



Aalto University  
School of Science



UNIVERSITY OF HELSINKI



# Case Study: Analysis of genre evolution in Finnish fiction literature using BookSampo's annotated genres

*Annastina Ahola, Aalto University*

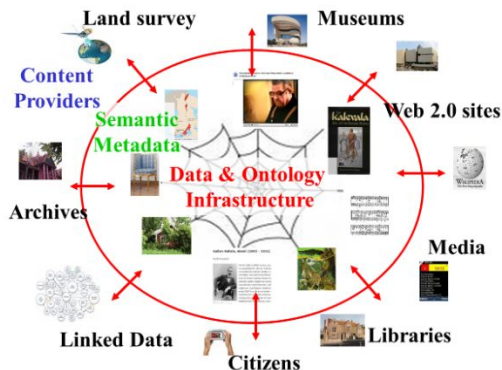
*25.3.2026*

# Sampo Model for 1) Human Users and 2) Machines



## LINKED DATA SERVICE

Findable, Accessible, Inteoperable, Re-usable  
FAIR data!



Linked Data Finland  
Living Laboratory Data Service for the Semantic Web

<https://ldf.fi/>

## APPLICATIONS & TOOLS

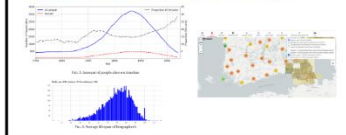
Use Cases

### Sampo Portals & Applications



Sampos: <https://seco.cs.aalto.fi/applications/sampo/>

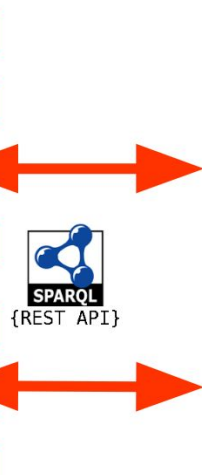
### Digital Humanities Research



Tools

Sampo-UI, etc.

Google Colab, Jupyter,  
YASGUI, etc.



### Sampo Model Principles P1–P6

P1	Support collaborative data creation and publishing
P2	Use a shared open ontology infrastructure
P3	Make clear distinction between the LOD service and the user interface (UI)
P4	Provide multiple perspectives to the same data
P5	Standardize portal usage by a simple filter-analyze two-step cycle
P6	Support data analysis and knowledge discovery in addition to data exploration

### References

Eero Hyvönen: **How to Create a National Cross-domain Ontology and Linked Data Infrastructure and Use It on the Semantic Web.** *Semantic Web*, 2024. DOI: 10.3233/SW-243468.

Eero Hyvönen: **Digital Humanities on the Semantic Web: Sampo Model and Portal Series.** *Semantic Web*, vol. 14, no. 4, pp. 729-744, 2023.

Esko Ikkala, Eero Hyvönen, Heikki Rantala and Mikko Koho: **Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces.** *Semantic Web*, vol. 13, no. 1, 69-84, 2022.

# Sampo Series of LOD Services & Portals



A. Gallen-Kallela: The Defense of the Sampo, 1896, Turku Art Museum

1. **MuseumFinland** – Finnish Museums on the Semantic Web (2004) [39 000 users]
2. **CultureSampo** – Finnish Culture on the Semantic Web (2008) [107 000 users]
3. **TravelSampo** - Mobile Contextualized Services of Cultural Tourism (2011)
4. **BookSampo** – Finnish Fiction Literature on the Semantic Web (2011) [1.6 million annual users]
5. **WW1LOD** – World War I Linked Open Data (2014)
6. **WarSampo** – Finnish World War 2 on the Semantic Web (2015-19) [1 200 000 users]
7. **Norssi Alumni on the Semantic Web** – Historical person registry using LOD (2017)
8. **U.S. Congress Prosopographer** – U.S. Congress Legislators 1789-2018
9. **BiographySampo** - Finnish Biographies on the Semantic Web (2018-20) [381 000 users]
10. **NameSampo** – Linked Data Workbench for Toponomastic Research (2019) [55 000 users]
11. **WarVictimSampo 1914-1922** – National War History [80 000 users]
12. **Mapping Manuscript Migrations (MMM)** – medieval manuscripts (2020) [9100 users]
13. **AcademySampo** – Finnish Academic People 1640 – 1899 (2021) [26 000 users]
14. **FindSampo** – Archaeological Finds on the Semantic Web (2021) [7 000 users]
15. **WarMemoirSampo** – Memoirs of Finnish WW2 veterans (2021) [4 400 users]
16. **LetterSampo** – Early Modern Letters on the Semantic Web (2022)
17. **ParliamentSampo** – Parliament of Finland on the Semantic Web (2023)
18. **LawSampo** – Finnish Legislation and Case Law on the Semantic Web (2023)
19. **BookSampo II** – Semantic Search, Browsing, and Data-analyses (2023)
20. **OperaSampo** – Opera and music theater performances in Finland 1830-1960 (2023)
21. **ConfermentSampo** – Conferment Ceremonies of the University of Helsinki 1640-1899 (2023)
22. **CoinSampo** – Finnish Numismatic Public Finds 2013 – 2023 (2024)
23. **LetterSampo Finland** – Finnish 19<sup>th</sup>-century Letters on the Semantic Web (2025)

Work in progress: **Samposampo**, **ArtSampo**, **NomismaSampo**, **PASampo**, .....



**Sampo** = Mythical artifact of the Finnish Epic Kalevala that gives to its owner riches and good fortune. A metaphor of amazing ancient technology. Eero Hyvönen: **Digital Humanities on the Semantic Web: Sampo Model and Portal Series**. *Semantic Web*, 14(4), 729-744, 2023.

# BookSampo – Finnish Fiction Literature on the Semantic Web

- Covers metadata about practically **all Finnish fiction literature available at Finnish Public Libraries** on a work level
  - Contains metadata on *genres, themes, characters, settings, etc.*
- Initially published in 2011 with its Drupal-based portal (<https://www.kirjasampo.fi/>) provided and maintained by the Finnish Public Libraries
  - offers **text-based search** for its work and author data

The screenshot shows the Kirjasampo website interface. At the top, there is a navigation bar with the logo 'Kirjasampo Sivupäivi Boksampo' and the tagline 'LÄITEN JA KUOSIENKUKAT WEBPÄLÄTIN PÄIVÄKKA'. Below the navigation bar, there is a search bar with the text 'HAKU' and a search icon. The search results section is titled 'Kirjasammon semanttinen haku ja data-analyysi' and includes the text 'Analyysi kirjasampo.fi:n vaihtoehtoinen hakulittymä Kirjasammon teos- ja tekijätiedoille.' Below the search results, there is a section titled 'KIRJAVINKIT JA -HYLLYT' featuring a book recommendation for 'Taikavuori' by Thomas Mann. The book cover is shown with the text 'Taikavuori TAKY 27.11.2024' and a description: 'Tässä kuussa tuli kuluneeksi 100 vuotta saksalaisen Thomas Mannin (1875-1955) romaanin Taikavuori (WSOY 1957, suom. Kai Kaila) ilmestymisestä. Romaani on yksi saksalaisen ja eurooppalaisen romaanikirjallisuuden suurimmista klassikoista. Se kertoo pankkimyönteisestä lääninrakennusinsinööriä Hans Castorpista, joka lähtee Alpeilla Davosissa sijaitsevaan parantolaan. Siellä hänen on tarkoitus lomailia ja tavata tuberkuloosia sairastavaa... Lue lisää ja jaa täky Lisää täkyjä | Teostiedot'. To the right of the book recommendation, there is a section titled 'KIRJAVINKIT JA -HYLLYT' with a grid of book covers and the text 'Kaunokirjallisuutta epäonnistumisesta KIRJAVINKIT JA -HYLLYT Epäonnea vai saamattomuutta? Tavallista tunarointia, mokaamista tai elämän kokosis virheitä ja erehdyksiä -kukapa ei olisi kokenut niitä jossain mittakaavassa. Tutustu hyllyyn Kaikki hyllyt | Hyllyt teemoitain | Hyllynteen ohje'. At the bottom right, there is a section titled 'LUMILOIDEN OMAT KIRJAVINKIT' with a grid of book covers and the text 'Romaaneja evakoloista'.

# BookSampo 2.0: Revisiting the Knowledge Graph

- However, the data had not really been used in **Digital Humanities research** despite its semantic richness and potential
- BookSampo 2.0 had with the following two goals:
  - a. Develop a prototype user interface for the BookSampo knowledge graph based on the **Sampo Model** and **Sampo-UI** that facilitate semantic search and browsing integrated seamlessly with tools for data analysis (<https://analyysi.kirjasampo.fi/>)
  - b. Study Finnish fiction literature **using Digital Humanities methods** based on the BookSampo knowledge graph, linked open data services, and the portal

The left screenshot shows the Kirjasampo 2.0 interface with search results for 'Romaanit'. The results are displayed in a table with columns for 'Nimi', 'Tekijä', 'Kirjallisuudenlaji', and 'Teema'. The right screenshot shows the same interface but with a pie chart titled 'Lajitustenuste: alkukielii' and a legend showing the distribution of books by language.

Language	Count	Percentage
englanti	4556	45.69 %
suomi	2392	23.88 %
ranska	2186	21.85 %
norja	173	1.73 %
ranska	150	1.50 %
italia	125	1.25 %
ruotsi	120	1.20 %
unknown	81	0.81 %
espanja	54	0.54 %
japani	39	0.39 %
venäjä	39	0.39 %
italia	26	0.26 %
hollanti	14	0.14 %
ruusi	9	0.09 %

# Using linked data for data analytic literary research: Case BookSampo

- The following examples partially draw from my and Telma's work presented in this paper:



RESEARCH ARTICLE |  Open Access |  

## Using linked data for data analytic literary research: Case BookSampo—Finnish fiction literature on the semantic web

Annastiina Ahola  Telma Peura, Eero Hyvönen

First published: 31 January 2025 | <https://doi.org/10.1002/asi.24984>

Annastiina Ahola, Telma Peura, and Eero Hyvönen. Using linked data for data analytic literary research: Case BookSampo - Finnish fiction literature on the semantic web. Journal of the Association for Information Science and Technology, 2025

# Data model

PREFIX kaunokki: <<http://www.yso.fi/onto/kaunokki#>>

PREFIX foaf: <<http://xmlns.com/foaf/0.1/>>

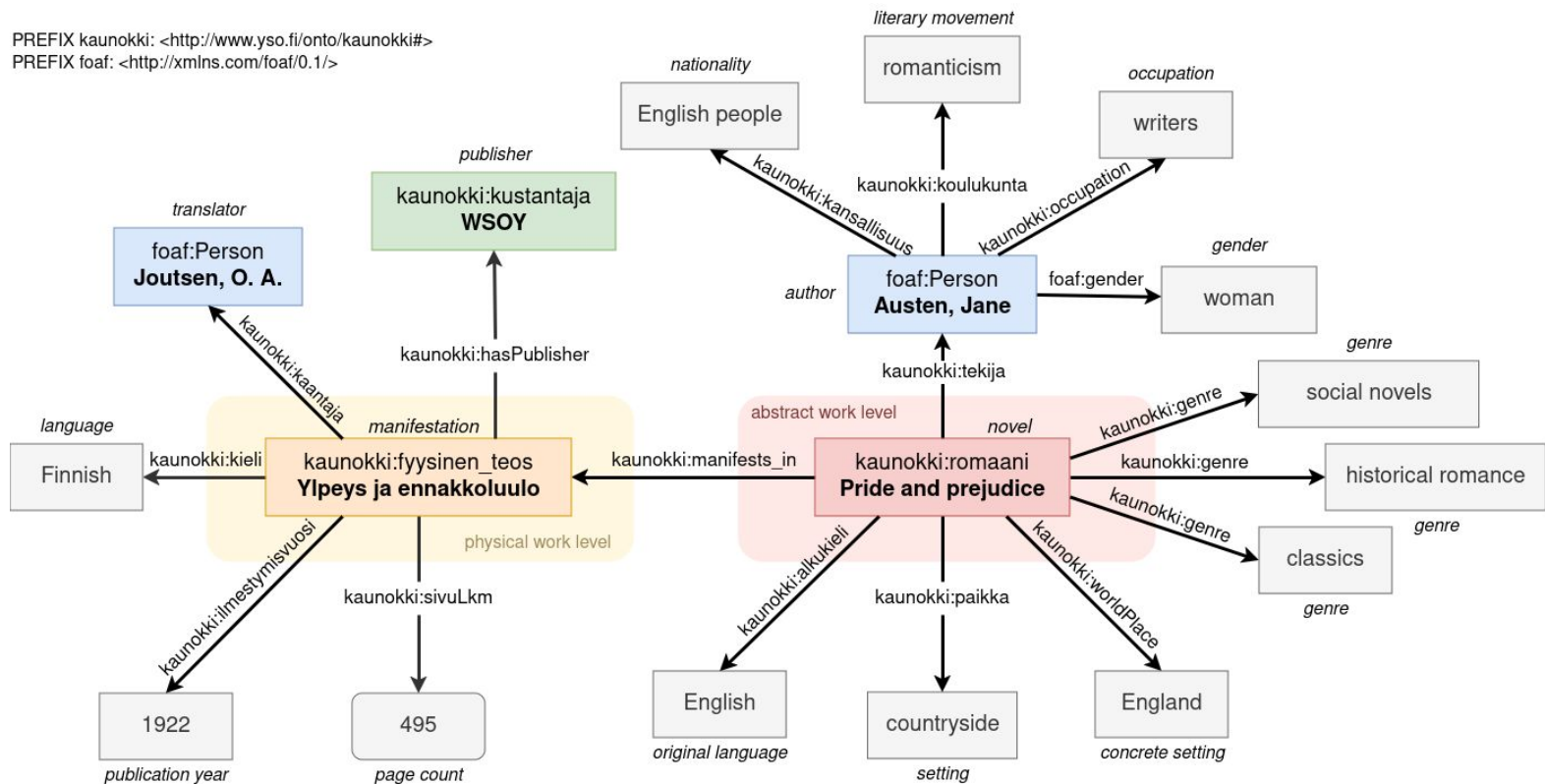


Image from:

Annastiina Ahola, Telma Peura, and Eero Hyvönen. Using linked data for data analytic literary research: Case BookSampo - Finnish fiction literature on the semantic web. *Journal of the Association for Information Science and Technology*, 2025

# Top annotated genres

## Top genres of all novels (91,315 total)

juvenile literature	15,365
children's literature	12,965
suspense fiction	10,474
Finnish language literature	10,204
psychological novels	8,749
<i>[no genre(s) defined]</i>	8,743
crime fiction	8,278
Swedish language literature	6,926
historical novels	6,188
light reading	5,444

## Top genres of novels originally written in Finnish (26,727 total)

Finnish language literature	10,080
juvenile literature	4,111
children's literature	3,426
psychological novels	3,318
suspense fiction	2,601
crime fiction	2,151
debut works	1,946
historical novels	1,770
humour	1,703
novels	1,536

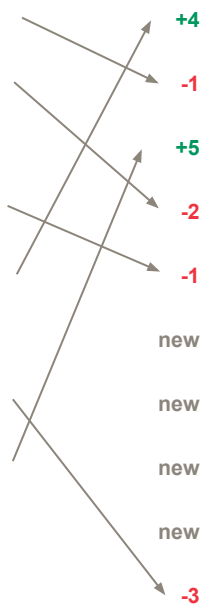
Since novels have publication years annotated, we can also how genres change for novels published in, e.g., different decades:

### Top genres of novels from 1990–1999 (9,395 total)

1	juvenile literature	1,834
2	suspense fiction	1,155
3	<i>[no genre(s) defined]</i>	989
4	crime fiction	893
5	children's literature	891
6	romantic novels	655
7	historical novels	625
8	psychological novels	616
9	light reading	569
10	detective literature	566

### Top genres of novels from 2010–2019 (22,150 total)

1	children's literature	4,302
2	juvenile literature	4,169
3	psychological novels	3,009
4	suspense fiction	2,979
5	crime fiction	2,408
6	easy readers	2,097
7	fiction and poetry	1,993
8	debut works	1,740
9	novels	1,709
10	historical novels	1,505

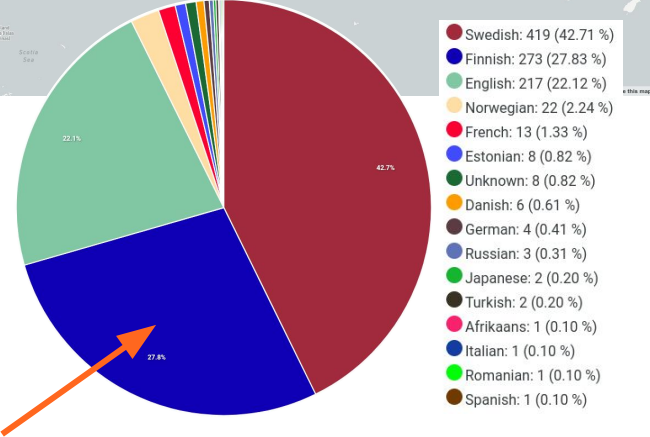
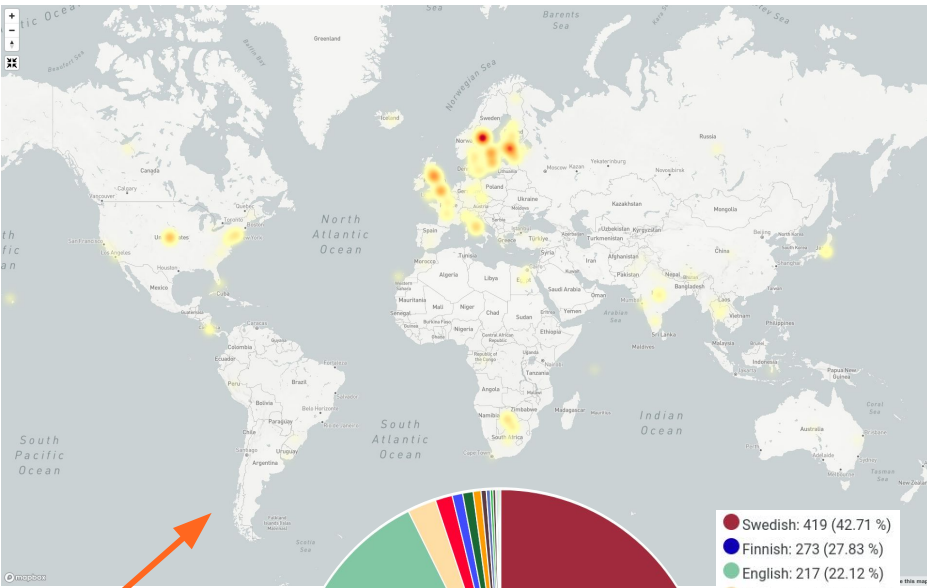
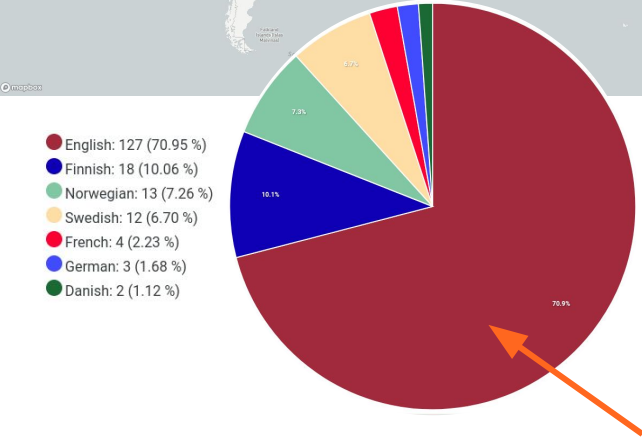
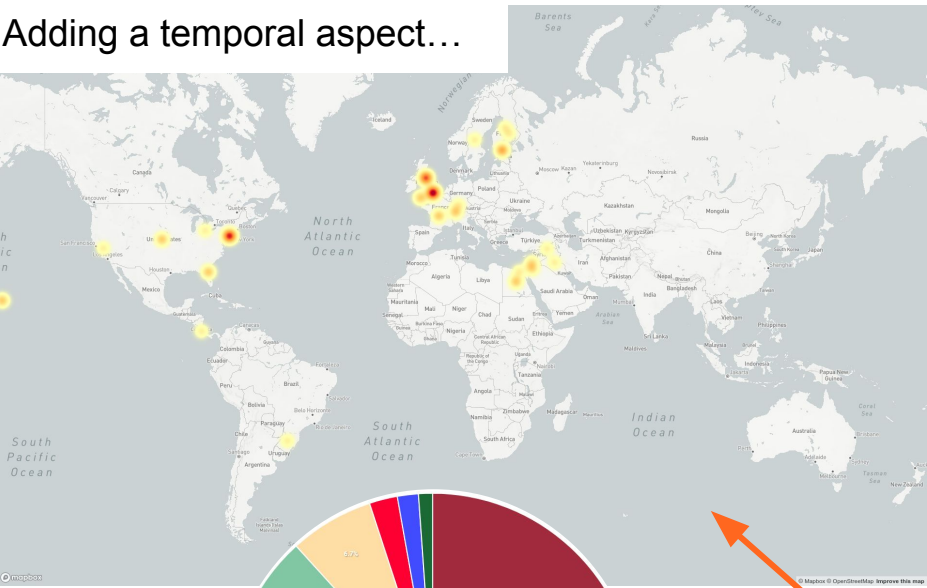


# Genres and settings

For novels with annotated concrete settings (e.g., “New York” or “France”), settings can be visualized on **maps** (e.g., as heatmaps) to see potential differences in the most common settings between different genres:



# Adding a temporal aspect...



Concrete settings (e.g., "New York") of the works visualized on a map

Original languages of the novels

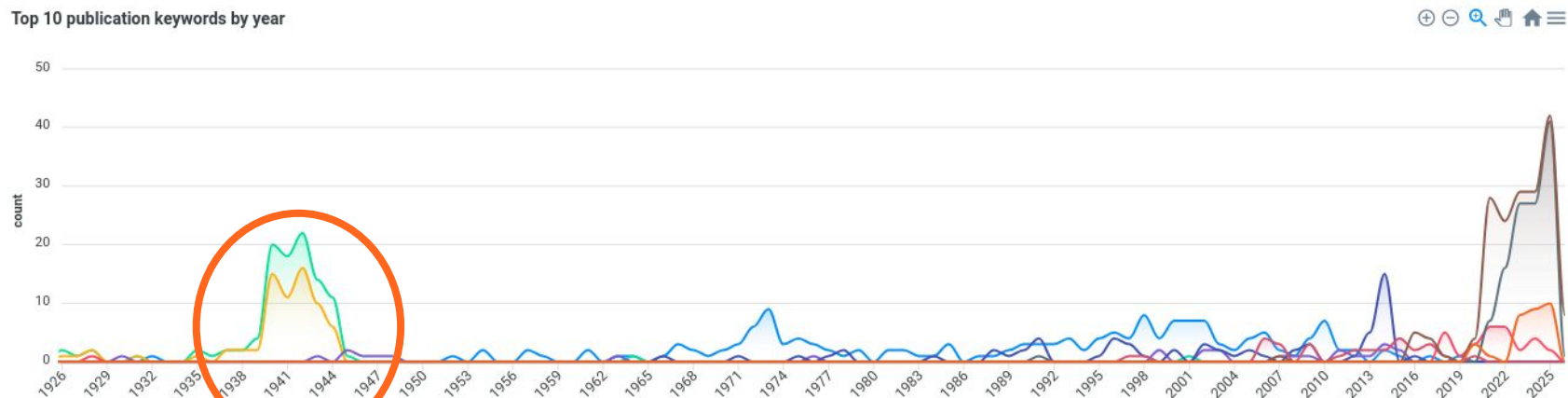
**Detective literature novels, 1900–1939 (179 works)**

**Detective literature novels, 2010– (980 works)**

# Keywords

## Top keywords for original Finnish publications by year

Top 10 publication keywords by year



“first person narrator”

● minäkertoja

“Finnish Civil War”

● prohibited books

“poison cabinet”

cabinet in which books considered  
“inappropriate” were stored  
(=censored book collections)

● myrkykaappi

● Suomen sisällissota

year

● Ostrobothnian dialects (Finnish)

“early youth literature”

(~ ages 9–12)

● amateur writers

● varhaisnuorten kirjat

“crime investigators”

● Advent calendars

“waiting for Christmas”

● rikostutkijat

● joulunodotus

# Genres and translations?

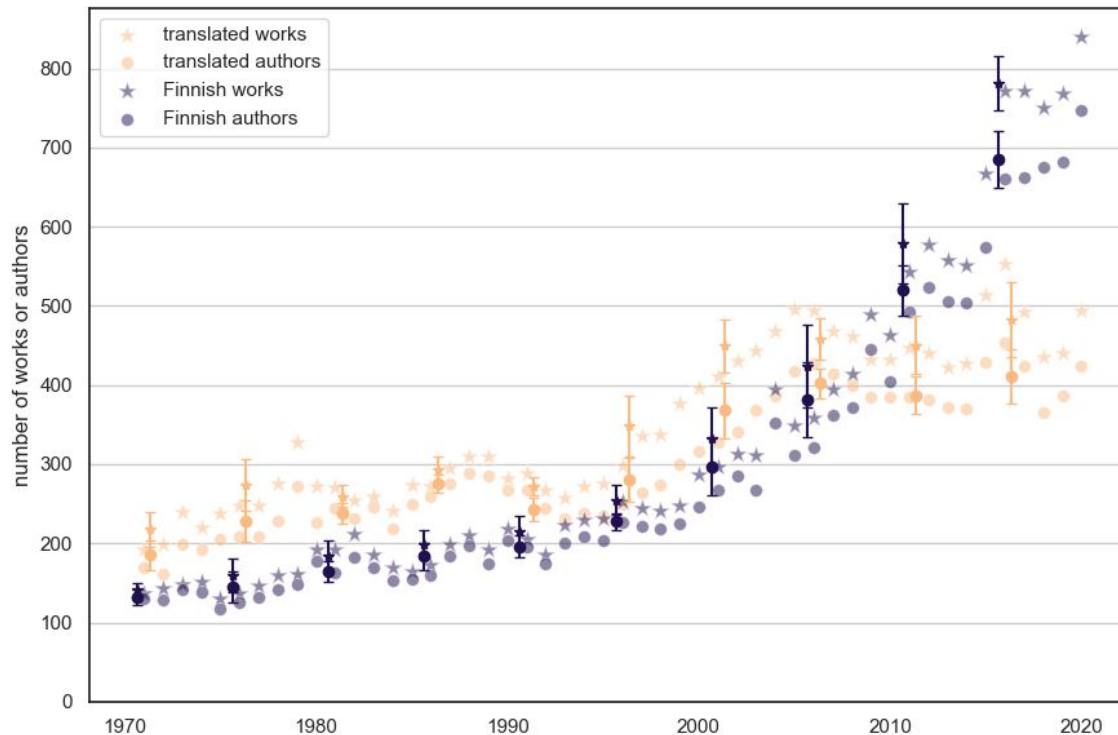
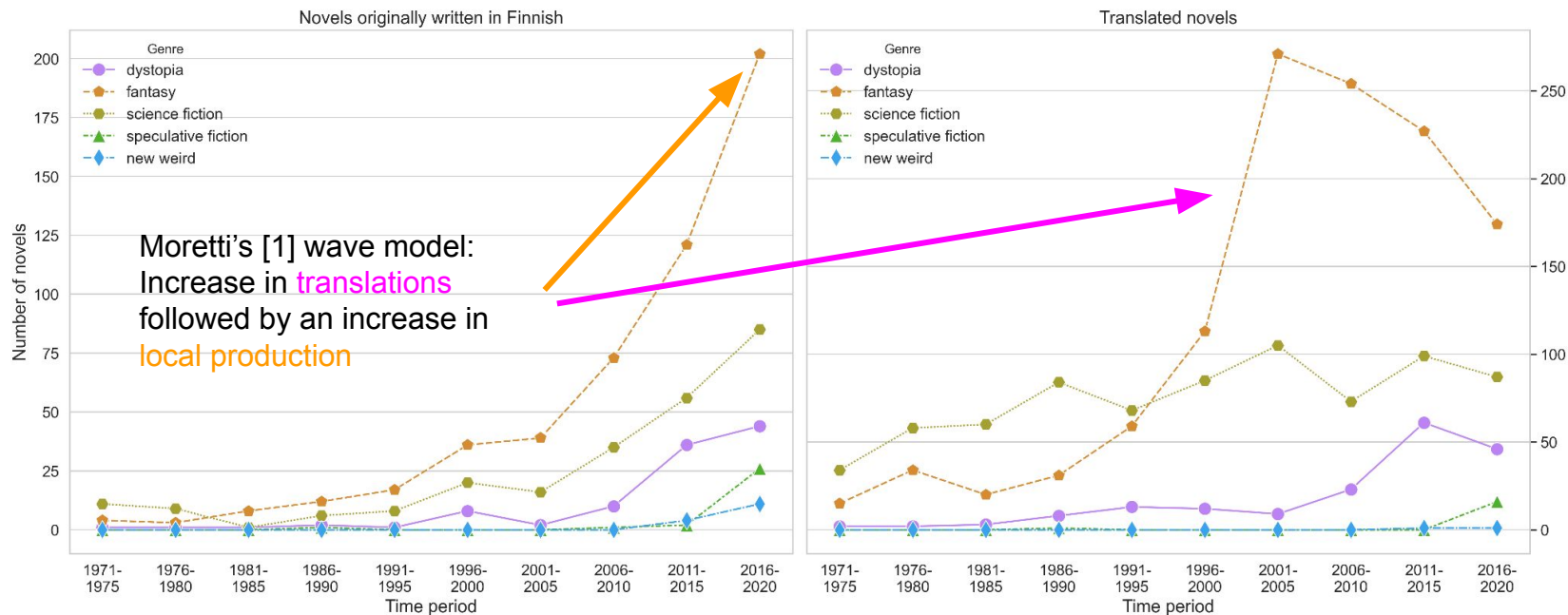


Image from:

Annastiina Ahola, Telma Peura, and Eero Hyvönen. Using linked data for data analytic literary research: Case BookSampo - Finnish fiction literature on the semantic web. *Journal of the Association for Information Science and Technology*, 2025

# Genres and translations



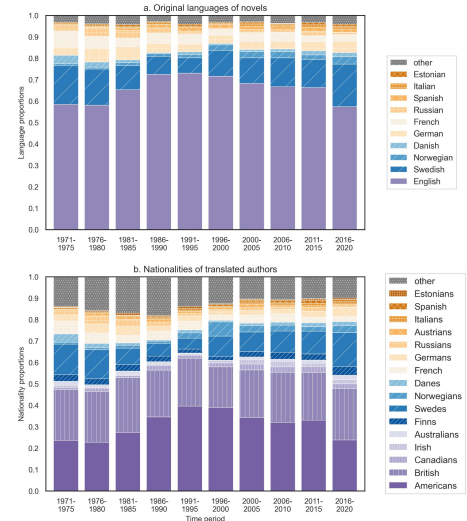
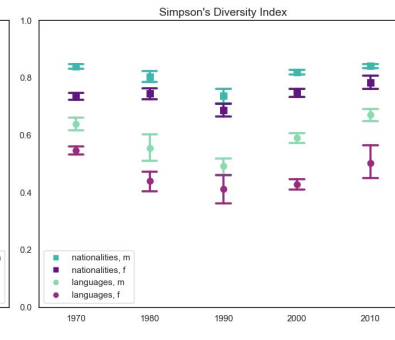
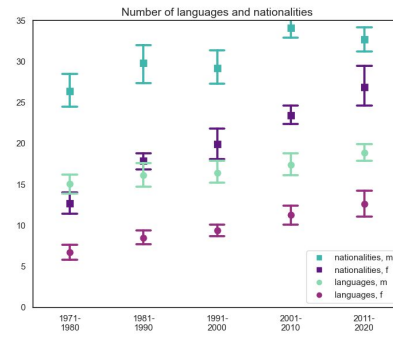
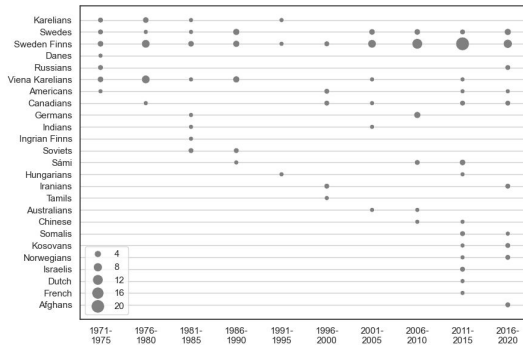
Underlying image from:

Annastiina Ahola, Telma Peura, and Eero Hyvönen. Using linked data for data analytic literary research: Case BookSampo - Finnish fiction literature on the semantic web. *Journal of the Association for Information Science and Technology*, 2025

[1] Moretti, F. (2005). *Graphs, maps, trees: abstract models for a literary history*. Verso.

# ... And many more possibilities!

- Focus here was on **genres**, but other annotations in the data also provide many opportunities for interesting analysis that is only possible due to the effort having gone (and still going) into annotating books (and data related to them)!



Images from:

Annastiina Ahola, Telma Peura, and Eero Hyvönen. Using linked data for data analytic literary research: Case BookSampo - Finnish fiction literature on the semantic web. *Journal of the Association for Information Science and Technology*, 2025



Thank you!