this by marking all types of meaning shift, while taking the demands of the text type into account. Although the texts may contain errors against the Dutch language, this is not your concern, unless the errors also bring about a shift in meaning (in which case you are required to mark the shift). To facilitate the task, you can use the brat rapid annotation tool. In this tool, the different types of meaning shift are predefined.

In the following paragraphs, the basics of the tool are explained, followed by the annotation rules and an overview of all the categories.

Using the brat-tool

Hover over the word 'brat' at the top right hand corner to be able to select 'log in' and use your username and password to log in.

You are now ready to start annotating! Just double-click a word or click and drag to select smaller/larger pieces of text and the tool will give you an overview of the possible categories. Just click 'ok' when you're done or 'cancel' when you've selected a piece of text that you didn't want to select. To change an annotation, double-click the label above the word. You can change the category and notes, you can decide to delete the annotation or you can move the annotation. To move an annotation, first select 'move' and then select the text span where you want the annotation to move to.

You can select a word or span more than once, so it is possible to assign different problem categories to the same word.

A special type of annotation: the empty span

Sometimes, you'll have to make an annotation that concerns the entire sentence instead of a single word or a short span of text. Rather than selecting the entire sentence in this case, you should insert an empty span at the beginning of the sentence.

An empty span is created by holding down the shift-key while you click.

Remember to only use it at the very beginning of a sentence (before the first word of the sentence) and to only use it for annotations that impact the entire sentence. The guidelines contain information on the cases in which you are allowed/required to use an empty span.

<u>Linking spans</u>

When you encounter a problem that concerns more than one word, but these words are split up by other words, you'll have to make sure that those non-relevant words are not contained in your span. You do this by selecting the two parts of the sentence that contain the problem separately (so you make two different annotations of the same category) and then you link them together with an arrow. You do this by clicking the first annotation and dragging your mouse pointer to the second annotation. You'll see an arrow appear with the words 'belongs_to'. The guidelines contain information on when you are allowed or required to insert a link between spans.

General annotation rules

1) Annotations are made on the Dutch target text, unless otherwise specified.
2) Always specify why you made a certain annotation in the 'notes' section. You are allowed to use either English or Dutch for your comments.
3) If an item contains more than one problem, highlight the item as many times as there are problems, once for each problem.
4) Only select 'other' if there is no other category that describes the problem better.
5) If you are not sure about your annotation, end your comment in the notes section with a double question mark '??' to indicate this insecurity.
6) If the exact same problem occurs multiple times in one sentence or throughout the text, select each item separately.

If you are not sure whether or not something is an error, don't be afraid to consult external sources. You are perfectly allowed to use a dictionary or a search engine to look things up. You can refer to external sources in the 'notes' section to support your decisions. It's also allowed to look back to previous texts to verify how you annotated a specific problem.

<u>Categorization</u>

A detailed explanation of each category can be found below the overview on this page. The information consists of the category name and colour in the brat-tool, followed by the full name, a definition, important remarks, guidelines for annotation and examples. The words that should be annotated are underlined (empty spans are indicated with an asterisk - *) and the information after the arrow sign is an example of a possible annotation note.

Meaning shift
Contradiction
Word sense disambiguation
Hyponymy
Hyperonymy
Terminology
Quantity
Time
Meaning shift caused by punctuation
Meaning shift caused by misplaced word
Deletion
Addition
Explicitation
Coherence
Inconsistent terminology
Other meaning shift

Contradiction (pastel red)

Definition   'The target text contradicts the source text.'

Annotation   Select the entire constituent / section of the sentence that contains the contradiction. If the contradiction is split up by non-relevant words, select the two parts separately and link them together with an arrow.

- e.g.:

  EN: The clearing of natural habitats for agriculture causes climate change

  NL: <u>Het verdwijnen van landbouwgebieden</u> veroorzaakt klimaatverandering

  → In the source text, the habitats disappear so that there can be agriculture, in the target text it is the agriculture that disappears.

Word_sense (pastel blue) - word sense disambiguation

Definition   'The Dutch word is a possible translation of the word in the ST, but not of the meaning the word has in this context.'

Annotation   Select the word(s). If the translation is split up by non-relevant words, select the two parts separately and link them together with an arrow.

- e.g.:

  EN: The frequency of this phenomenon should be appreciated so that claims of apparent cure by novel treatment strategies can be seen in an appropriate context.

  NL: … zodat <u>vorderingen</u> van schijnbare genezing in de juiste context gezien kunnen worden

  →'claim' can mean 'vordering' in a legal context, but here 'bewering' is meant.

- e.g.:

  EN: Climate change aggravates the threats of infectious diseases by further stressing habitats

  NL: Klimaatverandering verhoogt het risico op besmettelijke ziektes door meer <u>de nadruk te leggen op</u> habitats

  →'stress' can mean 'nadruk leggen op', but here 'onder druk zetten' is meant.

Hyponymy (pastel yellow)

Definition   'The target text contains a hyponym of the word/expression used in the source text, even though it could have been translated in an equally specific way.

Annotation   Select the hyponym.

- e.g.:

  EN: he sent her love messages

  NL: hij stuurde haar <u>liefdesbrieven</u>

→ 'love letters' is a hyponym of 'love messages' and there was no mention of 'letters' in the text

Hyperonymy (orange)

Definition 'The translation contains a superordinate term or expression even though it could have been translated in an equally specific/explicit way.'

Annotation Select the hypernym.

- e.g.:

  EN: infectious diseases

  NL: ziektes

- e.g.:

  EN: herbal potions

  NL: kruidengeneesmiddelen

  → potions are liquids, 'geneesmiddelen' is a hypernym

Terminology (dark green)

Definition 'Non-adherence to the specified terminology guidelines: The source text word was an entry in the glossary, yet the translator chose a different word as a translation. '

Annotation Select the term.

Be careful! If the translation is an incorrect translation of a word that was present in the glossary, annotate it both as a 'wrong word sense' or 'other' problem (depending on the nature of the incorrect translation) and as a 'terminology' problem.

Be careful! (2) If the translation is one of the possible translations of the word according to the glossary, but not the one needed in this context, select 'wrong word sense', not terminology.

Be careful! (3) If the translation is problematic when looking at the glossary, but it is also inconsistent when looking at previous occurences of the same source text term, select both 'terminology' and 'inconsistent terminology'.

- e.g.:

  EN: application

  NL: applicatie

  Glossary: toepassing

  → 'toepassing' is the appropriate term to use here, according to the glossary.

Quantity (purple)

Definition 'The number or amount mentioned in the target text is different from that in the source text.'

Annotation   If the number is different, select the word. If the amount is different, select the quantifying element

- e.g.:
  EN: I brought the books back to the library
  NL: Ik bracht het <u>boek</u> terug naar de bibliotheek
- e.g.:
  EN: He paid her 500 euros.
  NL: Hij betaalde haar <u>50</u> euro.

Time (beige)

Definition   'The time of the target text is different from the one in the source text, which changes the meaning. This can be caused by an incorrect verb tense or temporal element.'

Be careful!   Only select this category if there is a difference in meaning between source text and target text, not if the verb is grammatically incorrect.

Annotation   Select the verb or the temporal element. If the verb is split up by non-relevant elements, select the two parts separately and connect them with an arrow.

- e.g.:
  EN: We receive a lot of requests
  NL: We <u>kregen</u> veel verzoeken
  → The translation makes it sound as if it is something from the past, whereas the present tense in the source text indicates that the statement is still true.
- e.g.:
  EN: He always drinks coffee in the evening
  NL: Hij drinkt <u>'s ochtends</u> altijd koffie

Punctuation (light pink)- meaning shift caused by punctuation

Definition   'The punctuation mark alters the meaning as expressed in the source text.'

Annotation   Select the punctuation mark and specify the different meaning. If the punctuation mark is missing, insert an empty span at the beginning of the sentence and specify the change in meaning.

- e.g.: Exclamation/statement/question

Misplaced_word (grey) - meaning shift caused by misplaced word

Definition   'The element in the target text modifies a different word/constituent than the one in the source text, thus changing the meaning.'

Annotation   Select the entire constituent affected by the misplaced word or constituent and specify the problem in the notes.

- e.g.:

  EN: the reproduction of the handwritten report

  NL: <u>de handgeschreven weergave van het rapport</u>

  → the report is handwritten, not the reproduction

Deletion (red)

Definition   'A meaningful element of the source text or relationship present in the source text is deleted in the target text.'

Be careful!   Cases of hyperonymy must be annotated as such, not as cases of deletion.

Annotation   Select the elements that are deleted in the <u>English</u> source text. If two parts of deleted information are split up by non-relevant words, select the two parts separately and link them together with an arrow. If the deletion is more complex, insert an empty span at the beginning of the sentence and specify the deletion in the 'notes' section.

Addition (sea green)

Definition   'A meaningful element that cannot be derived from the source text is added in the target text.'

Annotation   Select the added information. If the addition is split up by non-relevant words, select the two elements separately and link them together with an arrow.

Explicitation (cyan)

Definition   'Information that can be derived from the source text but that is not necessary for the reader to understand the information presented in the text is added to the translation.'

Annotation   Select the entire constituent / fragment that contains the information. If the fragment is split up by non-relevant words, select the two parts separately and connect them with an arrow.

- e.g.:

  EN: The United Nations Environment Programme has warned that changes to the environment bring about a rise in diseases.

  NL: Het VN-milieuprogramma <u>(UNEP: United Nations Environment Programme)</u> heeft gewaarschuwd dat milieuveranderingen een stijging van ziektes met zich meebrengen

  → The English name does not give the reader extra information that he/she needs in order to understand what 'VN-milieuprogramma' is, and no meaningful elements are added, as 'UNEP' and 'VN-milieuprogramma' refer to the same organisation.

Coherence (yellow)

Definition   'The conjunction expresses a different relation in the target text than in the source text.'

Annotation   Select the conjunction and specify the difference in relation.

- e.g.:

  EN: new diseases arise as environments destroyed

  NL: nieuwe ziektes komen op <u>terwijl/wanneer</u> omgevingen worden kapotgemaakt

  → shift of causal to temporal relationship

- e.g.:

  EN: the judge granted the request, noting the restraining order was granted merely as a matter of proper pleading

  NL: de rechter keurde het verzoek goed, <u>maar</u> wees erop dat...

  → the judge simply gave an explanation, there is no relationship of contrast in the original text.

Inconsistent_terminology (bright purple)

Definition: 'The same term was used throughout the source text, but the translator uses different terms throughout the target text.'

Annotation: Select the first inconsistent occurrence of the term and specify the problem in the notes-section.

Other - other meaning shift

Choose this category when you encounter a meaning shift that does not belong to any of the abovementioned categories. Always explain why you choose this category. If two elements of the meaning shift are split up by non-relevant elements, select the two parts separately and connect them with an arrow. If the problem affects the entire sentence, insert an empty span at the beginning of the sentence and specify the problem in the notes.