

The First Dictionaries in Esperanto. Towards the Creation of a Parallel Corpus



Denis Eckert

Géographie-Cités / Centre

Marc Bloch

CNRS, Paris / Berlin

eckert@parisgeo.cnrs.fr



Francesca Frontini

Istituto di Linguistica

Computazionale - A. Zampolli

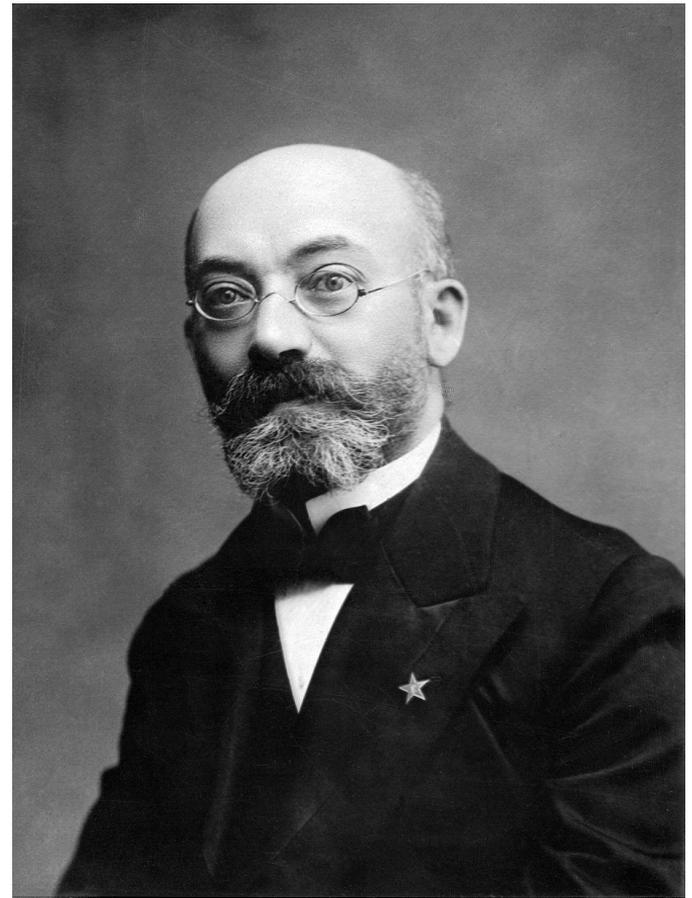
CNR - Pisa

francesca.frontini@ilc.cnr.it

Mid-**1887** in Warsaw an anonymous **Dr Esperanto** published a brochure presenting a ***Lingvo internacia*** - the international language.

Ludwik Lazar Zamenhof (1859-1917), was an ophthalmologist, born in Białystok to a Jewish family of intellectuals

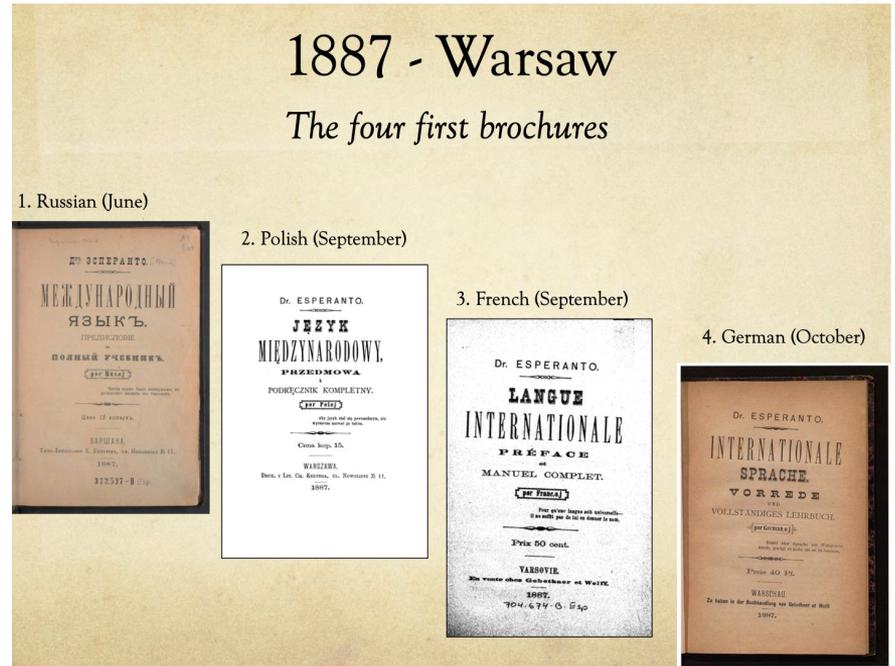
Esperanto is today **the only constructed language that has not disappeared** several years after its birth.



The early years of Esperanto: the “First Book”

Within five months in 1887, the first brochure presenting the new language was published in four main European languages (Russian, Polish, French, and German).

By the end of 1890 textbooks and booklets were available in **Russian, Polish, French, German, English, Hebrew, Yiddish, Swedish, Spanish, Latvian, Romanian, Danish, Bulgarian, Italian, Czech, Lithuanian.**



The early years of Esperanto constitute a remarkable moment of creation of a universalist contributing project in the age of 'First Globalization' (Berger, 2003).

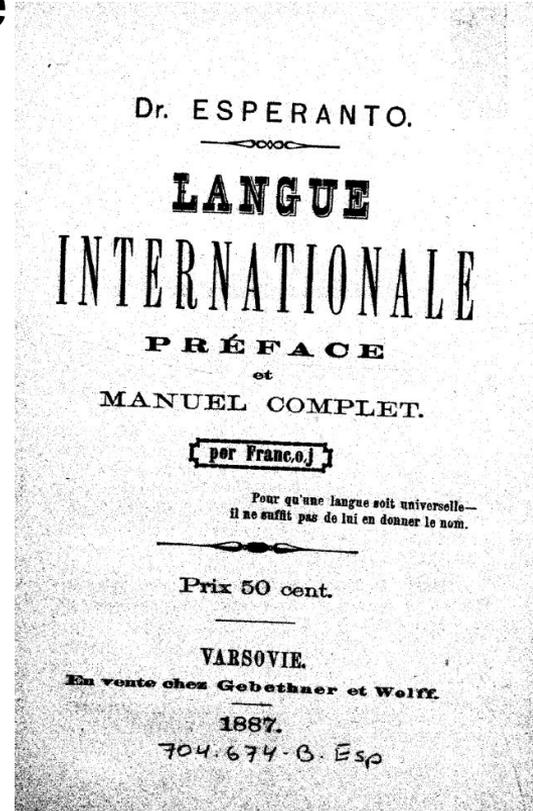
First Book (*Unua Libro*) - structure

The first publications have a similar structure:

- Aims and means of the project
- Basic rules and grammar
- A small dictionary - *vortaro*

The dictionaries all contain the same list of **Esperanto roots** (with small variations, 914 - 920 entries).

- Prepared by non specialist and early adopters, non coordinated



The *vortaroj*

This corpus of **early dictionaries** constitute an original and interesting **parallel corpus**, which has so far neither been put together nor analyzed.

Interesting for a great variety of scholars

(historians, linguists, digital humanists, cultural studies scholars, etc.), especially the global community of Esperantists.

A resource to study both the **evolution of Esperanto and of the various target languages**, e.g. previous stages of orthography (e.g. old style Russian, Hebrew, Baltic languages) and different scripts (Latvian).

Russian	1887	digitised-csv
Polish	1887	digitised-csv
French	1887	digitised-csv
German	1887	digitised-csv
English v1	1888	digitised-csv
English v2	1889	digitised-csv
Hebrew	1888	image format
English v3	1889	digitised-csv
Yiddish	1888	digitised-csv
Spanish	1889	digitised-csv
Swedish	1889	digitised-csv
Latvian	1889	image format
Romanian	1889	image forma
Bulgarian	1890	image format
Italian	1890	digitised-csv
Czech	1890	digitised-csv
Lithuanian	1890	image format

List of the early editions containing dictionaries.

The project timeline



Document collection: **Electronic facsimiles** of all publications containing a dictionary have been **collected** from *Department of Planned Languages* of the Österreichische Nationalbibliothek in Vienna, and other libraries in Europe, the USA and Israel.



OCR: **twelve *vortaroj*** (in 10 different languages) have been fully digitized and are available in .csv format.



Encoding: conversion into TEI, using the lexicographic module TEI-lex0. Currently, a proposal of **TEI modelling** has been made for two ***vortaroj*** (**French and German**).



Publication: make the digital editions corpora available for research.



Collaboration with CLARIN

- finding and digitizing further editions,
- establishing partnerships and collaborations for the encoding of specific languages and editions,
- consolidating the choice of data model and annotations,
- providing visibility to the corpus (which will be made available via a CLARIN center) and encouraging linguistic analysis,
- offering guidance as to copyright issues.

Furthermore, help creating of an international multidisciplinary network, and proceed to digitizing other early texts.

Dankon pro via atento

Further info at:

<https://halshs.archives-ouvertes.fr/halshs-02555912v1>
hal.archives-ouvertes.fr/ESPERANTO-HISTORIO