



Digitale Methodologie ist für mich *Modellierung*

„ALL MODELS ARE WRONG BUT SOME ARE USEFUL“ (BOX 1979, 202)

- Bewusste, zweckgerichtete Abstraktion, Reduktion, Erweiterung oder/und Verschmelzung
- *Fiktional*, häufig formalisiert, verschieden realisierbar
- Es entsteht etwas Neues und Virtuelles, dessen Instanzierungen komplementäre oder/und analog nicht realisierbare Ergebnisse ermöglichen.

Box, George E. P. (1979): Robustness in the strategy of scientific model building. In: Launer, Robert L. und Wilkinson, Graham N (Hrsg.): Robustness in Statistics. New York, San Francisco und London: Academic Press. 201–236.



Modellierung des Designs für ein Benchmark-Korpus

Sprachliche Modellierung
 (Designhandbuch) von
 korpuslinguistischen/literarischen
 Designparametern + Realisierung in u.a.
 TEI XML zum Aufbau und Monitoring
 eines balancierten Romane-Korpus des
 19. Jahrhunderts

Criteria: Level 1 Eligibility.
 In order to be included, a text must...

- have been first published as a book between 1850 and 1920, sub-grouped and equally split by decades.
- have first been published in a European country. [maybe not "first": within that decade]
- be a novel, i.e. a fictional prose narrative of a minimum length of 10,000 words.
- have originally been written in the language of the given subcollection

ELTeC-core Language	TEXTS		AUTHORSHIP				LENGTH			TIME SLOT					REPRINT COUNT		
	Texts	Words	Male	Female	1-title	3-title	Short	Medium	Long	1840-59	1860-79	1880-99	1900-20	range	Frequent	Rare	E5C
cze	100	5621667	88	12	62	6	43	49	8	12	21	39	28	27	1	19	80.00
deu	100	12738842	67	33	35	9	20	37	43	25	25	25	25	0	48	46	96.92
eng	100	12227703	49	51	70	10	27	27	46	21	22	31	26	10	32	68	100.00
fra	100	8712219	66	34	58	10	32	38	30	25	25	25	25	0	44	56	101.54

Carolin Odebrecht, Lou Burnard, & Christof Schöch. (2021). European Literary Text Collection (ELTeC): April 2021 release with 14 collections of at least 50 novels. (Version v1.1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.4662444> , Abbildung unten aus <https://distantreading.github.io/ELTeC/>, Abbildung rechts https://distantreading.github.io/sampling_proposal.pdf Seite 5, Zugriff am 17.07.2023

