



HAL
open science

Developing Keyboards for the Endangered Livonian Language

Mika Hämäläinen, Khalid Alnajjar

► **To cite this version:**

Mika Hämäläinen, Khalid Alnajjar. Developing Keyboards for the Endangered Livonian Language. Widening NLP, Nov 2021, Punta Cana, Dominican Republic. hal-03431557

HAL Id: hal-03431557

<https://hal.science/hal-03431557>

Submitted on 16 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Developing Keyboards for the Endangered Livonian Language

Mika Hämäläinen and Khalid Alnajjar
Rootroo Ltd & University of Helsinki, Finland

Abstract

We present our current work on developing keyboard layouts for a critically endangered Uralic language called Livonian. Our layouts work on Windows, MacOS and Linux. In addition, we have developed keyboard apps with predictive text for Android and iOS. This work has been conducted in collaboration with the language community.

1 Introduction

Livonian is a Uralic language that is classified as critically endangered according to Unesco (Moseley, 2010). The number of speakers is around 250 people according to a recent census in Latvia, although only 20 people speak the language fluently (Ernštreits, 2019). In this paper, we describe our efforts in building keyboards for the Livonian language¹.

Language users of any majority language take something as fundamental as keyboards for granted. For us, it is evident that we can type our language with ease on any operating system or device. This, however, is not the case for Livonian, despite of the availability of keyboard layouts for some other endangered Uralic languages (Moshagen et al., 2019). Livonian has letters that are not present in the superstrate language (Ernštreits, 2007), Latvian, spoken in the same country as Livonian, nor in the other larger Finnic languages, Finnish and Estonian. This means that many of the Livonian vowels such as \bar{a} , \bar{o} , \acute{o} and consonants l , k and η are simply impossible to type.

It is evident that for any revitalization and documentation efforts, it is important to make it possible to type the language on modern devices. We have, in collaboration with the language community, developed keyboard layouts for Windows, MacOS and Linux together with keyboard apps for iOS and Android.

2 Layout for computers

The keyboard layout was developed based on the wishes of the Livonian Institute², a national institute in Latvia that is dedicated to studying the Livonian language. The keyboard layout is based on the Latvian layout and it consists of two dead keys that provide access to the special characters in Livonian (see Figure 1).

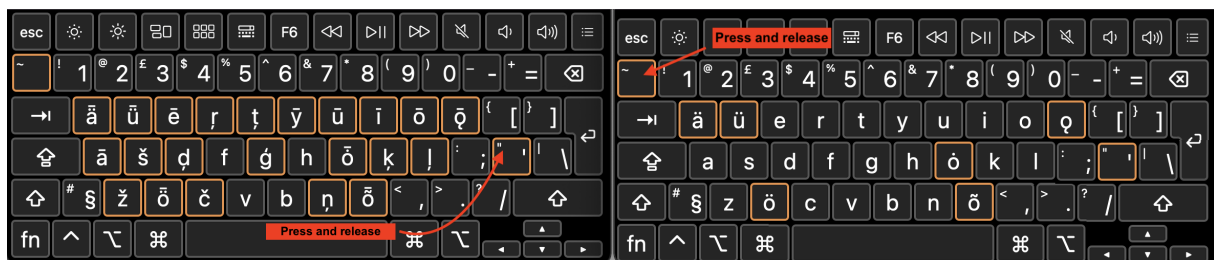


Figure 1: The layout on computers

¹Available here <https://rootroo.com/en/livonian-keyboard/>

²<http://www.livones.net/>

For the practical matter of installing the keyboard on a computer, we ran into an interesting problem that many endangered languages are bound to face. While MacOS accepted the keyboard to be installed under the Livonian language, Windows forced us to have our keyboard install itself as a Latvian keyboard. This is because Microsoft maintains their own list of numerical language codes³. This list does not recognize Livonian at all as a language.

3 Layout for phones

For Android and iOS (see Figure 2), we included the special characters under their unaccented counterparts. In some languages, such as Finnish, the accented letters appear as their own dedicated keys (*ä*, *ö* and *å*). However, this is not a practical solution for Livonian because the layout has altogether 28 different accented characters. What increases the number of characters in Livonian is the orthographic decision of marking long vowels with a macron (*ē* vs. *e*). As all vowels in Livonian can be made long, this means that macron can be added to all vowels, even the ones that already have an accent (*ō* vs *õ*).

| tierōdadrōt | | | tierābi | | | | tierāviļa | | |
|-------------|---|------|---------|---|---|---|-----------|---|---|
| q | w | e | r | t | y | u | i | o | p |
| a | s | d | f | g | h | j | k | l | |
| ⇧ | z | x | c | v | b | n | m | ⌘ | |
| 123 | 🌐 | vait | , | . | ↵ | | | | |

Figure 2: The keyboard on phones

As opposed to the computer layouts, the mobile keyboards need to be able to predict text. Currently, we use the Livonian words from the Finnish-Livonian dictionary⁴ to predict text. This is not an optimal solution as we can only recommend lemmas as opposed to inflectional forms. However, inflectional forms are problematic from the point of view of normativity. Even if we had access to data with inflections, not all inflectional forms have been studied well enough to derive a normative recommendation for them despite of the relatively extensive linguistic research on Livonian (cf. (Tuisk and Pajusalu, 2018)). Naturally, there are several dialectal ways of inflecting words, but the normative forms are yet to be established by the Livonian Institute.

Recommending forms that are dialectal or erroneous can be harmful for the language in the long run (cf. (Zeps, 1974)). This is a serious problem because a great part of the Livonian speaking community is not fluent in the language in the L1 level. This means that using a corpus-driven approach to derive the recommendations might result in non-normative or "bad" Livonian (cf. (Hämäläinen, 2021)).

4 Conclusions and future work

In this paper, we have described our practical work in building the most essential tool for any language in the modern era, namely a keyboard. We have described some of the challenges that are hindering the development such as the lack of recognition of Livonian by large companies and the reliability of any existing corpora as being representative of "good" or normative Livonian. Also, the language is under a process of getting normative forms, which means that currently it is not always clear what should be considered "good" and "bad" Livonian.

Our immediate future direction is to develop spell checkers in close collaboration with the language community. Although, it is to be remembered that the spell checkers can only recommend corrections for the forms that have already been established in the normative language. There is an extensive work on rule-based morphology for Livonian that has been conducted in the past (Rueter, 2014). We plan to build the spell checking using this resource as a starting point.

³https://docs.microsoft.com/en-us/openspecs/windows_protocols/ms-icid/70feba9f-294e-491e-b6eb-56532684c37f

⁴<https://gtsvn.uit.no/langtech/trunk/words/dicts/finliv/>

Acknowledgments

This work on Livonian keyboards was funded by the Livonian Institute. We would like to thank everyone who attended the Livonian Summer University in 2021 for their valuable feedback. We would like to acknowledge the hard work Valts Ernštreits is putting to building resources for Livonian.

References

- Valts Ernštreits. 2007. Livonian orthography. *Linguistica Uralica*, 43(1):11–22.
- Valts Ernštreits. 2019. Electronical resources for livonian. In *Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages*, pages 184–191.
- Mika Hämmäläinen. 2021. Endangered languages are not low-resourced! *Multilingual Facilitation*. Rootroo Ltd.
- Christopher Moseley, editor. 2010. *Atlas of the World's Languages in Danger*. UNESCO Publishing, 3rd edition. Online version: <http://www.unesco.org/languages-atlas/>.
- Sjur N Moshagen, Trond Trosterud, and Lene Antonsen. 2019. Language technology for indigenous languages: Achievements and challenges. In *Proceedings of the Language Technologies for All (LT4All)*.
- Jack Rueter. 2014. The livonian-estonian-latvian dictionary as a threshold to the era of language technological applications. *Journal of Estonian and Finno-Ugric Linguistics*, 5(1):251–259. ESUKA – JEFUL 2013, 5–1: 253–261.
- Tuuli Tuisk and Karl Pajusalu. 2018. A journey of discovery through the livonian world: the scholarly work of tiit-rein viitso. *Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics*, 9(2):13–34.
- Valdis J Zeps. 1974. A misleading example from livonian. *International Journal of American Linguistics*, 40(2):140–141.