WATCH · VEDIC AND UPANIṢADIC TEXTS / STUDY NOTES

Dating the Vedic Corpus - Oliver Hellwig (University of Zurich)

Study notes by Sanskrit AI Universe, based on this video

Video by Oliver Hellwig (University of Zurich) · Heidelberg - KRS · 9 Feb 2024

Not the creator's own slides

Sanskrit AI Universe

About these notes

— Official title: Dating the Vedic Corpus - Oliver Hellwig (University of Zurich)

— Speaker and channel: Oliver Hellwig (University of Zurich) · Heidelberg - KRS

— Published 9 Feb 2024 · Running time 1 h 01 min · https://www.youtube.com/watch?v=CQI3nR1dra0

— Official description: In this talk I will discuss what quantitative methods can reveal about the chronology of the Vedic corpus. Starting with a discussion of non-chronological influence factors and methods for detecting them, the presentation moves forward to a Bayesian approach that derives a chronological structure from the development of linguistic features. In addition, I will sketch how recent advances in Natural Language Processing can be applied to this question. Dr. Oliver Hellwig is a PostDoc researcher at

— Catalogue caution: Dates are statistical estimates, with error margins.

— These pages are study notes drawn from the video's own title, description and captions, plus the journal's already published Watch note. They are not a transcript and not a substitute for watching the film.

— The journal treats videos as leads, not as sources you should cite on their own. Check the papers and editions behind a claim.

— Caption source: English (auto-generated). Auto-captions can mis-hear words; obvious proper-name spellings were given in IAST where the journal already uses them. Distinctive phrases follow the captions closely.

From the official description

— In this talk I will discuss what quantitative methods can reveal about the chronology of the Vedic corpus.

— Starting with a discussion of non-chronological influence factors and methods for detecting them, the presentation moves forward to a Bayesian approach that derives a chronological structure from the development of linguistic features.

— In addition, I will sketch how recent advances in Natural Language Processing can be applied to this question.

— Dr. Oliver Hellwig is a PostDoc researcher at the Department of Comparative Linguistics, University of Zurich.

From the talk · 0:09–6:54

— In other places in the world this is our final talk that we're having this semester we will continue in April and today I'm very happy.

— These statements are always based on a very shaky text historical ground because we do not understand the text history and the other way around it's equally complicated.

— Because some researchers including me have tried to derive claims about the textual history from linguistic developments but if we do not really understand the linguistic trajectories.

— We cannot really derive good historical claims from these from these shaky grounds again and this is basically the basic situation in which our project started.

From the talk · 6:57–13:06

— They really interacted with the linguistics surface and this is really unpleasant there are a few papers like from Jameson or Hawk about these topics but it's nothing.

— That time or is this rather a genre effect this genre Sūtra literature has this genre influence the use of compounds we also have as.

— In other Indian traditions a very fragmentary transmission so much of the interesting texts have not been transmitted like texts written by women text written.

— In natural language processing are not really interested in dionic corpor they are not really interested in dating texts there has been some recent research in this.

From the talk · 13:10–19:05

— These guys from Survivor analysis do not care about cases like thetic chronology which is really missing or lacking data and it's it's has some interesting ideas but.

— This is a system was introduced byit and later on by adapted by h Martin KL and this system stratifies so to say.

— The Vedic literature into five or six layers starting with old vtic this is basically the core of the Rick vaa then we have Mantra language.

— I call Flat models and until recently this was dominating approach in all quantitative studies what you have to do here you have to write down formally.

From the talk · 19:09–25:19

— That it also fails foric and I should do something else so yeah that's why I will mention Latin and Icelandic later basically as evaluation cases so cases so.

— In general is syntax so basically we focused on dionic changes in syntax and to be able to do so we first needed some syntactic data and that's.

— Then just correct the errors and unfortunately interannotator agreement is quite low we have a paper about this means when you ask to human annotators to annotate.

— The rck vaa 19% of our syntactic data come from later metrical literature and so on so on and here is one point I want to mention here.

From the talk · 25:21–31:45

— In later Sanskrit and the central question I asked in this paper was if you use a lot of compounds something else must become less frequent so.

— The right ritual knowledge you will be successful or whatever so and here you have this modification or the specification of this the is I Vian.

— How do these factors influence the frequency of using compounds versus using some kind of other construction and construction and sorry a few words with technical stuff so how.

— We asked okay we can input time just as these six levels I mentioned so all old vaa later Mantra stuff and so on but we could also ask.

From the talk · 31:49–38:07

— I want to say here is when you see for instance an increasing Trend in some kind of compound or so it could well be the case that perhaps.

— The Shaka became more frequent over time it's not it's may not be a time effect may be an effect of school affiliation and for instance one thing that's.

— This model what we call poetical is absolutely or quite irrelevant so this really has no influence on how these constructions are used.

— So you can see you can find out quite a lot of interesting information when you use such basian approaches and this is one last Point here as I said.

From the talk · 38:09–44:10

— The best you can do is to move these points away from these centers of the periods so that they lie on these on this estimated line and this basically.

— This is a very trivial idea but this gives you gives you improved temporal or improved dionic or chronological estimates of the text and then you have these improved estimates and.

— Those who are interested this basically nothing else than a binomial glm so here you have the estimate in the linear space then you do omal stuff and the gradient and.

— We have these initial estimates and then we move these points so that they lie on this estimated curve and you see some points are moved only a bit.

From the talk · 44:13–50:20

— We get these results you see here at the in the left half of this table you see the results for the errors made for individual text so for Icelandic.

— For instance you can expect like on average 50 years of Errors when you apply this model to these texts and for Latin you can expect in.

— The situation is much better for the boundary estimation or for the period estimation so in both for both languages please just ignore the first line here for both languages.

— This is not really truly relevant so basically the model predicts okay when we work with these periods.

From the talk · 50:23–56:10

— Next Period and quite notably these old parts of these texts are placed right at the end of these of this old po period but.

— For delving into the details of very textual history but I find it interesting so and this one starts with the old s brahmanas andat so basically also when you.

— This model seems to support it also for instance when you just a few details when you look at this here Chand is placed more to the right and.

— The whole of the argument and argument and at least the model seems to support the second the second idea it seems not to be very early last slide.

From the talk · 56:11–1:01:46

— One I just sketched with it so we just take all VC text all sunr text and throw it in this model and train it.

— Now we ask the model these relations were clear but we have a lot of relations that are we are not sure about like what's the relation between brah.

— The same time as the Sha stuff we can ask the model about this or is hyada older or younger than shaaka and whatever so this is quite useful.

— And it could be additional evidence from another point of view to support you in getting a better more accurate image of the vtic timeline.

Sources and credit

— Study notes by Sanskrit AI Universe, based on the video “Dating the Vedic Corpus - Oliver Hellwig (University of Zurich)”. These are not the creator's own slides.

— Video by Oliver Hellwig (University of Zurich), posted on the YouTube channel Heidelberg - KRS.

— Watch the video: https://www.youtube.com/watch?v=CQI3nR1dra0

— Captions used: English (auto-generated) (en), retrieved on 24 November 2026 via a public Invidious instance. Auto-captions can mis-hear words; obvious proper-name spellings were given in IAST where the journal already uses them.

— Also used: the official video title and description, and the Watch catalogue note already published in this journal (presenter and why the video is listed).

— If the Founding Editor later supplies his own deck for this video, replace the PPTX and PDF in src_assets/slides/ (see the journal README, “How to swap in RK's own slides”).