



The UNDL Foundation has released a version in UNL of “Le Petit Prince” (The Little Prince), the famous novella by Antoine de Saint-Exupéry, published in 1943. The corpus is available under an Attribution Share Alike (CC-BY-SA) Creative Commons license at the [UNL arium](#), and may be used for researchers and developers interested in semantic annotation of natural language texts.

**W**

### What is UNL?

The UNL is a knowledge representation language that has been used for several different tasks in natural language engineering, such as machine translation, multilingual document generation, summarization, information retrieval and semantic reasoning. It has been originally proposed by the Institute of Advanced Studies of the United Nations University, in Tokyo, and has been currently promoted by the UNDL Foundation, in Geneva, Switzerland, under a mandate of the United Nations. [ [read more about UNL](#) ]

### Why *Le Petit Prince*?

*Le Petit Prince* is one of the best-selling books ever (more than 80 million copies), and has been translated to more than 180 languages, providing thus the possibility of contrasting and evaluating a wide range of UNL-based translations. Additionally, the text offers the chance of experimenting UNL in three situations that have not been explored so often: French original, narrative and literature. Our main goal is to “UNL-plicate” the text in at least three different directions: replication, summarization and simplification, in as many languages as possible. [ [read more about UNLplication](#) ]

]

### How the text was UNLized?

The integral version of *Le Petit Prince*, which has been released under public domain in Canada, was obtained from [http://wikilivres.info/wiki/Le Petit Prince](http://wikilivres.info/wiki/Le_Petit_Prince). The whole text comprises 15,513 word forms (tokens) and 1,684 sentences. The UNLization of the text was carried out in a fully-manual way through the UNL Editor, a graph-based authoring tool developed by the UNDL Foundation. The sentences have been divided into two main different groups: a) the training corpus, which comprises the first 53 sentences of the book (dedication and first chapter), including the title; and b) the application corpus, which comprises the remaining 1,548 sentences. The training corpus was addressed collectively by a group of four human UNLizers in order to synchronize and normalize the UNLization strategies. The application corpus was organized according to the similarity of sentences (and not to the order of appearance) and was addressed from December 2009 to February 2010 according to the guidelines resulting from the training exercise (and which are available at [http://www.unlweb.net/wiki/index.php/UNLization Guidelines](http://www.unlweb.net/wiki/index.php/UNLization_Guidelines) ).

).

### Further information

## Le Petit Prince in UNL

Written by Ronaldo Martins

Monday, 15 March 2010 07:43 - Last Updated Monday, 15 March 2010 08:21

---

For further information, please contact

[Ronaldo MARTINS](#)

Language Resources Manager

UNDL Foundation

48, route de Chancy, CH-1213, Petit-Lancy, Geneva, Switzerland

+41 22 879 8090

### **About**

The UNDL Foundation is a non-profit organization based in Geneva, Switzerland, which has received, from the United Nations, the mandate for implementing the Universal Networking Language (UNL). The UNL Programme is a collaborative effort to create natural language resources and technology to reduce language barriers and strengthen cross-cultural communication in the framework of the United Nations. Participation in the Programme is free and open to individuals and institutions, either as researchers or as developers. Special funds are available for some languages.