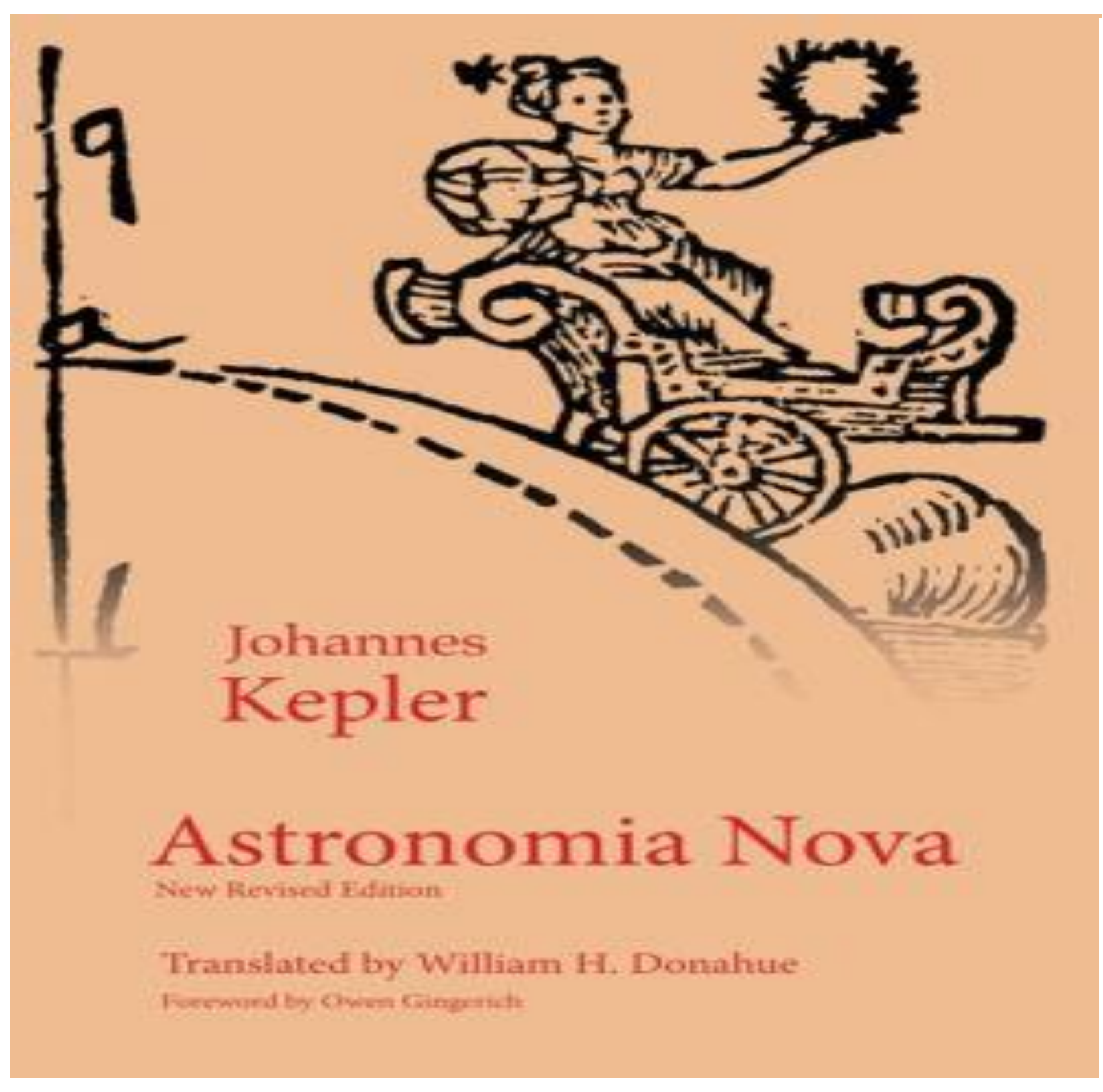
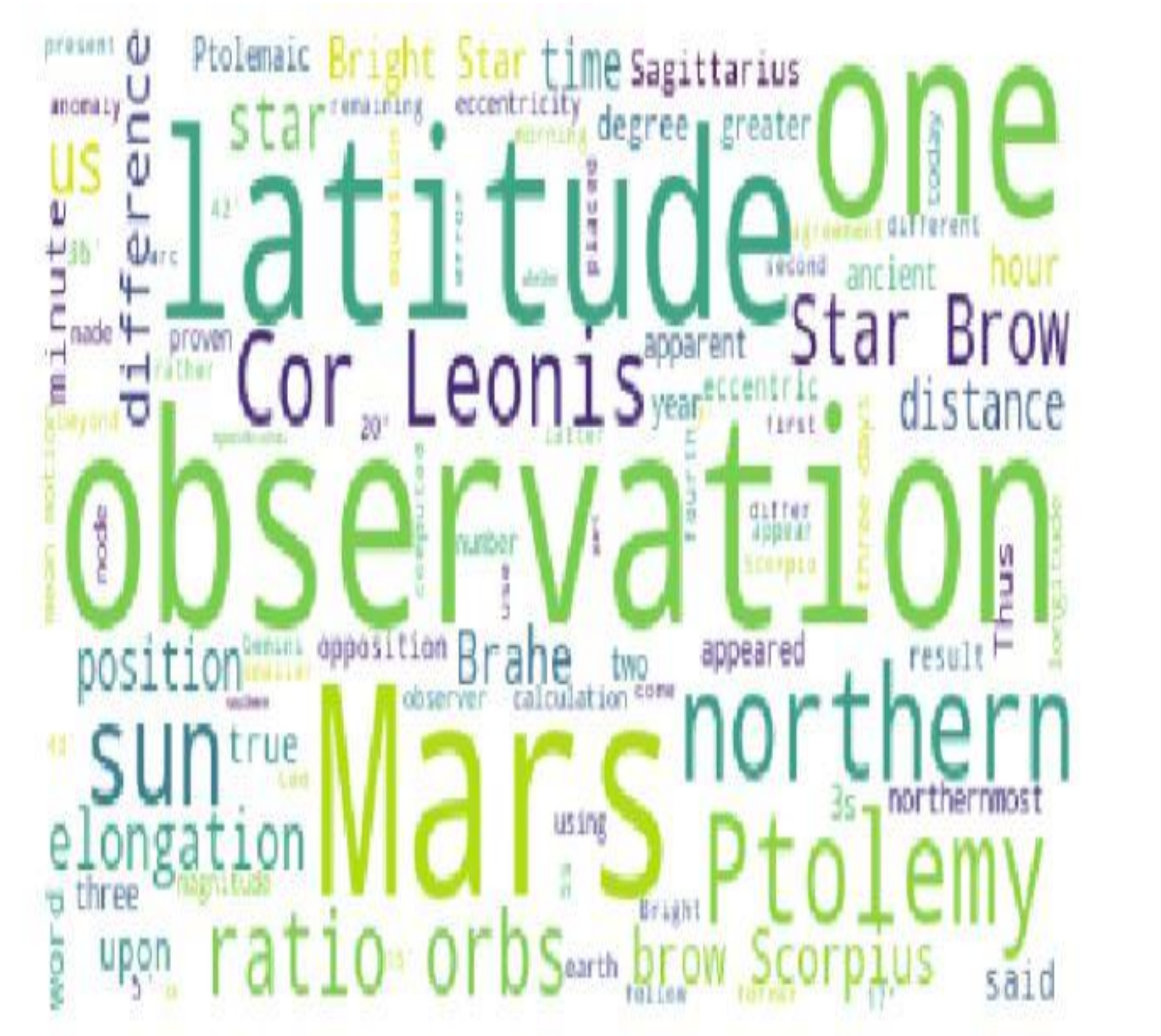
Tackling Semantic :
Semantic Analysis of Text Corpora using AI

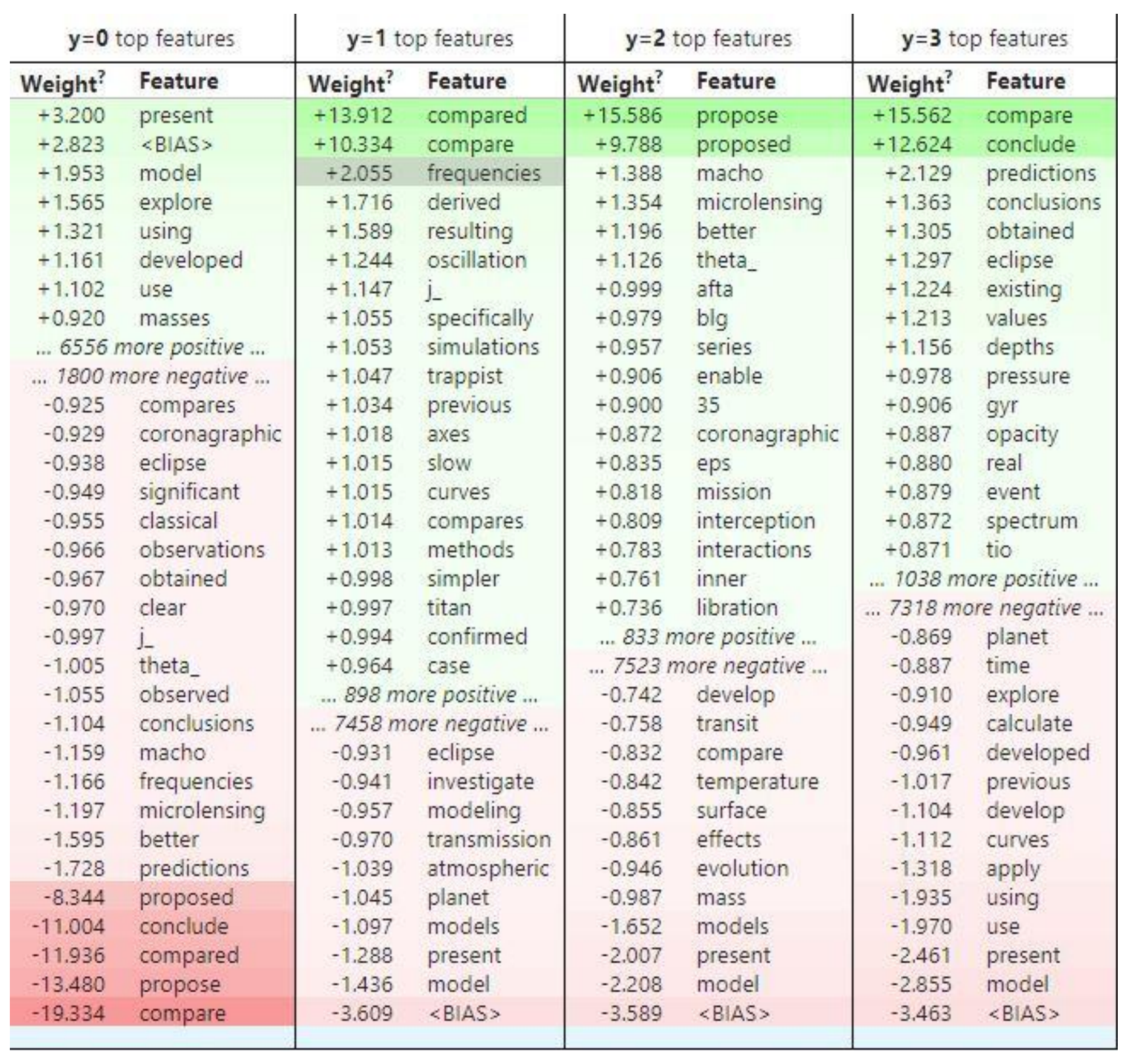
HU Berlin

Analysis of Text Corpora

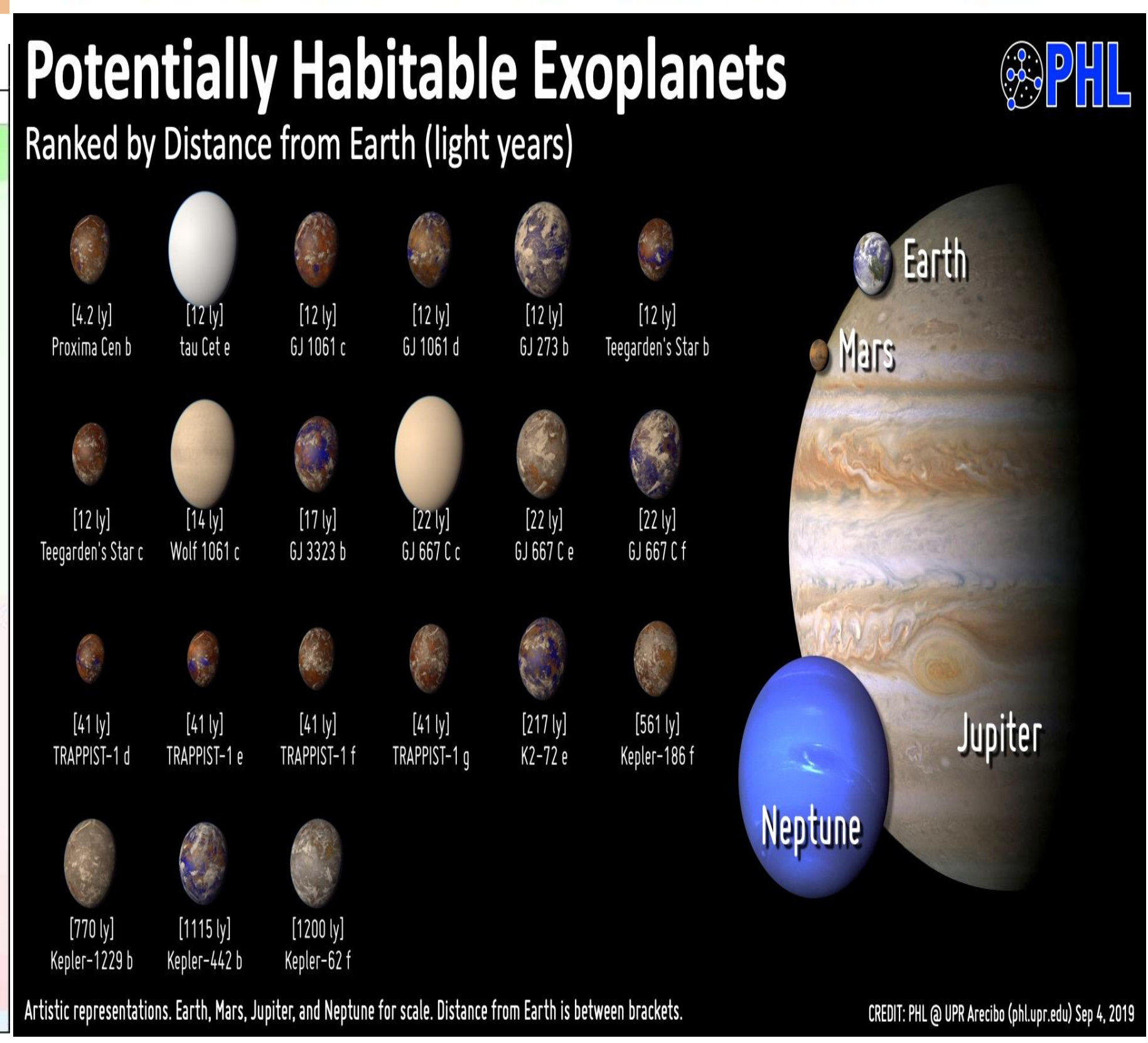
1) Kepler's book "Astronomia nova" Corpus
-
-
-
2) Exoplanet Publication Corpus [1]
-
-
3) Corona Virus Publication Corpus
-







y=0 top features		y=1 top features		y=2 top features		y=3 top features	
Weight?	Feature	Weight?	Feature	Weight?	Feature	Weight?	Feature
+3.200	present	+13.912	compared	+15.586	propose	+15.562	compare
+2.823	<BIAS>	+10.334	compare	+9.788	proposed	+12.624	conclude
+1.953	model	+2.055	frequencies	+1.388	macho	+2.129	predictions
+1.565	explore	+1.716	derived	+1.354	microlensing	+1.363	conclusions
+1.321	using	+1.589	resulting	+1.196	better	+1.305	obtained
+1.161	developed	+1.244	oscillation	+1.126	theta_	+1.297	eclipse
+1.102	use	+1.147	j_	+0.999	afta	+1.224	existing
+0.920	masses	+1.055	specifically	+0.979	blg	+1.213	values
… 6556 more positive …		+1.053	simulations	+0.957	series	+1.156	depths
… 1800 more negative …		+1.047	trappist	+0.906	enable	+0.978	pressure
-0.925	compares	+1.034	previous	+0.900	35	+0.906	gyr
-0.929	coronagraphic	+1.018	axes	+0.872	coronagraphic	+0.887	opacity
-0.938	eclipse	+1.015	slow	+0.835	eps	+0.880	real
-0.949	significant	+1.015	curves	+0.818	mission	+0.879	event
-0.955	classical	+1.014	compares	+0.809	interception	+0.872	spectrum
-0.966	observations	+1.013	methods	+0.783	interactions	+0.871	tio
-0.967	obtained	+0.998	simpler	+0.761	inner	… 1038 more positive …	
-0.970	clear	+0.997	titan	+0.736	libration	… 7318 more negative …	
-0.997	j_	+0.994	confirmed	… 833 more positive …		-0.869	planet
-1.005	theta_	+0.964	case	… 7523 more negative …		-0.887	time
-1.055	observed	… 898 more positive …		-0.742	develop	-0.910	explore
-1.104	conclusions	… 7458 more negative …		-0.758	transit	-0.949	calculate
-1.159	macho	-0.931	eclipse	-0.832	compare	-0.961	developed
-1.166	frequencies	-0.941	investigate	-0.842	temperature	-1.017	previous
-1.197	microlensing	-0.957	modeling	-0.855	surface	-1.104	develop
-1.595	better	-0.970	transmission	-0.861	effects	-1.112	curves
-1.728	predictions	-1.039	atmospheric	-0.946	evolution	-1.318	apply
-8.344	proposed	-1.045	planet	-0.987	mass	-1.935	using
-11.004	conclude	-1.097	models	-1.652	models	-1.970	use
-11.936	compared	-1.288	present	-2.007	present	-2.461	present
-13.480	propose	-1.436	model	-2.208	model	-2.855	model
-19.334	compare	-3.609	<BIAS>	-3.589	<BIAS>	-3.463	<BIAS>



Structural Corpora

1) Multi Index Dataframe

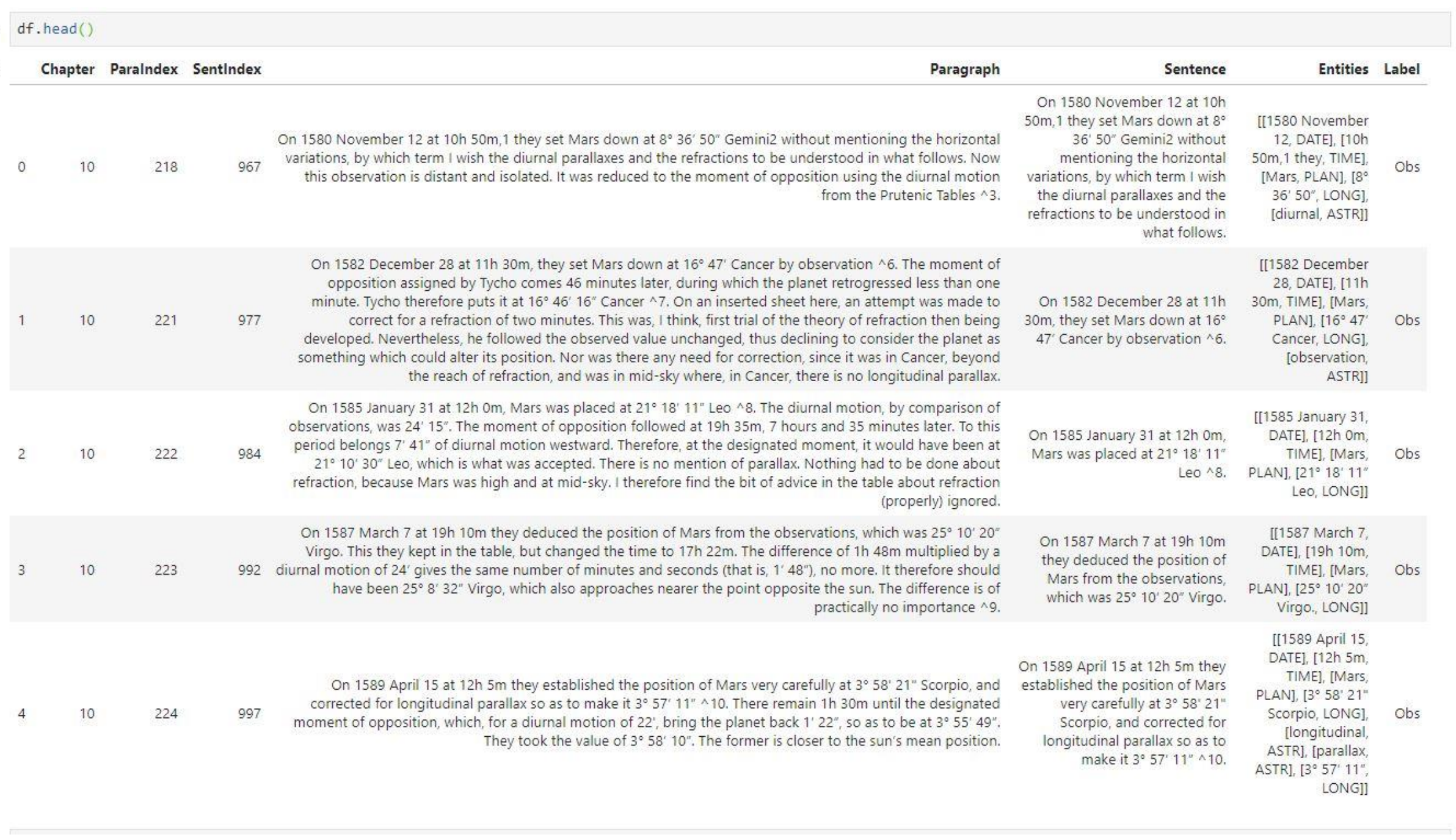
2) Sentence Structure with Spacy

a) Customized Entities

b) Dependency Parsing

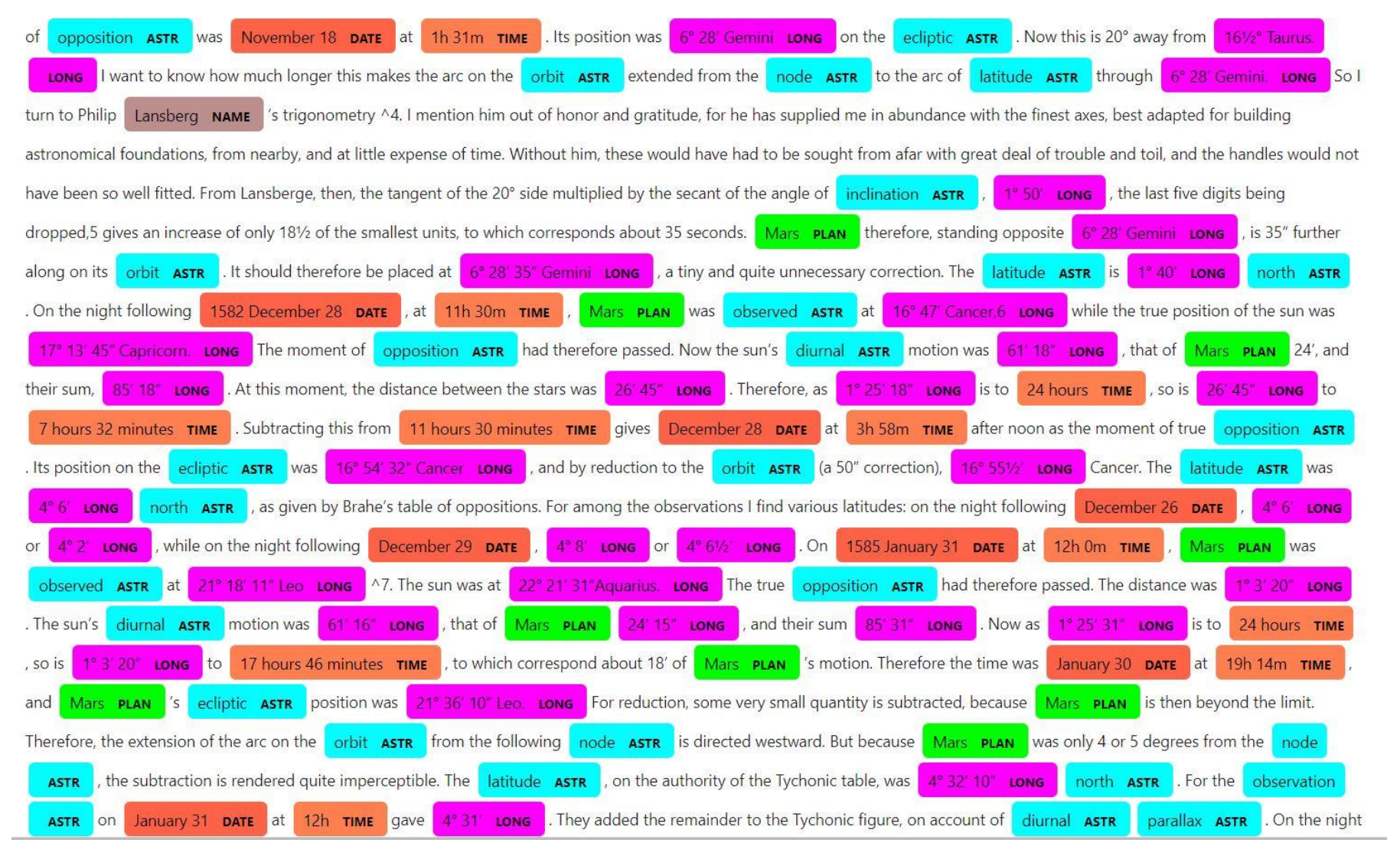
Customized Named Entity Recognition (NER) [2]

Key Idea. Hybrid approach; Pattern based & deep learning to provide NER Model by developing spaCy.



```
12]: df.head()
```

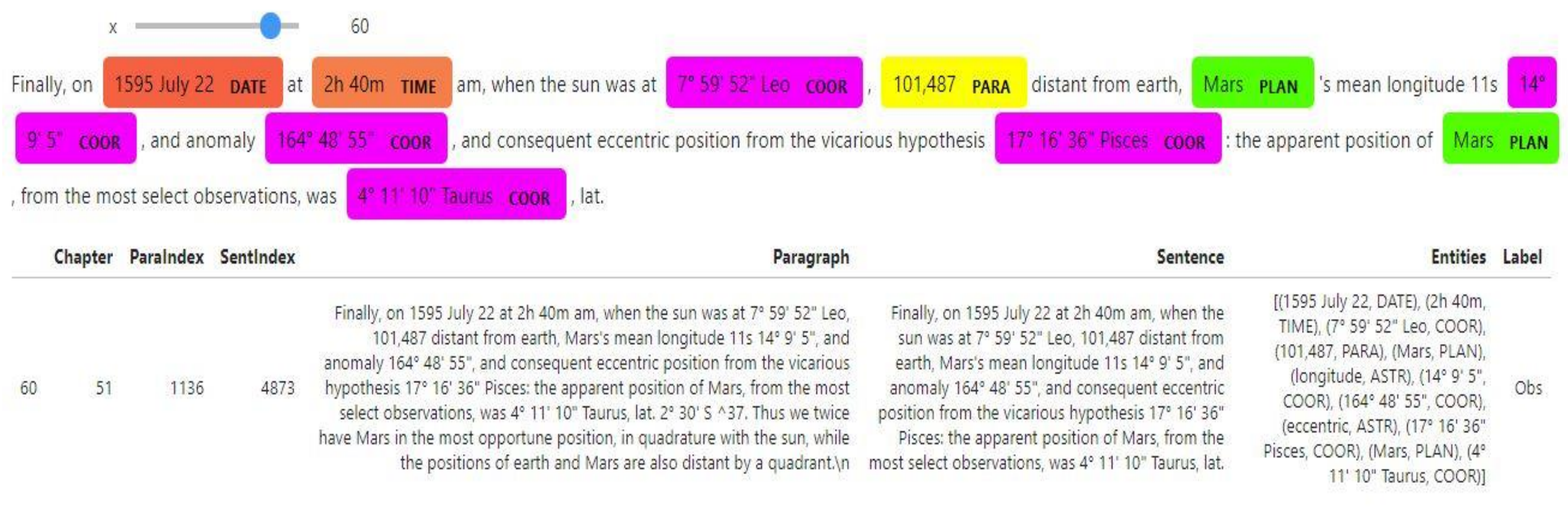
12]:	Chapter	ParaIndex	SentIndex	Paragraph	Sentence	Entities	Label
0	10	218	967	On 1580 November 12 at 10h 50m,1 they set Mars down at 8° 36′ 50″ Gemini2 without mentioning the horizontal variations, by which term I wish the diurnal parallaxes and the refractions to be understood in what follows. Now this observation is distant and isolated. It was reduced to the moment of opposition using the diurnal motion from the Prutenic Tables ^3.	On 1580 November 12 at 10h 50m,1 they set Mars down at 8° 36′ 50″ Gemini2 without mentioning the horizontal variations, by which term I wish the diurnal parallaxes and the refractions to be understood in what follows.	[[1580 November 12, DATE], [10h 50m,1 they, TIME], [Mars, PLAN], [8° 36′ 50″, LONG], [diurnal, ASTR]]	Obs
1	10	221	977	On 1582 December 28 at 11h 30m, they set Mars down at 16° 47′ Cancer by observation ^6. The moment of opposition assigned by Tycho comes 46 minutes later, during which the planet retrogressed less than one minute. Tycho therefore puts it at 16° 46′ 16″ Cancer ^7. On an inserted sheet here, an attempt was made to correct for a refraction of two minutes. This was, I think, first trial of the theory of refraction then being developed. Nevertheless, he followed the observed value unchanged, thus declining to consider the planet as something which could alter its position. Nor was there any need for correction, since it was in Cancer, beyond the reach of refraction, and was in mid-sky where, in Cancer, there is no longitudinal parallax.	On 1582 December 28 at 11h 30m, they set Mars down at 16° 47′ Cancer by observation ^6.	[[1582 December 28, DATE], [11h 30m, TIME], [Mars, PLAN], [16° 47′ Cancer, LONG], [observation, ASTR]]	Obs
2	10	222	984	On 1585 January 31 at 12h 0m, Mars was placed at 21° 18′ 11″ Leo ^8. The diurnal motion, by comparison of observations, was 24′ 15″. The moment of opposition followed at 19h 35m, 7 hours and 35 minutes later. To this period belongs 7′ 41″ of diurnal motion westward. Therefore, at the designated moment, it would have been at 21° 10′ 30″ Leo, which is what was accepted. There is no mention of parallax. Nothing had to be done about refraction, because Mars was high and at mid-sky. I therefore find the bit of advice in the table about refraction (properly) ignored.	On 1585 January 31 at 12h 0m, Mars was placed at 21° 18′ 11″ Leo ^8.	[[1585 January 31, DATE], [12h 0m, TIME], [Mars, PLAN], [21° 18′ 11″ Leo, LONG]]	Obs
3	10	223	992	On 1587 March 7 at 19h 10m they deduced the position of Mars from the observations, which was 25° 10′ 20″ Virgo. This they kept in the table, but changed the time to 17h 22m. The difference of 1h 48m multiplied by a diurnal motion of 24′ gives the same number of minutes and seconds (that is, 1′ 48″), no more. It therefore should have been 25° 8′ 32″ Virgo, which also approaches nearer the point opposite the sun. The difference is of practically no importance ^9.	On 1587 March 7 at 19h 10m they deduced the position of Mars from the observations, which was 25° 10′ 20″ Virgo.	[[1587 March 7, DATE], [19h 10m, TIME], [Mars, PLAN], [25° 10′ 20″ Virgo., LONG]]	Obs
4	10	224	997	On 1589 April 15 at 12h 5m they established the position of Mars very carefully at 3° 58′ 21" Scorpio, and corrected for longitudinal parallax so as to make it 3° 57′ 11″ ^10. There remain 1h 30m until the designated moment of opposition, which, for a diurnal motion of 22′, bring the planet back 1′ 22″, so as to be at 3° 55′ 49″. They took the value of 3° 58′ 10″. The former is closer to the sun's mean position.	On 1589 April 15 at 12h 5m they established the position of Mars very carefully at 3° 58′ 21" Scorpio, and corrected for longitudinal parallax so as to make it 3° 57′ 11″ ^10.	[[1589 April 15, DATE], [12h 5m, TIME], [Mars, PLAN], [3° 58′ 21" Scorpio, LONG], [longitudinal, ASTR], [parallax, ASTR], [3° 57′ 11″, LONG]]	Obs



of opposition ASTR was November 18 DATE at 1h 31m TIME . Its position was 6° 28′ Gemini LONG on the ecliptic ASTR . Now this is 20° away from 16½° Taurus. LONG I want to know how much longer this makes the arc on the orbit ASTR extended from the node ASTR to the arc of latitude ASTR through 6° 28′ Gemini. LONG So I turn to Philip Lansberg NAME 's trigonometry ^4. I mention him out of honor and gratitude, for he has supplied me in abundance with the finest axes, best adapted for building astronomical foundations, from nearby, and at little expense of time. Without him, these would have had to be sought from afar with great deal of trouble and toil, and the handles would not have been so well fitted. From Lansberge, then, the tangent of the 20° side multiplied by the secant of the angle of inclination ASTR , 1° 50′ LONG , the last five digits being dropped,5 gives an increase of only 18½ of the smallest units, to which corresponds about 35 seconds. Mars PLAN therefore, standing opposite 6° 28′ Gemini LONG , is 35″ further along on its orbit ASTR . It should therefore be placed at 6° 28′ 35″ Gemini LONG , a tiny and quite unnecessary correction. The latitude ASTR is 1° 40′ LONG north ASTR . On the night following 1582 December 28 DATE , at 11h 30m TIME , Mars PLAN was observed ASTR at 16° 47′ Cancer,6 LONG while the true position of the sun was 17° 13′ 45″ Capricorn. LONG The moment of opposition ASTR had therefore passed. Now the sun's diurnal ASTR motion was 61′ 18″ LONG , that of Mars PLAN 24′, and their sum, 85′ 18″ LONG . At this moment, the distance between the stars was 26′ 45″ LONG . Therefore, as 1° 25′ 18″ LONG is to 24 hours TIME , so is 26′ 45″ LONG to 7 hours 32 minutes TIME . Subtracting this from 11 hours 30 minutes TIME gives December 28 DATE at 3h 58m TIME after noon as the moment of true opposition ASTR . Its position on the ecliptic ASTR was 16° 54′ 32″ Cancer LONG , and by reduction to the orbit ASTR (a 50″ correction), 16° 55½′ LONG Cancer. The latitude ASTR was 4° 6′ LONG north ASTR , as given by Brahe's table of oppositions. For among the observations I find various latitudes: on the night following December 26 DATE , 4° 6′ LONG or 4° 2′ LONG , while on the night following December 29 DATE , 4° 8′ LONG or 4° 6½′ LONG . On 1585 January 31 DATE at 12h 0m TIME , Mars PLAN was observed ASTR at 21° 18′ 11″ Leo LONG ^7. The sun was at 22° 21′ 31″Aquarius. LONG The true opposition ASTR had therefore passed. The distance was 1° 3′ 20″ LONG . The sun's diurnal ASTR motion was 61′ 16″ LONG , that of Mars PLAN 24′ 15″ LONG , and their sum 85′ 31″ LONG . Now as 1° 25′ 31″ LONG is to 24 hours TIME , so is 1° 3′ 20″ LONG to 17 hours 46 minutes TIME , to which correspond about 18′ of Mars PLAN 's motion. Therefore the time was January 30 DATE at 19h 14m TIME , and Mars PLAN 's ecliptic ASTR position was 21° 36′ 10″ Leo. LONG For reduction, some very small quantity is subtracted, because Mars PLAN is then beyond the limit. Therefore, the extension of the arc on the orbit ASTR from the following node ASTR is directed westward. But because Mars PLAN was only 4 or 5 degrees from the node ASTR , the subtraction is rendered quite imperceptible. The latitude ASTR , on the authority of the Tychonic table, was 4° 32′ 10″ LONG north ASTR . For the observation ASTR on January 31 DATE at 12h TIME gave 4° 31′ LONG . They added the remainder to the Tychonic figure, on account of diurnal ASTR parallax ASTR . On the night

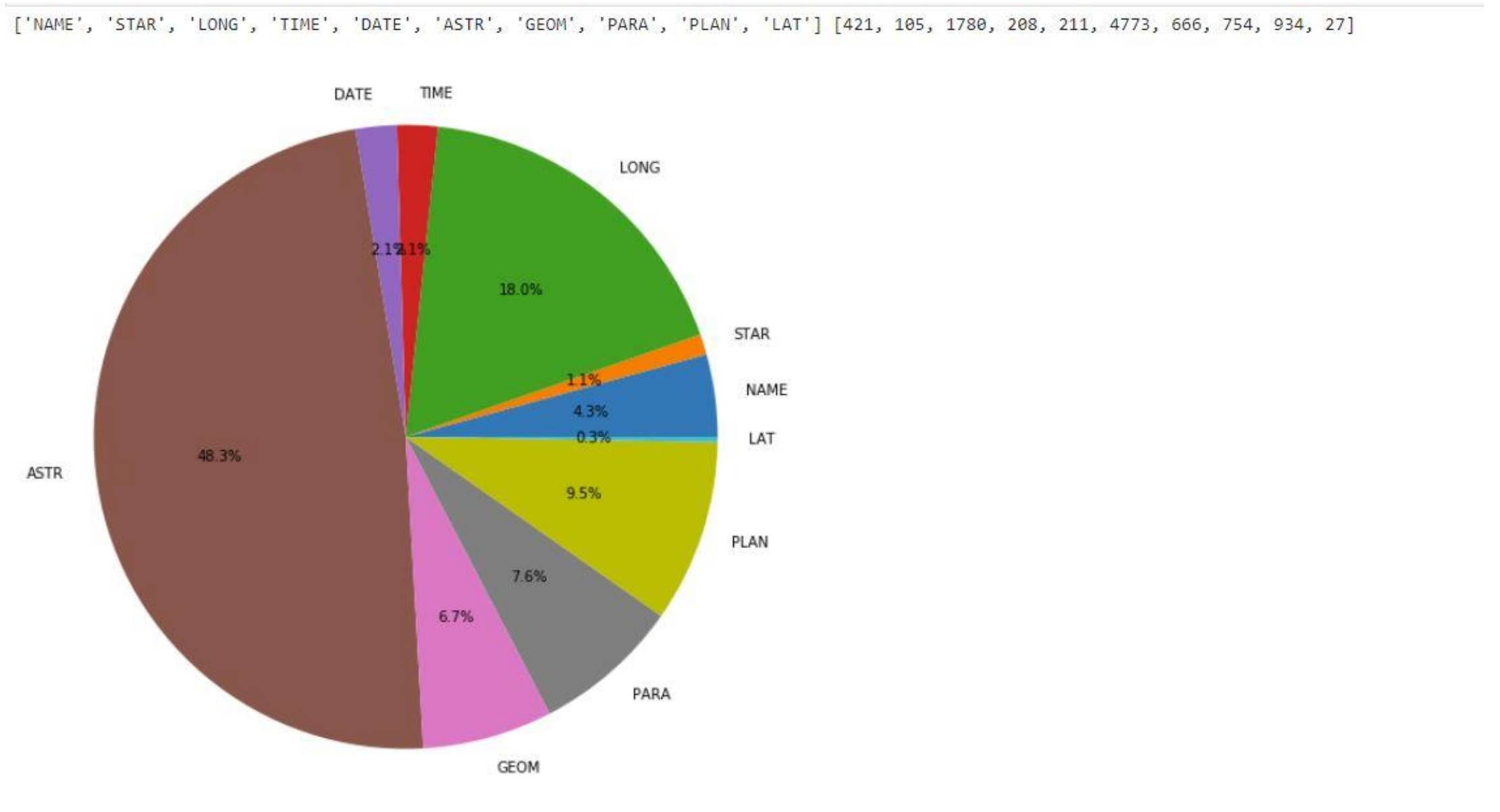
Explicit Observation Extraction [2]

Key Idea. Extract Semantic from the corpus using customized NER model.



Finally, on 1595 July 22 DATE at 2h 40m TIME am, when the sun was at 7° 59' 52" Leo COOR , 101,487 PARA distant from earth, Mars PLAN 's mean longitude 11s 14° 9' 5" COOR , and anomaly 164° 48' 55" COOR , and consequent eccentric position from the vicarious hypothesis 17° 16' 36" Pisces COOR : the apparent position of Mars PLAN , from the most select observations, was 4° 11' 10" Taurus COOR , lat.

	Chapter	ParaIndex	SentIndex	Paragraph	Sentence	Entities	Label
60	51	1136	4873	Finally, on 1595 July 22 at 2h 40m am, when the sun was at 7° 59' 52" Leo, 101,487 distant from earth, Mars's mean longitude 11s 14° 9' 5", and anomaly 164° 48' 55", and consequent eccentric position from the vicarious hypothesis 17° 16' 36" Pisces: the apparent position of Mars, from the most select observations, was 4° 11' 10" Taurus, lat. 2° 30' S ^37. Thus we twice have Mars in the most opportune position, in quadrature with the sun, while the positions of earth and Mars are also distant by a quadrant.\n	Finally, on 1595 July 22 at 2h 40m am, when the sun was at 7° 59' 52" Leo, 101,487 distant from earth, Mars's mean longitude 11s 14° 9' 5", and anomaly 164° 48' 55", and consequent eccentric position from the vicarious hypothesis 17° 16' 36" Pisces: the apparent position of Mars, from the most select observations, was 4° 11' 10" Taurus, lat.	[(1595 July 22, DATE), (2h 40m, TIME), (7° 59' 52" Leo, COOR), (101,487, PARA), (Mars, PLAN), (longitude, ASTR), (14° 9' 5", COOR), (164° 48' 55", COOR), (eccentric, ASTR), (17° 16' 36" Pisces, COOR), (Mars, PLAN), (4° 11' 10" Taurus, COOR)]	Obs



```
['NAME', 'STAR', 'LONG', 'TIME', 'DATE', 'ASTR', 'GEOM', 'PARA', 'PLAN', 'LAT'] [421, 105, 1780, 208, 211, 4773, 666, 754, 934, 27]
```

Future Direction: Semantic Objects

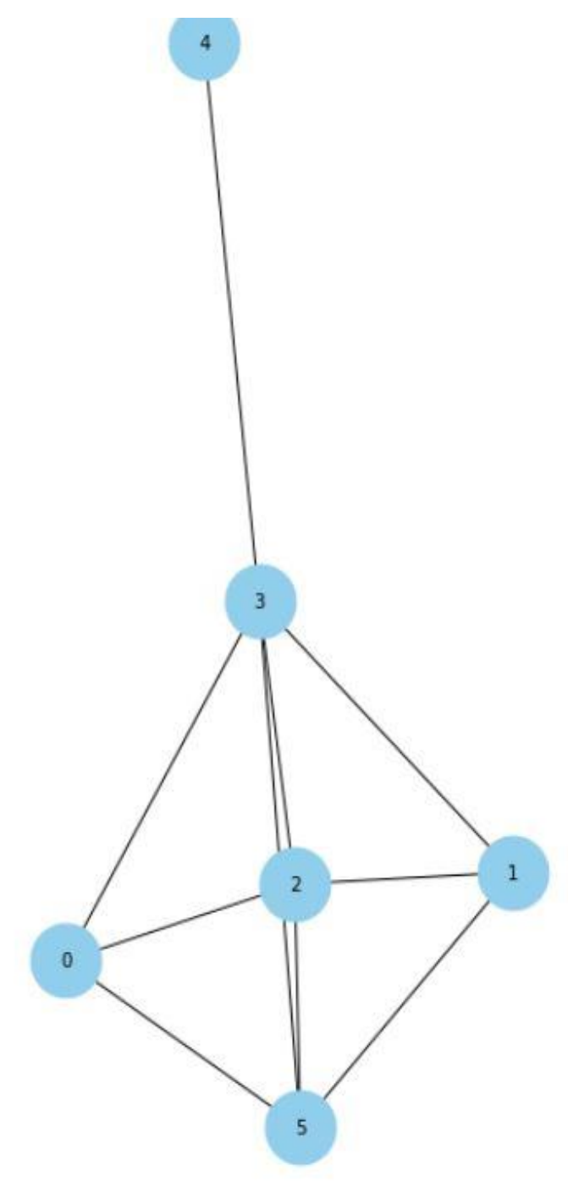
- *Tree structure*

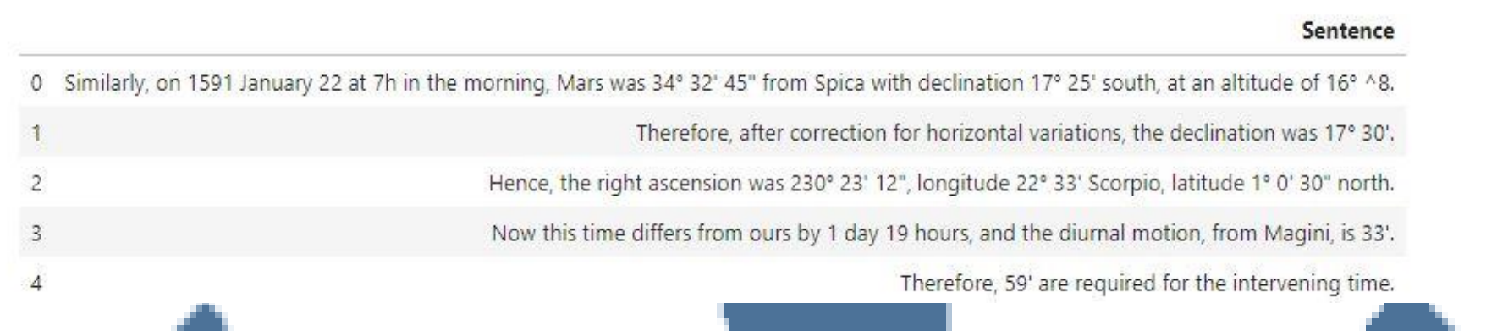
Idea. Integrate sentences using xml format to have semantic annotation for extracted observations as a first use case

<APPELL> REICH MIR MAL DEN 10-ER SCHLÜSSEL!</APPELL>
<APPELL> REICH MIR MAL DEN <WERKZ> 10-ER SCHLÜSSEL! </WERKZ>!</APPELL>
?
!

- *Semantic Relation Extraction*

Idea. Extract structure of argumentation components using semantic objects





	Sentence
0	Similarly, on 1591 January 22 at 7h in the morning, Mars was 34° 32' 45" from Spica with declination 17° 25' south, at an altitude of 16° ^8.
1	Therefore, after correction for horizontal variations, the declination was 17° 30'.
2	Hence, the right ascension was 230° 23' 12", longitude 22° 33' Scorpio, latitude 1° 0' 30" north.
3	Now this time differs from ours by 1 day 19 hours, and the diurnal motion, from Magini, is 33'.
4	Therefore, 59' are required for the intervening time.

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).