Explore Intra-Belgian book translation flows with a SAMPO-UI

Sven Lieber[1,*], Ann Van Camp[1], Joren Six[2] and Julie M. Birkholz[1,2]

[1]KBR -Royal Library of Belgium, Keizerslaan 4, 1000 Brussels, Belgium
[2]Department of History & Ghent Centre for Digital Humanities, Ghent Univerisity, Sint-Pietersnieuwstraat 33, 9000 Ghent, Belgium

Abstract

In the field of Big Translation History, (historical) book translation flows are studied based on publication metadata. This research data is often kept internally or only provided as domain-specific MARC21 records, a standard library format. Within the recently completed BELTRANS project, we created a large integrated data corpus of worldwide translations between Dutch and French between 1970 and 2020, where at least one Belgian was involved as author, scenarist, illustrator or editor, to study intra-belgian book translation flows, and published the corpus as FAIR data. In this paper we present the SAMPO-UI-based platform that can be used by non-technical users to explore the underlying BELTRANS Linked Data corpus. The off-the-shelf software SAMPO-UI was configured to build a user interface allowing the exploration of our data related to book publications, authors, publishers and linked temporal and geospatial information. The different search perspectives, filters and visualizations enable researchers to explore the corpus or download self-chosen subsets, without having to process the large underlying data dumps. Both the application and corpus data are opened up in a sustainable way, i.e. besides deposits at research repositories, the application is hosted in the long term at the Royal Library of Belgium (KBR) and the integrated data is equipped with provenance information allowing to click-through to the related metadata records at different national libraries. Future work can extend the application itself with novel visualization techniques or different download formats if necessary. Also, this application can serve as a role model to inspire other researchers to distribute their corpora as Linked Data and via such a SAMPO-UI platform.

Keywords

Big Translation History, Semantic Portal, Library Science

1. Introduction

Literature serves, among its many purposes, as a mean to convene stories, its translations to break down language barriers and bridge cultures. Translation scholars largely focus on the content of these translated texts and less on the volume and trends. Recently, the term Big Translation History (BTH) was proposed, defined as a translation history that can be analyzed computationally [1] to aid us in understanding historical patterns of translations [2].

A crucial aspect in BTH is the urgent need for interdisciplinarity and data sharing, covering also sustainable infrastructures [1]. In this paper we focus on these needs based on the use case of the BELTRANS project. This project was launched to study cultural transfers via intra-Belgian book translation flows between Dutch and French in the period between 1970 and 2020: a period of increasing regionalization of the multilingual state of Belgium. A major outcome of this 4-year project is a semantically annotated corpus of book translation flows. Even though several interdisciplinary vocabularies were used to annotate the corpus, and the corpus was made available according to the FAIR [1] principles [3], effective data sharing, especially reuse is still hampered for non technical users. Thus, we explored

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium
*Corresponding author.
sven.lieber@kbr.be (S. Lieber); ann.vancamp@kbr.be (A. V. Camp); joren.six@ugent.be (J. Six); julie.birkholz@kbr.be/ugent.be (J. M. Birkholz)
https://sven-lieber.org/en (S. Lieber)
0000-0002-7304-3787 (S. Lieber); 0000-0002-1915-5956 (A. V. Camp); 0000-0001-7671-1907 (J. Six); 0000-0003-1193-0847 (J. M. Birkholz)

[1]Findable, Accessible, Interoperable and Reusable

the implementation of a graphical user interface for querying, processing and filtering these data, and hence addressing the BTH needs.

This paper, answering the demo session's call for *new publicly available datasets*, is centered around the following three resources advancing the state of the art in BTH:

- BELTRANS corpus: https://doi.org/10.5281/zenodo.20668316
- BELTRANS user interface (UI) configuration: https://github.com/kbrbe/beltrans-ui-config
- BELTRANS UI: https://w3id.org/kbr/beltrans/ui

In this demo paper, we present a user interface to explore the BELTRANS corpus. In section 2 we briefly discuss related work. In section 3 we elaborate on our architecture, in section 4 we describe the demonstration of our platform, and in section 5 we outline the contributions, limitations and highlight future work.

2. Related Work

A major target group of the BELTRANS corpus are scholars in the Humanities, and, thus, the starting point for our work is how non-technical users can make use of Linked Data to explore the corpus and perform research on translation flows specifically.

Different systems can be used to import the data and analyze it, including general tools like Heurist[2] or nodegoat[3], to import and process relational data, or domain-specific platforms like DLBT [4] to manage, store and organize digital data.

Whereas both general and domain-specific tooling can and may be used on top of the FAIR data, neither is suited to consume the already semantically enabled data of the BELTRANS corpus. Ikkala et al. [5] reviewed the current semantically enabled platforms with user interfaces and introduced SAMPO: a declarative, configuration-based frontend framework designed to make RDF data from SPARQL endpoints accessible via a UI [5]. It maintains a clean separation between data and view. Data lives behind a SPARQL endpoint and is consumed by a configurable front-end through application perspectives, which makes the same data reusable across portals and the same front-end system reusable across datasets [6]. It has been used in many SAMPO portals [7] and also has gained international attention [8, 9]. It's Technology Readiness Level (TRL) is thus arguably high and makes it a good candidate for our BTH use case BELTRANS.

3. Architecture

The proposed architecture is based on Linked Data principles, open source software and the SAMPO model [6]. The model proposes a clear separation between data publication via a SPARQL endpoint and data presentation via SAMPO-UI [10], a configurable interface to filter and explore data.

The BELTRANS architecture builds upon QLever [11] a triple store which also provides a SPARQL endpoint. The FAIR BELTRANS corpus is loaded into this triple store and can be consumed from the SAMPO-UI system. We use the SAMPO-UI fork provided by GhentCDH [4]. The fork provides updated dependencies, a new plug-in system, a clear separation between the SAMPO-UI core and application specific configuration and versioned releases of the SAMPO-UI core in the form of OCI containers. The fork is developed in close collaboration with the main SAMPO-UI team in Finland and is planned to flow into the SAMPO-UI main branch in the near future.

The SAMPO-UI platform is based on a server and a client component. The server provides an API based on SPARQL queries and the client is a React application that can be configured via JSON files.

[2]Heurist: https://heuristnetwork.org/
[3]nodegoat: https://nodegoat.net/
[4]SAMPO-UI fork: https://github.com/GhentCDH/sampo-ui

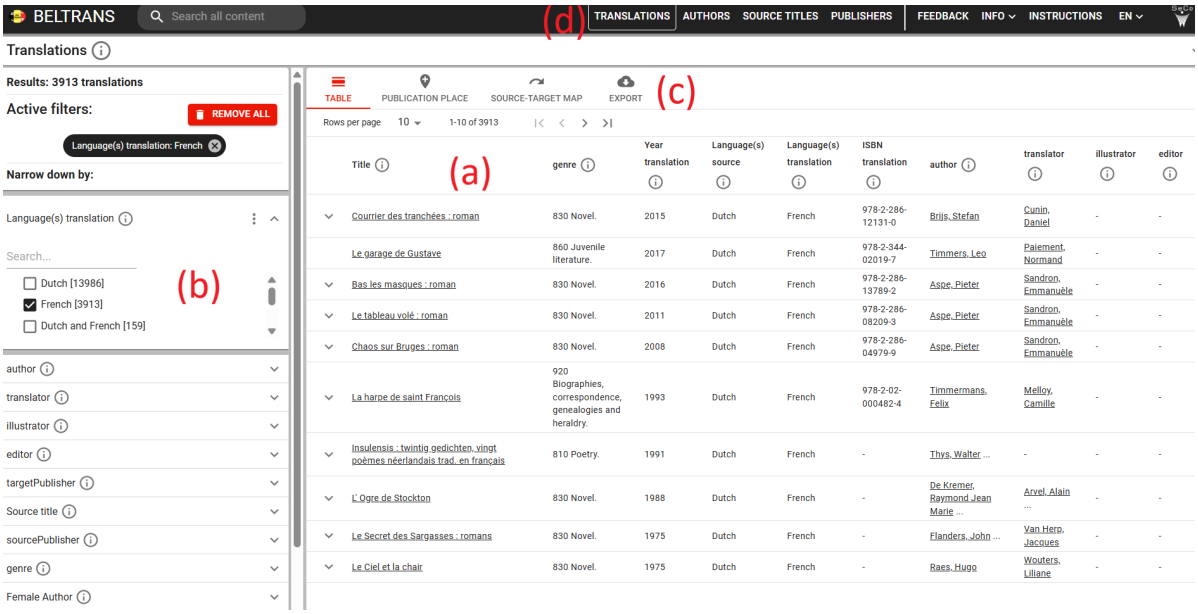

Figure 1: The tabular view of the BELTRANS SAMPO interface: (a) tabular view of selected data in the main panel; (b) configurable search facets in the left panel; (c) tabs to switch to between views; (d) navigation bar to switch to between perspectives

4. Demonstration Description

This section elaborates on the different possibilities to (visually) explore the BELTRANS corpus. First we briefly introduce the core concepts of SAMPO-UI and then we focus on four main interactions with the UI, accompanied by examples from BELTRANS.

A SAMPO-UI instance is centered around the principle of *search perspectives*, each perspective represents one entity type, e.g. a *book translation* or a *person*. The main panel either shows the instance page for such an entity or a view on all entities of a type (fig. 1, (a)), optionally filtered by search facets that are shown in the left panel (fig. 1, (b)). For the BELTRANS corpus we currently offer four search perspectives[5]: (i) Translations, (ii) Source titles, (iii) Persons, and (iv) Organizations. Each search perspective is configured with different JSON files and SPARQL queries.

4.1. Search and filter entities in a tabular view

When clicking on a perspective from the main homepage or via the navigation bar (fig. 1, (d)), the user is presented a tabular view of all entities of that perspective. With the faceted search on the left (fig. 1, (b)), users can filter based on different pre-configured criteria. It is configurable which data fields are shown in the tabular view, the instance page or whether and how the field should be used as search facet. This is a handy usability trade-off, possibly detailed information can be made available only via the instance page, without cluttering the main tabular or visualization view. Figure 1 shows the tabular view of the *Translations* search perspective as well as search facets such as year of publication, author, language, etc. on the left.

4.2. Traverse the graph: navigate between entities

Besides literal values such as *year of publication*, columns may contain links to other entities. Those entities are either from a different search perspective or are externally managed entities from the web. A click on an author in the *Translations* perspective leads to the related instance page in the BELTRANS *Person* perspective. A click on person's ISNI identifier [12] in the *Person* perspective, leads to the external

[5]Please note that we currently investigate the possibility to add or remove certain perspectives

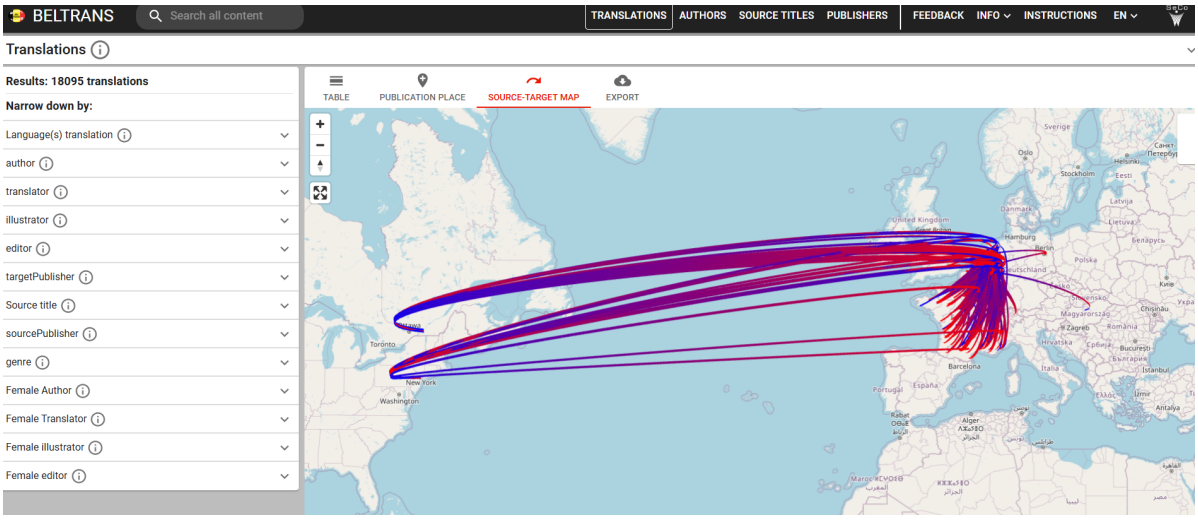

Figure 2: The source-target map view of the BELTRANS SAMPO interface: one arc is one translation, connecting the place of publication of the source title with the place of publication of the translation.

international ISNI database. This way, users can intuitively apply the *Follow-Your-Nose* method and traverse the underlying Linked Data graph.

4.3. Search and filter entities with different visualizations

The main view in the center shows all entities of the selected perspective, eventually filtered by selected search facets on the left. Different types of views can be implemented and can be accessed via the corresponding tab (fig. 1, (c)).

We provide the following tabs/views for the BELTRANS corpus, both the tabular view and the export view are provided for all perspectives, publication places and source target map are only provided for the perspectives *Translations* and *Source Titles*.

1. **Tabular view** This is the default view showing one entity and selected columns per row
2. **Publication place view** This map shows where a translation (or source title) was published
3. **Source-Target map view** This map shows arcs connecting the place of publication of a translation and the place of publication of the corresponding source title (fig. 2)
4. **Chart view** This view allows to show a pie chart or a bar chart of particular search facets.
5. **Export view** The export functionality is shown as a separate view in the main panel. When clicking on the button *Export CSV*, the current selection of entity types is downloaded.

5. Conclusion and Future Work

This paper presented the use of a SAMPO-UI to explore book translation flows of the recently published BELTRANS corpus. The corpus is published according to the FAIR principles, and the presented open source platform is hosted at KBR - the Royal Library of Belgium. Based on the deployment of this system several insights and future avenues emerge.

Instead of scrolling through CSV or RDF data, non-technical users can use the UI to filter and visualize the content by human-readable labels and explore relationships between translations, authors and publishers by clicking on links. This makes the BELTRANS corpus more accessible to the target audience of humanities and translation scholars.

The BELTRANS corpus relates to the collections and mission of KBR, however, developing and maintaining a dedicated UI was out of the scope of the BELTRANS project. Using externally maintained off-the-shelf and open source software like QLever or SAMPO-UI allowed us to build project-independent

infrastructure which eventually addresses crucial data sharing and infrastructure needs of Big Translation History (BTH).

One limitation of using SAMPO-UI in a Cultural Heritage use case is that its setup still requires technical knowledge and underlying infrastructure to provide RDF data via SPARQL. Whereas GhentCDH's code improvements represent a step towards more easily configurable front-ends, future work could explore more user friendly deployment methods integrating SAMPO-UI-based interfaces into commodity cloud computing services for Linked Open Data or generally into data publication pipelines.

The scope of SAMPO-UI is to provide user interfaces on top of semantically annotated data, like our BELTRANS corpus. However, the aspect of data curation is not addressed. Whereas the corpus can be curated externally and different SAMPO-UI instances could be used to visualize different corpus versions, future work could explore how newly built SAMPO-UI components can be used to further annotate data via SPARQL UPDATE queries. This would allow for further data maintenance by data curators or support citizen science initiatives to improve/correct corpora.

Finally, the improved version of SAMPO-UI implements a plug-in system where own components can be added. Future work can investigate the creation and maintenance of a plugin library, such that commonly used functionality can be shared to further increase the reusability of the software.

Credit authorship contribution statement

Sven Lieber Writing - Original Draft, Conceptualization, Software. **Ann Van Camp**: Writing - Review & Editing, Validation. **Joren Six**: Writing - Review & Editing, Software. **Julie M. Birkholz**: Writing - Review & Editing.

Acknowledgments

We would like to thank the BELTRANS team including interns and volunteers who helped curating and improving the data corpus. Furthermore we would like to thank the SeCo research group for creating the initial SAMPO-UI application and GhentCDH for upgrading and extending it. This work is financially supported by BELSPO BRAIN BELTRANS project (B2/202/P2/BELTRANS) (Lieber & Van Camp), CLARIAH-VL+ (I001525N) (Six & Birkholz), the BELSPO FED-tWIN project Digital Research Lab (Prf-2019-040, Birkholz).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] D. Roig-Sanz, L. Fólica, Big translation history: Data science applied to translated literature in the spanish-speaking world, 1898–1945, Translation Spaces 10 (2021) 231–259.

[2] J. Heilbron, Responding to globalization, Beyond descriptive translation studies (2008) 187–197.

[3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. doi:10.1038/sdata.2016.18.

[4] H. Van Uffelen, Dlbt: Digital library and bibliography for literature in translation and adaptation, in: Phaidracon 2020, University of Vienna, Vienna, Austria, 2020. URL: https://phaidra.univie.ac.at/o:1129342, conference presentation (PDF) in PHAIDRA digital repository; Handle: 11353/10.1129342.

[5] E. Ikkala, E. Hyvönen, H. Rantala, M. Koho, Sampo-ui: A full stack javascript framework for developing semantic portal user interfaces, Semantic Web 13 (2021) 69–84.

[6] E. Hyvönen, Digital humanities on the semantic web: Sampo model and portal series 14 (2023) 729–744. URL: https://journals.sagepub.com/doi/full/10.3233/SW-223034. doi:10.3233/SW-223034.

[7] E. Hyvönen, A. Ahola, P. Leskinen, J. Tuominen, Samposampo: A portal for studying enriched data and semantic connections on a cultural heritage linked open data cloud, in: The Semantic Web: ESWC 2025 Satellite Events, Portoroz, Slovenia, June 1 - 5, 2025, Proceedings, volume 15832 of *Lecture Notes in Computer Science*, Springer-Verlag, 2025, pp. 67–74. URL: https://doi.org/10.1007/978-3-031-99554-5_13.

[8] F. Pannach, L. Cheng, F. Pianzola, The golem-knowledge graph and search interface: Perspectives into narrative and fiction., in: CHR, 2024, pp. 462–471.

[9] B. Vercruysse, V. Ducatteeuw, J. Poukens, J. M. Birkholz, C. Verbruggen, Belhisfirm: Belgian company records as linked data, in: eabh annual meeting 2026, The European Association for Banking and Financial History (eabh), 2026.

[10] H. Rantala, A. Ahola, E. Ikkala, E. Hyvönen, How to create easily a data analytic semantic portal on top of a SPARQL endpoint: introducing the configurable Sampo-UI framework, in: Proceedings of 8th International Workshop on the Visualization and Interaction for Ontologies and Linked Data co-located with the 22nd International Semantic Web Conference (ISWC 2023) in Athens, Greece, CEUR Workshop Proceedings, Vol. 3508, 2023. URL: https://ceur-ws.org/Vol-3508/paper3.pdf.

[11] H. Bast, B. Buchhold, QLever: A Query Engine for Efficient SPARQL+Text Search, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, ACM, Singapore Singapore, 2017, pp. 647–656. URL: https://dl.acm.org/doi/10.1145/3132847.3132921. doi:10.1145/3132847.3132921.

[12] International Organization for Standardization, Information and documentation — International Standard Name Identifier (ISNI), Standard ISO 27729:2024, International Organization for Standardization, Geneva, CH, 2024. URL: https://www.iso.org/standard/87177.html.