Into the *Perryverse*: A CL Journey to the Realm of Lexical Complexity

Andreas Blombach, Thomas Proisl, Stefan Evert, Philipp Heinrich, Natalie Dykes
Computational Corpus Linguistics Group
Bismarckstr. 6, 91054 Erlangen, Germany
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
{andreas.blombach, thomas.proisl, stefan.evert, philipp.heinrich, natalie.mary.dykes}@fau.de

Perry Rhodan is the eponymous hero of a German science fiction series which has been published continuously and uninterrupted since 1961. (There have also been uninterrupted runs of translations in France, Japan and the Netherlands, as well as more limited runs in several other countries.) Short novels are published weekly in *Heftroman* form (roughly equivalent to dime novels or pulp magazines). To date, the main series comprises over 3,000 of these novels, which justifies its own designation as the world's "biggest science fiction series". The series has also spawned several spin-offs and "proper" novels in its so-called *Perryverse*, as well as hardcover collections, audio and comic books.

In previous research, we evaluated measures for different dimensions of textual complexity (e.g. lexical diversity, lexical disparity and syntactic complexity; for the latter, see Proisl et al. 2019) by comparing i.a. German highbrow and lowbrow literature.[1] Books classified as highbrow literature (selected for having been nominated for prestigious literature prizes) could quite reliably be separated from dime novels from the crime ("Jerry Cotton"), romance ("Julia Extra") and horror ("Geisterjäger John Sinclair") genres.[2] However, Perry Rhodan novels, representing science fiction dime novels, appeared to be much more complex, especially lexically, than their counterparts from other genres (and even more complex than many highbrow novels, according to some measures).

We now want to take a closer look at this apparent complexity, focussing especially on the series' idiosyncratic vocabulary (e.g. *Raumer*, *Impulsstrahler*, *Mausbiber*, *Arkonide*, *Linearraum*, *Zellaktivator*). Is it just the sheer amount of invented technical terms, species, planets and proper names that increases measured complexity, or is there more to it? In other words, does Perry Rhodan – and possibly science fiction in general – only appear especially complex to the uninitiated, or is it similar for regular readers familiar with its peculiar language? How can we characterise and categorise this special vocabulary?

---

[1] Using our toolbox available at https://github.com/tsproisl/Linguistic_and_Stylistic_Complexity.
[2] We want to thank our colleagues and collaborators at the University of Würzburg for compiling this corpus: Fotis Jannidis, Leonard Konle and Steffen Pielström.

To find out to which extent our results may be generalised (e.g. whether they are language-specific or even limited to Perry Rhodan), we compare English science fiction texts with texts from other genres, using both published works and fan fiction.

We also want to track changes over time in style, complexity and vocabulary, taking into account differences between individual authors. Our German corpus currently consists of approx. 650 Perry Rhodan novels from 1961 to 1975. The novels from this period, especially the earlier ones, are known for their focus on human expansionism and war, whilst later installments more often feature diplomacy and peaceful resolutions of conflicts. We plan to investigate if we can identify such changes using techniques like semantic tagging and topic modelling. Additionally, we want to compare the older novels to Perry Rhodan Neo, a modern retelling of the original story, which has been published biweekly since 2011, in parallel to the main series.

Proisl, Thomas, Leonard Konle, Stefan Evert, and Fotis Jannidis. 2019. “Dependenzbasierte syntaktische Komplexitätsmaße.” In *DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts*, edited by Patrick Sahle, 270–273. https://doi.org/10.5281/zenodo.2596095.

Leitfragen des Workshops

Potentiell besonders interessante Aspekte habe ich hervorgehoben:

- How can we characterize the language of Science Fiction, concerning individual authors as well as individual subgenres and text types?
- How has the language of Science Fiction changed over time? Which linguistic levels contribute most to such change?
- Which corpus linguistic methods lend themselves to an analysis of the language of Science Fiction (e.g., corpus stylistics, corpus-based approaches to metaphor)?
- Is it possible to quantify the notion of 'estrangement' characterizing Science Fiction texts?
- In how far has the language of Science Fiction entered popular culture?
- How are linguistic aspects of artificial intelligence (AI) represented in Science Fiction (e.g., by sentient androids, talking machines, etc., cf. Epstein et al. (2008))? How does this relate to present day AI developments, such as chatbots, Alexa, etc.?
- What do corpus linguistic methods contribute to analyzing world building in SF?
- How can multimodal material be productively included in the corpus-based analysis of SF texts?

Hier noch ein wunderes PR-Zitat:
»Der Grigoroff-Projektor stottert, Grigoroff-Feld zeigt Implosions-Symptome auf ultrahochfrequenter Basis. Bricht zusammen. Grigoroff-Unfall! Entstofflichung, dann sofortige Rematerialisierung, jedoch in einer anderen Galaxis. Nein, es müßte noch der Hyperraum sein. Es gleicht einem Arkoniden-Transitionsschock mit modifizierter Linearauslegung. Ist das etwa die Kalupsche Librationszone im Realisierungseffekt?« (PR 1343, Kap. 9)