

# Luxembourgish machine translation

*M.T. Carrasco Benitez, 12 March 2017, version 1*

## Overview

This is a proposal for creating a Luxembourgish machine translation that could be ready before the end of 2017. It is a very practical, concrete and immediate component that could be part of the long-term plan to promote the Luxembourgish language.

## Governance

Assuming that the proposal is accepted, the first step must be that the Government of Luxembourg creates the *Lëtzebuenger Iwwersetzung Büro* (LIB) with the mandate and resources to implement and maintain this initiative. The legal form is secondary: similar to the *Lëtzebuenger Online Dictionnaire* [LOD], an ASBL, etc. LIB must provide a free public service.

## Kickstarting

One should use existing components. It is recommended to use Apertium [AP], a free and open software developed by the firm Prompsit [PRO].

To get quickly the know-how, LIB must sign a contract with Prompsit. A rough cost estimation is in the range of 10k to 100k euros, according to the services requested. LIB can contact directly Prompsit or I can be an introducer: I had previous contacts with Prompsit as organiser of the *Luxembourg Apertium Workshop* [LAW]. I do not have any commercial linking to Prompsit and I will not intervene in any contractual negotiations or accept payment from any party. Though, I am willing to help in this project.

Prompsit will do the preparatory work: preparation of languages, server installation, interfaces, training, hackathon [HA], and wiki similar to the one for the Luxembourg Apertium Workshop [LAWW]. All the results should be open source.

The *Official Journal of the European Union* [OJ] and the *Journal officiel du Grand-Duché de Luxembourg* [JO] could be translated into Luxembourgish; previous publications can be done in bulk; the new publications daily as they become available. Other relevant publications could be added.

## Languages

Aim for making Luxembourgish the language with the most translating pairs in Apertium. Start with the low hanging fruits: the first objective will be making a very high quality translation pair from German to Luxembourgish. The other languages in LOD can follow, in both directions: to and from Luxembourgish. The next languages as per the best return on investments (ROI) and available resources, particularly human resources.

## Luxembourgish Apertium Hackathon

It can be organised the weekend of the 16-17 September 2017, open to all comers. Prompsit will be in charge of the tutorials. The main objective should be the most useful languages that Prompsit manages to prepare programmatically for Apertium and ready for human refinements. English is the language of the event.

Participants willing to work in other languages should be allowed: Luxembourg has a large population of professional linguists, though amateurs should also be allowed to participate. As an incentive, participants should be credited toward the Luxembourgish language requirements needed

to obtain the Luxembourgish nationality.

The translating of European Union legislation in bulk should be part of this event, particularly from German as it will probably be the best quality pair.

## **IT**

There should be at least the following interfaces:

- Command line interface (CLI)
- Direct URI
- Document upload
- Web form

## **URI**

*Uniform Resource Identifier* [URI] must be short and mnemonic; for example, one should avoid subdomain such as LIB.LOD.lu, even if LIB comes under the supervision of LOD.

The domain LEI.lu has been registered for this initiative: *Lëtzebuenger Iwwersetzung*. It will be passed free to LIB. The domain LIB.lu is not available.

Examples:

<code>http://lei.lu</code>	home page
<code>http://de-lb.lei.lu</code>	German-Luxembourgish translation - form
<code>http://de-lb.lei.lu/Morgen</code>	German-Luxembourgish translation - data in URI

## **Machine translation technologies**

One should use *rule-based machine translation* (RBMT). *Statistical machine translation* (SMT) should not be considered as this requires large parallel corpora between Luxembourgish and other languages.

Apertium is RBMT. It seems there is already some preliminary work for Luxembourgish [APL]. To add Luxembourgish to Apertium, language pairs in each direction will have to be implemented. For example, Luxembourgish<->German.

## **Aligned corpora**

Luxembourgish lacks large aligned corpora. Translating in bulk from other languages with high quality machine translation will facilitate the creation of aligned corpora, a mandatory resource for other machine translation technologies such as SMT and *neural machine translation* (NMT).

N-language alignment refers to aligning all the languages as one record, as opposed to bilingual alignment. Creating this type of corpora will be a very valuable resource for Luxembourgish and other languages.

## **Status of this document**

This is a draft personal proposal. Mentioned entities do not necessarily back the proposal. New more complete versions will follow if interest is demonstrated by the stakeholders. The style used is to propose concrete names, dates, etc that can be changed. The latest version is at `http://lei.lu`

## Author

Manuel Tomas CARRASCO BENITEZ  
ca@dragoman..org  
Telephone: +352 691 72 04 78

## References

[AP] Apertium

<https://www.apertium.org>

[APL] Apertium Luxembourgish resources

<http://wiki.apertium.org/wiki/Apertium-deu-ltz>

<http://wiki.apertium.org/wiki/Luxembourgish>

[HA] Hackathon

<https://en.wikipedia.org/wiki/Hackathon>

[JO] Journal officiel du Grand-Duché de Luxembourg

<http://legilux.lu>

[LAW] Luxembourg Apertium Workshop, May 2011

<http://big.prompsit.com/apertium.eu/law>

[LAWM] Luxembourg Apertium Workshop - manual

<http://big.prompsit.com/apertium.eu/law-manual.pdf>

[LAWW] Luxembourg Apertium Workshop - wiki

[http://big.prompsit.com/wiki.apertium.eu/index.php/Main\\_Page](http://big.prompsit.com/wiki.apertium.eu/index.php/Main_Page)

[LOD] Lëtzebuenger Online Dictionnaire

<http://lod.lu>

[LEI] Lëtzebuenger Iwwersetzung - reserved domain

<http://lei.lu>

[PRO] Promsit

<http://www.prompsit.com>

[OJ] Official Journal of the European Union

<http://eur-lex.europa.eu/oj/direct-access.html>

[URI] Uniform Resource Identifier (URI): Generic Syntax

<https://tools.ietf.org/html/rfc3986>