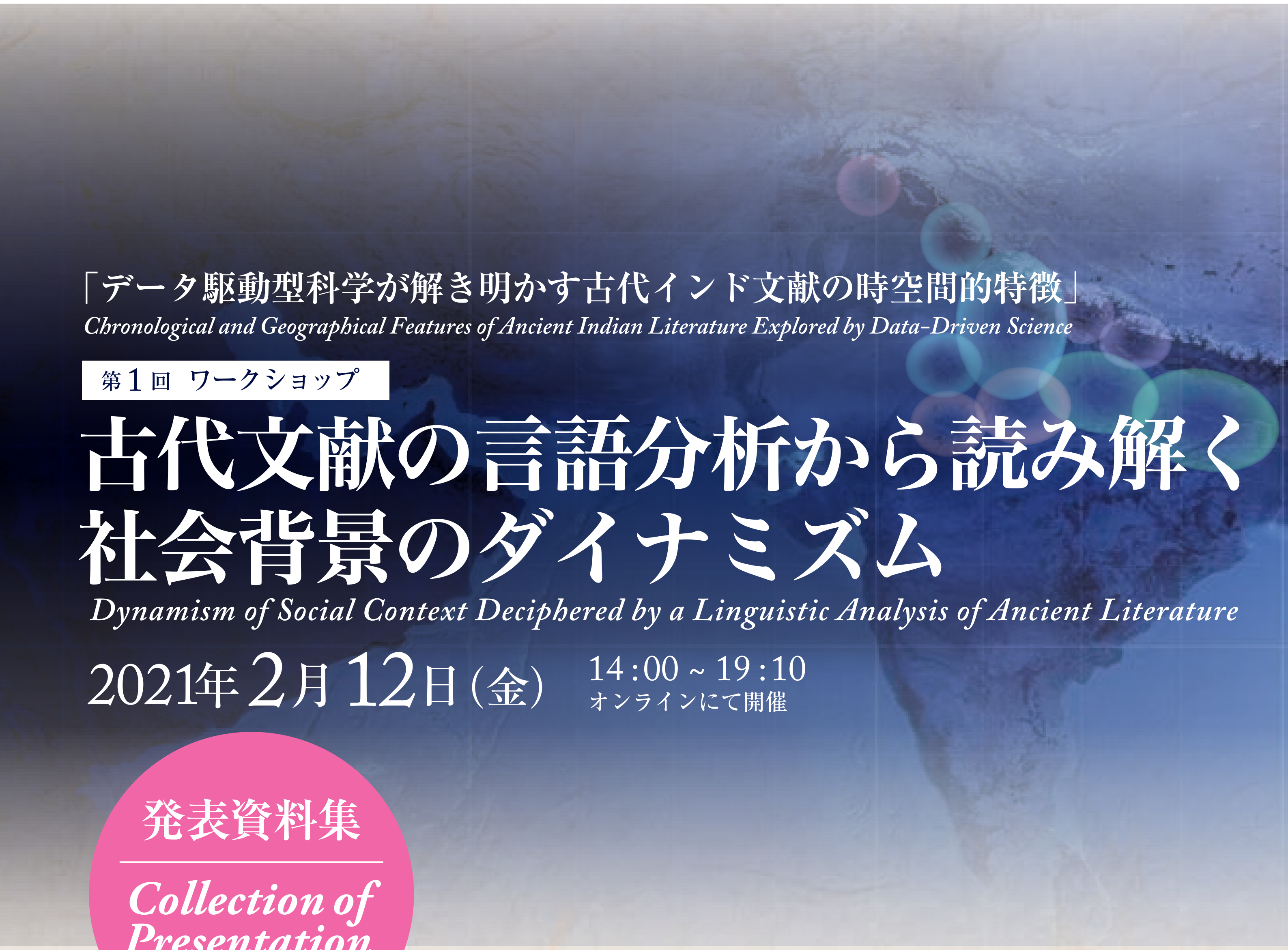
京都大学研究支援SPIRITS
「知の越境」融合チーム研究プログラム
2020-2021年度採択プロジェクト

京都大学
KYOTO UNIVERSITY

SPIRITS
SUPPORTING PROGRAM FOR INTERACTION-BASED INITIATIVE TEAM STUDIES



「データ駆動型科学が解き明かす古代インド文献の時空間的特徴」
Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science

第1回 ワークショップ

古代文献の言語分析から読み解く社会背景のダイナミズム

Dynamism of Social Context Deciphered by a Linguistic Analysis of Ancient Literature

2021年2月12日（金）　14:00 ~ 19:10
オンラインにて開催

発表資料集

Collection of Presentation Slides

主催 SPIRITSプロジェクト「データ駆動型科学が解き明かす古代インド文献の時空間的特徴」
（天野恭子、夏川浩明、Oliver Hellwig、京極祐希）

共催 京都大学学術情報メディアセンター、天野恭子白眉プロジェクト「古代インド祭式文献の言語および社会的・文化的成立背景の研究」、科研費挑戦的研究（萌芽）「古代インド文献成立過程解明に向けた文体計量分析のためのデータベース構築」（研究代表者：天野恭子、20K20697）

白眉
Kyoto Univ

Contents｜目次

■Opening｜オープニング

Problems in the Formation of the Vedas, Ancient Indian Religious Texts
古代インド宗教文献ヴェーダの成立を巡る諸問題
天野 恭子（京都大学 白眉センター･人文科学研究所）
Kyoko Amano（Kyoto University, Institute for Research in Humanities / Hakubi Center） ······ **2**

■Presentation 1｜報告１

The Possibility of Information Visualization and Data Analysis for Ancient Indian Literature
古代インド文献を対象とした情報可視化やデータ分析の可能性
夏川 浩明（京都大学 学術情報メディアセンター）
Hiroaki Natsukawa（Kyoto University, Academic Center for Computing and Media Studies） ······ **12**

■Presentation 2｜報告２

Relationship among Vedic Schools Deciphered by the Visualization of Mantra Collocation
マントラ共起関係の可視化から読み解くヴェーダ学派間の関係性
天野 恭子（京都大学 白眉センター･人文科学研究所）
Kyoko Amano（Kyoto University, Institute for Research in Humanities / Hakubi Center） ······ **21**

■Presentation 3｜報告３

Citation Prediction Using Academic Paper Data and Application for Surveys
学術論文データを用いた引用数予測とサーベイへの活用
濵地 瞬（京都大学 工学研究科）
Shun Hamachi（Kyoto University, Graduate School of Engineering） ······ **30**

■Presentation 4｜報告４

Measuring the Semantic Similarity between the Chapters of Taittirīya Saṃhitā Using a Vector Space Model
ベクトル空間モデルによる『タイッティリーヤ・サンヒター』の章間類似度比較
京極 祐希（Leipzig University, Indology）
Yuki Kyogoku ······ **41**

■Presentation 5｜報告５

Dating Vedic Texts with Computational Models: Algorithmic Considerations and Data Selection
Oliver Hellwig（University of Zurich, Department of Comparative Language Science） ······ **50**

■Presentation 6｜報告６

morogram: Background, History, and Purpose of a Tool for East Asian Text Analysis
morogram: 東アジア文献分析ツールの開発の経緯と目的
師 茂樹（花園大学 文学部）
Shigeki Moro（Hanazono University, Faculty of Letter） ······ **65**

■Reference｜資料

Notes and references for presentations by Kyoko Amano ······ **69**

Opening｜オープニング

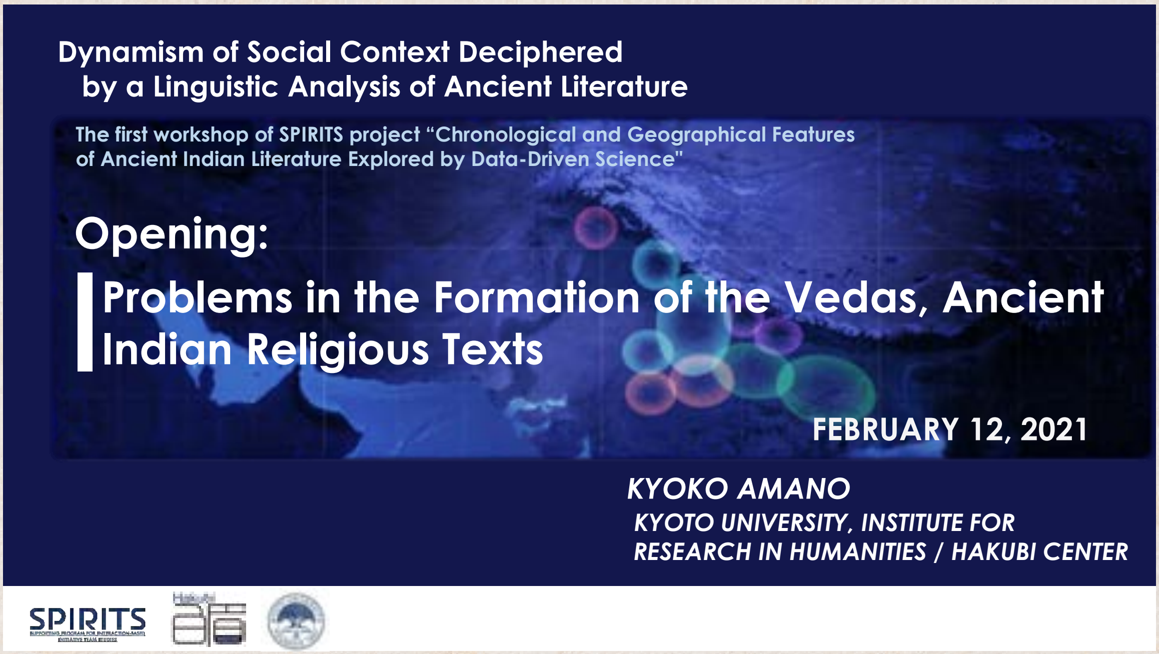
Problems in the Formation of the Vedas, Ancient Indian Religious Texts

古代インド宗教文献ヴェーダの成立を巡る諸問題

天野恭子（京都大学 白眉センター・人文科学研究所）

Kyoko Amano (Kyoto University, Institute for Research in Humanities / Hakubi Center)

1



Dynamism of Social Context Deciphered
by a Linguistic Analysis of Ancient Literature

The first workshop of SPIRITS project "Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science"

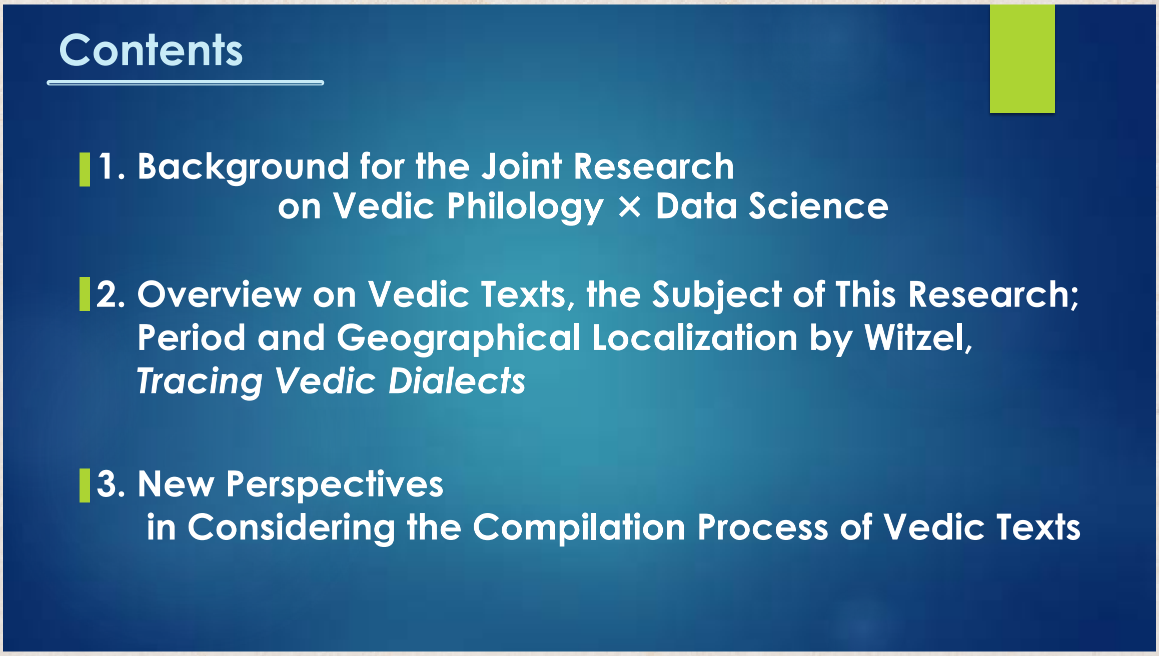
Opening:

Problems in the Formation of the Vedas, Ancient Indian Religious Texts

FEBRUARY 12, 2021

KYOKO AMANO
KYOTO UNIVERSITY, INSTITUTE FOR RESEARCH IN HUMANITIES / HAKUBI CENTER

2



Contents

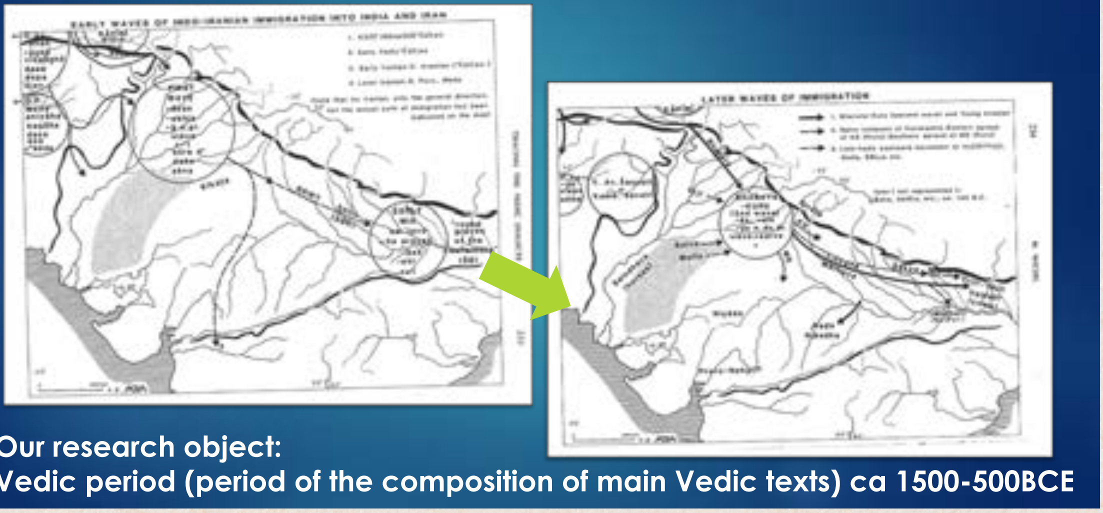
1. Background for the Joint Research on Vedic Philology × Data Science

2. Overview on Vedic Texts, the Subject of This Research; Period and Geographical Localization by Witzel, *Tracing Vedic Dialects*

3. New Perspectives in Considering the Compilation Process of Vedic Texts

3

Vedas: religious texts of Indo-Aryan people who immigrated in South Asia since ca 1500BCE


EARLY WAVES OF INDO-IRANIAN IMMIGRATION INTO INDIA AND IRAN
LATER WAVES OF IMMIGRATION

Our research object:
Vedic period (period of the composition of main Vedic texts) ca 1500-500BCE

4

Vedic texts

- oral tradition; since ca 10CE or later written in manuscripts
- Vedas describe religion and mythology
  (no direct description of historical events)

Vedic period

- no other historical materials
- no archeological evidence for towns, cities and kingdoms
  transitional period from nomadic lifestyle to sedentism,
  half-pastoral, half-agricultural, then more and more agriculture

5

Linguistic Analysis and Visualization

- Database of Vedic texts with morpho-lexical annotation
  in the Digital Corpus of Sanskrit (DCS) by Oliver Hellwig

Texts

Help

Select a text:

Select a chapter:

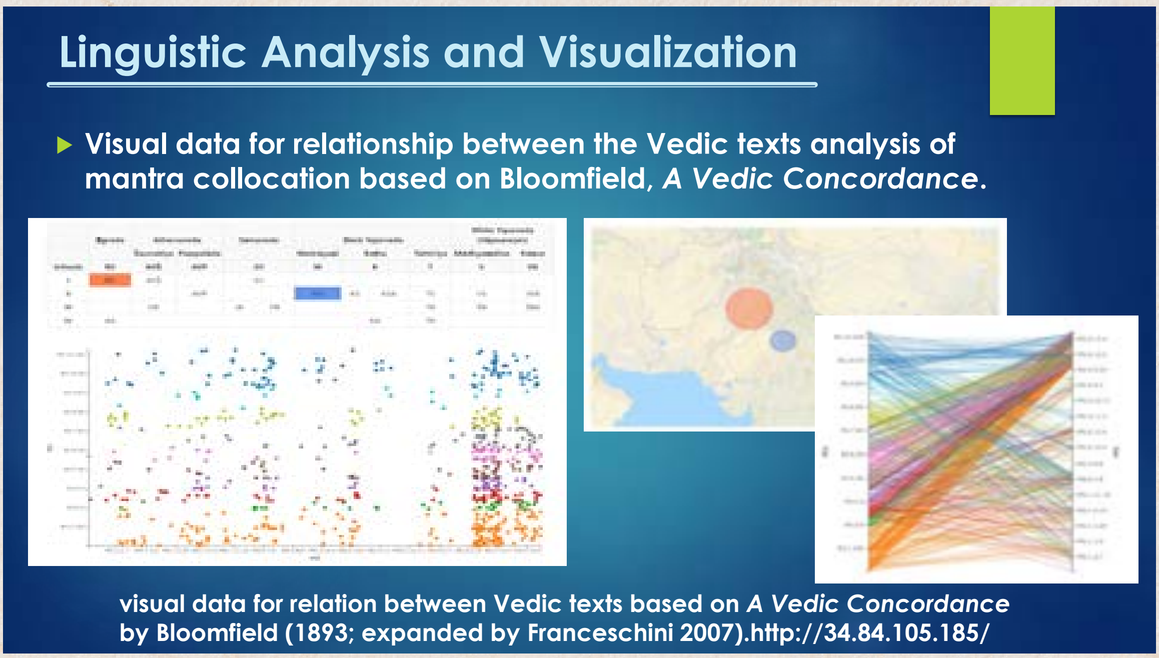
Show parallels Show headlines

Use dependency labeler

- Click on a sentence to show its analysis.
- Keep the mouse pointer over a lemma to show its meanings.

http://www.sanskrit-linguistics.org/dcs/index.php

6



Linguistic Analysis and Visualization

- Visual data for relationship between the Vedic texts analysis of mantra collocation based on Bloomfield, *A Vedic Concordance*.

visual data for relation between Vedic texts based on *A Vedic Concordance* by Bloomfield (1893; expanded by Franceschini 2007).http://34.84.105.185/

7

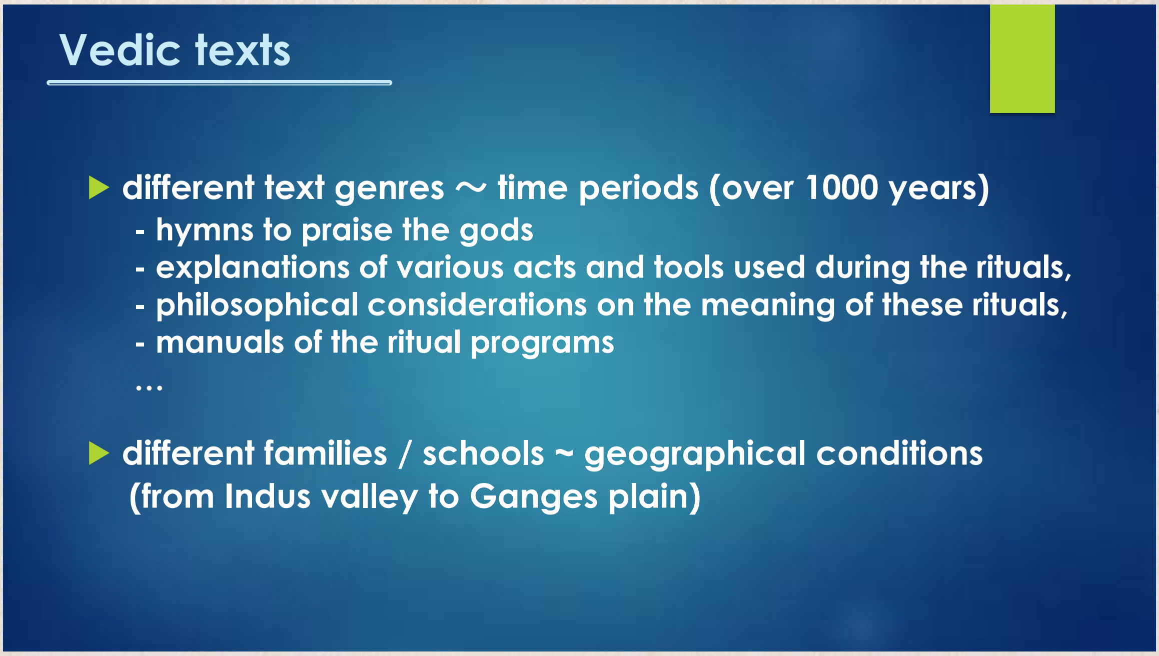
Contents

1. Background for the Joint Research
on Vedic Philology × Data Science

2. Overview on Vedic Text, the Subject of This Research;
Period and Geographical Localization by Witzel,
Vedic Dialects

3. New Perspectives
in Considering the Compilation Process of Vedic Texts

8



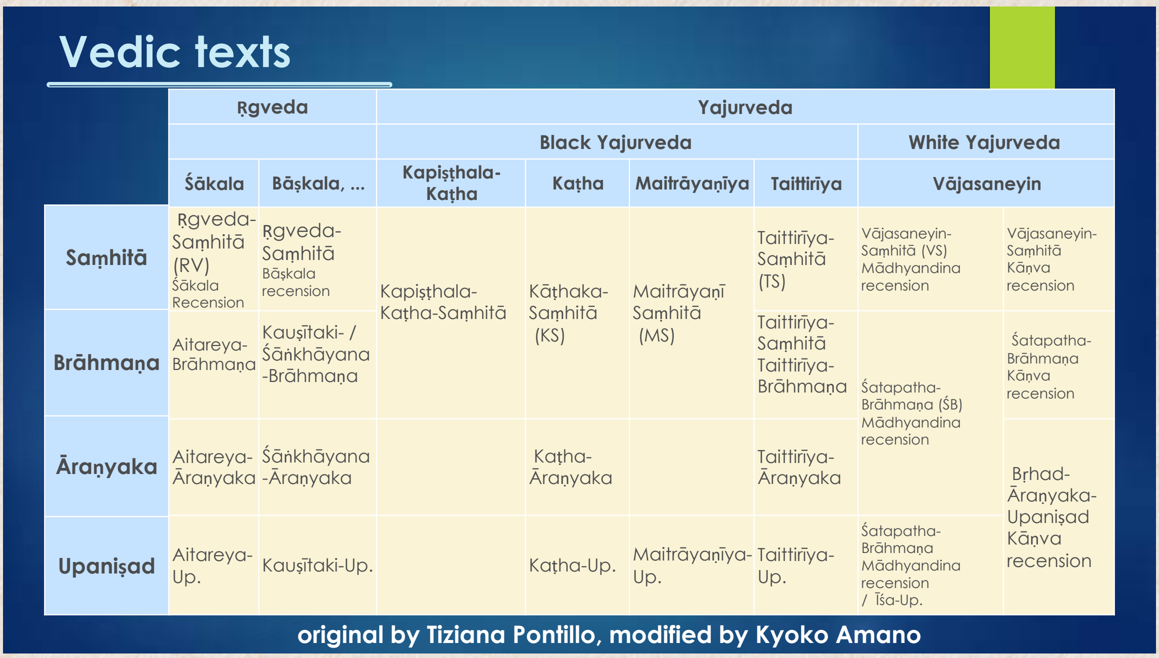
Vedic texts

- different text genres ～ time periods (over 1000 years)
  - hymns to praise the gods
  - explanations of various acts and tools used during the rituals,
  - philosophical considerations on the meaning of these rituals,
  - manuals of the ritual programs

  …

- different families / schools ~ geographical conditions
(from Indus valley to Ganges plain)

9

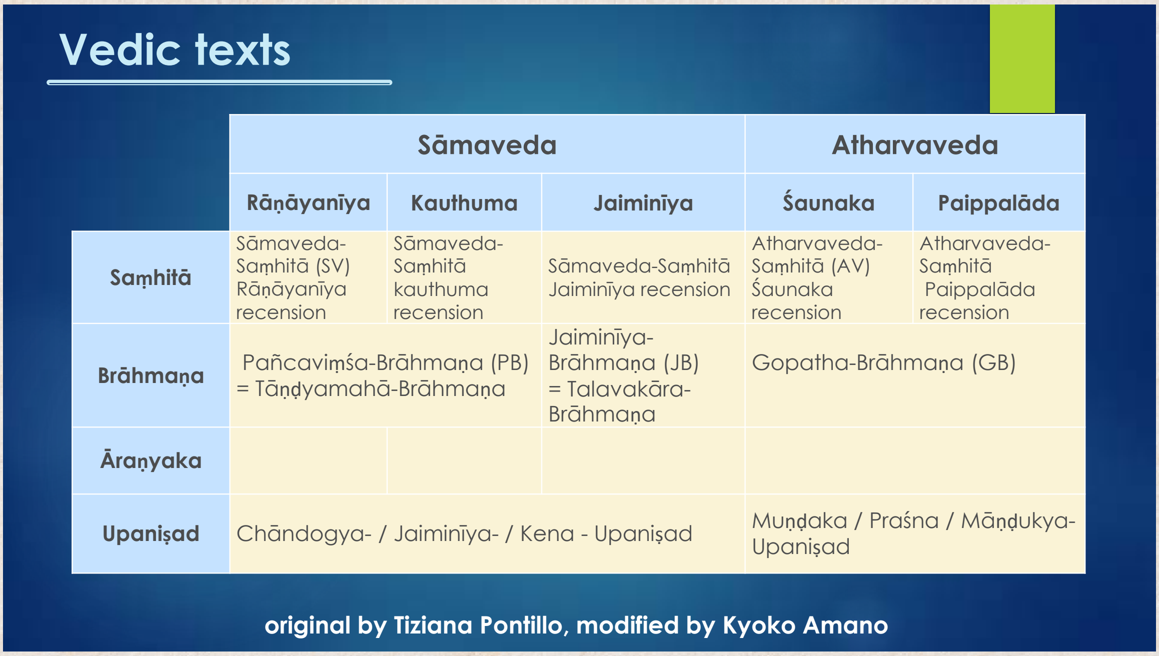


Vedic texts

	Ṛgveda		Yajurveda					
			Black Yajurveda				White Yajurveda	
	Śākala	Bāṣkala, …	Kapiṣṭhala-Kaṭha	Kaṭha	Maitrāyaṇīya	Taittirīya	Vājasaneyin	
Saṃhitā	Ṛgveda-Saṃhitā (RV) Śākala Recension	Ṛgveda-Saṃhitā Bāṣkala recension	Kapiṣṭhala-Kaṭha-Saṃhitā	Kāṭhaka-Saṃhitā (KS)	Maitrāyaṇī Saṃhitā (MS)	Taittirīya-Saṃhitā (TS)	Vājasaneyin-Saṃhitā (VS) Mādhyandina recension	Vājasaneyin-Saṃhitā Kāṇva recension
Brāhmaṇa	Aitareya-Brāhmaṇa	Kauṣītaki- / Śāṅkhāyana -Brāhmaṇa				Taittirīya-Saṃhitā Taittirīya-Brāhmaṇa	Śatapatha-Brāhmaṇa (ŚB) Mādhyandina recension	Śatapatha-Brāhmaṇa Kāṇva recension
Āraṇyaka	Aitareya-Āraṇyaka	Śāṅkhāyana -Āraṇyaka		Kaṭha-Āraṇyaka		Taittirīya-Āraṇyaka		Bṛhad-Āraṇyaka-Upaniṣad Kāṇva recension
Upaniṣad	Aitareya-Up.	Kauṣītaki-Up.		Kaṭha-Up.	Maitrāyaṇīya-Up.	Taittirīya-Up.	Śatapatha-Brāhmaṇa Mādhyandina recension / Īśa-Up.	

original by Tiziana Pontillo, modified by Kyoko Amano

10

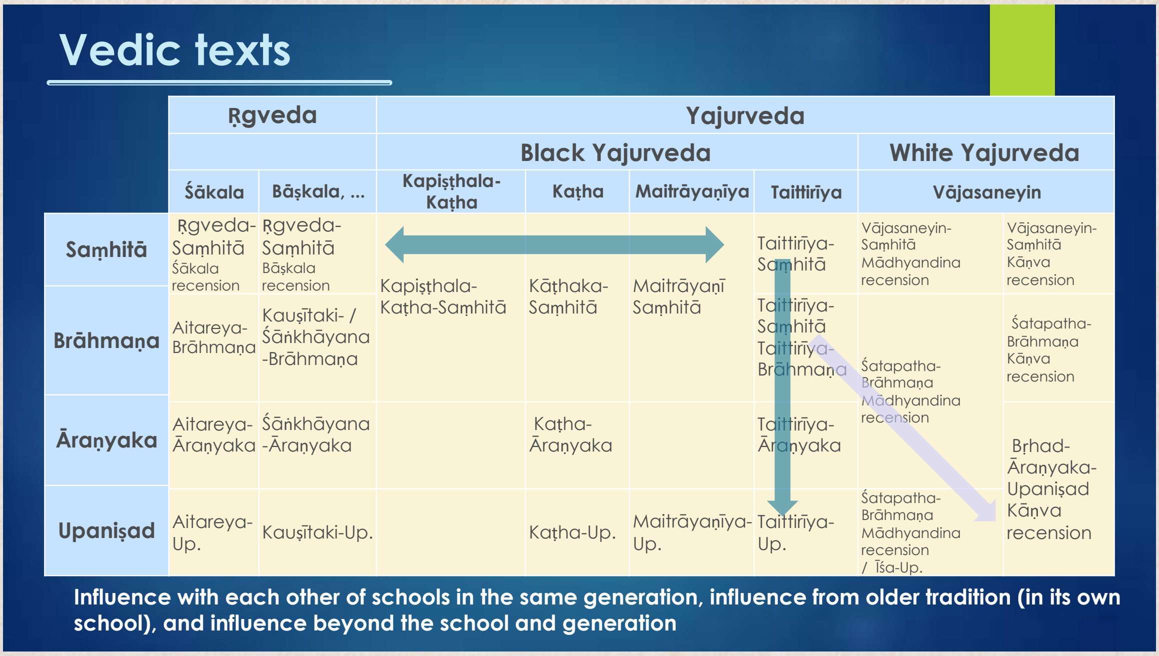


Vedic texts

	Sāmaveda			Atharvaveda	
	Rāṇāyanīya	Kauthuma	Jaiminīya	Śaunaka	Paippalāda
Saṃhitā	Sāmaveda-Saṃhitā (SV) Rāṇāyanīya recension	Sāmaveda-Saṃhitā kauthuma recension	Sāmaveda-Saṃhitā Jaiminīya recension	Atharvaveda-Saṃhitā (AV) Śaunaka recension	Atharvaveda-Saṃhitā Paippalāda recension
Brāhmaṇa	Pañcaviṃśa-Brāhmaṇa (PB) = Tāṇḍyamahā-Brāhmaṇa		Jaiminīya-Brāhmaṇa (JB) = Talavakāra-Brāhmaṇa	Gopatha-Brāhmaṇa (GB)	
Āraṇyaka					
Upaniṣad	Chāndogya- / Jaiminīya- / Kena - Upaniṣad			Muṇḍaka / Praśna / Māṇḍukya-Upaniṣad	

original by Tiziana Pontillo, modified by Kyoko Amano

11



Vedic texts

	Ṛgveda		Yajurveda					
			Black Yajurveda				White Yajurveda	
	Śākala	Bāṣkala, …	Kapiṣṭhala-Kaṭha	Kaṭha	Maitrāyaṇīya	Taittirīya	Vājasaneyin	
Saṃhitā	Ṛgveda-Saṃhitā Śākala recension	Ṛgveda-Saṃhitā Bāṣkala recension	Kapiṣṭhala-Kaṭha-Saṃhitā	Kāṭhaka-Saṃhitā	Maitrāyaṇī Saṃhitā	Taittirīya-Saṃhitā	Vājasaneyin-Saṃhitā Mādhyandina recension	Vājasaneyin-Saṃhitā Kāṇva recension
Brāhmaṇa	Aitareya-Brāhmaṇa	Kauṣītaki- / Śāṅkhāyana -Brāhmaṇa				Taittirīya-Saṃhitā Taittirīya-Brāhmaṇa	Śatapatha-Brāhmaṇa Mādhyandina recension	Śatapatha-Brāhmaṇa Kāṇva recension
Āraṇyaka	Aitareya-Āraṇyaka	Śāṅkhāyana -Āraṇyaka		Kaṭha-Āraṇyaka		Taittirīya-Āraṇyaka		Bṛhad-Āraṇyaka-Upaniṣad Kāṇva recension
Upaniṣad	Aitareya-Up.	Kauṣītaki-Up.		Kaṭha-Up.	Maitrāyaṇīya-Up.	Taittirīya-Up.	Śatapatha-Brāhmaṇa Mādhyandina recension / Īśa-Up.	

Influence with each other of schools in the same generation, influence from older tradition (in its own school), and influence beyond the school and generation

12

Vedic texts

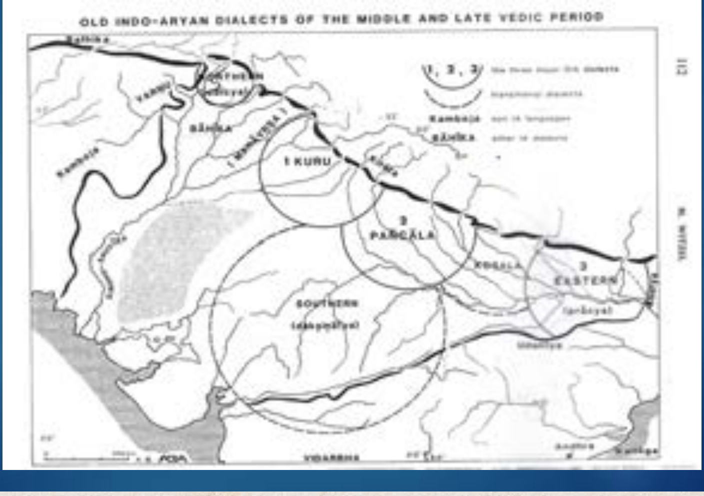
	Ṛgveda		Yajurveda					
			Black Yajurveda				White Yajurveda	
	Śākala	Bāṣkala, ...	Kapiṣṭhala-Kaṭha	Kaṭha	Maitrāyaṇīya	Taittirīya	Vājasaneyin	
Saṃhitā	Ṛgveda-Saṃhitā Śākala recension	Ṛgveda-Saṃhitā Bāṣkala recension	Kapiṣṭhala-Kaṭha-Saṃhitā	Kāṭhaka-Saṃhitā	Maitrāyaṇī Saṃhitā	Taittirīya-Saṃhitā	Vājasaneyin-Saṃhitā Mādhyandina recension	Vājasaneyin-Saṃhitā Kāṇva recension
Brāhmaṇa	Aitareya-Brāhmaṇa	Kauṣītaki- / Śāṅkhāyana-Brāhmaṇa				Taittirīya-Saṃhitā Taittirīya-Brāhmaṇa	Śatapatha-Brāhmaṇa Mādhyandina recension	Śatapatha-Brāhmaṇa Kāṇva recension
Āraṇyaka	Aitareya-Āraṇyaka	Śāṅkhāyana-Āraṇyaka		Kaṭha-Āraṇyaka		Taittirīya-Āraṇyaka		Bṛha-d Āraṇyaka-Upaniṣad Kāṇva recension
Upaniṣad	Aitareya-Up.	Kauṣītaki-Up.		Kaṭha-Up.	Maitrāyaṇīya-Up.	Taittirīya-Up.	Śatapatha-Brāhmaṇa Mādhyandina recension / Īśa-Up.	

The older the layers in these texts, the stronger the reflections on the families' (schools') geographical and social conditions.
The more recent the era, the greater the development of networks between social groups (or Vedic schools). Language, culture, ideas, and methods of rituals were shared, and the different schools would generally be standardized as *Brahmanism*.

13

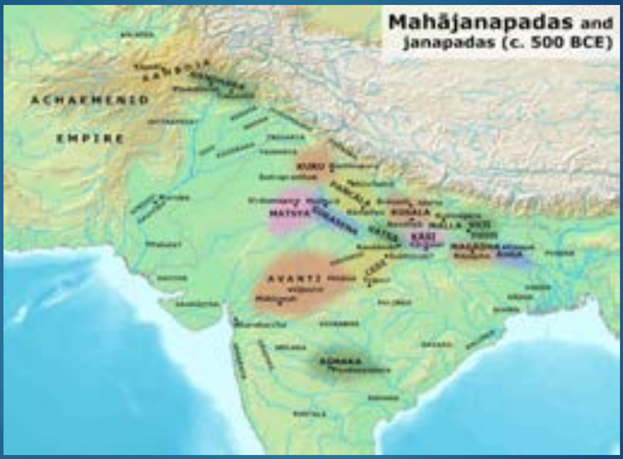
Localization and Dating of Vedic Texts

Witzel, Michael, Tracing the Vedic Dialects (1989)



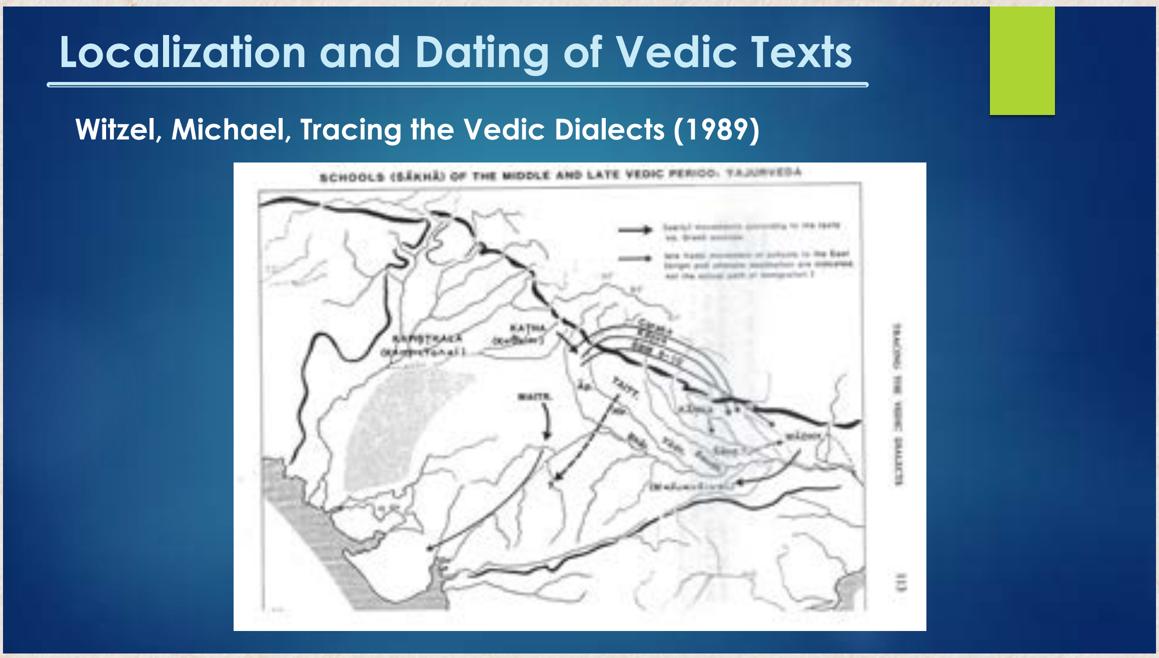
14

Localization and Dating of Vedic Texts



https://upload.wikimedia.org/wikipedia/commons/2/28/ Mahajanapadas_%28c._500_BCE%29.png

15

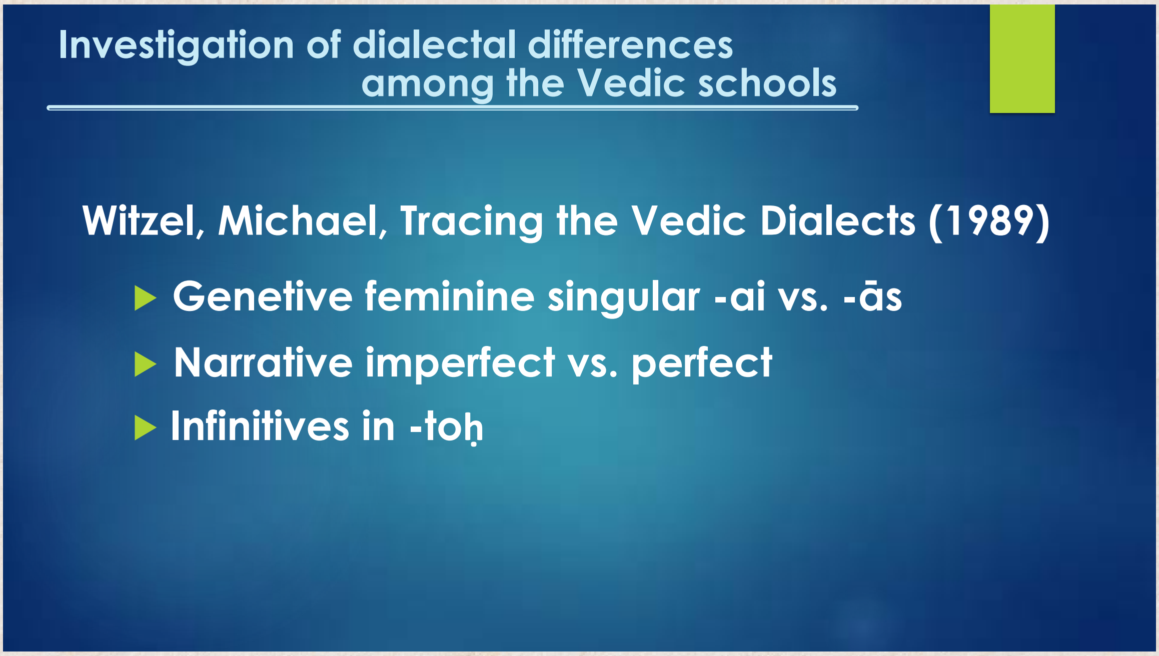


Localization and Dating of Vedic Texts

Witzel, Michael, Tracing the Vedic Dialects (1989)

SCHOOLS (ŚĀKHĀ) OF THE MIDDLE AND LATE VEDIC PERIOD: YAJURVEDA
KAPIṢṬHALA
KAṬHA
MAITR.
TAITT.
TRACING THE VEDIC DIALECTS

16

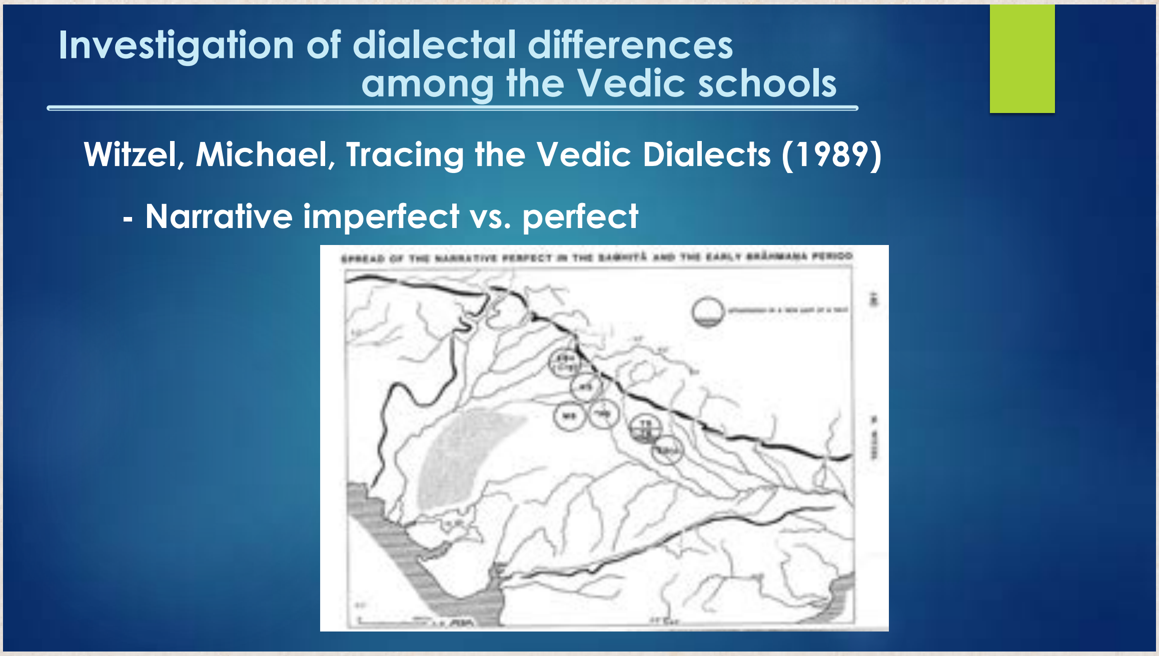


Investigation of dialectal differences among the Vedic schools

Witzel, Michael, Tracing the Vedic Dialects (1989)

- Genetive feminine singular -ai vs. -ās
- Narrative imperfect vs. perfect
- Infinitives in -toḥ

17



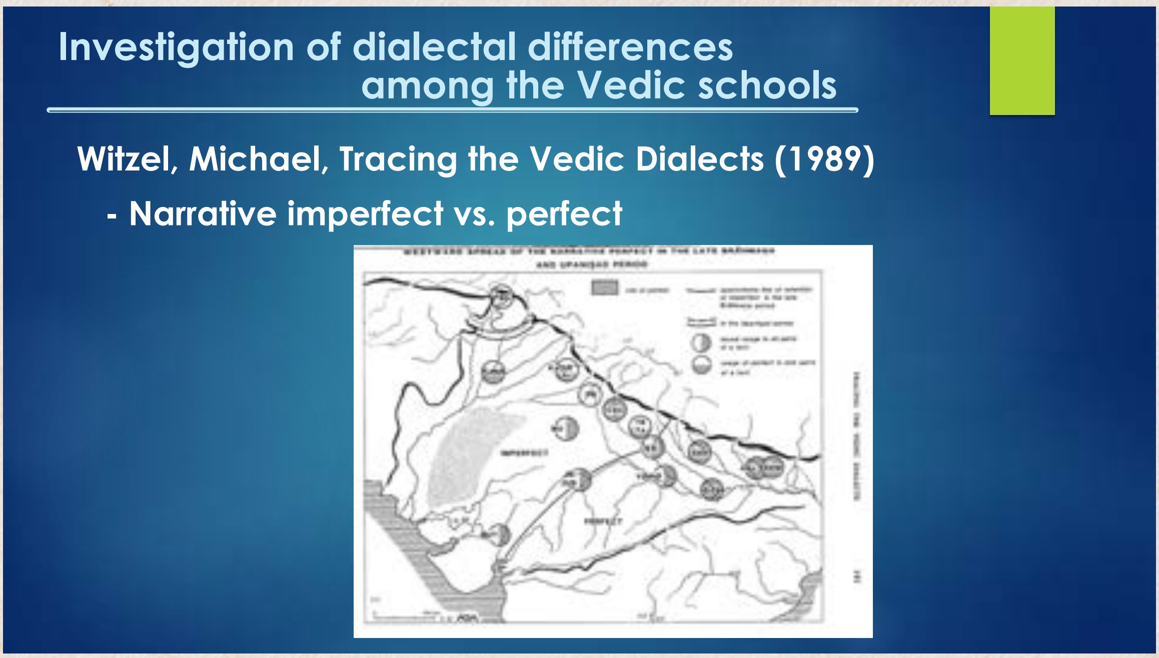
Investigation of dialectal differences among the Vedic schools

Witzel, Michael, Tracing the Vedic Dialects (1989)

- Narrative imperfect vs. perfect

SPREAD OF THE NARRATIVE PERFECT IN THE SAṂHITĀ AND THE EARLY BRĀHMAṆA PERIOD
MS

18



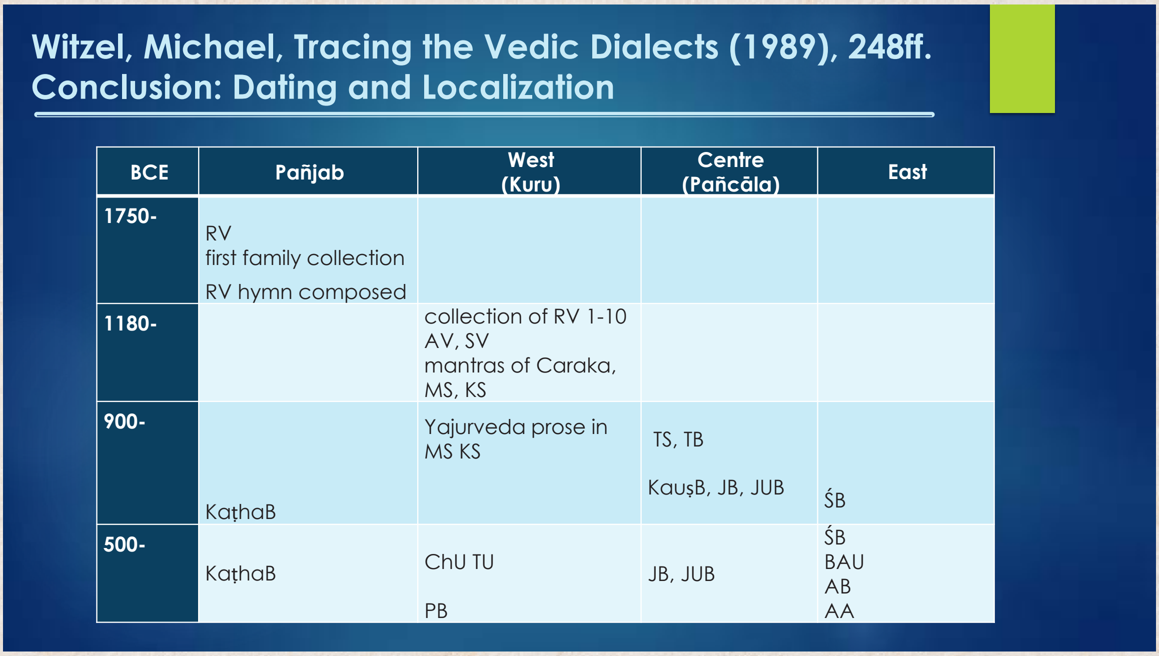
Investigation of dialectal differences among the Vedic schools

Witzel, Michael, Tracing the Vedic Dialects (1989)

- Narrative imperfect vs. perfect

WESTWARD SPREAD OF THE NARRATIVE PERFECT IN THE LATE BRĀHMAṆA AND UPANIṢAD PERIOD
IMPERFECT
PERFECT

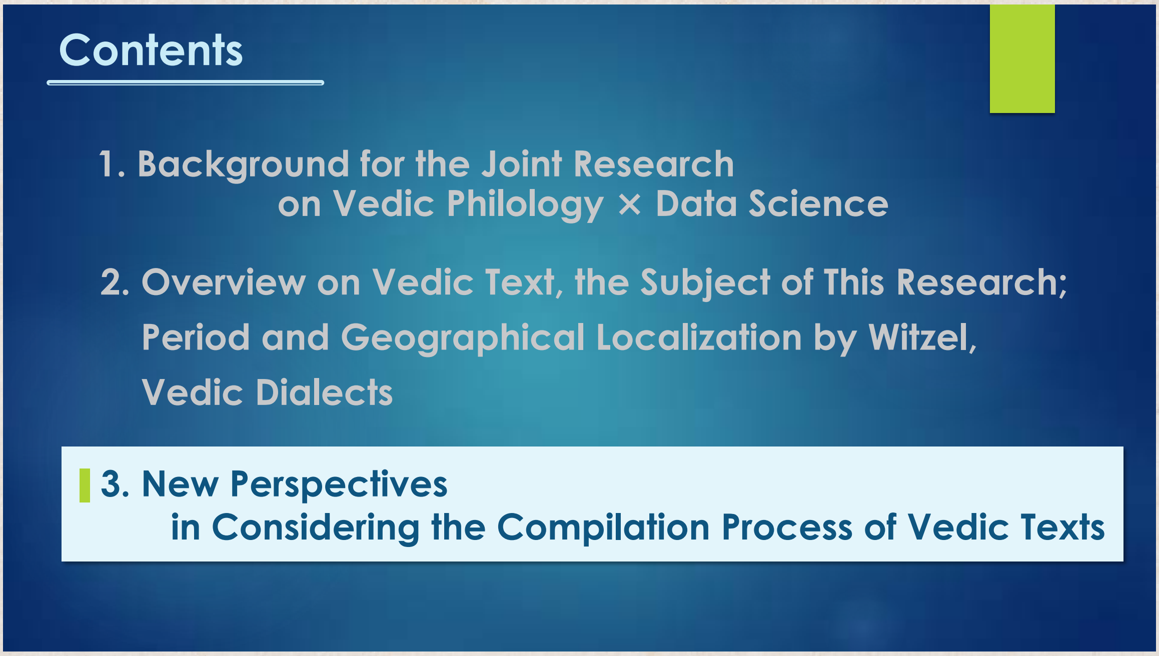
19



Witzel, Michael, Tracing the Vedic Dialects (1989), 248ff. Conclusion: Dating and Localization

BCE	Pañjab	West (Kuru)	Centre (Pañcāla)	East
1750-	RV first family collection RV hymn composed			
1180-		collection of RV 1-10 AV, SV mantras of Caraka, MS, KS		
900-	KaṭhaB	Yajurveda prose in MS KS	TS, TB KauṣB, JB, JUB	ŚB
500-	KaṭhaB	ChU TU PB	JB, JUB	ŚB BAU AB AA

20



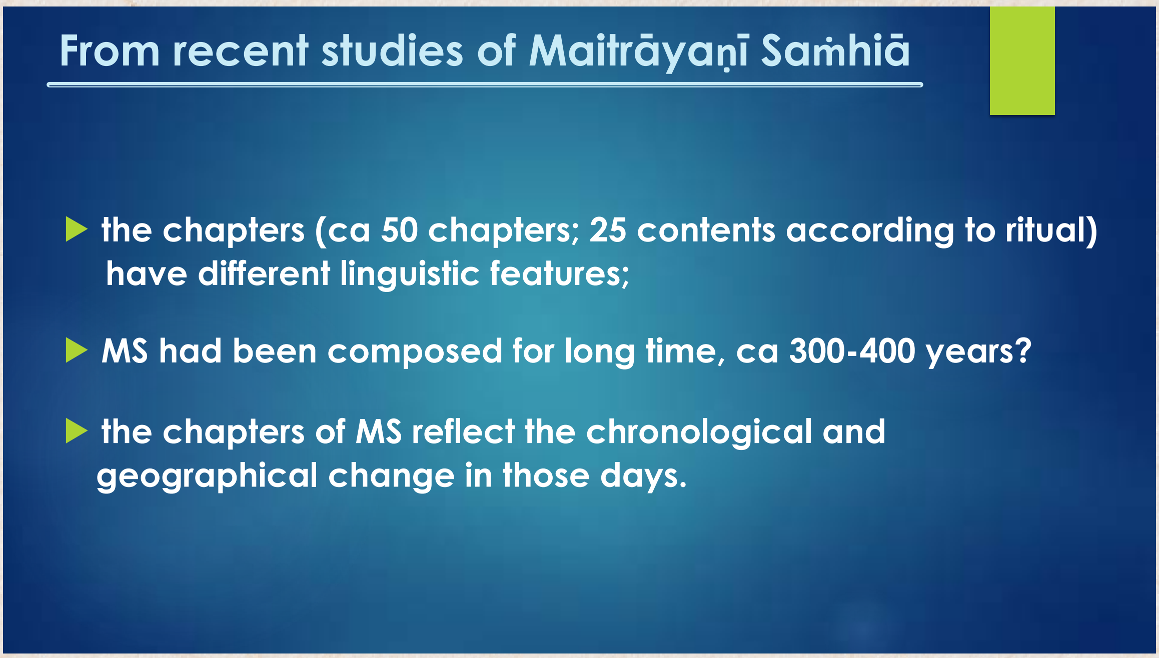
Contents

1. Background for the Joint Research on Vedic Philology × Data Science

2. Overview on Vedic Text, the Subject of This Research; Period and Geographical Localization by Witzel, Vedic Dialects

3. New Perspectives in Considering the Compilation Process of Vedic Texts

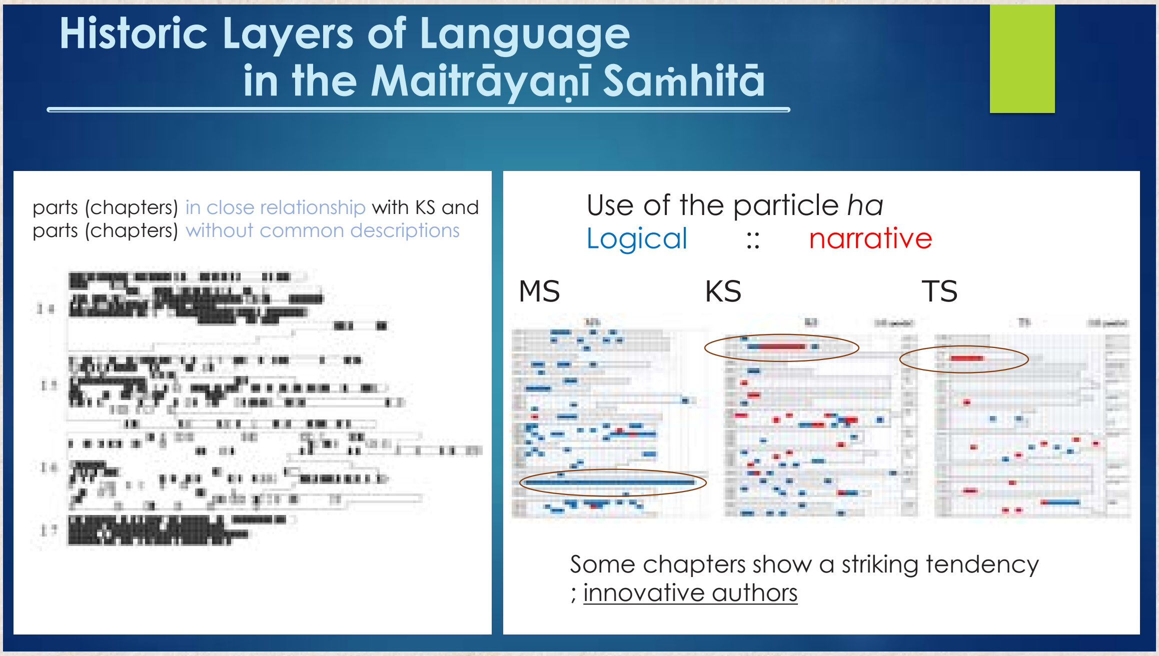
21



From recent studies of Maitrāyaṇī Saṁhiā

- the chapters (ca 50 chapters; 25 contents according to ritual) have different linguistic features;
- MS had been composed for long time, ca 300-400 years?
- the chapters of MS reflect the chronological and geographical change in those days.

22



Historic Layers of Language in the Maitrāyaṇī Saṁhitā

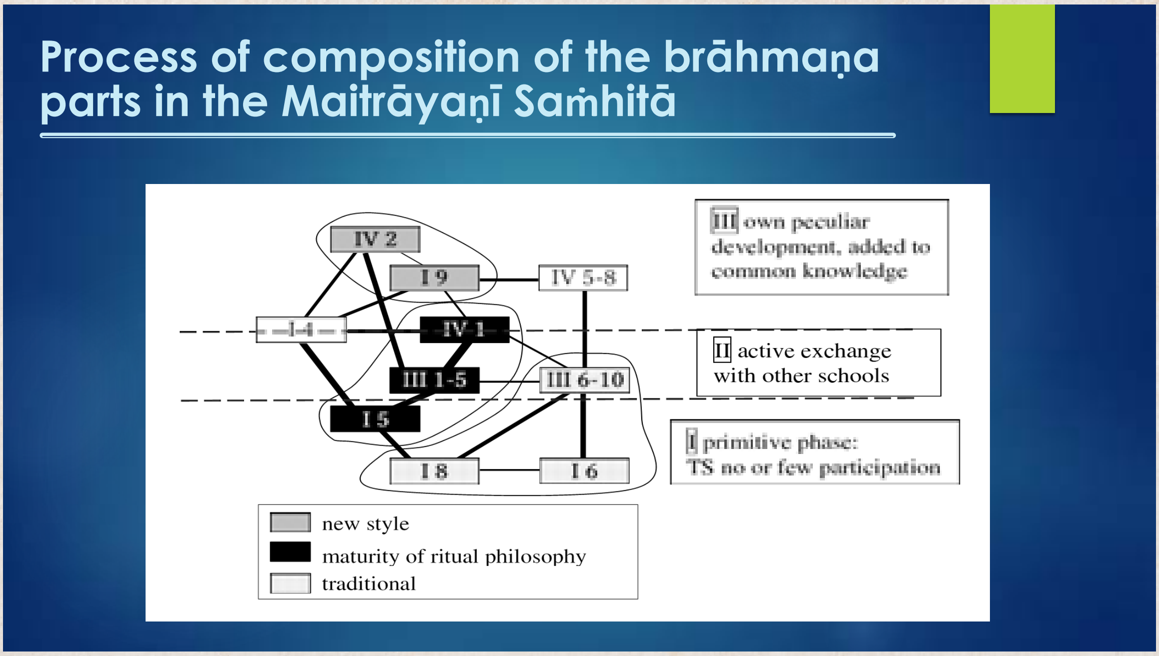
parts (chapters) in close relationship with KS and
parts (chapters) without common descriptions

Use of the particle *ha*
Logical :: narrative

MS KS TS

Some chapters show a striking tendency
; innovative authors

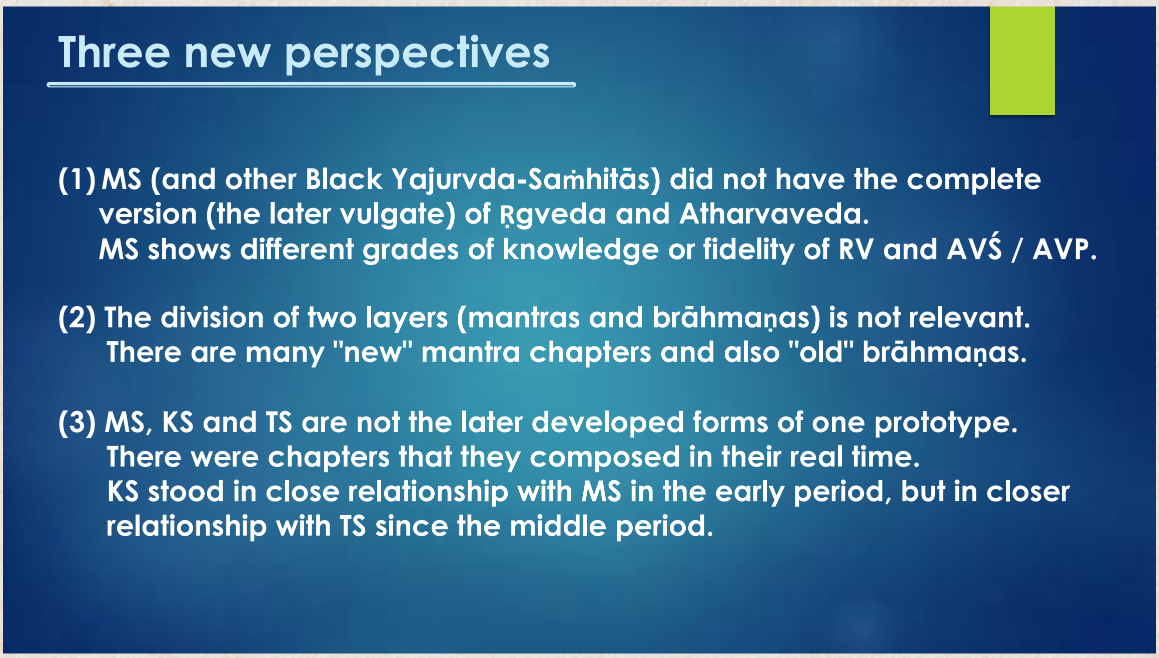
23



Process of composition of the brāhmaṇa parts in the Maitrāyaṇī Saṁhitā

IV 2
I 9
IV 5-8
I 4
IV 1
III 1-5
III 6-10
I 5
I 8
I 6
III own peculiar development, added to common knowledge
II active exchange with other schools
I primitive phase: TS no or few participation
new style
maturity of ritual philosophy
traditional

24



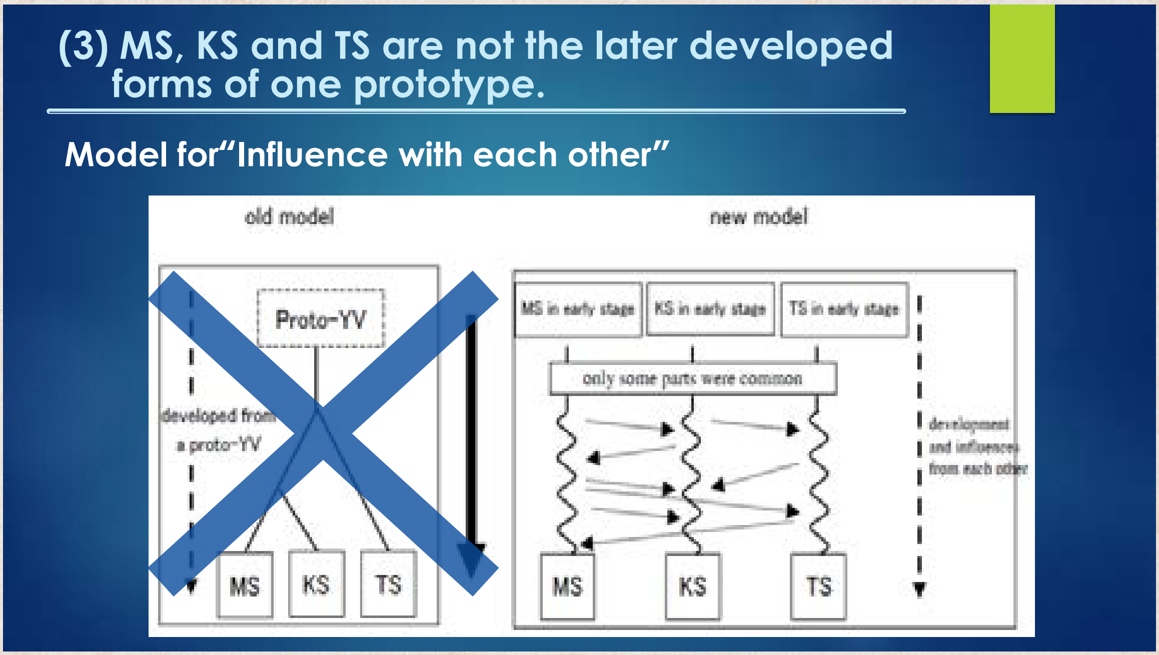
Three new perspectives

(1) MS (and other Black Yajurvda-Saṁhitās) did not have the complete version (the later vulgate) of Ṛgveda and Atharvaveda.
MS shows different grades of knowledge or fidelity of RV and AVŚ / AVP.

(2) The division of two layers (mantras and brāhmaṇas) is not relevant.
There are many "new" mantra chapters and also "old" brāhmaṇas.

(3) MS, KS and TS are not the later developed forms of one prototype.
There were chapters that they composed in their real time.
KS stood in close relationship with MS in the early period, but in closer relationship with TS since the middle period.

25

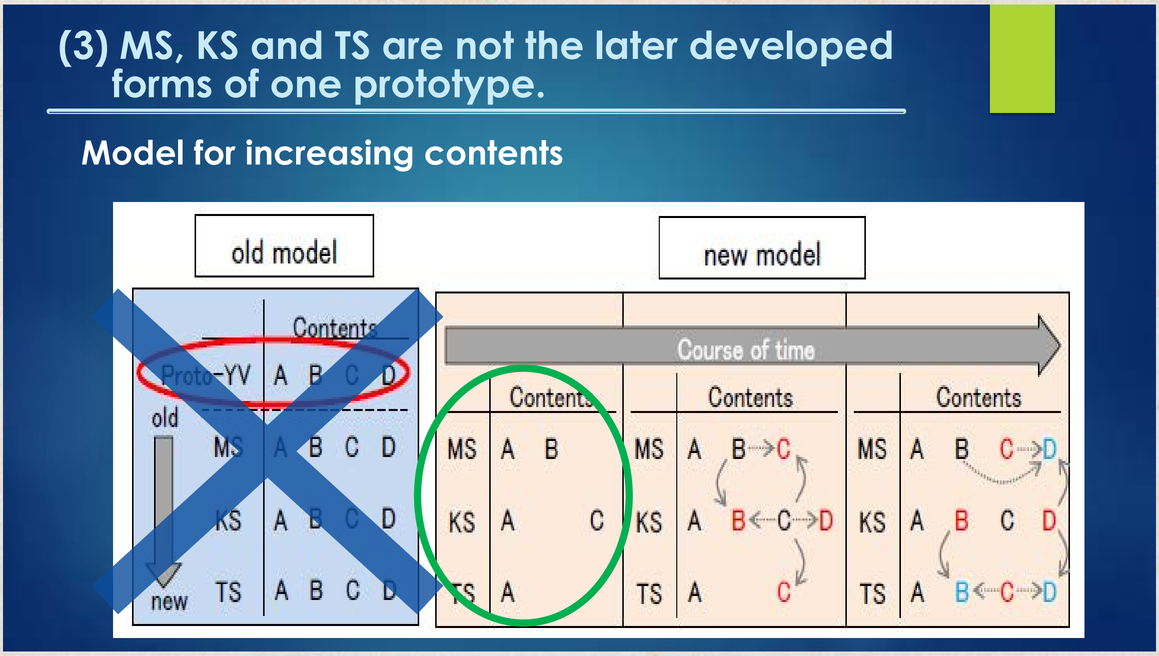


(3) MS, KS and TS are not the later developed forms of one prototype.

Model for"Influence with each other"

old model
Proto-YV
developed from a proto-YV
MS
KS
TS
new model
MS in early stage
KS in early stage
TS in early stage
only some parts were common
development and influences from each other
MS
KS
TS

26



(3) MS, KS and TS are not the later developed forms of one prototype.

Model for increasing contents

old model

	Contents			
Proto-YV	A	B	C	D
MS	A	B	C	D
KS	A	B	C	D
TS	A	B	C	D

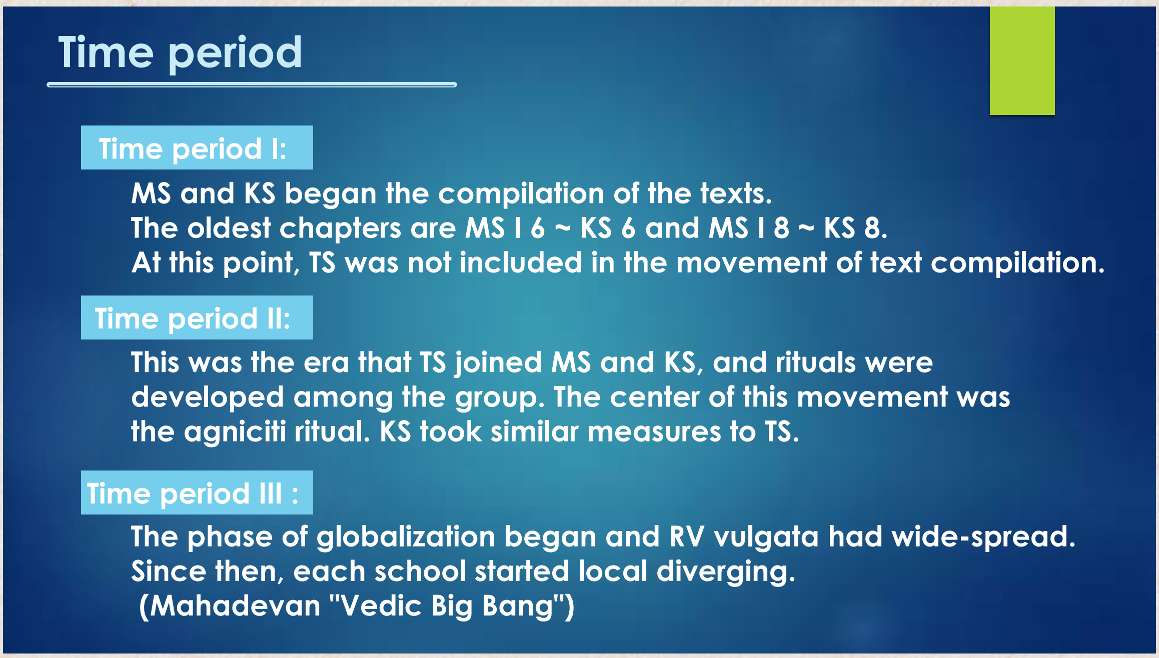
new model

	Contents		
MS	A	B	
KS	A		C
TS	A		

	Contents			
MS	A	B→C		
KS	A	B←C→D		
TS	A	C		

	Contents			
MS	A	B	C→D	
KS	A	B	C	D
TS	A	B←C→D		

27



Time period

Time period I:

MS and KS began the compilation of the texts.
The oldest chapters are MS I 6 ~ KS 6 and MS I 8 ~ KS 8.
At this point, TS was not included in the movement of text compilation.

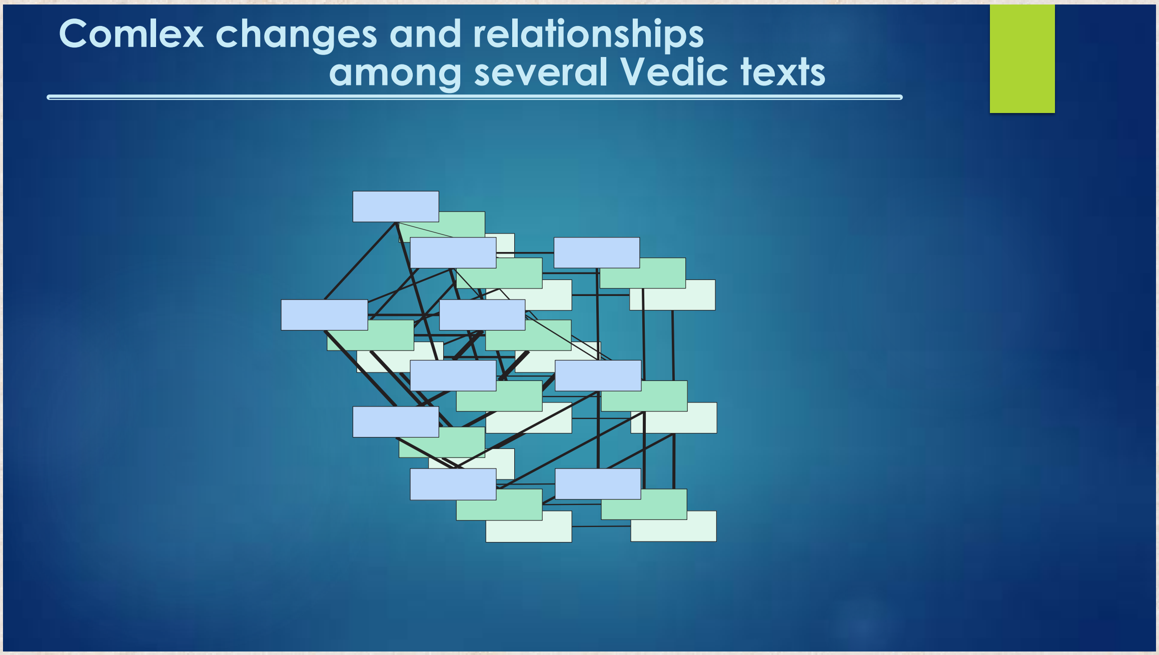
Time period II:

This was the era that TS joined MS and KS, and rituals were developed among the group. The center of this movement was the agniciti ritual. KS took similar measures to TS.

Time period III :

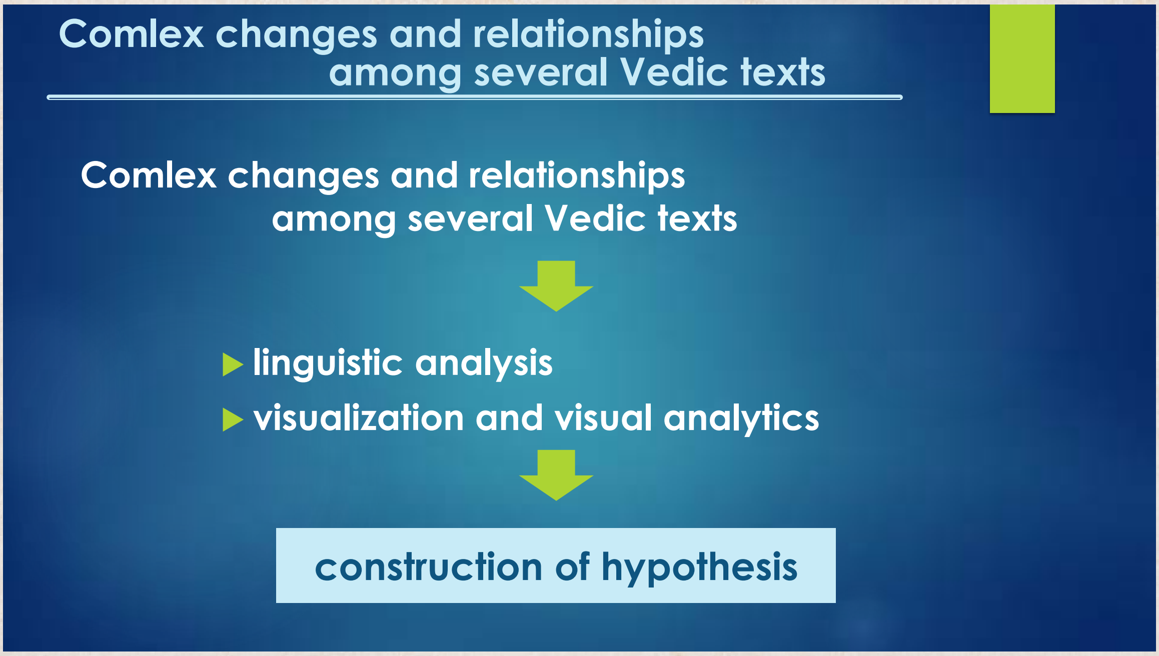
The phase of globalization began and RV vulgata had wide-spread.
Since then, each school started local diverging.
(Mahadevan "Vedic Big Bang")

28



Comlex changes and relationships among several Vedic texts

29



Comlex changes and relationships among several Vedic texts

Comlex changes and relationships
among several Vedic texts

- linguistic analysis
- visualization and visual analytics

construction of hypothesis

Presentation 1 ｜報告 1

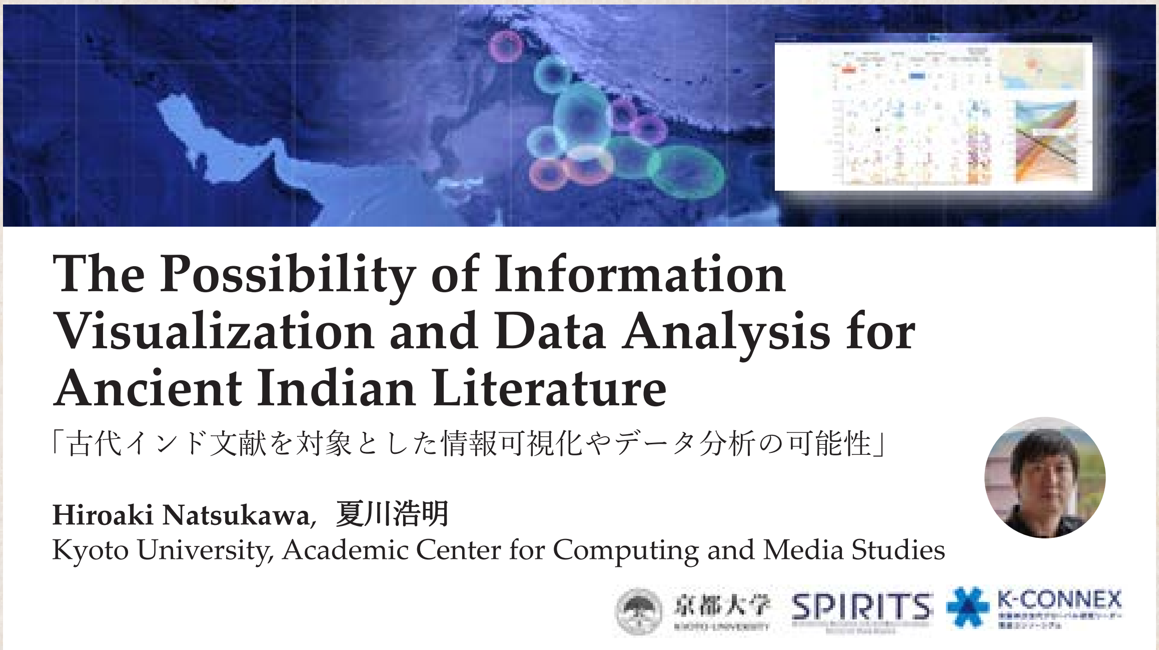
The Possibility of Information Visualization and Data Analysis for Ancient Indian Literature

古代インド文献を対象とした情報可視化やデータ分析の可能性

夏川浩明（京都大学 学術情報メディアセンター）

Hiroaki Natsukawa (Kyoto University, Academic Center for Computing and Media Studies)

1

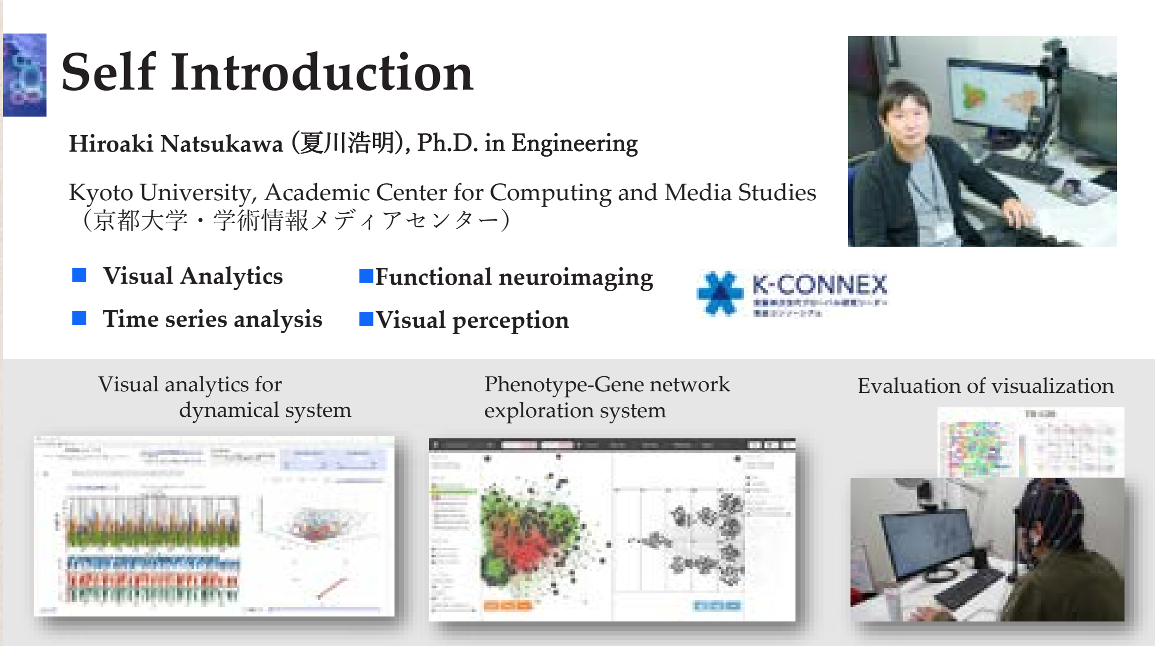


The Possibility of Information Visualization and Data Analysis for Ancient Indian Literature

「古代インド文献を対象とした情報可視化やデータ分析の可能性」

Hiroaki Natsukawa, 夏川浩明
Kyoto University, Academic Center for Computing and Media Studies

2



Self Introduction

Hiroaki Natsukawa (夏川浩明), Ph.D. in Engineering

Kyoto University, Academic Center for Computing and Media Studies
（京都大学・学術情報メディアセンター）

- Visual Analytics
- Time series analysis
- Functional neuroimaging
- Visual perception

Visual analytics for dynamical system

Phenotype-Gene network exploration system

Evaluation of visualization

3

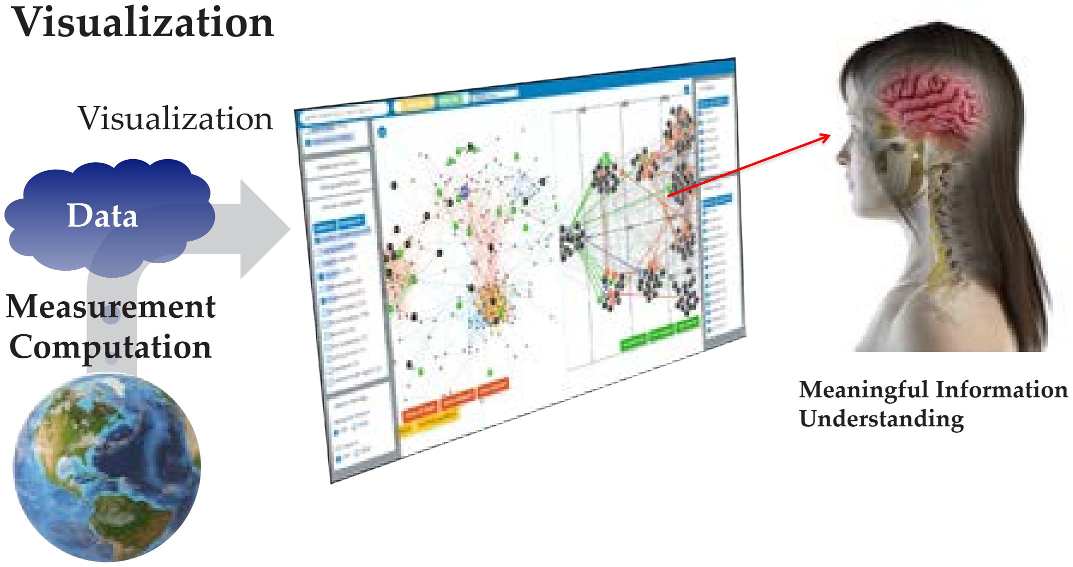
The Possibility of Information Visualization and Data Analysis for Ancient Indian Literature

「古代インド文献を対象とした情報可視化やデータ分析の可能性」

- Visualization & Visual Analytics
- Exploranation
- Ancient Indian Literature
- Visualization Tool for Ancient Indian Literature
- Analysis of Hypergraph

4

Visualization



Visualization

Data

Measurement
Computation

Meaningful Information
Understanding

5

Visualization & Visual Analytics

```
1 0 2 2 7 6 3 3 0 8 8 0 2 1 8 8 1 2 1 7 5 2 9 3 5 8 3 2 5
6 8 5 1 6 3 4 3 3 1 6 0 5 0 2 3 6 8 3 0 1 0 5 3 3 3 8 4 4
5 3 7 8 1 6 3 7 7 5 0 2 6 4 3 1 0 3 2 8 0 6 4 4 1 8 2 3 0
4 7 7 9 7 6 5 6 5 5 7 8 8 5 4 5 5 8 3 6 7 2 0 4 2 5 7 3 7
6 8 4 8 6 4 3 8 8 9 3 3 4 6 1 0 4 1 5 5 1 3 3 5 1 4 2 0 5
2 1 7 0 5 6 0 6 4 8 7 5 8 2 0 9 2 0 7 8 1 0 0 2 7 7 1 1 2
0 4 6 5 0 1 1 2 1 6 1 5 1 4 6 2 6 5 2 6 7 7 8 2 1 2 5 8 2
1 8 8 5 0 7 1 5 0 8 7 8 1 5 4 1 7 3 7 8 2 4 1 9 5 1 7 1 0
1 2 7 3 6 4 0 4 6 2 4 2 3 8 8 8 8 2 0 2 9 7 5 7 8 6 6 4 7
7 6 3 3 6 3 5 0 0 4 0 5 2 5 4 2 9 1 4 7 4 4 2 7 1 2 0 4 4
5 2 3 4 4 6 3 6 4 2 7 2 6 1 3 2 8 3 2 3 8 3 5 3 5 3 0 5 5
4 8 2 3 0 4 7 5 4 5 8 3 1 8 8 7 3 5 8 4 8 2 6 6 4 6 6 3 4
6 4 5 8 7 3 4 1 6 8 5 7 4 6 6 2 0 7 7 7 1 1 0 2 2 0 7 7 4
5 2 4 3 6 7 7 6 9 4 5 3 3 5 4 2 1 4 5 4 6 7 1 3 1 5 4 8 0
0 7 8 5 3 3 2 7 1 7 4 2 0 9 3 0 0 6 7 0 7 6 4 5 6 2 2 6 8
3 2 8 4 2 0 6 4 3 7 2 8 1 3 4 7 0 6 7 0 5 9 4 5 3 1 3 4 2
2 3 0 3 4 4 8 0 0 4 1 7 4 7 6 4 4 0 8 6 5 7 1 6 1 5 0 5 0
0 2 8 7 4 8 3 6 5 1 1 3 2 5 7 7 0 2 8 7 7 8 3 8 2 7 5 7 8
2 7 0 3 0 2 0 3 3 6 7 5 2 3 6 5 2 0 6 7 2 3 1 2 0 6 4 0 7
0 2 0 7 2 3 2 5 0 2 7 7 3 8 3 0 4 1 2 8 6 7 8 1 0 3 0 3 3
```

How many 9 can you find in this dataset?

6

Visualization & Visual Analytics

```
1 0 2 2 7 6 3 3 0 8 8 0 2 1 8 8 1 2 1 7 5 2 9 3 5 8 3 2 5
6 8 5 1 6 3 4 3 3 1 6 0 5 0 2 3 6 8 3 0 1 0 5 3 3 3 8 4 4
5 3 7 8 1 6 3 7 7 5 0 2 6 4 3 1 0 3 2 8 0 6 4 4 1 8 2 3 0
4 7 7 9 7 6 5 6 5 5 7 8 8 5 4 5 5 8 3 6 7 2 0 4 2 5 7 3 7
6 8 4 8 6 4 3 8 8 9 3 3 4 6 1 0 4 1 5 5 1 3 3 5 1 4 2 0 5
2 1 7 0 5 6 0 6 4 8 7 5 8 2 0 9 2 0 7 8 1 0 0 2 7 7 1 1 2
0 4 6 5 0 1 1 2 1 6 1 5 1 4 6 2 6 5 2 6 7 7 8 2 1 2 5 8 2
1 8 8 5 0 7 1 5 0 8 7 8 1 5 4 1 7 3 7 8 2 4 1 9 5 1 7 1 0
1 2 7 3 6 4 0 4 6 2 4 2 3 8 8 8 8 2 0 2 9 7 5 7 8 6 6 4 7
7 6 3 3 6 3 5 0 0 4 0 5 2 5 4 2 9 1 4 7 4 4 2 7 1 2 0 4 4
5 2 3 4 4 6 3 6 4 2 7 2 6 1 3 2 8 3 2 3 8 3 5 3 5 3 0 5 5
4 8 2 3 0 4 7 5 4 5 8 3 1 8 8 7 3 5 8 4 8 2 6 6 4 6 6 3 4
6 4 5 8 7 3 4 1 6 8 5 7 4 6 6 2 0 7 7 7 1 1 0 2 2 0 7 7 4
5 2 4 3 6 7 7 6 9 4 5 3 3 5 4 2 1 4 5 4 6 7 1 3 1 5 4 8 0
0 7 8 5 3 3 2 7 1 7 4 2 0 9 3 0 0 6 7 0 7 6 4 5 6 2 2 6 8
3 2 8 4 2 0 6 4 3 7 2 8 1 3 4 7 0 6 7 0 5 9 4 5 3 1 3 4 2
2 3 0 3 4 4 8 0 0 4 1 7 4 7 6 4 4 0 8 6 5 7 1 6 1 5 0 5 0
0 2 8 7 4 8 3 6 5 1 1 3 2 5 7 7 0 2 8 7 7 8 3 8 2 7 5 7 8
2 7 0 3 0 2 0 3 3 6 7 5 2 3 6 5 2 0 6 7 2 3 1 2 0 6 4 0 7
0 2 0 7 2 3 2 5 0 2 7 7 3 8 3 0 4 1 2 8 6 7 8 1 0 3 0 3 3
```

How many 9 can you find in this dataset?

7

Visualization & Visual Analytics

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

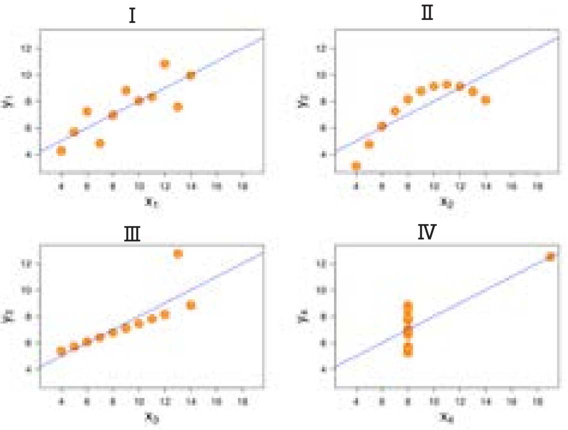
Property	Value
Mean of x in each case	9 (exact)
Variance of x in each case	11 (exact)
Mean of y in each case	7.50 (to 2 decimal places)
Variance of y in each case	4.122 or 4.127 (to 3 decimal places)
Correlation between x and y in each case	0.816 (to 3 decimal places)
Linear regression line in each case	y = 3.00 + 0.500x (to 2 and 3 decimal places, respectively)

"Anscombe's Quartet"
Source: Wikipedia

8

Visualization & Visual Analytics

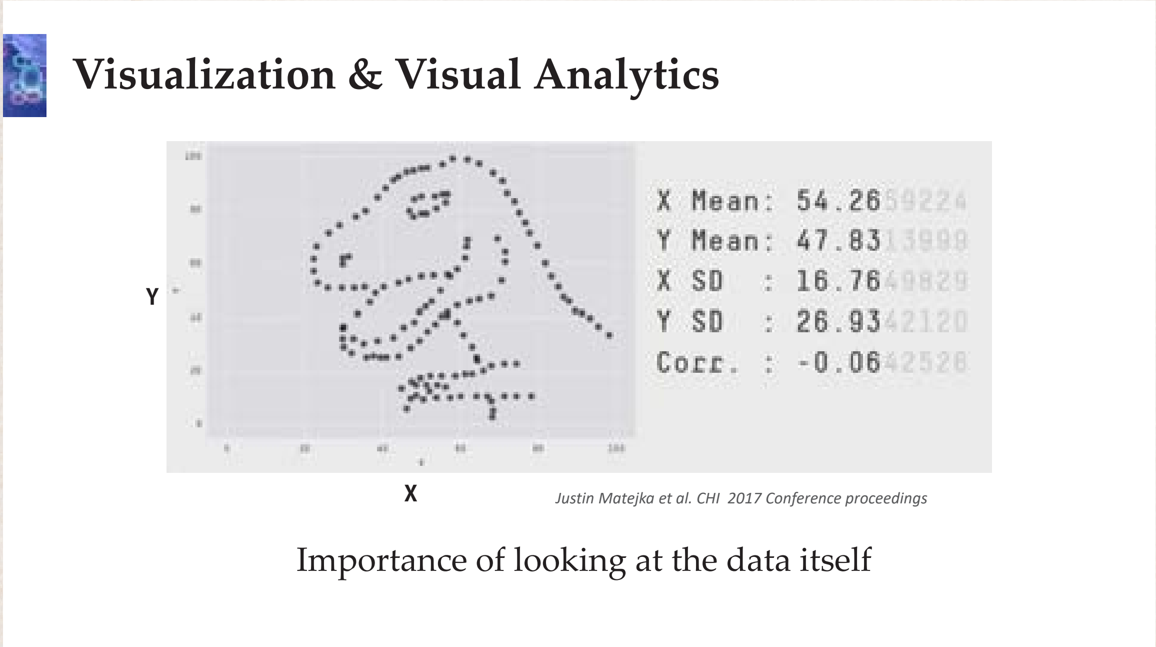
I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89



"Anscombe's Quartet"
Source: Wikipedia

Importance of looking at the data itself

9



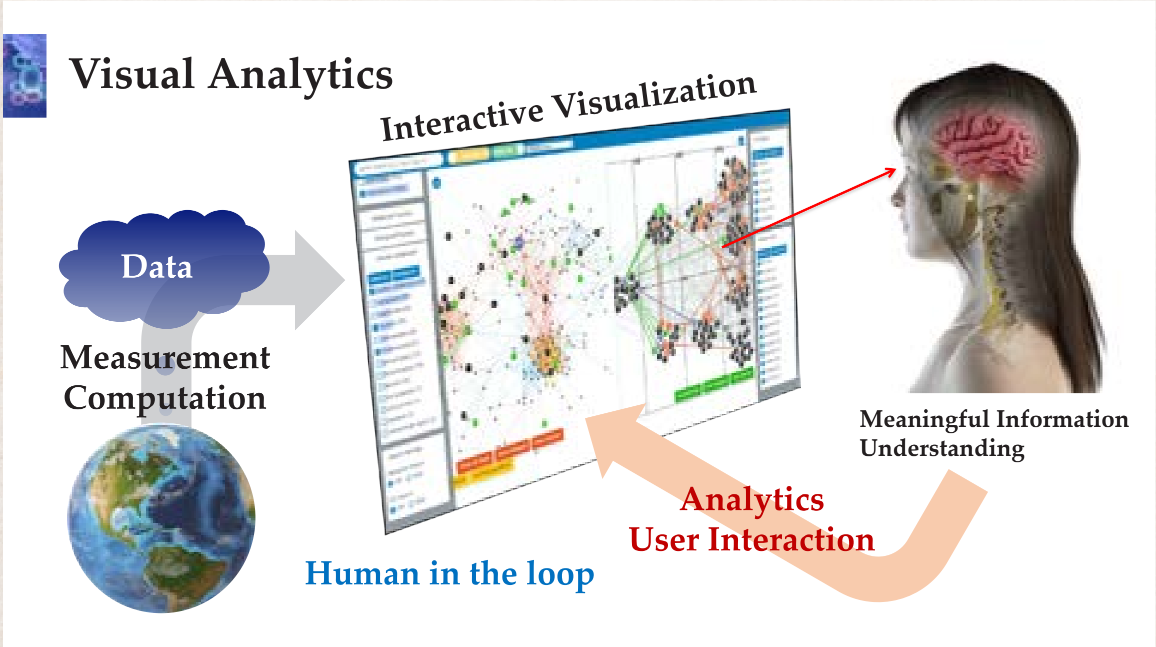
Visualization & Visual Analytics

X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

Justin Matejka et al. CHI 2017 Conference proceedings

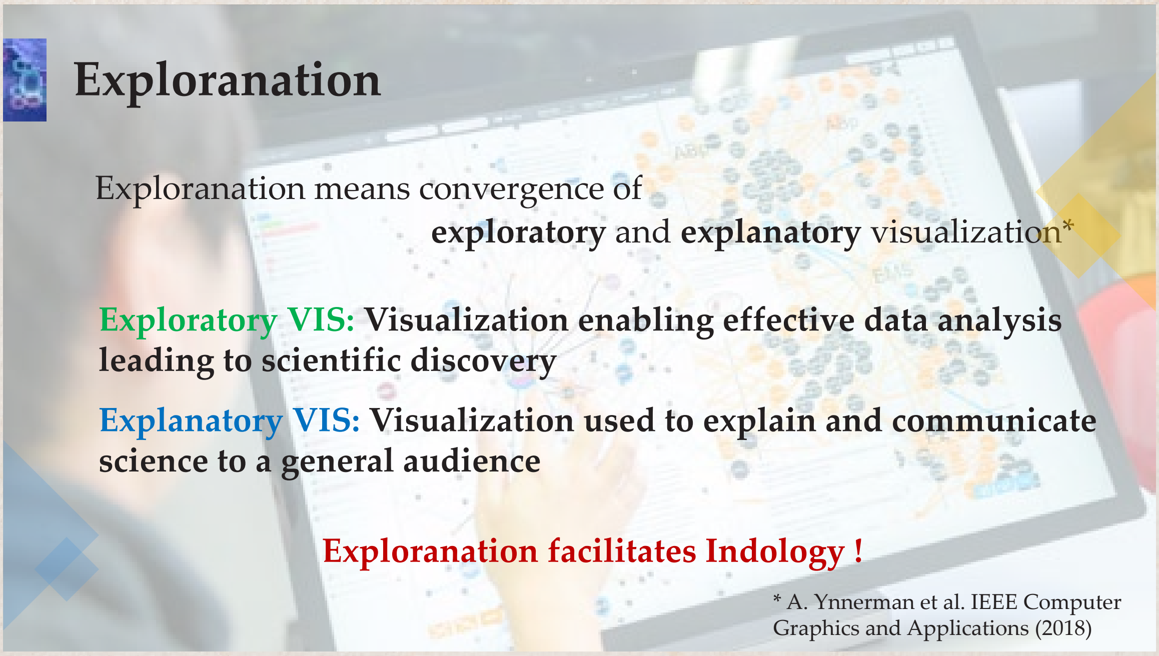
Importance of looking at the data itself

10



Visual Analytics

11



Exploranation

Exploranation means convergence of
exploratory and **explanatory** visualization*

Exploratory VIS: Visualization enabling effective data analysis leading to scientific discovery

Explanatory VIS: Visualization used to explain and communicate science to a general audience

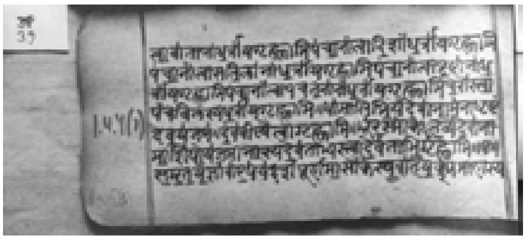
Exploranation facilitates Indology !

* A. Ynnerman et al. IEEE Computer Graphics and Applications (2018)

12

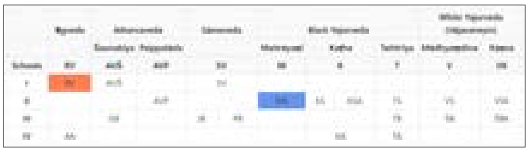
Visualization Contributing to the Analysis of Ancient Indian Literature

Examining the origins of literature through the relationship of mantras in 19 documents



Ancient Indian ritual texts BC1200-500

- Mantras
- Historical classification of literature
- Schools of literature
- Geographical characteristics of the schools



	Ṛgveda	Atharvaveda		Sāmaveda		Black Yajurveda				White Yajurveda (Vājasaneyin)	
		Śaunakīya	Paippalāda			Maitrāyaṇī	Kaṭha		Taittirīya	Mādhyandina	Kāṇva
Schools	RV	AVŚ	AVP	SV		M	K		T	V	VK
I	RV	AVŚ		SV							
II			AVP			MS	KS	KpS	TS	VS	VSK
III		GB		JB	PB				TB	ŚB	ŚBK
IV	AA						KA		TA		

13

Database

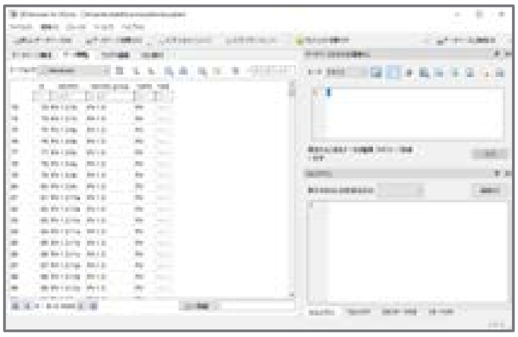
Ancient Indian ritual texts BC1200-500

- avāryāṇi pakṣmāṇi pāryā ikṣavaḥ TS.1.6.1.1 MS.1.1.2 KS.1.10 BŚ.3.16

We've tried to look at the co-occurrence of mantras in each literature

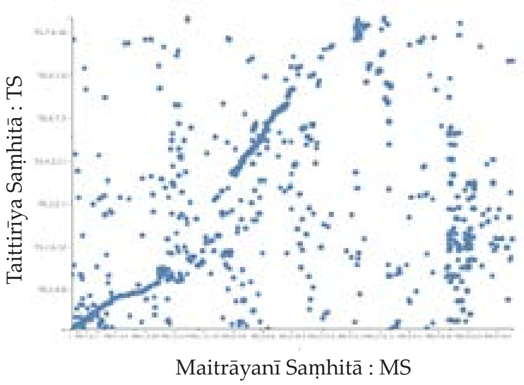
- Relational Database using SQLite
- Co-occurrence relationships between 19 literatures
- Chapter structure of literature

→Relationships among about 150 sets of documents

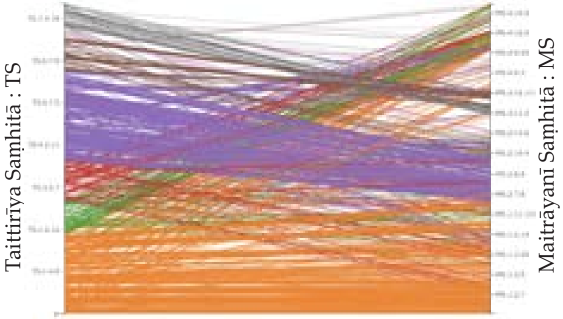


14

Visualizing the Co-occurrence of Mantras in Ancient Indian Literature



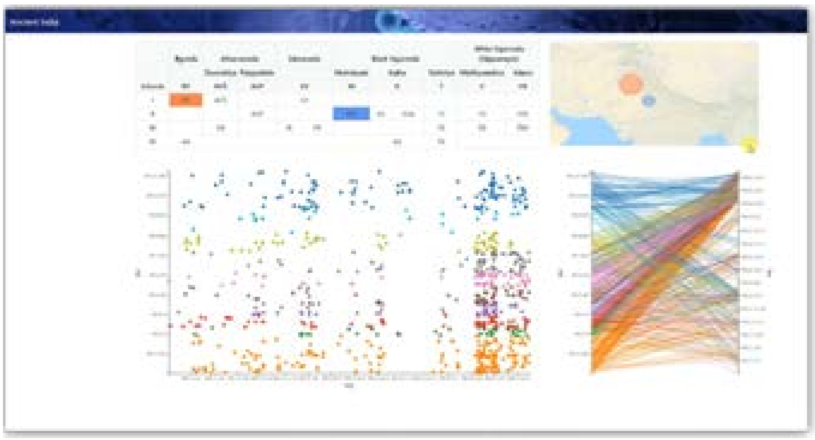
Scatter plot



Parallel coordinate plot

15

A tool for visualizing the co-occurrence of Mantras

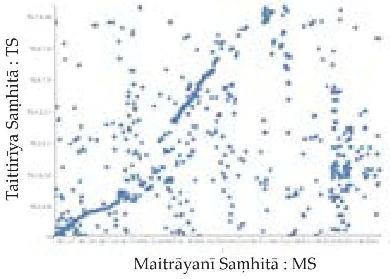


16

A tool for visualizing the co-occurrence of Mantras

Challenges

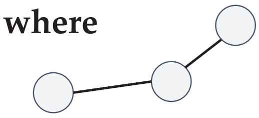
- How do we handle ritual information?
- Analyzing features across multiple references, not just one-to-one literature relationships
- Further development requirements



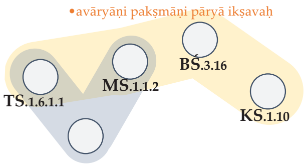
17

Hypergraph Analysis

Hypergraph is an extension of the graph concept where a link can connect several nodes.



Represented as a Hypergraph with literature as nodes and mantras as links



As a challenge for information visualization
→Scalability and Interaction issues

As a challenge in the analysis of ancient Indian literature
→To analyze in detail the relationships among multiple documents and contexts.

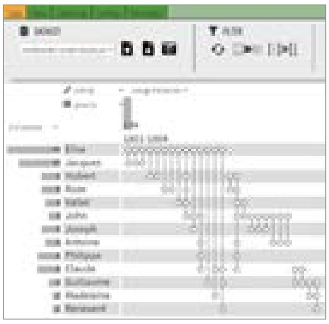
Hypergraph representation that solves these problems?

18

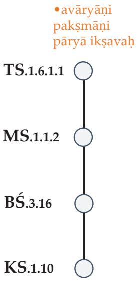
Hypergraph Analysis

French historian's analysis of the role of Marie Boucher, a female merchant in the 16th and 17th centuries*.

PAOH Vis



http://paovis.ddns.net/paoh.html



Time-varying Hypergraph representation with Scalability

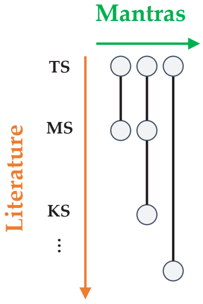
Nodes are arranged vertically, and links are represented vertically (BioFabric)

*P. Valdivia et al. IEEE TVCG (2020)

19

Hypergraph Analysis

The representation of PAOH Vis looks good, but can it be applied to ancient Indian literature data?



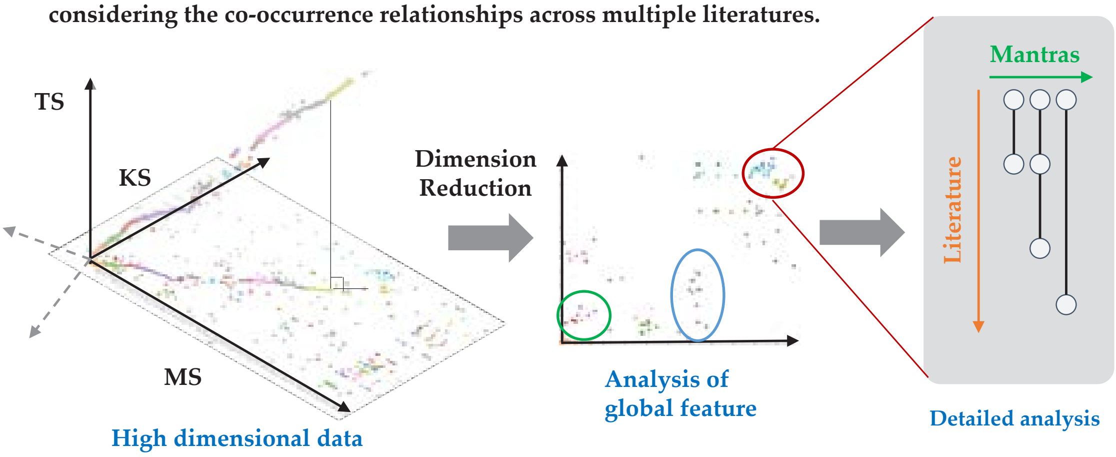
The number of literature is 19,
but how to handle the 3452 sections?
→Size and hierarchical representation
Mantra also exists for 88919 links
→Difficult to see all of them

PAOH Vis (BioFabric) representation can be used for data interpretation, but global features need to be summarized beforehand.

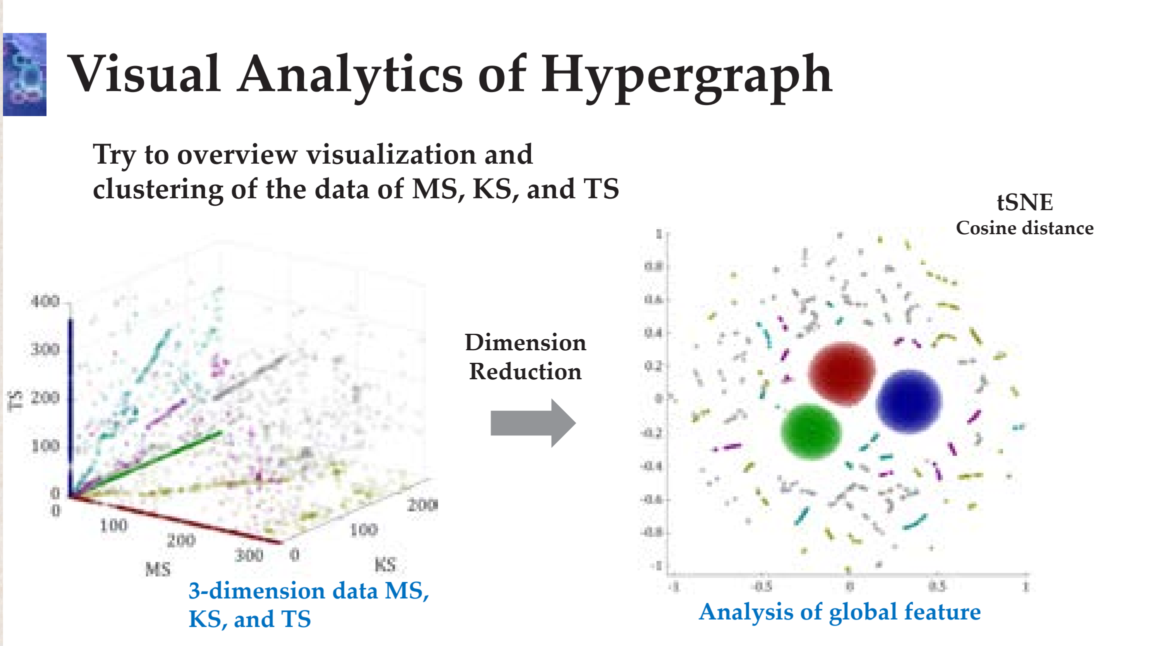
20

Visual Analytics of Hypergraph

Overview visualization and clustering of the data of 88919 links, considering the co-occurrence relationships across multiple literatures.



21



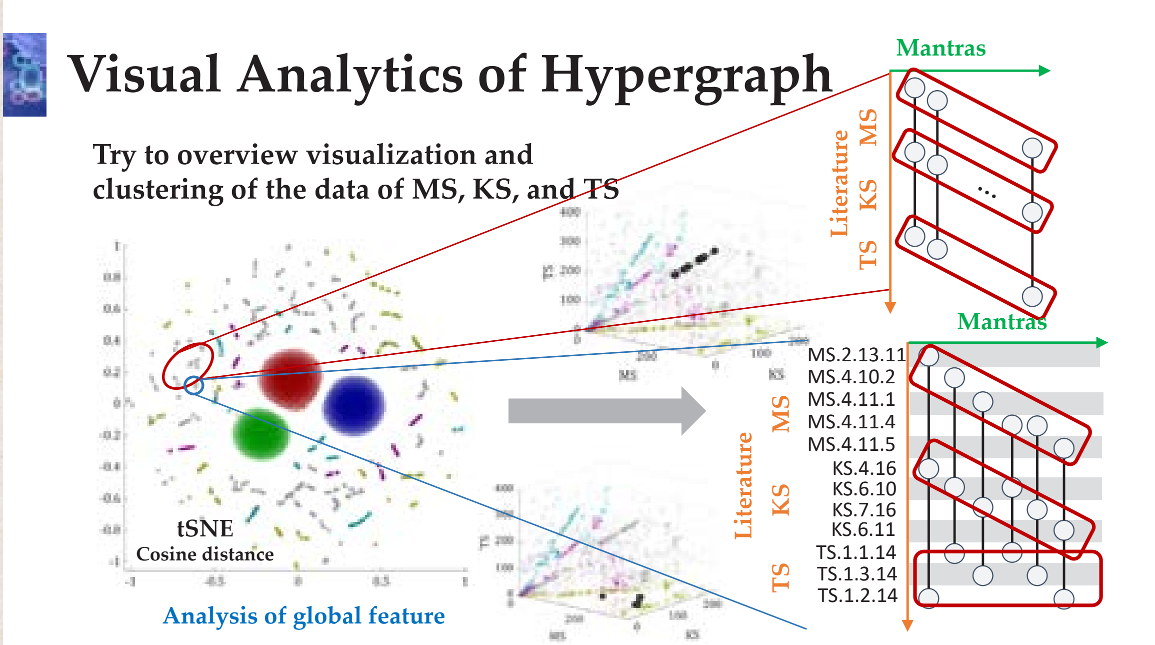
Visual Analytics of Hypergraph

Try to overview visualization and clustering of the data of MS, KS, and TS

3-dimension data MS, KS, and TS

Analysis of global feature

22

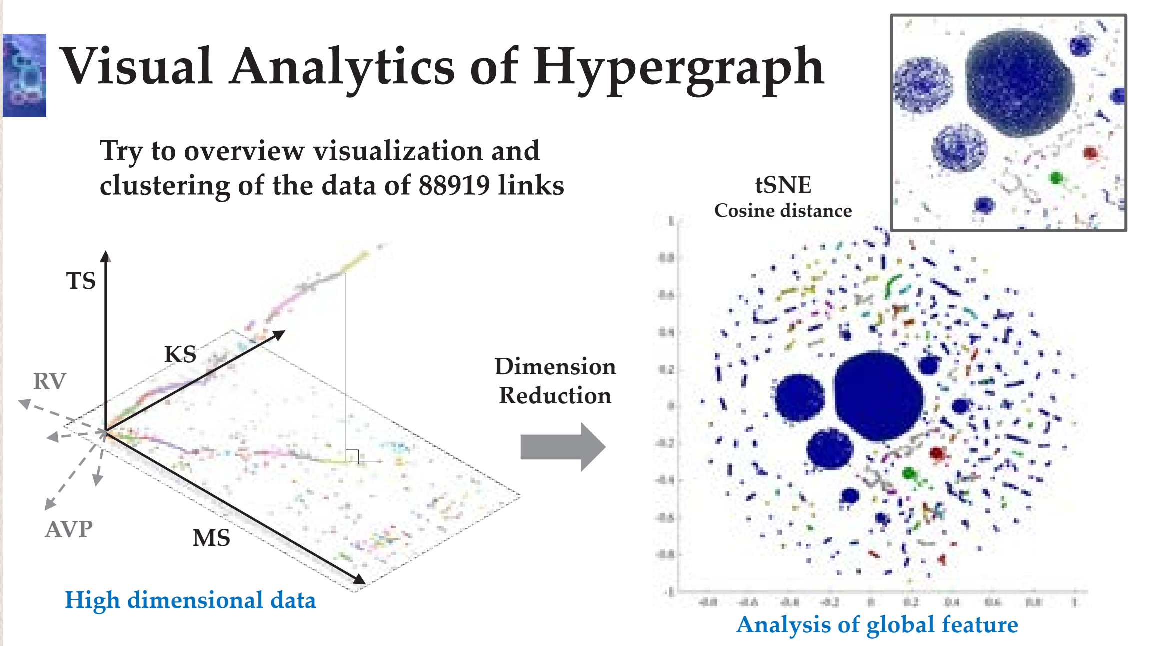


Visual Analytics of Hypergraph

Try to overview visualization and clustering of the data of MS, KS, and TS

Analysis of global feature

23



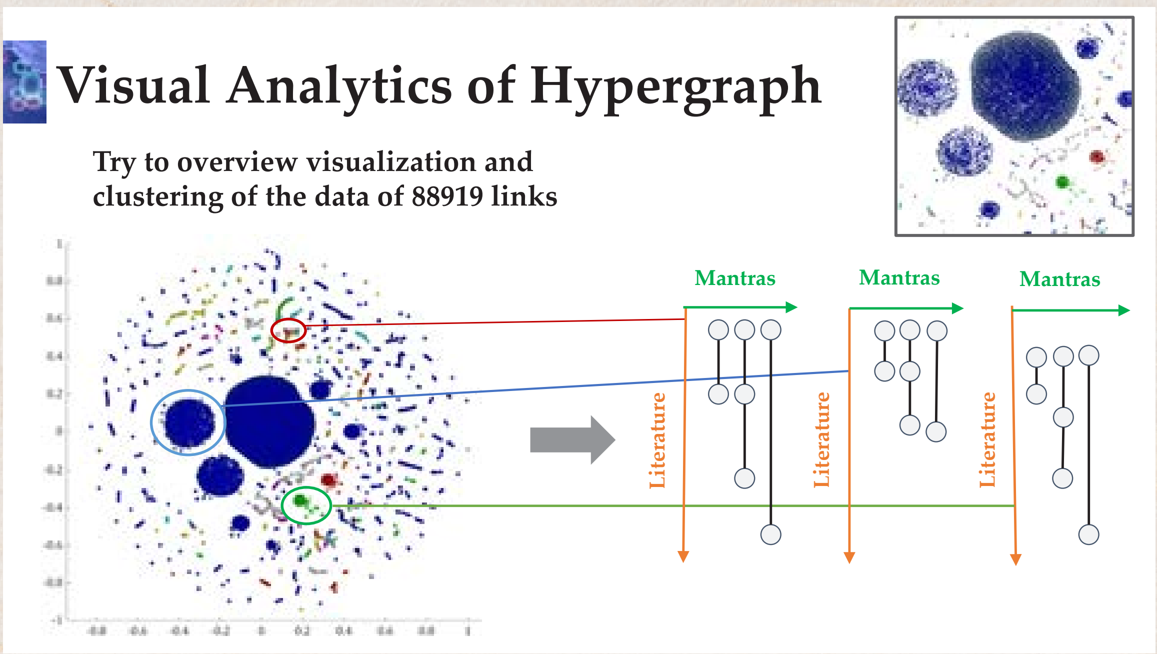
Visual Analytics of Hypergraph

Try to overview visualization and clustering of the data of 88919 links

High dimensional data

Analysis of global feature

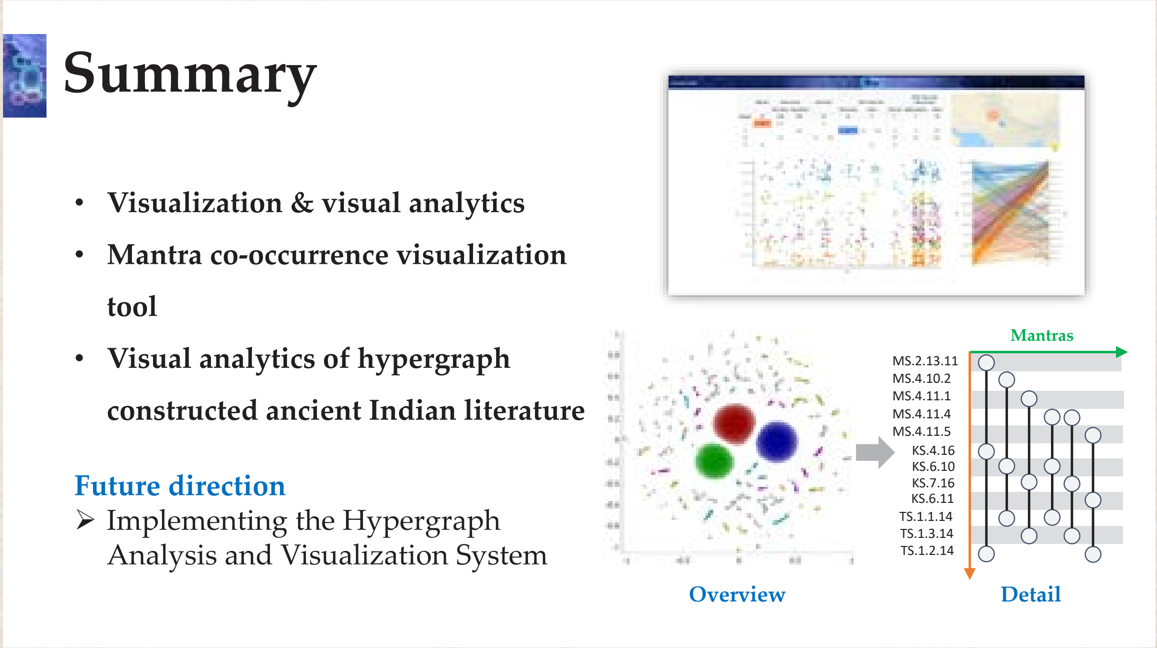
24



Visual Analytics of Hypergraph

Try to overview visualization and clustering of the data of 88919 links

25



Summary

- Visualization & visual analytics
- Mantra co-occurrence visualization tool
- Visual analytics of hypergraph constructed ancient Indian literature

Future direction

➢ Implementing the Hypergraph Analysis and Visualization System

26

Thank you for your attention

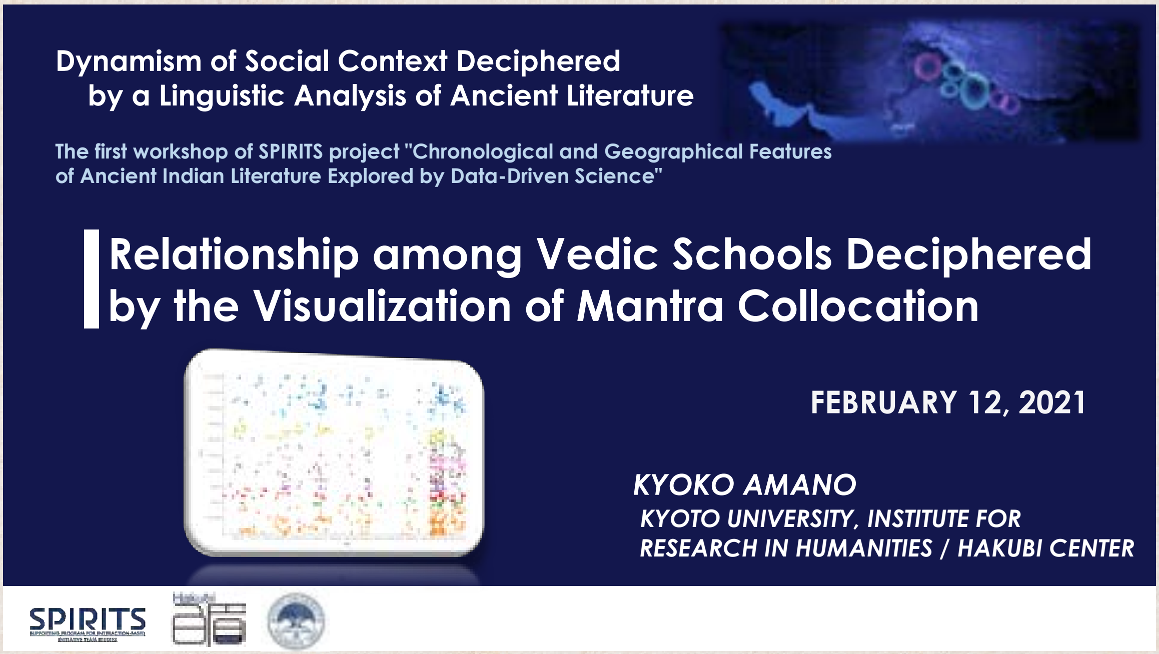
Relationship among Vedic Schools Deciphered by the Visualization of Mantra Collocation

マントラ共起関係の可視化から読み解くヴェーダ学派間の関係性

天野恭子（京都大学 白眉センター・人文科学研究所）

Kyoko Amano (Kyoto University, Institute for Research in Humanities / Hakubi Center)

1



Dynamism of Social Context Deciphered
by a Linguistic Analysis of Ancient Literature

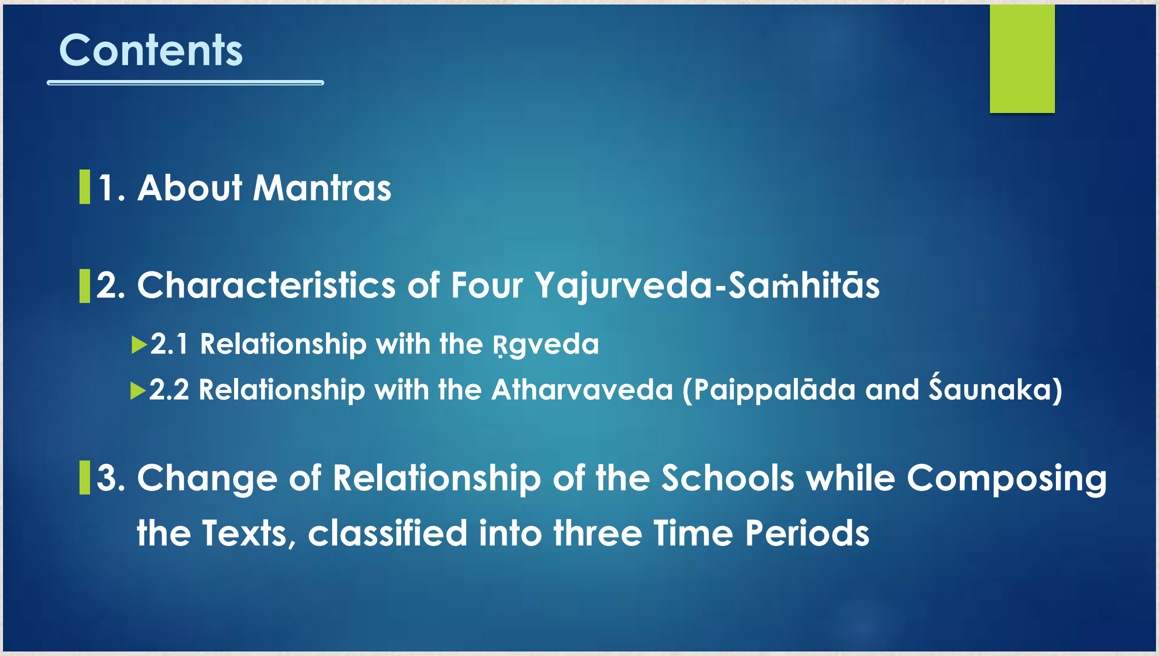
The first workshop of SPIRITS project "Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science"

Relationship among Vedic Schools Deciphered by the Visualization of Mantra Collocation

FEBRUARY 12, 2021

KYOKO AMANO
KYOTO UNIVERSITY, INSTITUTE FOR RESEARCH IN HUMANITIES / HAKUBI CENTER

2



Contents

1. About Mantras

2. Characteristics of Four Yajurveda-Saṁhitās
   - ▶2.1 Relationship with the Ṛgveda
   - ▶2.2 Relationship with the Atharvaveda (Paippalāda and Śaunaka)

3. Change of Relationship of the Schools while Composing the Texts, classified into three Time Periods

3

What is Mantra?

- mantras are ritual formula;
- hymns (verses) recited to invoke and praise the gods;
- ritual formula to give a symbolic role to ritual too

4

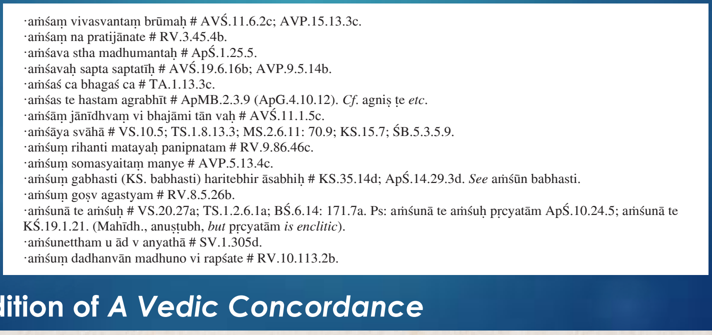
Collocation of mantras / Differece of mantras

- in several texts of a certain family / school
  (chronological axis)
- in several texts of several schools
  (geographical axis / confluent relation / status of networking)
  for example: mantra A in MS :: mantra B in KS, TS
- used for several rituals (development of rituals)
  the rituals were related closely with each other,
  a ritual influenced another, etc.

5

Data for investigation of mantra collocation

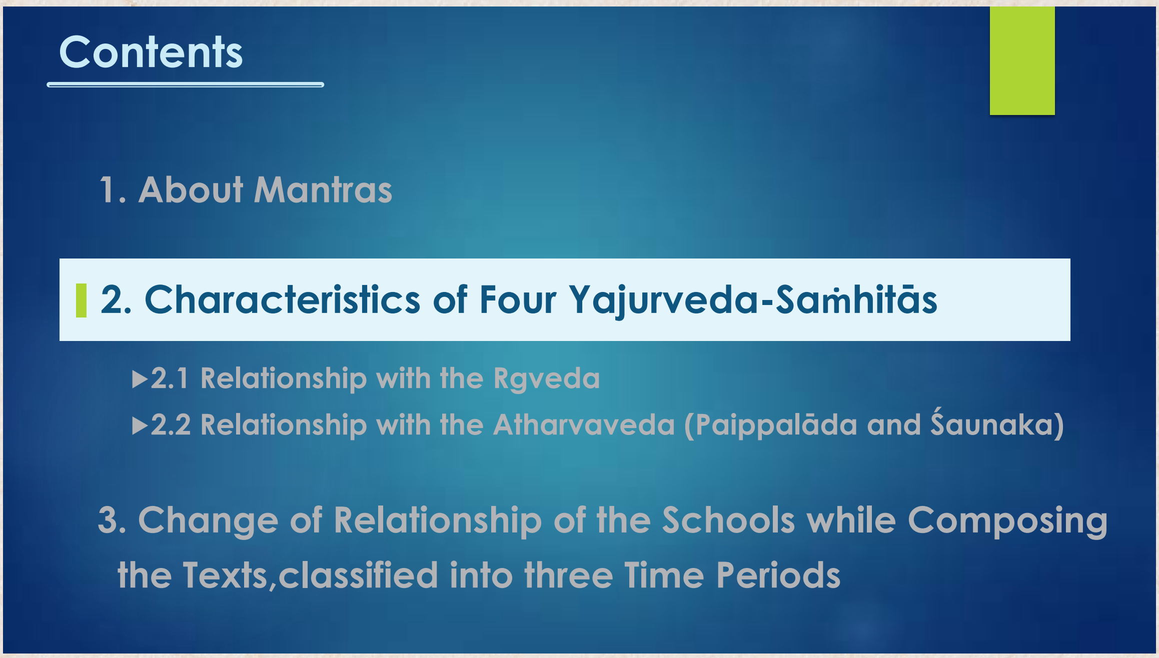
Bloomfield, Maurice (1893):
A Vedic Concordance. [Harvard Oriental Series 10]. Cambridge – Mass.
Franceschini, Marco (2007):
An updated Vedic concordance : Maurice Bloomfield's A Vedic concordance enhanced with new material taken from seven Vedic texts. Cambridge: Dept. of Sanskrit and Indian Studies, Harvard University



·aṁśaṃ vivasvantaṃ brūmaḥ # AVŚ.11.6.2c; AVP.15.13.3c.
·aṁśaṃ na pratijānate # RV.3.45.4b.
·aṁśava stha madhumantaḥ # ApŚ.1.25.5.
·aṁśavaḥ sapta saptatīḥ # AVŚ.19.6.16b; AVP.9.5.14b.
·aṁśaś ca bhagaś ca # TA.1.13.3c.
·aṁśas te hastam agrabhīt # ApMB.2.3.9 (ApG.4.10.12). *Cf.* agniṣ ṭe *etc*.
·aṁśāṃ jānīdhvaṃ vi bhajāmi tān vaḥ # AVŚ.11.1.5c.
·aṁśāya svāhā # VS.10.5; TS.1.8.13.3; MS.2.6.11: 70.9; KS.15.7; ŚB.5.3.5.9.
·aṁśuṃ rihanti matayaḥ panipnatam # RV.9.86.46c.
·aṁśuṃ somasyaitaṃ manye # AVP.5.13.4c.
·aṁśuṃ gabhasti (KS. babhasti) haritebhir āsabhiḥ # KS.35.14d; ApŚ.14.29.3d. *See* aṁśūn babhasti.
·aṁśuṃ goṣv agastyam # RV.8.5.26b.
·aṁśunā te aṁśuḥ # VS.20.27a; TS.1.2.6.1a; BŚ.6.14: 171.7a. Ps: aṁśunā te aṁśuḥ pṛcyatām ApŚ.10.24.5; aṁśunā te KŚ.19.1.21. (Mahīdh., anuṣṭubh, *but* pṛcyatām *is enclitic*).
·aṁśunettham u ād v anyathā # SV.1.305d.
·aṁśuṃ dadhanvān madhuno vi rapśate # RV.10.113.2b.

electronic edition of *A Vedic Concordance*

6



Contents

1. About Mantras

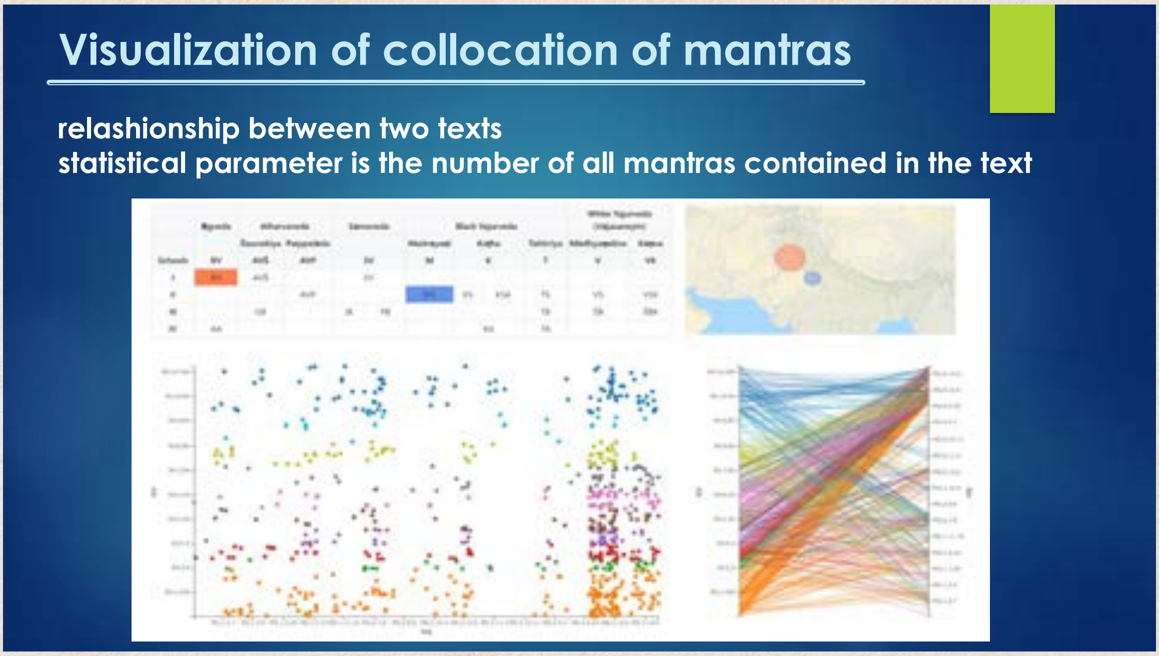
2. Characteristics of Four Yajurveda-Saṁhitās

▶2.1 Relationship with the Rgveda

▶2.2 Relationship with the Atharvaveda (Paippalāda and Śaunaka)

3. Change of Relationship of the Schools while Composing the Texts,classified into three Time Periods

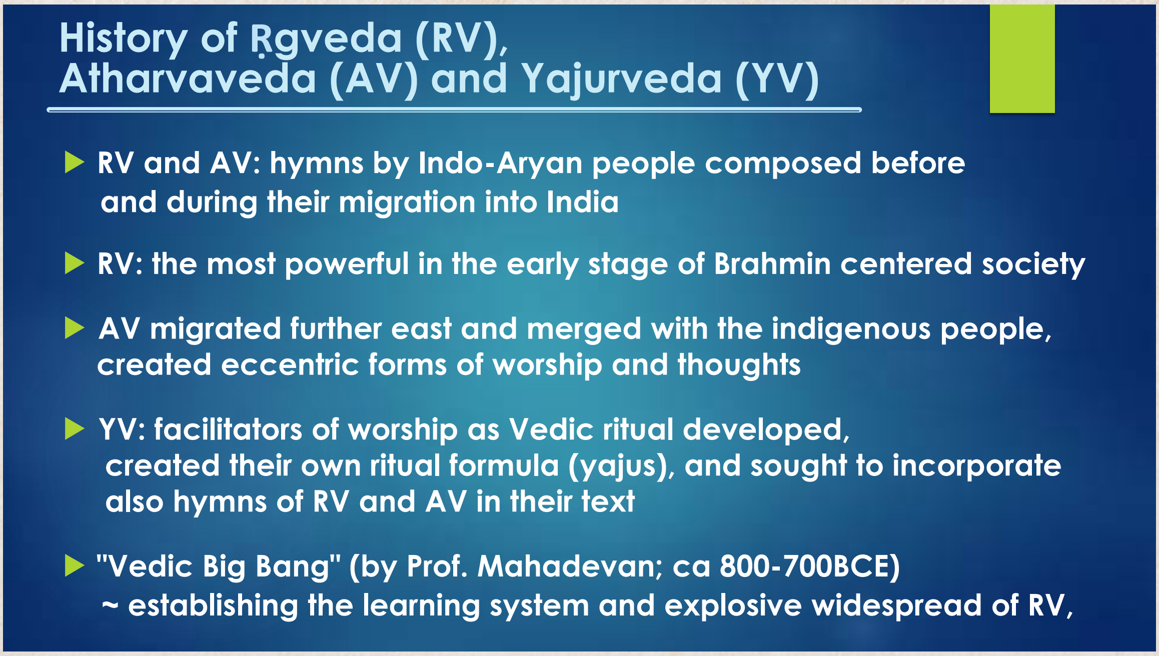
7



Visualization of collocation of mantras

relashionship between two texts
statistical parameter is the number of all mantras contained in the text

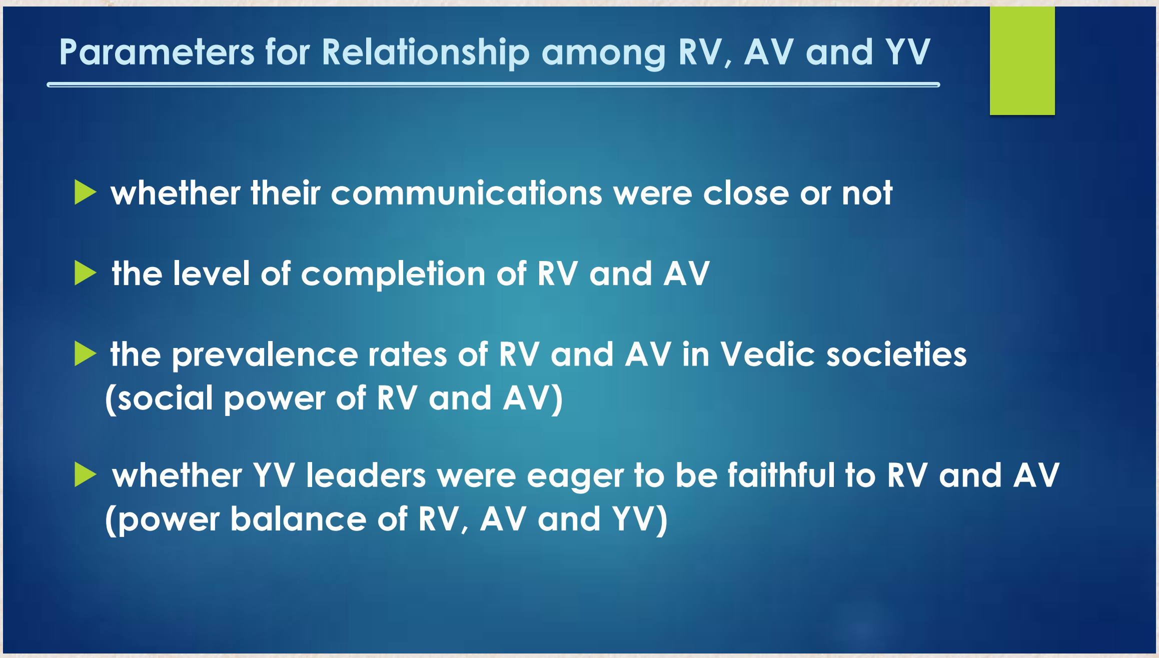
8



History of Ṛgveda (RV), Atharvaveda (AV) and Yajurveda (YV)

- RV and AV: hymns by Indo-Aryan people composed before and during their migration into India
- RV: the most powerful in the early stage of Brahmin centered society
- AV migrated further east and merged with the indigenous people, created eccentric forms of worship and thoughts
- YV: facilitators of worship as Vedic ritual developed, created their own ritual formula (yajus), and sought to incorporate also hymns of RV and AV in their text
- "Vedic Big Bang" (by Prof. Mahadevan; ca 800-700BCE) ~ establishing the learning system and explosive widespread of RV,

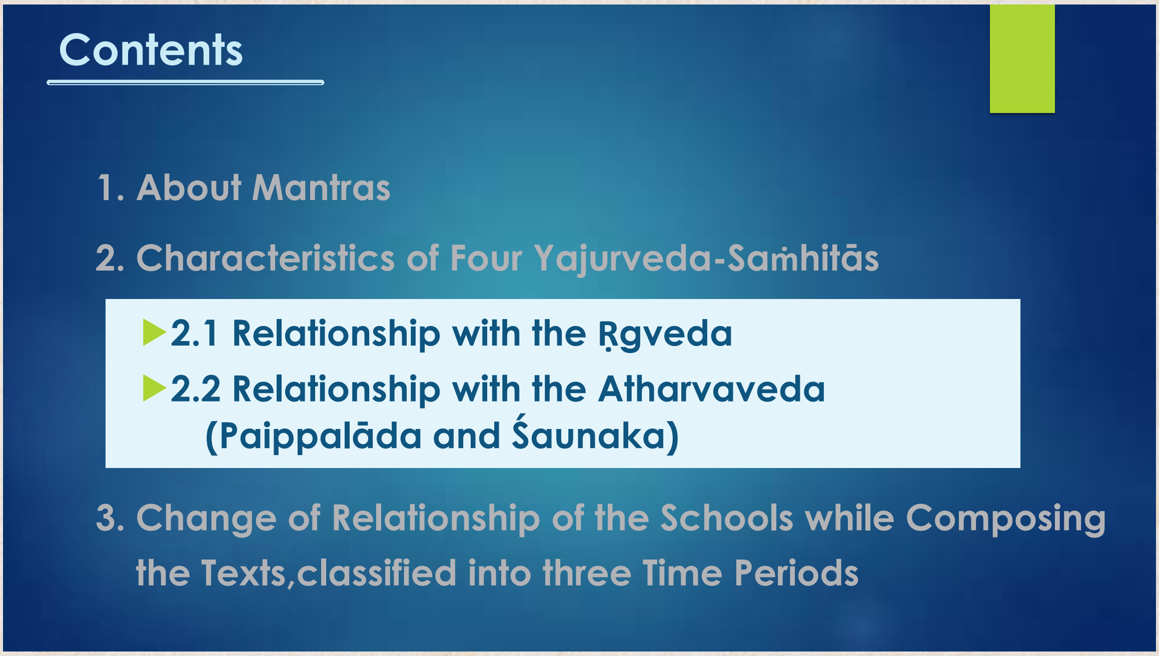
9



Parameters for Relationship among RV, AV and YV

- whether their communications were close or not
- the level of completion of RV and AV
- the prevalence rates of RV and AV in Vedic societies (social power of RV and AV)
- whether YV leaders were eager to be faithful to RV and AV (power balance of RV, AV and YV)

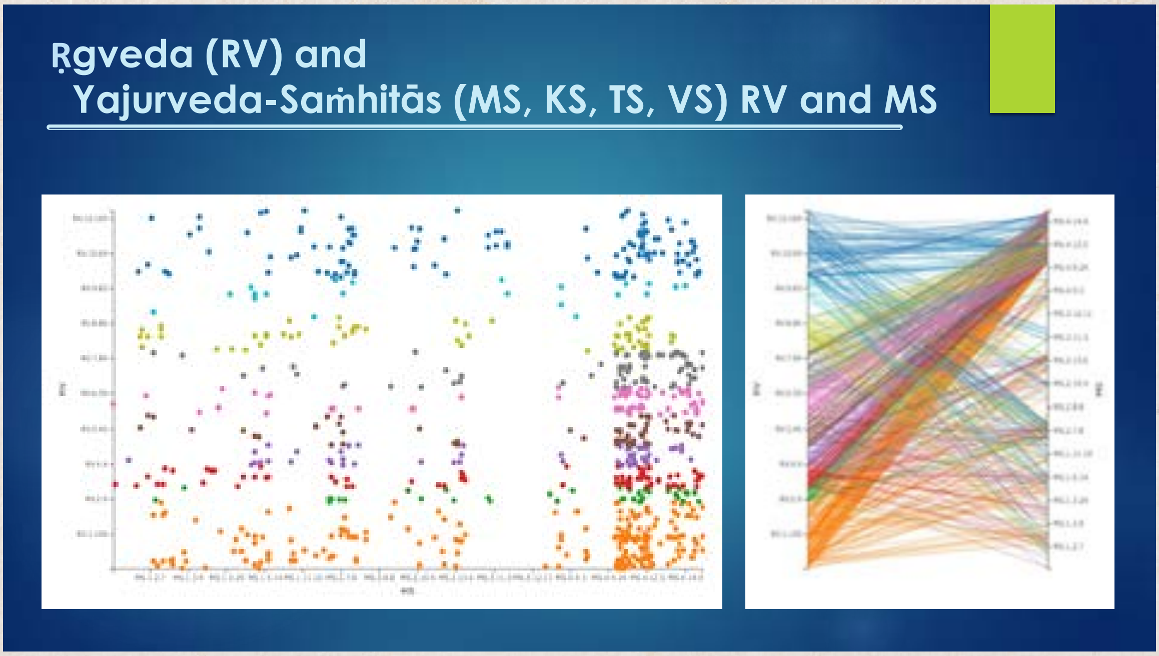
10



Contents

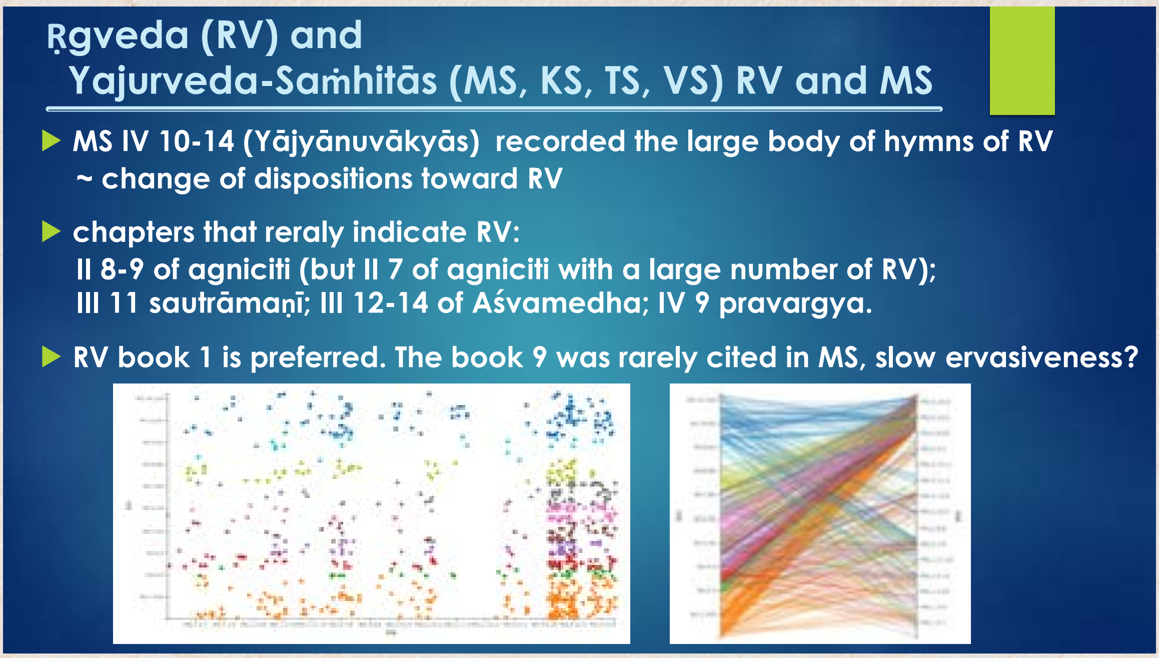
1. About Mantras
2. Characteristics of Four Yajurveda-Saṁhitās
   - 2.1 Relationship with the Ṛgveda
   - 2.2 Relationship with the Atharvaveda (Paippalāda and Śaunaka)
3. Change of Relationship of the Schools while Composing the Texts,classified into three Time Periods

11



Ṛgveda (RV) and Yajurveda-Saṁhitās (MS, KS, TS, VS) RV and MS

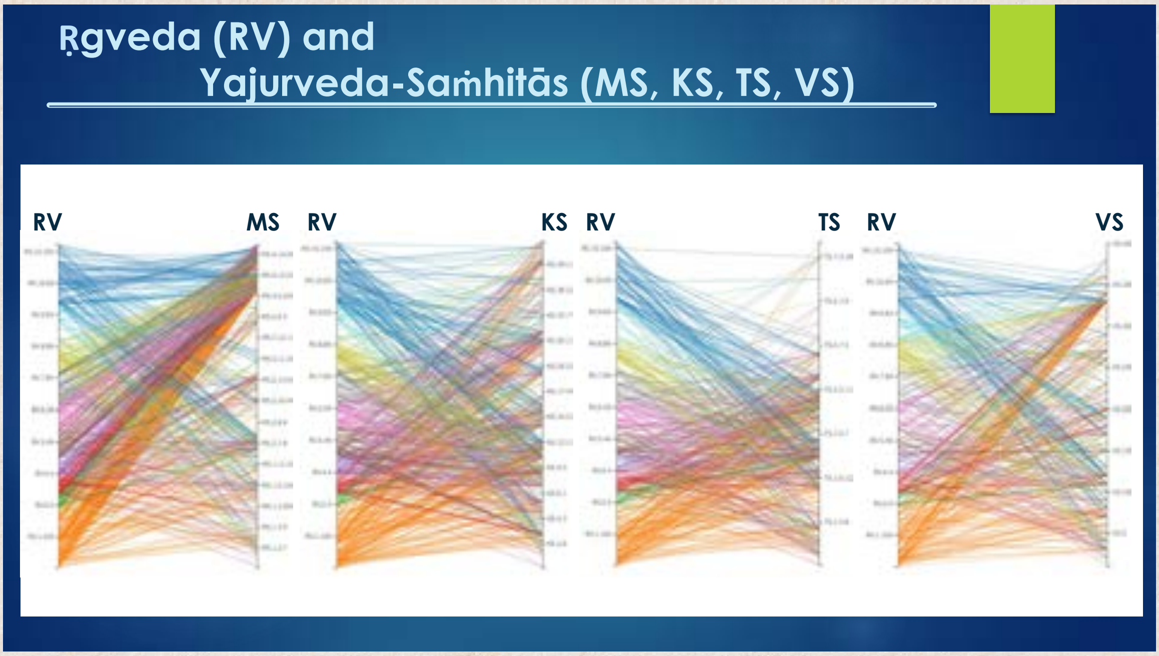
12



Ṛgveda (RV) and Yajurveda-Saṁhitās (MS, KS, TS, VS) RV and MS

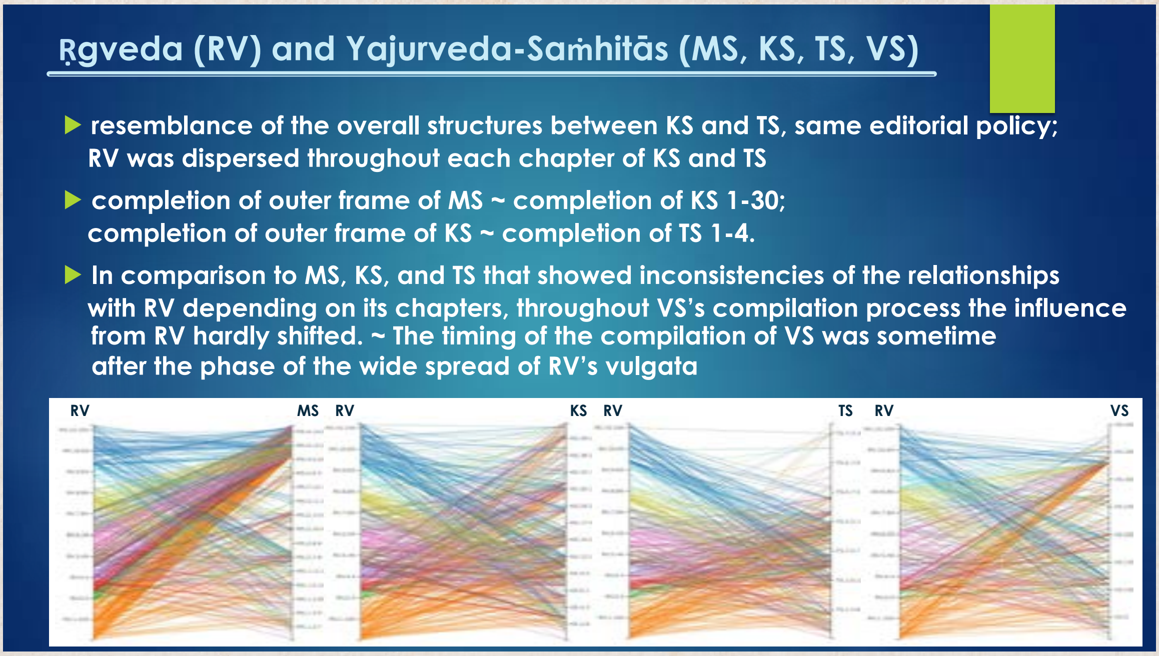
- MS IV 10-14 (Yājyānuvākyās) recorded the large body of hymns of RV ~ change of dispositions toward RV
- chapters that reraly indicate RV:
  II 8-9 of agniciti (but II 7 of agniciti with a large number of RV);
  III 11 sautrāmaṇī; III 12-14 of Aśvamedha; IV 9 pravargya.
- RV book 1 is preferred. The book 9 was rarely cited in MS, slow ervasiveness?

13



Ṛgveda (RV) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

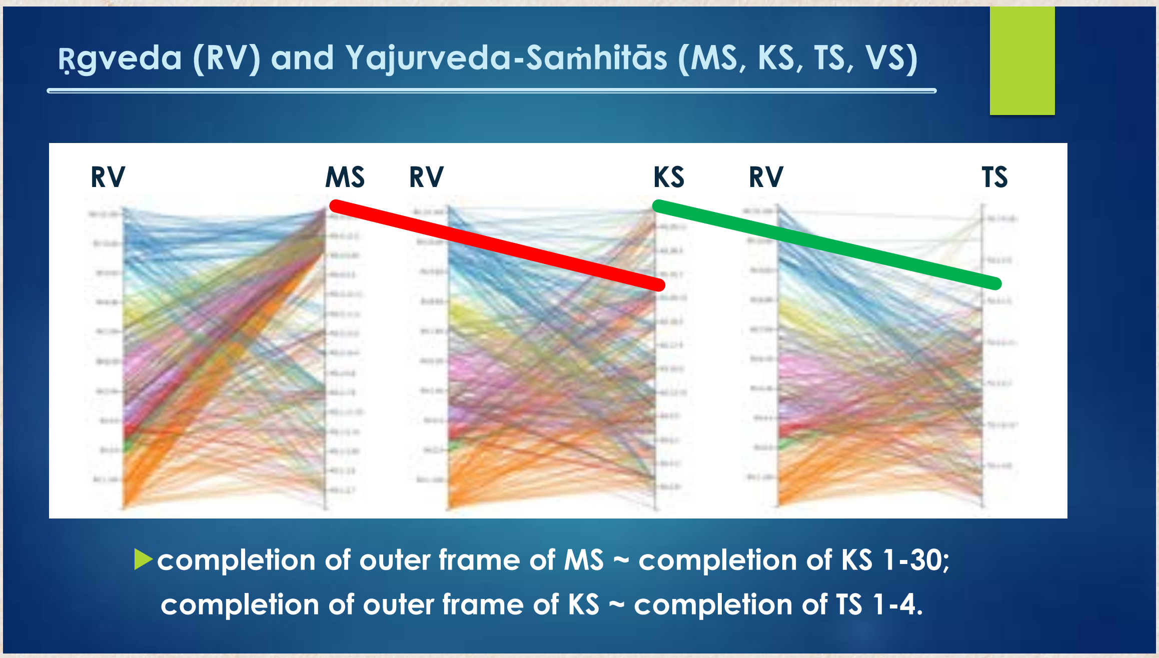
14



Ṛgveda (RV) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

- resemblance of the overall structures between KS and TS, same editorial policy; RV was dispersed throughout each chapter of KS and TS
- completion of outer frame of MS ~ completion of KS 1-30; completion of outer frame of KS ~ completion of TS 1-4.
- In comparison to MS, KS, and TS that showed inconsistencies of the relationships with RV depending on its chapters, throughout VS's compilation process the influence from RV hardly shifted. ~ The timing of the compilation of VS was sometime after the phase of the wide spread of RV's vulgata

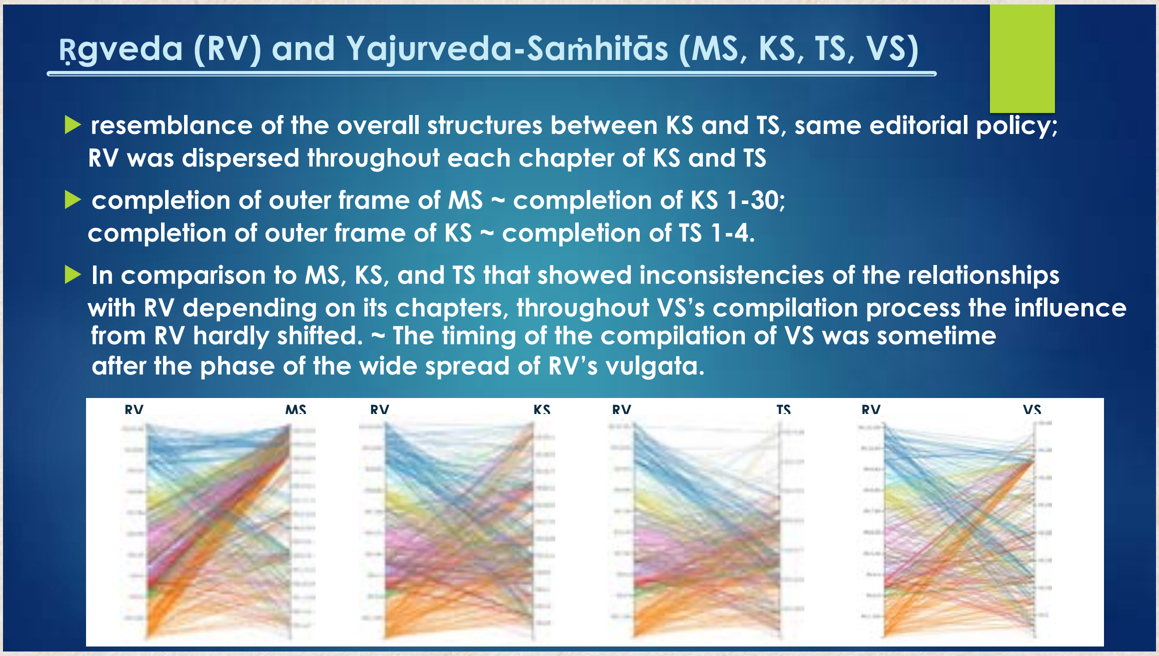
15



Ṛgveda (RV) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

▶completion of outer frame of MS ~ completion of KS 1-30;
completion of outer frame of KS ~ completion of TS 1-4.

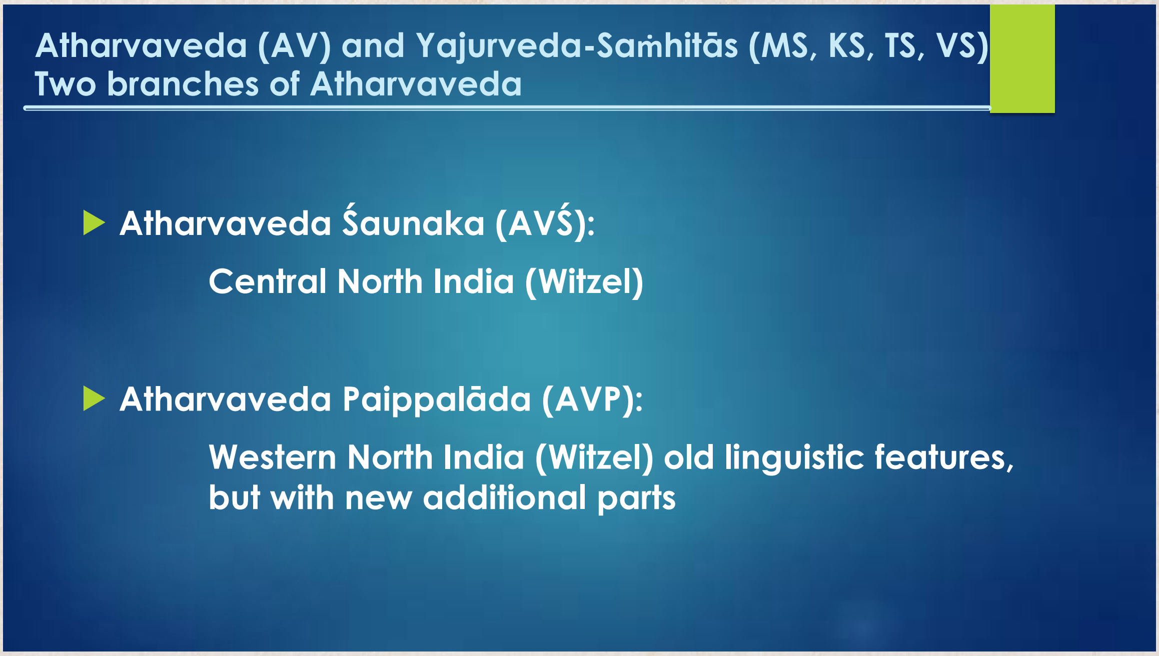
16



Ṛgveda (RV) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

- ▶ resemblance of the overall structures between KS and TS, same editorial policy; RV was dispersed throughout each chapter of KS and TS
- ▶ completion of outer frame of MS ~ completion of KS 1-30; completion of outer frame of KS ~ completion of TS 1-4.
- ▶ In comparison to MS, KS, and TS that showed inconsistencies of the relationships with RV depending on its chapters, throughout VS's compilation process the influence from RV hardly shifted. ~ The timing of the compilation of VS was sometime after the phase of the wide spread of RV's vulgata.

17



Atharvaveda (AV) and Yajurveda-Saṁhitās (MS, KS, TS, VS)
Two branches of Atharvaveda

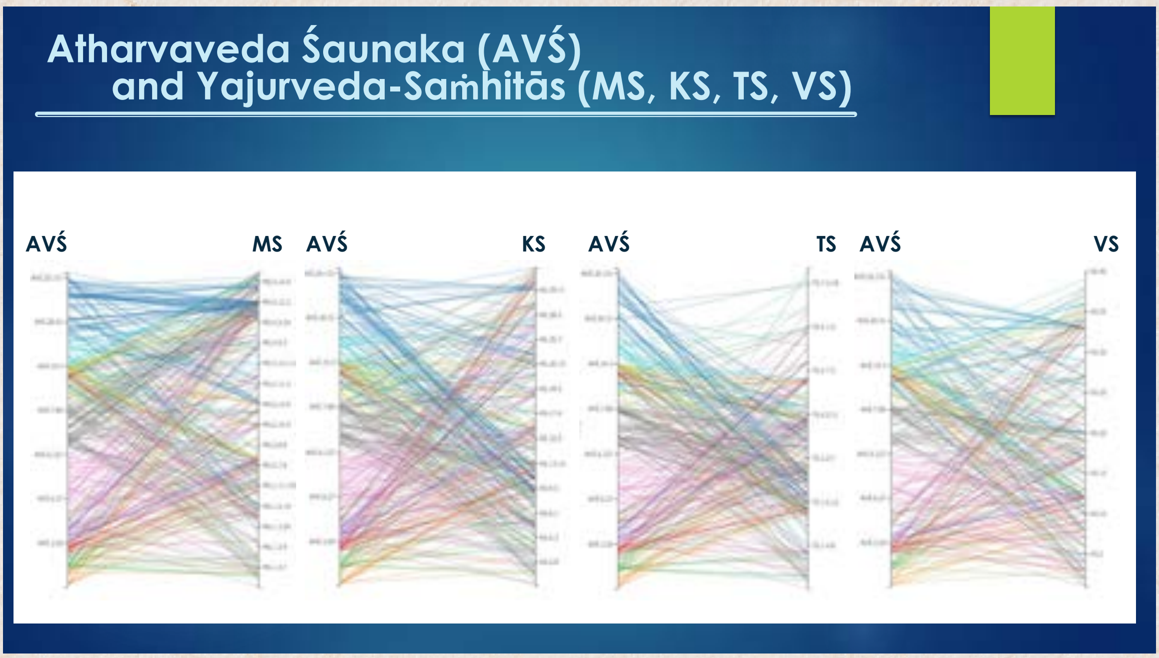
▶ Atharvaveda Śaunaka (AVŚ):

Central North India (Witzel)

▶ Atharvaveda Paippalāda (AVP):

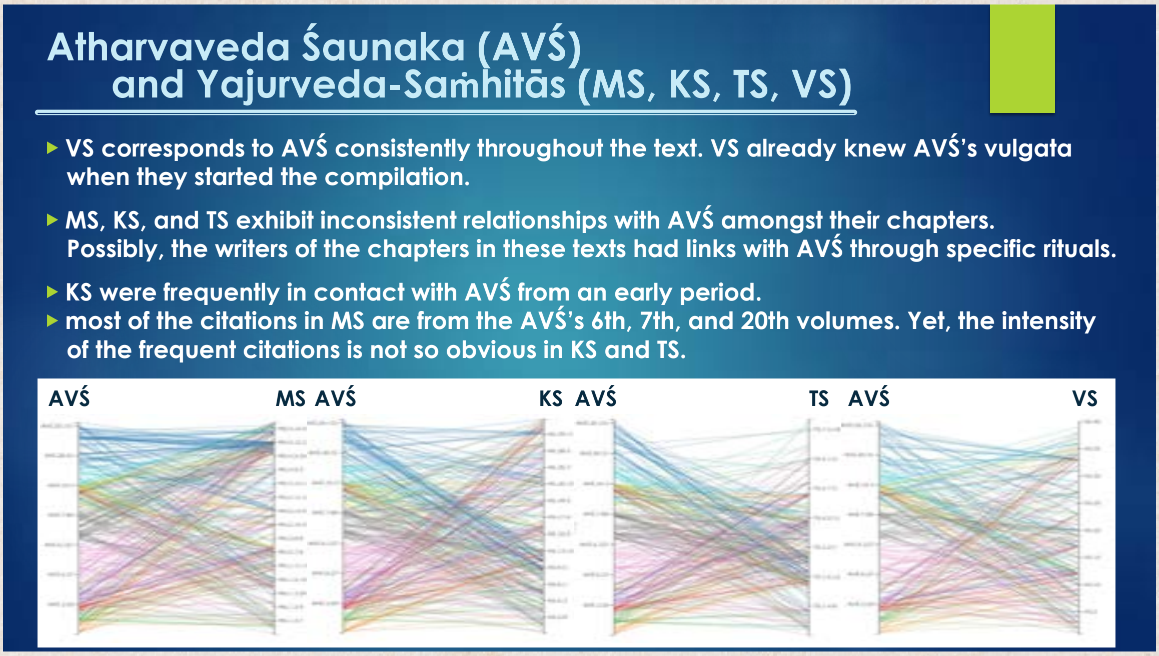
Western North India (Witzel) old linguistic features,
but with new additional parts

18



Atharvaveda Śaunaka (AVŚ) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

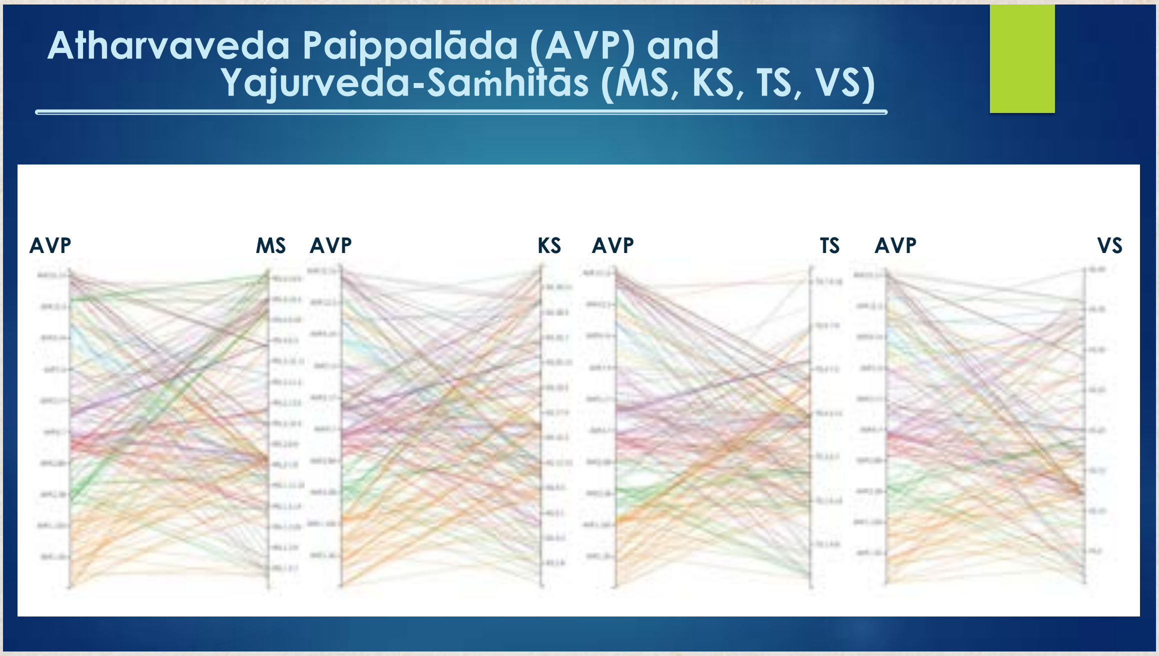
19



Atharvaveda Śaunaka (AVŚ) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

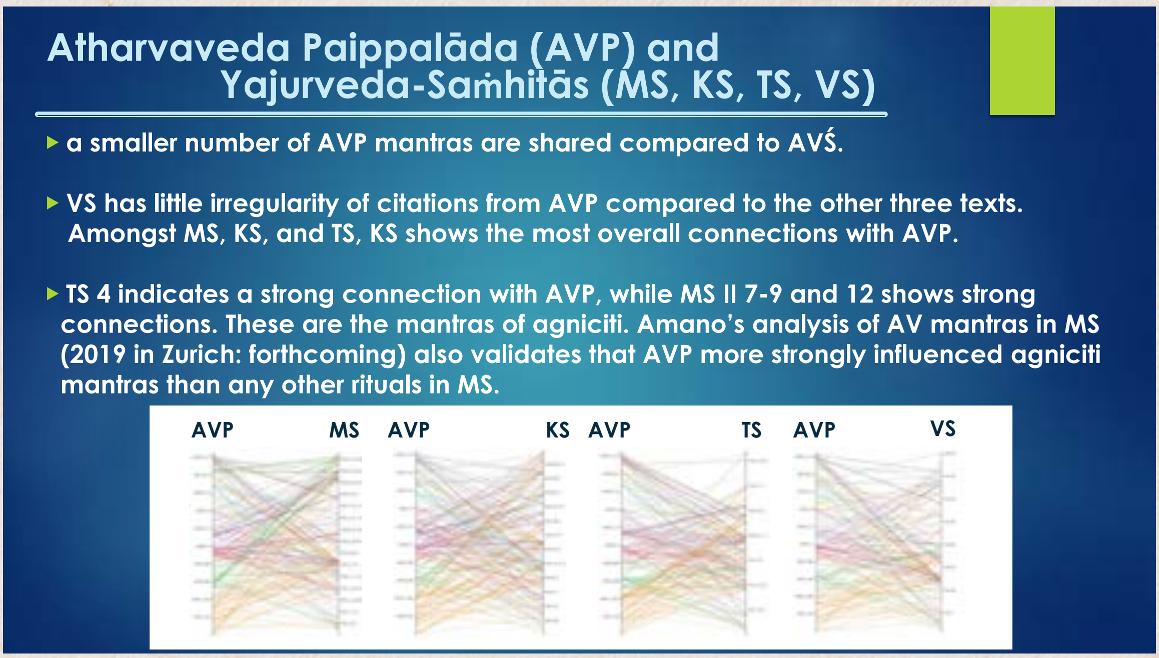
- VS corresponds to AVŚ consistently throughout the text. VS already knew AVŚ's vulgata when they started the compilation.
- MS, KS, and TS exhibit inconsistent relationships with AVŚ amongst their chapters. Possibly, the writers of the chapters in these texts had links with AVŚ through specific rituals.
- KS were frequently in contact with AVŚ from an early period.
- most of the citations in MS are from the AVŚ's 6th, 7th, and 20th volumes. Yet, the intensity of the frequent citations is not so obvious in KS and TS.

20



Atharvaveda Paippalāda (AVP) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

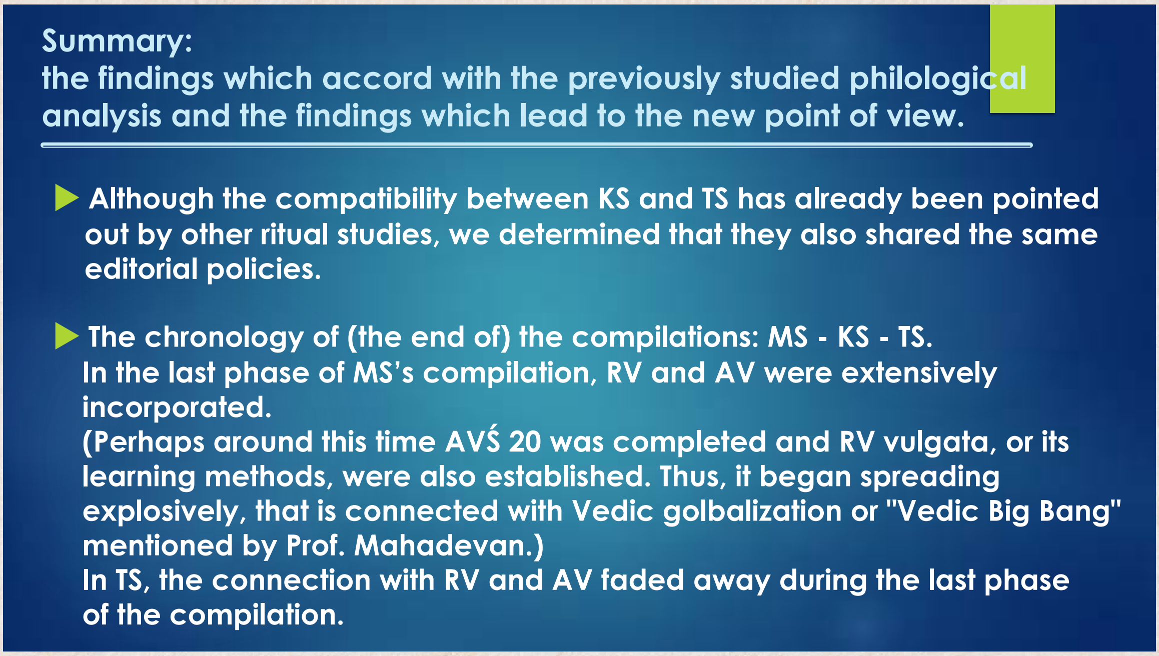
21



Atharvaveda Paippalāda (AVP) and Yajurveda-Saṁhitās (MS, KS, TS, VS)

- a smaller number of AVP mantras are shared compared to AVŚ.
- VS has little irregularity of citations from AVP compared to the other three texts. Amongst MS, KS, and TS, KS shows the most overall connections with AVP.
- TS 4 indicates a strong connection with AVP, while MS II 7-9 and 12 shows strong connections. These are the mantras of agniciti. Amano’s analysis of AV mantras in MS (2019 in Zurich: forthcoming) also validates that AVP more strongly influenced agniciti mantras than any other rituals in MS.

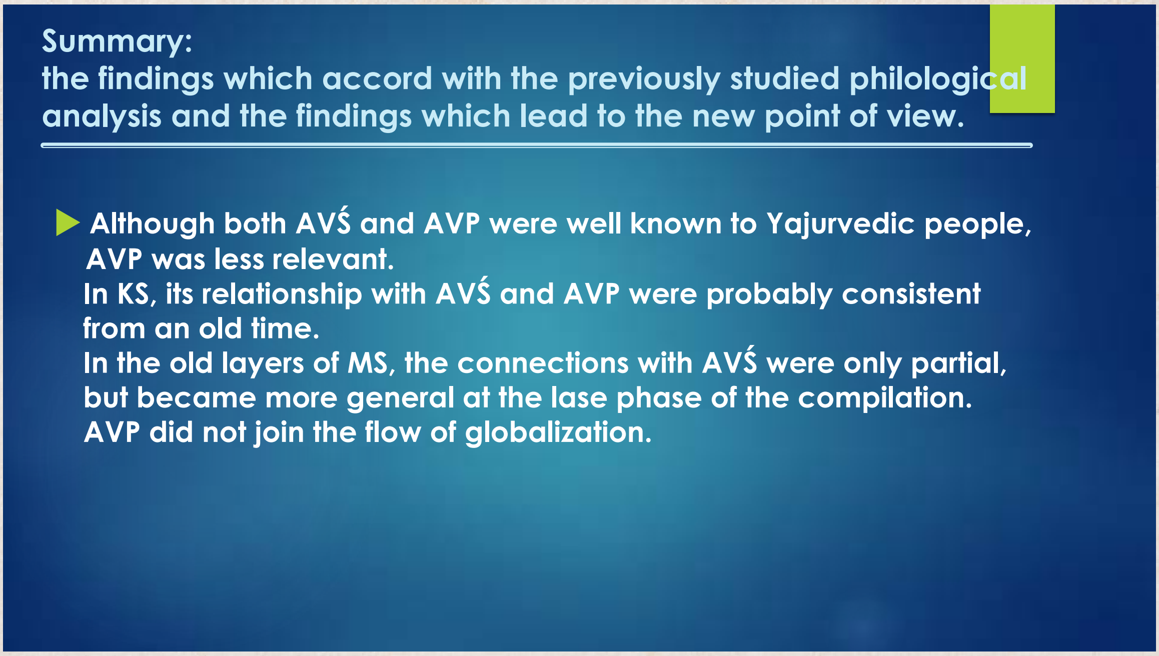
22



Summary: the findings which accord with the previously studied philological analysis and the findings which lead to the new point of view.

- Although the compatibility between KS and TS has already been pointed out by other ritual studies, we determined that they also shared the same editorial policies.
- The chronology of (the end of) the compilations: MS - KS - TS.
  In the last phase of MS’s compilation, RV and AV were extensively incorporated.
  (Perhaps around this time AVŚ 20 was completed and RV vulgata, or its learning methods, were also established. Thus, it began spreading explosively, that is connected with Vedic golbalization or "Vedic Big Bang" mentioned by Prof. Mahadevan.)
  In TS, the connection with RV and AV faded away during the last phase of the compilation.

23



Summary: the findings which accord with the previously studied philological analysis and the findings which lead to the new point of view.

- Although both AVŚ and AVP were well known to Yajurvedic people, AVP was less relevant.
  In KS, its relationship with AVŚ and AVP were probably consistent from an old time.
  In the old layers of MS, the connections with AVŚ were only partial, but became more general at the lase phase of the compilation.
  AVP did not join the flow of globalization.

24

SPIRITS
SUPPORTING PROGRAM FOR INTERACTION-BASED INITIATIVE TEAM STUDIES
Hakubi
白眉
Kyoto Univ.
KYOTO UNIVERSITY
KYOTO JAPAN
FOUNDED 1897

data reduction by
Naoko Oshiro, Chihiro Ueda, Sousuke Tanaka

Presentation 3 ｜報告 3

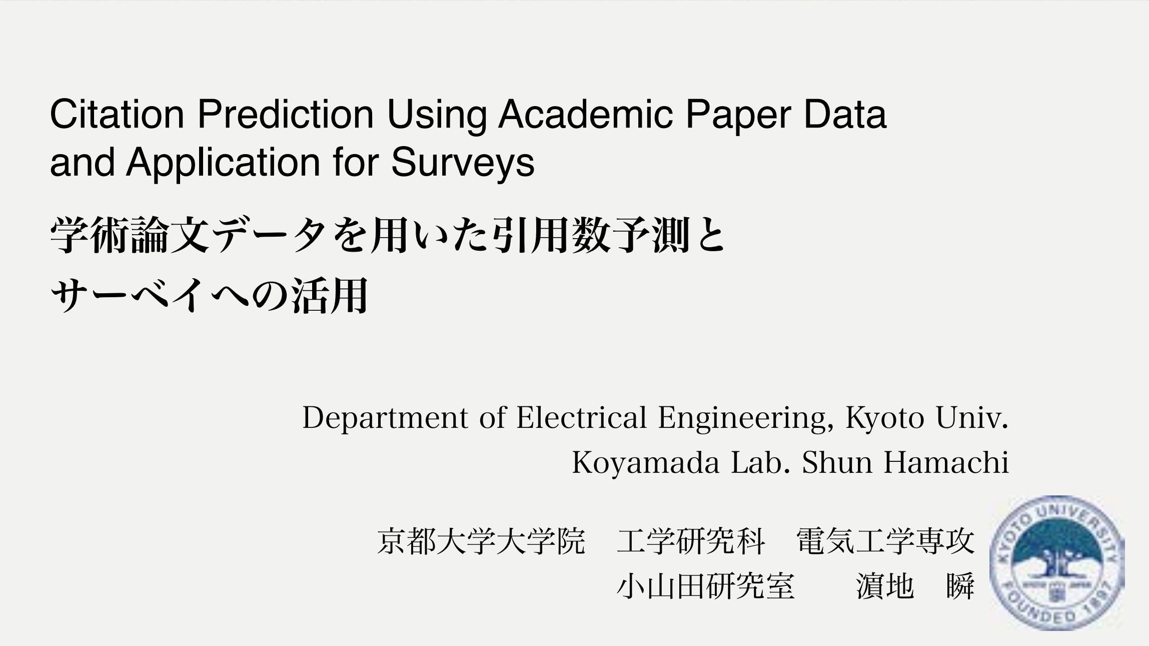
Citation Prediction Using Academic Paper Data and Application for Surveys

学術論文データを用いた引用数予測とサーベイへの活用

濵地瞬（京都大学 工学研究科）

Shun Hamachi (Kyoto University, Graduate School of Engineering)

1



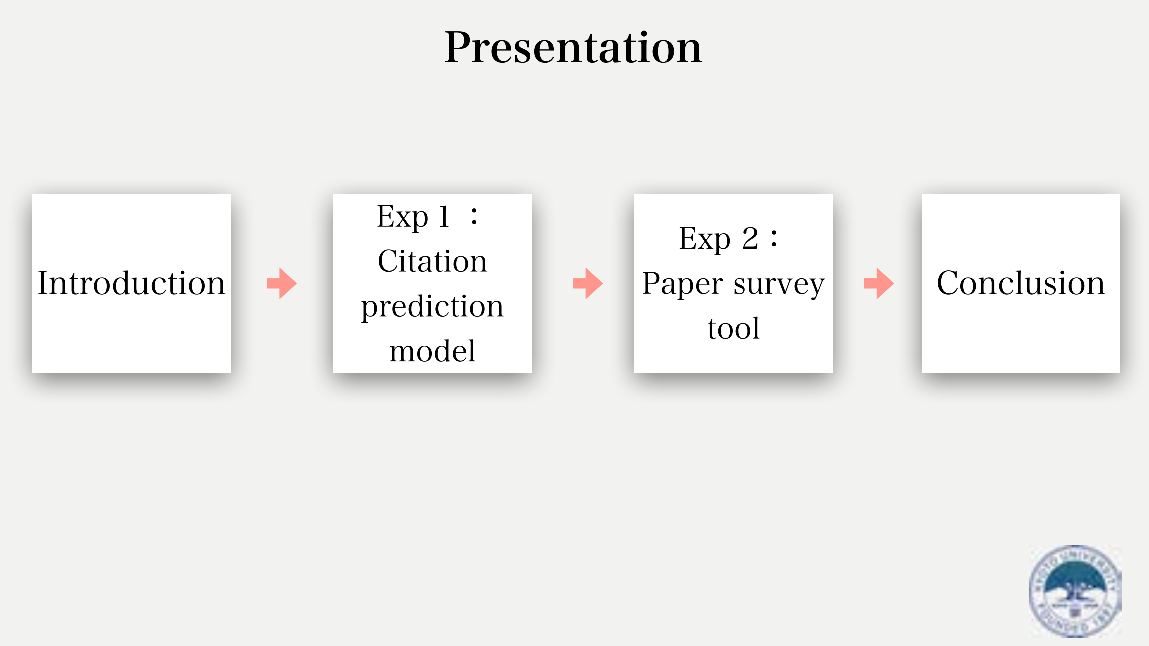
Citation Prediction Using Academic Paper Data
and Application for Surveys

学術論文データを用いた引用数予測と
サーベイへの活用

Department of Electrical Engineering, Kyoto Univ.
Koyamada Lab. Shun Hamachi

京都大学大学院　工学研究科　電気工学専攻
小山田研究室　　濵地　瞬

2

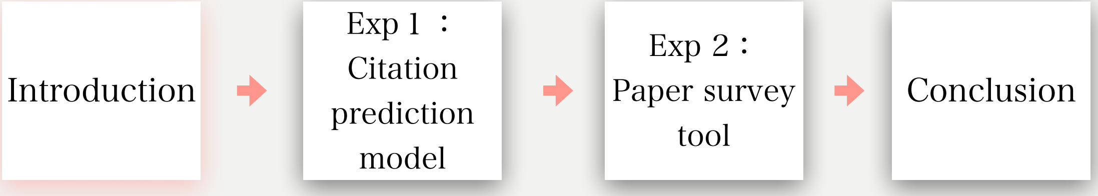


Presentation

Introduction → Exp 1 ：Citation prediction model → Exp 2：Paper survey tool → Conclusion

3

Presentation


Introduction

Exp 1 ： Citation prediction model

Exp 2 : Paper survey tool

Conclusion

4

What is the paper survey ?

Introduction

For researchers

To search and study related papers complehensively

Purpose of the paper survey

- Find tips for your research theme
- Find the elemental technology you need
- Show novelty and superiority of your research

etc . . .

5

Ordinary method of the paper survey

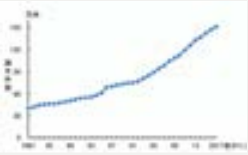
Introduction

Academic research database

Example：Web of Science

Challenges

- The number of papers is huge



The number of science papers in the world is increasing
(National Institute of Science and Technology Policy)

- The survey can be biased
- Difficult to understand related technics and community distribution

➜ **There is a risk of missing important authors and papers unless researchers look at the field of study from a higher perspective**

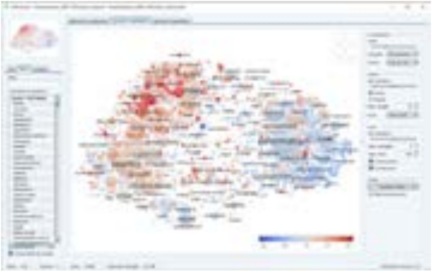
6

Academic paper network

Introduction

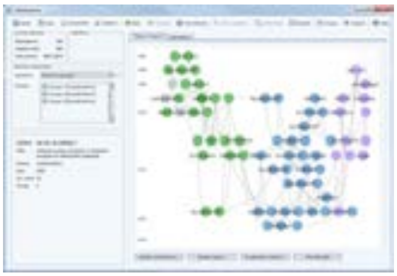
Network visualization is effective to look at the field of study from a higher perspective

Visualization of research field clusters



Software survey: VOSviewer, a computer program for bibliometric mapping (2010)

Visualization of citation network



CitNetExplorer: A new software tool for analyzing and visualizing citation networks (2014)

It is still difficult to select the most notable papers

➡ **It is necessary to pay attention to notable papers by associating the evaluation metrics of the papers.**

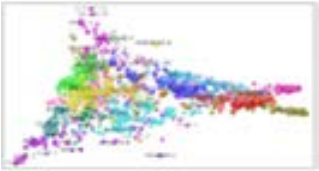
7

Citation count

Introduction

The most common paper evaluation metric is a citation count

The node size means the number of citations



A unified approach to mapping and clustering of bibliometric networks (2010)

Challenges of a citation count

- Inconsistent with the long-term value of the paper
- Time lag before starting to be cited

If we survey papers based solely on citation counts, we may miss new or important hidden papers.

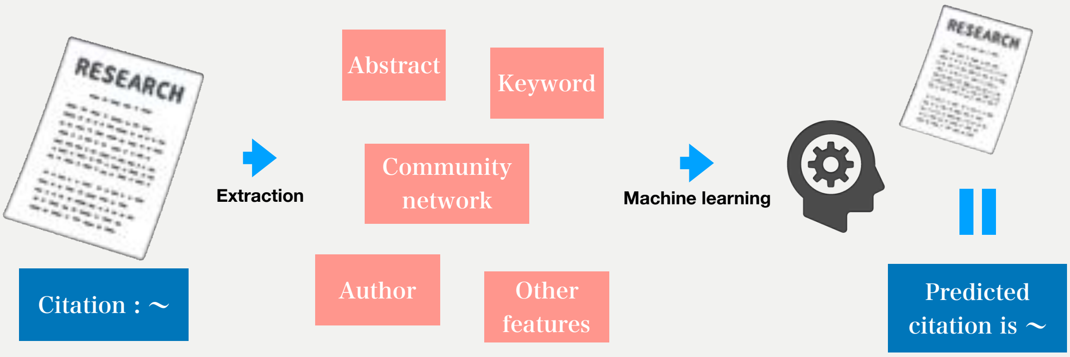
⬇

New metrics are needed to evaluate the essential importance of papers

8

Method : Paper survey tool utilizing citation count prediction

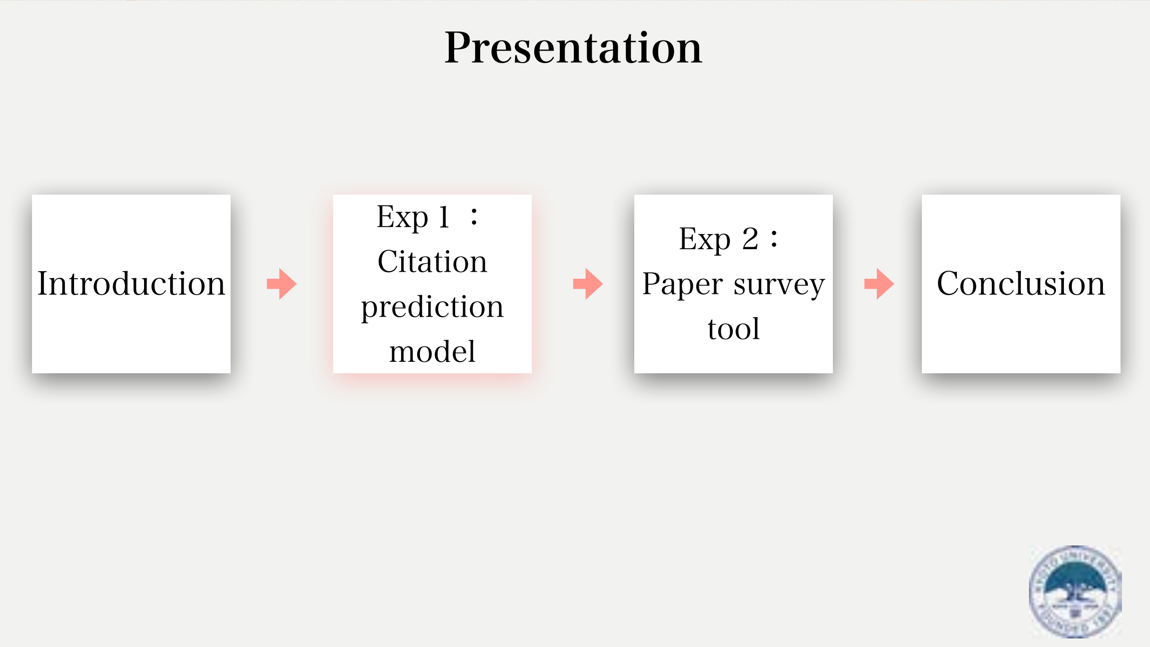
Introduction



➡ Predicted citations：evaluates papers from multiple perspectives
There is no time lag

Metrics that complement the number of citations

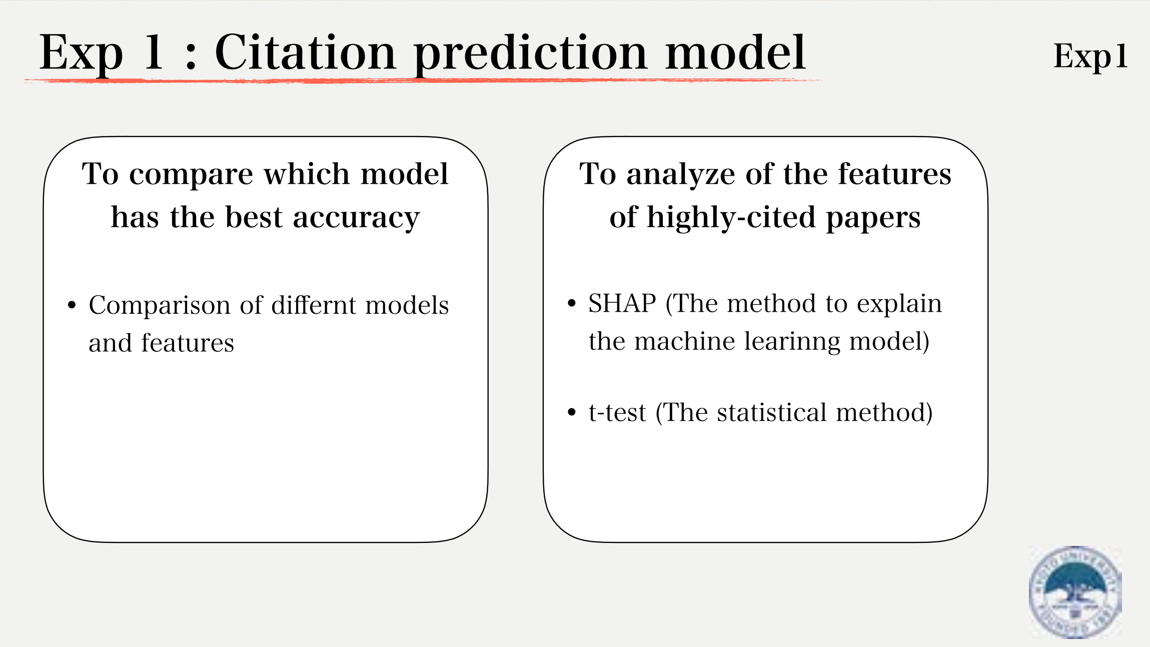
9



Presentation

Introduction → Exp 1 ： Citation prediction model → Exp 2 : Paper survey tool → Conclusion

10



Exp 1 : Citation prediction model

Exp1

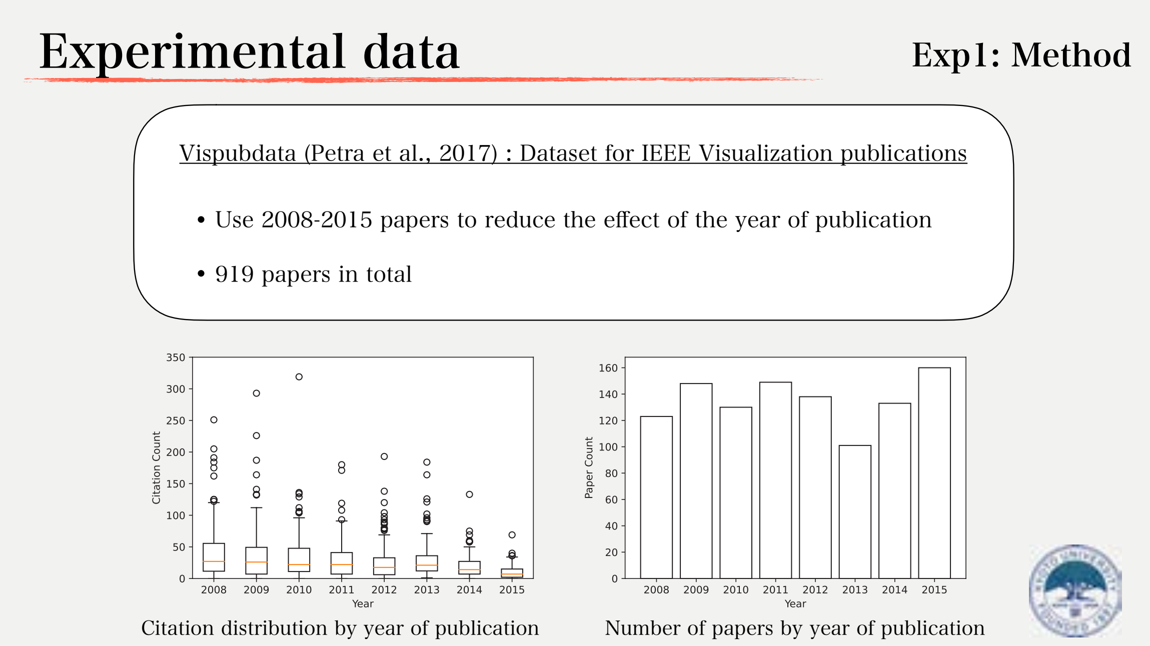
To compare which model has the best accuracy

- Comparison of differnt models and features

To analyze of the features of highly-cited papers

- SHAP (The method to explain the machine learinng model)
- t-test (The statistical method)

11



Experimental data

Exp1: Method

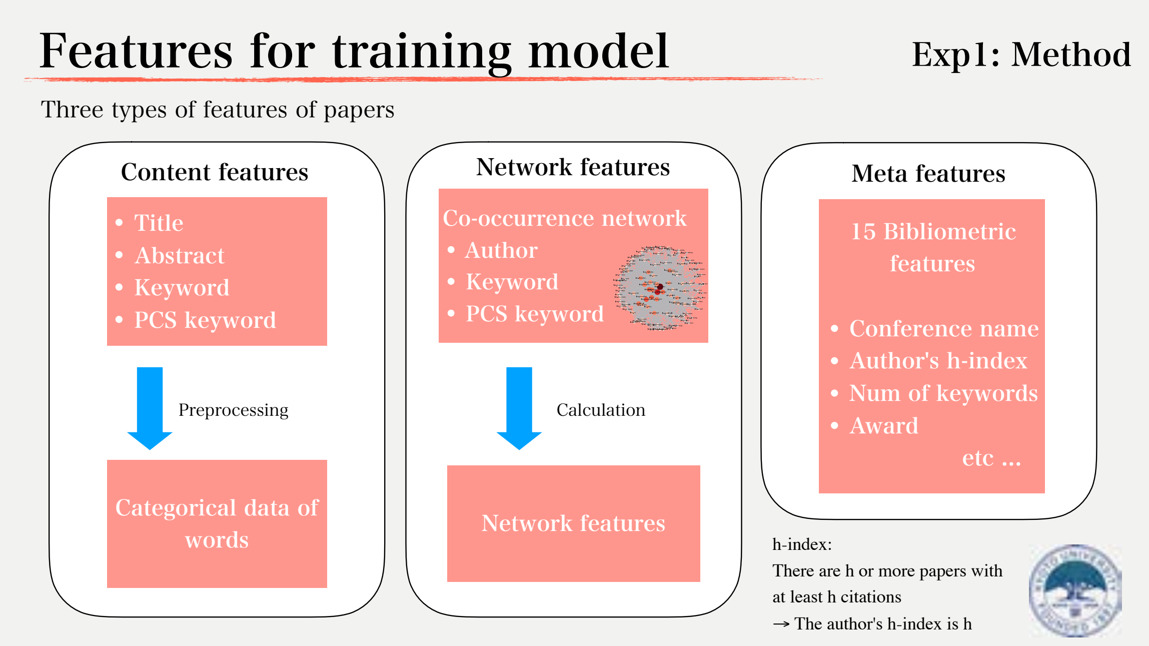
Vispubdata (Petra et al., 2017) : Dataset for IEEE Visualization publications

- Use 2008-2015 papers to reduce the effect of the year of publication
- 919 papers in total

Citation distribution by year of publication

Number of papers by year of publication

12



Features for training model

Exp1: Method

Three types of features of papers

Content features

- Title
- Abstract
- Keyword
- PCS keyword

Preprocessing

Categorical data of words

Network features

Co-occurrence network

- Author
- Keyword
- PCS keyword

Calculation

Network features

Meta features

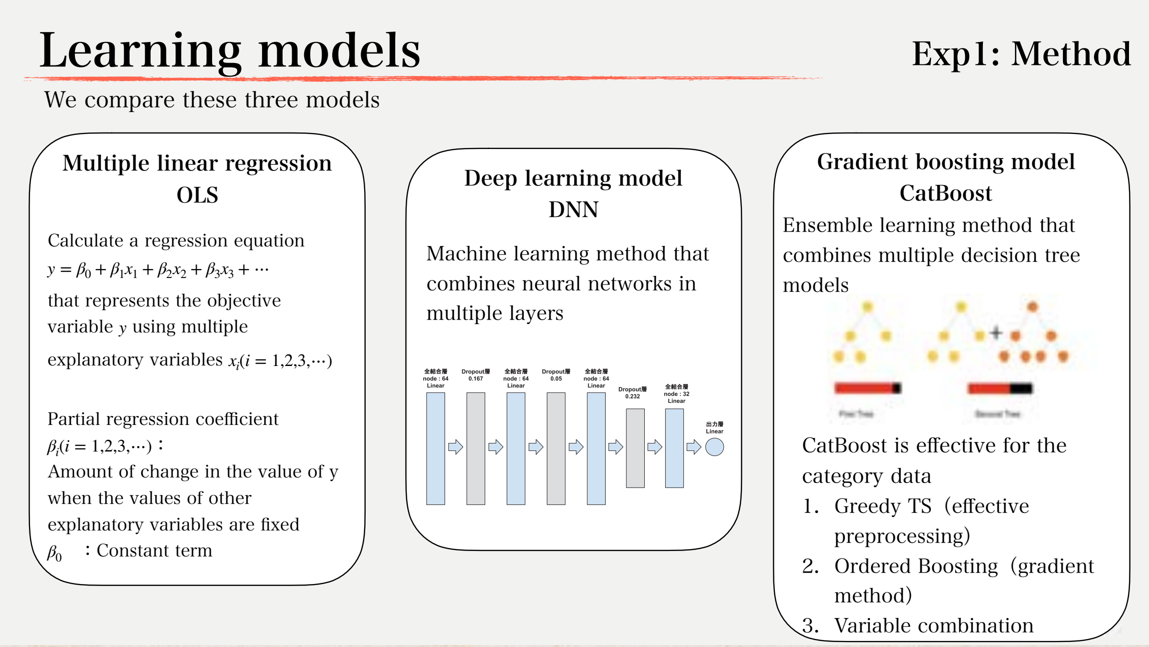
15 Bibliometric features

- Conference name
- Author's h-index
- Num of keywords
- Award

etc …

h-index:
There are h or more papers with at least h citations
→ The author's h-index is h

13



Learning models

Exp1: Method

We compare these three models

Multiple linear regression OLS

Calculate a regression equation

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots$

that represents the objective variable y using multiple explanatory variables $x_i (i = 1,2,3,\cdots)$

Partial regression coefficient $\beta_i (i = 1,2,3,\cdots)$：
Amount of change in the value of y when the values of other explanatory variables are fixed

β_0 ：Constant term

Deep learning model DNN

Machine learning method that combines neural networks in multiple layers

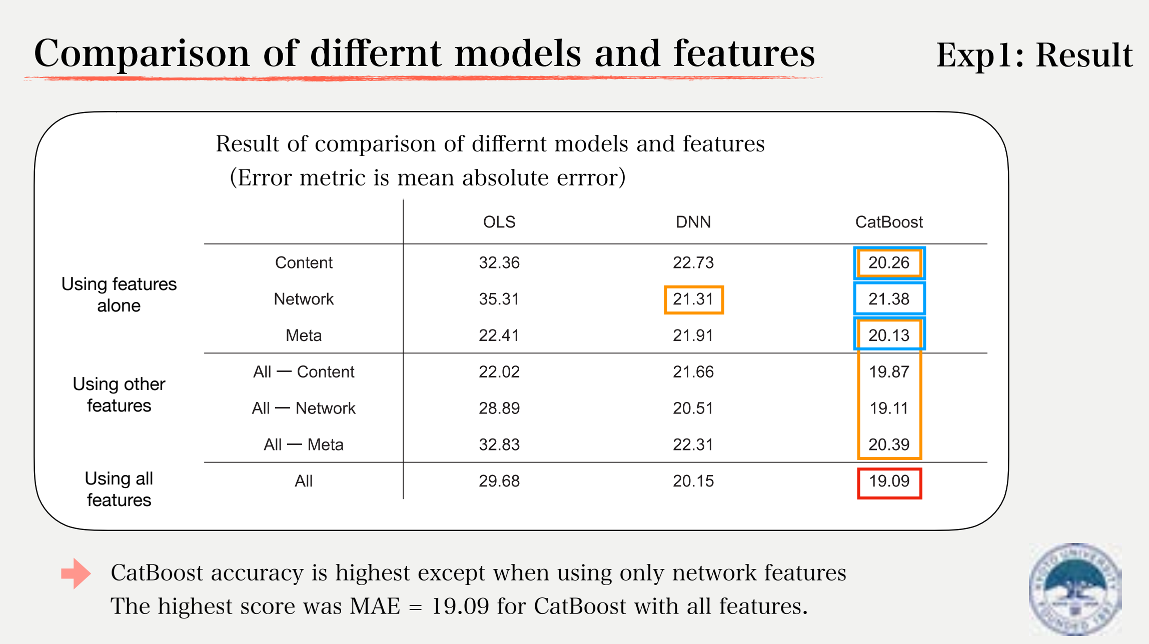
Gradient boosting model CatBoost

Ensemble learning method that combines multiple decision tree models

CatBoost is effective for the category data

1. Greedy TS（effective preprocessing）
2. Ordered Boosting（gradient method）
3. Variable combination

14



Comparison of diffelnt models and features

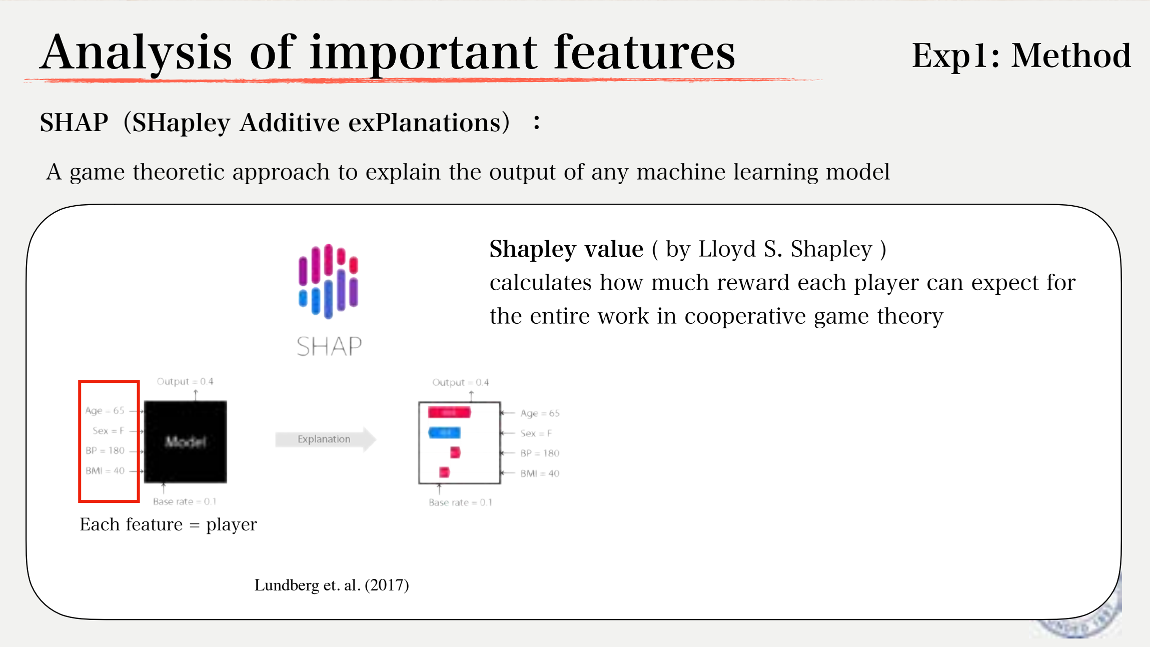
Exp1: Result

Result of comparison of differnt models and features
(Error metric is mean absolute errror)

		OLS	DNN	CatBoost
Using features alone	Content	32.36	22.73	20.26
	Network	35.31	21.31	21.38
	Meta	22.41	21.91	20.13
Using other features	All — Content	22.02	21.66	19.87
	All — Network	28.89	20.51	19.11
	All — Meta	32.83	22.31	20.39
Using all features	All	29.68	20.15	19.09

CatBoost accuracy is highest except when using only network features
The highest score was MAE = 19.09 for CatBoost with all features.

15



Analysis of important features

Exp1: Method

SHAP（SHapley Additive exPlanations）：

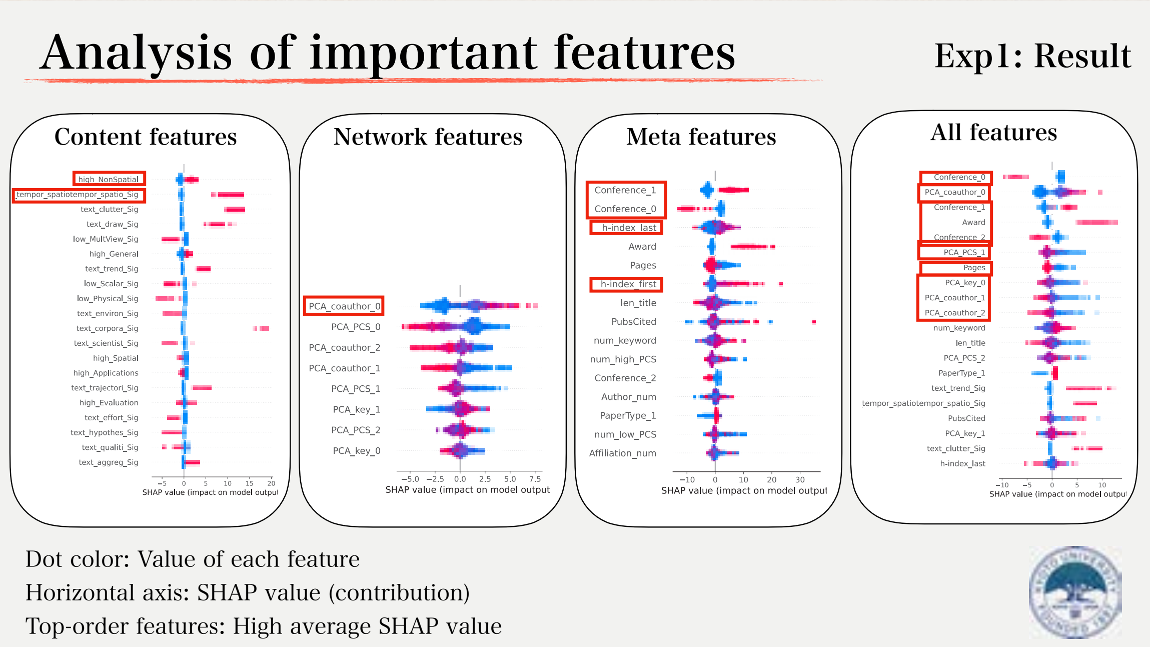
A game theoretic approach to explain the output of any machine learning model

Shapley value (by Lloyd S. Shapley)
calculates how much reward each player can expect for the entire work in cooperative game theory

Each feature = player

Lundberg et. al. (2017)

16

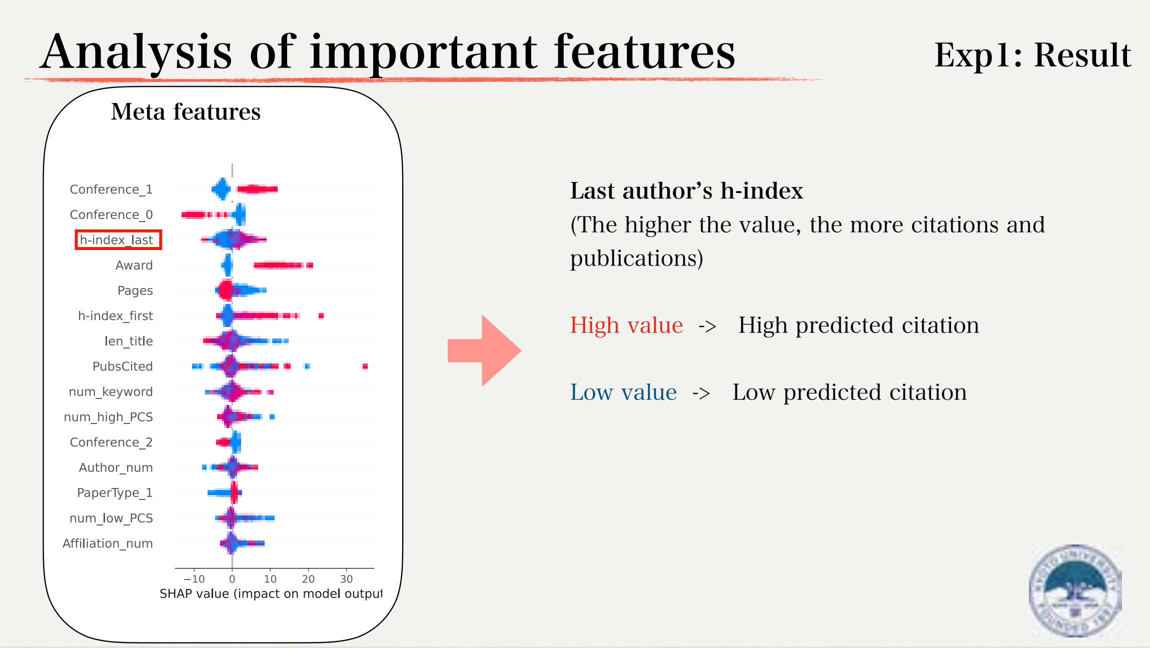


Analysis of important features

Exp1: Result

Dot color: Value of each feature
Horizontal axis: SHAP value (contribution)
Top-order features: High average SHAP value

17



Analysis of important features

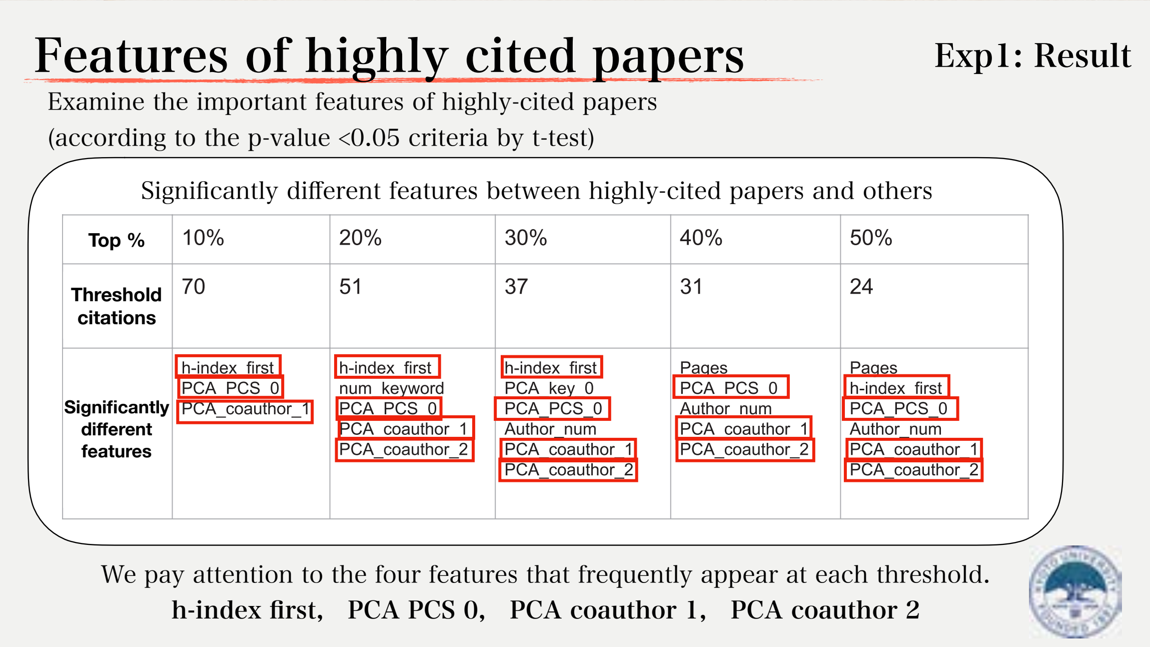
Exp1: Result

Last author’s h-index
(The higher the value, the more citations and publications)

High value -> High predicted citation

Low value -> Low predicted citation

18



Features of highly cited papers

Exp1: Result

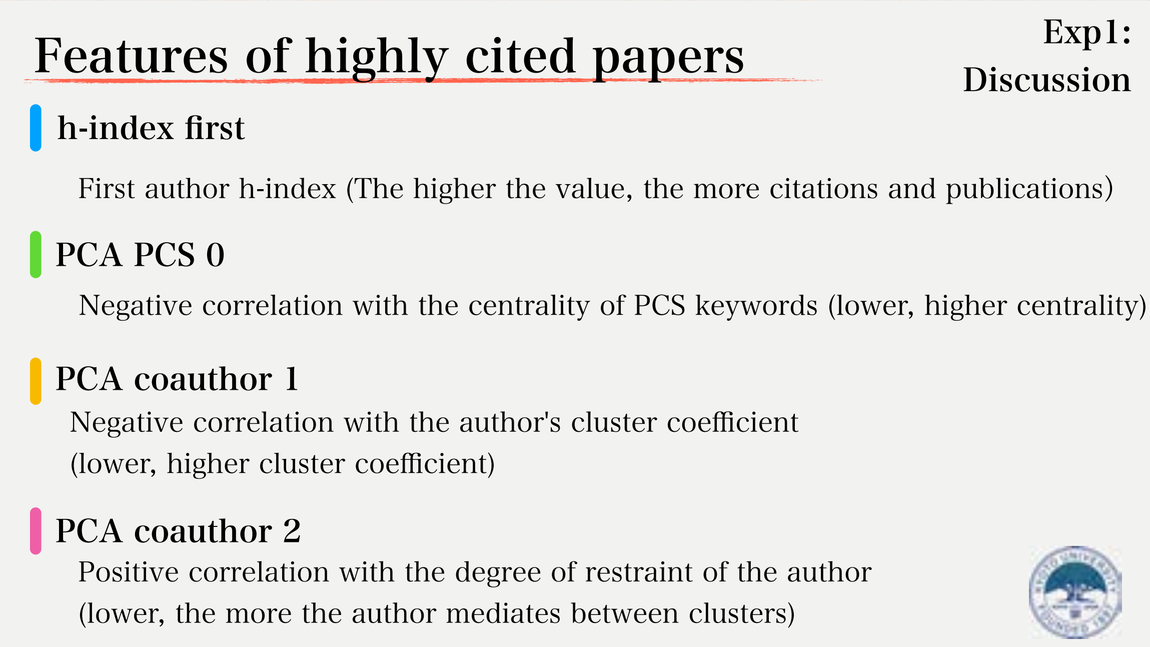
Examine the important features of highly-cited papers
(according to the p-value <0.05 criteria by t-test)

Significantly different features between highly-cited papers and others

Top %	10%	20%	30%	40%	50%
Threshold citations	70	51	37	31	24
Significantly different features	h-index_first PCA_PCS_0 PCA_coauthor_1	h-index_first num_keyword PCA_PCS_0 PCA_coauthor_1 PCA_coauthor_2	h-index_first PCA_key_0 PCA_PCS_0 Author_num PCA_coauthor_1 PCA_coauthor_2	Pages PCA_PCS_0 Author_num PCA_coauthor_1 PCA_coauthor_2	Pages h-index_first PCA_PCS_0 Author_num PCA_coauthor_1 PCA_coauthor_2

We pay attention to the four features that frequently appear at each threshold.
h-index first, PCA PCS 0, PCA coauthor 1, PCA coauthor 2

19



Features of highly cited papers

Exp1: Discussion

h-index first

First author h-index (The higher the value, the more citations and publications)

PCA PCS 0

Negative correlation with the centrality of PCS keywords (lower, higher centrality)

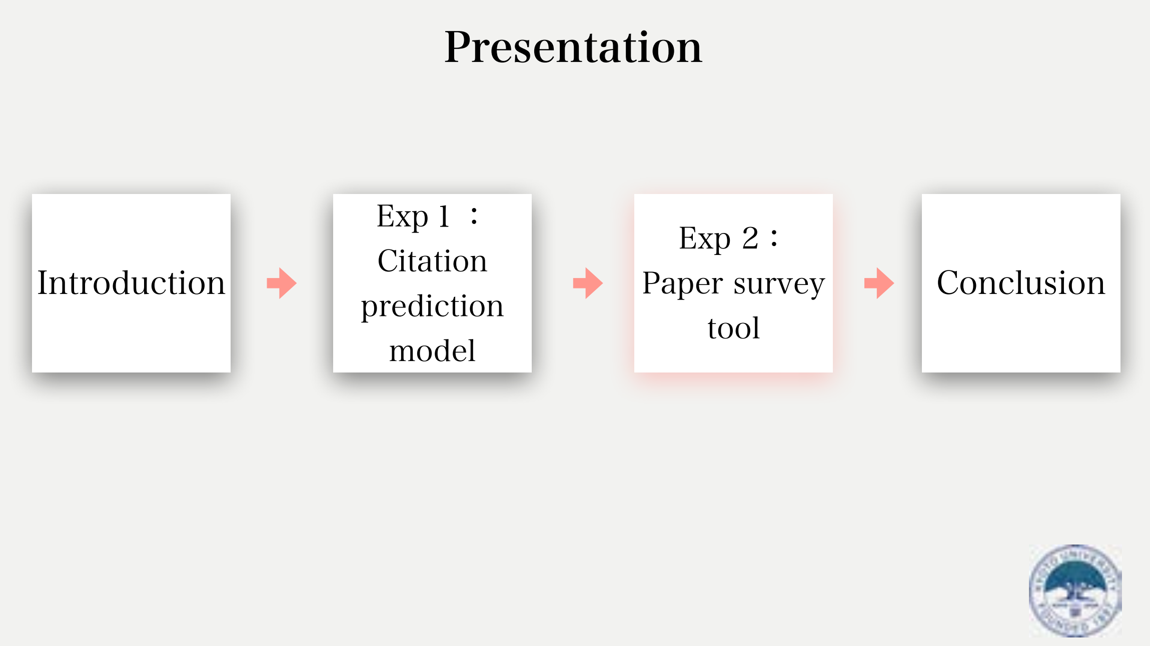
PCA coauthor 1

Negative correlation with the author's cluster coefficient
(lower, higher cluster coefficient)

PCA coauthor 2

Positive correlation with the degree of restraint of the author
(lower, the more the author mediates between clusters)

20



Presentation

Introduction → Exp 1 ： Citation prediction model → Exp 2 : Paper survey tool → Conclusion

21

Exp 2 : Paper survey tool utilizing citation count prediction

Exp2

Development of paper survey tool that allows you to explore papers based on predicted citations

Evaluation of the effectiveness of the predicted number of citations in the survey

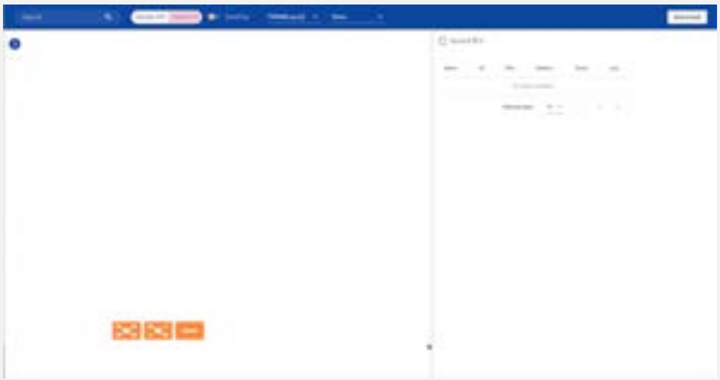
We compare the features of papers when users survey

- based solely on citation counts
- based on predicted citation counts

22

Demonstration of our paper survey tool

Exp2: Method



Web application we developed
http://35.221.115.40/

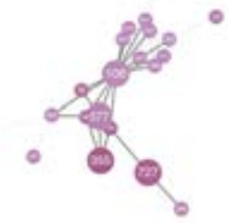
Back end：Python
Front end：JavaScript
Graph data generation：networkx
Coordinate calculation：EgRenderer

Node: Papers
Links: Authors / Keywords / Citations

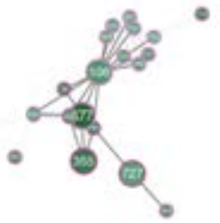
23

Examples of possible strategies

Exp2: Method



Nodes colored by Citation



Nodes colored by Score + Citation

Strategy①： Nodes' color

		Predicted citation counts（Score）	
		High (Dark green)	Low (Light green)
Citation counts (Citation)	High (Dark purple)	Has important features and high citation count (355)	More citations than expected（727）
	Low (Light purple)	Low citation counts but has important features (677)	Does not have the feature of increasing citation counts (106)

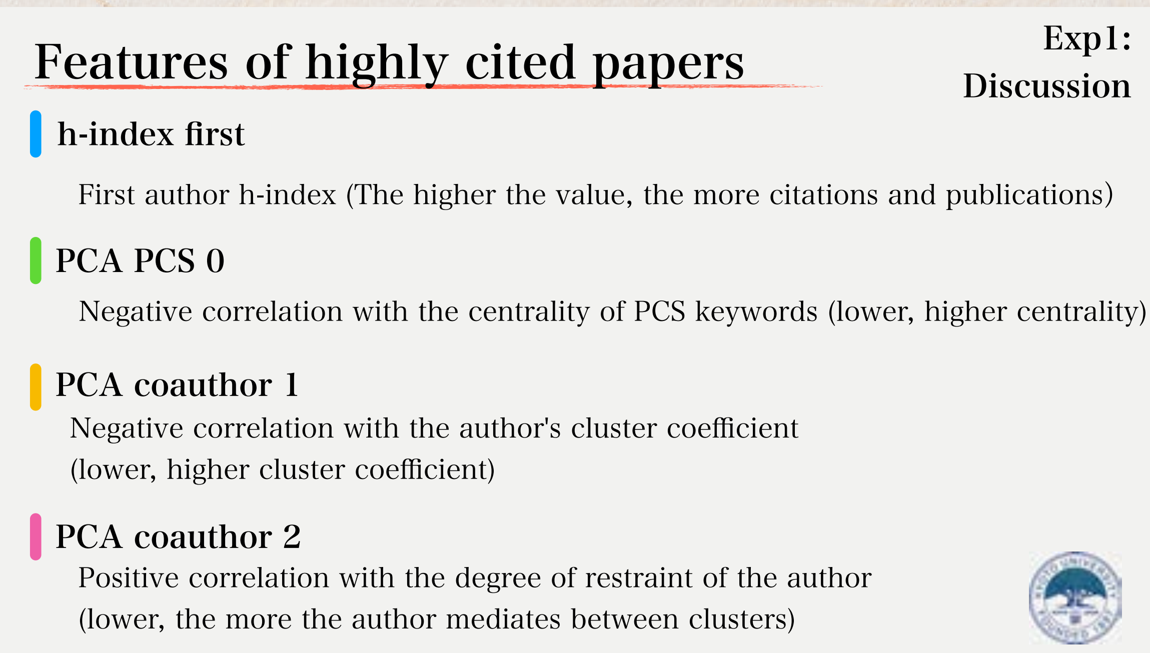
Strategy②：Network exploration perspective

Central or mediating between clusters in the network

Strategy③：Own interest

➡ When users survey papers by freely combining these strategies, how do the user's attention change with or without predicted citation counts ?

24



Features of highly cited papers

Exp1: Discussion

- h-index first

  First author h-index (The higher the value, the more citations and publications)

- PCA PCS 0

  Negative correlation with the centrality of PCS keywords (lower, higher centrality)

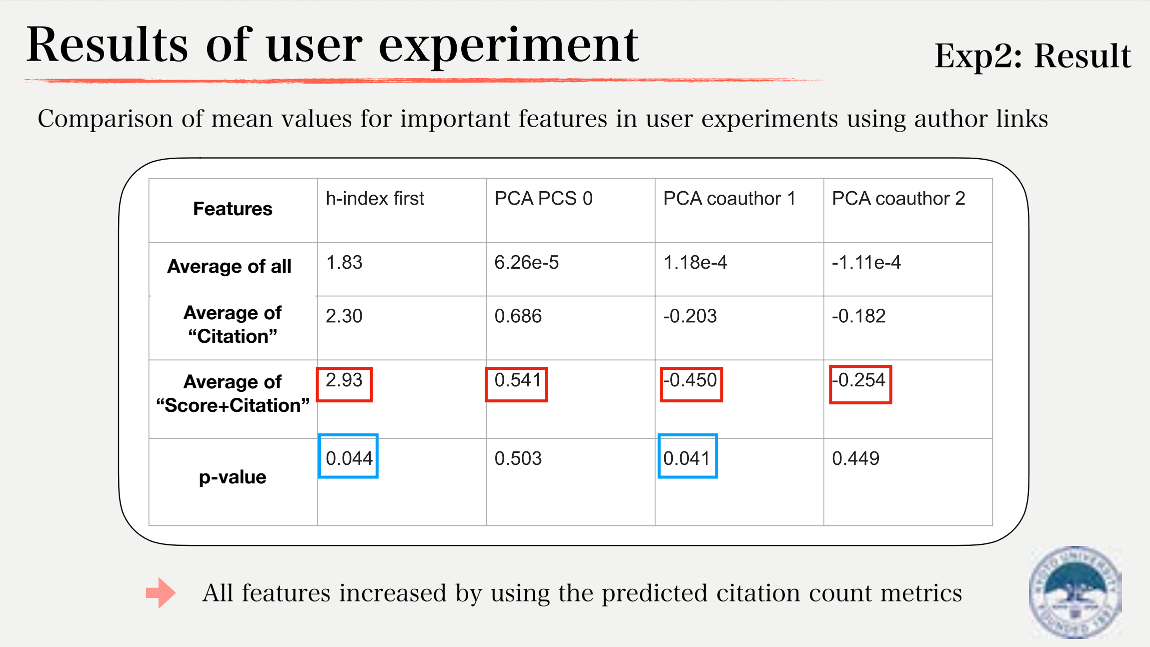
- PCA coauthor 1

  Negative correlation with the author's cluster coefficient (lower, higher cluster coefficient)

- PCA coauthor 2

  Positive correlation with the degree of restraint of the author (lower, the more the author mediates between clusters)

25



Results of user experiment

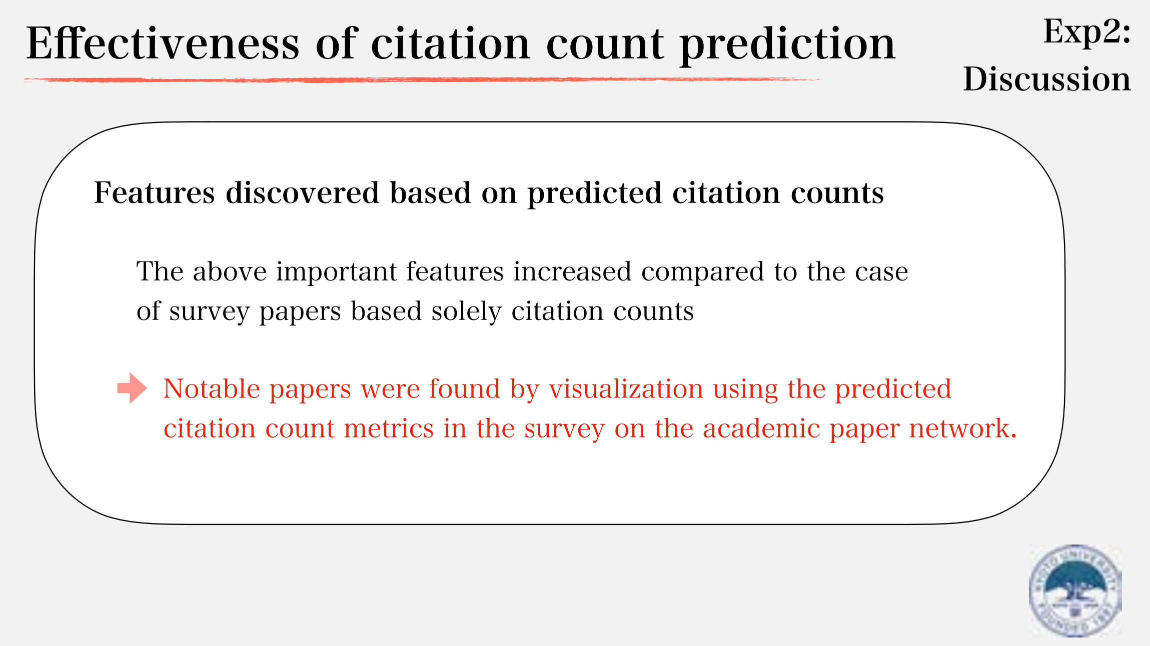
Exp2: Result

Comparison of mean values for important features in user experiments using author links

Features	h-index first	PCA PCS 0	PCA coauthor 1	PCA coauthor 2
Average of all	1.83	6.26e-5	1.18e-4	-1.11e-4
Average of “Citation”	2.30	0.686	-0.203	-0.182
Average of “Score+Citation”	2.93	0.541	-0.450	-0.254
p-value	0.044	0.503	0.041	0.449

→ All features increased by using the predicted citation count metrics

26



Effectiveness of citation count prediction

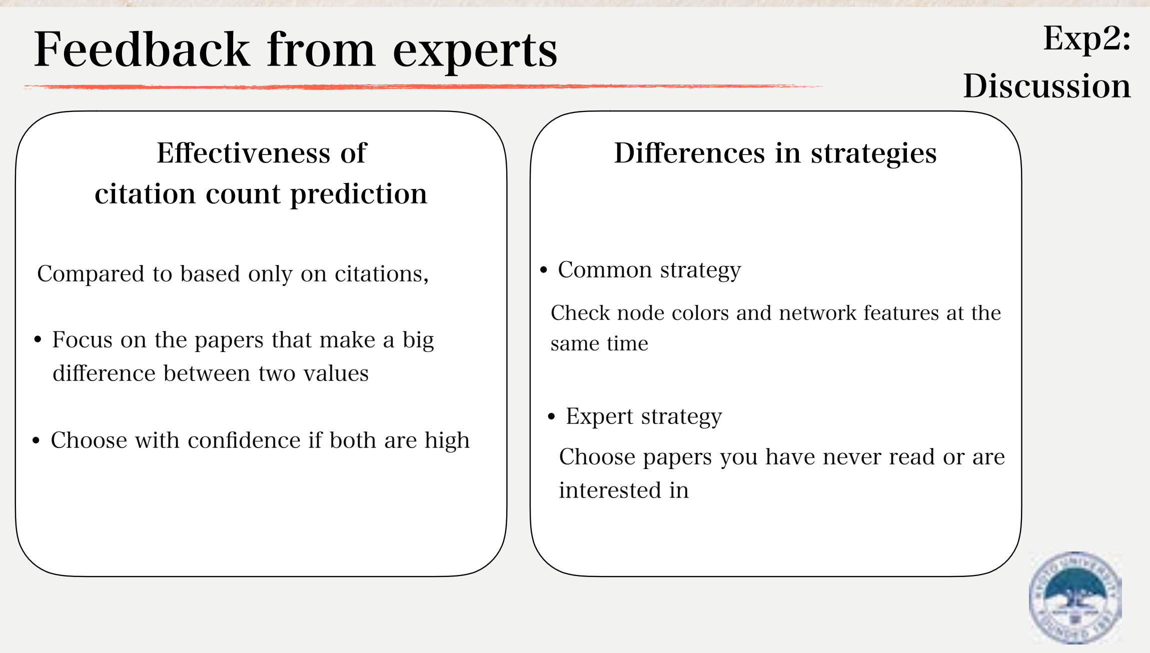
Exp2: Discussion

Features discovered based on predicted citation counts

The above important features increased compared to the case of survey papers based solely citation counts

→ Notable papers were found by visualization using the predicted citation count metrics in the survey on the academic paper network.

27



Feedback from experts

Exp2:
Discussion

Effectiveness of citation count prediction

Compared to based only on citations,

- Focus on the papers that make a big difference between two values
- Choose with confidence if both are high

Differences in strategies

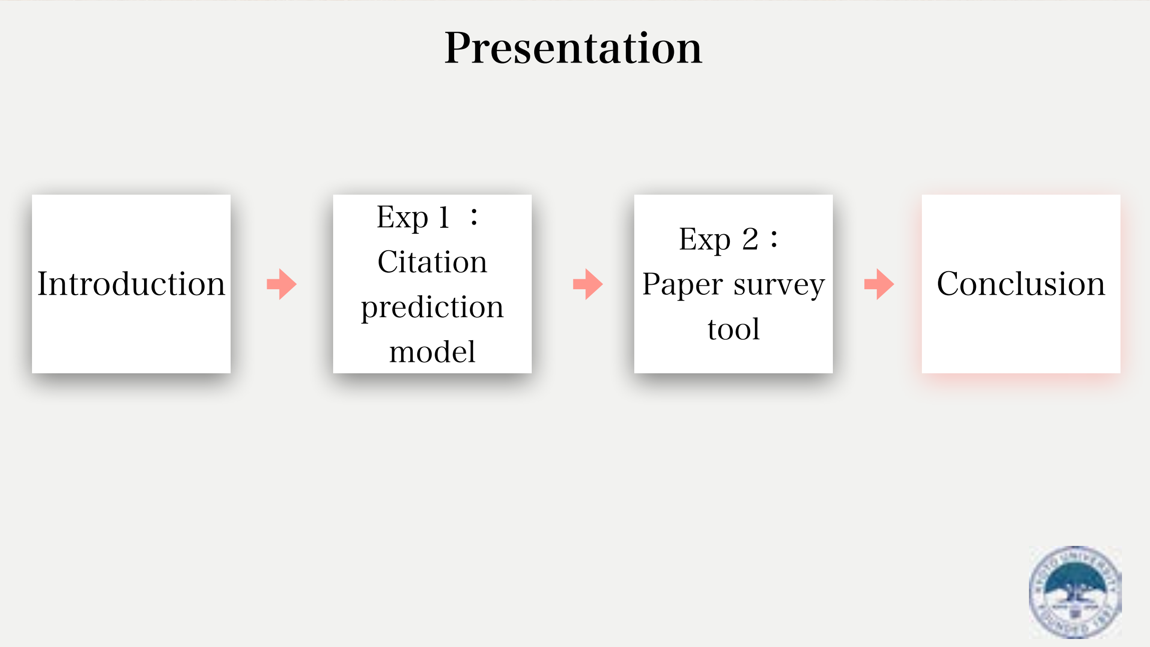
- Common strategy

Check node colors and network features at the same time

- Expert strategy

Choose papers you have never read or are interested in

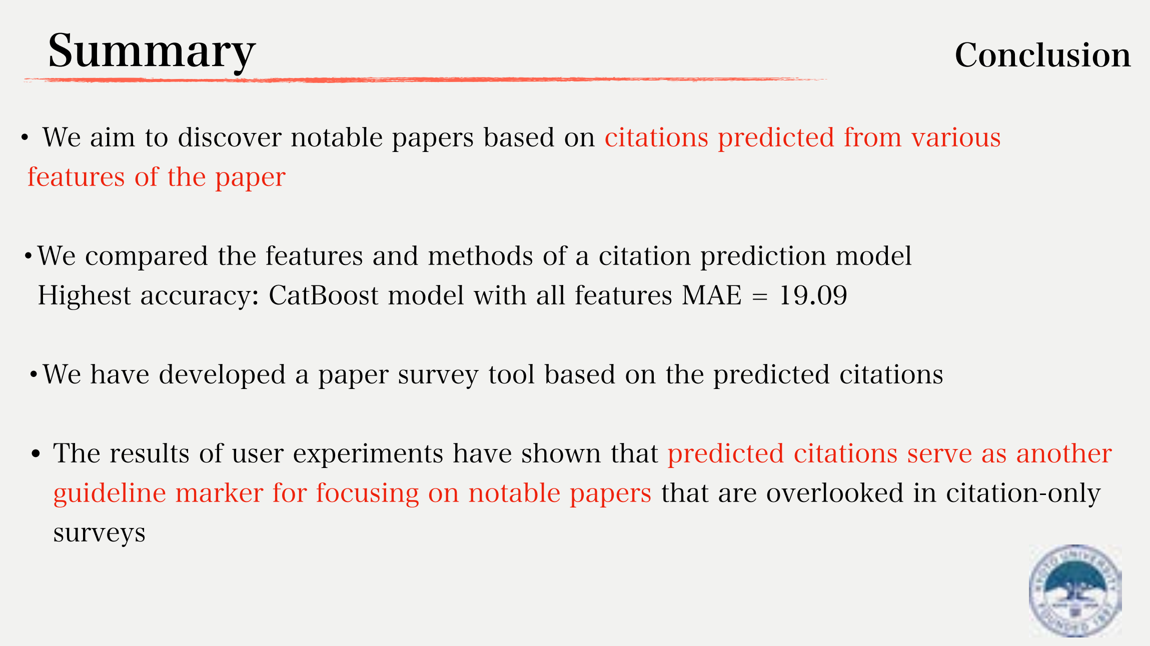
28



Presentation

Introduction → Exp 1 ： Citation prediction model → Exp 2 : Paper survey tool → Conclusion

29



Summary

Conclusion

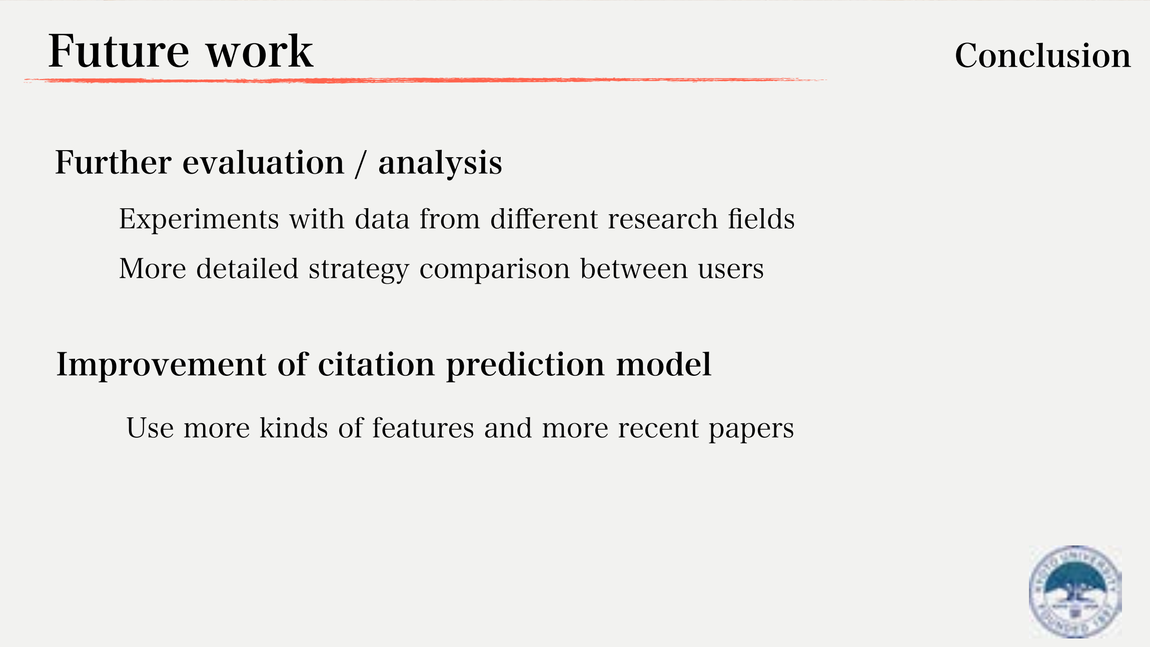
- We aim to discover notable papers based on citations predicted from various features of the paper

- We compared the features and methods of a citation prediction model
  Highest accuracy: CatBoost model with all features MAE = 19.09

- We have developed a paper survey tool based on the predicted citations

- The results of user experiments have shown that predicted citations serve as another guideline marker for focusing on notable papers that are overlooked in citation-only surveys

30



Future work

Conclusion

Further evaluation / analysis

Experiments with data from different research fields

More detailed strategy comparison between users

Improvement of citation prediction model

Use more kinds of features and more recent papers

31

Conclusion

Thank you

Presentation 4｜報告 4

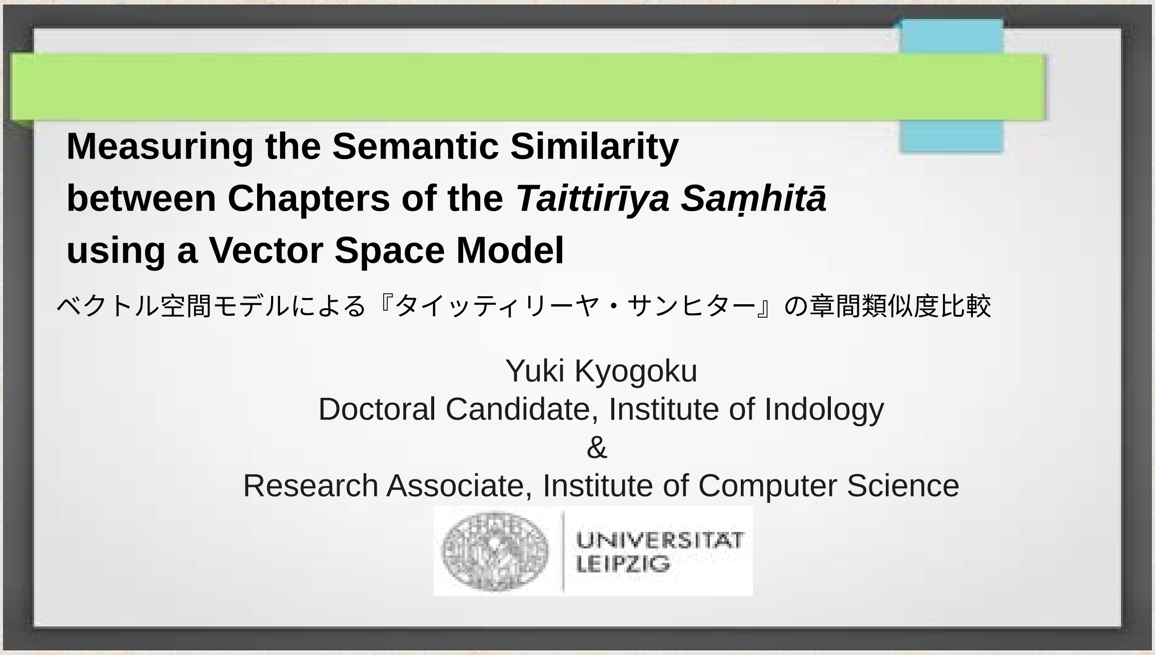
Measuring the Semantic Similarity between the Chapters of Taittirīya Saṃhitā Using a Vector Space Model

ベクトル空間モデルによる『タイッティリーヤ・サンヒター』の章間類似度比較

京極祐希（Leipzig University, Indology）

Yuki Kyogoku

1



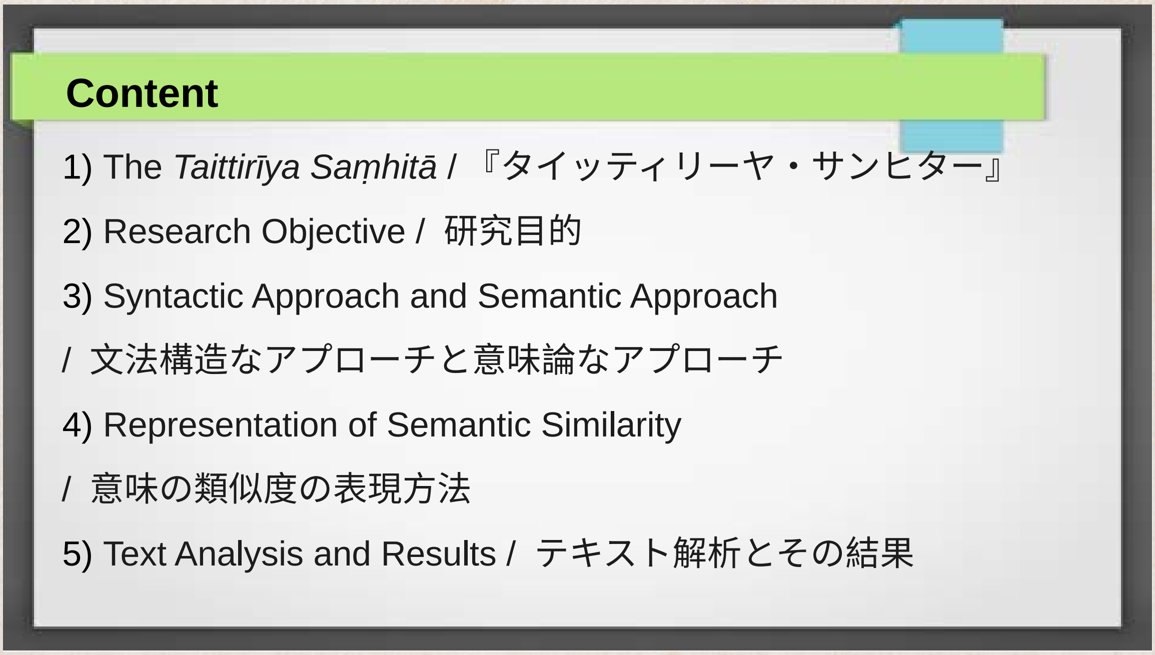
Measuring the Semantic Similarity
between Chapters of the *Taittirīya Saṃhitā*
using a Vector Space Model

ベクトル空間モデルによる『タイッティリーヤ・サンヒター』の章間類似度比較

Yuki Kyogoku
Doctoral Candidate, Institute of Indology
&
Research Associate, Institute of Computer Science

UNIVERSITAT LEIPZIG

2



Content

1) The *Taittirīya Saṃhitā* / 『タイッティリーヤ・サンヒター』

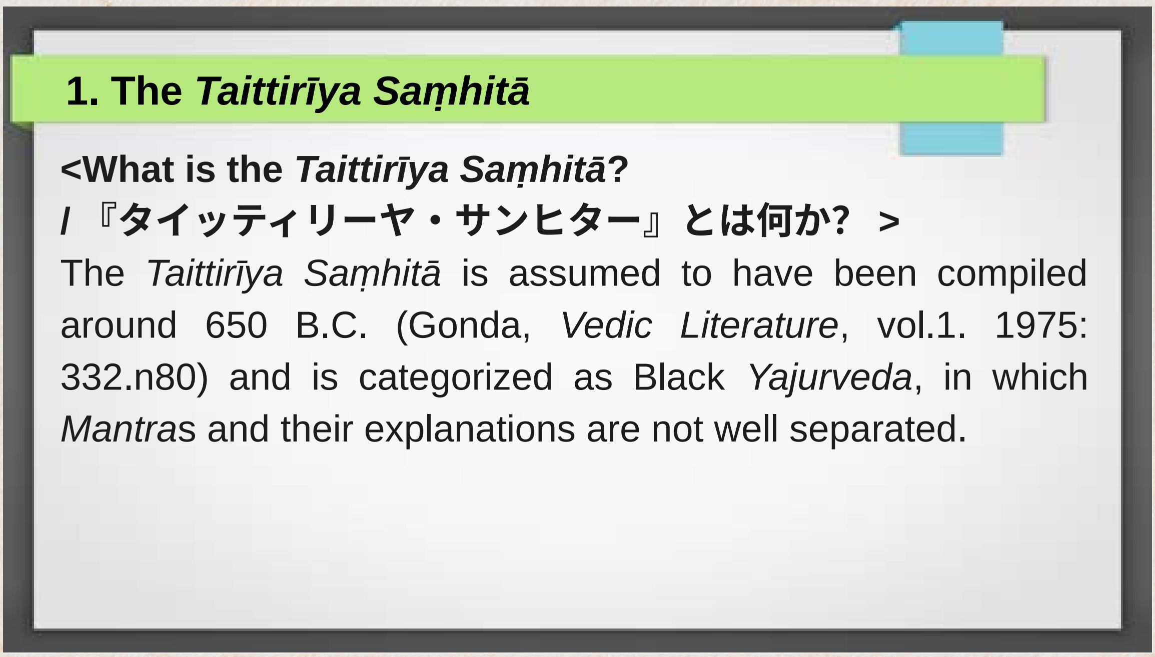
2) Research Objective / 研究目的

3) Syntactic Approach and Semantic Approach
/ 文法構造なアプローチと意味論なアプローチ

4) Representation of Semantic Similarity
/ 意味の類似度の表現方法

5) Text Analysis and Results / テキスト解析とその結果

3

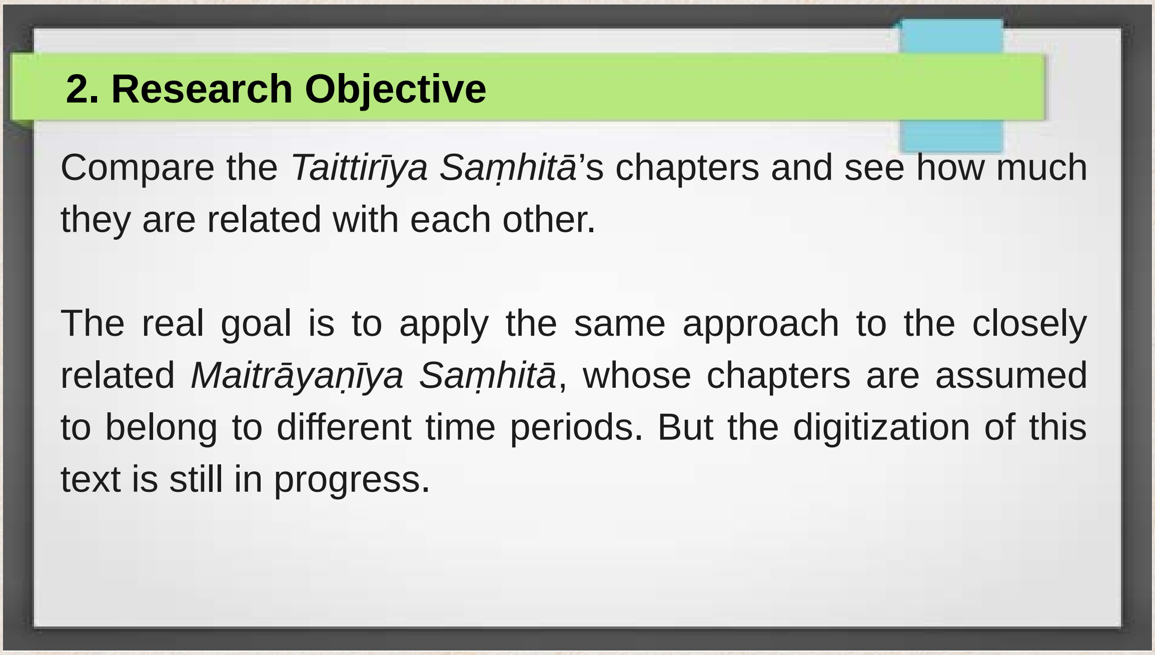


1. The *Taittirīya Saṃhitā*

<What is the *Taittirīya Saṃhitā*?
/『タイッティリーヤ・サンヒター』とは何か? >

The *Taittirīya Saṃhitā* is assumed to have been compiled around 650 B.C. (Gonda, *Vedic Literature*, vol.1. 1975: 332.n80) and is categorized as Black *Yajurveda*, in which *Mantras* and their explanations are not well separated.

4

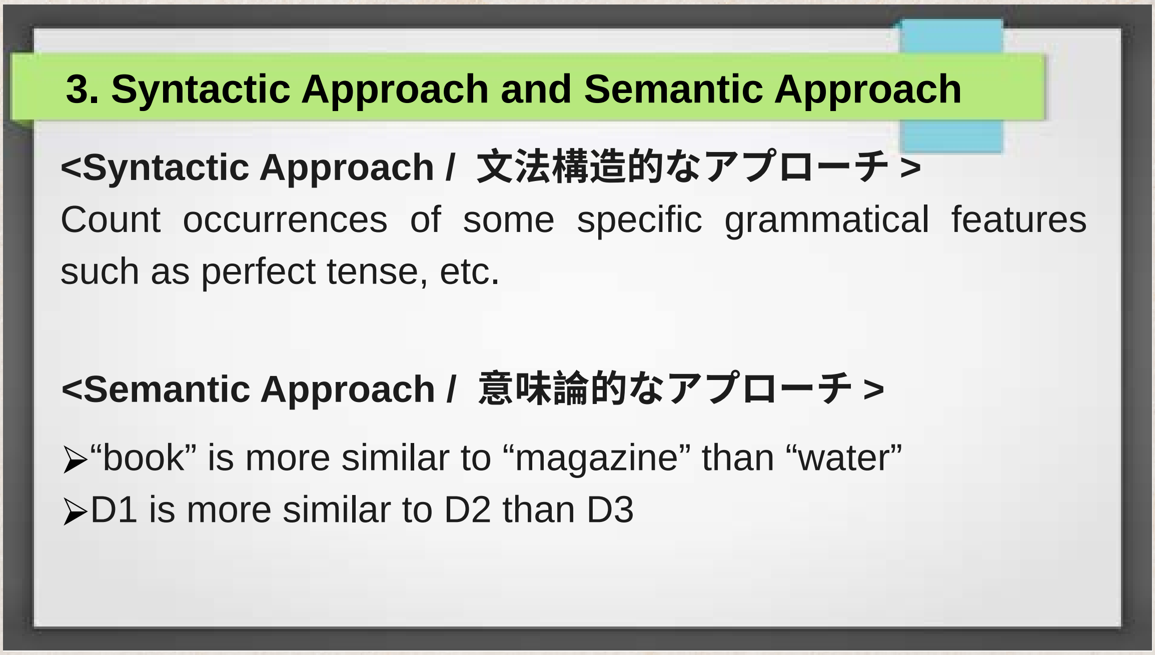


2. Research Objective

Compare the *Taittirīya Saṃhitā*'s chapters and see how much they are related with each other.

The real goal is to apply the same approach to the closely related *Maitrāyaṇīya Saṃhitā*, whose chapters are assumed to belong to different time periods. But the digitization of this text is still in progress.

5



3. Syntactic Approach and Semantic Approach

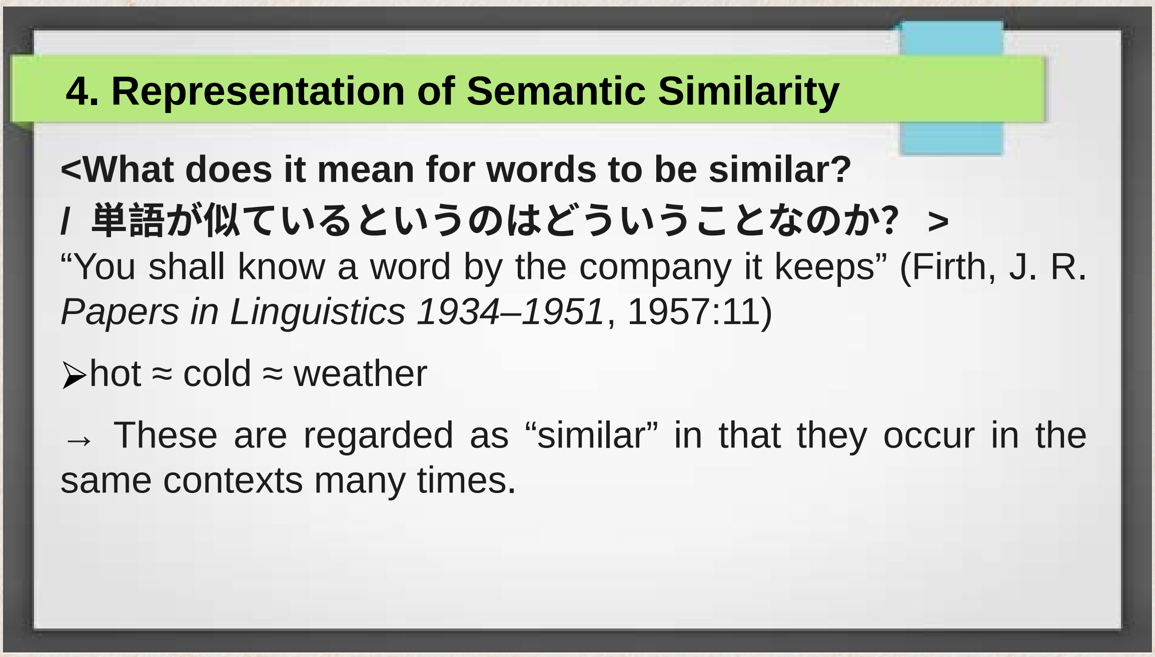
<Syntactic Approach / 文法構造的なアプローチ >

Count occurrences of some specific grammatical features such as perfect tense, etc.

<Semantic Approach / 意味論的なアプローチ >

- "book" is more similar to "magazine" than "water"
- D1 is more similar to D2 than D3

6



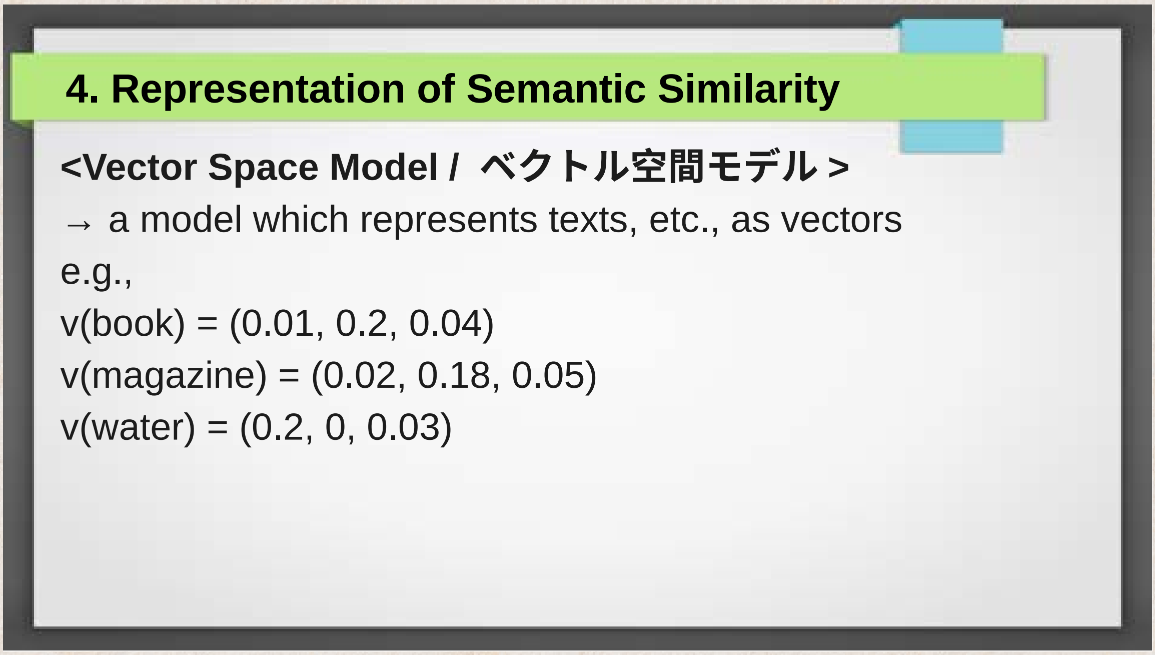
4. Representation of Semantic Similarity

<What does it mean for words to be similar?
/ 単語が似ているというのはどういうことなのか? >
“You shall know a word by the company it keeps” (Firth, J. R. *Papers in Linguistics 1934–1951*, 1957:11)

➢hot ≈ cold ≈ weather

→ These are regarded as “similar” in that they occur in the same contexts many times.

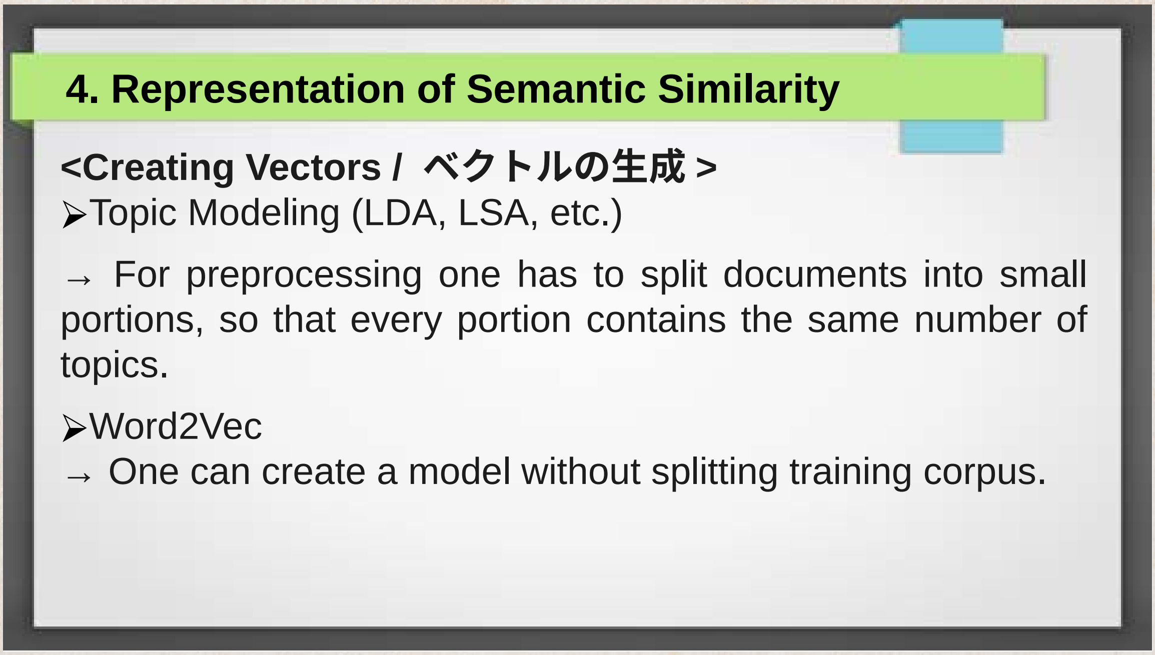
7



4. Representation of Semantic Similarity

<Vector Space Model / ベクトル空間モデル >
→ a model which represents texts, etc., as vectors
e.g.,
v(book) = (0.01, 0.2, 0.04)
v(magazine) = (0.02, 0.18, 0.05)
v(water) = (0.2, 0, 0.03)

8



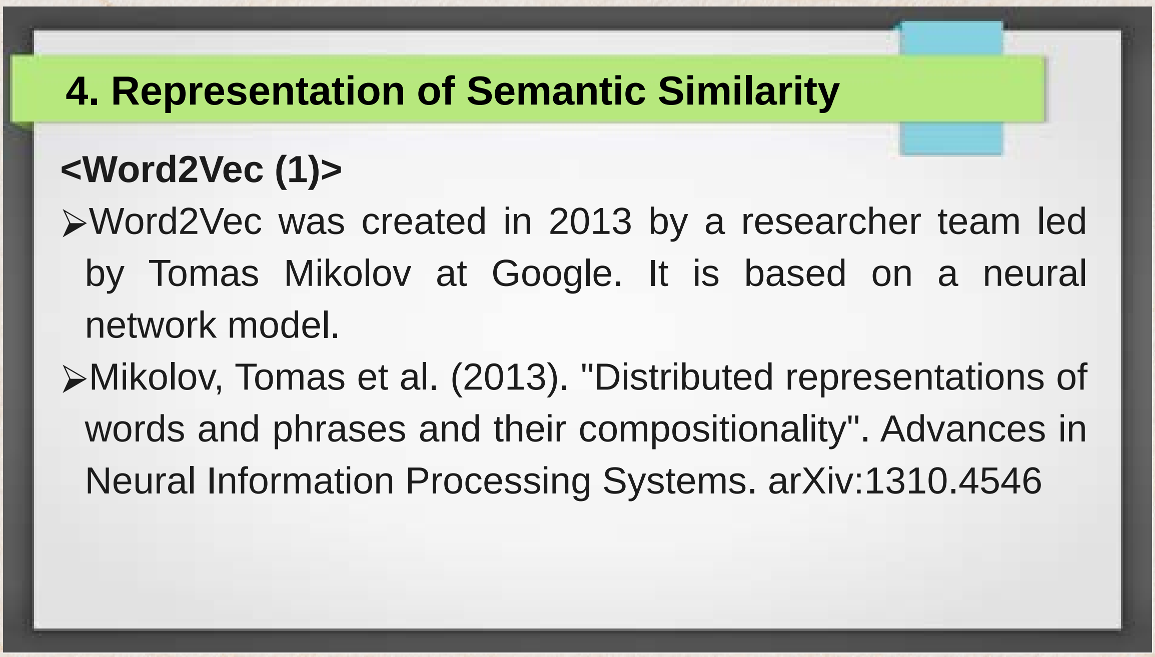
4. Representation of Semantic Similarity

<Creating Vectors / ベクトルの生成 >
➢Topic Modeling (LDA, LSA, etc.)

→ For preprocessing one has to split documents into small portions, so that every portion contains the same number of topics.

➢Word2Vec
→ One can create a model without splitting training corpus.

9



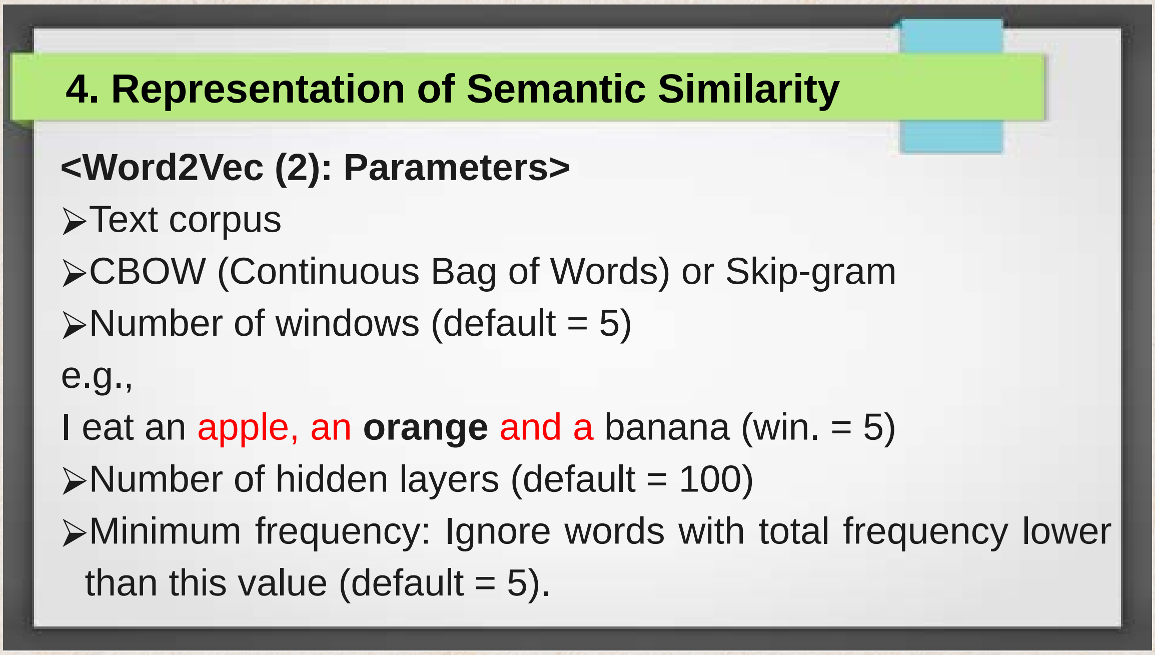
4. Representation of Semantic Similarity

<Word2Vec (1)>

➢Word2Vec was created in 2013 by a researcher team led by Tomas Mikolov at Google. It is based on a neural network model.

➢Mikolov, Tomas et al. (2013). "Distributed representations of words and phrases and their compositionality". Advances in Neural Information Processing Systems. arXiv:1310.4546

10



4. Representation of Semantic Similarity

<Word2Vec (2): Parameters>

➢Text corpus

➢CBOW (Continuous Bag of Words) or Skip-gram

➢Number of windows (default = 5)

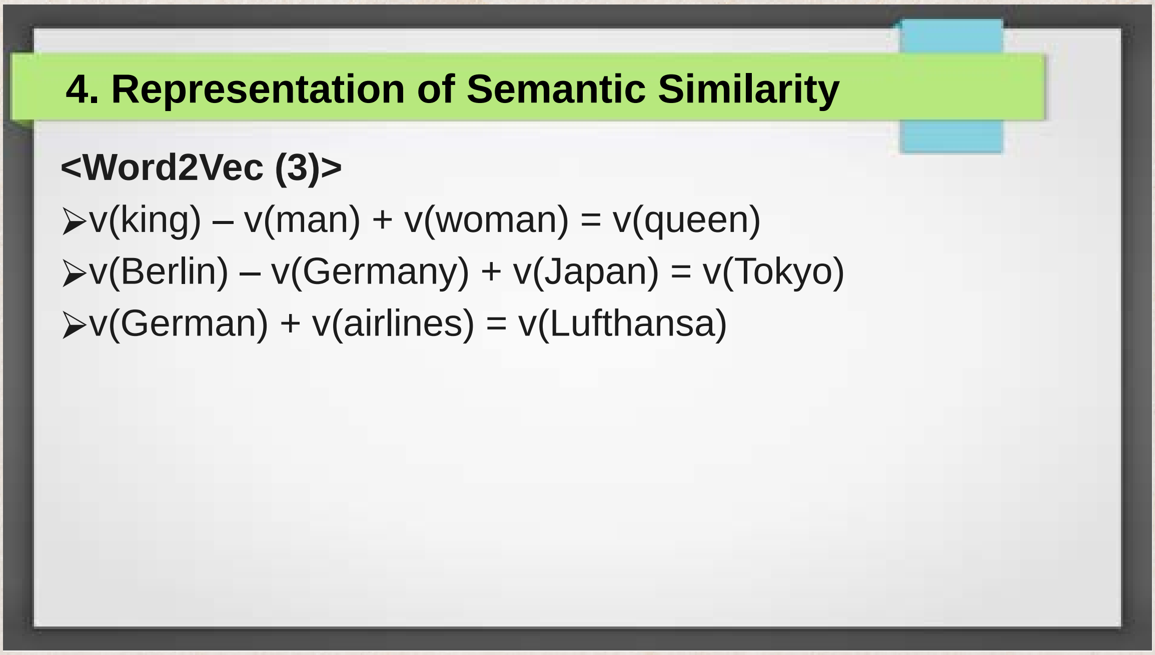
e.g.,

I eat an apple, an **orange** and a banana (win. = 5)

➢Number of hidden layers (default = 100)

➢Minimum frequency: Ignore words with total frequency lower than this value (default = 5).

11



4. Representation of Semantic Similarity

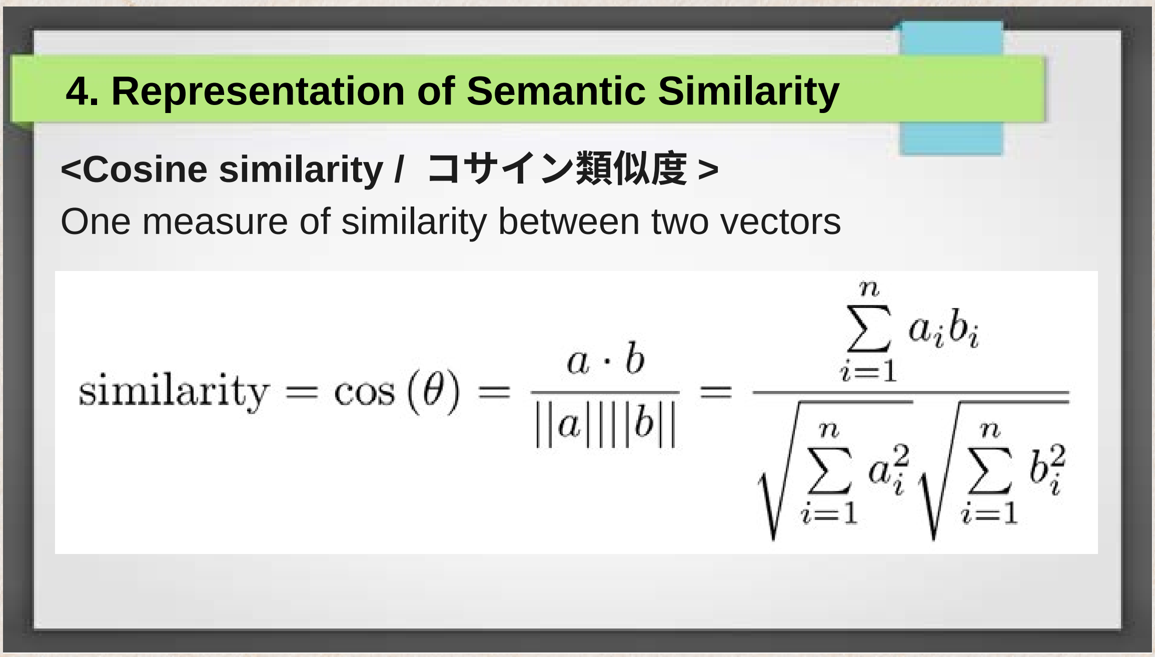
<Word2Vec (3)>

➢v(king) – v(man) + v(woman) = v(queen)

➢v(Berlin) – v(Germany) + v(Japan) = v(Tokyo)

➢v(German) + v(airlines) = v(Lufthansa)

12



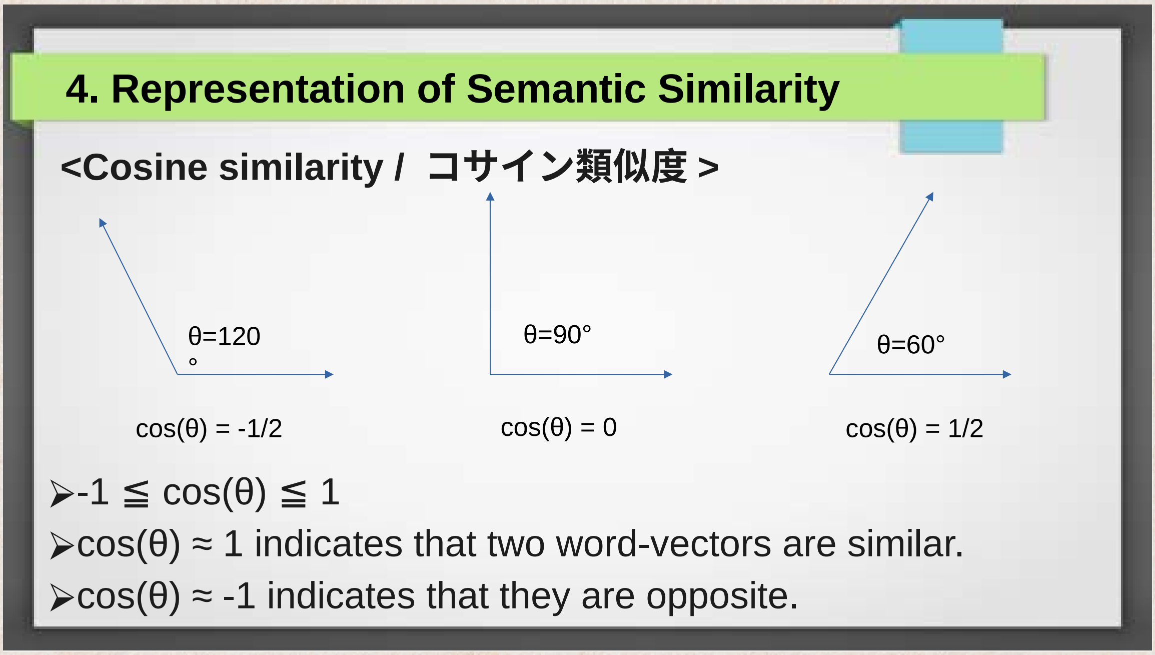
4. Representation of Semantic Similarity

<Cosine similarity / コサイン類似度 >

One measure of similarity between two vectors

$$\text{similarity} = \cos(\theta) = \frac{a \cdot b}{||a||||b||} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{\sum_{i=1}^{n} a_i^2}\sqrt{\sum_{i=1}^{n} b_i^2}}$$

13



4. Representation of Semantic Similarity

<Cosine similarity / コサイン類似度 >

θ=120 °

cos(θ) = -1/2

θ=90°

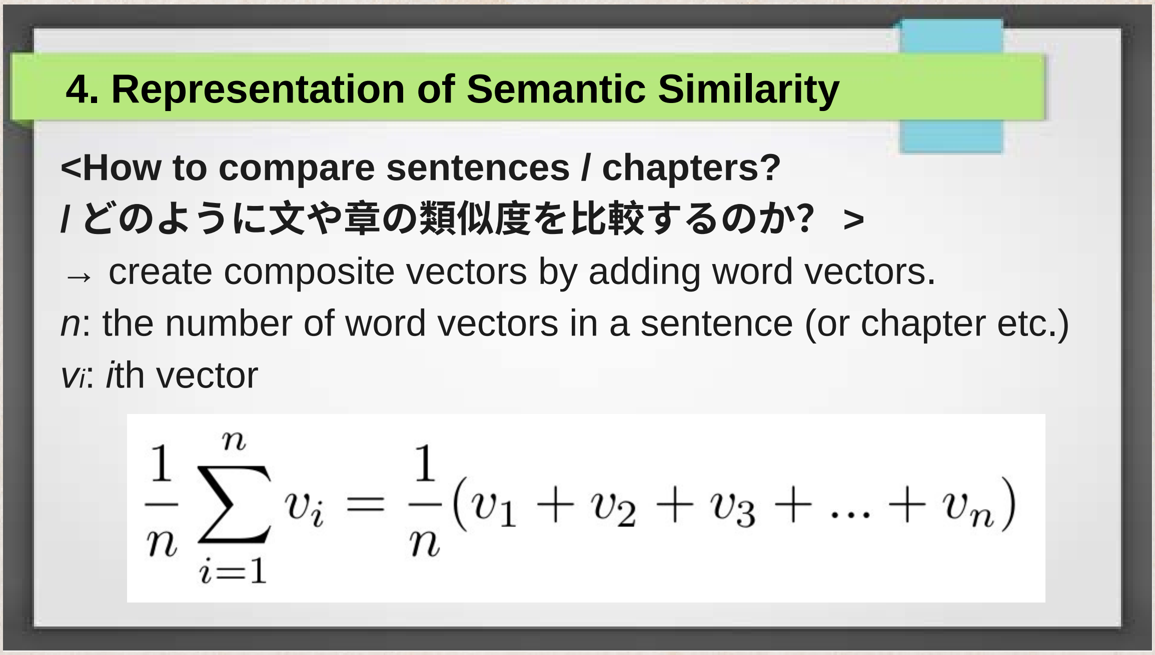
cos(θ) = 0

θ=60°

cos(θ) = 1/2

- -1 ≦ cos(θ) ≦ 1
- cos(θ) ≈ 1 indicates that two word-vectors are similar.
- cos(θ) ≈ -1 indicates that they are opposite.

14



4. Representation of Semantic Similarity

<How to compare sentences / chapters?
/ どのように文や章の類似度を比較するのか? >

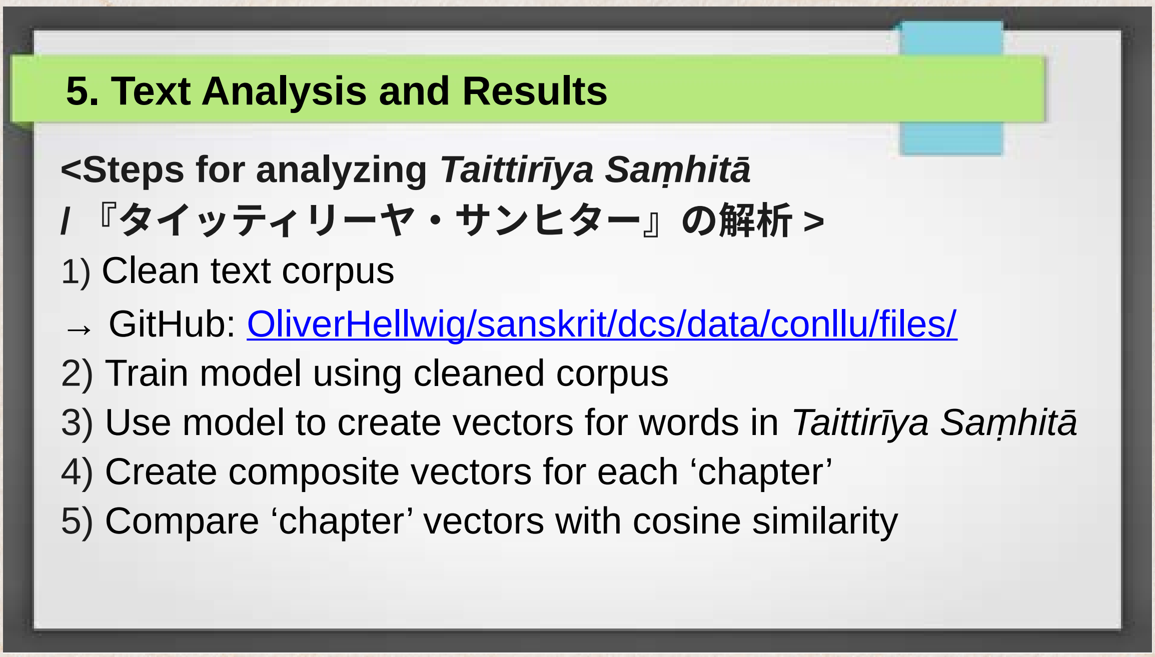
→ create composite vectors by adding word vectors.

n: the number of word vectors in a sentence (or chapter etc.)

v_i: *i*th vector

$$\frac{1}{n}\sum_{i=1}^{n} v_i = \frac{1}{n}(v_1 + v_2 + v_3 + \ldots + v_n)$$

15

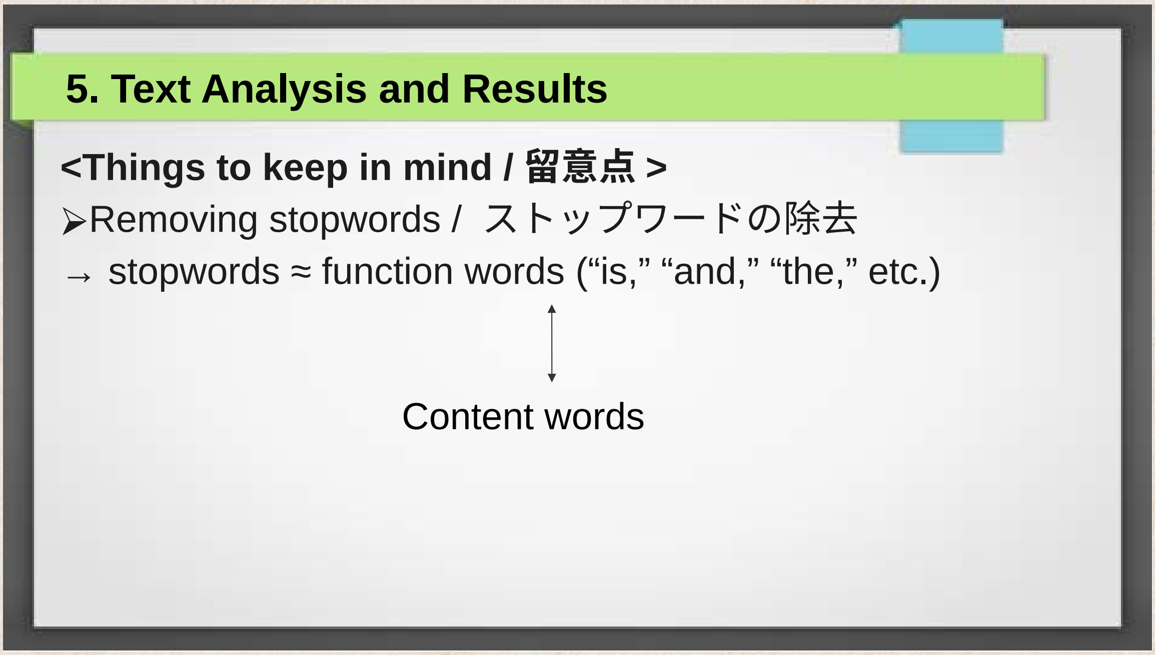


5. Text Analysis and Results

<Steps for analyzing *Taittirīya Saṃhitā*
/ 『タイッティリーヤ・サンヒター』の解析 >

1) Clean text corpus
→ GitHub: OliverHellwig/sanskrit/dcs/data/conllu/files/
2) Train model using cleaned corpus
3) Use model to create vectors for words in *Taittirīya Saṃhitā*
4) Create composite vectors for each ‘chapter’
5) Compare ‘chapter’ vectors with cosine similarity

16



5. Text Analysis and Results

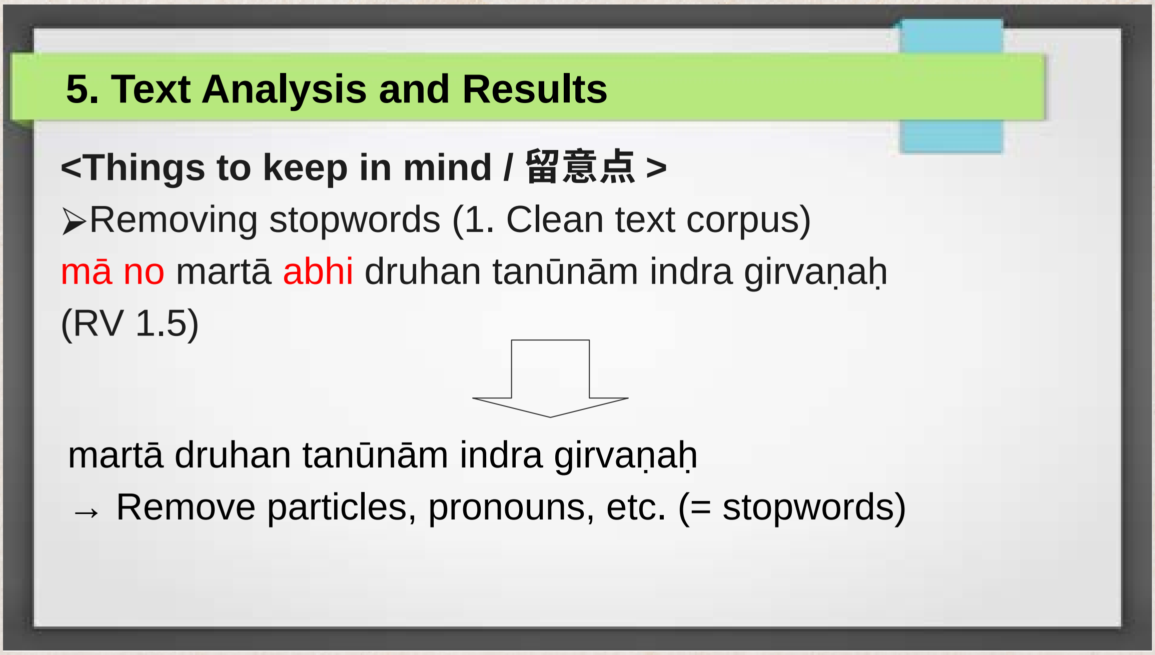
<Things to keep in mind / 留意点 >

➢Removing stopwords / ストップワードの除去
→ stopwords ≈ function words (“is,” “and,” “the,” etc.)

↕

Content words

17



5. Text Analysis and Results

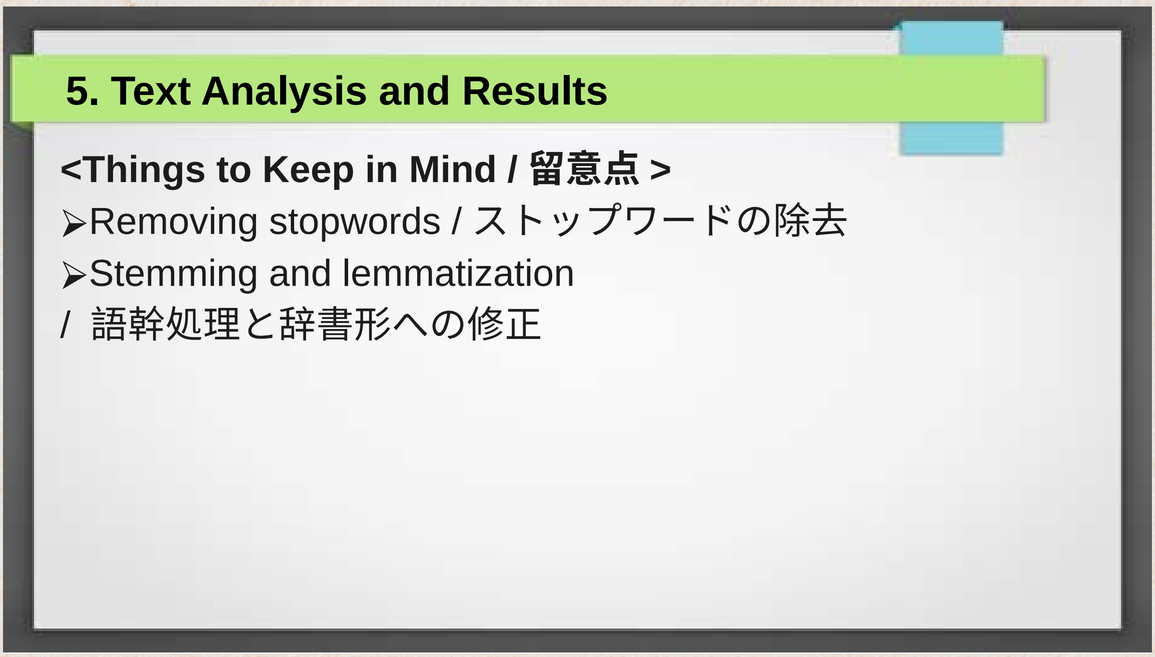
<Things to keep in mind / 留意点 >

➢Removing stopwords (1. Clean text corpus)
mā no martā abhi druhan tanūnām indra girvaṇaḥ
(RV 1.5)

⇩

martā druhan tanūnām indra girvaṇaḥ
→ Remove particles, pronouns, etc. (= stopwords)

18

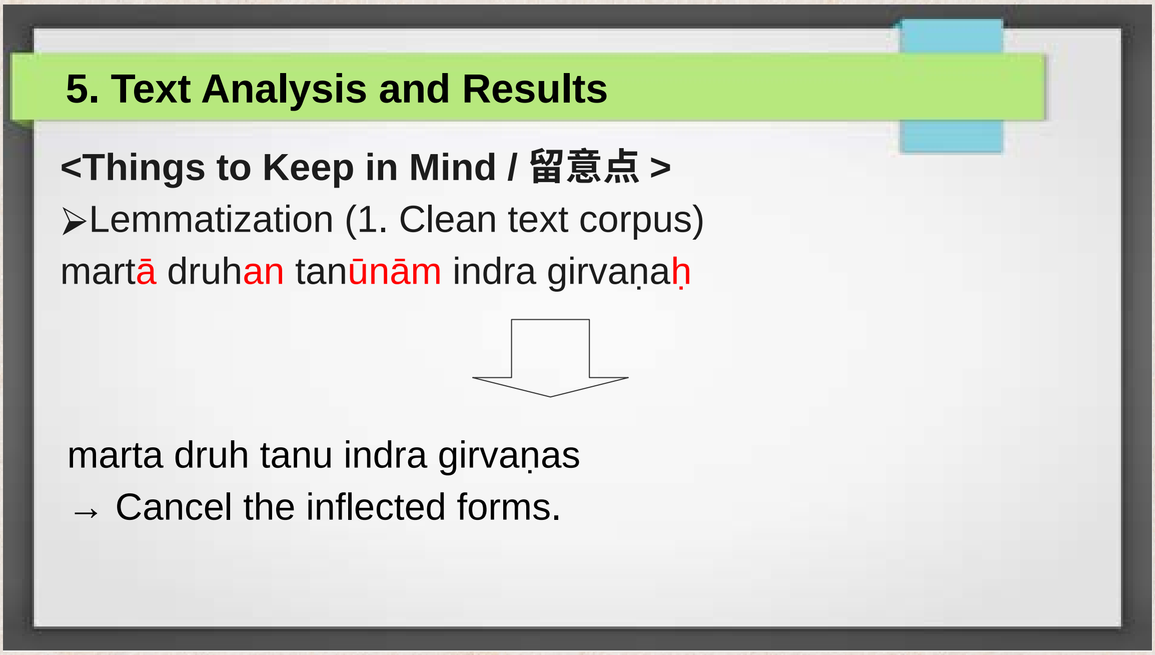


5. Text Analysis and Results

<Things to Keep in Mind / 留意点 >

- ➢Removing stopwords / ストップワードの除去
- ➢Stemming and lemmatization
/ 語幹処理と辞書形への修正

19



5. Text Analysis and Results

<Things to Keep in Mind / 留意点 >

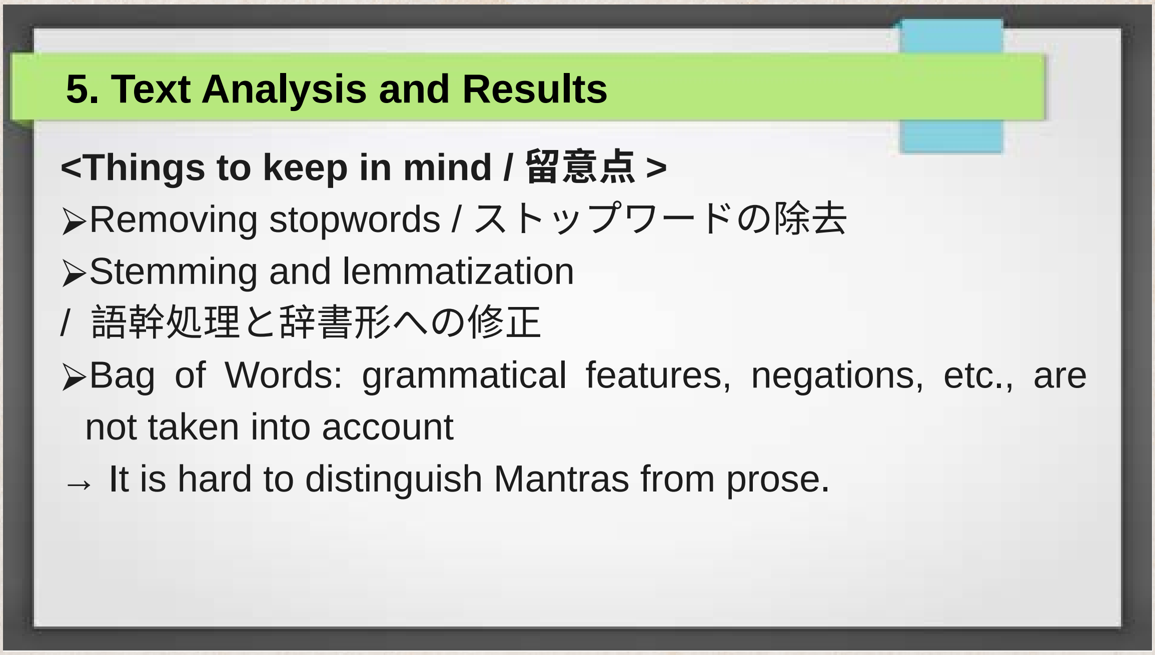
- ➢Lemmatization (1. Clean text corpus)

martā druhan tanūnām indra girvaṇaḥ

marta druh tanu indra girvaṇas

→ Cancel the inflected forms.

20



5. Text Analysis and Results

<Things to keep in mind / 留意点 >

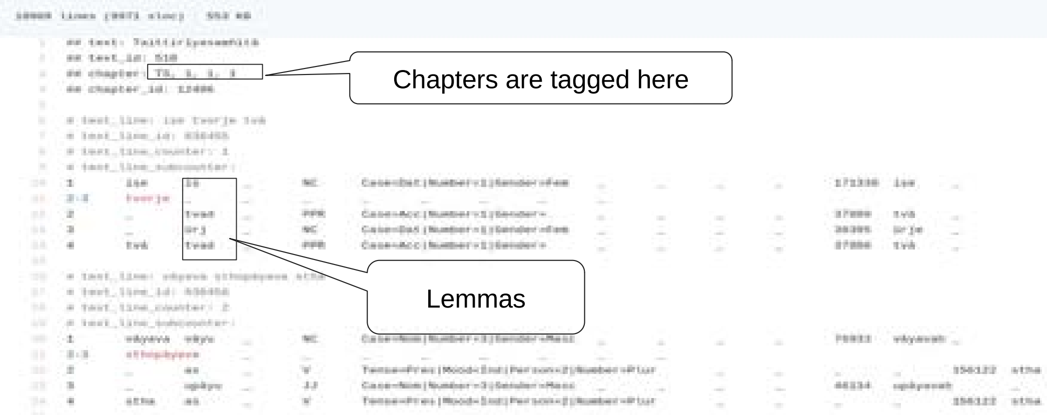
- ➢Removing stopwords / ストップワードの除去
- ➢Stemming and lemmatization
/ 語幹処理と辞書形への修正
- ➢Bag of Words: grammatical features, negations, etc., are not taken into account

→ It is hard to distinguish Mantras from prose.

21

5. Text Analysis and Results

<Universal Dependencies>



```
10008 lines (9971 sloc)   552 KB
## text: TaittirIyasamhitA
## text_id: 518
## chapter: TS, 1, 1, 1
## chapter_id: 12486

# text_line: iSe tvorje tvA
# text_line_id: 838455
# text_line_counter: 1
# text_line_subcounter:
1      iSe      iS     _    NC    Case=Dat|Number=1|Gender=Fem   _   _   _   _   171338   iSe    _
2-3    tvorje   _      _    _     _                              _   _   _   _
2      _        tvad   _    PPR   Case=Acc|Number=1|Gender=      _   _   _   _   37000    tvA    _
3      _        Urj    _    NC    Case=Dat|Number=1|Gender=Fem   _   _   _   _   38395    Urje   _
4      tvA      tvad   _    PPR   Case=Acc|Number=1|Gender=      _   _   _   _   37000    tvA    _

# text_line: vAyava sthopAyava stha
# text_line_id: 838456
# text_line_counter: 2
# text_line_subcounter:
1      vAyava   vAyu   _    NC    Case=Nom|Number=3|Gender=Masc  _   _   _   _   75931    vAyavaH  _
2-3    sthopAyava _    _    _     _                              _   _   _   _
2      _        as     _    V     Tense=Pres|Mood=Ind|Person=2|Number=Plur   _   _   _   _   156122   stha
3      _        upAyu  _    JJ    Case=Nom|Number=3|Gender=Masc  _   _   _   _   46134    upAyavaH  _
4      stha     as     _    V     Tense=Pres|Mood=Ind|Person=2|Number=Plur   _   _   _   _   156122   stha
```

Chapters are tagged here

Lemmas

22

5. Text Analysis and Results

<Corpus Information / コーパス情報 >

[Training corpus]

➢Total number of texts: 6,277

➢Avg. sentences / text: 98.71

➢Avg. tokens / sentence: 5.57

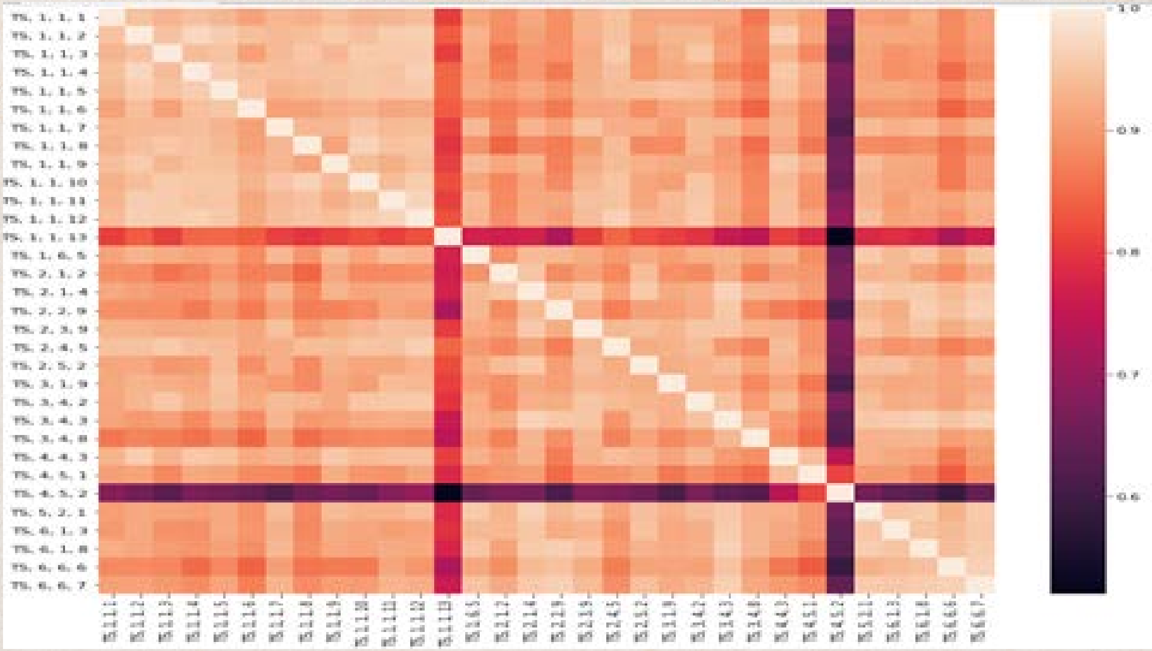
[Corpus of *Taittirīya Saṃhitā*]

➢Total number of chapters: 32

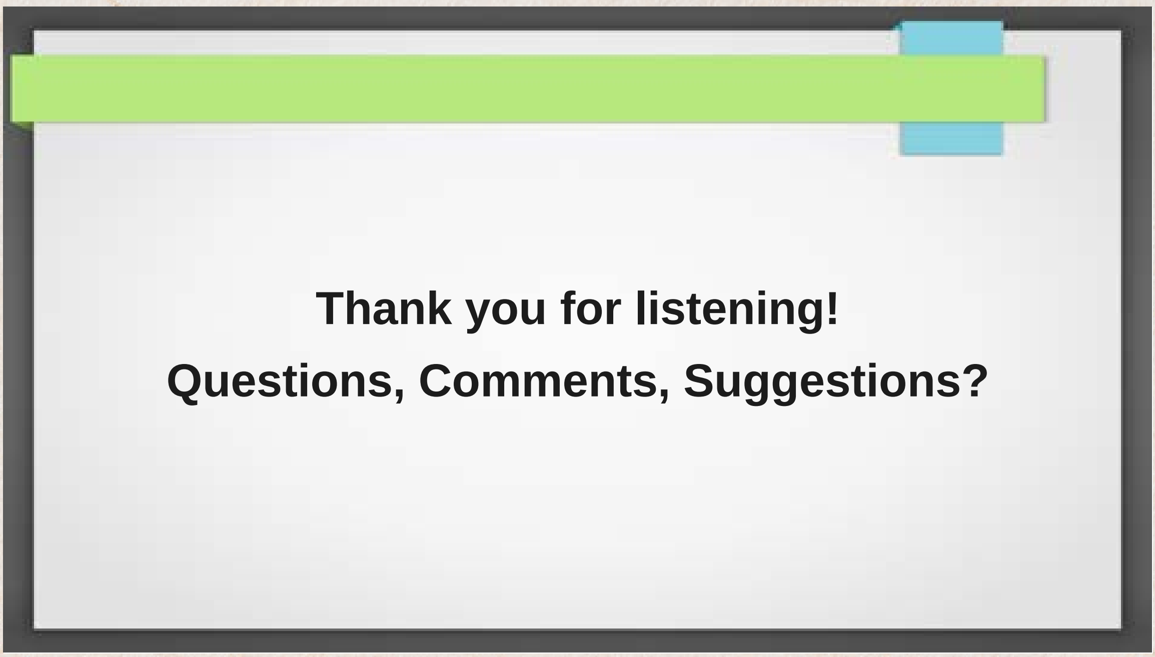
➢Avg. sentences / chapter: 26.15

➢Avg. tokens / sentence: 4.18

23



24



Thank you for listening!

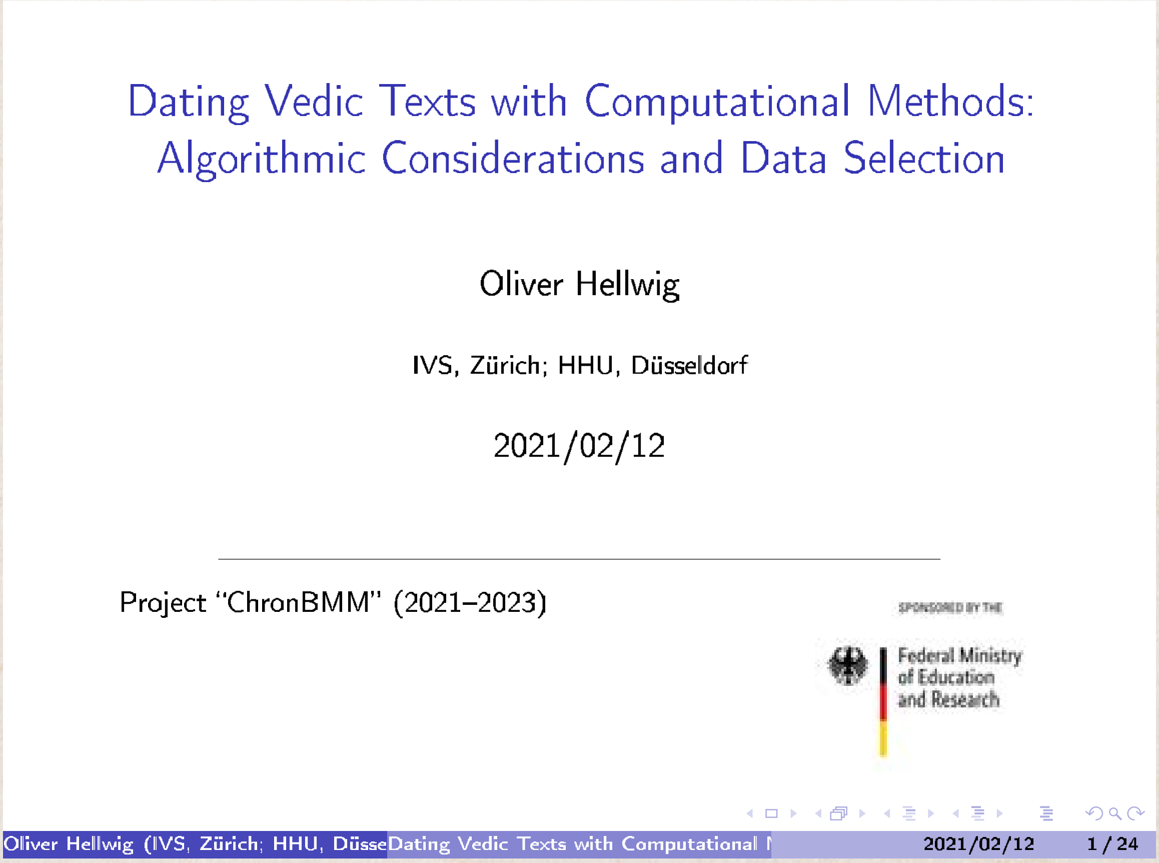
Questions, Comments, Suggestions?

Presentation 5 ｜報告 5

Dating Vedic Texts with Computational Models: Algorithmic Considerations and Data Selection

Oliver Hellwig (University of Zurich, Department of Comparative Language Science)

1



Dating Vedic Texts with Computational Methods: Algorithmic Considerations and Data Selection

Oliver Hellwig

IVS, Zürich; HHU, Düsseldorf

2021/02/12

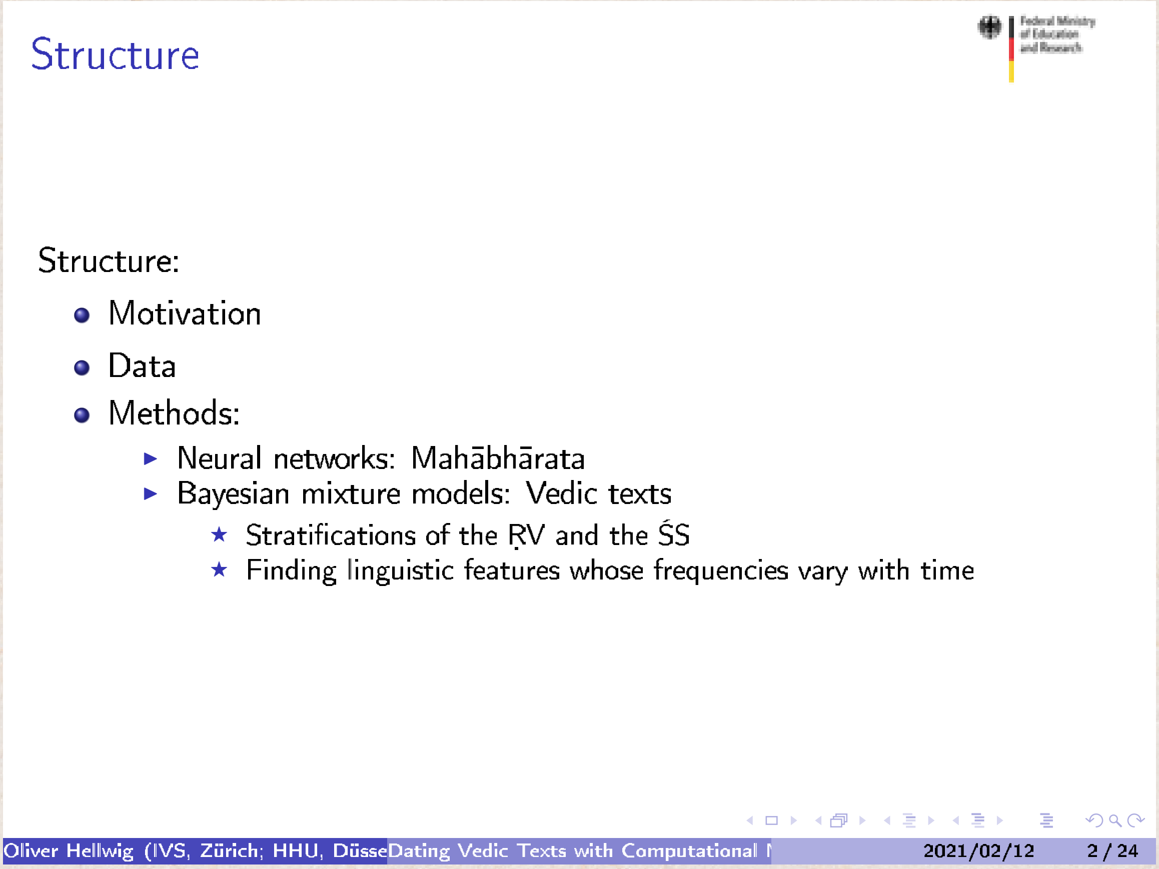
Project "ChronBMM" (2021–2023)

SPONSORED BY THE

Federal Ministry
of Education
and Research

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational I 2021/02/12 1 / 24

2



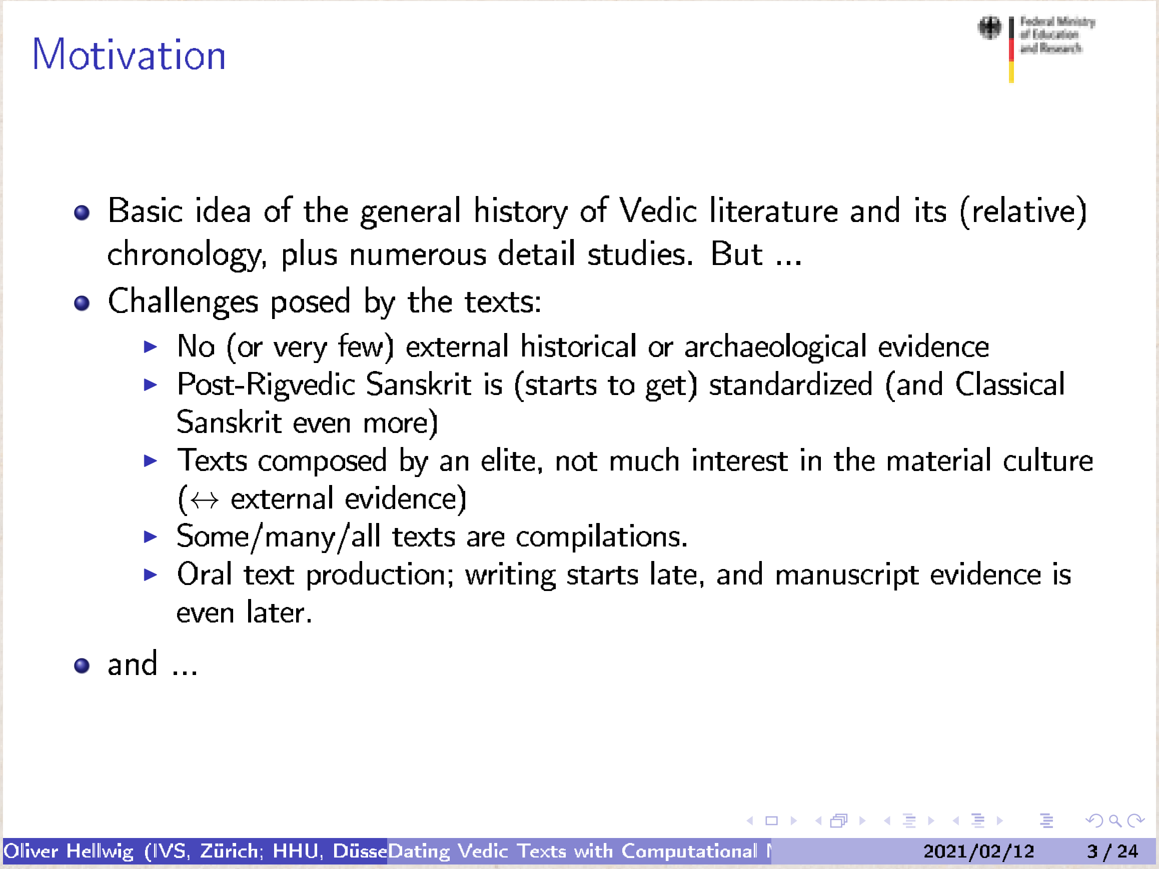
Structure

Structure:

- Motivation
- Data
- Methods:
  - Neural networks: Mahābhārata
  - Bayesian mixture models: Vedic texts
    - Stratifications of the ṚV and the ŚS
    - Finding linguistic features whose frequencies vary with time

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 2 / 24

3

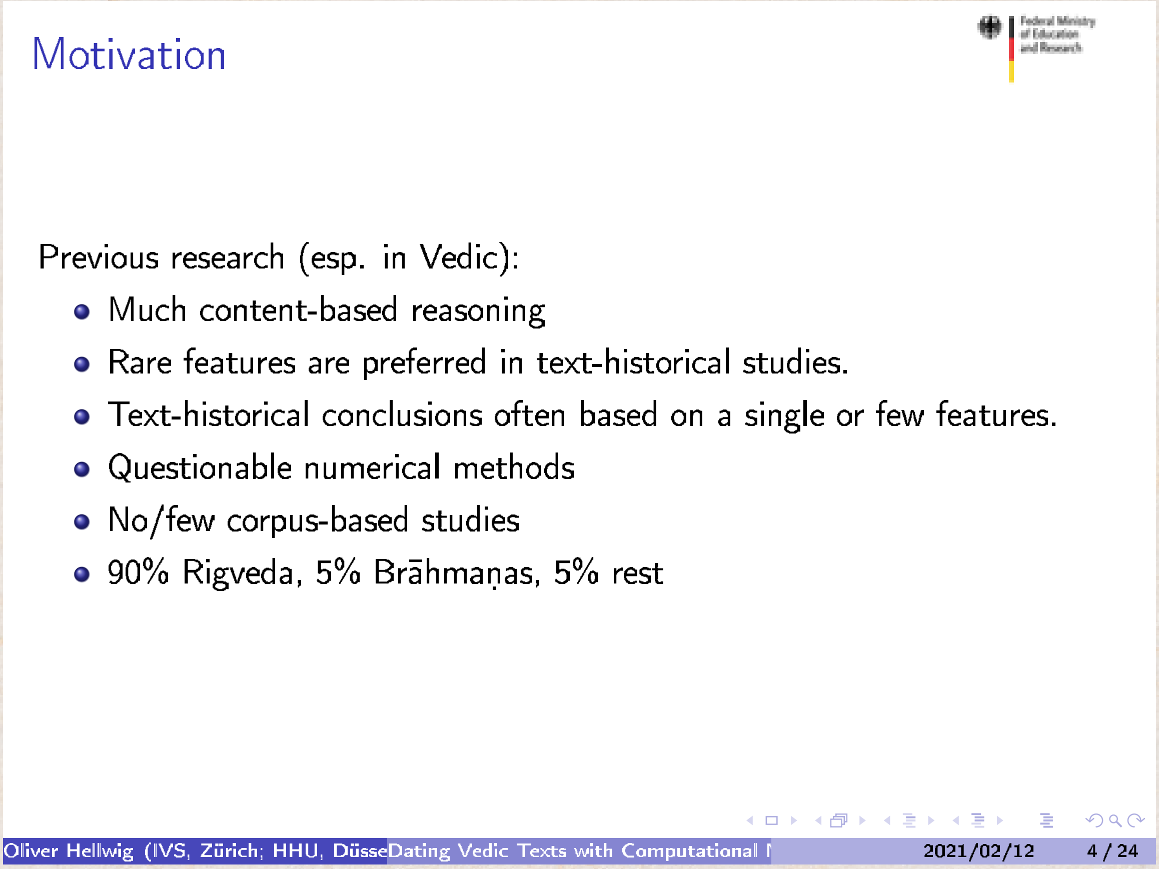


Motivation

- Basic idea of the general history of Vedic literature and its (relative) chronology, plus numerous detail studies. But ...
- Challenges posed by the texts:
  - No (or very few) external historical or archaeological evidence
  - Post-Rigvedic Sanskrit is (starts to get) standardized (and Classical Sanskrit even more)
  - Texts composed by an elite, not much interest in the material culture (↔ external evidence)
  - Some/many/all texts are compilations.
  - Oral text production; writing starts late, and manuscript evidence is even later.
- and ...

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 3 / 24

4



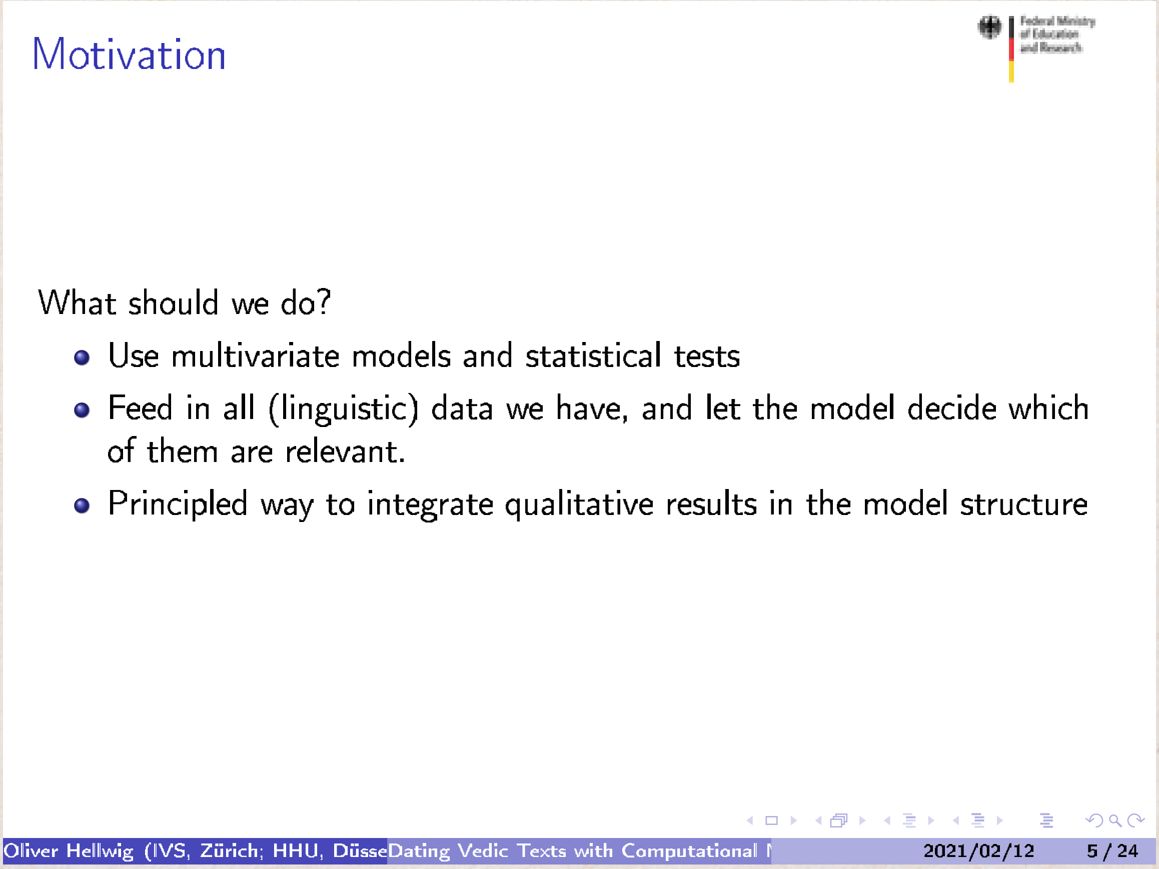
Motivation

Previous research (esp. in Vedic):

- Much content-based reasoning
- Rare features are preferred in text-historical studies.
- Text-historical conclusions often based on a single or few features.
- Questionable numerical methods
- No/few corpus-based studies
- 90% Rigveda, 5% Brāhmaṇas, 5% rest

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 4 / 24

5

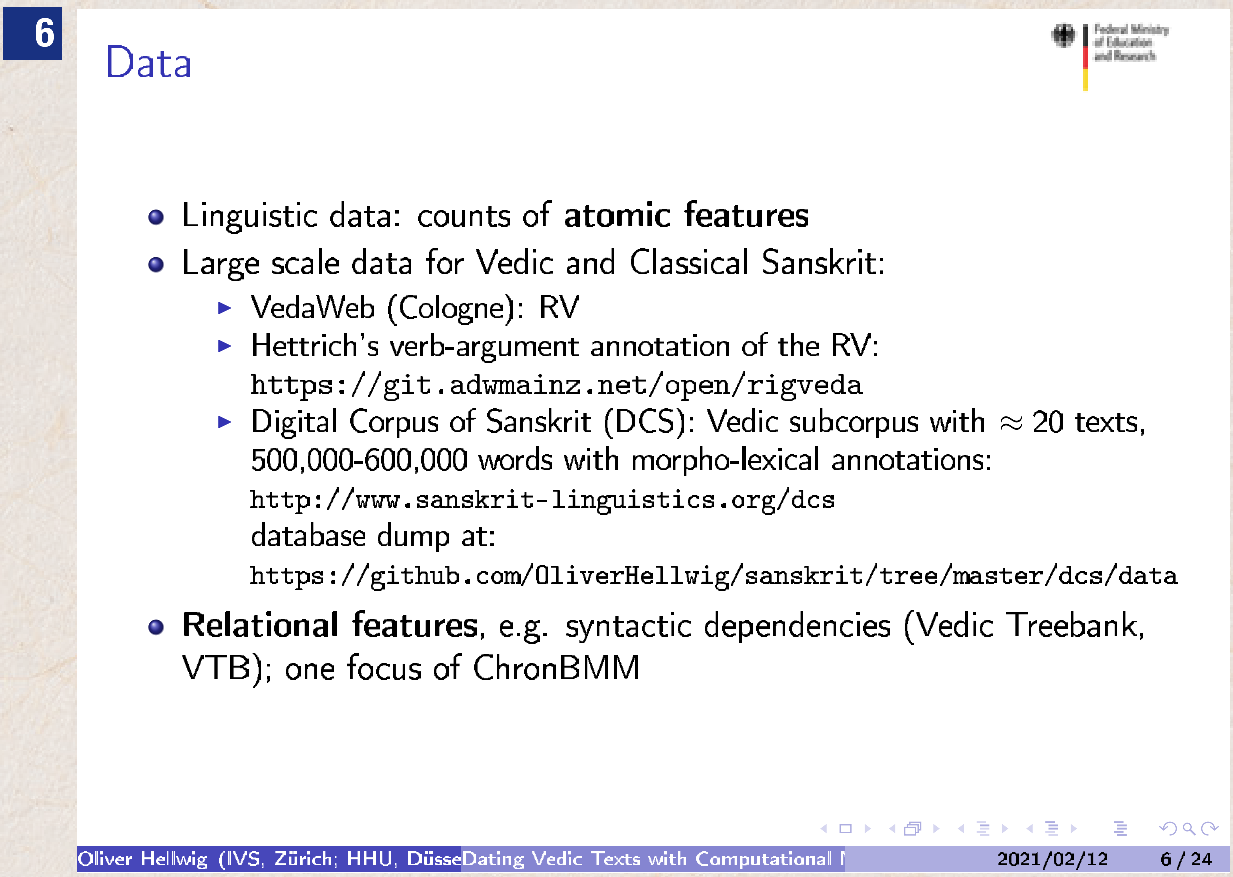


Motivation

What should we do?

- Use multivariate models and statistical tests
- Feed in all (linguistic) data we have, and let the model decide which of them are relevant.
- Principled way to integrate qualitative results in the model structure

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 5 / 24

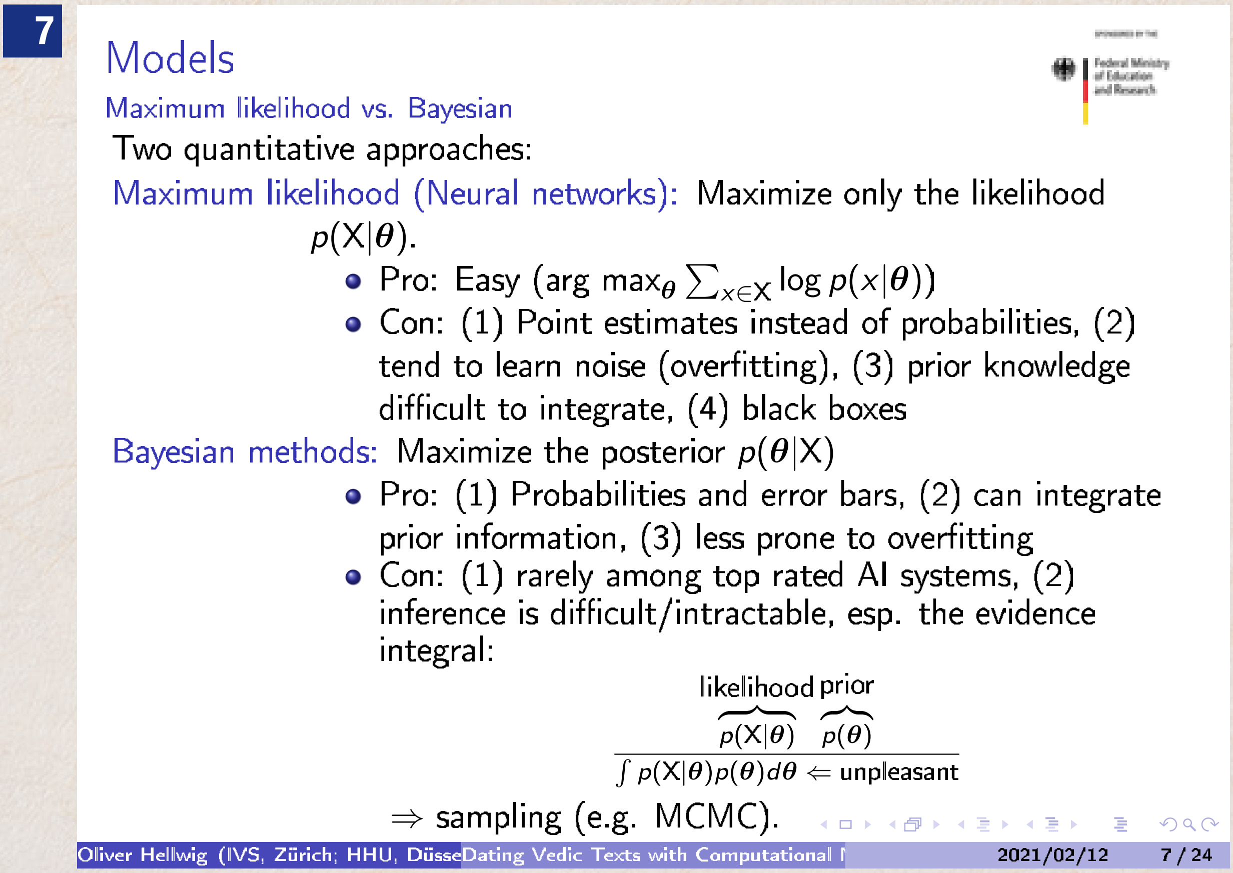


6

Data

- Linguistic data: counts of **atomic features**
- Large scale data for Vedic and Classical Sanskrit:
  - VedaWeb (Cologne): RV
  - Hettrich's verb-argument annotation of the RV: `https://git.adwmainz.net/open/rigveda`
  - Digital Corpus of Sanskrit (DCS): Vedic subcorpus with ≈ 20 texts, 500,000-600,000 words with morpho-lexical annotations: `http://www.sanskrit-linguistics.org/dcs`
    database dump at: `https://github.com/OliverHellwig/sanskrit/tree/master/dcs/data`
- **Relational features**, e.g. syntactic dependencies (Vedic Treebank, VTB); one focus of ChronBMM

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 6 / 24



7

Models

Maximum likelihood vs. Bayesian

Two quantitative approaches:

Maximum likelihood (Neural networks): Maximize only the likelihood $p(\mathrm{X}|\boldsymbol{\theta})$.

- Pro: Easy (arg $\max_{\boldsymbol{\theta}} \sum_{x \in \mathrm{X}} \log p(x|\boldsymbol{\theta})$)
- Con: (1) Point estimates instead of probabilities, (2) tend to learn noise (overfitting), (3) prior knowledge difficult to integrate, (4) black boxes

Bayesian methods: Maximize the posterior $p(\boldsymbol{\theta}|\mathrm{X})$

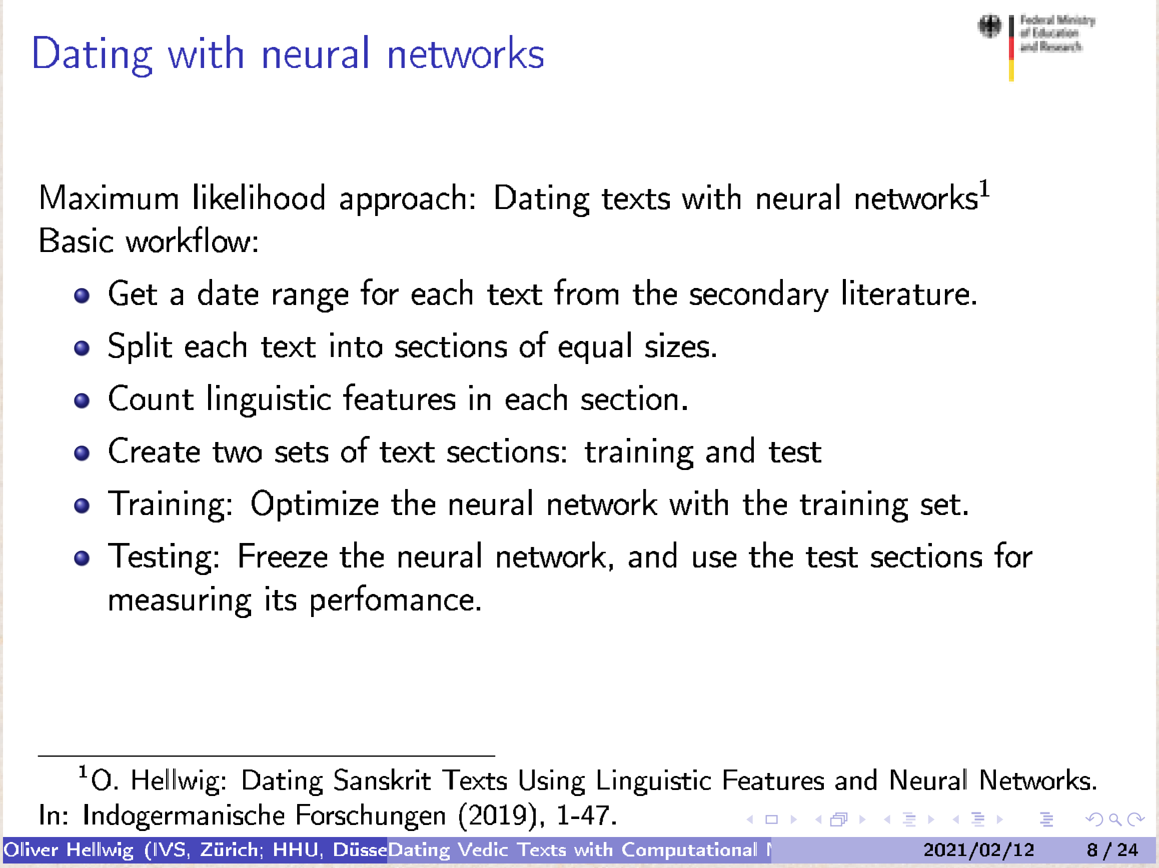
- Pro: (1) Probabilities and error bars, (2) can integrate prior information, (3) less prone to overfitting
- Con: (1) rarely among top rated AI systems, (2) inference is difficult/intractable, esp. the evidence integral:

$$\frac{\overbrace{p(\mathrm{X}|\boldsymbol{\theta})}^{\text{likelihood}}\ \overbrace{p(\boldsymbol{\theta})}^{\text{prior}}}{\int p(\mathrm{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta} \Leftarrow \text{unpleasant}}$$

⇒ sampling (e.g. MCMC).

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 7 / 24

8



Dating with neural networks

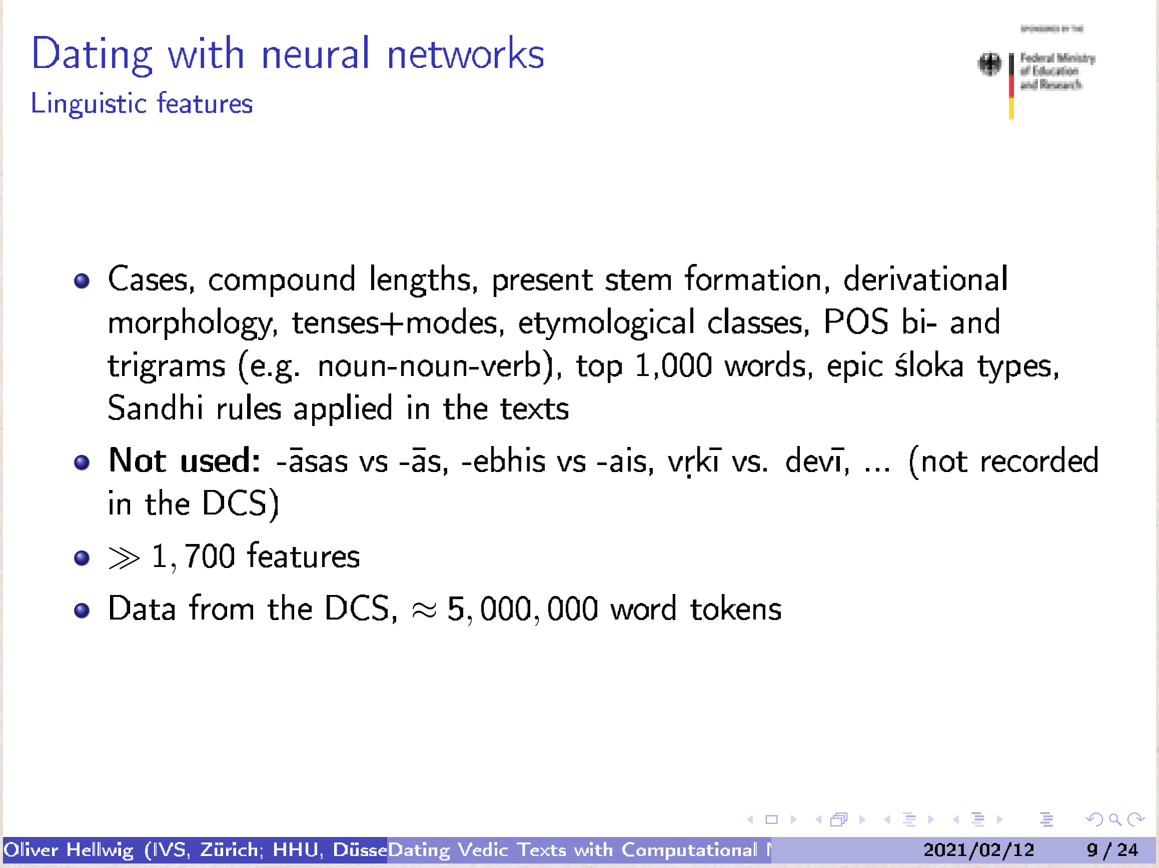
Maximum likelihood approach: Dating texts with neural networks[1]
Basic workflow:

- Get a date range for each text from the secondary literature.
- Split each text into sections of equal sizes.
- Count linguistic features in each section.
- Create two sets of text sections: training and test
- Training: Optimize the neural network with the training set.
- Testing: Freeze the neural network, and use the test sections for measuring its perfomance.

[1] O. Hellwig: Dating Sanskrit Texts Using Linguistic Features and Neural Networks. In: Indogermanische Forschungen (2019), 1-47.

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 8 / 24

9



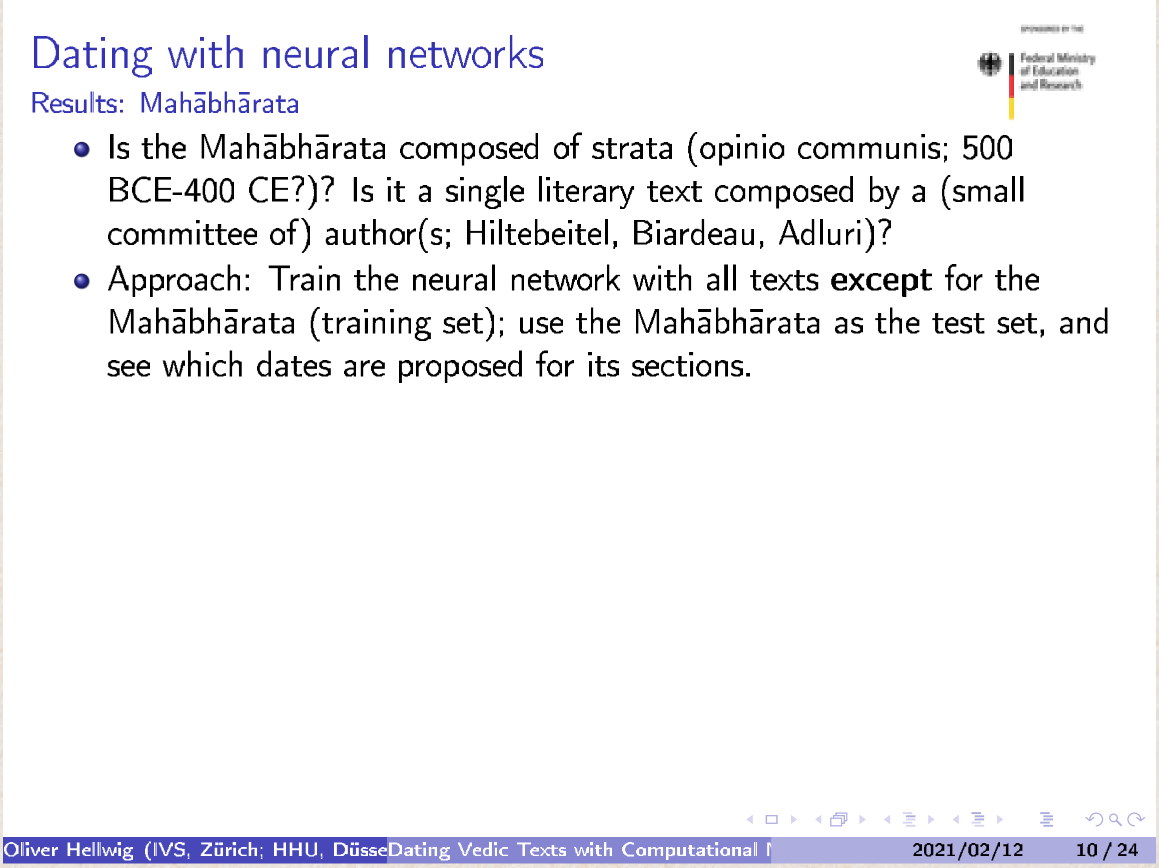
Dating with neural networks

Linguistic features

- Cases, compound lengths, present stem formation, derivational morphology, tenses+modes, etymological classes, POS bi- and trigrams (e.g. noun-noun-verb), top 1,000 words, epic śloka types, Sandhi rules applied in the texts
- **Not used:** -āsas vs -ās, -ebhis vs -ais, vṛkī vs. devī, ... (not recorded in the DCS)
- $\gg 1,700$ features
- Data from the DCS, $\approx 5,000,000$ word tokens

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 9 / 24

10



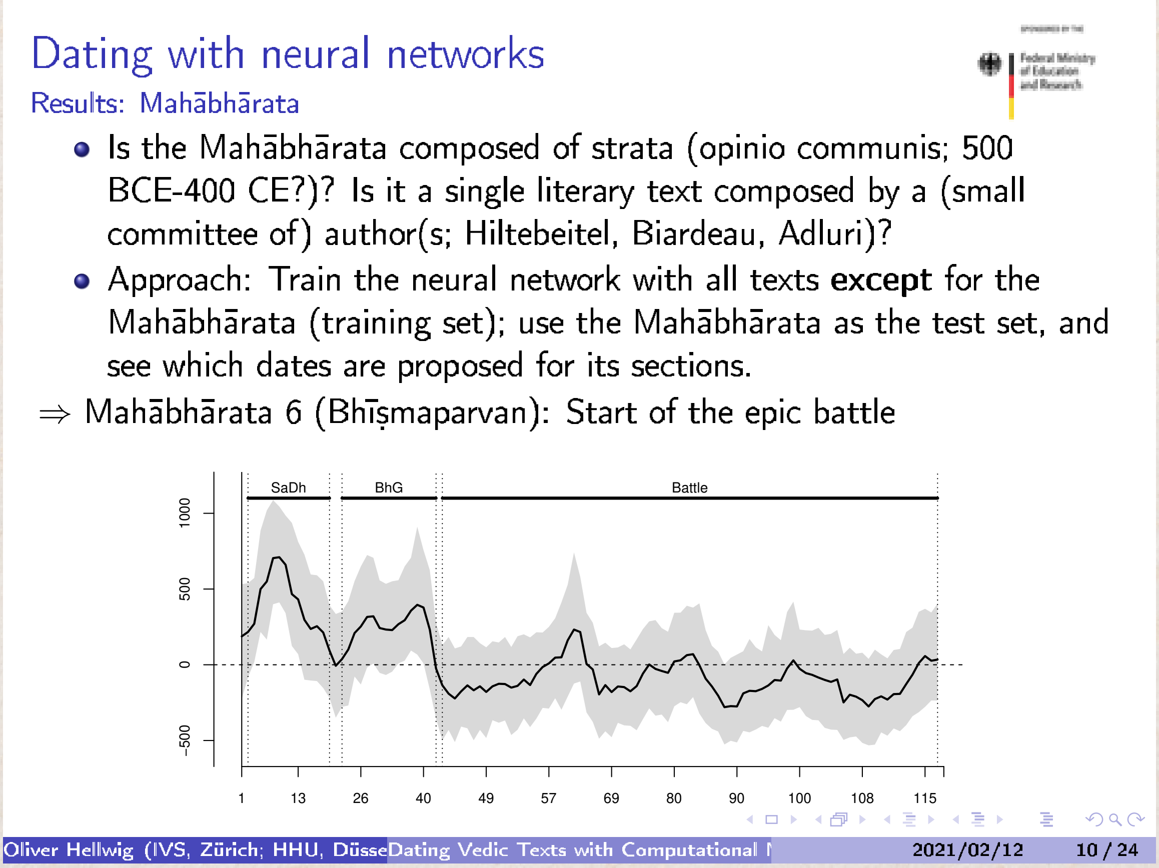
Dating with neural networks

Results: Mahābhārata

- Is the Mahābhārata composed of strata (opinio communis; 500 BCE-400 CE?)? Is it a single literary text composed by a (small committee of) author(s; Hiltebeitel, Biardeau, Adluri)?
- Approach: Train the neural network with all texts **except** for the Mahābhārata (training set); use the Mahābhārata as the test set, and see which dates are proposed for its sections.

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational 2021/02/12 10 / 24

11



Dating with neural networks

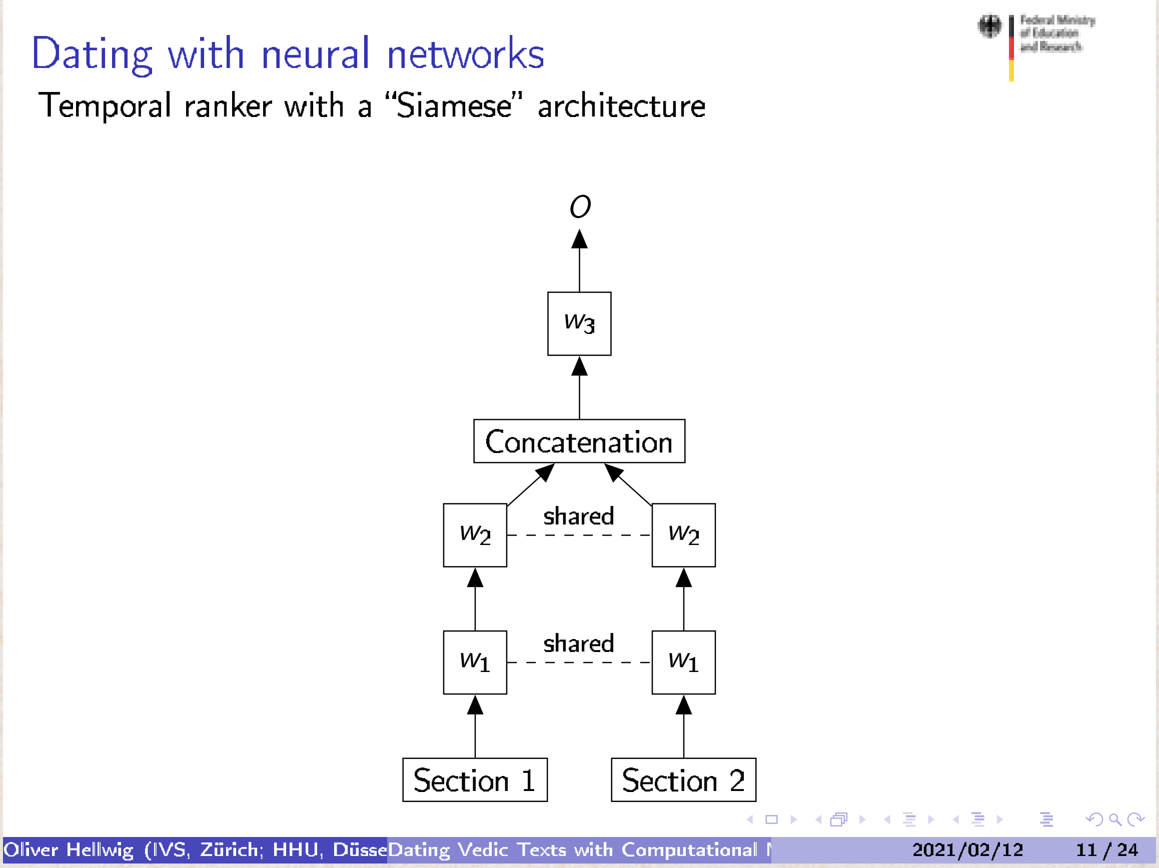
Results: Mahābhārata

- Is the Mahābhārata composed of strata (opinio communis; 500 BCE-400 CE?)? Is it a single literary text composed by a (small committee of) author(s; Hiltebeitel, Biardeau, Adluri)?
- Approach: Train the neural network with all texts **except** for the Mahābhārata (training set); use the Mahābhārata as the test set, and see which dates are proposed for its sections.

⇒ Mahābhārata 6 (Bhīṣmaparvan): Start of the epic battle

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational 2021/02/12 10 / 24

12



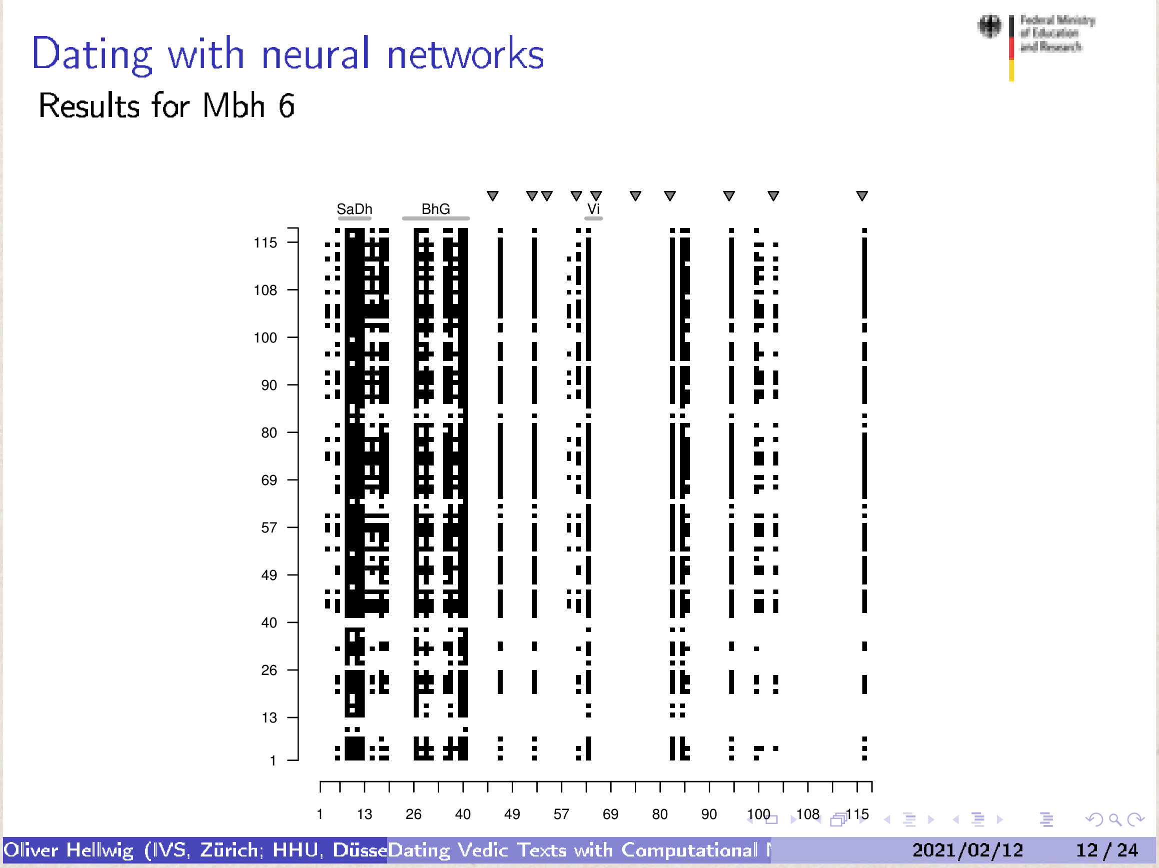
Federal Ministry of Education and Research

Dating with neural networks

Temporal ranker with a “Siamese” architecture

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 11 / 24

13



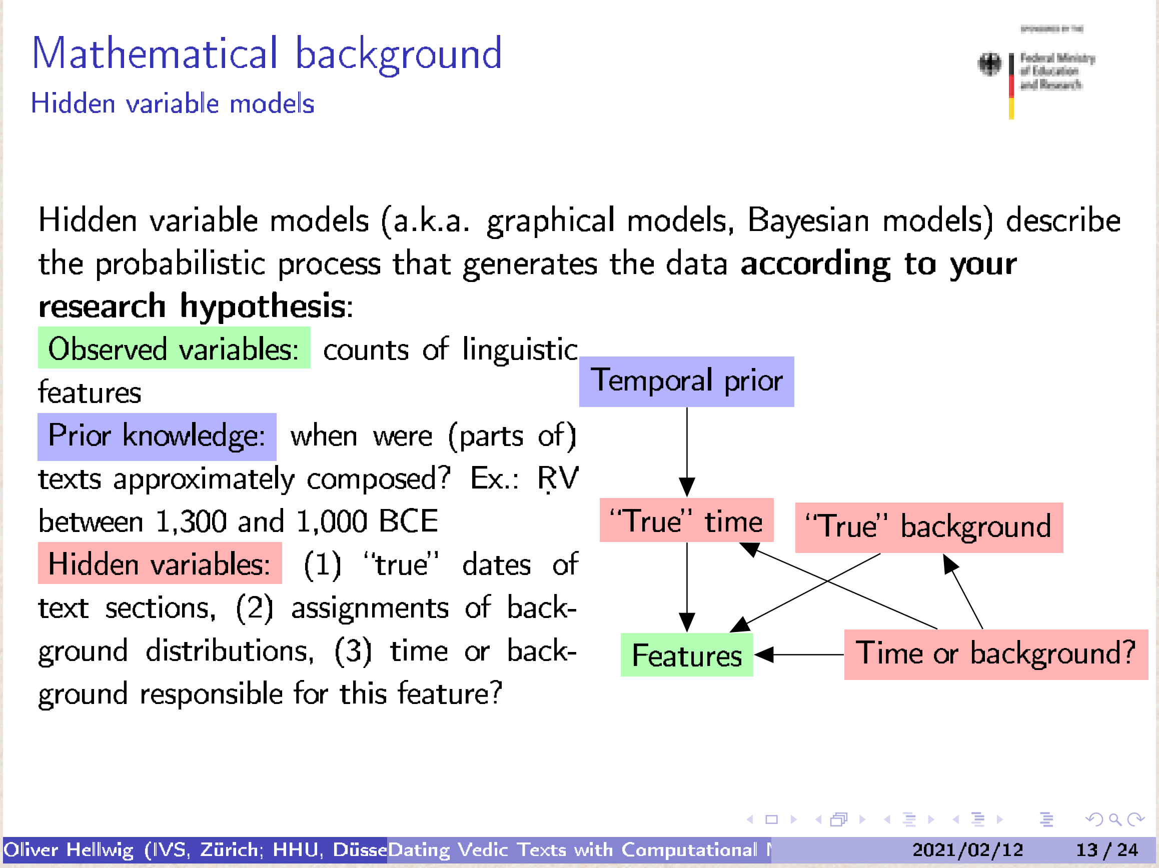
Federal Ministry of Education and Research

Dating with neural networks

Results for Mbh 6

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 12 / 24

14



Mathematical background

Hidden variable models

Hidden variable models (a.k.a. graphical models, Bayesian models) describe the probabilistic process that generates the data **according to your research hypothesis**:

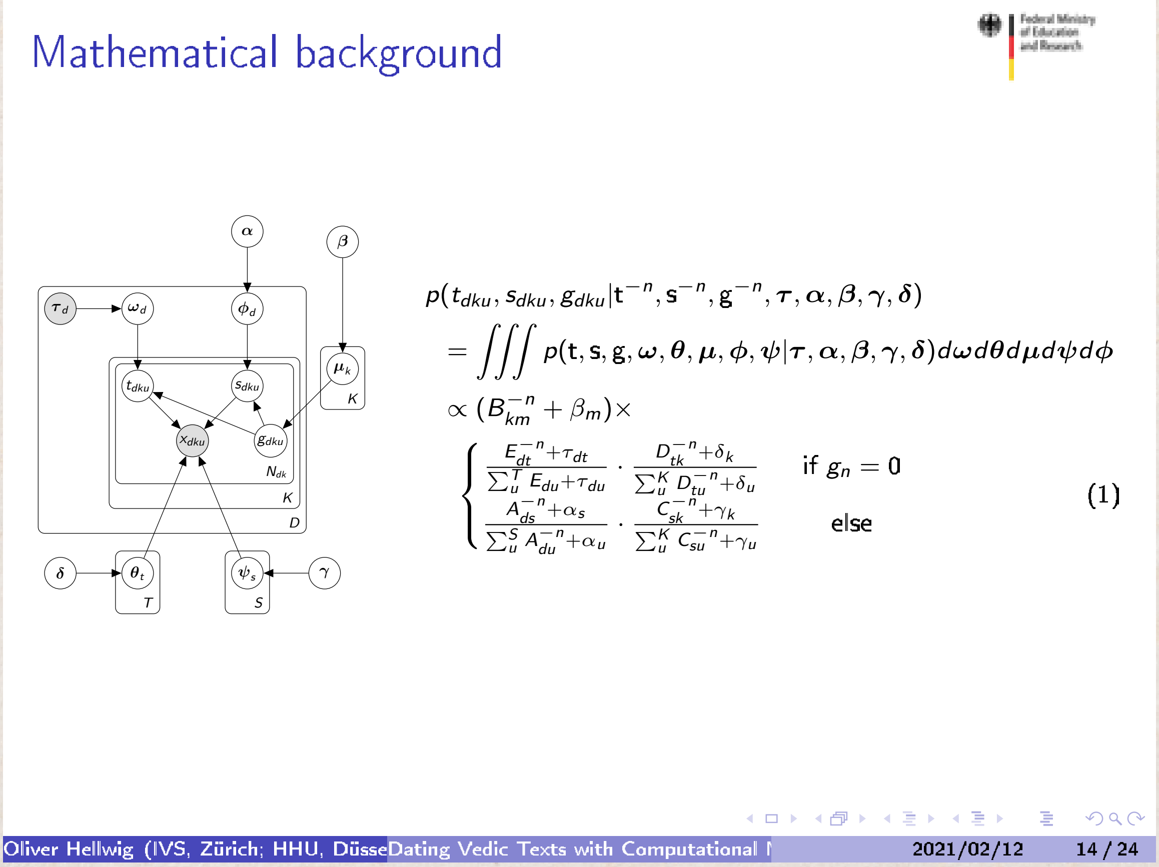
Observed variables: counts of linguistic features

Prior knowledge: when were (parts of) texts approximately composed? Ex.: ṚV between 1,300 and 1,000 BCE

Hidden variables: (1) "true" dates of text sections, (2) assignments of background distributions, (3) time or background responsible for this feature?

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 13 / 24

15

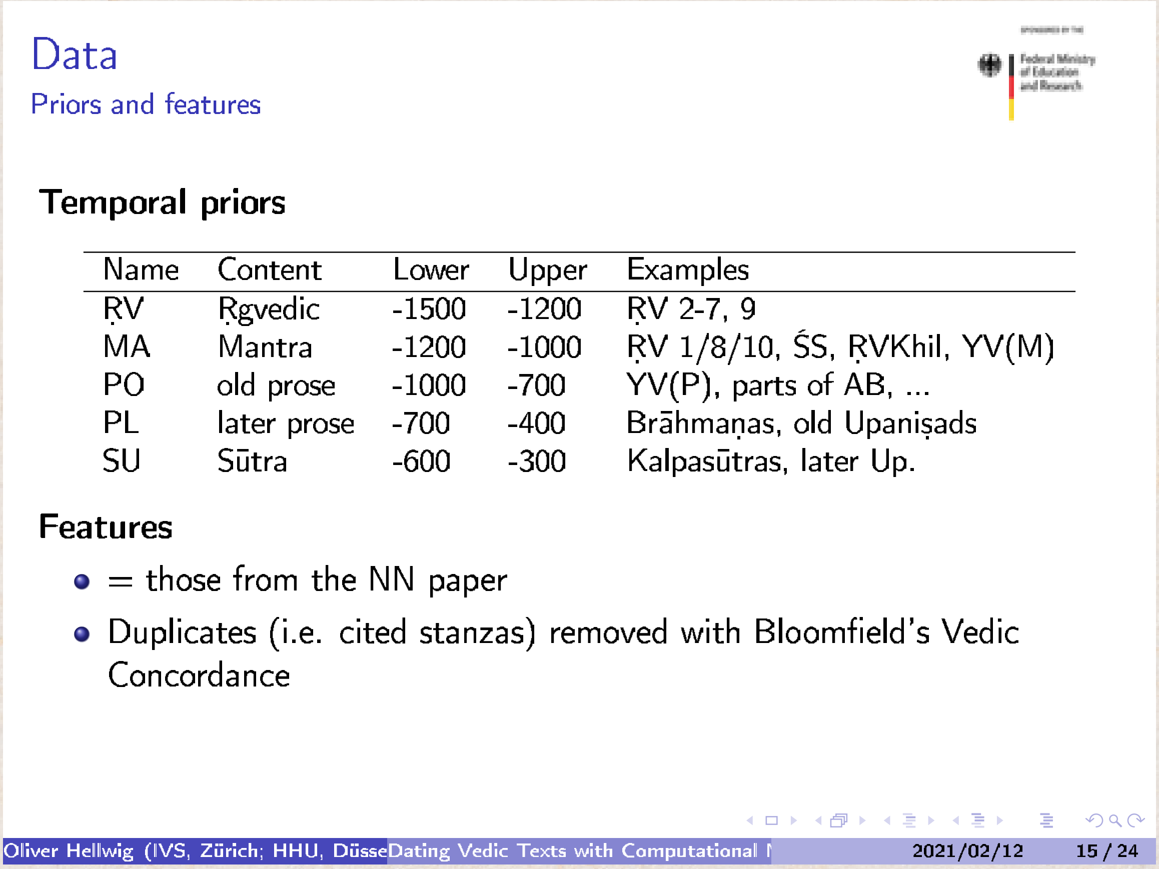


Mathematical background

$$
\begin{aligned}
&p(t_{dku}, s_{dku}, g_{dku} | \mathbf{t}^{-n}, \mathbf{s}^{-n}, \mathbf{g}^{-n}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) \\
&\quad = \iiint p(\mathbf{t}, \mathbf{s}, \mathbf{g}, \boldsymbol{\omega}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\phi}, \psi | \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}) d\boldsymbol{\omega} d\boldsymbol{\theta} d\boldsymbol{\mu} d\psi d\phi \\
&\quad \propto (B_{km}^{-n} + \beta_m) \times \\
&\quad \begin{cases} \frac{E_{dt}^{-n} + \tau_{dt}}{\sum_u^T E_{du} + \tau_{du}} \cdot \frac{D_{tk}^{-n} + \delta_k}{\sum_u^K D_{tu}^{-n} + \delta_u} & \text{if } g_n = 0 \\ \frac{A_{ds}^{-n} + \alpha_s}{\sum_u^S A_{du}^{-n} + \alpha_u} \cdot \frac{C_{sk}^{-n} + \gamma_k}{\sum_u^K C_{su}^{-n} + \gamma_u} & \text{else} \end{cases}
\end{aligned} \tag{1}
$$

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 14 / 24

16



Data

Priors and features

Temporal priors

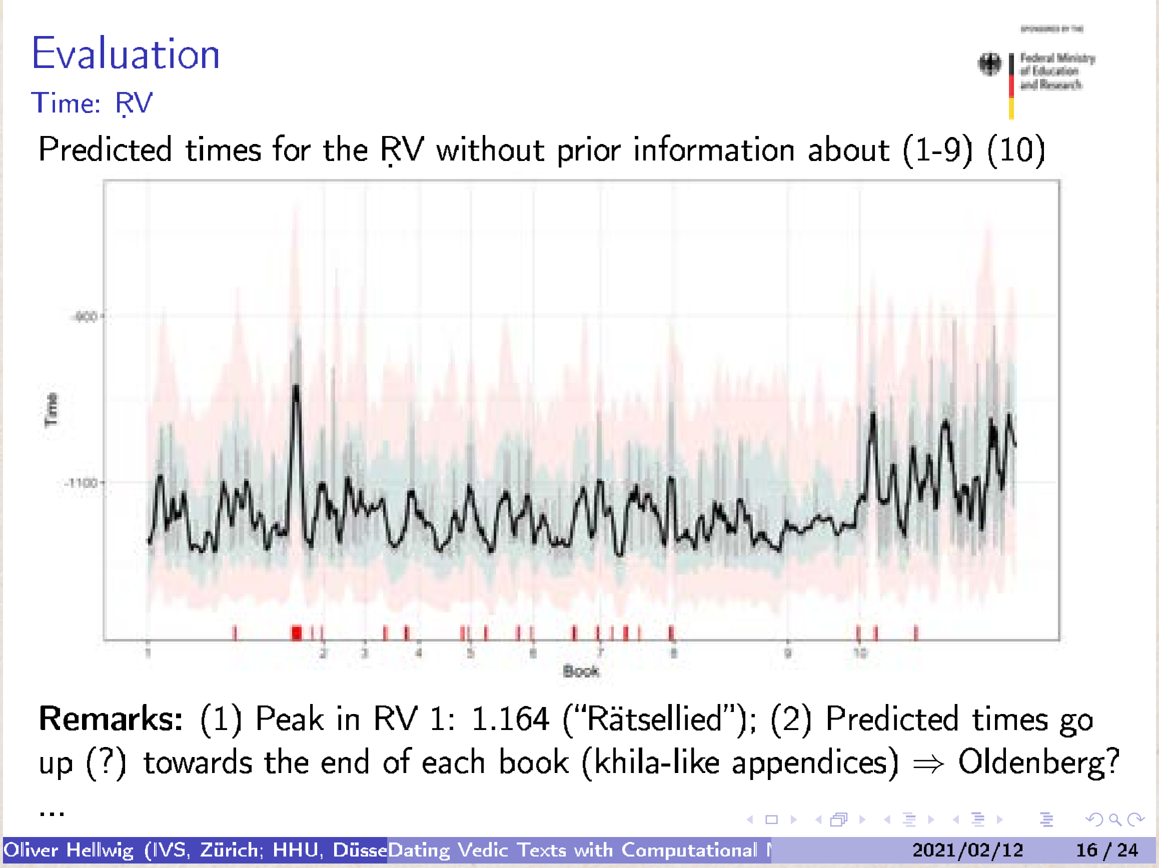
Name	Content	Lower	Upper	Examples
ṚV	Ṛgvedic	-1500	-1200	ṚV 2-7, 9
MA	Mantra	-1200	-1000	ṚV 1/8/10, ŚS, ṚVKhil, YV(M)
PO	old prose	-1000	-700	YV(P), parts of AB, ...
PL	later prose	-700	-400	Brāhmaṇas, old Upaniṣads
SU	Sūtra	-600	-300	Kalpasūtras, later Up.

Features

- = those from the NN paper
- Duplicates (i.e. cited stanzas) removed with Bloomfield's Vedic Concordance

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 15 / 24

17



Evaluation

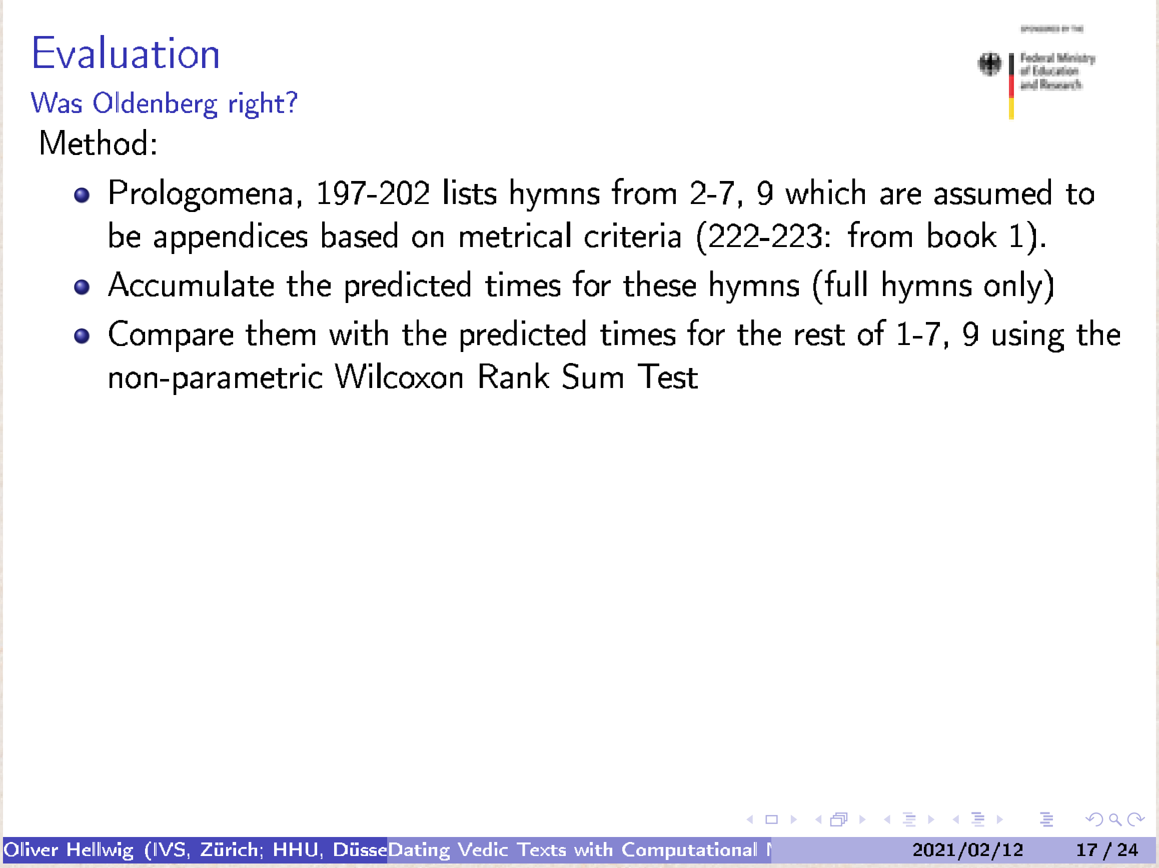
Time: ṚV

Predicted times for the ṚV without prior information about (1-9) (10)

Remarks: (1) Peak in RV 1: 1.164 ("Rätsellied"); (2) Predicted times go up (?) towards the end of each book (khila-like appendices) ⇒ Oldenberg? ...

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 16 / 24

18



Evaluation

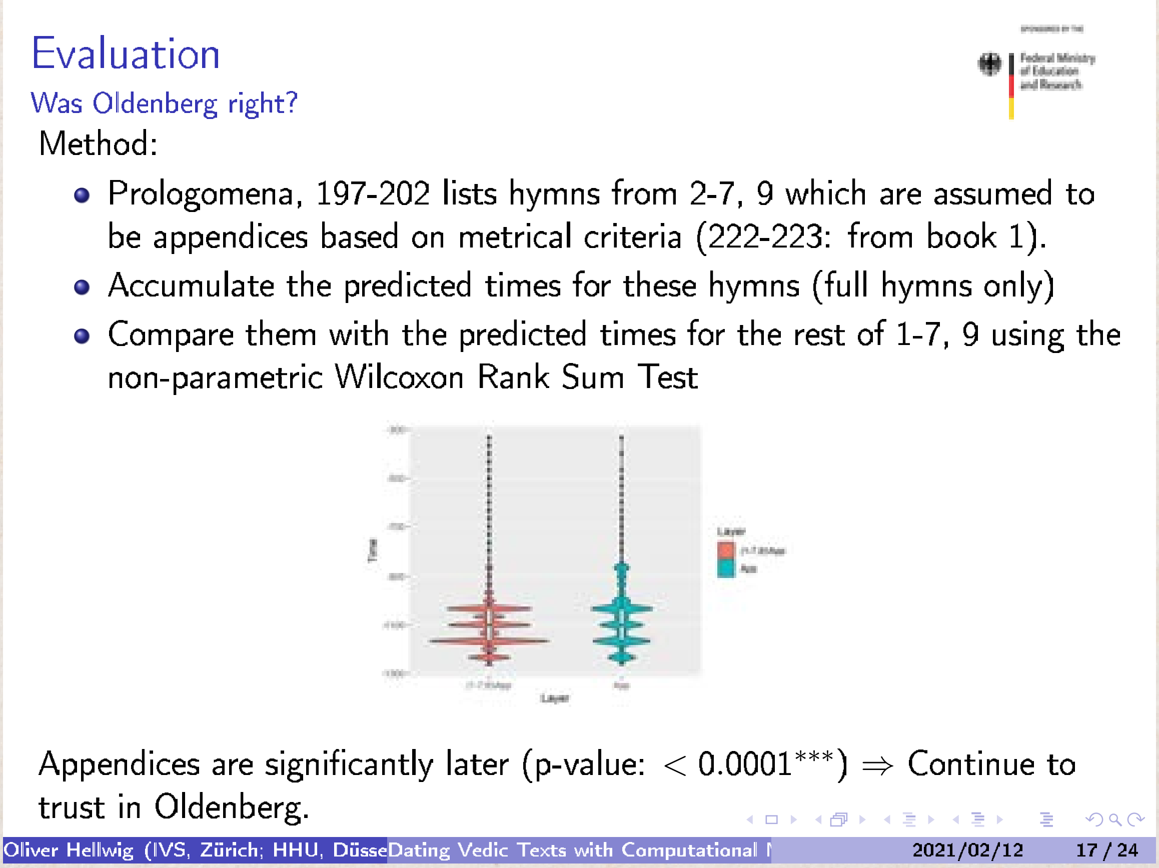
Was Oldenberg right?

Method:

- Prologomena, 197-202 lists hymns from 2-7, 9 which are assumed to be appendices based on metrical criteria (222-223: from book 1).
- Accumulate the predicted times for these hymns (full hymns only)
- Compare them with the predicted times for the rest of 1-7, 9 using the non-parametric Wilcoxon Rank Sum Test

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 17 / 24

19



Evaluation

Was Oldenberg right?

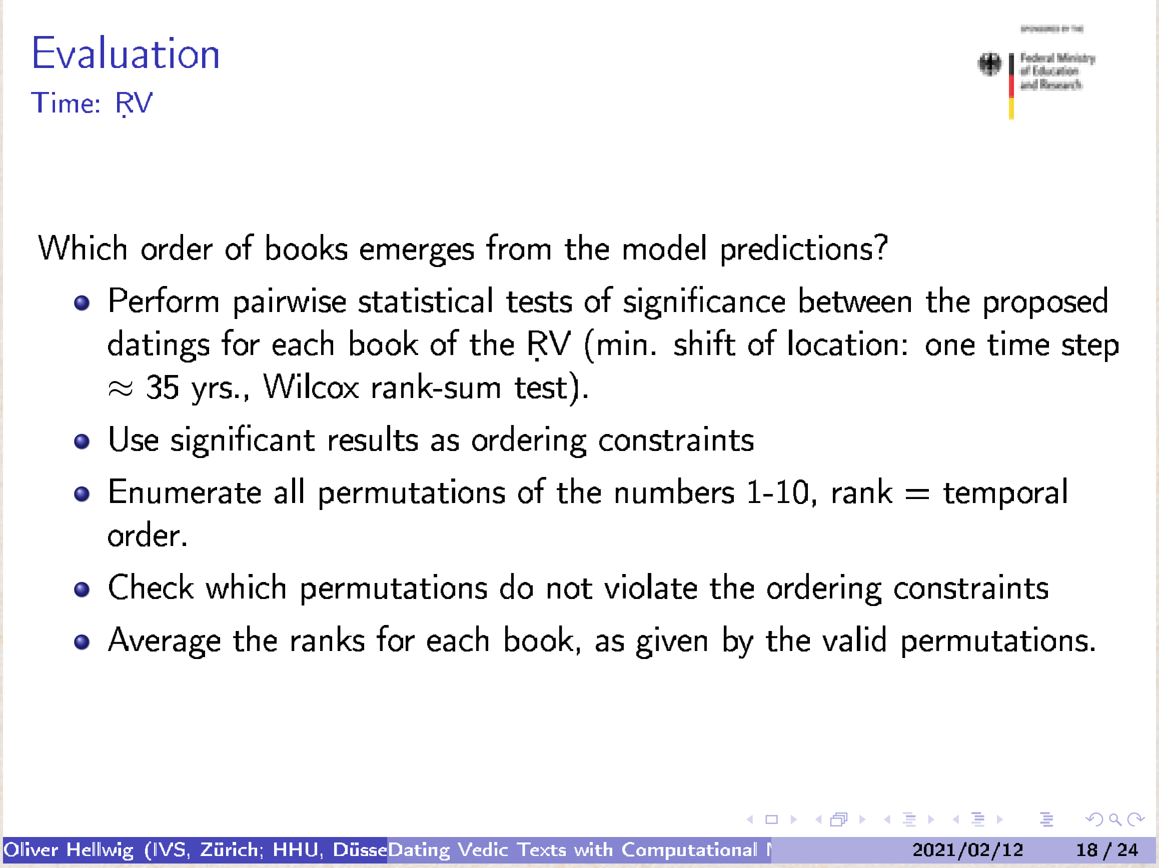
Method:

- Prologomena, 197-202 lists hymns from 2-7, 9 which are assumed to be appendices based on metrical criteria (222-223: from book 1).
- Accumulate the predicted times for these hymns (full hymns only)
- Compare them with the predicted times for the rest of 1-7, 9 using the non-parametric Wilcoxon Rank Sum Test

Appendices are significantly later (p-value: $< 0.0001^{***}$) ⇒ Continue to trust in Oldenberg.

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 17 / 24

20



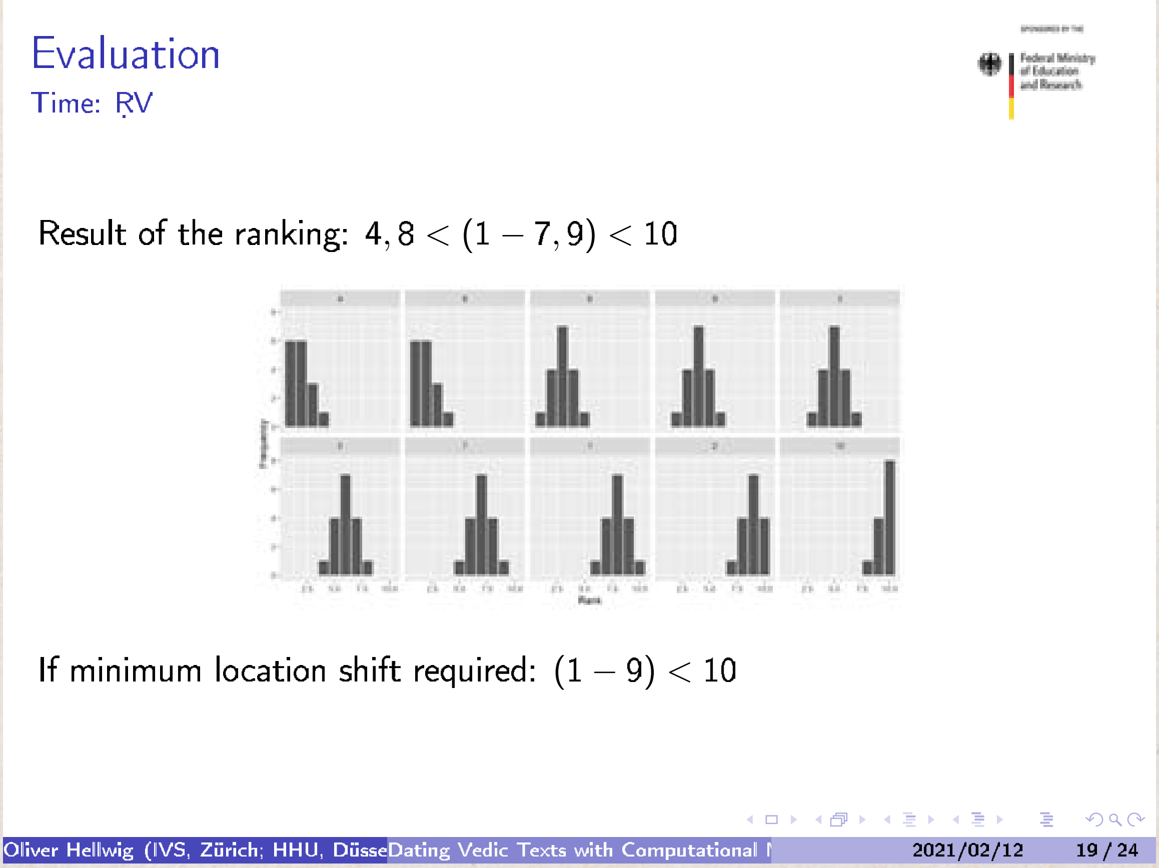
Evaluation

Time: ṚV

Which order of books emerges from the model predictions?

- Perform pairwise statistical tests of significance between the proposed datings for each book of the ṚV (min. shift of location: one time step ≈ 35 yrs., Wilcox rank-sum test).
- Use significant results as ordering constraints
- Enumerate all permutations of the numbers 1-10, rank = temporal order.
- Check which permutations do not violate the ordering constraints
- Average the ranks for each book, as given by the valid permutations.

21



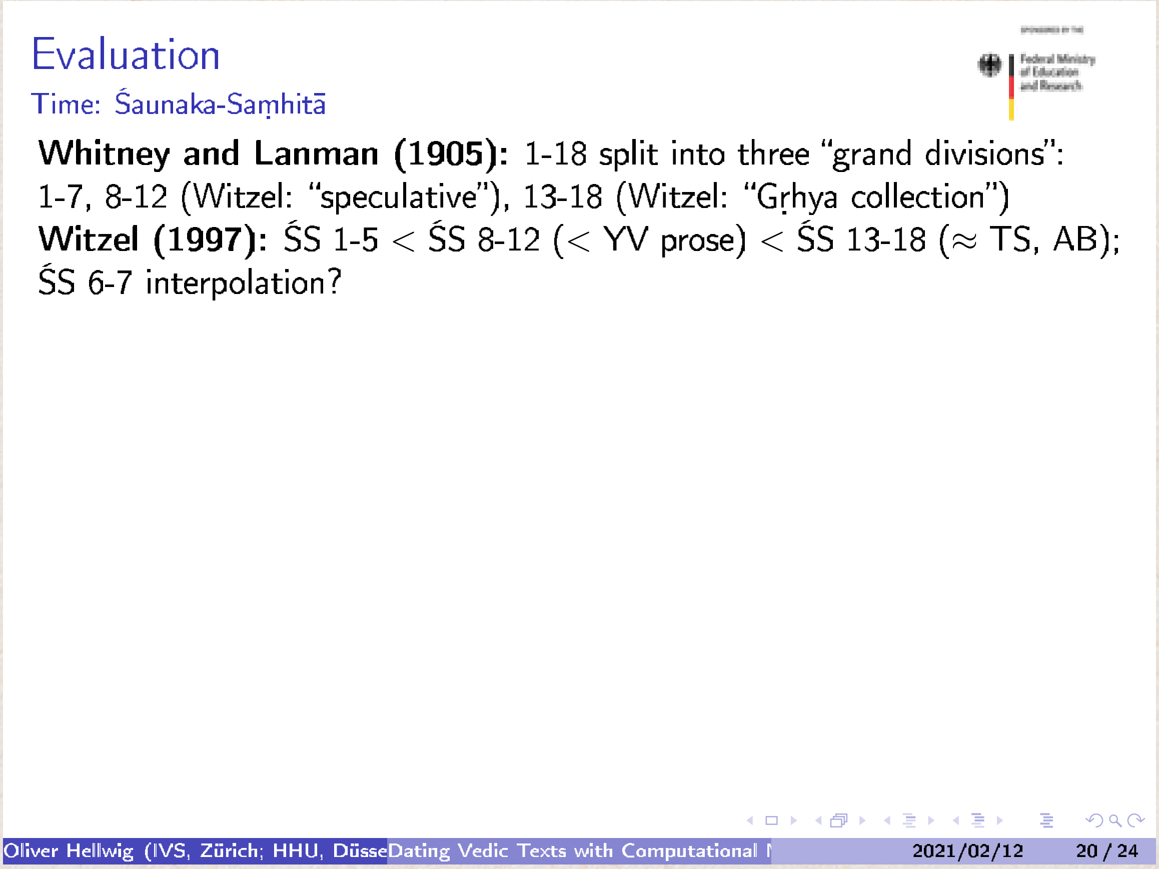
Evaluation

Time: ṚV

Result of the ranking: $4, 8 < (1 - 7, 9) < 10$

If minimum location shift required: $(1 - 9) < 10$

22



Evaluation

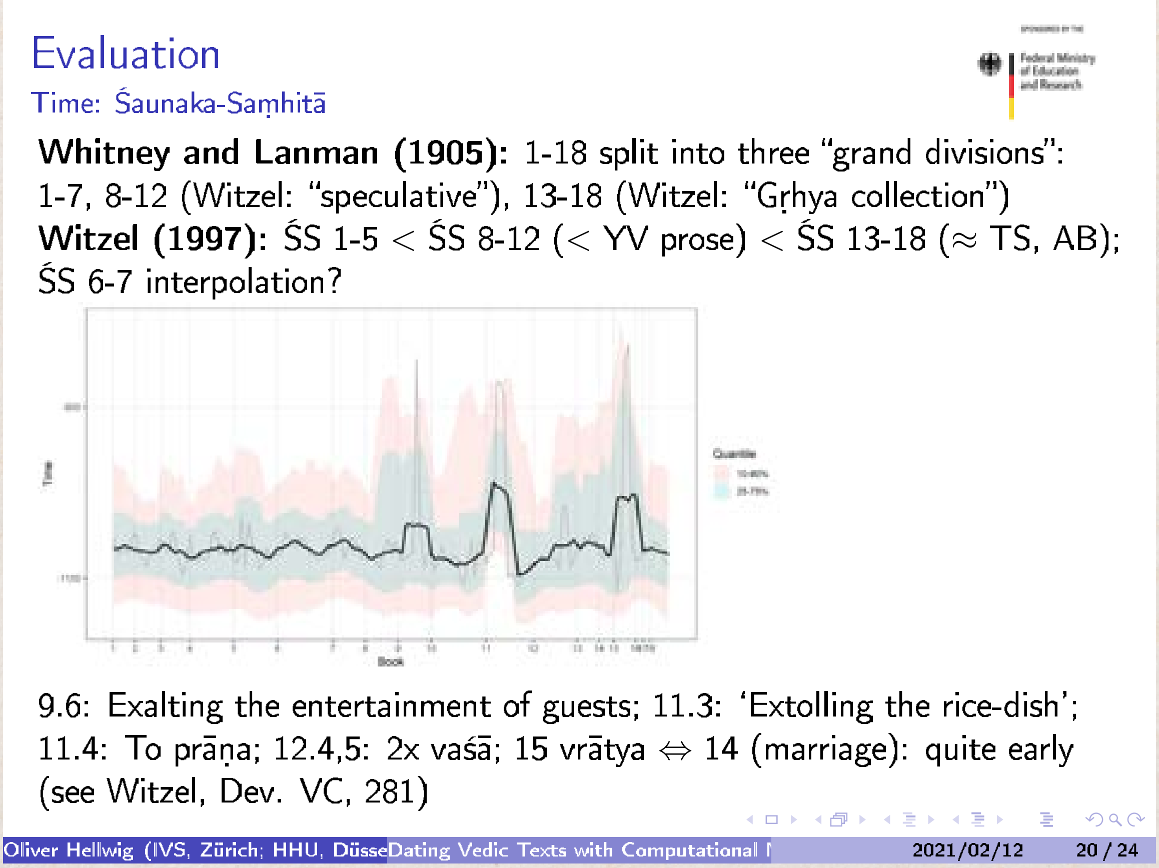
Time: Śaunaka-Saṃhitā

SPONSORED BY THE
Federal Ministry of Education and Research

Whitney and Lanman (1905): 1-18 split into three “grand divisions”: 1-7, 8-12 (Witzel: “speculative”), 13-18 (Witzel: “Gṛhya collection”)
Witzel (1997): ŚS 1-5 < ŚS 8-12 (< YV prose) < ŚS 13-18 (≈ TS, AB); ŚS 6-7 interpolation?

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational I 2021/02/12 20 / 24

23



Evaluation

Time: Śaunaka-Saṃhitā

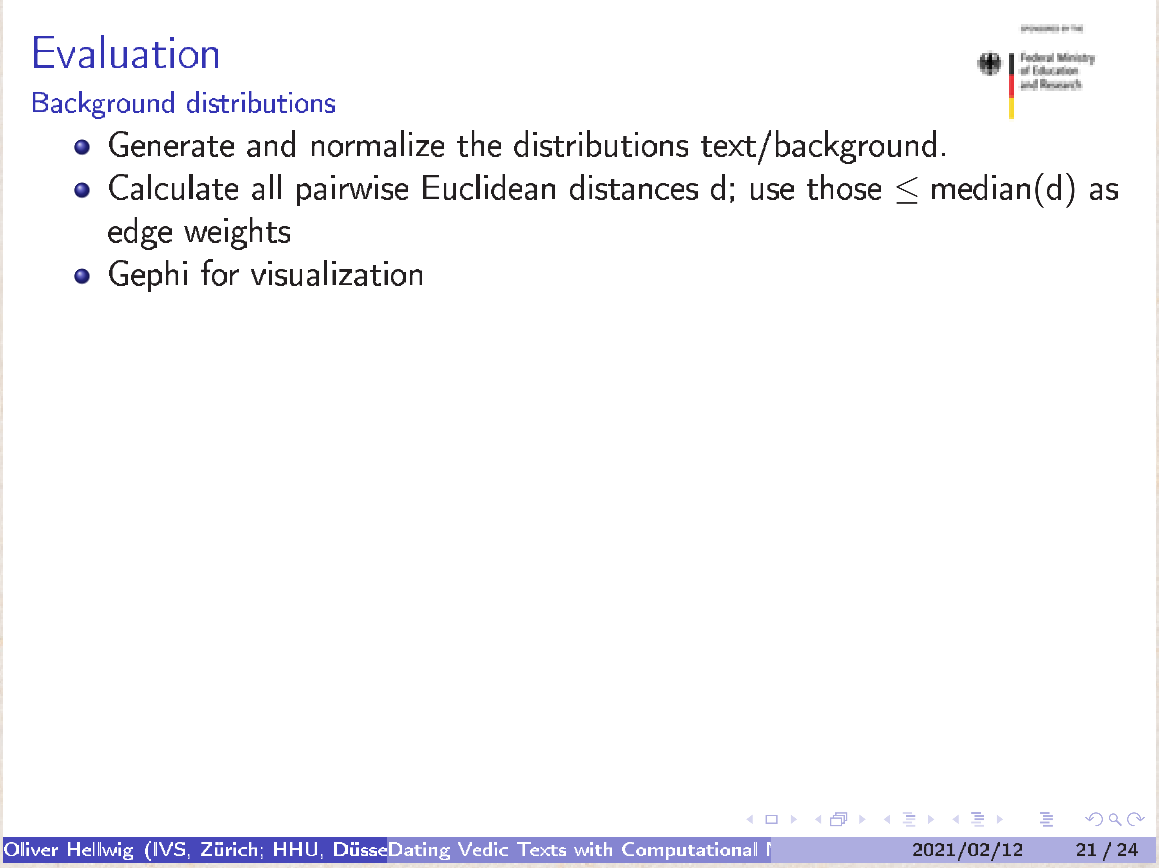
SPONSORED BY THE
Federal Ministry of Education and Research

Whitney and Lanman (1905): 1-18 split into three “grand divisions”: 1-7, 8-12 (Witzel: “speculative”), 13-18 (Witzel: “Gṛhya collection”)
Witzel (1997): ŚS 1-5 < ŚS 8-12 (< YV prose) < ŚS 13-18 (≈ TS, AB); ŚS 6-7 interpolation?

9.6: Exalting the entertainment of guests; 11.3: ‘Extolling the rice-dish’; 11.4: To prāṇa; 12.4,5: 2x vaśā; 15 vrātya ⇔ 14 (marriage): quite early (see Witzel, Dev. VC, 281)

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational I 2021/02/12 20 / 24

24



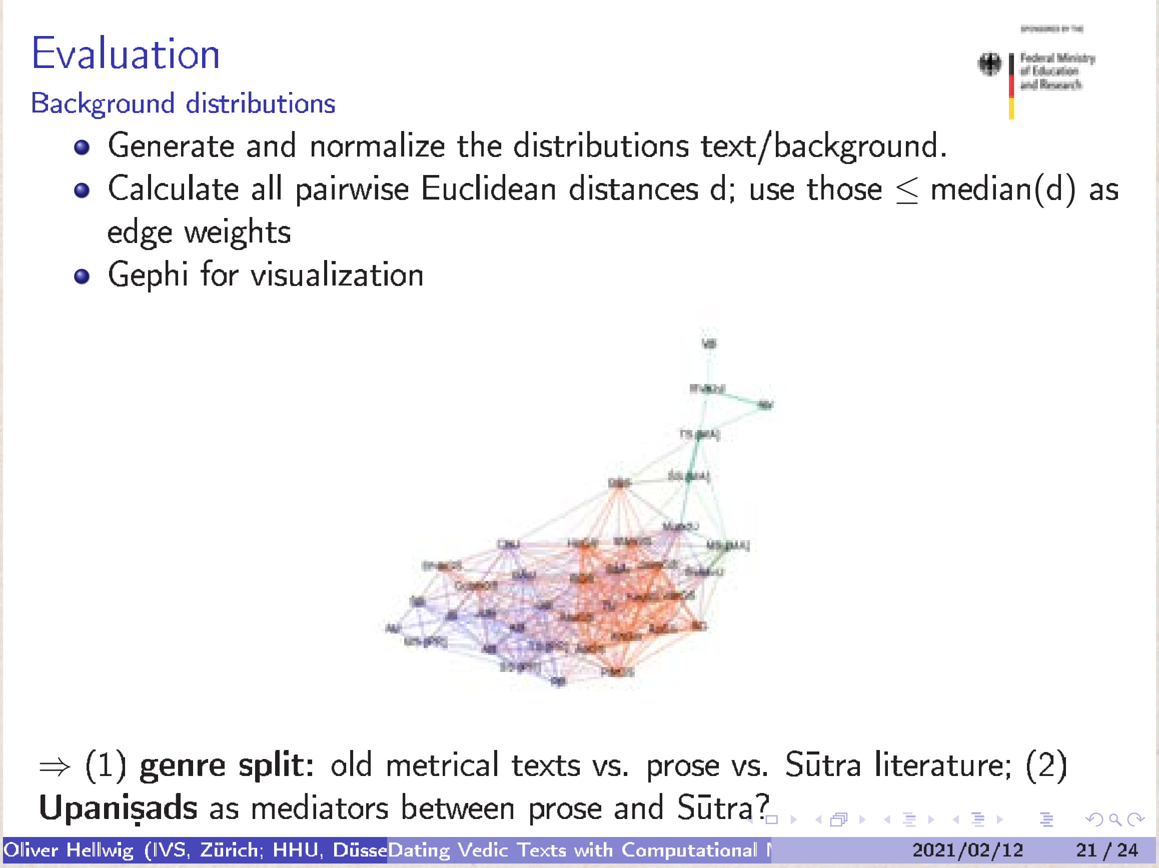
Evaluation

Background distributions

- Generate and normalize the distributions text/background.
- Calculate all pairwise Euclidean distances d; use those ≤ median(d) as edge weights
- Gephi for visualization

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 21 / 24

25



Evaluation

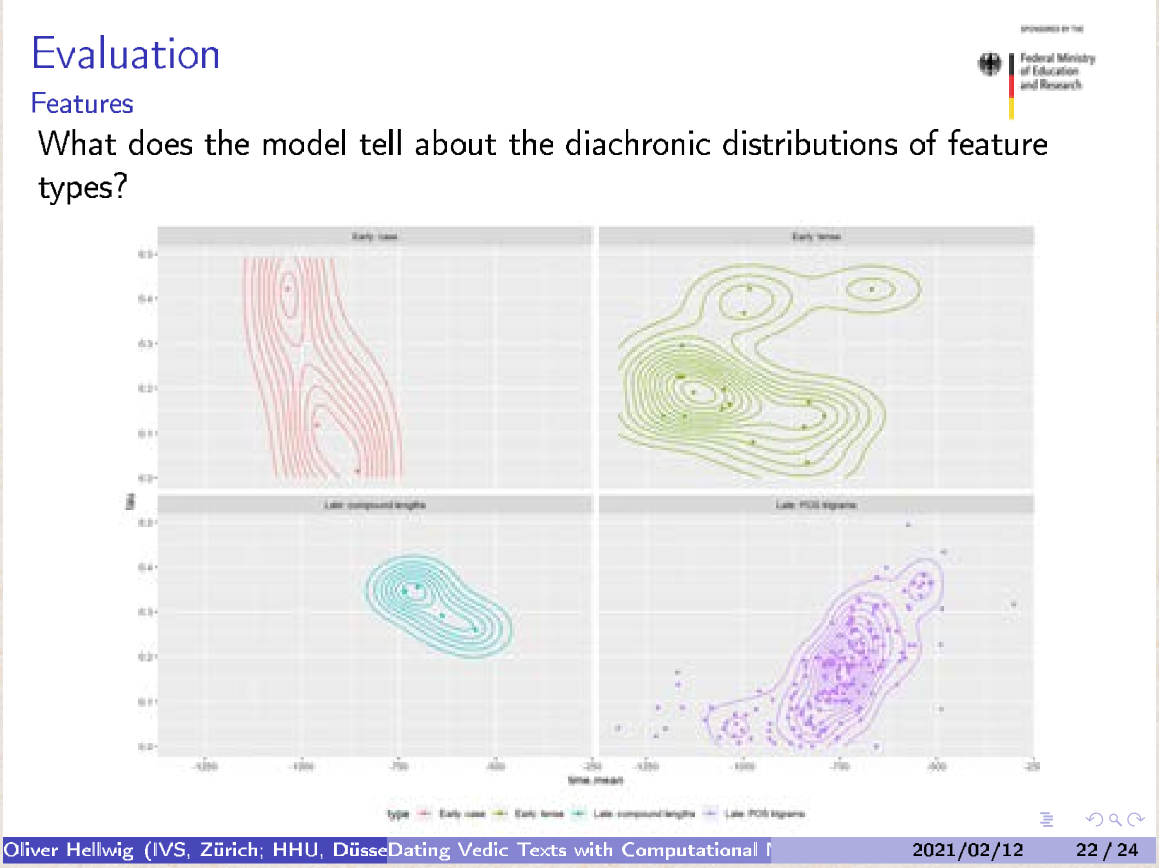
Background distributions

- Generate and normalize the distributions text/background.
- Calculate all pairwise Euclidean distances d; use those ≤ median(d) as edge weights
- Gephi for visualization

⇒ (1) **genre split:** old metrical texts vs. prose vs. Sūtra literature; (2) **Upaniṣads** as mediators between prose and Sūtra?

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational M 2021/02/12 21 / 24

26



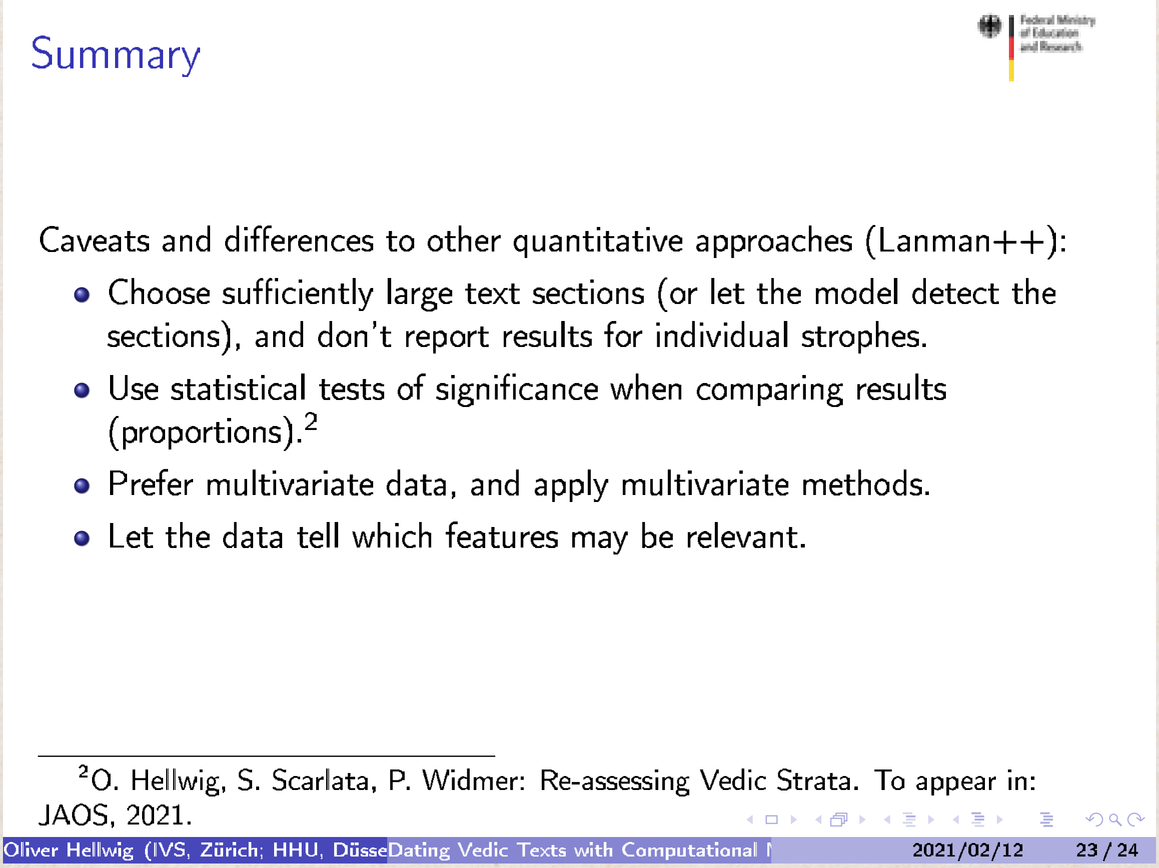
Evaluation

Features

What does the model tell about the diachronic distributions of feature types?

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational I 2021/02/12 22 / 24

27



Summary

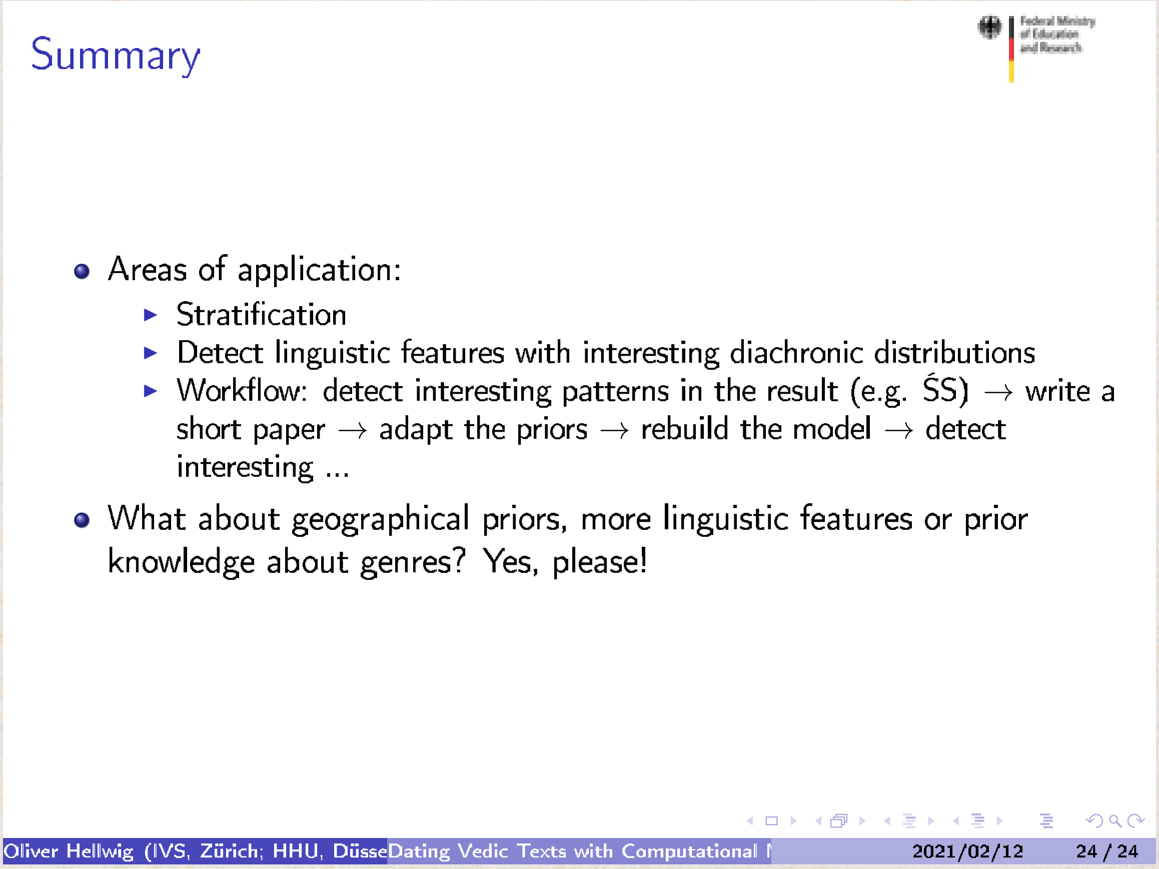
Caveats and differences to other quantitative approaches (Lanman++):

- Choose sufficiently large text sections (or let the model detect the sections), and don't report results for individual strophes.
- Use statistical tests of significance when comparing results (proportions).[2]
- Prefer multivariate data, and apply multivariate methods.
- Let the data tell which features may be relevant.

[2] O. Hellwig, S. Scarlata, P. Widmer: Re-assessing Vedic Strata. To appear in: JAOS, 2021.

Oliver Hellwig (IVS, Zürich; HHU, Düsse Dating Vedic Texts with Computational I 2021/02/12 23 / 24

28



Summary

- Areas of application:
  - Stratification
  - Detect linguistic features with interesting diachronic distributions
  - Workflow: detect interesting patterns in the result (e.g. ŚS) → write a short paper → adapt the priors → rebuild the model → detect interesting ...
- What about geographical priors, more linguistic features or prior knowledge about genres? Yes, please!

Oliver Hellwig (IVS, Zürich; HHU, DüsseDating Vedic Texts with Computational I 2021/02/12 24 / 24

Presentation 6 ｜報告 6

morogram: Background, History, and Purpose of a Tool for East Asian Text Analysis

morogram: 東アジア文献分析ツールの開発の経緯と目的

師茂樹（花園大学 文学部）

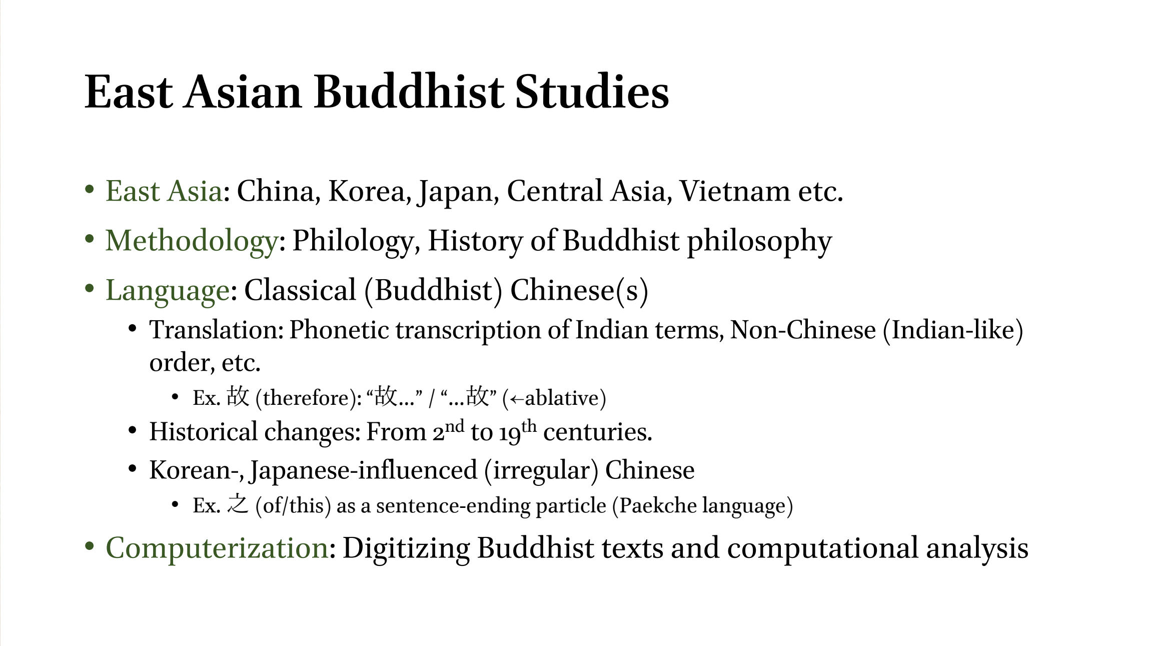
Shigeki Moro (Hanazono University, Faculty of Letter)

1

morogram: Background, History, and Purpose of a Tool for East Asian Text Analysis

Shigeki Moro (Hanazono University)

2



East Asian Buddhist Studies

- East Asia: China, Korea, Japan, Central Asia, Vietnam etc.
- Methodology: Philology, History of Buddhist philosophy
- Language: Classical (Buddhist) Chinese(s)
  - Translation: Phonetic transcription of Indian terms, Non-Chinese (Indian-like) order, etc.
    - Ex. 故 (therefore): “故…” / “…故” (←ablative)
  - Historical changes: From 2nd to 19th centuries.
  - Korean-, Japanese-influenced (irregular) Chinese
    - Ex. 之 (of/this) as a sentence-ending particle (Paekche language)
- Computerization: Digitizing Buddhist texts and computational analysis

3

East Asian Text Analysis based on N-gram model

- N-gram: a sequence of *n* characters or words
- Frequency of N-gram: Characteristics of a language, text, author etc.
- Applications: Index for Full-text search engines
- History:
  - Miyuki & Yasuhiro Kondo (2000–)
  - Kosei Ishii & Moro: development of *morogram* & NGSM (N-Gram based System for Multiple document comparison and analysis) tools (2002–)
    - Now: https://github.com/moroshigeki/ngsm

4

Tri-gram (3-gram) of 摩訶般若波羅蜜多心経

摩	**訶**	**般**	若	波	羅	蜜	多	心	経
摩	**訶**	**般**	**若**	波	羅	蜜	多	心	経
摩	訶	**般**	**若**	**波**	羅	蜜	多	心	経
摩	訶	般	**若**	**波**	**羅**	蜜	多	心	経
摩	訶	般	若	**波**	**羅**	**蜜**	多	心	経
摩	訶	般	若	波	**羅**	**蜜**	**多**	心	経
摩	訶	般	若	波	羅	**蜜**	**多**	**心**	経
摩	訶	般	若	波	羅	蜜	**多**	**心**	**経**

摩訶	mahā
般若	prajñā
波羅蜜多	pāramitā
心	hṛdaya
経	sutra

(*Heart sutra*)

5

Kondo Miyuki (1960–2019)

- *Ōchō waka kenkyū no hōhō* [Methodology of the study of Dynastic *waka* poems] (2015)
  - *waka*: a traditional Japanese poem of thirty-one syllables.

- Gender analysis of Japanese classical literatures using N-gram

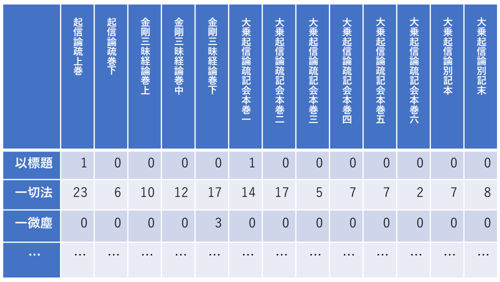
6

Purposes of N-gram Analysis

- Statistical analysis to make a hypothesis for philological studies
  - Cf. “distant reading” (Franco Moretti)
- Searching for hidden characteristics
- Searching for hidden (silent) quotations

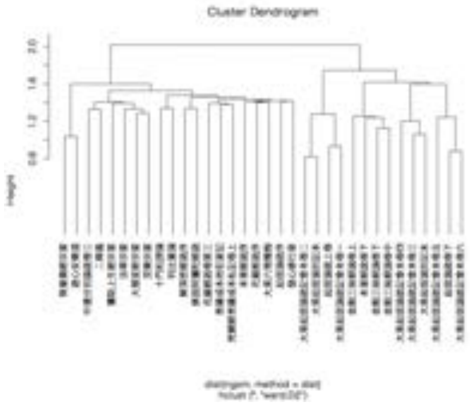
7

Making a hypothesis for philological studies



	起信論疏上巻	起信論疏巻下	金剛三昧経論巻上	金剛三昧経論巻中	金剛三昧経論巻下	大乗起信論疏記会本巻一	大乗起信論疏記会本巻二	大乗起信論疏記会本巻三	大乗起信論疏記会本巻四	大乗起信論疏記会本巻五	大乗起信論疏記会本巻六	大乗起信論別記本	大乗起信論別記末
以標題	1	0	0	0	0	1	0	0	0	0	0	0	0
一切法	23	6	10	12	17	14	17	5	7	7	2	7	8
一微塵	0	0	0	0	3	0	0	0	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…	…

NGSM table of Wonhyo 元暁 (617-686)

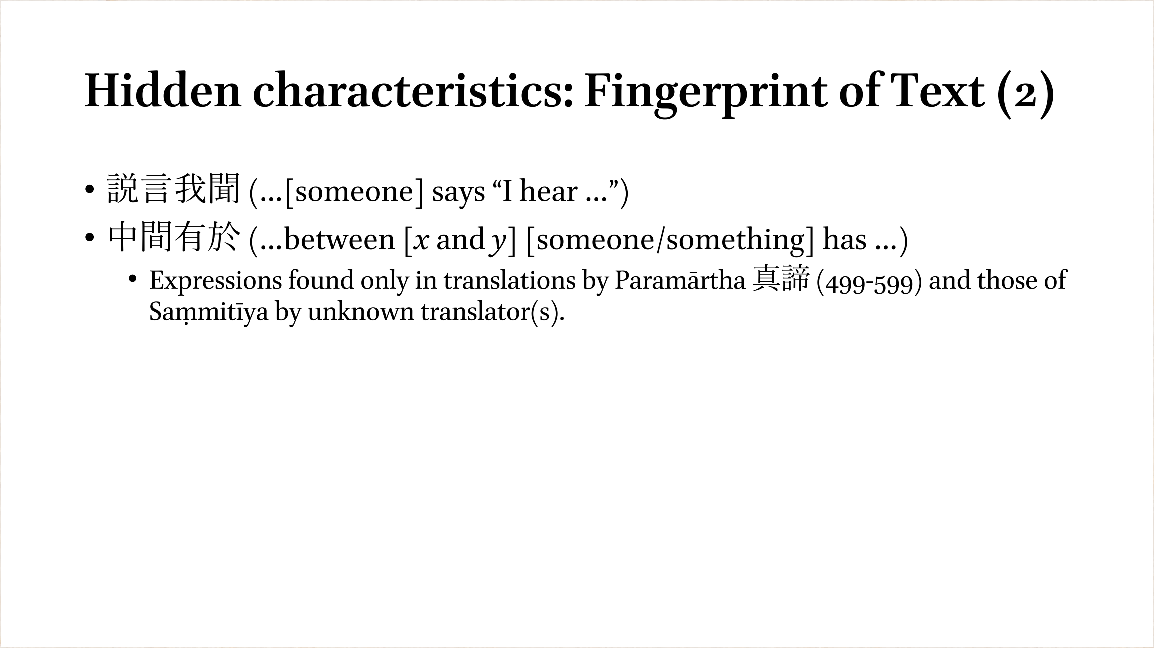


8

Hidden characteristics: Fingerprint of Text (1)

- “What would a stylistic fingerprint be? It would be a feature of an author’s style––a combination perhaps of very humble features such as the frequency of *such as*––no less unique to him than a bodily fingerprint is. Being a trivial and humble feature of style would be no objection to its use for identification purposes: the whorls and loops at the ends of our fingers are not valuable or striking parts of our bodily appearance” (Kenny 1982: 12-13).
- “Comprehensive research of a text will allow us to explore the structure and regularity of the text, which is usually unrecognizable to modern people. This will contribute to the study of literature by making up for our lack of ‘introspection’ (word sense) towards classical languages.” (Kondo, Y. 2001)

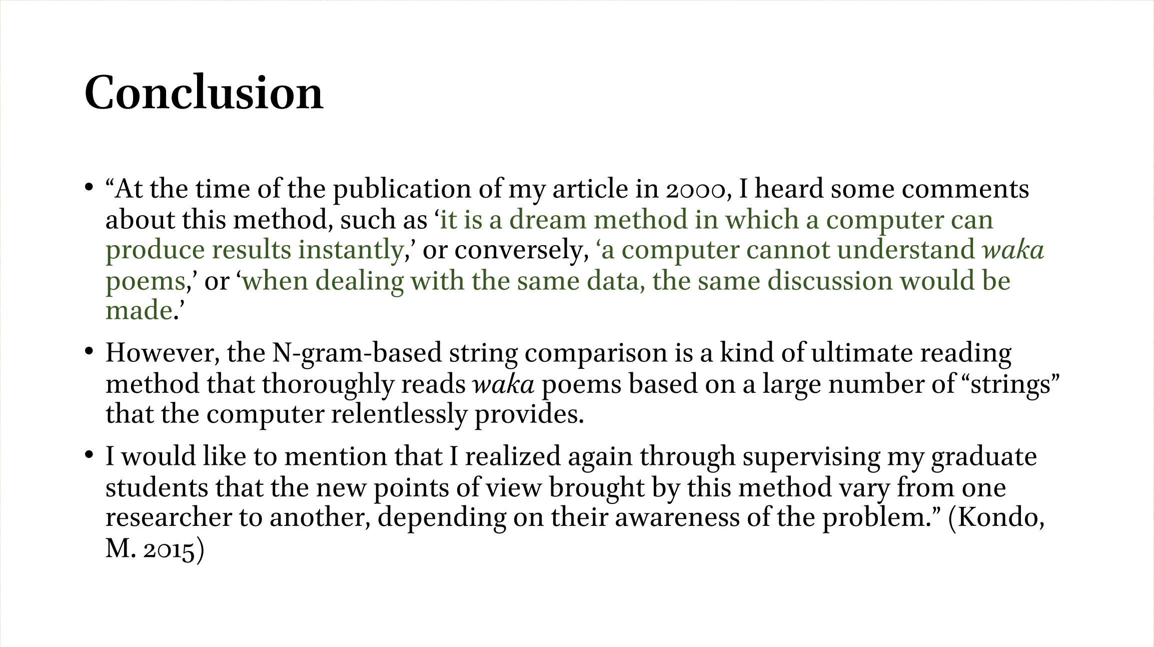
9



Hidden characteristics: Fingerprint of Text (2)

- 說言我聞 (...[someone] says “I hear ...”)
- 中間有於 (...between [*x* and *y*] [someone/something] has ...)
    - Expressions found only in translations by Paramārtha 真諦 (499-599) and those of Saṃmitīya by unknown translator(s).

10



Conclusion

- “At the time of the publication of my article in 2000, I heard some comments about this method, such as ‘it is a dream method in which a computer can produce results instantly,’ or conversely, ‘a computer cannot understand *waka* poems,’ or ‘when dealing with the same data, the same discussion would be made.’
- However, the N-gram-based string comparison is a kind of ultimate reading method that thoroughly reads *waka* poems based on a large number of “strings” that the computer relentlessly provides.
- I would like to mention that I realized again through supervising my graduate students that the new points of view brought by this method vary from one researcher to another, depending on their awareness of the problem.” (Kondo, M. 2015)

Notes and references for presentations by Kyoko Amano

Dynamism of Social Context Deciphered by a Linguistic Analysis of Ancient Literature
February 12, 2021. Kyoto University (online).

The first workshop of SPIRITS project "Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science"
https://ancientindia-datascience.hakubi.kyoto-u.ac.jp/en/index-en/

Opening：Problems in the Formation of the Vedas, Ancient Indian Religious Texts
Kyoko Amano (Kyoto University, Institute for Research in Humanities / Hakubi Center)

Contents:
1. Background for the Joint Research of Vedic Text and Data Science
2. Overview on Vedic Texts, the Subject of This Research;
   Period and Geographical Localization by Witzel, *Tracing Vedic Dialects*
3. New Perspectives in Considering the Compilation Process of Vedic Texts

1.
references:
Waves of immigration of Indo-Aryan people:

Witzel, Michael (1989a). Tracing the Vedic dialects. In: *Dialectes dans les littératures indo-aryennes : Actes du Colloque International (Paris, 16-18 septembre 1986)*, ed. by Colette Caillat, 97-264. Paris: Collége de France, Institut de Civilisation Indienne. Maps on p. 233-234.

Linguistic Analysis and textual layers:

Kobayashi, Masato (2012). "Information Structure and the Particles vái and evá in Vedic Prose". In: *Indic across the Millennia. From the Rigveda to Modern Indo-Aryan.* 14th World Sanskrit Conference, Kyoto, Japan, September 1st-5th, 2009. Proceedings of the Linguistic Section, ed. by J. S. Klein / K. Yoshida, Kyoto, 77-92.

Andrijanic̄, Ivan (2013). Historical Analysis of Textual Layers in Ancient Indian LIterature and Indian Cultural History. In: *CEENIS Current Research Series* vol. 1, 21-43.

Hellwig, Oliver. (2008). "Frequent phrases and their Application to Text Segmentation". In: *Studien zur Indologie und Iranistik* 25, S. 55–72.

Hellwig, Oliver (2016). "A Computational Approach to the Text History of the Rāmāyaṇa". In: *Proceedings of the DICSEP 2008*. Hrsg. von Ivan Andrijanić und Sven Sellmer. Zagreb: Croatian Academy of Sciences und Arts, S. 41–62.

Database of Vedic texts with annotation in Digital Corpus of Sanskrit
http://www.sanskrit-linguistics.org/dcs/index.php?contents=texte

Our visual data of mantra collocation (http://34.84.105.185/) is based on

Bloomfield, Maurice (1893). *A Vedic Concordance*. [Harvard Oriental Series 10]. Cambridge - Mass.

The expanded edition (with electronic data) was used for this research:

Franceschini, Marco (2007). *An updated Vedic concordance : Maurice Bloomfield's A Vedic concordance enhanced with new material taken from seven Vedic texts*. Cambridge: Dept. of Sanskrit and Indian Studies, Harvard University.

2.
About Vedic texts:
see Witzel (1989a).
Witzel, Michael (1997). "The Development of the Vedic Canon and its Schools: The Social and Political Milieu. (Materials on Vedic Sakhas 8)." In: I*nside the Texts, Beyond the Texts. New Approaches to the Study of the Vedas.* Harvard Oriental Series. Opera Minora, vol. 2. Cambridge, 257-345

Localization of Vedic texts:
Witzel (1989a), 110 with n. 34 indicates that a few tentative localizations had been made by Weber, Caland and others.
Other Witzel's works are the followings:
—— (1987). "On the localisation of Vedic texts and schools (Materials on Vedic sakhas, 7)." In: *India and the Ancient world. History, Trade and Culture before A.D. 650.* P.H.L. Eggermont Jubilee Volume, ed.by G. Pollet, Orientalia Lovaniensia Analecta 25, Leuven, pp. 173-213,
—— (1989b). "The Realm of the Kurus: Origins and Development of the First State in India." *Nihon Minami Ajia Gakkai Zenkoku Taikai, Hokoku Yoshi*, [Summaries of the Congress of the Japanese Association for South Asian Studies], Kyoto 1989, pp. 1-4
—— (1997). "Early Sanskritization. Origins and development of the Kuru State." B. Kölver (ed.). *Recht, Staat und Verwaltung im klassischen Indien. The state, the Law, and Administration in Classical India.* München : R. Oldenbourg 1997 : 27-52

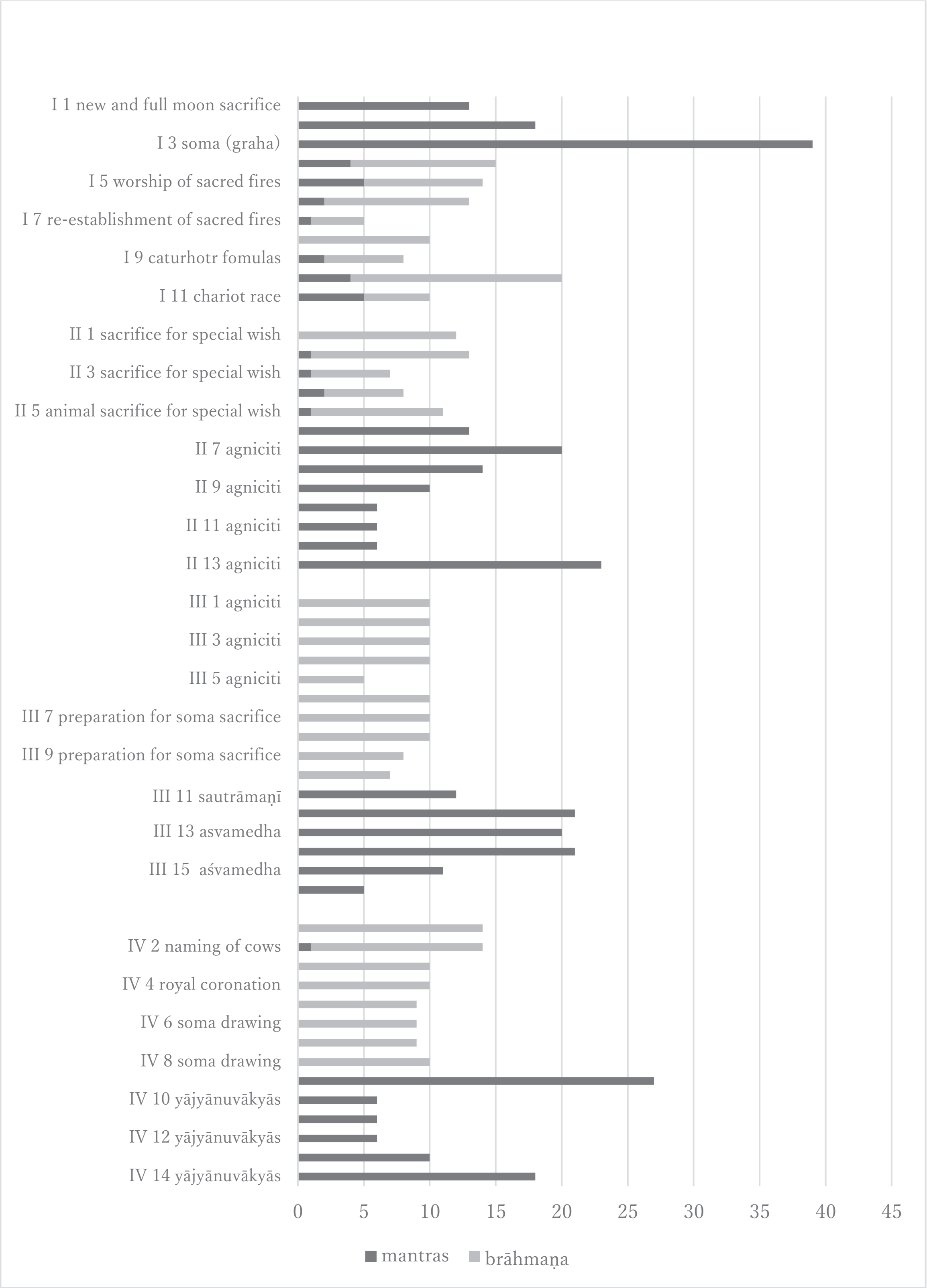
3.
To linguistic layers in the Maitrāyaṇī Saṁhitā:
Amano, Kyoko (2014-2015). "Zur Klärung der Sprachschichten in der Maitrāyaṇī Sam̊hitā." *Journal of Indological Studies* 26/27: 1-36.
——— (2015). "Style and Language of the *Agniciti* Chapter in the *Maitrāyaṇī Sam̊hitā* (III 1-5)." *Journal of Indian and Buddhist Studies* 63-3: 1161-1167.
——— (2016a). "Saishiki wo urazukeru chishiki wo megutte." *Machikaneyamaronso* 50 (Philosophy): 29-56.
——— (2016b). "Indication of Divergent Ritual Opinions in the Maitrāyaṇī Sam̊hitā." In *Vedic Śākhās: Past, Present, Future. Proceedings of the Fifth International Vedic Workshop, Bucharest 2011*, ed. by J. E. M. Houben, J. Rotaru and M. Witzel, 461-490. Cambridge: Harvard University Press.
——— (2016c). "Ritual Contexts of *Sattra* Myths in the Maitrāyaṇī Sam̊hitā." In *Vrātya culture in Vedic sources. Select Papers from the Panel on "Vrātya culture in Vedic Sources" at the 16th World Sanskrit Conference (28 June - 2 July 2015) Bangkok,* by Tiziana Pontillo, Moreno Dore and Hans Heinrich Hock, 35-72. Bangkok: DK Publishers.
——— (2017). "A Ritual Explanation Concealing its Name. Maitrāyaṇī Sam̊hitā I 9 (*caturhotr̥* chapter)". *Journal of Indian and Buddhist Studies* 65-3, 1039-1046 (1)-(8).
——— (2019a). "A Non-Śrauta Ritual in the Oldest Yajurveda Text. Maitrāyaṇī Sam̊hitā IV 2 (Gonāmika Chapter)." In *Proceedings of the 17th World Sanskrit Conference, Vancouver, Canada, July 9-13, 2018, Section 1: Veda,* ed. Bahulkar, Sh., Jurewicz, J., 1-27. Published by the Department of Asian Studies, University of British Columbia, on behalf of the International Association for Sanskrit Studies. DOI: 10.14288/1.0379840. URI: http://hdl.handle.net/2429/70986.
——— (2019b). "The Development of the Uses of *ha / ha vái / ha sma vái* with or without the Narrative Perfect and Language Layers in the Old Yajurveda-Sam̊hitā Texts." *Lingua Posnaniensis* 61, ed. by Chandotti, M. P. / Pontillo, T. Sciendo: Warszawa. 11-24.

——— (2020). "What is ‘knowledge’ justifying a ritual action? Uses of *yá evám̐ véda / yá evám̐ vidvā́n* in the Maitrāyaṇī Sam̐hitā.” In *Aux sources des liturgies indo-iraniennes*, ed. by Redard, C. / Ferrer-Losilla, J. / Moein, H. / Swennen, P., Collection Religions, Comparatisme - Histoire - Anthropologie 10, 39-68. Liège: Presses Universitaeire de Liège.

——— (forthcoming1). "*etád vā́ eṣā́bhyánūktā* in the Maitrāyaṇī Sam̐hitā. The Beginning of Didactical Verse Embedded in Narrative Prose." *Myth, Language, and Prehistory: A Celebratory Conference in Honor of Prof. Michael Witzel*, 2019 Sep. 8, Harvard University.

Mantra parts and brāhmaṇa parts in the MS:

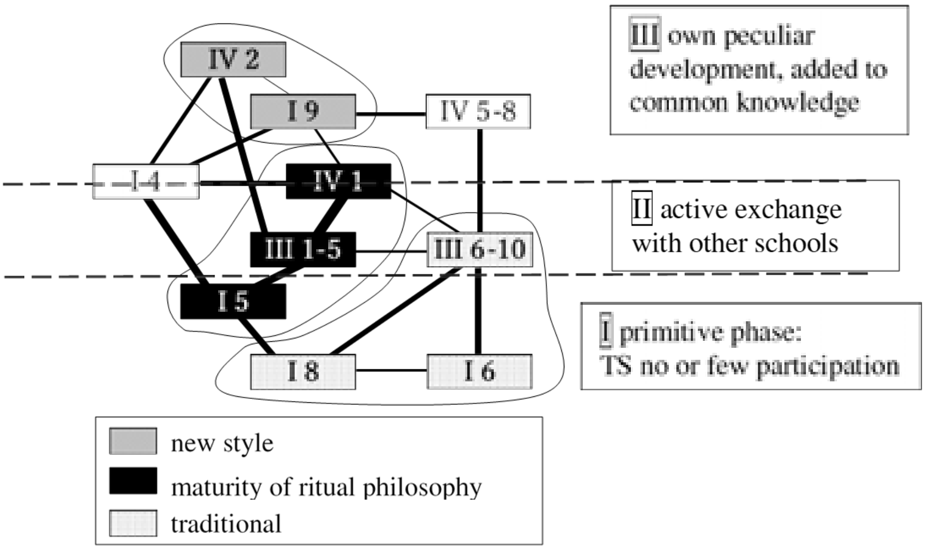
Amano (forthcoming2) "Composition of the Mantra Parts in the Maitrāyaṇī Sam̐hitā". 7th International Vedic Workshop Dubrovnik, 2019/8/20. Inter-University Centre, Centre for Advanced Academic Studies Dubrovnik, and
Amano (forthcoming3) "Influence from the Atharvaveda on Rituals in the Maitrāyaṇī Sam̐hitā." The Atharvaveda and its South Asian Contexts: 3rd Zurich International Conference on Indian Literature and Philosophy (ZICILP), 2019/9/27. University of Zurich.


I 1 new and full moon sacrifice
I 3 soma (graha)
I 5 worship of sacred fires
I 7 re-establishment of sacred fires
I 9 caturhotr fomulas
I 11 chariot race
II 1 sacrifice for special wish
II 3 sacrifice for special wish
II 5 animal sacrifice for special wish
II 7 agniciti
II 9 agniciti
II 11 agniciti
II 13 agniciti
III 1 agniciti
III 3 agniciti
III 5 agniciti
III 7 preparation for soma sacrifice
III 9 preparation for soma sacrifice
III 11 sautrāmaṇī
III 13 asvamedha
III 15 aśvamedha
IV 2 naming of cows
IV 4 royal coronation
IV 6 soma drawing
IV 8 soma drawing
IV 10 yājyānuvākyās
IV 12 yājyānuvākyās
IV 14 yājyānuvākyās
0
5
10
15
20
25
30
35
40
45
mantras
brāhmaṇa

brāhmaṇa chapters in MS and their parallels in KS and TS (Amano 2019b, 14f.):

brāhmaṇa chapters in MS			KS/TS parallels	
			KS	TS
I 4	yajamāna	duty of a sacrificer	32	I 6-7
I 5	agnyupasthāna	worship of sacred fires	7	I 5
I 6	ādhāna	establishment of sacred fires	8	
I 7	punarādhāna	re-establishment of sacred fires	9	I 5
I 8	agnihotra	daily offering to sacred fires	6	
I 9	caturhotr̥	caturhotr̥ formulas	9	
I 10	cāturmāsya	seasonal rites	36	
I 11	vājapeya	soma drinking for winning a chariot race	14	
II 1-4	kāmyā-iṣṭi	rites for special wish (with cake and gruel)	10-12	II 2-4
II 5	kāmya-paśu	rites for special wish (with sacrificial animal)	13	II 1
III 1-5	agniciti	piling of fire altar	19-22	V 1-7
III 6-10	soma adhvara	preparation for soma ritual	23-26	VI 1-3
IV 1	darśapūrṇamāsa	new and full moon sacrifice	31	
IV 2	gonāmika	rite for naming cows		
IV 3-4	rājasūya	royal coronation		
IV 5-8	soma graha	soma drawing	27-30	VI 4-6

Process of composition of the brāhmaṇa parts in the Maitrāyaṇī Saṁhitā (Amano forthcoming1):


IV 2
I 9
IV 5-8
I 4
IV 1
III 1-5
III 6-10
I 5
I 8
I 6
III own peculiar development, added to common knowledge
II active exchange with other schools
I primitive phase: TS no or few participation
new style
maturity of ritual philosophy
traditional

Three new perspectives:

(1) MS and RV / AV

The consensus had been that RV and Atharvaveda had been compiled in their currently existing form before Yajurveda. However, careful consideration of RV and Atharvaveda hymns quoted in

Yajurvedas shows consistency with RV and AV at a different level according to chapter in Yajurvedas. They are likely to indicate the degree of the spread of completed books of RV and learning system, or variances in their loyalty to RV. Perhaps the old layers had not fully mastered RV in its entirety (maybe only hymns used in rituals had been known according to their use; or perhaps some chapters trace RV while deliberately avoiding direct quotations). There are also significant variances between chapters on the knowledge of AVŚ and AVP in MS. MS chapters may have only had partial knowledge of AVŚ and AVP. Quotations from AV in MS are sporadic in the older layers, and large amounts of quotations are seen only in the additional parts at the end of each chapter, which may mean that the MS got the full-fledged knowledge about AV in the newer layers of its editing process.
See Amano (forthcoming 2 and 3)

Different grades of conformity with the RV and the AV in different layers in the MS:
table 3: Number of cited verses and conformity[1] of citations according to the chronological classification of chapters (Amano forthcoming 3 "Influence from the Atharvaveda ..."):

		RV		AVŚ		AVP	
		number of citations	conformity	number of citations	conformity	number of citations	conformity
I 1-II 6	old chapters	123 (+21)	78%	**54** (+25)	52%	21 (+17)	62%
II 7-12	agniciti	134 (+13)	72%	38 (+17)	56%	**61** (+16)	51%
II 13	agniciti additional	43 (+23)	**88%**	11 (+29)	**74%**	**14** (+8)	48%
III 11-13	new mantras (sautrāmaṇī and aśvamedha)	11 (+6)	77%	8 (+6)	56%	2 (+1)	---
III 16 + IV 9	additional to aśvamedha + pravargya	42 (+22)	**86%**	10 (+22)	**60%**	22 (+17)	49%

MS II 13, III 16 and IV 9 indicate high grade of conformity with RV, probably the period with good learning system of RV, and these chpaters show also high grade of conformity with AVŚ (not higher than RV). MS II 7-13, mantras for agniciti, include more citations from AVP than other chapters in MS do.

<u>(2) Mantras and brāhmaṇas in the Yajurveda-Saṁhitās:</u>
The Yajurveda-Saṁhitās contains different styles and categories of mantras and interpretations of rituals (brāhmaṇa). Conventional belief had been that mantras were old (before 800 to 900 BC), and brāhmaṇa chapters were newer than mantras. In other words, eras were discussed by a breakdown of two layers: mantra and brāhmaṇa. However, recent studies have pointed out that mantra chapters contain both new and old texts and that not all mantras are older than Brahmana. Examples for probably new mantra chapters include MS II 9 and II 13 (part of the Agniciti mantra), III 11 (the Sautrāmaṇī mantra), and IV 9 (the Pravargya mantra). They may be considered to have been added after the brāhmaṇ chapters s had been created. The compilation process of the Yajurveda-Saṁhitās is not a two-layer structure of mantras and brāhmaṇas; we should eliminate the premise that mantras are older than brāhmaṇas. We should position all mantra and brāhmaṇa chapters in the same way somewhere in the compilation process.

[1] Here I valued the grade of conformity with the following scoring: same verses × 5p + slightly varying verses × 3p + varying verses × 1p / full scores (all verses × 5p).

About the "new" mantras in the MS see Amano (2016c) and (forthcoming3).

(3) MS, KS and TS developed from a prototype text?
The three Black Yajurveda-Saṁhitās, MS, KS, and TS had been believed to have branched out from a single prototype, with MS-KS and TS first being separated, followed by MS and KS branching out. The compilation order of texts had been believed to be MS, KS, then TS. But recent research has indicated that this phylogenetic tree model is not appropriate for expressing the compilation of these three texts. The argument is that MS and KS show the low rates of similarities in the older layers and the high rates in the newer layers, where much borrowing appears to occur. We may presume that some chapters had been individually written, and some in the networks where the authors of the three texts shared their philosophies and rites. That reflects the change of relationship that occurred in real time of compilation. MS and KS have been considered closely related, but it has been revealed that KS and TS had became closer since a certain period.

To closer relationship of KS with TS than with MS see
Izawa, Atsuko (2009. "The Position of the Kāṭhaka Saṃhitā - Kapiṣṭhalakaṭha Saṃhitā among the Black Yajur Veda Saṃhitās in the Section about the Brick-piling of the Fourth Layer of the Agnicayana". 14th World Sanskrit Conference, Kyoto, handout.

Three time periods with change of relationship among the Yajurveda-Saṁhitās:

Time period I: MS and KS began the compilation of the texts. The oldest chapters are MS I 6 ~ KS 6 (ādhāna) and MS I 8 ~ KS 8 (agnihotra). At this point, TS was not included in the movement of text compilation. MS I 1-3 ~ KS 1-4 (mantras for full and new moon rituals, soma preparation ceremony, and soma ritual) may have been compilated around this time. (Perhaps TS might have extracted them during the time period II.) The chapters of cāturmāsya and vājapeya in MS (I 10 and 11) also could have been compilated around this time.

Time period II: This was the era that TS joined MS and KS, and rituals were developed among the group. The center of this movement was the agniciti ritual. KS took similar measures to TS.

Time period III: The phase of globalization began and RV vulgata had spread. Since then, each school started local diverging more strongly. (Mahadevan "Vedic Big Bang") MS added more chapters. Hymns from RV were collectively added at the last part of MS. KS and TS added hymns from RV by inserting them between the chapters. KS added missing chapters. KS-TS added the original chapter (sattra chapter). TS added its own original chapter.

Note to the compilation of Black Yajurveda-Saṁhitās:
Three Black Yajurveda-Saṁhitās, or if we think about the origin and the process of the composition of the texts, we should consider it with the forth Black Yajurveda school, namely the Carakas. I am thinking that it may be possible that MS, KS and TS did not always have direct exchange with each other, but there was an intermediation that brought information about (new) rituals to MS, which was the Carakas.
To Caraka school and an unknown Yajurveda-Saṁhitā:
Witzel, Michael (1981). "Materialien zu den vedischen Schulen: I. Über die Caraka-Schule." *StII* 7, 109-132. [=Pt.1: History of the Caraka School].
—— (1982). "Über die Caraka-Schule [continuation. of No.14: ch.2-4: position of the school, texts, present state of this lost Sakha]." *StII* 8/9 (1982), 171-240
—— (1984). "An unknown Yajurveda-Samhita (Materials on Vedic sakhas, 6)." *IIJ* 27, 105-106.

Relationship among Vedic Schools Deciphered by the Visualization of Mantra Collocation

Contents:

1. About Mantras
2. Characteristics of Four Yajurveda-Saṁhitās
    2.1 Relationship with the R̥gveda
    2.2 Relationship with the Atharvaveda (Paippalāda and Śaunaka)
3. New Points of View about Relationship of the Texts and Schools while Composing the Texts

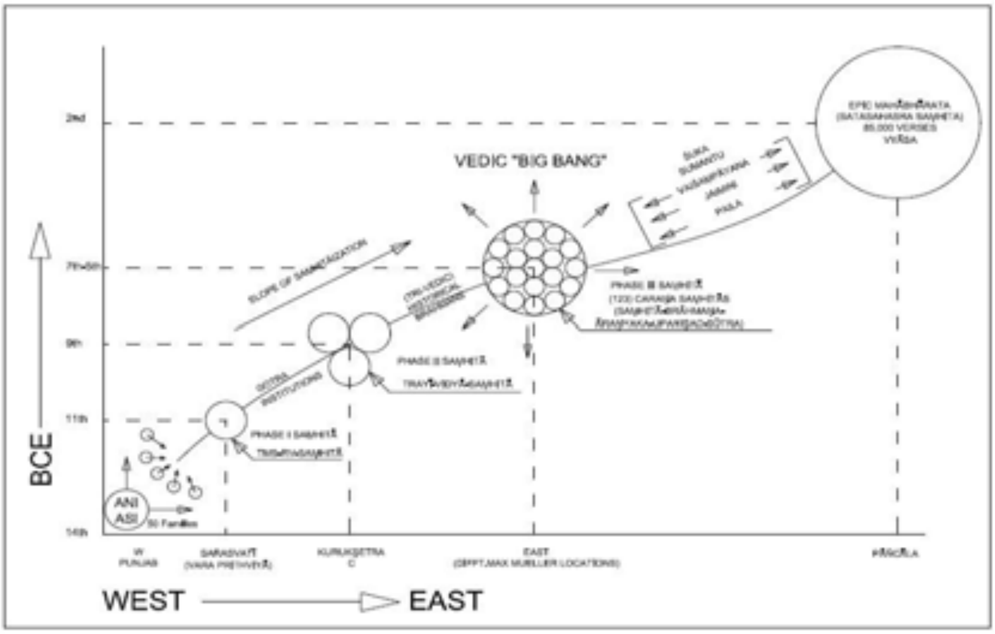
2.

History of RV, AV and YV:

"Vedic Big Bang" by Prof. Mahadevan:

Mahadevan, Thennilapuram Parasuram Iyer (2019). "The Indo-European oral tradition, the Śravas akṣitam, from its first appearance in Punjab, ca. 15th century, to the 3rd century BCE in the Pāncāla country." presentation at the 7th International Vedic Workshop, August 19–24, 2019 Inter-University Centre, Centre for Advanced Academic Studies, Dubrovnik.

Slope of Saṃhitā-ization



slide from his presentation materials that Prof. Mahadevan kindly provided for me.

2.1.

To RV and MS, KS, TS, especially about KS 38-40:

Although the relationship with RV reappears in the last portion of KS (38-40), this part is unusual and un-Vedic even within the mantras of agniciti. The agniciti mantras are collected in the 15 to 18 in KS, so this portion is set apart. From its segregated location, it is reasonable to assume that this portion was added later. The correspondence of MS regarding this portion is incorporated into II 7-13, chapters of the agniciti mantras. The correspondence graph of KS and MS shows this clearly. When we look at the correspondence of KS and TS, mantras that correspond to KS 38-40 can be seen in TS 4 that is for the agniciti mantras. Same as MS, the (possibly) new mantras are incorporated into the original collections of mantras. Scrutinizing the meaning of mantras and their use in rituals is necessary to understand how to explain these contexts, so we'd like to address the topic at our next opportunity.

possible chronological evidence for the chapters for agniciti mantras and aśvamedha mantras in the MS is the mention of a new name of a season or a month *mádhu-* and *mā́dhava-*: in I 3,16:36,9, II 8,12:116,3, III 12,13:164,5, III 16,4:187,14 IV 6,7:89,6.

京都大学
KYOTO UNIVERSITY

SPIRITS
SUPPORTING PROGRAM FOR INTERACTION-BASED INITIATIVE TEAM STUDIES

Dynamism of Social Context Deciphered by a Linguistic Analysis of Ancient Literature

The first workshop of SPIRITS project
"Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science"
2020-2021 Interdisciplinary type project, in the priority area of humanities and social sciences
SPIRITS: Supporting Program for Interaction-based Initiative Team Studies

Friday, February 12, 2021 | 14 : 00 ~ 19 : 10

The workshop will be held online and in English.

Understanding the social background of text formation is a basic requirement to accurately understand documents. However, the background of ancient societies is often hidden in a veil of mystery, which makes it difficult to understand the process of text formation. The Vedas, religious texts in Ancient India, are among these documents. In this workshop, we will seek to decipher the social movements, geographical mobility, and change in the spheres of influence in ancient India through a language analysis of the Vedic texts. The discussion will address the question how quantitative methods and data science can be applied to this field.

Part1	14:00 ~ 14:30	**Opening :** Problems in the Formation of the Vedas, Ancient Indian Religious Texts **Kyoko Amano** (Kyoto University, Institute for Research in Humanities / Hakubi Center)
	14:30 ~ 15:10	The Possibility of Information Visualization and Data Analysis for Ancient Indian Literature **Hiroaki Natsukawa** (Kyoto University, Academic Center for Computing and Media Studies)
	15:10 ~ 15:50	Relationship among Vedic Schools Deciphered by the Visualization of Mantra Collocation **Kyoko Amano** (Kyoto University, Institute for Research in Humanities / Hakubi Center)
	15:50 ~ 16:30	Citation Prediction Using Academic Paper Data and Application for Surveys **Shun Hamachi** (Kyoto University, Graduate School of Engineering)
Part2	16:50 ~ 17:30	Measuring the Semantic Similarity between the Chapters of Taittirīya Saṃhitā Using a Vector Space Model **Yuki Kyogoku** (Leipzig University, Indology)
	17:30 ~ 18:10	Dating Vedic Texts with Computational Models: Algorithmic Considerations and Data Selection **Oliver Hellwig** (University of Zurich, Department of Comparative Language Science)
	18:10 ~ 18:50	morogram: Background, History, and Purpose of a Tool for East Asian Text Analysis **Shigeki Moro** (Hanazono University, Faculty of Letter)
	18:50 ~ 19:10	**Discussion** (Moderator: Hiroaki Natsukawa)

Registration

Please register using the Google Form on the official website of the project. The Zoom Meeting ID and password will be sent to you by e-mail.

URL: https://ancientindia-datascience.hakubi.kyoto-u.ac.jp

Registration is available untill the end of the workshop.

No registrant limit. No registration fee.

Organizer: SPIRITS Project "Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science" (Kyoko Amano, Hiroaki Natsukawa, Oliver Hellwig, Yuki Kyogoku)

Co-Organizer: Kyoto University, Academic Center for Computing and Media Studies; Kyoko Amano Hakubi Project "Language and Social-Cultural Background of the Ancient Indian Ritual Literature"; Grant-in-Aid for Challenging Research (Exploratory) "Constructing a Database for Quantitative Analysis of Style Toward Elucidation of the Formation Process of Ancient Indian Texts" (Representative researcher: Kyoko Amano, 20K20697)

Hakubi

京都大学研究支援SPIRITS：「知の越境」融合チーム研究プログラム
2020-2021年度採択プロジェクト

京都大学 KYOTO UNIVERSITY　SPIRITS SUPPORTING PROGRAM FOR INTERACTION-BASED INITIATIVE TEAM STUDIES

「データ駆動型科学が解き明かす古代インド文献の時空間的特徴」
Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science

第1回 ワークショップ

古代文献の言語分析から読み解く社会背景のダイナミズム

Dynamism of Social Context Deciphered by a Linguistic Analysis of Ancient Literature

2021年2月12日（金）｜14:00 ~ 19:10 オンラインにて開催｜発表はすべて英語で行われます

およそ文献を正しく読む上で、文献成立の背景となる社会への理解は根底となる要件である。しかし古代社会の場合は多くの場合において実態が謎に包まれ、そこでどのような過程によって文献が成立したかも明らかでない。古代インドの宗教文献ヴェーダはそのような例の一つである。本ワークショップでは、ヴェーダ文献の言語を分析することで、古代インド社会の動き、地理的な移動や勢力圏の変化をどのように読み解くことができるのか、この分野への情報科学の応用の方法を検討しながら議論したい。

第1部	14:00 ~ 14:30	オープニング： *Problems in the Formation of the Vedas, Ancient Indian Religious Texts* 「古代インド宗教文献ヴェーダの成立を巡る諸問題」 **天野恭子**（京都大学 白眉センター・人文科学研究所）
	14:30 ~ 15:10	*The Possibility of Information Visualization and Data Analysis for Ancient Indian Literature* 「古代インド文献を対象とした情報可視化やデータ分析の可能性」 **夏川浩明**（京都大学 学術情報メディアセンター）
	15:10 ~ 15:50	*Relationship among Vedic Schools Deciphered by the Visualization of Mantra Collocation* 「マントラ共起関係の可視化から読み解くヴェーダ学派間の関係性」 **天野恭子**（京都大学 白眉センター・人文科学研究所）
	15:50 ~ 16:30	*Citation Prediction Using Academic Paper Data and Application for Surveys* 「学術論文データを用いた引用数予測とサーベイへの活用」 **濵地瞬**（京都大学 工学研究科）
第2部	16:50 ~ 17:30	*Measuring the Semantic Similarity between the Chapters of Taittirīya Samhitā Using a Vector Space Model* 「ベクトル空間モデルによる『タイッティリーヤ・サンヒター』の章間類似度比較」 **京極祐希**（Leipzig University, Indology）
	17:30 ~ 18:10	*Dating Vedic Texts with Computational Models: Algorithmic Considerations and Data Selection* **Oliver Hellwig**（University of Zurich, Department of Comparative Language Science）
	18:10 ~ 18:50	*morogram: Background, History, and Purpose of a Tool for East Asian Text Analysis* 「morogram: 東アジア文献分析ツールの開発の経緯と目的」 **師茂樹**（花園大学 文学部）
	18:50 ~ 19:10	ディスカッション（司会：夏川浩明）

参加登録

本プロジェクトウェブサイト上のGoogleフォームより参加登録をお願いいたします。登録いただいた皆様に、e-mailにてZoomミーティングIDおよびパスワードをお知らせいたします。

URL｜https://ancientindia-datascience.hakubi.kyoto-u.ac.jp
お申込み締切｜ワークショップ終了まで、いつでもご登録いただけます
定員なし、参加費無料

主催｜SPIRITSプロジェクト「データ駆動型科学が解き明かす古代インド文献の時空間的特徴」（天野恭子、夏川浩明、Oliver Hellwig、京極祐希）

共催｜京都大学学術情報メディアセンター、天野恭子白眉プロジェクト「古代インド祭式文献の言語および社会的・文化的成立背景の研究」、科研費挑戦的研究（萌芽）「古代インド文献成立過程解明に向けた文体計量分析のためのデータベース構築」（研究代表者：天野恭子、20K20697）

Hakubi 白眉