Chapter 26
Language Report Luxembourgish

Dimitra Anastasiou

Abstract The Grand Duchy of Luxembourg is a small and multilingual country. The national language is Luxembourgish, and the legislative language is French. French, German and Luxembourgish are the three administrative and judicial languages. There are about 650,000 inhabitants and the majority of Luxembourgers speak four languages. As of March 2021, there were 59,000 Wikipedia articles written in Luxembourgish. Luxembourgish is very under-resourced when it comes to data resources and tools. This chapter provides a brief overview of the current level of support that Luxembourgish receives through technology (Anastasiou 2022).

1 The Luxembourgish Language

Luxembourg is a very small, but highly multilingual country. At various times, it was part of different European empires. Today, Luxembourg is the third European capital, along with Brussels and Strasbourg. It has the honour of hosting many of the EU's important institutions, including the Publications Office of the EU, the Directorate-General for Translation, and the Translation Centre for the Bodies of the EU.

As for the population of Luxembourg, the Statistics portal of the Grand Duchy of Luxembourg (STATEC) published a demographic atlas in 2019. According to this atlas, between 1981 and 2018, the Luxembourgish population increased by about 65%, from 364,597 to 602,005. There are 12 officially declared towns and 102 municipalities. Luxembourg City has the highest percentage of foreigners with 70.8%.

The languages spoken vary depending on the social situation or region. The regions with the highest density of Luxembourgish speakers include the north (85%) and the east (81%) of the country. According to STATEC, three out of four residents work in a multilingual environment and 25% of the population has to speak four or more languages at work. French is the most spoken language at work (78%), followed by English (51%) and Luxembourgish (48%). Luxembourgish is the most widely spoken language at home (53%), followed by French (32%) and Portuguese

Dimitra Anastasiou

Luxembourg Institute of Science and Technology, Luxembourg, dimitra.anastasiou@list.lu

(19%). It should be noted that Luxembourgish is not an official language of the EU. The Luxembourgish language is a Moselle-Franconian dialect, which was historically the mainly spoken language up to the 19th century in Luxembourg. On 24 February 1984, a law was enacted which made Luxembourgish an officially recognised language. In September 2018, the law was amended to add German sign language as an official language of Luxembourg. According to the provisions of the Languages Law of 1984, French, German or Luxembourgish may be used in administrative and judicial matters. Citizens can interact with the administration in any of these three languages, and officials must attempt 'as far as possible' to respond in the language used by the applicant. Legislative documents are written in French and an important consequence of this on the judicial level is that only the text in French is deemed authentic for all levels of public administration.

In terms of vocabulary, Luxembourgish has a substantial number of loan words from French and German, but its morpho-syntax follows Germanic patterns (Gilles and Trouvain 2013). With the exception of the alveolo-palatal fricatives and the approximant [w], the consonant inventory of Luxembourgish is quite similar to Standard German. In addition, Luxembourgish has a set of eight diphthongs, which is considerably larger than for Standard German which has just one (Gilles and Trouvain 2013). Gilles (2019) examines the complex language situation of Luxembourg. There is an officially recognised system with regards to the orthography of Luxembourgish, called "OLO" (ofizjel lezebuurjer ortografi); it can be found at the Zenter fir d'Lëtzebuerger Sprooch (ZLS)/Centre for the Luxembourgish Language.[1]

2 Technologies and Resources for Luxembourgish

Luxembourgish-specific tools include a grapheme-to-phoneme conversion for Luxembourgish based on 30,000 manually phonetically transcribed words, two spell-checkers, a PoS-tagger (including a tokenizer and lemmatizer), and sentence splitter (Sirajzade and Schommer 2019), and a mobile application called Schnëssen.[2] This crowdsourcing app collects data on the present-day language situation of Luxembourgish; users can participate in a large set of audio recordings tasks and in sociolinguistic surveys. A recently published tool, LëtzRead,[3] is a free browser extension to integrate Luxembourgish-learning just by browsing the web (displaying certain words in Luxembourgish). Moreover, The library spaCy for advanced NLP has been trained for Luxembourgish, and the text-to-speech (TTS) tool MaryTTS has also been extended to support Luxembourgish. Luxembourgish data resources are mainly monolingual corpora, but there is also a Luxembourgish COVID glossary as well as an orthography trainer. The biggest text corpus in Luxembourgish contains 170 million words from a wide range of genres (Parliamentary debates, lit-

---

[1] https://portal.education.lu/zls/ORTHOGRAFIE

[2] https://infolux.uni.lu/schnessen/

[3] https://www.letzread.com

erature, transcripts of conversations, and media texts including articles from news outlets like RTL.lu, radio100,7, eldoradio, and social media). All texts are annotated and orthographically normalised. This corpus is owned by the University of Luxembourg and is for internal use only. Many lexical Luxembourgish-specific resources, including corpora, dictionaries, material for phonetics, applications, etc. are available at Infolux,[4] which is the research portal about Luxembourgish developed and maintained by the Institute of the Luxembourgish Language and Literature at the University of Luxembourg. Another important resource is the Luxembourgish Online Dictionary (LOD),[5] managed by the ZLS, a multilingual dictionary with 30,000 entries, in which Luxembourgish words are translated into German, French, English and Portuguese and illustrated by examples.

Among the recent research projects related to language technology (LT) including Luxembourgish are ENRICH4ALL, STRIPS, and Lingscape. ENRICH4ALL (E-goverNment [RI] CHatbot for ALL)[6] is a CEF-funded project (06/21-05/23) coordinated by the Luxembourg Institute of Science and Technology, and its objective is to have a multilingual chatbot through integrating the CEF AT core service platform eTranslation to existing AI-based chatbot technology. The chatbot service will be deployed in public administration in Luxembourg, Denmark, and Romania. STRIPS[7] was a three-year project (02/18-01/21), funded by the University of Luxembourg, that aimed to develop a semantic search toolbox for the retrieval of similar patterns in documents written in Luxembourgish. Lingscape[8] is a mobile application researching linguistic landscapes all over the world by collecting photos of signs and lettering on an interactive map.

3 Recommendations and Next Steps

Digitalisation plays a big role in the government of Luxembourg, the Ministry for Digitalisation was created on 11th December 2018. Luxembourg's national AI Vision initiative underlines the country's unique ability to become a living lab of real-world AI applications.

Mainly because of the lack of underlying data resources, there are gaps in many aspects of Luxembourgish Language Technology. What is currently missing are available bilingual corpora, e. g., Luxembourgish – English, German, French. The availability of such data sets would facilitate the development of many LT applications, such as named entity recognition, machine translation, virtual agents, recommender systems, etc. All of these applications are mainly statistically-based, so typically require a large amount of manually annotated training data. Regarding language mod-

---

[4] https://infolux.uni.lu

[5] https://www.lod.lu

[6] https://www.enrich4all.eu

[7] https://acc.uni.lu/Research/strips/

[8] https://lingscape.uni.lu

els which can be used for natural language understanding and generation, the multilingual BERT covers many languages, including Luxembourgish; however, a BERT model trained specifically on large Luxembourgish data would yield better results. Another important aspect is that written Luxembourgish is not well standardised; while both German and French are intensively taught in schools, Luxembourgish, although the first language of around 60% of the population, forms part of the school curriculum only rudimentarily. This has an impact on the correctness of Luxembourgish in the development of LT applications. It is noteworthy that Luxembourgish has become more important in secondary schools with changes incorporated for the academic school years 2021-2022 and 2022-2023.

Luxembourg needs united forces for efficient collaboration. Since most people are multilingual, various stakeholders do not see the need to invest either time or budget in creating or sharing Luxembourgish resources. The EU, the government, research institutions, and language service providers have to work together to achieve the desired results. Important action points to improve the Luxembourgish LT landscape are: 1. reaching a status that Luxembourgish can be used in many administrative procedures; 2. raising awareness among various stakeholders in public and private sectors about the impact of Luxembourgish data; 3. advancing the standardisation, use and study of Luxembourgish, and 4. dedicated and collaborative national and EU funding programmes for both basic and applied research on Luxembourgish.

References

Anastasiou, Dimitra (2022). *Deliverable D1.24 Report on the Luxembourgish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-langu age-equality.eu/reports/language-report-luxembourgish.pdf.

Gilles, Peter (2019). "39. Komplexe Überdachung II: Luxemburg. Die Genese Einer Neuen Nationalsprache". In: *Deutsch*. De Gruyter Mouton, pp. 1039–1060.

Gilles, Peter and Jürgen Trouvain (2013). "Luxembourgish". In: *Journal of the International Phonetic Association* 43.1, pp. 67–74. DOI: 10.1017/S0025100312000278.

Sirajzade, Joshgun and Christoph Schommer (2019). "The LuNa Open Toolbox for the Luxembourgish Language". In: *Advances in Data Mining, Applications and Theoretical Aspects, Poster Proceedings 2019*. Ibai publishing.