

Manual Text Simplification of the TICO-19 Dataset

You are given **sentences** that need to be rewritten so that they use **simpler language** and can be **understood by the general public** (non-specialists). This means that you should reduce the number of difficult words or idioms, simplify complex phrasing, delete information that may not be relevant, and make the sentence more straight forward. This could be accomplished by applying different transformations to the original sentences, such as lexical paraphrasing, compression, sentence splitting, etc. We explain some of these transformations below and provide examples.

- Lexical Paraphrasing

This involves changing complex words or phrases for simpler synonyms. For example:

Original	Simplification
From its inception , it was designated a duty-free port and vied with the neighboring Sultanate of Pattani for trade.	From the start , it was chosen to be a duty-free port to compete with the neighboring Sultanate of Pattani for trade.

*the bold facing here is for ease of readability and will not be used in the task.

When paraphrasing, try to preserve as much of the original meaning as possible. Proper names (e.g. John, Microsoft, Apple) and geographical locations (e.g. Northern Europe, Sultanate of Pattani) should be kept whenever possible. Due to the domain of the texts you will be simplifying, it is possible that some specialized technical terms appear. (e.g. polymorphism, electro-pneumatic). In that case, you should attempt to find a simpler synonym word or phrase using external resources. However, if you are unsure about the substitution, leave the word as-is and submit a comment explaining your doubt. You should use simpler concepts/words/phrases in your paraphrasing. This means your sentence could potentially have more words than the original.

- Compression

This involves deleting information that may not be relevant to understand the main idea of the sentence. For example:

Original	Simplification
Think of all the ways everyone in your household will benefit from your membership in Audubon.	Everyone in your household will benefit from membership in Audubon.
Mr Usta was examined by Dr Raymond Crockett, a Harley Street physician specialising in kidney disease.	Mr Usta was examined by Dr Raymond Crockett, a physician specialising in kidney disease.

*the bold facing here is for ease of readability and will not be used in the task.

When compressing, it is important that the resulting simplification does not alter the core meaning of the original sentence. It's ok if the resulting meaning is a subset of the original meaning as long as the lost information is of low-relevance to the key concept in the sentence. When deleting some content, it is possible to make other small changes to keep the resulting sentence grammatical and fluent.

- Sentence Splitting

This involves splitting a sentence as much as possible into several sentences in places where such a splitting could make the sentence simpler. The resulting sentences, taken together, should mean the same as the original sentence, be grammatical and fluent. For example:

Original	Simplification
He was involved in two conversions which turned out to be crucial.	He was involved in two conversions. These conversions turned out to be crucial.
The federal government suspended sales of U.S. savings bonds because Congress hasn't lifted the ceiling on government debt.	Congress hasn't lifted the ceiling on government debt. So, the federal government suspended sales of U.S. savings bonds.

As in the examples shown, when splitting, it is also possible to make other small changes to keep the resulting text grammatical. Also, if no splitting is possible in a given sentence, leave the sentence as it is and focus on other possible changes. The examples above only show a few possible ways in which a sentence could be split. Feel free to try out other ways.

- Putting it all together

All the transformations explained before should be applied whenever possible to simplify the given original sentences. We provide an example where all three transformations were performed:

Original	Simplification
If you are under the age of 18, you are required to complete at least 65 hours of behind-the-wheel skill-building including 10 hours of nighttime driving.	If you are under 18, you must complete 65 hours of practice driving. This must include at least 10 hours of driving at night.

The transformations performed were:

Original	Simplification	Transformation
<i>under the age of 18</i>	<i>under 18</i>	compression
<i>behind-the-wheel skill-building</i>	<i>practice driving</i>	paraphrasing
<i>nighttime driving</i>	<i>driving at night</i>	paraphrasing

<i>you are required to</i>	<i>you must</i>	paraphrasing
<i>including</i>	<i>. This must include</i>	splitting

*this table is just for illustration purposes and is not part of the annotation you should submit.

The text transformations described above are the ones that professional editors use the most. However, **other types of transformations can also be performed**, such as changing sentences from passive to active voice. As such, it is possible for you to apply other transformations that you think are necessary. In such cases, we would appreciate it if you keep a record so that we can incorporate them into the guidelines if necessary.

Finally, refrain from making complete rewrites of the original sentences. The simplified sentence should be kept as similar to the original (i.e. similar lexis, etc.) wherever possible, with changes only being made where necessary. This is because highly edited sentences may be more difficult for automatic models to learn from.

Additional Resources

Here is a list of web pages with information on Plain Language that may be helpful to read to complement the guidelines provided.

English

- [How to write in Plain English](#)
- [How to write medical information in plain English](#)
- [COVID-19 in Plain Language](#)
- [Easy Read information from Public Health England](#)
- [Plain Language Commission style guide](#)
- [How to write clearly \(Publications Office of the EU\)](#)
- [Before and after \(Examples from plainlanguage.gov\)](#)

Spanish

- [Guía de Lenguaje Claro para Servidores Públicos de Colombia](#) (Sección 4.3)
- [Cómo redactar textos de salud que todos entiendan \(2\): estilo y lenguaje](#)
- [COVID-19 en un lenguaje claro](#)