Lexicography in Language Documentation and Education

Prof. Dr. habil. Nicole Nau

Adam Mickiewicz University in Poznan
Faculty of Modern Languages and Literatures

Documentation for Education of Lesser Used Languages

COLING Workshop, Rēzekne, 1-4 May 2023

Lexicography: Dictionaries, their making and their users

Outline of this talk

- ❖ Short introduction

- ❖ Dictionary making in language documentation

- ❖ My dictionary of words from Latgalian fairytales

- ❖ Using dictionaries in language education

- ❖ Education ⇔ Documentation: Learners as compilers

Dictionary making in language documentation

Traditions and goals before modern Documentary linguistics

❖ «**Colonial** tradition» and «missionary tradition»

❖ «**ethnographic** tradition»,

❖ «**pedagogical** tradition» (Haviland 2006)

«In what I call the Colonial tradition, collecting vocabularies was always a vocation of imperialists, often an accidental byproduct of exploration and conquest. Explorers collected flora and fauna, and often they also collected words.» (Haviland 2006)

Dictionaries in Himmelmann's concept of language documentation

Primary data	Apparatus	
	Per session	For documentation as a whole
recordings/records of observable linguistic behavior and metalinguistic knowledge	*Metadata* *Annotations* – transcription – translation – further linguistic and ethnographic glossing and commentary	*Metadata* *General access resources* – introduction – orthographical conventions – glossing conventions – indices – links to other resources … *Descriptive analysis* – ethnography – descriptive grammar – dictionary

But:

«dictionaries of endangered languages have begun to blur the boundaries between documentation and description. More and more, they have become repositories for primary data which include images, sound, and video»

«The advent of documentary linguistics has encouraged lexicographers to integrate documentary materials into the text so as to create multimedia dictionaries which are more like cultural encyclopedias in their range.»

(Ogilvie 2010)

Two examples

Ulrike Mosel's Teop dictionary

https://dictionaria.clld.org/contributions/teop

Yurok dictionary (and more)

https://linguistics.berkeley.edu/~yurok/web/search.php

Dictionaries of lesser vs better resourced languages

«Twenty years ago, field lexicography was lagging behind commercial lexicography on all levels, and my experience of dictionary making in the field as opposed to the office certainly provided a stark contrast to my future experiences in lexicography»

«[a job at the *Australian Concise Oxford Dictionary*] did little to prepare me for writing a dictionary of a language that 1) I had never heard before, 2) had never been written down before, and 3) was spoken by two last speakers who were rarely sober enough to teach me their language.»

(Ogilvie 2010)

«The past ten years have seen the emergence of new lexicographic policies and practices around the world that can be characterized by an innovative exploitation of new technologies, predominant use of oral as well as written sources, incorporation of pedagogical materials, and collaborative involvement of members of the speech community.»

«Using innovative lexicographic policies, practices, and technologies, the lexicographer is able to produce dictionaries that are useful to both communities and scholars; dictionaries that not only describe and preserve an endangered language – as was the goal of linguists in the past – but also help in the processes of maintenance and revitalization.»

(Ogilvie 2010)

So – what do we need?
(and often still lack)

❖ Collaboration, collaboration, collaboration… - of people with different kind of knowledge and skills

❖ Time, money, and well-organized projects with realistic goals

it is unrealistic to plan a comprehensive dictionary to be finished in less than ten years (Mosel 2004)

❖ Accessible technology that allows collaborative compiling and editing of multimedia dictionaries and offers a long-time perspective
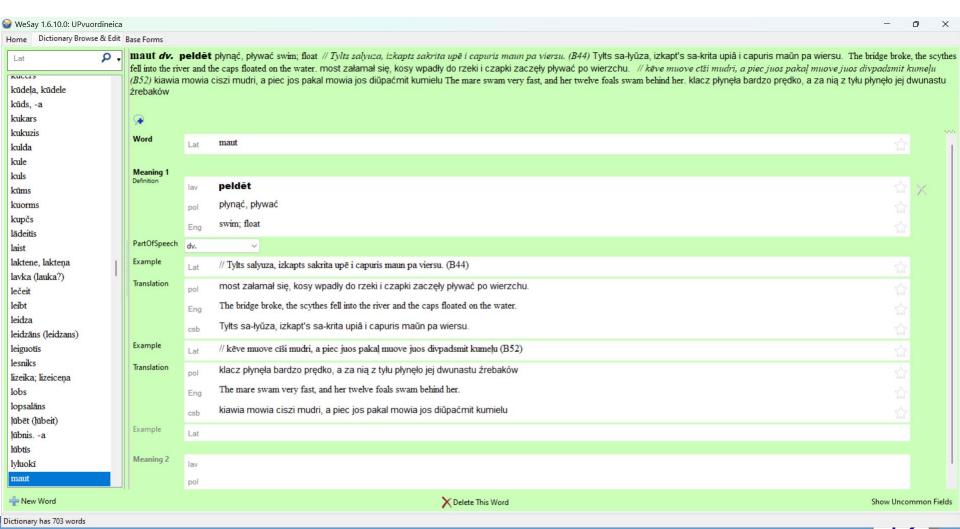
Dictionary making software – some examples

❖ WeSay (SIL) – non-commercial, free, proprietary, easy to use for simple dictionaries (PDF) – no maintenance after 2019

❖ FLEX (SIL) – more possibilities than WeSay, but more difficult to use, especially editing and publishing dictionaries is clumsy

❖ TLEx – semi-commercial, more professional

❖ Lexonomy – non-commercial, free, non-proprietary, relatively easy to use for online dictionaries

Example of a dictionary made with WeSay

❖ Idea: document words used in Stefanija Ulanowska's collection fairytales (1895)

❖ Microstructure: Headword in modern orthography, Latvian, English, and Polish translation equivalents; Examples in modern and original orthography, English and Polish translation

❖ To be published electronically (PDF, maybe e-book formats) and on paper

Editing view in WeSay

Word		
	Lat	maut

Meaning 1
Definition

	lav	**peldēt**
	pol	płynąć, pływać
	Eng	swim; float

PartOfSpeech | dv. ⌄

Example

	Lat	// Tylts salyuza, izkapts sakrita upē i capuris maun pa viersu. (B44)

Translation

	pol	most załamał się, kosy wpadły do rzeki i czapki zaczęły pływać po wierzchu.
	Eng	The bridge broke, the scythes fell into the river and the caps floated on the water.
	csb	Tyłts sa-łyŭza, izkapt's sa-krita upiâ i capuris maŭn pa wiersu.

Example

	Lat	// kēve muove cĭši mudri, a piec juos pakaļ muove juos divpadsmit kumeļu (B52)

Translation

	pol	klacz płynęła bardzo prędko, a za nią z tyłu płynęło jej dwunastu źrebaków
	Eng	The mare swam very fast, and her twelve foals swam behind her.
	csb	kiawia mowia ciszi mudri, a piec jos pakal mowia jos diŭpaćmit kumielu

Automatic output to LibreOffice document

M

maut DV. swim; float peldēt płynąć, pływać // Tylts salyuza, izkapts sakrita upē i capuris maun pa viersu. (B44) **Tylts sa-łyŭza, izkapt's sa-krita upiâ i capuris maŭn pa wiersu.** The bridge broke, the scythes fell into the river and the caps floated on the water. most załamał się, kosy wpadły do rzeki i czapki zaczęły pływać po wierzchu. // kēve muove cīši mudri, a piec juos pakal muove juos divpadsmit kumeļu (B52) **kiawia mowia ciszi mudri, a piec jos pakal mowia jos diŭpaćmit kumielu** The mare swam very fast, and her twelve foals swam behind her. klacz płynęła bardzo prędko, a za nią z tyłu płynęło jej dwunastu źrebaków

medebnīks v. hunter mednieks łowca // Kaids tys kačs lobs medēbnīks! (B34) **Kajds tys kaczsz łobs miediebniks!** What a good hunter

gons par znutu! but alright! Somehow we will manage so that I won't have a shepherd as a son-in-law! no, ale pomiarkujemy jakoś, żeby pastuch nie był moim zięciem!

miert zemē VĀRDKOPA die mirt (nost) umierać // Buoba soka, ka palaide jū iz cjotku, atzasproščātīs, ka miers zemē (B17) **Boba soka, ka pa-łajdia ju iz ciotku, aca-sproszczatiś, ka mierś ziamia** The woman said that she had let her go to bid farewell to her aunt, as she was dying Baba mówi, że puściła ją do ciotki umierającej, na pożegnanie

miša s. mass mise msza // Pazataiseja par kapleicenu, a jū pataiseja par baznīckungu, kur mišu turēja. (B49) **Paza-tajsieja par kaplejcieniu, a ju pa-tajsieja par baźnijc-kungu, kur miszu turiaja.** She transformed herself into a small chapel, and him she transformed into a priest that was reading mass. Stała się kapliczka, a jego zrobiła księdze

14

After minimal editing in Word
(PDF file for corrections)

M

maut DV. swim; float peldēt płynąć, pływać

// Tylts salyuza, izkapts sakrita upē i capuris maun pa viersu. (B44) **Tyłts sa-łyŭza, izkapt's sa-krita upiâ i capuris maŭn pa wiersu.** The bridge broke, the scythes fell into the river and the caps floated on the water. most załamał się, kosy wpadły do rzeki i czapki zaczęły pływać po wierzchu.

// kēve muove cīši mudri, a piec juos pakaļ muove juos divpadsmit kumeļu (B52) **kiawia mowia ciszi mudri, a piec jos pakal mowia jos diŭpaćmit kumielu** The mare swam very fast, and her twelve foals swam behind her. klacz płynęła bardzo prędko, a za nią z tyłu płynęło jej dwunastu źrebaków

Uses and users

Questions about my dictionary:

❖ Who needs such a dictionary?

❖ How could it be used in education?

Questions about dictionaries in general?

❖ What are dictionaries used for?

❖ How are dictionaries used in education?

«When writing a dictionary for an endangered language, it is not enough to just say we want to document the lexicon of the language for future generations and generate a dictionary automatically from a text corpus.

Rather, as with any other dictionary project, one has to identify the prospective users of the dictionary and what they will use the dictionary for.»

(Mosel 2004)

Cristinoi & Nemo (2013): «Endangered language dictionaries will thus be used:

- for research purposes;
- to document a specific language and more importantly a specific culture;
- to preserve a linguistic and cultural heritage which would disappear without written material;
- to help indigenous people communicate in a foreign (often dominant) language by finding the proper equivalents for indigenous words;
- to help non-native speakers understand native-speakers and their cultural background;
- to provide a stable orthography for the whole lexicon.»

Which dictionaries for Latgalian?

❖ Existing?

❖ Lacking?

❖ Most needed / wanted?

❖ Project DHELI (state funded project for development of digital humanities), working group at Rēzekne Academy of Techologies (RTA): planing a lexical database and future dictionaries

Why dictionaries?

❖ What makes dictionaries (potentially) so useful for language documentation and for education?

   ▫ Various kind of information about words

   ▫ Word as point of departure for language knowledge

Dictionary making in education

❖ Can be integrated into language and culture teaching at all levels and in almost all age groups.

❖ Supports learner autonomy, learning individually and in groups.

❖ Combines linguistic, artistic, and IT skills.

❖ Furthers collaboration between learners and the community.

Example: Making electronic dictionaries with Lexonomy

https://www.lexonomy.eu/

mauduotīs V

maudojās, mauduojuos / mauduovuos, mauduosīs

https://forvo.com/word/mauduot%C4%ABs/#ltg

=> mauduot; maut

— LLV peldēties, peldēt

ENG bath, swim

Te ir prūds. Soka, ka te beistami mauduotīs, daži ir nūsleikuši (Sandra Ūdre)

ENG There is a pond. They say it is dangerous to swim here, some people have drowned.

Itūšaļt baseinā maudovuos .. desmit mauduotuoji, nu tūs leluokuo daļa pensionari voi skūlnīki. (Aldis Bukšs)

ENG At that moment, ten swimmers were swimming in the pool, for the greatest part pensioners or school children.

Paldies par uzmanību!

E-mail

Nicole.Nau@amu.edu.pl

Homepage

www.naunicol-e.home.amu.edu.pl

References

Cristinoi, Antonia & François Nemo (2013) Challenges in endangered language lexicography. In *Lexicography and Dictionaries in the Information Age,* eds. Deny A.Kwary, Nur Wulan & Lilla Musyahda. Airlangga University Press. 126-132.

Haviland, John. 2006. Documenting lexical knowledge. In: *Essentials of language documentation,* eds. Jost Gippert, Nikolaus P. Himmelmann, Ulrike Mosel. Mouton de Gruyter,

Himmelmann, Nikolaus. 2006. Language documentation: What is it and what ist it good for? In *Essentials of language documentation*, eds. Jost Gippert, Nikolaus P. Himmelmann, Ulrike Mosel. Mouton de Gruyter, 1-30.

Mosel, Ulrike (2004) Dictionary making in endangered speech communities. In *Language Documentation and Description* 2: 39–54.

Ogilvie, Sarah (2011) Linguistics, lexicography, and the revitalization of endangered languages. *International Journal of Lexicography* 24.4, 389–404