

The “Historical Luxembourgish Bilingual Public Notices Database”¹

Peter Gilles / Evelyn Ziegler

Abstract

Bilingual parallel corpora have been gaining increasing importance as a solid basis for contrastive linguistics both from a synchronic and diachronic perspective. The “Historical Luxembourgish Bilingual Public Notices Database” presents a diachronic single genre corpus and stores French-German parallel texts covering the years from 1795 to 1920. This paper gives an overview of the underlying existent text-corpus, specifies the features of the genre “public notices” and explains the criteria for text selection. Building upon that, the compilation and presentation of text data and image data stored in the “Historical Luxembourgish Bilingual Public Notices Database” are detailed. Finally, the technical tools for indexing, searching and managing the text and image data are described.

1. Introduction

The *Archive of the City of Luxembourg* hosts a large number (around 7000) of public notices, i.e. printed proclamations presented to the public, between the end of the 18th and the beginning of the 20th century. These large-scale printed public announcements served as a central means of information for the city administration during that time, e.g. to communicate laws, regulations, organizational matters etc., concerning the public life of the city of Luxembourg. As most of the public notices are written in two languages (French and German) they form an ideal basis² for a large parallel corpus. Parallel corpora are defined as follows: “Parallel corpora consist of a source text and its translation into one or more languages” (Aijmer 2008: 276). They guarantee comparability, but also allow studying monolingual features and developments. However, parallel corpora are still a desideratum in historical linguistics, as Claridge (2008: 256) points out: “from a cultural-historical perspective, it would be interesting to link several European

¹ This paper was made possible by a project funded by the *Fonds National de la Recherche*, Luxembourg.

² We are very grateful to Dr. Evamarie Bange/Archive of the City of Luxembourg, for her permission to use the Affichen-Corpus.

languages in a corpus of parallel texts or translations”.

The corpus of the “Historical Luxembourgish Bilingual Public Notices Database” provides a resource for studies in contrastive linguistics (e.g. similarities and differences concerning information structure phenomena, lexicography, metaphorical conceptualizations, variation and change of features and constructions in the process of standardization) and contact linguistics (transfers of lexical, grammatical and orthographic features), as both languages played (and still play) a long and important role in the history of Luxembourg.

The selection of the genre “public notice” (see example below) is driven by several facts:

1. The text genre represents a central top-down strategy of information policy carried out by the municipality of Luxembourg. Therefore, these culturally significant documents offer new opportunities for linguistic inquiry into strategies of public language use at the interface of standard language and technical language use.
2. The genre is a quite unique genre, which has so far not been considered as a primary research source in linguistics, political science, historical, legal or cultural studies.
3. A single genre corpus opens up the opportunity to study the relation between language and genre and how a single genre is affected by language change over time (macro-text level approaches, micro-text level approaches).

As mentioned above, the genre “public notice” formed a central means of “top-down” communication³ in the regulatory discourse⁴ of the municipality of Luxembourg City during the 18th and 19th century. In a socio-cultural perspective, the development of the genre “public notice” is closely tied to a development that has been referred to as the “linguistic colonization of the public sphere” (“Die Kolonialisierung des öffentlichen Raums durch die Schrift”, Auer 2010: 295). Hereby Auer means that the producers of signs displayed in the public sphere exert agentivity and power, the more so as not every member of society is granted the authority and right to do so. Governmental and administrative institutions considered the emerging

³ The concept of “top down” communication goes back to Ben-Rafael et al. (2006). For a critical discussion of the concept see Auer (2010).

⁴ The concept of “regulatory discourse” goes back to Scollon/Scollon (2003).

public sphere as an arena to display and exercise power. One way of doing so was the posting of public notices at the town hall or other prominent places, which provided general public access (e.g. church doors). Public notices were used as an instrument of social control, i.e. they aimed at civil obedience and order of society. As Rickards (1973: 7) puts it: “the printed public announcement is essentially an instrument of public control. ... it represents an extension of the power of the ruler - authority mass-produced.” The authoritative gesture is expressed on the text-linguistic level in devices such as royal emblems (e.g. royal Luxembourg lion emblem), and naming of honorifics (e.g. Grand-Duché du Luxembourg).

2. Development and characteristics of the text genre “public notice”

The history of the genre begins with orders that were proclaimed for the illiterate by a town crier. Rickards (1973: 8) describes the presentation as follows: “Summoned by drum or trumpet to the market-place, the people saw the actual document unrolled ... and heard its oracular pronouncement.” In those days the production of a public notice was an individual effort and documents with the hand-written orders of monarchs are typical examples.

Due to increasing “pedagogicalization” (Mattheier 2003) and “popularization” of writing and reading as well as technical innovations such as mass printing and colour typography, town criers were no longer needed and the public notice as an efficient means of information management rose to general importance, contributing to what is called in modern sociolinguistic terms the “linguistic landscape” of urban spaces (Shohamy et al. 2010). As a consequence, the production of public notices became a process demanding a number of people involved in producing the text (composition, revision, error correction, layout design, and printing).

The genre “public notice” can be described according to text external and text internal characteristics (cf. Esser 2009). The defining text external characteristics are the following:

- Mode of visual communication,
- Written to be viewed at close range, i.e. for “targeted perception”,
- Temporarily exhibited documents in places where the public is likely to come to its notice,

- Primary text functions/intentions: information about government and municipal activities and appellation directed at the citizens (instruction, order, prohibition),
- Text producers: mayors of Luxembourg City and/or their secretaries,
- Text recipients: inhabitants of Luxembourg City.

The defining text internal criteria are the following:

- Wide range of themes: health, education, election, public affairs, and commerce,
- Mostly bilingual documents written in French and German,
- Use of language for specific purposes (stylistic features, technical lexis) and special notice vocabulary,
- Typographic distinction between French and German through the use of Roman and Gothic (black letter) fonts,
- Variation of layout and typography:
 - alternating horizontal and vertical text vector, changing special notice typography, use of national emblems according to the political changes,
 - professionalization of layout and typography over time, e.g. use of display typefaces from the early 1800s,
- Emergence of a global macro-structure, specialization and differentiation of the genre over time,
- Size of text length varies between 700 and 51,000 characters.

Although the names of the writers or those in charge of the texts are generally recorded at the bottom of the documents, nothing is known about the language biographies of the writers or translators. Therefore, the question must remain open whether there was one single bilingual writer or two and if so, in which direction the translation went. Only for documents published between 1795 and 1814 do contextualization cues in the German text column make it possible to determine the direction of translation. Among these contextualization cues are dates referring to the *French Revolutionary Calendar*⁵ or references to Luxembourg as *Département des Forêts*, which indicate that the direction of translation went from French to German. In fact, from the current point of analysis it seems obvious that hardly one single author or one single writer could be identified. Rather, for the

⁵ The *French Revolutionary Calendar*, invented during the French Revolution, was in use in France from 1793 until 1805.

production of these texts, several stages of conceptualization and the deployment of prefabricated text modules and routines, with several writers involved, have to be assumed.

3. Compilation of the Corpus

The “Historical Luxembourgish Bilingual Public Notices Database“ draws on a single-genre corpus, which consists in its entirety of almost 7000 documents stored in the *Archive of the City of Luxembourg*. This corpus covers the years 1795 to 1920. As the corpus compilation for the database was mainly driven by the intention of allowing a fairly broad scope of research interests, the start and end dates of the corpus are defined by socio-historical landmarks rather than by linguistic criteria: namely, the French annexation of Luxembourg and the end of the First World War. Thus, the start and end dates not only cover a long period (the so called “long 19th century”), but they also link up to the traditional periodization of the history of Luxembourg including important times of transition.

Nevertheless, sociolinguistic criteria are also taken into account as the popularization of the text genre “public notice” in Luxembourg is most closely connected to the effects of the French Revolution, in particular the introduction of democratic ideas, such as public information and participation in decision-making concerning public affairs. Public notices were used to communicate transparency for those who wanted to know more about government and municipal activities. In this vein, public notices were not only an instrument to control the public, but also an instrument to allow the public to form their opinions on actions taken by the administration (e.g. estate actions, plans for road building etc.).

The production and publication of the public notices is embedded in a changing socio-historical context. To allow investigation of the relationship between linguistic choices and socio-cultural contexts, the corpus is subdivided according to the established periodization of the history of Luxembourg. The content of the corpus encompasses four periods based on important landmarks in the history of Luxembourg (Period LU II: 1795-1814, Period LU III: 1815-1843, Period LU IV, 1: 1844-1890, Period LU IV, 2: 1891-1920). From 1795 to 1814, Luxembourg was under French rule and became part of the *Départements des Forêts*. In 1815, the Congress of Vienna declared Luxembourg a Grand Duchy and allocated it as personal property to William I, king of the Netherlands. Luxembourg remained under

the sovereignty of the Orange-Nassau dynasty until 1890, the year of the death of William III. Between 1815 and 1867, it was also member of the German Confederation (*Deutscher Bund*) and the City of Luxembourg was a confederate fortress with a Prussian garrison. During the Belgian Revolution (1830-1839), Luxembourg was divided and lost half of its territory. The Treaty of London (1839) established Luxembourg's present-day borders and ruled that the western, francophone part should go to Belgium, while the eastern, germanophone part⁶ remained to constitute the Grand-Duchy. In the same year, Luxembourg was granted more independence and became a nation state. With the foundation of the nation state, Luxembourg was confined to the germanophone territory. Nevertheless, French and German continued to be used as national languages and as administrative languages (see Fehlen 2009, Gilles/Moulin 2003, Weber 2000).

4. Structure of the database

The database comprises 2000 documents. It can be regarded as representative and sufficient for sociolinguistic research⁷, e.g. for documenting and analyzing language management, language policy, language change and language contact. The database allows for typological and genetic approaches as well as quantitative analysis for less frequent linguistic features and structures. According to standard corpus definitions, the database falls in the category of a systematically compiled corpus. The following selection parameters were chosen to compile the corpus:

- Language (only bilingual French-German documents were chosen)
- Topic (health, education, public affairs, commerce)
- Period of time (1795-1814, 1815-1843, 1844-1890, 1891-1920)
- Quantity (approx. 500 documents per period)

⁶ The concepts of “francophone” and “germanophone” are used to refer to a spectrum of varieties (standard and non-standard) realized in the speech communities. Furthermore, the concept of “germanophone” is intended as a superordinate concept including German and emerging Luxembourgish. It would not be sufficient to conceive of the germanophone part as a “German speaking” part. Finally, it should be pointed out that the distinction between a francophone and a germanophone part is not meant as a clear cut distinction between homogeneous entities but as a distinction that includes cultural diversity.

⁷ For a detailed discussion of sociolinguistic research perspectives in corpus linguistics see Romaine (2008) and Meir (2009), of the use of corpus linguistics to study language change see Curzan (2009).

As an equal number of texts per period were included, the internal composition of the corpus is balanced and allows synchronic and diachronic studies as well as inter-period cross-linguistic comparisons.

In a first step, the selection of 2000 public notices, with their sizes ranging between letter format and DIN A0 (approx. 800 x 1200 mm), have been scanned with a large-scale, high-definition scanner. Several of them consist of more than one page. Information about these image files along with their metadata (shelf mark, date, title, content, issuing authority, used languages) is stored in a relational database (MySQL). Through the web-based frontend, all this information is retrievable through searching facilities. Figure 1 provides an overview of the frontend. In this case the database is filtered to show only those public notices of the year 1899.

----- place figure 1 here -----

When a specific public notice is selected, users will get access to the full-size image, where they can zoom and pan into sections (figure 2). Technically the image player draws on functionality provided by the Zoomify function of the OpenLayers software package.⁸ This image viewer

⁸ <http://openlayers.org>.

is especially useful to navigate in large-size documents with varying font sizes.

----- place figure 2 here -----

Besides the scanned image, the user has also access to the full text of the public notice. All documents have been text-digitized by a company specializing in the text-digitization of old documents. Instead of using OCR techniques, which are known to be especially error-prone for old documents, the more traditional way of typing all documents by hand was chosen. In fact, all documents were double-keyed and automatically compared for discrepancies. This technique guarantees extremely low error rates (Büdenbender 2011). In a next step, this raw text data is transformed into XML-format. The format is in accordance with the principles of the Text Encoding Initiative (TEI) to guarantee sustainable data exchange with similar research projects or software. In order to keep as much as possible of the original document structure, the full text contains extensive tagging. This comprises tagging of the different kinds of headings, the switch between font-face and font-type (French text is printed in Roman, while German is printed in Gothic) and the differentiation between sub- and superscript text. As most documents are bilingual, it is of special importance that the different languages (mostly French and German) can be distinguished in the full-text data. Since the XML conversion has not been completed yet, it is not possible to provide definite figures about the size of the corpus. As the

average word count for one public notice lies at around 800 words, it can be projected that the final corpus will contain approx. 2 million running words.

The full text of all public notices is further available for concordance and text retrieval software to facilitate linguistic analyses. The next figure shows the result of a pattern search using the tICorpus programme.⁹ This corpus organization and search software utilizes regular expressions, and this allows finding nearly all the desired text patterns. In the example given in Figure 3, ‘.*ntern’ was searched for, matching all words (i.e. strings between spaces) ending in the string ‘ntern.’ The list with the search results shows the search word in the centre, with an amount of context on the right and on the left (so called ‘Key-word-in-context’ concordance). Additionally, the name of the document, where the search term has been found, is indicated on the very left.

---- place figure 3 here ----

As most of the public notices are bilingual, it is of interest to see how certain elements in one language are replicated in the other. Because of the tagging of the corpus, the immediate link between the two languages can easily be established. This interlinked corpus structure allows finding equivalent

⁹ <http://tshwanedje.com/corpus/>

words/structures in the two languages quite easily. Due to the fine-grained structure of the corpus, it will be possible to analyze the evolution of certain linguistic features, especially in the domain of language contact over the time-span of the available data.

5. Conclusion

The aim of this paper was to present the structure of an extensive corpus of a hitherto largely neglected text genre in studies on language history, i.e. historical public notices of the city of Luxembourg, which is currently being compiled at the Universities of Luxembourg and Duisburg-Essen. It comprises the period of time from the end of the 18th to the beginning of the 20th century and thus covers the pivotal period for establishing the language of administration and the development of official communication in the public sphere. Moreover, the corpus is a parallel corpus as most of the public notices are bilingual (German/French). The data are available as an image database to study the visual arrangement and layout (e.g., the use of emblems, handwritten additions) and as a searchable full text database, giving the user a flexible tool to analyze the data.

With about 2 million words of running text, this corpus permits studying several aspects of language use in the administration of the city of Luxembourg: In terms of sociolinguistic research, questions of language policies and language standardization in the context of multilingualism can be addressed. Linguistically, the various aspects of language contact are especially promising. For most of the public notices, the direction of translation seems to be from French to German and the (mostly lexical and syntactical) transfers from French can be studied in situ in the German version.¹⁰

6. References

Aijmer, Karin (2008): Parallel and comparable corpora. In: Lüdeling, Anke/Kytö, Merja (eds.): *Corpus Linguistics. An International Handbook*. Vol.1. Berlin: de Gruyter, 275–292.

Auer, Peter (2010): Sprachliche Landschaften. Die Strukturierung des öffentlichen Raums durch die Schrift. In: Deppermann, Arnulf/Linke,

¹⁰ For first analyses of the data see Ziegler (2011).

Angelika (eds.): Sprache intermedial. Stimme und Schrift, Bild und Ton. Berlin: de Gruyter, 271–298.

Ben-Rafael, Eliezer/Shohamy, Elana/Amara, Muhammad Hasan/Trumper-Hecht, Nira (2006): Linguistic landscape as symbolic construction of the public space: The case of Israel. In: Gorter, Durk (ed.): Linguistic Landscape: A New Approach to Multilingualism. Clevedon: Multilingual Matters, 7-30.

Büdenbender, Stefan (2011): LexicoLux: EDV-philologische Perspektiven bei der Erstellung eines Wörterbuchnetzes der Grossregion. In: Gilles, Peter/Wagner, Melanie (eds.): Linguistische und soziolinguistische Bausteine der Luxemburgistik, Mikrolottika. Minority language studies. Frankfurt a.M. [et al]: Peter Lang, 261–275.

Claridge, Claudia (2008): Historical Corpora. In: Lüdeling, Anke/Kytö, Merja (eds.): Corpus Linguistics. An International Handbook. Vol.1. Berlin: de Gruyter, 242–259.

Curzan, Anne (2009): Historical corpus linguistics and evidence of language change. In: Lüdeling, Anke/Kytö, Merja (eds.): Corpus Linguistics. An International Handbook. Vol. 2. Berlin: de Gruyter, 1091–1108.

Esser, Jürgen (2009): Introduction to English Text-linguistics. Frankfurt a.M.: Peter Lang.

Fehlen, Fernand (2009): BalaineBis. Une enquête sur une marché linguistique multilingue en profonde mutation. Luxemburgs Sprachenmarkt im Wandel. (Recherche Etude Documentation No 12). Luxembourg: SESOPI Centre Intercommunautaire.

Gilles, Peter/Moulin, Claudine (2003): Luxembourgish. In: Deumert, Ana/Vandenbussche, Wim (eds.): Germanic Standardizations. Past and Present. Amsterdam: Benjamins, 303–329.

Mair, Christian (2009): Corpus Linguistics Meets Sociolinguistics: the Role of Corpus Evidence in the Study of Sociolinguistic Variation and Change. In: Renouf, Antoinette/Kehoe, Andrew (eds.): Corpus Linguistics: Refinements and Reassessments – Proceedings of the 2007 ICAME Conference – Stratford-upon-Avon. Amsterdam: Rodopi, 1–26.

Mattheier, Klaus (2003): German. In: Deumert, Ana/Vandenbussche, Wim (eds.): Germanic Standardizations. Past to Present. Philadelphia. Benjamins, 211-244.

Rickards, Maurice (1973): The Public Notice. An Illustrated History. New York: Potter.

Romaine, Suzanne (2008): Corpus linguistics and sociolinguistics. In: Lüdeling, Anke/Kytö, Merja (eds.): Corpus Linguistics. An International Handbook. Vol. 1. Berlin: de Gruyter, 96–111.

Scollon, Ron/Scollon, Suzie Wong (2003): Discourses in Place. Language in the material world. New York: Routledge.

Shohamy, Elana/Ben-Rafael, Eliezer/Barni, Monica (eds.) (2010): Linguistic Landscape in the City. Bristol: Multilingual Matters.

Weber, Nico (2000): Multilingualism and Language Policy in Luxembourg. In: Deprez, Kas (ed.): Multilingualism and government. Belgium, Luxembourg, Switzerland, Former Yugoslavia, South Africa. Pretoria: Van Schaik, 82-91.

Ziegler, Evelyn (2011): Sprachenpolitik und Sprachenmanagement in Luxemburg (1795-1920). In: Gilles, Peter/Wagner, Melanie (eds.): Linguistische und soziolinguistische Bausteine der Luxemburgistik. [=Mikroglottika. Minority language studies 4]. Frankfurt a.M. [et al.]: Peter Lang, 177-203.

Figures (use separate files delivered with this article)

Thumb	No.	Date	Title	Content	Author	Language
		1899				
	LU Imp. IV.2_0052	1899 Janvier 26.	Arrêté concernant la police pendant le carnaval		collège échevinal	allemand/français
	LU Imp. IV.2_0151	1899 Octobre 25.	Kandidatenliste für die Gemeinderathswahl vom Dienstag den 31. Oktober 1899		MOUSEL E., président bureau de vote ppal	allemand
	LU Imp. IV.2_0254	1899 Juillet 15.	Anniversaire de la Naissance de S.A.R. Monseigneur le Grand-Duc	Anniversaire de la Naissance de S.A.R. Monsieur le Grand-Duc		allemand/français
	LU Imp. IV.2_0261	1899 octobre 20.	Bekanntmachung	Bekanntmachung: Wahl von 7 Mitgliedern für den Gemeinderat		allemand
	LU Imp. IV.2_0293	1899 Janvier 26.	Arrêté concernant la police pendant le carnaval		MOUSEL E., président; FABER J., secrétaire	allemand/français

Figure 1 Database frontend of the database for public notices

Affiche Signatur LU Imp. III_1303 - -



Figure 2 Detailed view on a public notice (with zoom and pan function)

Line	Text	Context
112	Das Gesetz vom 9ten Vendemiaire 4ten Jahrs vereinigtu	Departementern
113	in Republik und hauptsächlich in den neun vereinigten	Departementern
114	om 12ten September 1791, welcher in den vereinigten	Departementern
115	„ dieß noch nicht geschehen wäre, in den vereinigten	Departementern
116	„ Gesetz vom 9ten Vendemiaire 4ten Jahrs vereinigtu	Departementern
117	ich königliche Meuchelmörderbände in verschiedenen	Departementern
118	nen wegen jenen Schäden, so in det Gränzund West =	Departementern
119	iten Artikel des obangezogenen Gesetzes in besagten	Departementern
120	die neun am 9ten Vendemiaire 4ten Jahrs vereinigten	Departementern
121	die neun am 9ten Vendemiaire 4ten Jahrs vereinigten	Departementern
122	par extraits et affichés dans les lieux indiqués nelle.	Mantern
123	alten. Franken, bereits scheinen in einigen West- und	tags=Departement
124	ie sich zum geistlichen Stande bestimmen. 4o. Die bei	Regimentern
125	ntimeter hervorragen, wenn die Schornsteine sich am	untern
126	nen zu fügen. Der obere Rand dieser Röhren muß den	untern
127	Eingesehen den Brief des einstweiligen Jury's für den	untern
128	Samstags, den 28ten um 9 Uhr des Vormittags die der	untern
129	en in drei Klassen eingetheilt, so daß die Zöglinge der	untern

Figure 3 Example for keyword-in-context (KWIC) concordance of the full-text data of the database.