The First Mini-Dictionaries in Esperanto (1887-1890). An Original Multilingual Corpus for Comparative and Interdisciplinary Research

Denis Eckert

The first mini-dictionaries in Esperanto (1887-1890)

An original multilingual corpus for comparative and interdisciplinary research

Denis Eckert, Centre National de la Recherche Scientifique
UMR Géographie-Cités, Paris-Aubervilliers / Centre Marc Bloch, Berlin
eckert@parisgeo.cnrs.fr

Keywords

Esperanto - Dictionaries - Digital Humanities - Translation - Interlinguistics - Parallel Corpuses

Abstract

During the first years of existence of the *esperanto* language (1887-1892) , many different texts have been published by its promoters. Among them, published separately or included in texbooks or presentation brochures of the *lingvo internacia*, , 37 documents in at least 16 different languages were prepared by Zamenhof himself or by early adopters of Esperanto, in order to present the new "international language" to the broadest possible public. This unique multilingual corpus is scattered across many national libraries. We have systematically collected digital versions of this set of documents and begun to analyse them in parallel. Many of them (17) contain the same basic dictionary, elaborated by people who mostly had a limited, or absolutely no knowledge of philology. These 17 dictionaries encompass about 920 entries in Esperanto, each time translated in a given target language. We are progressively digitizing the whole corpus of these small dictionaries (so far, 12 versions have been digitized and encoded) and aim at making it accessible to scholars of various disciplines in an Open Access repository.

INTRODUCTION

During the first years of existence of Esperanto, (1887-1890), its creator, the Warsaw ophtalmologist Ludwik Zamenhof, as well as his early followers, published many texts to promote and put in practice the ideal of a *lingvo internacia*, a neutral language for international communication (Korzhenkov, 2009)[1].

Though Esperanto, as the new language was soon named, never became the universal standard for international communication it was meant to be, it is still in use worldwide. Esperanto is the only constructed language that has not disappeared several years after its birth, as *Pasilingua, Volapük*, *Ido*, *Langue bleue* and many other similar projects did (Haupenthal, 2013). The esperantists, albeit few in number, continue to promote the ideal of a neutral language for international communication: the Universala Esperanto-Asocio (UEA) is on the list of non-governmental organizations officially recognized by the United Nations, and many books and texts have been written or translated - and published - in Esperanto during the 133 years of existence of the language.

We intend to make a contribution to one of the less studied aspects of the early history of Esperanto (approx. 1887-1890), by documenting the diffusion of the first descriptions of the new language in natural (mostly European) languages. After having gathered electronic facsimiles of many of the first documents published by the movement, we could identify a specific corpus: the first Esperanto-target languages dictionaries. Dictionaries were obviously a strategic tool for the diffusion of the new language: Zamenhof published the very first one (around 920 entries) together with the first brochure released in June 1887. This small unidirectional (Esperanto-Russian) dictionary had, curiously, a long career, despite its limited scope. It was adapted, translated, or reprinted 17 times between 1887 and 1890[2], and made available in 15 different target languages.

---

1. A more comprehensive biography of Zamenhof, containing a detailed description of the first years of the Esperantist movement was published in Esperanto by the same author (Korzhenkov, 2011)
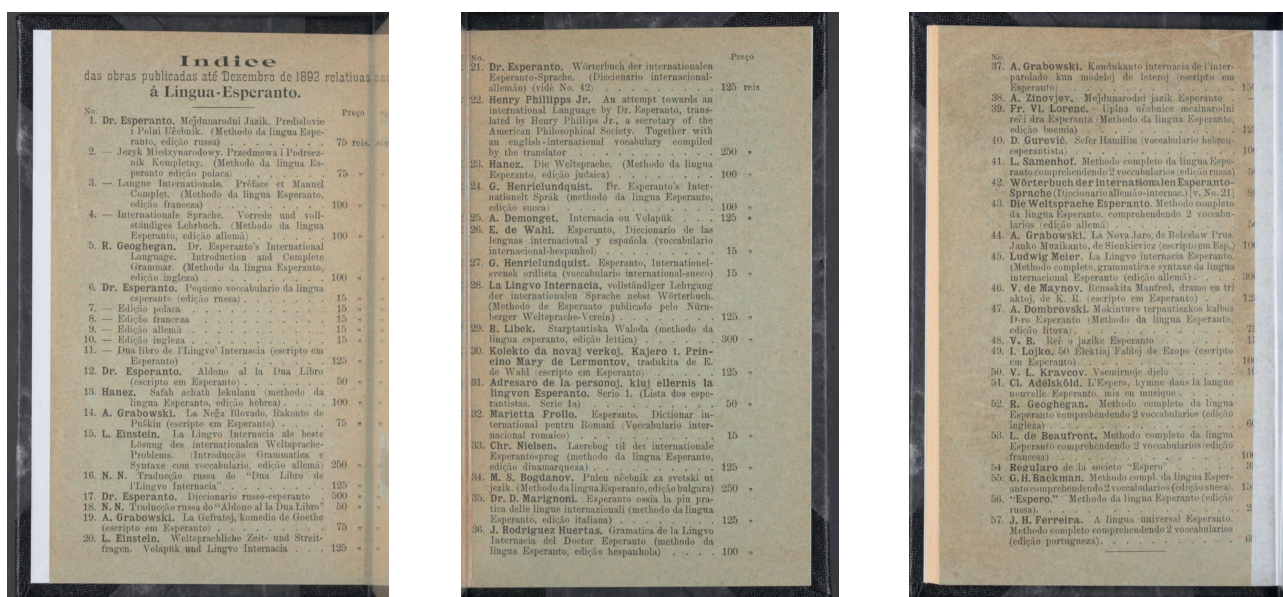
2. The 920-roots *vortaro* was re-used, probably for the last time, in 1893 (Esperanto-French textbook, de Beaufront 1893, ESN 53), though at that time much more complete dictionaries had been published.

The objective of the present paper is to present this corpus of early dictionaries, and to discuss the possibilities of further research and systematic online editing of their contents. Some examples of the full-text digitization are provided for 12 of these dictionaries.

1. THE CORPUS OF EARLY PUBLICATIONS OF ESPERANTO: TOWARDS A SYSTEMATIC CATALOGING

The early Esperantists publications are far from constituting an unknown corpus. Actually, from the very beginning, the leaders of the movement established, and maintained throughout decades, a detailed catalogue (Fig.1) of approved publications (below *Nomaro*[3]), all of them being assigned a serial number (below *Esperanto Serial Number* or ESN)[4].

Figure 1. the *Nomaro* (list of approved publications) as of December, 1892.
Extract of *A língua universal esperanto : methodo completo* (Ferreira 1892)

Original : Österreichische Nationalbibliothek (Vienna). Available online. http://data.onb.ac.at/rec/AC04212758

But this list is only seemingly complete. 2 documents, published in 1888, soon disappeared from the *Nomaro* and were replaced by new publications. The history of the replacement is well documented (Korzhenkov, 2009, p.89). Initially, a young medical student from Warsaw, Julius Steinhaus (1865-1922), an acquaintance of L. Zamenhof, volunteered for the preparation of a quick English translation of the initial text. *Dr. Esperanto's International Tongue* was published with a corresponding dictionary in January 1888 (see Table 1, items 5.1 and 10.1). But the translation was exceedingly bad, and Zamenhof decided to withdraw and destroy all stored copies. During the year 1888, an Oxford philologist, Richard Geoghegan, prepared a new translation of the same text, entitled *Dr. Esperanto's International Language*, and of the corresponding dictionary (Table 1, items 5.2 and 10.2); both were released in January of the next year.

Where the *Nomaro* shows only 57 items, by including the two early items translated by J. Steinhaus, we reach on our side a total of 59 works published.

---

3. Original title *: Nomaro pri l'verkoj pri la lingvo internacia Esperanto* (List of works on the international language Esperanto)

4. The present working paper was written during the Covid-19 lockdown in Europe (March-April, 2020). Therefore, due to the closing of all public and national libraries, it was impossible to consult several sources, in particular the classical bibliography of early Esperantist works first edited in 1929, and reprinted in 1973 (Stojan, 1929).

Table 1. Successive versions of the first English translation (1888 and 1889)

───

Item 5, first version (**5.1** in our counting) : textbook for English speakers
    Title *Dr. Esperanto's international tongue. Preface and complete method. Ed. for Englishmen by J. St.*
    Author: Steinhaus, Julius
    Date: released in January 1888, **removed from the catalogue the same year**
Item 10, first version (**10.1** in our counting): Esperanto-English dictionary
    Title *International-English Vocabulary. Vortaro por Angloj*
    Author. Steinhaus, Julius
    Date: released in January 1888, **removed from the catalogue the same year**
Item 5, second version (**5.2** in our counting): textbook for English speakers)
    Title *Dr. Esperanto's International Language, Introduction and Complete Grammar. por Angloj*
    Author Geoghegan, Richard
    Date: released in January 1889, **replacing immediately 5.1 in the catalogue**
Item 10, second version (**10.2** in our counting): Esperanto-English dictionary)
    Title *Dr Esperanto's International-English Vocabulary. Vortaro por Angloj*
    Author: Geoghegan, Richard
    Date: released in January 1889, **replacing immediately item 10.1 in the catalogue**

───

D. Eckert – CNRS, 2020

The Austrian National Library (*Sammlung für Plansprachen* — Department of Planned Languages) has a gathered an unique collection of these early Esperantist publications. Still, it holds only 40 out of the 59 items. The 19 others are scattered in many national and public libraries, sometimes not easily accessible, and, in the worst case, are not traceable at all. There is therefore no *complete* collection of early books and documents of the Esperantist movement anywhere in the world. Of course, these documents have never been considered together and even less systematically compared[5]. Our idea is to fill that gap, and to collect systematically electronic versions of all locatable documents.

The first phase of our research (2019-2020) implied several steps:

1. identifying and cataloging all the items of the corpus (including the missing ones)
2. downloading scanned copies of all available original documents
3. ordering scanned copies of rare items stored by national libraries but not available in digital form (Vienna, Jerusalem, Sofia, London, Lisbon, Madrid, Copenhague)
4. making direct pictures of an exemplar in a library (one case, Lyon).

Step 1 resulted in the building of a small bibliographic database, called by us *Nomaro-expanded*. We recorded all the characteristics of the 59 Esperantist items published from 1887 until 1892 (Table 2).

Steps 2 to 4 resulted in the retrieval of 49 items out of the total 59 (in facsimile or reprint form). So far, 8 items (ESN 17, 28, 38, 41, 50, 51, 52, 55) are stored in public libraries and need to be scanned upon request (Oxford, Stockholm, Saint-Petersburg, Vienna, Moscow). Four more items (ESN 42, 48, 49, 56) are, for the moment, out of our reach. ESN 42, a 1889 German-Esperanto dictionary disappeared during the war[6], while the remaining three are unknown publications (we only now their titles, corresponding to Esperanto Serial Numbers 48, 49, and 56).

─────────────────────────────

5.   Such an archive did exist before World War II. But all the documents gathered by Ludwik Zamenhof in Warsaw, stored and preserved by his family after his death (1917), were destroyed by the Nazis in 1939. The documents that were collected and published later in Japan by *Ludovikito* (Ito Kanzi) in his series *Ludovikologia Dokumentaro* are far from being complete, and often lack basic bibliographic indications.

6.   We found only reference of this document at the Staatsbibliothek zu Berlin, but the exemplar is reported as *Kriegsverlust* (lost during the war).

Table 2. Excerpt of the *Nomaro-expanded* – First twelve items

ESN*	Remark	Title	Language	Year	Authors - Translators	Library	Permalink
1	-	*Международный язык*	Russian	1887	Zamenhof	ÖNB**	http://data.onb.ac.at/rec/AC13760387
2	-	*Język międzynarodowy*	Polish	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215123
3	-	*Langue internationale*	French	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04217603
4	-	*Internationale Sprache*	German	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215124
5.1	replaced by 5.2 in the *Nomaro*	*International Tongue*	English	1888	Steinhaus & Zamenhof	British Library	http://explore.bl.uk/BLVU1:LSCOP-ALL:BLL01001164235
5.2	-	*International Language*	English	1889	Geoghegan & Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04212621
6	-	*Vortaro por Rusoj*	Russian	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215746
7	-	*Vortaro por Poloj*	Polish	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215123
8	-	*Vortaro por Francoj*	French	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04217603
9	-	*Vortaro por Gemanoj*	German	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215124
10.1	replaced by 10.2 in the *Nomaro*	*Vortaro por Angloj*	English	1888	Steinhaus & Zamenhof	British Library	http://explore.bl.uk/BLVU1:LSCOP-ALL:BLL01001164235
10.2	-	*Vortaro por Angloj*	English	1889	Geoghegan & Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04212621
							D. Eckert (CNRS) - 2020

* ESN stands for Esperanto Serial Number
** ÖNB stands for Österreichische Nationalbibliothek

2. Documenting the early international diffusion of Esperanto through the first dictionaries (1887-1890)

2.1 Multilingual aspects of the early diffusion of the Esperanto project

The usual history of the diffusion of Esperanto is centered on the development of the language itself (main characteristics, concurrence with *Volapük*), or on the key people (Zamenhof and the early adopters), and tells the story of the progressive organization of the movement (Korzhenkov, 2009, 2011). To sum up, these approaches are generally diachronic. The beginnings of the movement are typically described from 1887 (the *Genesis* momentum) up to 1905, date of the first World Congress in Boulogne-sur-Mer. 1905 is also the year of publication of the *Fundamento de Esperanto* by L. Zamenhof, a text immediately adopted as the unique authority over the language by the Congress participants.

On the contrary, our research is based on a synchronic comparative analysis of the first publications describing the project and the language to the international audience. The idea is to document the global diffusion of the project by analyzing simultaneously the publications that were released in many languages shortly after the birth of Esperanto.

From this perspective, the years 1887-1890 are of crucial importance. The international diffusion of Esperanto can be explained, among other factors, by the very quick translation/ adaptation of the initial brochure in many languages. Indeed, the first brochure describing the project of *lingvo internacia*, a short booklet, soon nicknamed *First Book* (Unua Libro) by the early Esperantists, was released almost simultaneously in Russian, Polish, French, and German (June— November 1887), and completed by four corresponding small dictionaries (Esperanto-target language). The sole author/tranlator of the 1887 documents is Zamenhof, who was a distinguished polyglot. The following year, four new brochures were released. The *First Book* was translated into English by Julius Steinhaus, and then in Hebrew and Yiddish by Naftali Najmanovich, two Warsaw acquaintances of Zamenhof (January—September 1888) and, at the end of the year, the first systematic textbook (for German speakers) was published in Germany by Leopold Einstein, an adopter from Nürnberg. The following years, many adaptations of the initial brochure were released, targeting people from different nationalities. By the end of 1890, a total of 36 booklets, textbooks or dictionaries aiming at presenting Esperanto to the international public had been published[7], in 16 different languages (Table 3).

**Table 3: The 36 documents published in natural languages
released between 1887 and 1890 for promoting Esperanto**

Year	Published	Languages (**bold**: new languages)
1887	8	**Russian**, **Polish**, **French**, **German**
1888	5	**English**, **Hebrew**, **Yiddish**, *German*
1889	14	**Swedish**, **Spanish**, **Latvian**, **Rumanian**, *English, Russian, German, French*
1890	9	**Danish**, **Bulgarian**, **Lithuanian**, **Italian**, **Czech,** *Spanish, Russian, Hebrew*

D. Eckert – CNRS (2020)

---

7. we exclude from this list of works published between 1887-1890 the twelve books written exclusively in Esperanto, whose target group was the people that already had a significant knowledge of the language.

Among these 36 items, we isolate a specific sub-corpus of particular interest: the first dictionaries. Dictionaries were obviously an essential tool for the spreading of the new language. The early Esperantists made considerable efforts for preparing as much dictionaries as possible. They were sometimes printed and published separately (as for the initial four 1887 brochures), and sometimes included in the textbooks. More than twenty dictionaries were released between 1887 and 1890. Seventeen of these dictionaries deserve, in our view, particular attention.

2.2 The Seventeen Early Small Dictionaries

The first brochures of 1887 were accompanied by a small vocabulary of around 900 words, or, in more precise terms, of Esperanto roots (*radikoj*) and affixes (*afiksoj*) that are the basis for the formation of a much bigger number of words (verbs, adjectives, adverbs, nouns...). This first vocabulary was meant to help people to understand the first texts elaborated by Zamenhof and to create their own texts (small letters, etc.). The very first one is the Esperanto-Russian dictionary (Fig.2).

Fig. 2. Excerpt of the *Vortaro por Rusoj* – Dictionary for Russian Speakers (Zamenhof, 1887, ESN 6)

Original : Österreichische Nationalbibliothek (Vienna). Available online. See Table 2 for the hyperlink

Many brochures include this short list of 900 entries[8]. Surprisingly, these *vortaroj* were most frequently unidirectional (Esperanto-target language). The first bi-directional small *vortaroj* (for Latvian speakers by Rudolf Libeks, and for English speakers by Henry Phillips, both in 1889) were released only two years after the first publication.

This limited document should have had a short life. And yet, while more more comprehesive

---

8.   Actually, the average number of entries in these small *vortaroj* is closer to 920 than to 900.

dictionaries had already been published in 1889 (including between 2000 and 3000 roots)[9], the original small vocabulary continued its international career. We identified a total of eighteen documents published between 1887 and 1890, where the initial 920-roots *vortaro* is present. For logical reasons, we exclude the second Esperanto-German *vortaro*, published in Nürnberg by Leopold Einstein in 1888 (it is a simple replication of the 1887 German *vortaro*), thus reducing our corpus to seventeen items (Table 4).

Table 4. The Parallel Corpus of the 17 small *Vortaroj* (1887-1890)

ESN	Language	Year	Authors - Translators	Library	Permalink or System number
6	Russian	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215746
7	Polish	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215123
8	French	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04217603
9	German	1887	Zamenhof	ÖNB	http://data.onb.ac.at/rec/AC04215124
10.1	English	1888	Steinhaus	British Library	http://explore.bl.uk/BLVU1:LSCOP-ALL:BLL01001164235
10.2	English	1889	Geoghegan	ÖNB	http://data.onb.ac.at/rec/AC04212621
13	Hebrew	1888	Najmanovich	INL	https://www.nli.org.il/en/books/NNL_ALEPH001909605/NLI
22	English	1889	Phillips	LoC	https://lccn.loc.gov/12040473
23	Yiddish	1888	Najmanovich	ÖNB	http://data.onb.ac.at/rec/AC04206542
26	Spanish	1889	de Wahl	ÖNB	http://data.onb.ac.at/rec/AC04216409
27	Swedish	1889	Henriclundquist *	ÖNB	http://data.onb.ac.at/rec/AC04209376
29	Latvian	1889	Libeks	ÖNB	http://data.onb.ac.at/rec/AC04218615
32	Romanian	1889	Frollo **	BML	https://catalogue.bm-lyon.fr System number 470092
34	Bulgarian	1890	Bogdanov	BNL	https://plus.bg.cobiss.net/opac7/bib/1051828964
35	Italian	1890	Marignoni	ÖNB	http://data.onb.ac.at/rec/AC04215418
39	Czech	1890	Lorenc	ÖNB	http://data.onb.ac.at/rec/AC10080420
47	Lithuanian	1890	Dąbrauskas K.A. ***	ÖNB	http://data.onb.ac.at/rec/AC04217283
					D. Eckert - CNRS (2020)

BML - Bibliothèque Municipale de Lyon
BNL - Bulgarian National Library
ESN - Esperanto Serial Number
INL - Israel National Library
Loc - Library of Congress
ÖNB - Österreichische Nationalbibliothek

* later known as Sébastien Voirol after he emigrated to France

** Marietta Frollo (Bucarest) is the unique female author present in the corpus

***pseudonym used by Adomas Jakštas

9. The Esperanto-Russian dictionary, with more than 3000 roots was released in January 1889 (ESN 17); and the Esperanto-German dictionary with around 2000 roots in February of the same year (ESN 21)

3. Digitizing work: methodology and results

To sum up, this set of documents has the characteristics of a parallel corpus that deserves specialized analysis, for instance by linguists. Such analysis is far beyond the reach of the present paper (and of the author). Though, we have made a consistent preparatory work for a later systematic encoding and for further analysis of the present material by scholars of different disciplines. Our objective is to make this corpus available to every interested scholar, regardless of the discipline. This material could be used in varied fields such as linguistics, interlinguistics, cultural history, etc.

But, in order to make it exploitable, we can not simply point to the catalogues of the libraries where the different *vortaroj* are stored (as indicated in Table 4). We know from direct and recent experience that gathering these copies can be an long process, that would cost anyone time, energy, and finally money. That is why we want to capitalize on our work (and on the remarkable collection constituted by the Austrian National Library) to offer a wider audience a structured output of this research.

After having systematically collected scans of the first *vortaroj*, it became obvious that a next step would be necessary. It is for instance legally impossible to deposit all these image scans in an Open Access Archive: there are varying limitations set by the different libraries to a public use of these digital versions. Some of them are already online, but many documents are subject to reuse restrictions. Besides, these document are in image format and are therefore not always easily readable, not to mention the linguistic heterogeneity of the corpus. On the other hand, the **textual content** of these documents **is in the public domain**.

That is the reason why we decided to carry out a full-text digitization of as many documents as possible. We created a simple uniform template for all dictionaries with 4 columns and 924 lines (Table 5):

1. Uniform code number; below [ID] for all Esperanto entries found in these dictionaries, from [1] to [924]
2. Original spelling of the Esperanto word in the first *vortaro* ever published (Esperanto-Russian, 1887), below [Esperanto ESN 1–1887)]
3. Spelling of the Esperanto word the current *vortaro*, below [Esperanto ESN 35–1890]
4. Entry in the target language

Table 5 — Standard format for the full-text digitization of a Vortaro

1	2	3	4
ID	**Esperanto** ESN 1 (1887)	**Esperanto** ESN 35 (1890)	**Italian**
482	*malgraŭ*	**malgrau*	malgrado
483	*man*	*man*	mano
484	*manĝ*	*man**g***	mangiare
			D. Eckert - CNRS (2020)

Source: ESN 35 (Marignoni, 1890)

We first encoded the four 1887 *vortaroj* (Russian, Polish, French, German). This helped us to solve methodological problems. For instance, we decided to make an edition as faithful as possible to the original. We adopted a general principle for our digitizing work: to achieve a **diplomatic edition** of all these documents. Original spellings are respected (outdated German spelling, old-

style Russian, among others), and typos are reproduced without correction. The Russian *vortaro* is an interesting example (Table 6). The 1887 document uses pre-revolutionary spelling (called *old-style spelling* in Russia). Instead of modernizing the entries (as in column 4), we kept the original (column 3).

Table 6: The original spelling in the Russian dictionary

1	2	3	4
ID	Esperanto root	Russian old-style spelling	Contemporary Russian spelling
871	*um*	можетъ имѣть различное значенie	может иметь различное значениe
			D. Eckert - CNRS (2020)

Source: ESN 6 (Zamenhof, 1887)

But the best example is the *Vortaro por Angloj* (Dictionary for English speakers) published by J. Steinhaus in Warsaw in 1888. This document is notoriously full of English mistakes (see p.2). We reproduced them systematically. If one compare one of Steinhaus' translations, with the equivalent work done by R. Geoghegan, the interest of the diplomatic edition is obvious (Table 7). In that specific case, J. Steinhaus invented the word frather (instead of **brother**).

Table 7: Comparing entries between two dictionaries for English speakers

ID	Esperanto root	Steinhaus' translation (1888)	Geoghegan's translation (1889)
76	*bo*	got by marriage (own or other people's); f. in. *patr* — father, *bo,patr* — father in law; *frat* — frather, *bo,frat* — frather in law	relation by marriage (own or other people's) e. g. *patr'* — father, *bo'patr'* — father in law; *frat'* — **brother**, *bo'frat'* — **brother** in law
			D. Eckert - CNRS (2020)

Source: ESN 10.1 (Steinhaus, 1888) and ESN 10.2 (Geoghegan, 1889). (in red, a word created by Steinhaus in **bold**, the correction by Geoghegan)

Other dictionaries posed difficult problems. The most challenging was the *Vortaro por Izraelidoj* (ESN 23) — Dictionary for Yiddish speakers. The document is a rare combination of different languages and scripts. One may find on the same page and sometimes in the same entry:

- two writing directions (European and Hebrew)
- three alphabets 1.2.3. and four languages a.b.c.d. : 1. Latin (a. Esperanto – b. Polish), 2. Cyrillic (c. Russian), 3. Hebrew (d. Yiddish)

Table 8: Yiddish, Russian, Polish and Esperanto – Dictionary for Yiddish speakers (1888)

ID	Esperanto	Yiddish
4	*ad*	צייגט אויף איין פעולה , אז עס איז נישט געמאלט ווארען אין איין מאהל נאר אין לענגערע צייט , למשל : *ir* (געהען) , אזוי ווי אין רוסישען идти אונד אין פּאָלנישען *iść* ir,ad (לאנג געהען) chodzić ходить .
		D. Eckert - CNRS (2020)

Source: ESN 23 (Najmanovich, 1888)

Usual word processors or spreadsheet software are not adapted to such a complex situation. We decided to use a specific word processor, *Mellel*, that allowed us to create multilingual tables of excellent quality[10]. In the end, a simple copy-and-paste (from Mellel to Excel) allowed us to create a .csv file.

To sum up, as of April 2020, we could fully digitize twelve different *vortaroj* (Table 9), and partially the Hebrew one. Some results of this work are displayed below (Table 10, p. 11).

Table 9 : state of the encoding work as of April, 2020

ESN	Target language	Encoding (percent)
6	Russian	100
7	Polish	100
8	French	100
9	German	100
10.1	English	100
10.2	English	100
13	Hebrew	68
22	English	100
23	Yiddish	100
26	Spanish	100
27	Swedish	100
29	Latvian	-
32	Romanian	-
34	Bulgarian	-
35	Italian	100
39	Czech	100
47	Lithuanian	-
	D. Eckert – CNRS 2020	

---

10. *Mellel* © runs only under Mac OS. More at https://www.mellel.com

Twelve early dictionaries of the Esperanto language compared (1887 - 1890)

ID	Esperanto	Russian	Polish	French	German	English (Steinhaus)	Swedish	English (Geoghegan)	Spanish	English (Phillips)	Czech	Italian	Yiddish
1	*a*	означаетъ прилагательное; напр. *hom* человѣкъ — *hom,a* человѣческій	oznacza przymiotnik; np. *hom* człowiek — *hom,a* ludzki	signifie un adjectif; par exemple: *hom* — *hom,a* humain	bezeichnet das Adjektiv; z. B. *hom* Mensch — *hom,a* menschlich	expresses an adjective; f. in. *hom* — man, *hom,a* — human	betecknar adjektivet; t. ex. *hom'* menniska — *hom'a* mensklig	expresses an adjective; e. g. *hom'* — man, *hom'a* — human	significa un adjetivo; por ejemplo: *hom'* — el hombre, *hom'a* — humano	expresses an adjective; f. in. *hom* — man, *homa* — human	(přípona přídav. jména)	des. aggettivi	קומט אם סוף ווארט פֿר אַ סימן אז דאס ווארט איז אײן אײגענשאָפֿטסווָארט (שם התיאר) . — למשל : *bon,a* אַ גוטער .
2	*acid*	кислый	kwaśny	acide, aigre	sauer	sour, acid	sur	sour, acid	agrio, ácido	sour, acid	kyselý	acido	זױער .
3	*ačet*	купить	kupować	acheter	kaufen	to buy	köpa	to buy	comprar	to buy	kupovati	comperare	קױפֿען .
4	*ad*	означаетъ продолжительность дѣйствія, напр. *ir* идти — *ir,ad* ходить; *danc* танецъ — *danc,ad* танцованіе	oznacza trwanie czynności; np. *ir* iść — *ir,ad* chodzić; *danc* taniec — *danc,ad* tańczenie	signifie la durée d'une action; par exemple: *paf* coup de fusil — *paf,ad* fusillade	bezeichnet die Dauer der Thätigkeit; z. B. *ir* gehen — *ir,ad* lange gehen; *danc* der Tanz — *danc,ad* das Tanzen	indicates the duration of an action; f. in. *ir* — go, *ir,ad* — to walk; *danc* — a danc, *danc,ad* — dancing	betecknar händelsens varaktighet; t. ex. *ir* gå — *ir'ad'* gå längre; *danc'* dans — *danc'ad'* dansandet	indicates the duration of an action; e. g. *ir'* — go, *ir'ad'* — to walk; *danc'* — a dance, *danc'ad'* — dancing	significa la duracion de la actividad; por ejemplo: *pens'* pensar, *pens'ad'* — andar pensando	indicates the duration of an action; f. in. ir — go, ir-ad — to walk; danc — a dance, danc-ad — dancing	(trvání děje)	s. az. continuata	צײגט אויף אײן פעולה , אז עס איז נישט געמאכט ווארט אין אײן מאהל נאר אין לענגערע צײט , למשל *ir* (געהן) , אזוי וויא אײן רוסישען אונד אין פֿאַלנישען *iśĉ ir,ad* (לאָבג) *chodziĉ* ходить (געהען) .
5	*adiaŭ*	прощай	adieu, bądź zdrów	adieu	adieu, lebe wohl!	adieu, good-bye	adjö, farväl	adieu, good-bye	adios	adieu, good-bye	s Bohem!	addio	אַריע ! לעבט װאָהל !
6	*aer*	воздухъ	powietrze	air	Luft	the air	luft	the air	aire	the air	vzduch	aria	לופֿט
7	*afer*	дѣло	sprawa, interes	affaire	Werk, Sache, Angelegenheit	affair, business	affär, sak	affair, business	obra, asunto, negocio	affair, business	věc, záležitosť	affare faccenda impresa cosa	זאַך , געשעפֿט .
8	*agl*	орелъ	orzeł	aigle	Adler	the eagle	örn	the eagle	águila	the eagle	orel	aquila	אָדלער
9	*agrabl*	пріятный	przyjemny	agréable	angenehm	agreeable	angenäm	agreeable	agradable	agreeable	příjemný	gradevole	אנגענעהם
10	*aĝ*	вѣкъ, возрастъ	wiek	âge	Alter	the age	ålder	the age	edad	the age	věk, stáří	età	אלטער (ימי הָאָדם) , לטשל ? : *Kia,n aĝ,o,n vi hav,as* (אײן וועלכען אלטער ביזט דוא ?)

First 10 entries of twelve different *Esperanto - target language* dictionaries - D. Eckert (CNRS - Centre National de la Recherche Scientifique) , 2020

Table 10 : The first ten entries of twelve Vortaroj

4. Research perspectives

There are two main perspectives for this research.

4.1. Digitizing the remaining 5 *vortaroj*

- So far, we have fully digitized 12 dictionaries in 10 languages (Table 9). We do not speak all these languages: specifically, our skills in Swedish, Italian, and Czech are close to zero. But for these three languages the digitizing work was rather easy, the scanned copies being of good quality, and our knowledge of related languages having helped us to decipher correctly the originals.

- For similar reasons, the next possible step would be the digitizing of the Romanian (ESN 32) and Bulgarian (ESN 34) dictionaries, a work we could carry out ourselves with limited supervision by a native speaker.

- The last three *vortaroj* pose more complex problems. The dictionaries in Hebrew (ESN 13), Latvian (ESN 29), and Lithuanian (ESN 47) cannot be digitized in full-text without expert help.

1. Hebrew: we have done some preliminary work on that *vortaro*, but this work needs expert supervision. By partially digitizing this dictionary, we specifically intended at identifying words or short sentences written in Yiddish (introduced by the author for exemplifying the Hebrew translation in a language more familiar to the Jewish readership of the late 19th century). We could document such Yiddish translations for all entries: Najmanovich completed 905 entries (out of 921) of his dictionary with Yiddish examples (Table 11).

Table 11: use of Yiddish equivalents in the Dictionary for Hebrew speakers (1888)

ID	Esperanto	Original Hebrew entry with Yiddish in brackets	Yiddish (transliterated)	English translation
126	*danc*	רָקד (טאַנצען)	tants'n	to dance
127	*danĝer*	סכנה (געפֿאהר)	gefa'r	danger
128	*dank*	הודות (דאַנקען)	dank'n	to thank
				D. Eckert – CNRS (2020)

Source: ESN 13 (Najmanovich, 1888)

The full digitizing of the Hebrew would imply further steps. An indispensable prerequisite is a new high-resolution (minimum 300 dpi) scanned copy, since the file at our disposal does not allow us for instance to read diacritical marks (*niqqud* and *dagesh*) with certainty, or to remove ambiguities between some letters (typically ב and כ). Besides, this work should be completed by a person with adequate skills in Hebrew. The late-19th century Hebrew used by Najmanovich is far from being the contemporary standard, so we would like to involve a scholar with a sufficient knowledge of *fin de siècle* written Hebrew.

2. Latvian: the document at our disposal is a low-resolution B&W scan, difficult to decipher (like the Hebrew dictionary). Furthermore, the pre-1918 Latvian spelling used is not the modern standard: it was then close to German, with no use of diacritical marks like in the contemporary language. Besides, the text is printed in a difficult-to-read *Fraktur* typesetting.

3. Lithuanian: the pre-1918 Lithuanian used is not the contemporary standard. Same remarks as for Hebrew and Latvian apply: a cooperation with a scholar having the required linguistic skills is indispensable.

4.2. Encoding the *vortaroj* according to the TEI standards.

The last step of this work is a systematic encoding in a format recognized by linguists, such as *Text Encoding Initiative* or TEI. A collaboration with specialists is a prerequisite. We already have begun to share our sources, metadata and the digitized *vortaroj* with Francesca Frontini (Praxiling, University of Montpellier), a researcher in computational linguistics.

Francesca Frontini and Denis Eckert have elaborated guidelines for a further common work:

„The next step in the project is the conversion into TEI, using the lexicographic module. In creating the data model we have adhered as closely as possible to the guidelines of TEI-lex0 (Bánski et al., 2017).The TEI version will faithfully describe the content of each vortaro, with precise metadata information about the original edition; at the same time it should allow and facilitate the exploration of the collection, by means of an interlinguistic alignment of each lexical entry as well as by enriching the corpus with explicit linguistic information, such as part of speech."

A proposal of TEI modelling has been elaborated by Frontini, and a preliminary test made for two *vortaroj* : French – ESN 8 (Annex 1), and German – ESN 9 (Annex 2).

5. Conclusion

The encoding of the *vortaroj* is the first step of a more comprehensive project. The early years of Esperanto constitute a remarkable moment of creation of an universalist contributing project in the age of "First Globalization" (Berger, 2003). It should be therefore systematically documented.

The project is to achieve, in the long run, the systematic full-text encoding of all the early works (1887-1893) of the Esperanto movement, and make their contents accessible online in a standardized format, thus avoiding the difficult question of the copyrights issues related to image-scanned items belonging to national libraries. This corpus would then be accessible to a great variety of scholars (in History, Linguistics, Digital Humanities, Cultural Studies, etc.), allowing them process this corpus with their respective methods, in many possible thematic directions.

References

Bánski, P., Bowers, J., & Erjavec, T. (2017). *TEI-Lex0 Guidelines for the Encoding of Dictionary Information on Written and Spoken Forms.* ELex2017.

Berger, S. (2003). *Notre première globalisation: Leçons d'un échec oublié*. Seuil.

Haupenthal, R. (2013). *Über die Startbedingungen zweier Plansprachen: Schleyers Volapük (1879/80) und Zamenhofs Esperanto (1887).* Ed. Iltis.

Korzhenkov, A. (2009). *Zamenhof: The Life, Works and Ideas of the Author of Esperanto* (H. Tompkin, Trans.). Mondial - Universal Esperanto Association.

Korzhenkov, A. (2011). *Homarano: La vivo, verkoj kaj ideoj de d-ro L. L. Zamenhof (2a eldono).* Sezono, Litova Esperanto-Asocio.

Stojan, P. E. (1929). *Bibliografio de internacia lingvo.* Bibliografia servo de Universala esperanto-asocio.

Annexes
TEI Coding proposals

```xml
<TEI xml:id="vortaro-fr-test">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title type="main">Unua Libro - French</title>
        <editor key="DE" role="scientific editor">Denis Eckert</editor>
      </titleStmt>
      <editionStmt>
        <edition>
          First edition of the French "Unua Libro"
          <date>1887</date>
        </edition>
        <respStmt>
          <resp>OCR, encoding</resp>
          <name key="DE">Denis Eckert</name>
        </respStmt>
        <respStmt>
          <resp>TEI, encoding</resp>
          <name key="FF">Francesca Frontini</name>
        </respStmt>
      </editionStmt>
      <publicationStmt>
        <publisher>UMR Géocités</publisher>
        <pubPlace>
          <address>
            <addrLine>Paris Aubervilliers</addrLine>
            <addrLine>Campus Condorcet, 5 cours des Humanités, </addrLine>
            <addrLine>F-93322 Aubervilliers (Cedex)</addrLine>
          </address>
        </pubPlace>
        <availability status="free">
          <licence target="http://creativecommons.org/licenses/by/4.0/deed.fr">
            Extraits : Attribution — Vous devez créditer l'œuvre, intégrer un lien vers la licence et
            indiquer si des modifications ont été effectuées. Vous devez indiquer ces informations
            par tous les moyens possibles mais vous ne pouvez pas suggérer que les éditeurs vous
            soutiennent ou soutiennent la façon dont vous avez utilisé leur œuvre. Les éditeurs
            demandent que toute publication dérivée de ses ressources électroniques comporte : 1.
            leur nom, ainsi que les noms des personnes auteurs de l'édition ; 2. l'URI permettant
            d'accéder à la page concernée du présent site web, ou à défaut à la page d'accueil de ce
            site ; 3. la date et la version du fichier source utilisé (le cas échéant). Pas de restrictions
            additionnelles — Vous n'êtes pas autorisé à appliquer des conditions légales ou des
            mesures techniques qui empêcheraient autrui d'utiliser l'œuvre dans les conditions
            décrites par la licence.
          </licence>
        </availability>
      </publicationStmt>
```

```xml
        −<sourceDesc>
          −<bibl>
              <author>Ludwik Lejzer Zamenhof</author>
            −<title>
                Dr. Esperanto - Langue Internationale - Préface et Manuel complet - Por Franc,o,j
              </title>
              <publisher>Gebethner et Wolff</publisher>
              <pubPlace>Kelter</pubPlace>
              <date>24/09/1887</date>
            −<distributor>
                <orgName>Österreichische Nationalbibliothek</orgName>
                <placeName>Vienna, Austria</placeName>
                <idno type="URI">http://data.onb.ac.at/rec/AC04217603</idno>
              </distributor>
            </bibl>
          </sourceDesc>
        </fileDesc>
      −<profileDesc>
        −<langUsage>
            <language ident="fr">French</language>
            <language ident="eo">Esperanto</language>
          </langUsage>
        </profileDesc>
      </teiHeader>
   −<text>
    −<front>
      −<div type="preface">
          <head>Préface</head>
        −<p>
            Ce ne sera peut-être pas sans une certaine méfiance .....
          </p>
        </div>
      −<div type="bulletin">
          <p> .....</p>
        </div>
      −<div type="grammaire">
          <head>MANUEL COMPLET DE LA LANGUE INTERNATIONALE</head>
          <p>....</p>
        </div>
      </front>
    −<body>
      −<entry xml:id="fre_1887_1" n="n_1">
        −<form type="lemma">
            <orth>a</orth>
          </form>
        −<gramGrp>
            <pos resp="DE,FF" value="suffix"/>
          </gramGrp>
        −<sense>
          −<def>
              signifie un adjectif par exemple:
              <mentioned>hom — hom,a</mentioned>
```

```xml
          humain
        </def>
      </sense>
    </entry>
  −<entry xml:id="fre_1887_2" n="n_2">
    −<form type="lemma">
        <orth>acid</orth>
      </form>
    −<gramGrp>
        <pos resp="DE,FF" value="adjective"/>
      </gramGrp>
    −<sense>
        <def>acide, aigre</def>
      </sense>
    </entry>
  −<entry xml:id="fre_1887_3" n="n_3">
    −<form type="lemma">
        <orth>aĉet</orth>
      </form>
    −<gramGrp>
        <pos resp="DE,FF" value="adjective"/>
      </gramGrp>
    −<sense>
        <def>acheter</def>
      </sense>
    </entry>
  −<entry xml:id="fre_1887_4" n="n_4">
    −<form type="lemma">
        <orth>ad</orth>
      </form>
    −<gramGrp>
        <pos resp="DE,FF" value="preposition"/>
      </gramGrp>
    −<sense>
      −<def>
          signifie la durée d'une action; par exemple:
          <mentioned>paf</mentioned>
          coup de fusil —
          <mentioned>paf</mentioned>
          ,
          <mentioned>ad fusillade</mentioned>
        </def>
      </sense>
    </entry>
  −<entry xml:id="fre_1887_5" n="n_5">
    −<form type="lemma">
        <orth>adiaŭ</orth>
      </form>
    −<gramGrp>
        <pos resp="DE,FF" value="interjection"/>
      </gramGrp>
    −<sense>
```

```
            <def>adieu</def>
         </sense>
      </entry>
     −<entry xml:id="fre_1887_6" n="n_6">
       −<form type="lemma">
            <orth>aer</orth>
         </form>
       −<gramGrp>
            <pos resp="DE,FF" value="noun"/>
         </gramGrp>
       −<sense>
            <def>air</def>
         </sense>
      </entry>
     </body>
   </text>
</TEI>
```

```xml
−<TEI xml:id="vortaro-de-test">
  −<teiHeader>
    −<fileDesc>
      −<titleStmt>
          <title type="main">Unua Libro - German</title>
          <editor key="DE" role="scientific editor">Denis Eckert</editor>
      </titleStmt>
      −<editionStmt>
        −<edition>
            First edition of the German "Unua Libro"
            <date>1887</date>
        </edition>
        −<respStmt>
            <resp>OCR, encoding</resp>
            <name key="DE">Denis Eckert</name>
        </respStmt>
        −<respStmt>
            <resp>TEI, encoding</resp>
            <name key="FF">Francesca Frontini</name>
        </respStmt>
      </editionStmt>
      −<publicationStmt>
          <publisher>UMR Géocités</publisher>
        −<pubPlace>
          −<address>
              <addrLine>Paris Aubervilliers</addrLine>
              <addrLine>Campus Condorcet, 5 cours des Humanités, </addrLine>
              <addrLine>F-93322 Aubervilliers (Cedex)</addrLine>
          </address>
        </pubPlace>
        −<availability status="free">
          −<licence target="http://creativecommons.org/licenses/by/4.0/deed.fr">
              Extraits : Attribution — Vous devez créditer l'œuvre, intégrer un lien vers la licence et
              indiquer si des modifications ont été effectuées. Vous devez indiquer ces informations
              par tous les moyens possibles mais vous ne pouvez pas suggérer que les éditeurs vous
              soutiennent ou soutiennent la façon dont vous avez utilisé leur œuvre. Les éditeurs
              demandent que toute publication dérivée de ses ressources électroniques comporte : 1.
              leur nom, ainsi que les noms des personnes auteurs de l'édition ; 2. l'URI permettant
              d'accéder à la page concernée du présent site web, ou à défaut à la page d'accueil de ce
              site ; 3. la date et la version du fichier source utilisé (le cas échéant). Pas de restrictions
              additionnelles — Vous n'êtes pas autorisé à appliquer des conditions légales ou des
              mesures techniques qui empêcheraient autrui d'utiliser l'œuvre dans les conditions
              décrites par la licence.
          </licence>
        </availability>
      </publicationStmt>
```

```xml
          −<sourceDesc>
            −<bibl>
                <author>Ludwik Lejzer Zamenhof</author>
              −<title>
                  Dr. Esperanto - Internationale Sprache. Vorrede und vollständiges Lehrbuch - Por
                  German,o,j
                </title>
                <publisher>Kelter</publisher>
                <pubPlace>Varsovie</pubPlace>
                <date>02/11/1887</date>
              −<distributor>
                  <orgName>Österreichische Nationalbibliothek</orgName>
                  <placeName>Vienna, Austria</placeName>
                  <idno type="URI">http://data.onb.ac.at/rec/AC04215124</idno>
                </distributor>
              </bibl>
            </sourceDesc>
          </fileDesc>
        −<profileDesc>
          −<langUsage>
              <language ident="de">German</language>
              <language ident="eo">Esperanto</language>
            </langUsage>
          </profileDesc>
        </teiHeader>
      −<text>
        −<front>
          −<div type="preface">
              <head>Préface</head>
            −<p>
                Der Leser wird warscheinlich die vorliegende Broschüre mit einigem Misstrauen .....
              </p>
            </div>
          −<div type="bulletin">
              <p> .....</p>
            </div>
          −<div type="grammaire">
              <head>VOLSTÄNDIGES LEHRBUCH DER INTERNATIONALEN SPRACHE</head>
              <p>....</p>
            </div>
          </front>
        −<body>
          −<entry xml:id="de_188_17" n="N_1">
            −<form type="lemma">
                <orth>a</orth>
              </form>
            −<gramGrp>
                <pos resp="DE,FF" value="suffix"/>
              </gramGrp>
            −<sense>
              −<def>
                  bezeichnet das Adjektiv; z. B.
```

```xml
          <mentioned>hom</mentioned>
          Mensch —
          <mentioned>hom</mentioned>
          ,a menschlich
        </def>
      </sense>
    </entry>
  −<entry xml:id="de_1887_2" n="n_2">
    −<form type="lemma">
        <orth>acid</orth>
      </form>
    −<gramGrp>
        <pos resp="DE,FF" value="adjective"/>
      </gramGrp>
    −<sense>
        <def>sauer</def>
      </sense>
    </entry>
  −<entry xml:id="de_1887_3" n="aĉet_3">
    −<form type="lemma">
        <orth>aĉet</orth>
      </form>
    −<gramGrp>
        <pos resp="DE,FF" value="adjective"/>
      </gramGrp>
    −<sense>
        <def>kaufen</def>
      </sense>
    </entry>
  −<entry xml:id="ade_1887_4" n="n_4">
    −<form type="lemma">
        <orth>ad</orth>
      </form>
    −<gramGrp>
        <pos resp="DE,FF" value="preposition"/>
      </gramGrp>
    −<sense>
      −<def>
          bezeichnet die Dauer der Thätigkeit; z. B.
          <mentioned>ir</mentioned>
          gehen —
          <mentioned>ir</mentioned>
          ,
          <mentioned>ad</mentioned>
          lange gehen;
          <mentioned>danc</mentioned>
          der Tanz —
          <mentioned>danc</mentioned>
          ,
          <mentioned>ad</mentioned>
          das Tanzen
        </def>
```

```xml
        </sense>
      </entry>
    <entry xml:id="de_1887_5" n="n_5">
      <form type="lemma">
        <orth>adiaŭ</orth>
      </form>
      <gramGrp>
        <pos resp="DE,FF" value="interjection"/>
      </gramGrp>
      <sense>
        <def>adieu, lebe wohl!</def>
      </sense>
    </entry>
    <entry xml:id="de_1887_6" n="n_6">
      <form type="lemma">
        <orth>aer</orth>
      </form>
      <gramGrp>
        <pos resp="DE,FF" value="noun"/>
      </gramGrp>
      <sense>
        <def>Luft</def>
      </sense>
    </entry>
  </body>
</text>
</TEI>
```