

D20.3 Structure of Arguments in Science: Modelling Kepler's "Astronomia nova"



Prof. Gerd Graßhoff (HU Berlin, MPIWG Berlin)
Mohammad Yeghaneh Abkenar (HU Berlin)

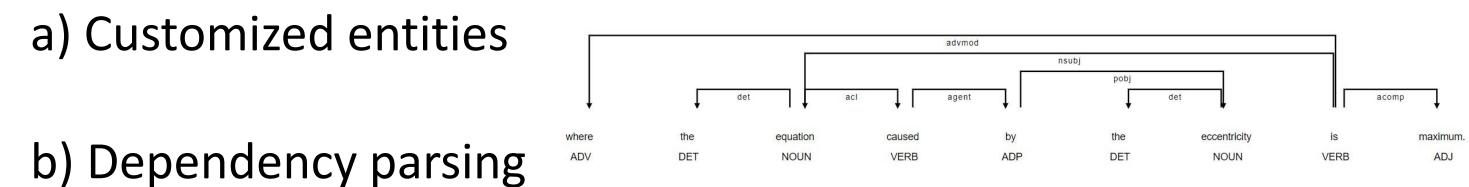
Scientific Publications, text corpora

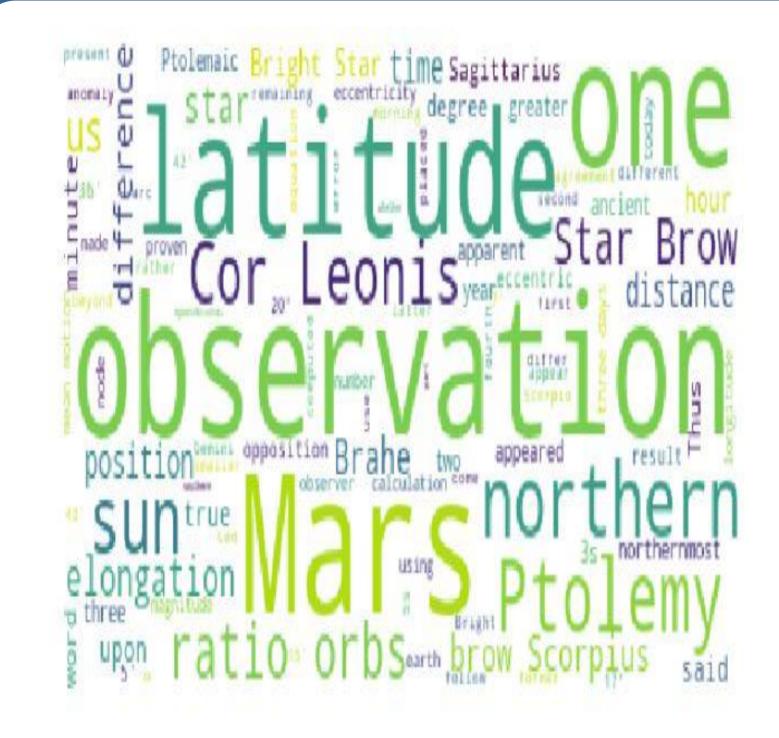
- Kepler's "Astronomia nova" Corpus
 - Revolutionary new theory of planetary motion
 - Purely geometrical epicyclic models by physical interpretations
 - → Aim: Computational model of Kepler's argumentation
- Exoplanet Publication Corpus [1]
 - Scientific articles on recent research on exoplanets 1990-2020
 - → Aim: Computational model to extract arguments
- Corona Virus Publication Corpus
 - The corpus analysis modelling understanding of COVID-19

Astronomia nova

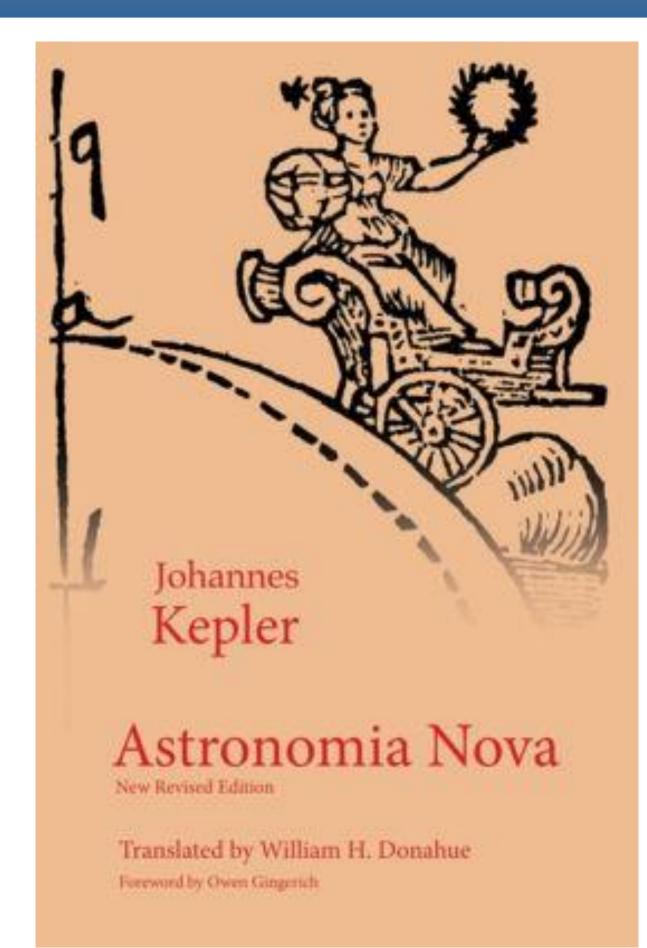
The corpora are represented in structural trees representing their textual division.

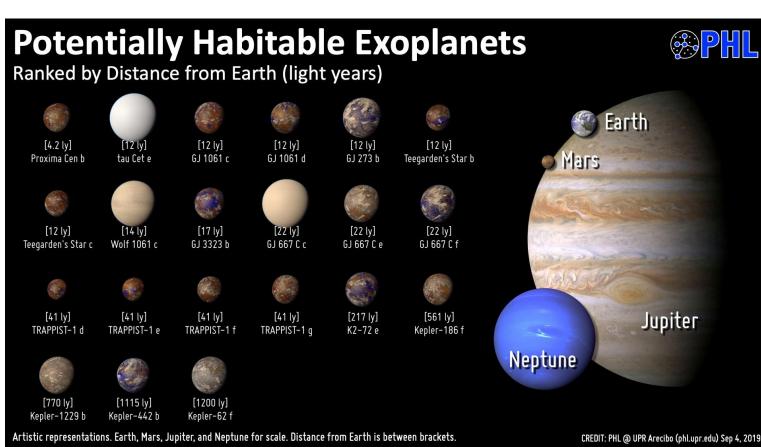
- 1) Parts, chapters, paragraphs, sentences, tables, diagrams
- 2) Sentence structure analysed with Spacy



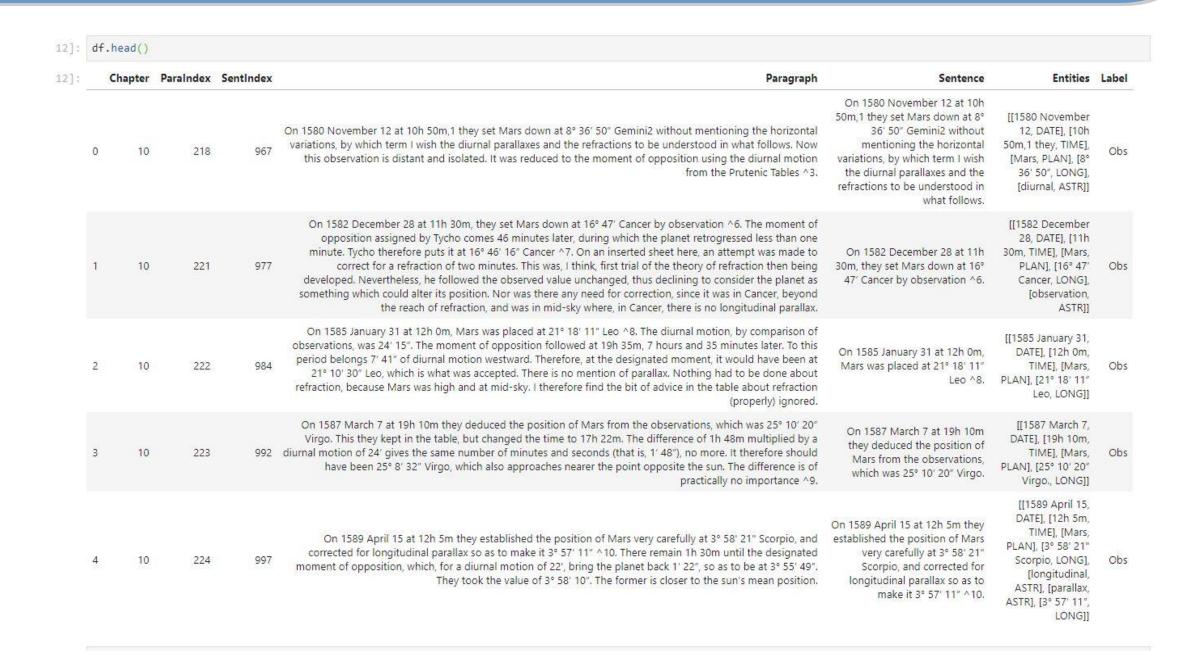


y=0 top features		y=1 top features		y=2 top features		y=3 top features	
Weight?	Feature	Weight?	Feature	Weight?	Feature	Weight?	Feature
+3.054	present	+21.317	compare	+14.381	propose	+13.982	conclude
+2.851	<bias></bias>	+12.742	compared	+9.543	proposed	+1.150	strong
+2.061	model	+2.287	predictions	+1.402	macho	+0.932	fully
+1.711	using	+1.631	spectra	+1.239	microlensing	+0.858	suggests
+1.207	use	+1.444	values	+1.106	theta_	+0.840	likely
+1.190	planet	+1.341	derived	+0.950	blg	+0.813	taken
+1,186	explore	+1.301	obtained	+0.936	better	+0.802	pulsation
+1.023	demonstrate	+1.240	results	+0.925	test	+0.799	presented
6401 more positive		+1.225	trappist	+0.923	coronagraphic	+0.696	nonequilibrium
1810 more negative		+1.144	observed	+0.901	statistical	+0.671	considered
-1.006	optically	+1.060	slow	+0.814	interception	+0.658	necessary
-1.011	coronagraphic	+1.048	j_	+0.787	extra	+0.643	explanation
-1.020	events	+1.023	t_	+0.764	collect	+0.641	normal
-1.024	conclusions	+1.016	000	+0.737	experimental	+0.621	114762b
-1.051	derived	+1.007	literature	752 n	nore positive	+0.604	difficult
-1.060	predicted	+1.007	frequencies	7459 more negative		555 more positive	
-1.069	spectra	+0.977	chromospheric	-0.732	different	7656 r	nore negative
-1.093	values	+0.972	existing	-0.742	spectra	-0.594	results
-1.099	support	+0.967	measurements	-0.746	surface	-0.601	new
-1.107	obtained	+0.967	compares	-0.757	planets	-0.643	apply
-1.110	microlensing	+0.951	spectrum	-0.799	stellar	-0.659	methods
-1.141 macho		1287 more positive		-0.819	temperature	-0.660	models
-1.143	trappist	6924 n	nore negative	-0.895	evolution	-0.682	based
-1.158	pulsation	-0.944	use	-0.930	light	-0.816	discuss
-1.233	better	-0.954	investigate	-0.935	using	-0.882	use
-1.237	literature	-0.977	composition	-1,001	earth	-0.935	planet
-2,106	predictions	-1.070	using	-1.015	modeling	-0.984	planets
-8,006	proposed	-1.210	conclude	-1.086	compare	-1.118	stellar
-10.899	conclude	-1.216	planet	-1.597	models	-1.245	present
-11.718	compared	-2.089	present	-1.912	present	-1.578	using
-12.273	propose	-2,091	model	-1.917	model	-2.369	model
-19.718	compare	-3.357	<bias></bias>	-3.698	<bias></bias>	-4.037	<bias></bias>





Customized named entity recognition (NER) [2]

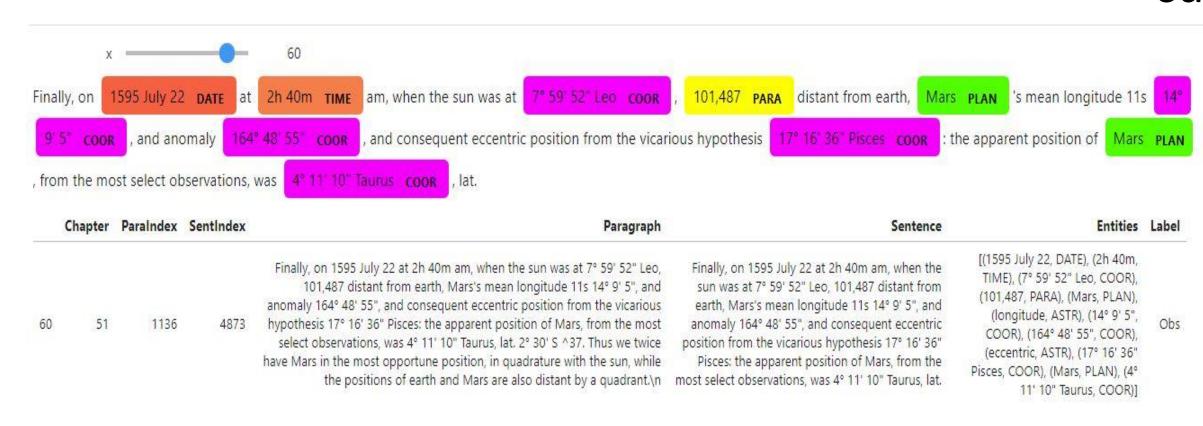


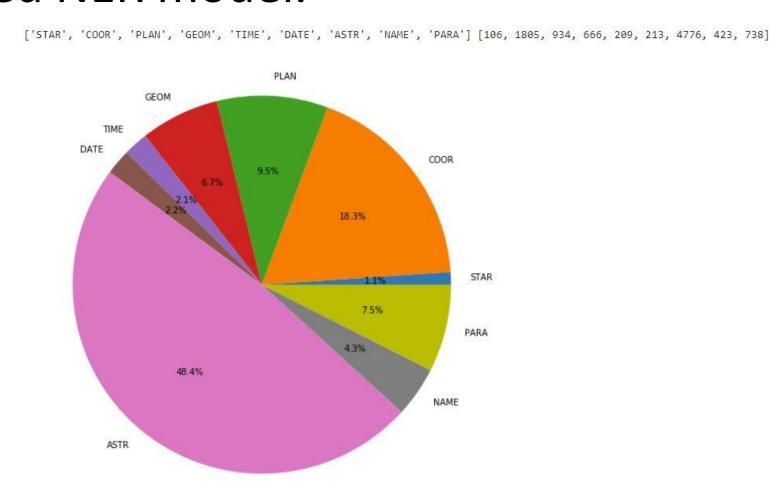
Key Idea. Hybrid approach; pattern(rule) based & deep learning approach to provide NER model using spaCy.



Explicit Observation Extraction [2]

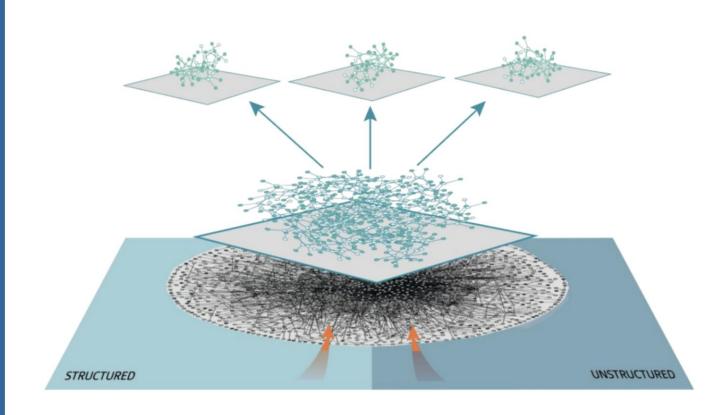
Key Idea. Extract Observations from the corpus by customized NER model.





Model between layers of textual, linguistic and semantic structures

- Textual structures are the ground layer interpreted by argument modeling
- Linguistic parsing extracts tokens, entities, dependency graphs
- Semantic tree structure
 contain semantic elements,
 statements and their
 relations



• Argument modeling extracts logical, causal, geometrical, scientific reasoning.

References

[1] Graßhoff, Gerd; Bier, Sabrina (2019): Database of abstracts in publications on exoplanets from the NASA archive. Zenodo. Dataset. https://doi.org/10.5281/zenodo.3269732

[2] Yeghaneh, Mohammad; https://github.com/myeghaneh

[3] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

