Swisscorpora.ch – dealing with copyrighted textual data in linguistic and literary research

Ruprecht von Waldenfels[†], Bruno Cartoni, Jacques Moeschler[‡]

[†]Slavic languages and literatures, U Bern & [‡]General linguistics, U Geneva

28.11.2013, digital humanities sagw — Neue Herausforderungen für den Forschungsplatz Schweiz

How can we use copyrighted texts in text-based research?

Electronic texts are increasingly important in linguistics and literary studies - but **what are the legal implications of using copyrighted texts as primary data?**
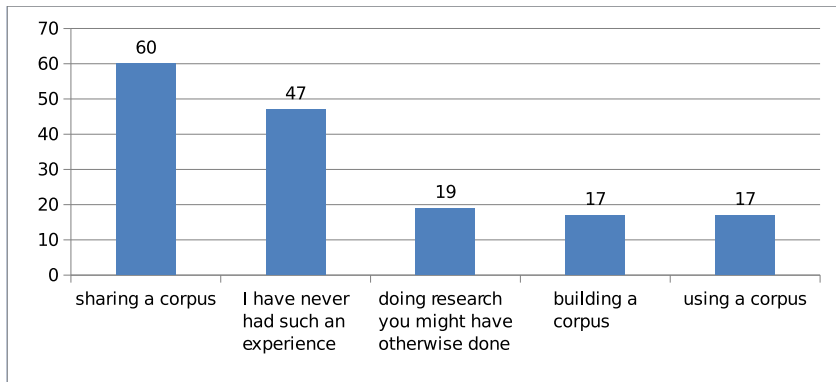
Objectives of the SwissCorpora project:

- do a survey to **analyze** current practices
- build case studies to **clarify** the legal status
- publish on a web site to **inform** the community and potentially spark a discussion

SwissCorpora.ch – a joint project of the Swiss Academic Society for Eastern European Studies, the Swiss Society of Linguistics, the Collegium Romanicum, supported by the Swiss Academy of Humanities and Social Sciences

Has your concern about legal/copyright issues ever caused you to refrain from... ?

*Problem is **relevant,** even though little concrete legal problems*

Results

Key legal issues:

- **digitalization, data collection and annotation is a potential problem** as it results in a derived work infringing on the copyright for the original work – even if the material was published on the web

- **for researchers working individually that's not a problem** since this infringement is admissable for private use

- **but sharing collected data with colleagues is a real problem** – it is admissable only for educational purposes, but not for research, not even sharing with colleagues working on related problems

*Web site **coming up** with **survey results**, **guidelines** and **case studies** (in French, German and English).*

General conclusions

- use of digital text corpora **is important for research** in linguistics and literary science (and beyond)
- however, the legal situation is **a potential impediment** especially **for sharing** the data, which **runs counter** to demands for **transparency and open access**

Questions to the community:

- can our research really **do without the possibility to legally share our primary data?**
- do we need to **lobby for exceptions for publicly funded research and *fair use*** of copyrighted textual material?

http://www.swisscorpora.ch

General conclusions

- use of digital text corpora **is important for research** in linguistics and literary science (and beyond)
- however, the legal situation is **a potential impediment** especially **for sharing** the data, which **runs counter** to demands for **transparency and open access**

Questions to the community:

- can our research really **do without the possibility to legally share our primary data?**
- do we need to **lobby for exceptions for publicly funded research and *fair use*** of copyrighted textual material?

Thank you for your attention!
http://www.swisscorpora.ch

Thank you for your attention!

SwissCorpora.ch – a joint project of the Swiss Academic Society for Eastern European Studies, the Swiss Society of Linguistics, the Collegium Romanicum, supported by the Swiss Academy of Humanities and Social Sciences

- Idea and project supervision: Ruprecht von Waldenfels, Jacques Moeschler
- Survey, case studies, web site: Bruno Cartoni
- Legal assessment: Marcello Baggi, Bertil Cottier, Stefano Tardini (DICE project)