Ancient India meets Data-Science

The 2nd and concluding Workshop of SPIRITS project

"Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science"

古代インドとデータサイエンス

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」

第２回（最終）ワークショップ

it's also

A Kick-off for Joint International Research

"A Study of Language Layers in Vedic Literature for the Development of a Program for Age-Estimation"

国際共同研究

「ヴェーダ文献における言語層の考察と それを利用した文献年代推定プログラムの開発」

のキックオフを兼ねて

Collection of Presentation Slides

発表資料集

Organizer: SPIRITS Project “Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science” (Kyoko Amano, Hiroaki Natsukawa, Oliver Hellwig, Yuki Kyogoku); Fostering Joint International Research (B) of KAKENHI "A Study of Language Layers in Vedic Literature for the Development of a Program for Age-Estimation" (Representativ researcher: Kyoko Amano)

主催：京都大学研究支援 SPIRITS　2021「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）、国際共同研究強化 (B)2021-2026「ヴェーダ文献における言語層の考察とそれを利用した文献年代推定プログラムの開発」（研究代表者：天野恭子、課題番号：21KK0004）

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)

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）2020-2022「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）

SPIRITS
SUPPORTING PROGRAM FOR INTERACTION-BASED INITIATIVE TEAM STUDIES

KYOTO UNIVERSITY KYOTO JAPAN FOUNDED 1897

Hakubi Kyoto Univ.

Contents｜目次

Presentation 1 ｜報告 1

The Result of the Two-Year SPIRITS Project and Our Vision for the Next Research.

２年間のSPIRITSプロジェクトの成果と今後の研究への展望

Kyoko Amano (Kyoto University, Hakubi Center / Institute for Research in Humanities)
天野恭子（京都大学 白眉センター／人文科学研究所）

1

The Result of the Two-Year SPIRITS Project and Our Vision for the Next Research

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

Workshop "Ancient India meets Data-Science"

2022/2/11

Data Science
Hakubi
SPIRITS

2

Our two Projects

Kyoto University SPIRITS project (FY2020-2021)

Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science

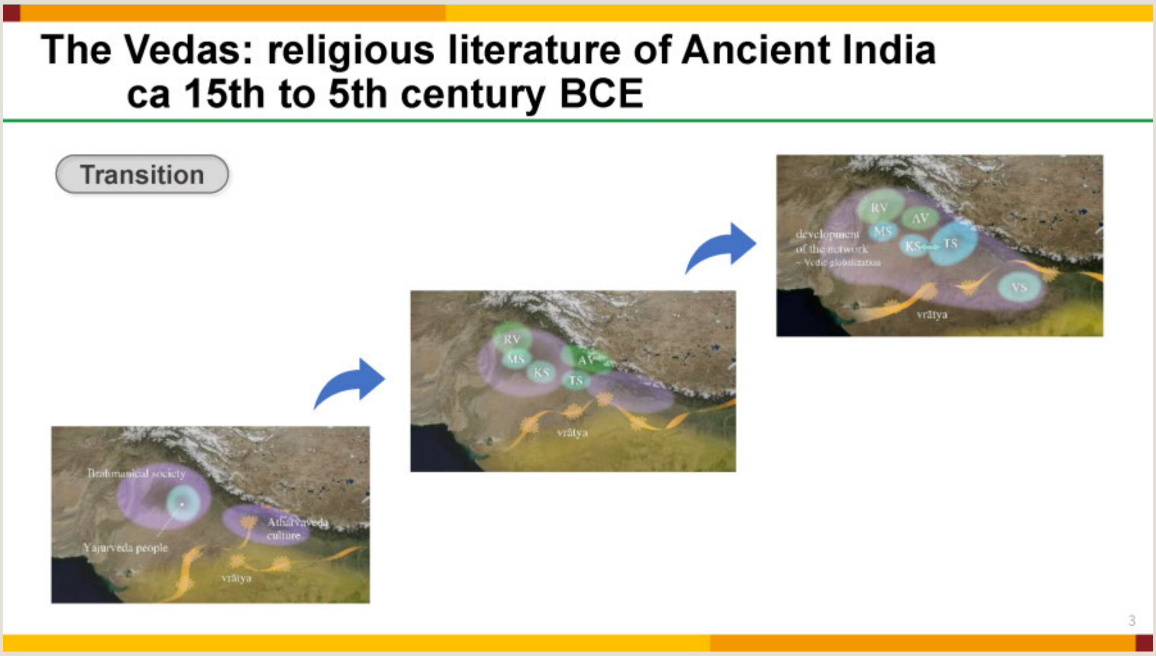
Fostering Joint International Research (B) of KAKENHI (FY2021-2026)

A Study of Language Layers in Vedic Literature for the Development of a Program for Age-Estimation

collaborating with "Chron-BMM - Bayesian Mixture Models für die Datierung von Textkorpora" lead by Oliver Hellwig

2

3

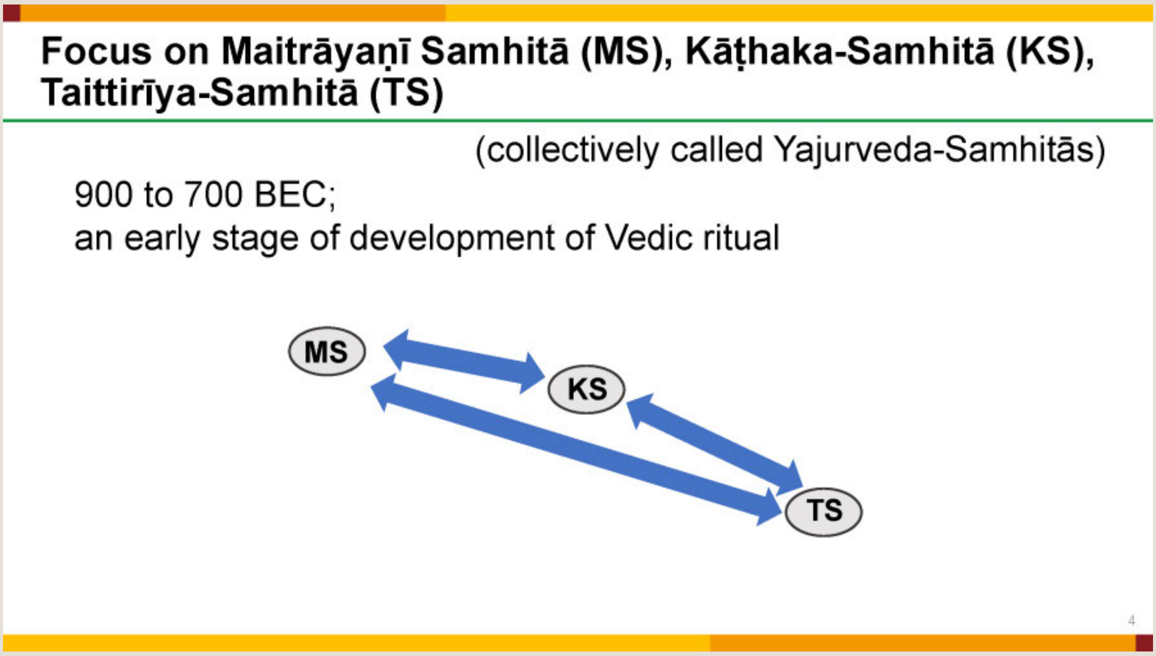


The Vedas: religious literature of Ancient India ca 15th to 5th century BCE

Transition

3

4



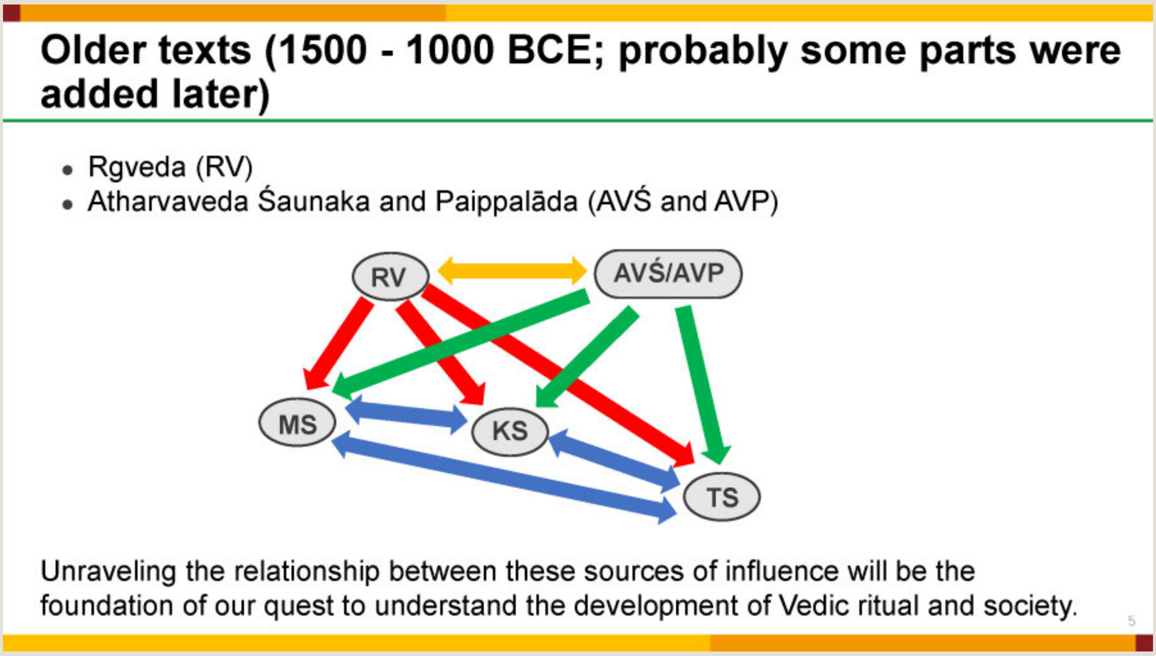
Focus on Maitrāyaṇī Samhitā (MS), Kāṭhaka-Samhitā (KS), Taittirīya-Samhitā (TS)

(collectively called Yajurveda-Samhitās)

900 to 700 BEC;
an early stage of development of Vedic ritual

4

5



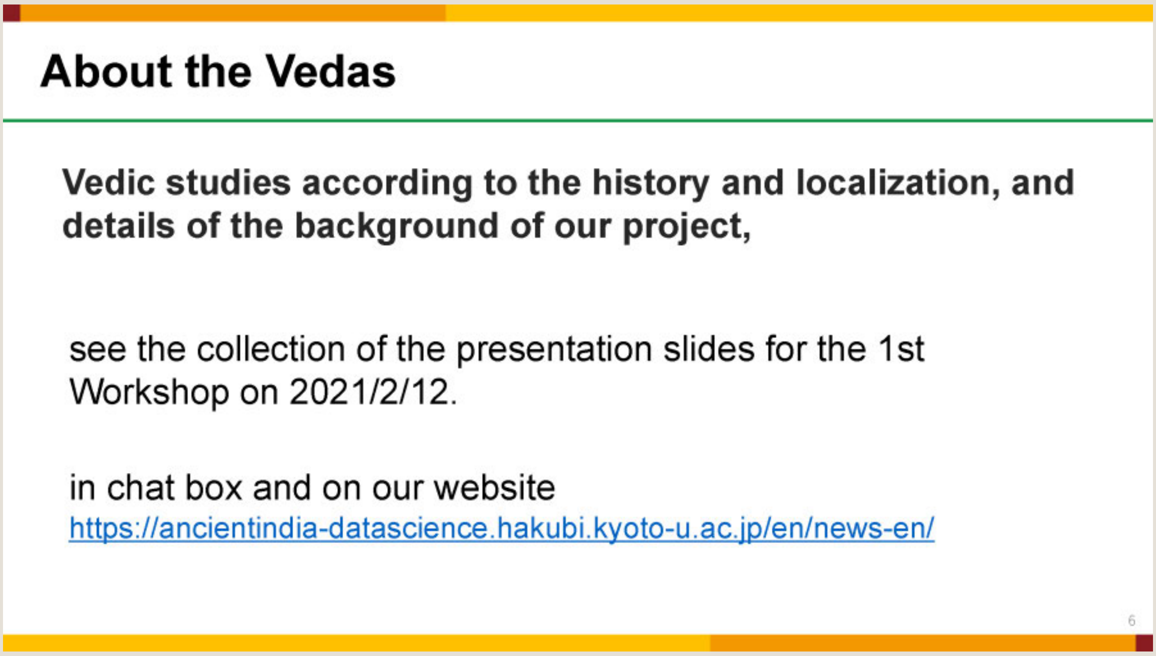
Older texts (1500 - 1000 BCE; probably some parts were added later)

- Rgveda (RV)
- Atharvaveda Śaunaka and Paippalāda (AVŚ and AVP)

Unraveling the relationship between these sources of influence will be the foundation of our quest to understand the development of Vedic ritual and society.

5

6



About the Vedas

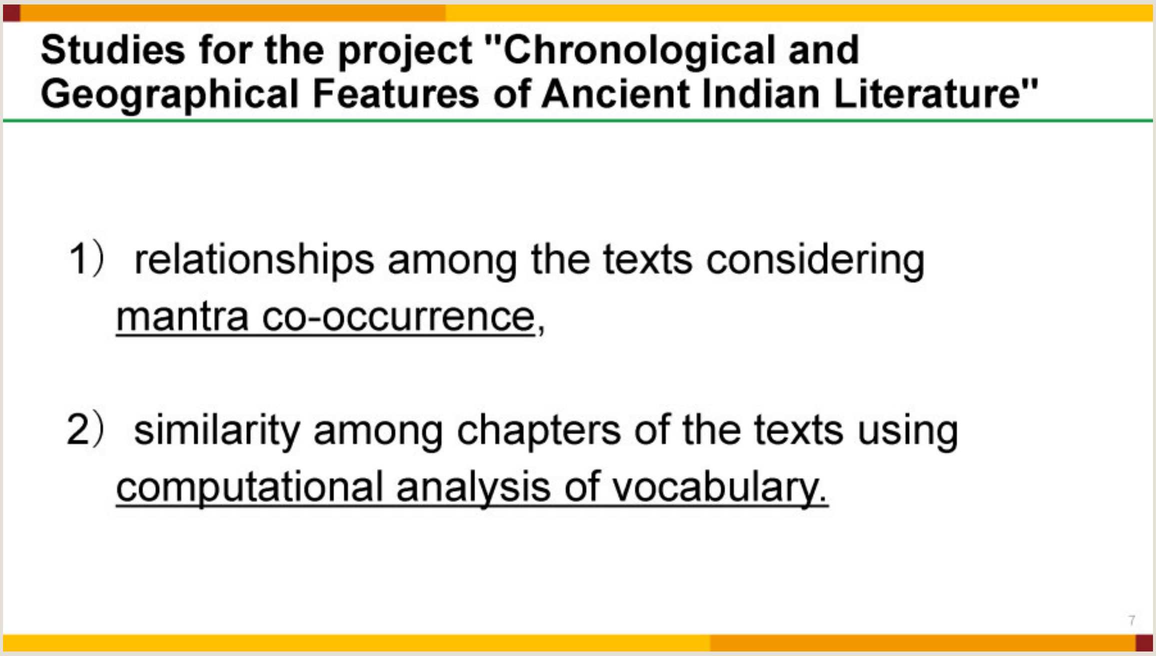
Vedic studies according to the history and localization, and details of the background of our project,

see the collection of the presentation slides for the 1st Workshop on 2021/2/12.

in chat box and on our website
https://ancientindia-datascience.hakubi.kyoto-u.ac.jp/en/news-en/

6

7



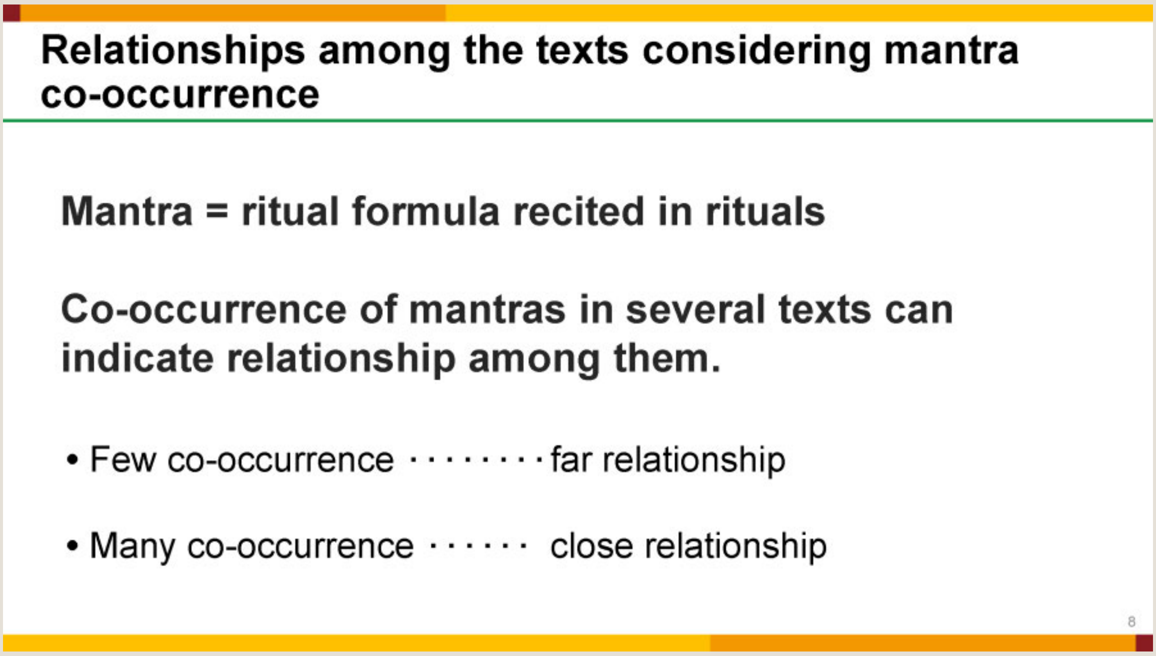
Studies for the project "Chronological and Geographical Features of Ancient Indian Literature"

1） relationships among the texts considering mantra co-occurrence,

2） similarity among chapters of the texts using computational analysis of vocabulary.

7

8



Relationships among the texts considering mantra co-occurrence

Mantra = ritual formula recited in rituals

Co-occurrence of mantras in several texts can indicate relationship among them.

- Few co-occurrence ········far relationship
- Many co-occurrence ······ close relationship

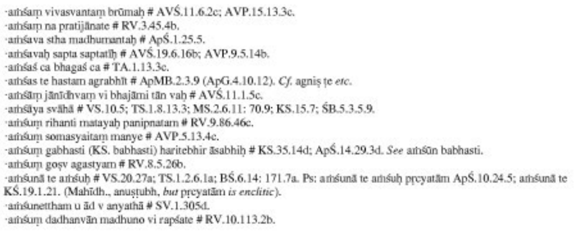
8

9

Relationships among the texts considering mantra co-occurrence

- **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.

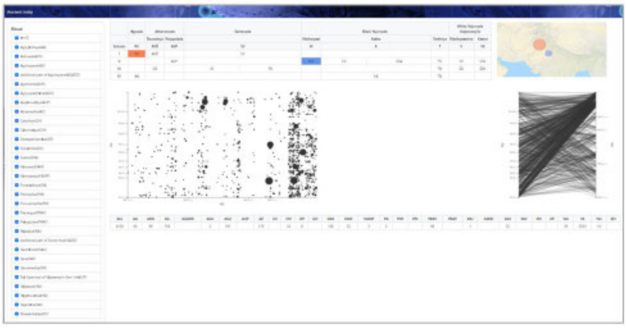
electronic edition of A Vedic Concordance
This index is used as a database.

Data for the investigation of the relationship among the texts:
Index for ca 90,000 mantras that appear in all Vedic literature, with names of literature and the places.

10

Relationships among the texts considering mantra co-occurrence

Visualizing the co-occurrence of mantras
- Relational Database using SQLite
- Co-occurrence relationships among the 19 texts
  → Identified 150 relationships involving these 19 important texts.
- Visualize the relationship between two selected texts.
- Scatter plot and parallel coordinate plot.
- Visualizing the chapter structure with colors.

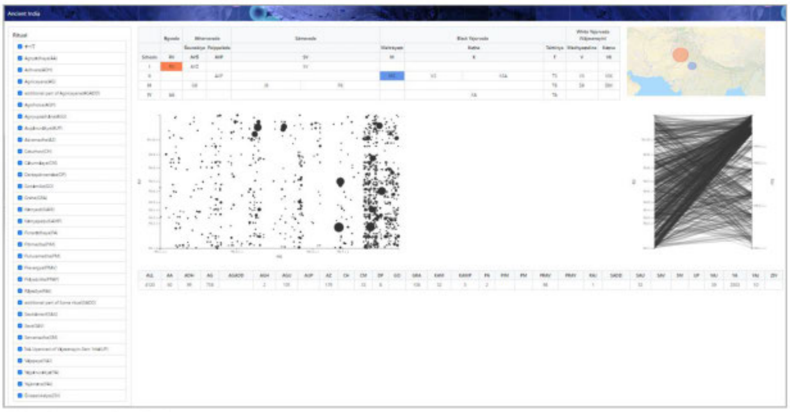


http://34.146.175.179/

11

Relationships among the texts considering mantra co-occurrence

New functions: Classification by ritual
Numerical Data of the number of co-occurred mantras

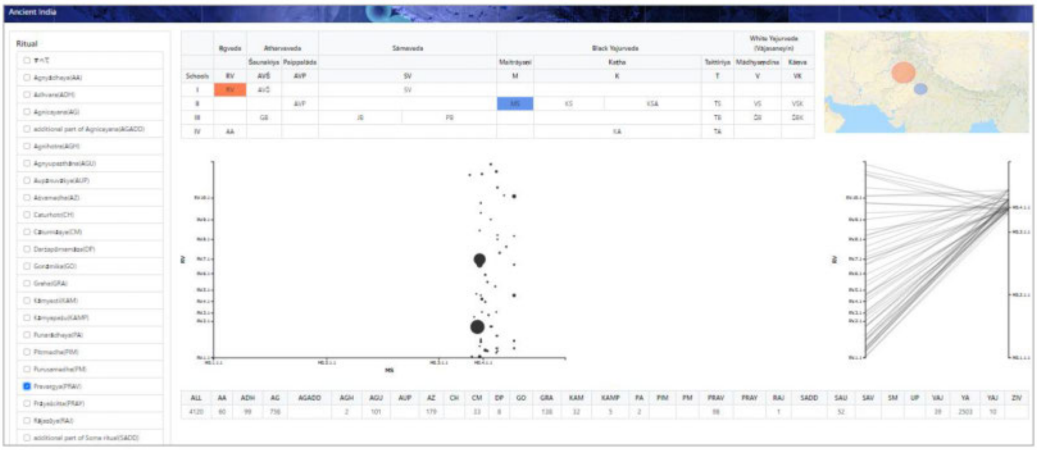


http://34.146.175.179/

12

Relationships among the texts considering mantra co-occurrence

choosing pravargya ritual



Ancient India

Ritual

- □ すべて
- □ Agnyādheya(AA)
- □ Adhvara(ADH)
- □ Agnicayana(AG)
- □ additional part of Agnicayana(AGADD)
- □ Agnihotra(AGH)
- □ Agnyupasthāna(AGU)
- □ Aupānuvākya(AUP)
- □ Aśvamedha(AZ)
- □ Caturhotṛ(CH)
- □ Cāturmāsya(CM)
- □ Darśapūrṇamāsa(DP)
- □ Gorāmika(GO)
- □ Graha(GRA)
- □ Kāmyeṣṭi(KAM)
- □ Kāmyapaśu(KAMP)
- □ Punarādheya(PA)
- □ Pitṛmedha(PIM)
- □ Puruṣamedha(PM)
- ☑ Pravargya(PRAV)
- □ Prāyaścitta(PRAY)
- □ Rājasūya(RAJ)
- □ additional part of Soma ritual(SADD)

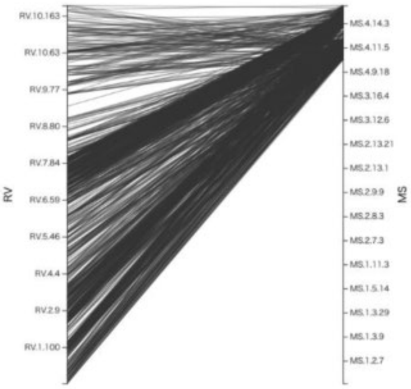
	Ṛgveda	Atharvaveda		Sāmaveda		Black Yajurveda				White Yajurveda (Vājasaneyin)	
		Śaunakīya	Paippalāda			Maitrāyaṇī	Kaṭha		Taittirīya	Mādhyaṃdina	Kāṇva
Schools	RV	AVŚ	AVP	SV		M	K		T	V	VK
I	RV	AVŚ		SV							
II			AVP			MS	KS	KSA	TS	VS	VSK
III		GB		JB	PB				TB	ŚB	ŚBK
IV	AA						KA		TA		

ALL	AA	ADH	AG	AGADD	AGH	AGU	AUP	AZ	CH	CM	DP	GO	GRA	KAM	KAMP	PA	PIM	PM	PRAV	PRAY	RAJ	SADD	SAU	SAV	SM	UP	VAJ	YA	YAJ	ZIV
4120	60	99	736		2	101		179		33	8		138	32	5	2			98		1		52				39	2503	10	

13

Relationships among the texts considering mantra co-occurrence

improving the data



·yūyaṃ pāta svastibhiḥ sadā naḥ # RV.7.1.20d,25d; 3.10d; 7.7d,8d; 9.6d; 11.5d; 12.3d; 13.3d; 14.3d; 19.11d; 20.10d; 21.10d; 22.9d; 23.6d; 24.6d; 25.6d; 26.5d; 27.5d; 28.5d; 29.5d; 30.5d; 34.25d; 35.15d; 36.9d; 37.8d; 39.7d; 40.6d; 41.7d; 42.6d; 43.5d; 45.4d; 46.4d; 47.4d; 48.4d; 51.3d; 53.3d; 54.3d; 56.25d; 57.7d; 58.6d; 60.12d; 61.7d; 62.6d; 63.6d; 64.5d; 65.5d; 67.10d; 68.9d; 69.8d; 70.7d; 71.6d; 72.5d; 73.5d; 75.8d; 76.7d; 77.6d; 78.5d; 79.5d; 80.3d; 84.5d; 85.5d; 86.8d; 87.7d; 88.7d; 90.7d; 91.7d; 92.5d; 93.8d; 95.6d; 97.10d; 98.7d; 99.7d; 100.7d; 101.6d; 9.90.6d; 97.3d,6d; 10.65.15d; 66.15d; 122.8d; AVŚ.3.16.7d; 19.11.5d; 20.12.6d; 17.12d; 37.11d; 87.7d; AVP.4.31.7d; 12.17.5d; SV.2.656d,751d,977d; VS.20.54d; 27.28d; 34.40d; TS.1.5.11.2d; 2.2.12.5d; 3.4.10.1d; MS.4.14.2d: 217.6; 4.14.7d: 226.8; 4.14.12d: 235.12; KS.6.10d; 8.16d; GB.2.4.2; JB.3.243; TB.2.5.6.4d; 8.5d; 8.1.2d; 4.1d; 9.9d; 3.5.2.3d; 6.1.3d; ApŚ.13.18.1d; 22.7.11d; MŚ.2.5.4.12d; PG.3.4.7d; ApMB.1.14.7d; 2.15.19d; MG.2.11.19d; JG.1.11d.

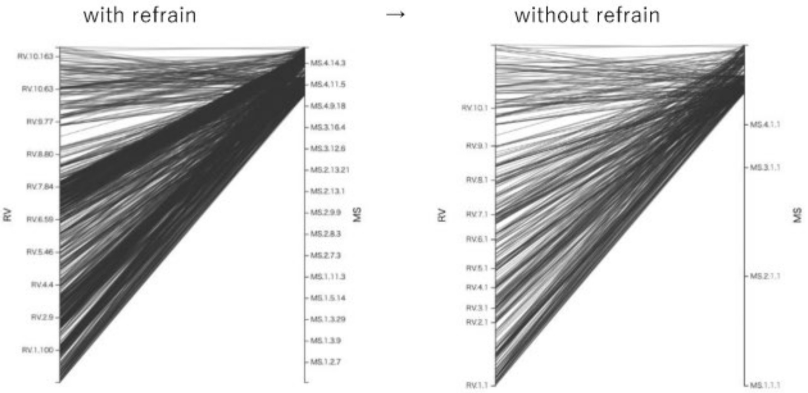
(Vedic Concordance)

Refrain･･･78 times in the 7th book!

14

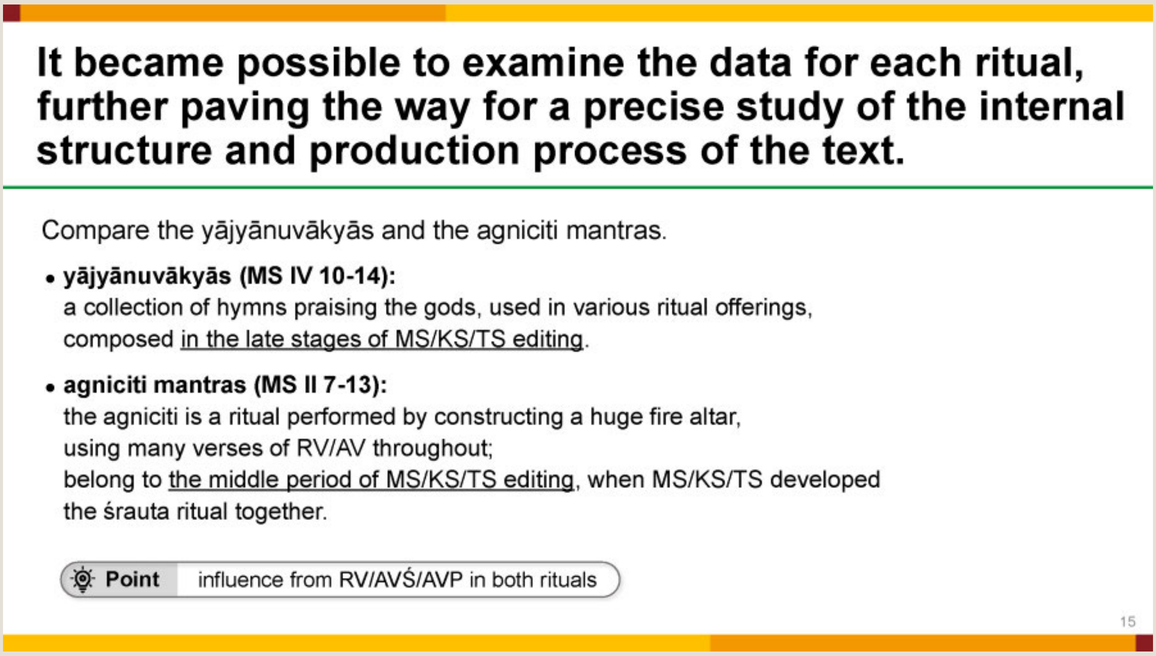
Relationships among the texts considering mantra co-occurrence

improving the data



It is important to check and examine the original texts!

15



It became possible to examine the data for each ritual, further paving the way for a precise study of the internal structure and production process of the text.

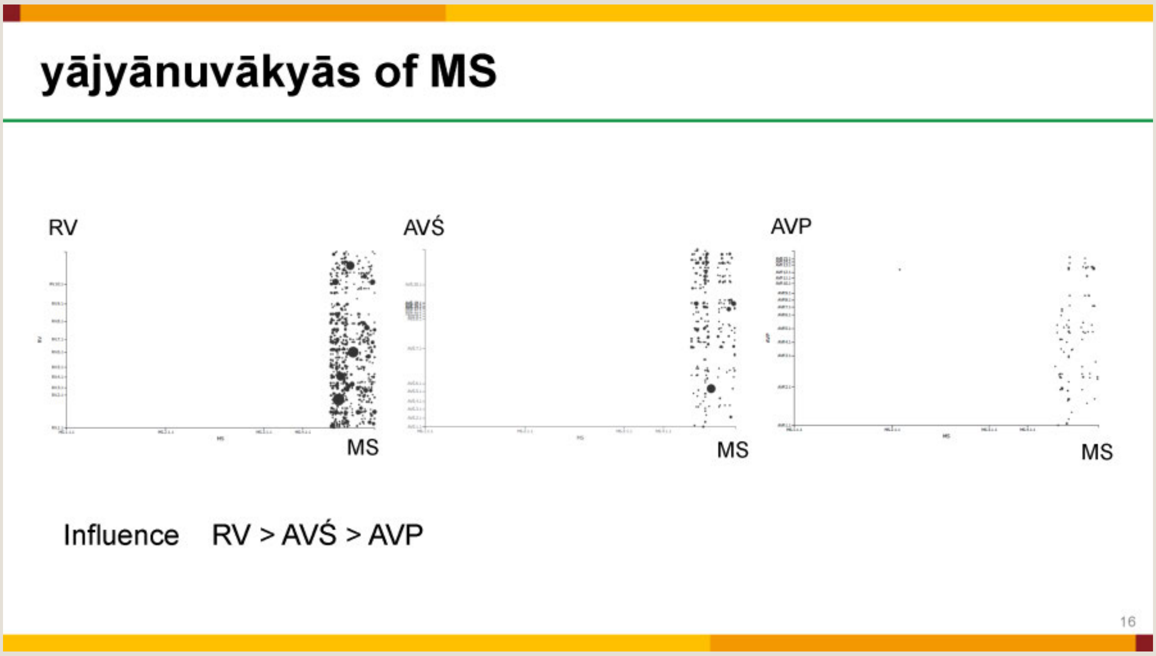
Compare the yājyānuvākyās and the agniciti mantras.

- **yājyānuvākyās (MS IV 10-14):**
  a collection of hymns praising the gods, used in various ritual offerings,
  composed in the late stages of MS/KS/TS editing.
- **agniciti mantras (MS II 7-13):**
  the agniciti is a ritual performed by constructing a huge fire altar,
  using many verses of RV/AV throughout;
  belong to the middle period of MS/KS/TS editing, when MS/KS/TS developed
  the śrauta ritual together.

Point influence from RV/AVŚ/AVP in both rituals

15

16

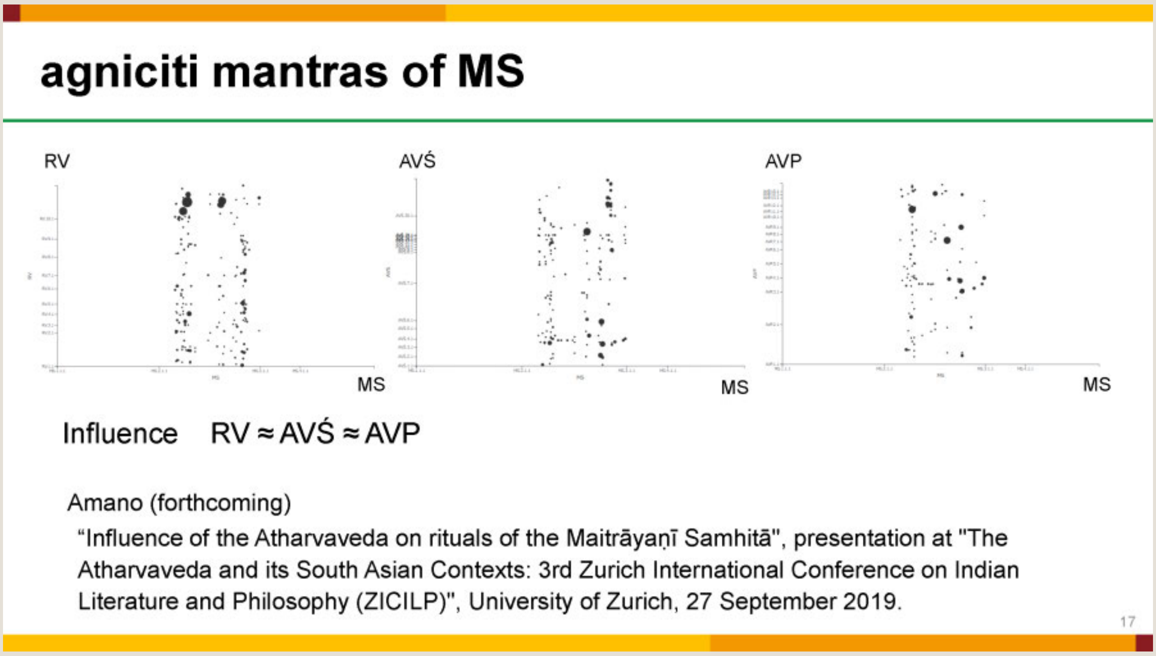


yājyānuvākyās of MS

Influence RV > AVŚ > AVP

16

17



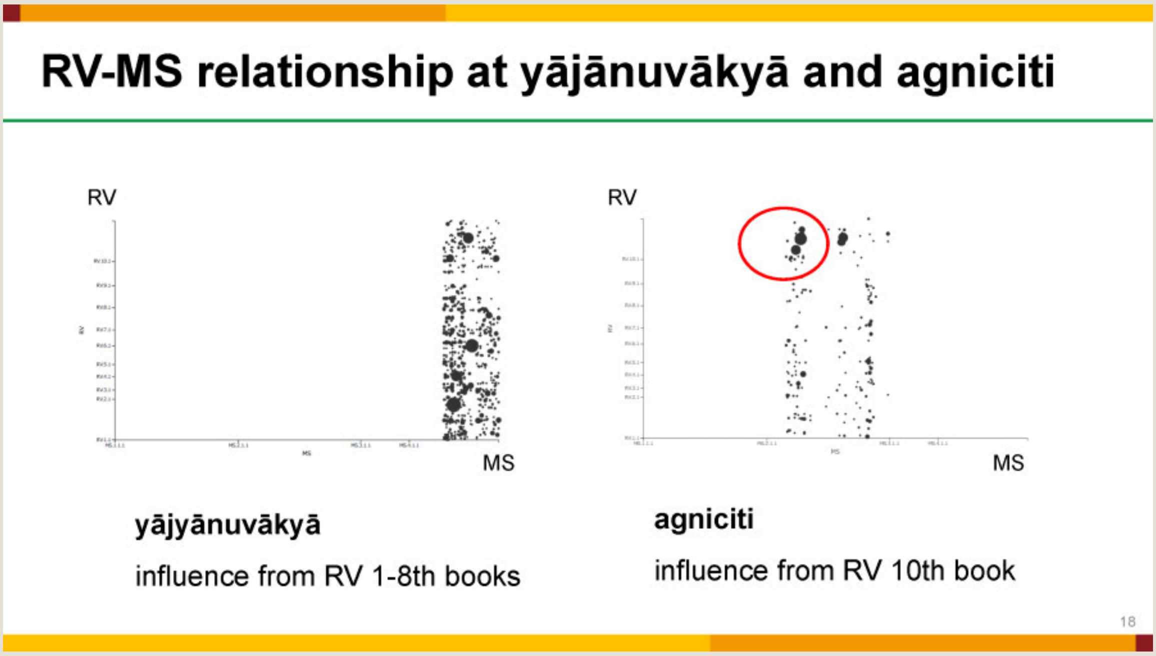
agniciti mantras of MS

Influence RV ≈ AVŚ ≈ AVP

Amano (forthcoming)
“Influence of the Atharvaveda on rituals of the Maitrāyaṇī Samhitā", presentation at "The Atharvaveda and its South Asian Contexts: 3rd Zurich International Conference on Indian Literature and Philosophy (ZICILP)", University of Zurich, 27 September 2019.

17

18



RV-MS relationship at yājānuvākyā and agniciti

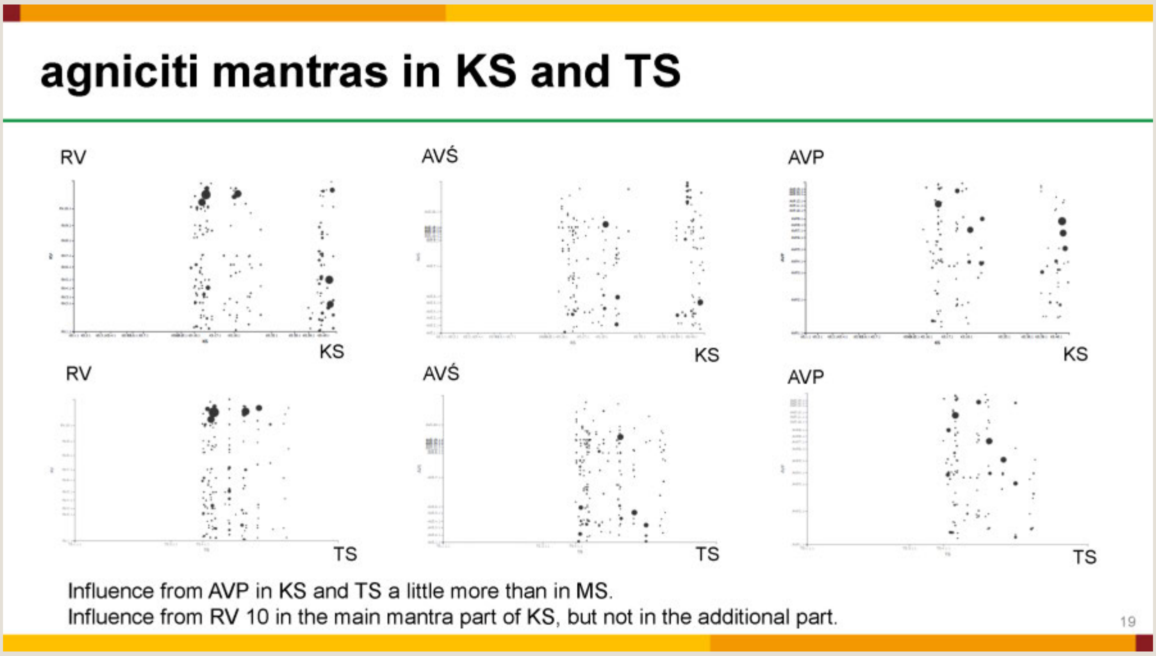
yājyānuvākyā

influence from RV 1-8th books

agniciti

influence from RV 10th book

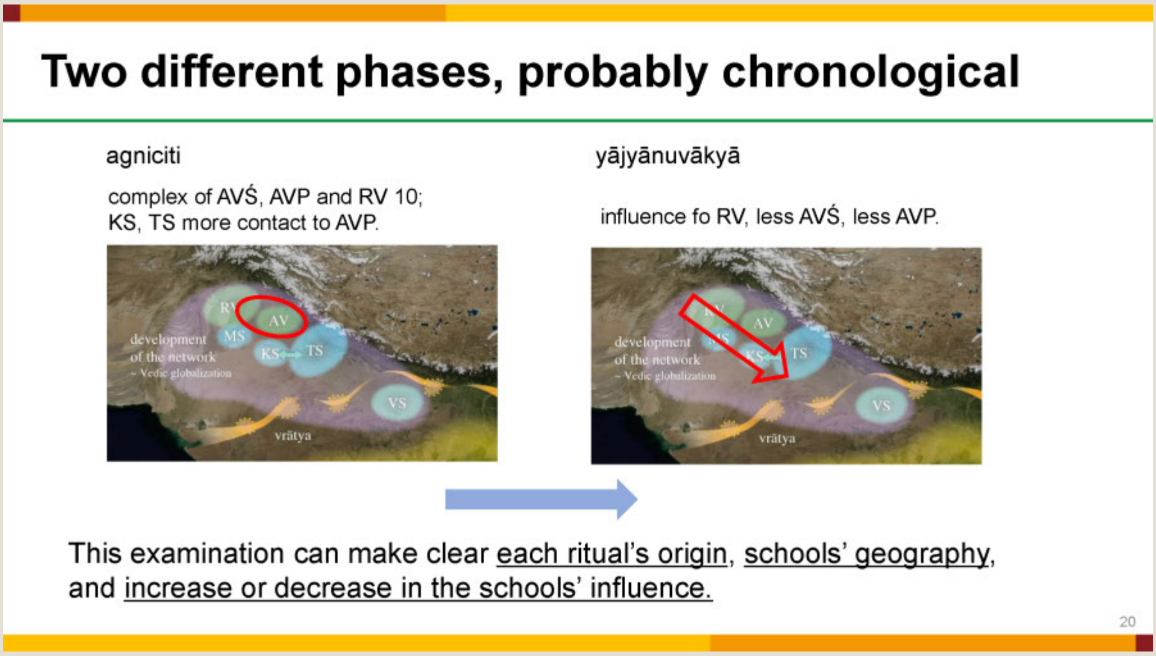
19



agniciti mantras in KS and TS

Influence from AVP in KS and TS a little more than in MS.
Influence from RV 10 in the main mantra part of KS, but not in the additional part.

20



Two different phases, probably chronological

agniciti

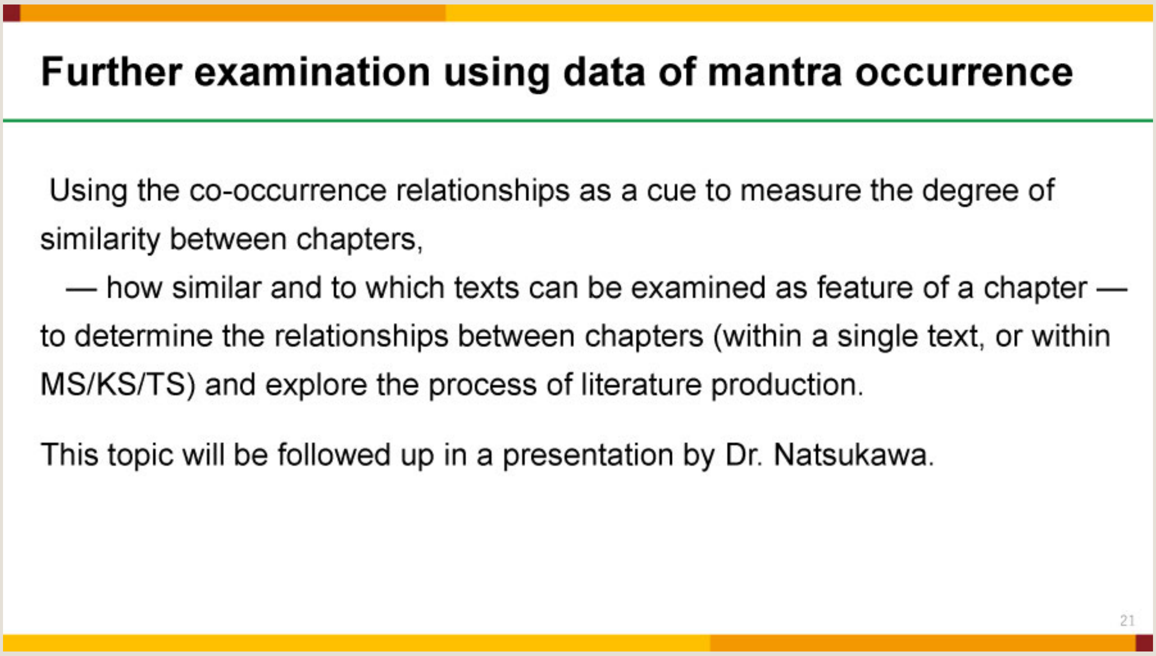
complex of AVŚ, AVP and RV 10;
KS, TS more contact to AVP.

yājyānuvākyā

influence fo RV, less AVŚ, less AVP.

This examination can make clear each ritual's origin, schools' geography, and increase or decrease in the schools' influence.

21



Further examination using data of mantra occurrence

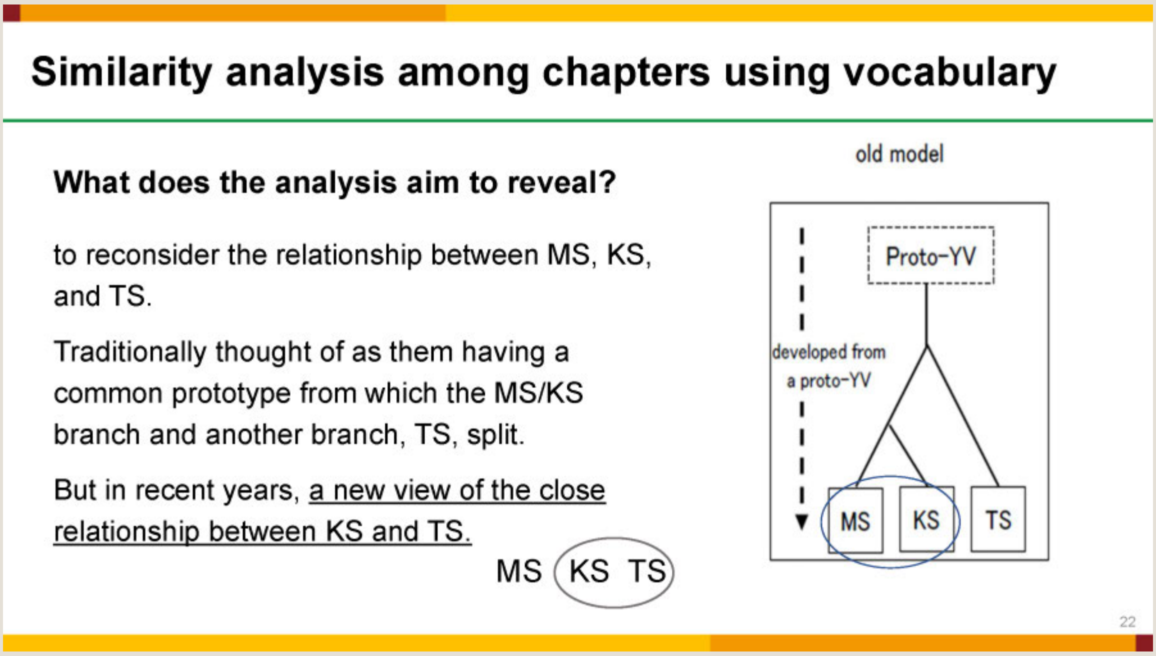
Using the co-occurrence relationships as a cue to measure the degree of similarity between chapters,

— how similar and to which texts can be examined as feature of a chapter — to determine the relationships between chapters (within a single text, or within MS/KS/TS) and explore the process of literature production.

This topic will be followed up in a presentation by Dr. Natsukawa.

21

22



Similarity analysis among chapters using vocabulary

What does the analysis aim to reveal?

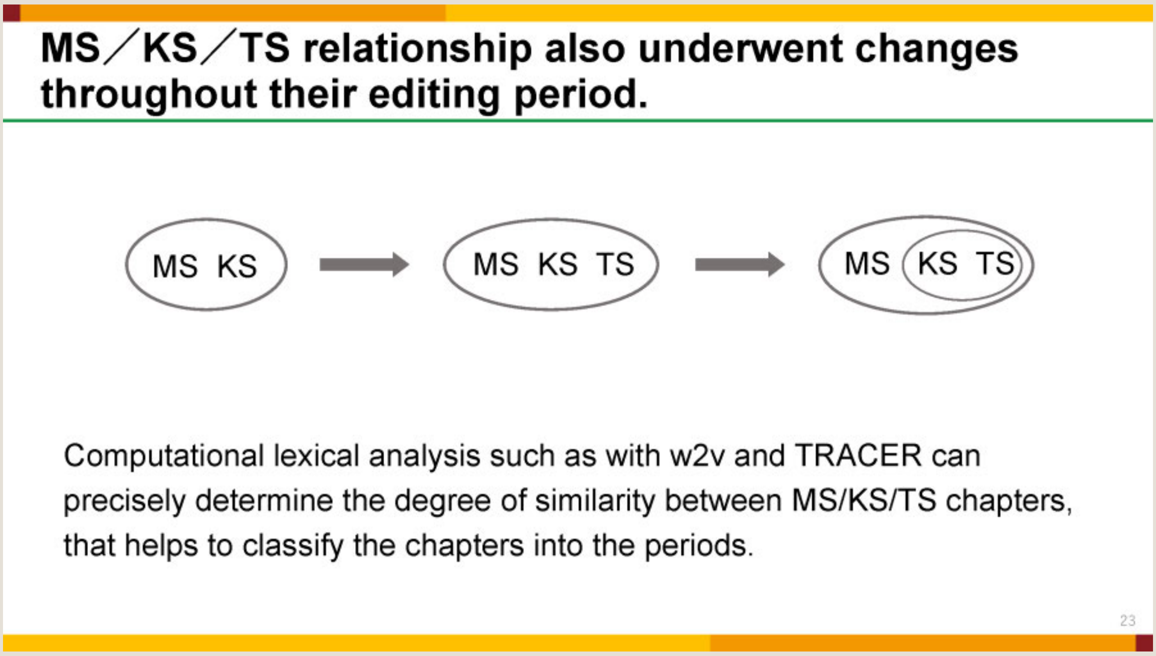
to reconsider the relationship between MS, KS, and TS.

Traditionally thought of as them having a common prototype from which the MS/KS branch and another branch, TS, split.

But in recent years, a new view of the close relationship between KS and TS.

22

23



MS／KS／TS relationship also underwent changes throughout their editing period.

Computational lexical analysis such as with w2v and TRACER can precisely determine the degree of similarity between MS/KS/TS chapters, that helps to classify the chapters into the periods.

23

24

Plan for the new project "A Study of Language Layers in Vedic Literature for the Development of a Program for Age-Estimation"

Analysis of syntax

Important : uses of particles and pronouns

: function of verb tense and mood.

These can be good indicators to stratify the linguistic layers

→ Examples from my recent studies of MS

25

Syntactic phenomena as indicator for linguistic layers

ha vai "consequently" or ***tad*** "so, then, thus" used with the phrase *ya evaṃ vidvān / veda* (Amano 2020)

sarvā ha vā asya yakṣyamāṇasya devatā yajñam āgacchanti ya evaṃ veda //

sarva	ha	vai	idam	yaj	devatā	yajña	āgam,	yad	evam	vid.
n.p.f.	indecl.	indecl.	g.s.m.	Fut., g.s.m.	n.p.f.	ac.s.m.	3. pl., Pre. ind.	n.s.m.	indecl.	3. sg., Perf.

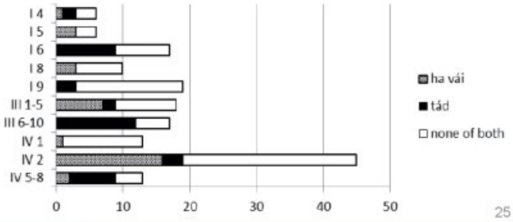
All deities consequently come to his sacrifice, as he is planning to hold a sacrifice, when he knows thus.

tad ya evaṃ veda bhavaty ātmanā // (23) Par

tad	yad	evam	vid,	bhū	ātman.
ac.s.n.	n.s.m.	indecl.	3. sg., Perf.	3. sg., Pre. ind.	i.s.m.

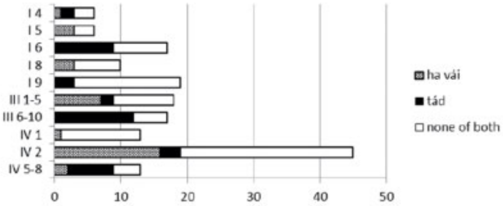
So, he himself successes, when he know thus.

The numbers of *ha vai* and *tad* at *ya evaṃ vidvān/veda* in the chapters of MS:

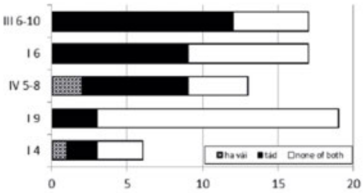


26

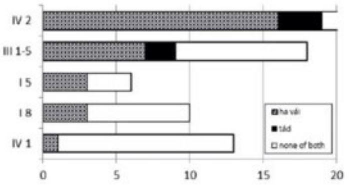
"*ha vai* chapters" and "*tad* chapters"



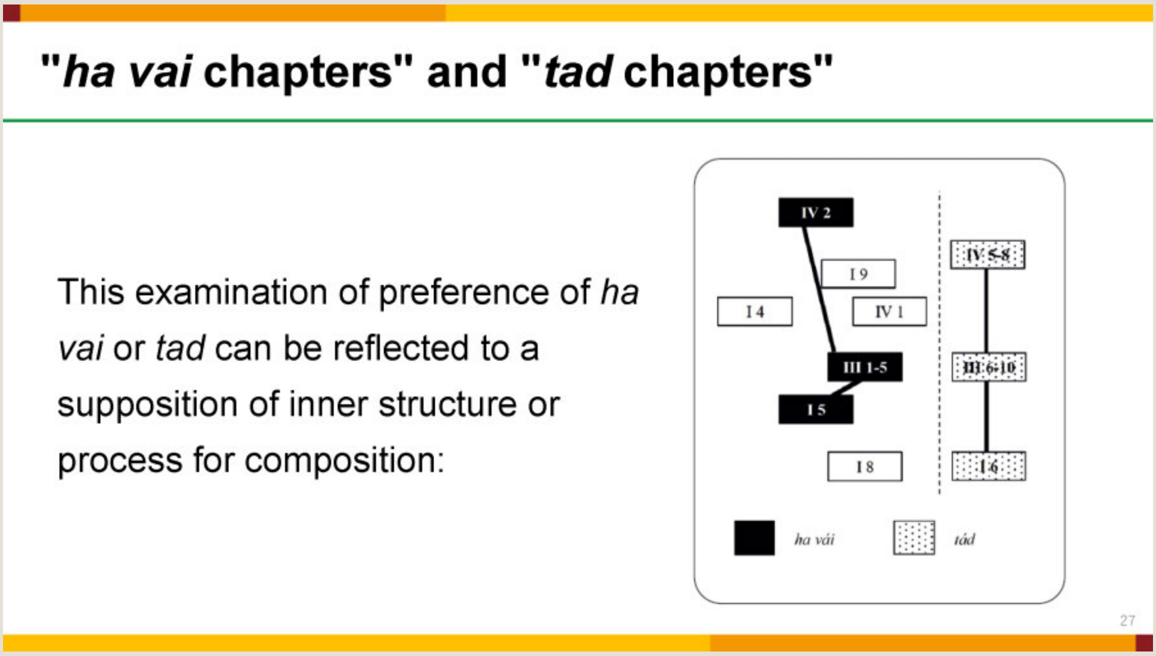
chapters with frequent use of **tad**



chapters with frequent use of **ha vai**



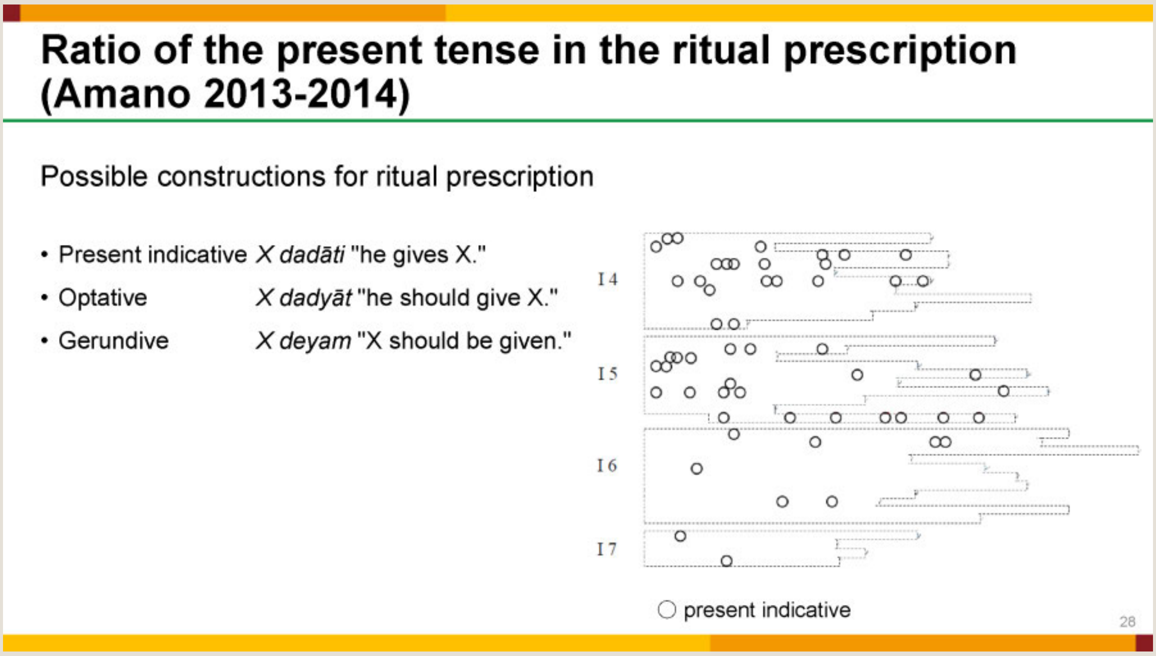
27



"*ha vai* chapters" and "*tad* chapters"

This examination of preference of *ha vai* or *tad* can be reflected to a supposition of inner structure or process for composition:

28

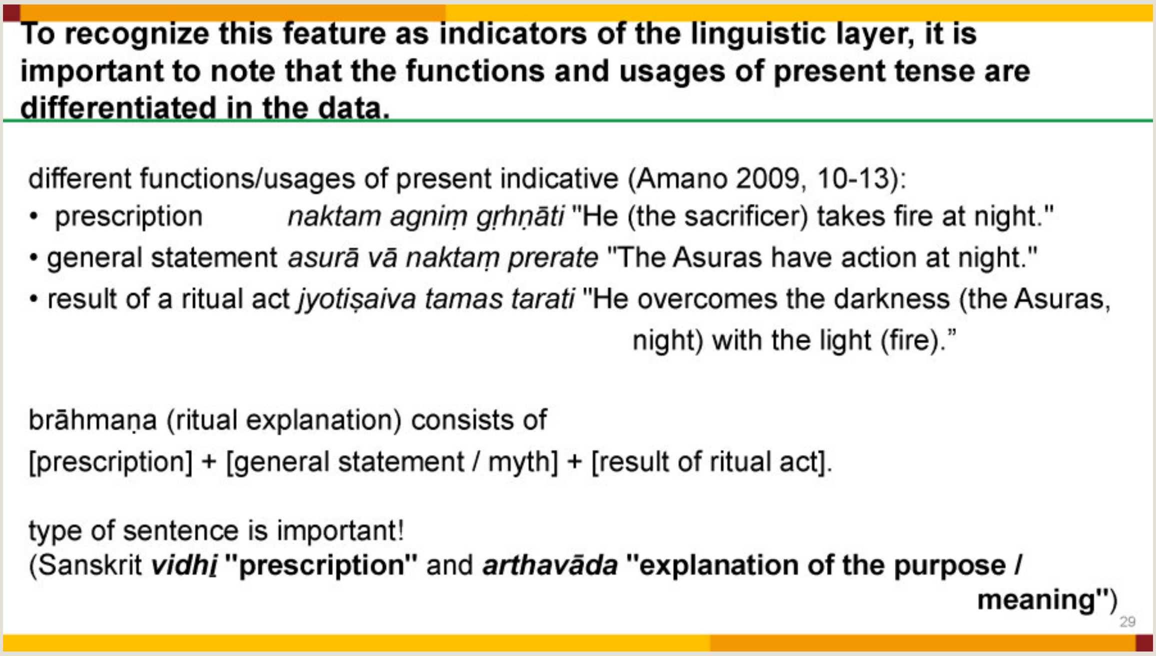


Ratio of the present tense in the ritual prescription (Amano 2013-2014)

Possible constructions for ritual prescription

• Present indicative	*X dadāti* "he gives X."
• Optative	*X dadyāt* "he should give X."
• Gerundive	*X deyam* "X should be given."

29



To recognize this feature as indicators of the linguistic layer, it is important to note that the functions and usages of present tense are differentiated in the data.

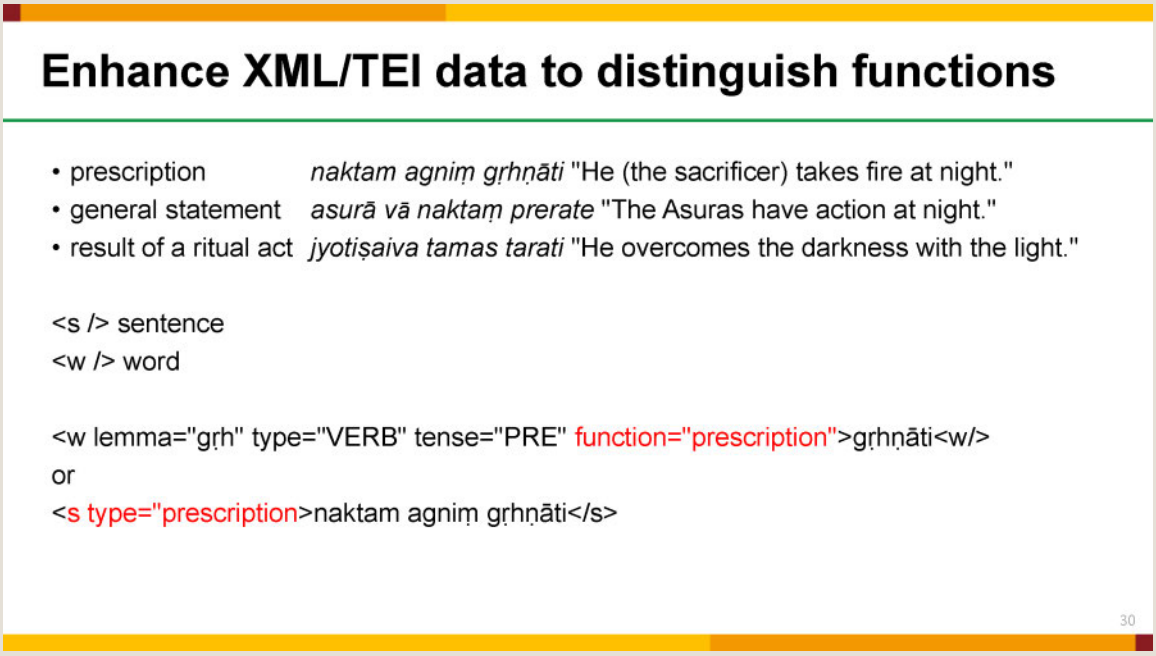
different functions/usages of present indicative (Amano 2009, 10-13):

- prescription *naktam agniṃ gṛhṇāti* "He (the sacrificer) takes fire at night."
- general statement *asurā vā naktaṃ prerate* "The Asuras have action at night."
- result of a ritual act *jyotiṣaiva tamas tarati* "He overcomes the darkness (the Asuras, night) with the light (fire)."

brāhmaṇa (ritual explanation) consists of
[prescription] + [general statement / myth] + [result of ritual act].

type of sentence is important!
(Sanskrit ***vidhi*** **"prescription"** and ***arthavāda*** **"explanation of the purpose / meaning"**)

30



Enhance XML/TEI data to distinguish functions

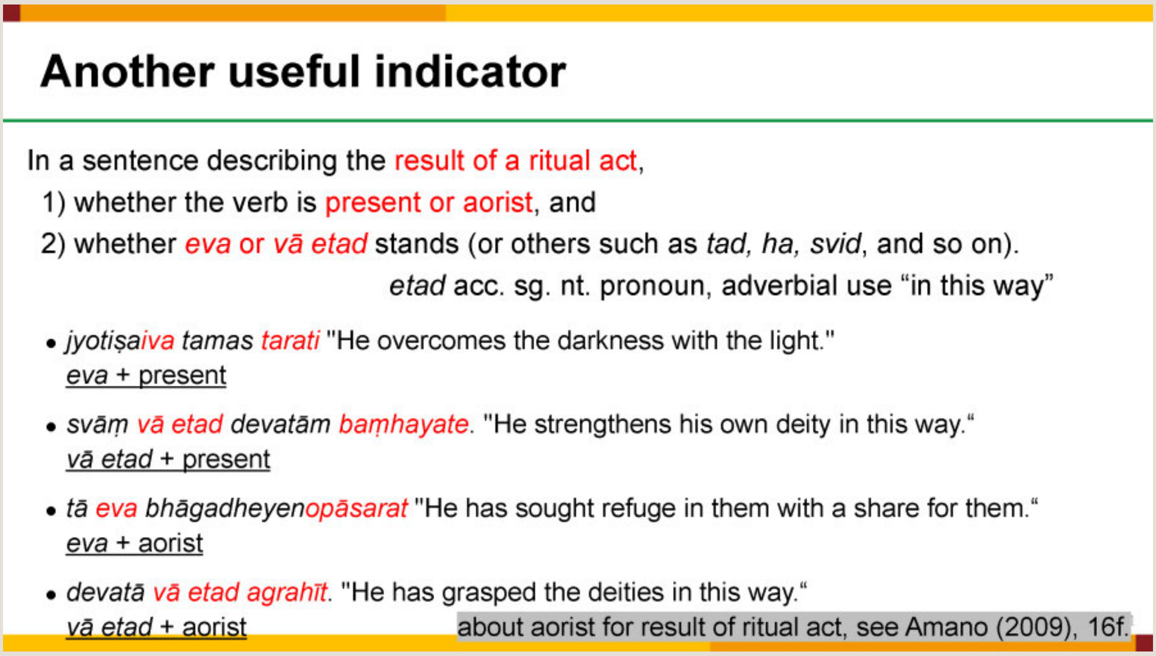
- prescription *naktam agniṃ gṛhṇāti* "He (the sacrificer) takes fire at night."
- general statement *asurā vā naktaṃ prerate* "The Asuras have action at night."
- result of a ritual act *jyotiṣaiva tamas tarati* "He overcomes the darkness with the light."

<s /> sentence
<w /> word

```
<w lemma="gṛh" type="VERB" tense="PRE" function="prescription">gṛhṇāti<w/>
or
<s type="prescription>naktam agniṃ gṛhṇāti</s>
```

30

31



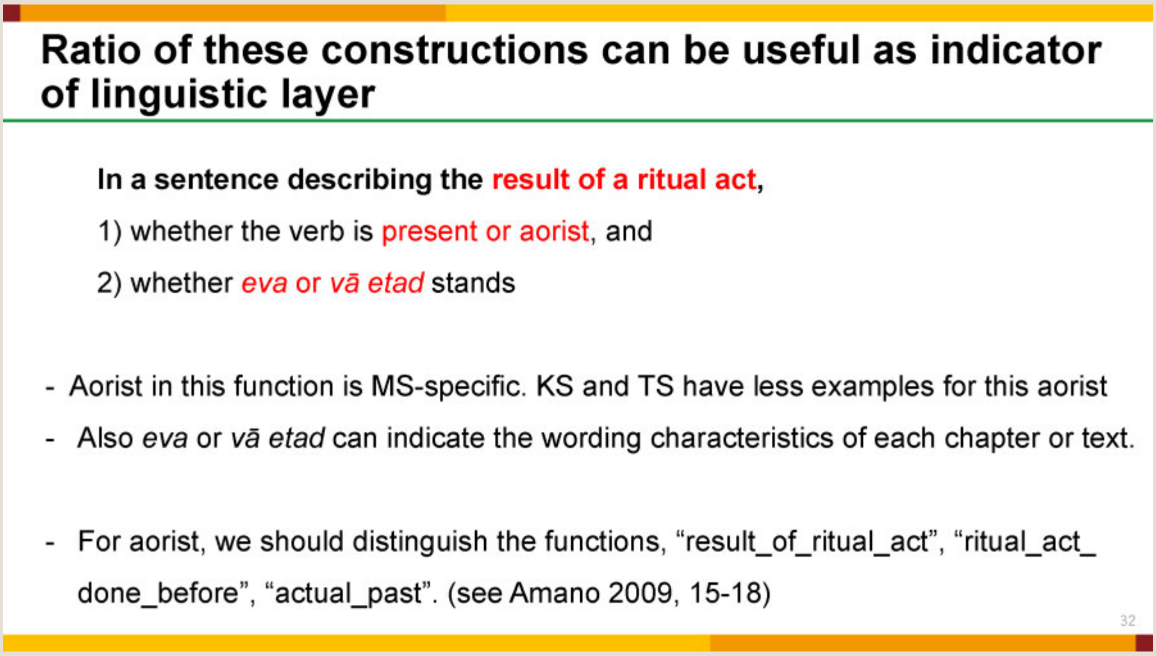
Another useful indicator

In a sentence describing the result of a ritual act,
1) whether the verb is present or aorist, and
2) whether *eva* or *vā etad* stands (or others such as *tad, ha, svid*, and so on).
etad acc. sg. nt. pronoun, adverbial use "in this way"

- *jyotiṣaiva tamas tarati* "He overcomes the darkness with the light."
  eva + present
- *svāṃ vā etad devatāṃ baṃhayate*. "He strengthens his own deity in this way."
  vā etad + present
- *tā eva bhāgadheyenopāsarat* "He has sought refuge in them with a share for them."
  eva + aorist
- *devatā vā etad agrahīt*. "He has grasped the deities in this way."
  vā etad + aorist

about aorist for result of ritual act, see Amano (2009), 16f.

32



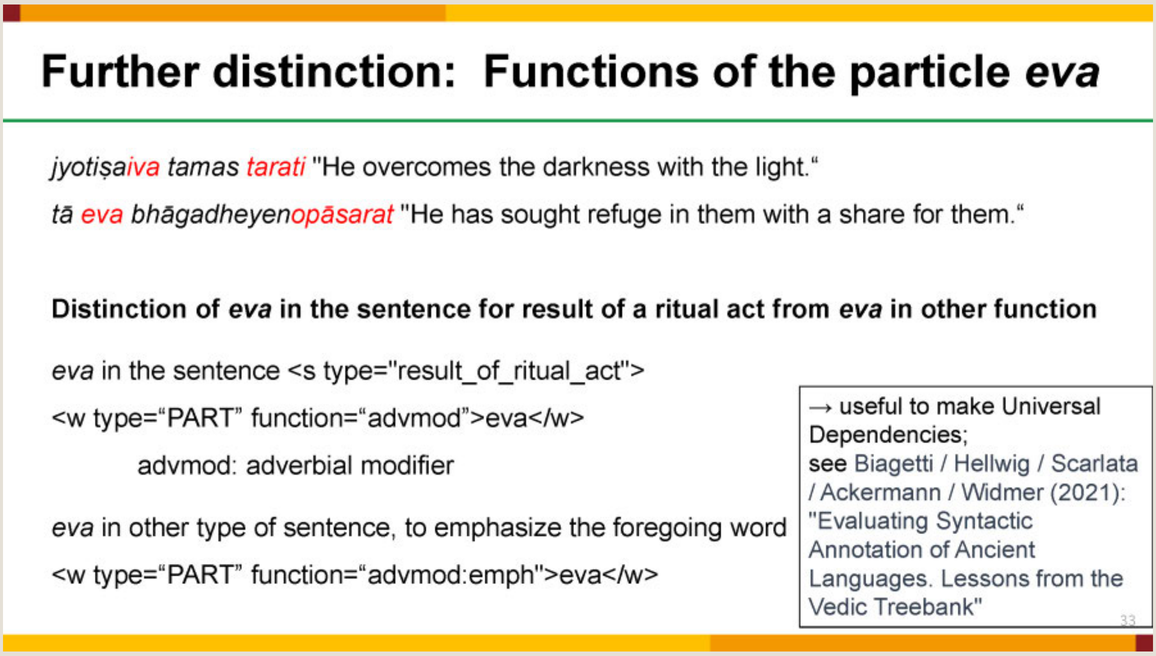
Ratio of these constructions can be useful as indicator of linguistic layer

In a sentence describing the result of a ritual act,
1) whether the verb is present or aorist, and
2) whether *eva* or *vā etad* stands

- Aorist in this function is MS-specific. KS and TS have less examples for this aorist
- Also *eva* or *vā etad* can indicate the wording characteristics of each chapter or text.
- For aorist, we should distinguish the functions, "result_of_ritual_act", "ritual_act_done_before", "actual_past". (see Amano 2009, 15-18)

32

33



Further distinction: Functions of the particle *eva*

jyotiṣaiva tamas tarati "He overcomes the darkness with the light.“

tā eva bhāgadheyenopāsarat "He has sought refuge in them with a share for them.“

Distinction of *eva* in the sentence for result of a ritual act from *eva* in other function

eva in the sentence <s type="result_of_ritual_act">

<w type="PART" function="advmod">eva</w>

advmod: adverbial modifier

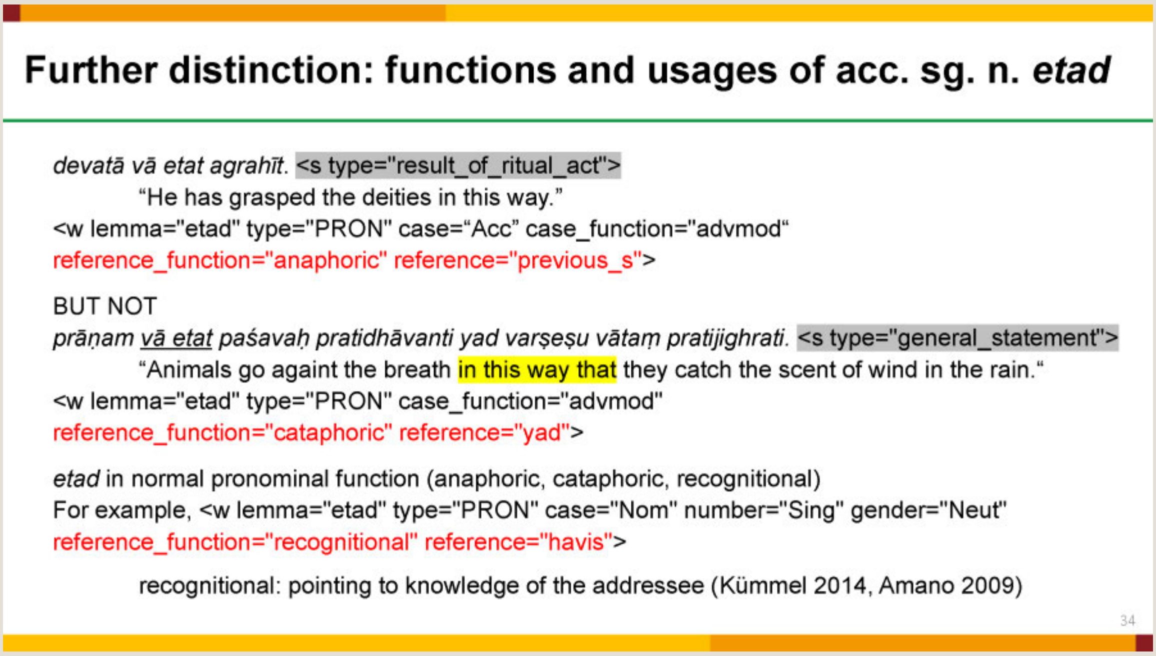
eva in other type of sentence, to emphasize the foregoing word

<w type="PART" function="advmod:emph">eva</w>

→ useful to make Universal Dependencies;
see Biagetti / Hellwig / Scarlata / Ackermann / Widmer (2021): "Evaluating Syntactic Annotation of Ancient Languages. Lessons from the Vedic Treebank"

33

34



Further distinction: functions and usages of acc. sg. n. *etad*

devatā vā etat agrahīt. <s type="result_of_ritual_act">

“He has grasped the deities in this way.”

<w lemma="etad" type="PRON" case="Acc" case_function="advmod“
reference_function="anaphoric" reference="previous_s">

BUT NOT

prāṇam vā etat paśavaḥ pratidhāvanti yad varṣeṣu vātaṃ pratijighrati. <s type="general_statement">

“Animals go againt the breath in this way that they catch the scent of wind in the rain.“

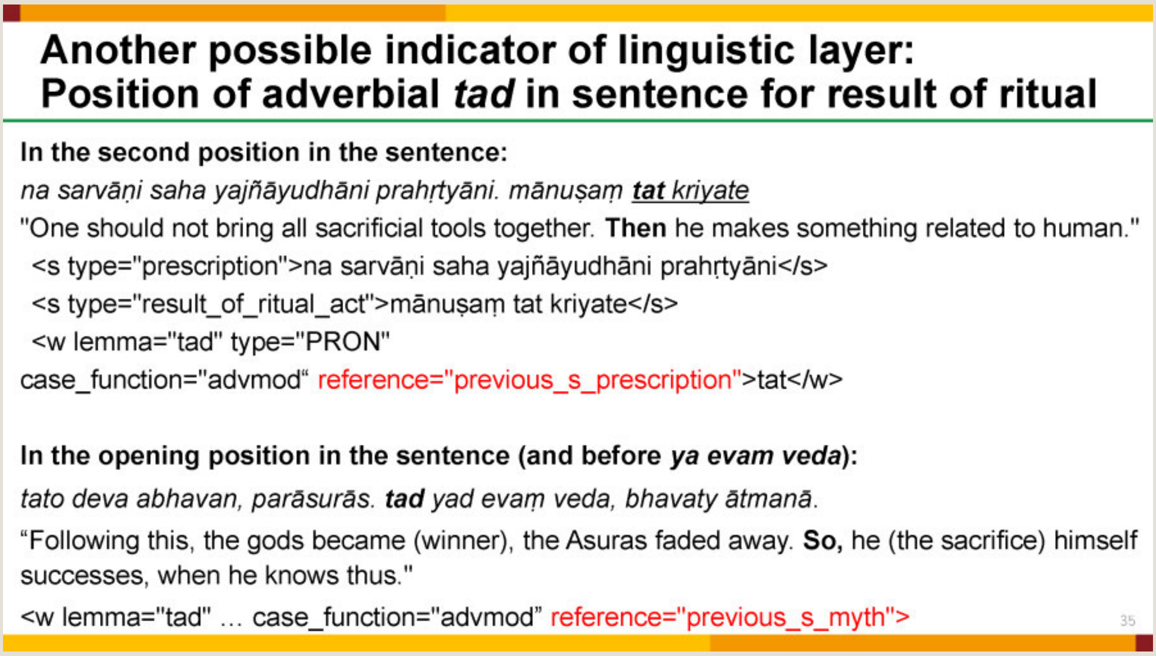
<w lemma="etad" type="PRON" case_function="advmod"
reference_function="cataphoric" reference="yad">

etad in normal pronominal function (anaphoric, cataphoric, recognitional)
For example, <w lemma="etad" type="PRON" case="Nom" number="Sing" gender="Neut"
reference_function="recognitional" reference="havis">

recognitional: pointing to knowledge of the addressee (Kümmel 2014, Amano 2009)

34

35



Another possible indicator of linguistic layer: Position of adverbial *tad* in sentence for result of ritual

In the second position in the sentence:

na sarvāṇi saha yajñāyudhāni prahṛtyāni. mānuṣaṃ ***tat*** *kriyate*

"One should not bring all sacrificial tools together. **Then** he makes something related to human."

<s type="prescription">na sarvāṇi saha yajñāyudhāni prahṛtyāni</s>

<s type="result_of_ritual_act">mānuṣaṃ tat kriyate</s>

<w lemma="tad" type="PRON"
case_function="advmod“ reference="previous_s_prescription">tat</w>

In the opening position in the sentence (and before *ya evam veda*):

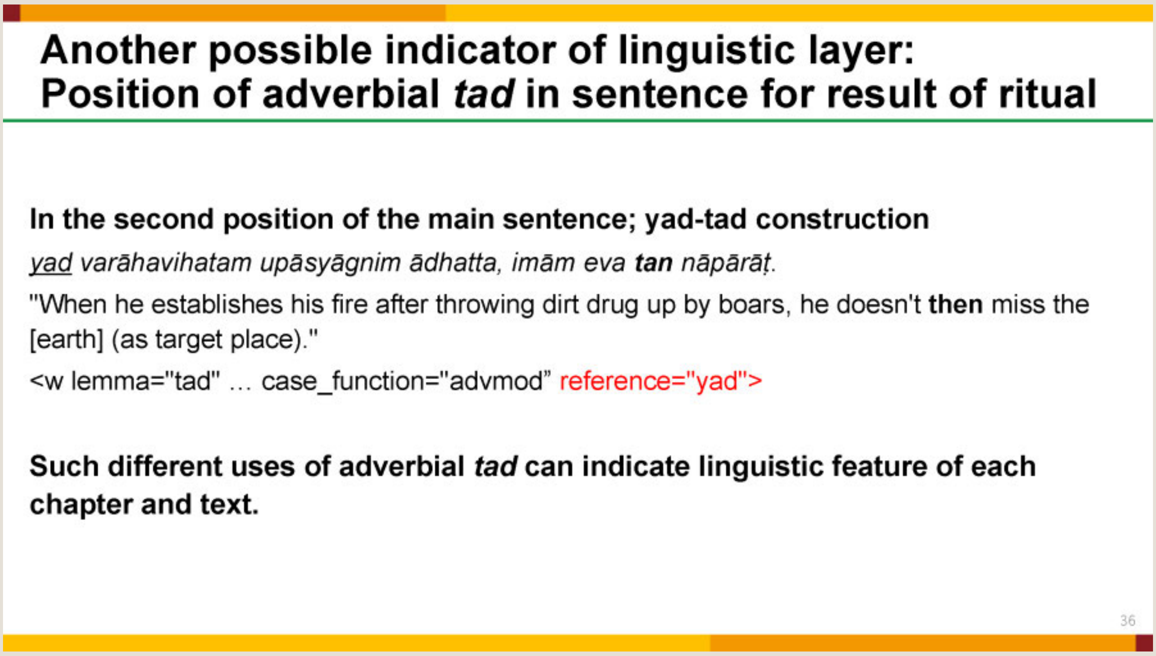
tato deva abhavan, parāsurās. ***tad*** *yad evaṃ veda, bhavaty ātmanā.*

“Following this, the gods became (winner), the Asuras faded away. **So,** he (the sacrifice) himself successes, when he knows thus."

<w lemma="tad" … case_function="advmod” reference="previous_s_myth">

35

36



Another possible indicator of linguistic layer: Position of adverbial *tad* in sentence for result of ritual

In the second position of the main sentence; yad-tad construction

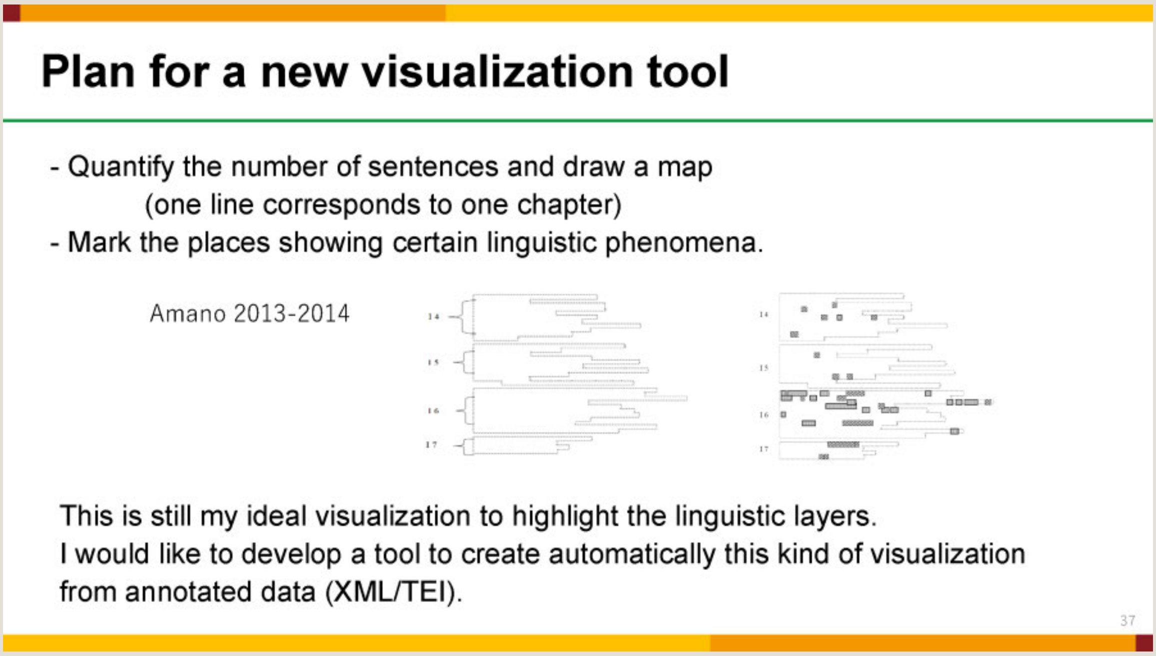
yad varāhavihatam upāsyāgnim ādhatta, imām eva ***tan*** *nāpārāṭ.*

"When he establishes his fire after throwing dirt drug up by boars, he doesn't **then** miss the [earth] (as target place)."

<w lemma="tad" … case_function="advmod" reference="yad">

Such different uses of adverbial *tad* can indicate linguistic feature of each chapter and text.

37



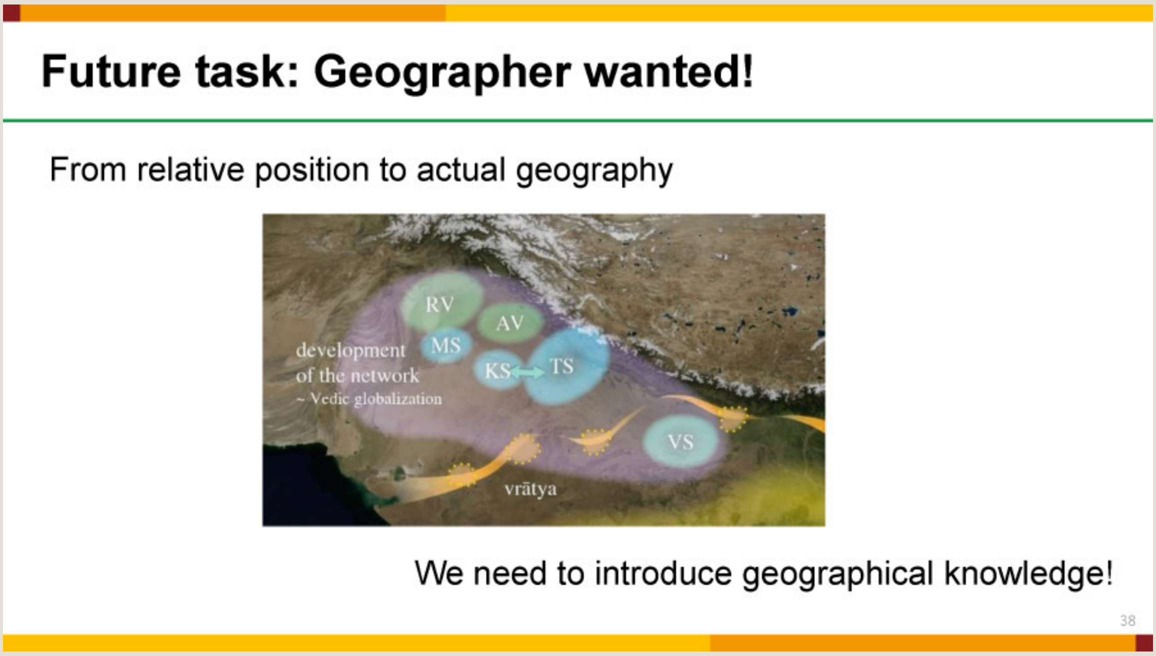
Plan for a new visualization tool

- Quantify the number of sentences and draw a map
  (one line corresponds to one chapter)
- Mark the places showing certain linguistic phenomena.

Amano 2013-2014

This is still my ideal visualization to highlight the linguistic layers.
I would like to develop a tool to create automatically this kind of visualization from annotated data (XML/TEI).

38



Future task: Geographer wanted!

From relative position to actual geography

We need to introduce geographical knowledge!

Presentation 2 ｜報告 2

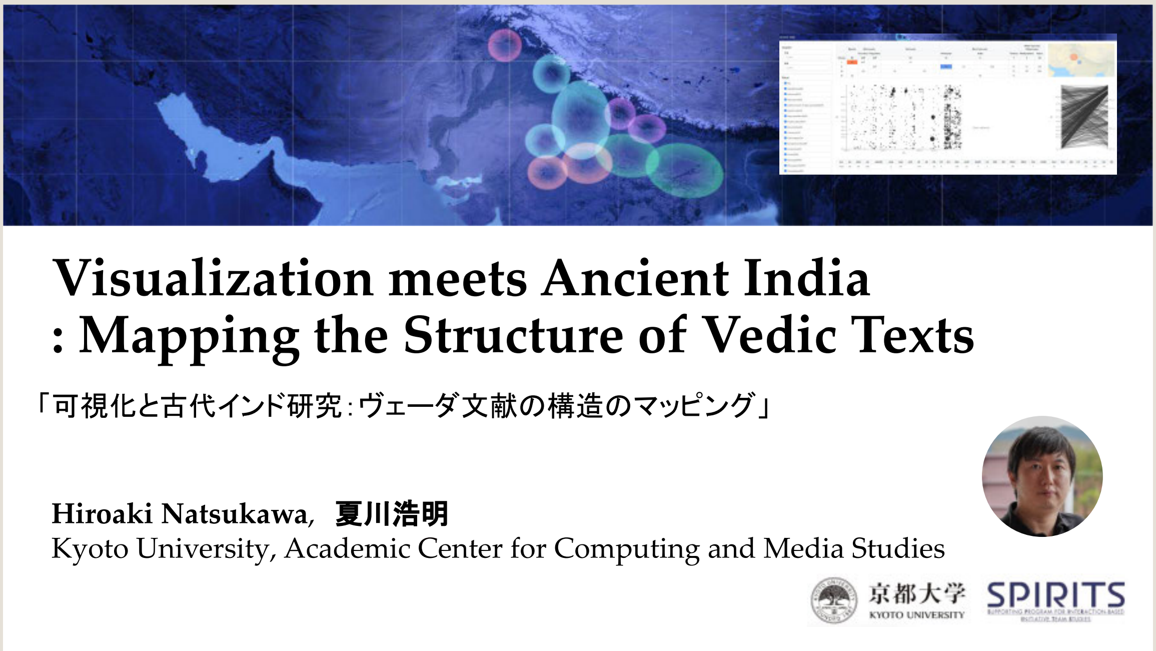
Visualization meets Ancient India: Mapping the Structure of Vedic Texts

可視化と古代インド研究：ヴェーダ文献の構造のマッピング

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

夏川浩明（京都大学 学術情報メディアセンター）

1



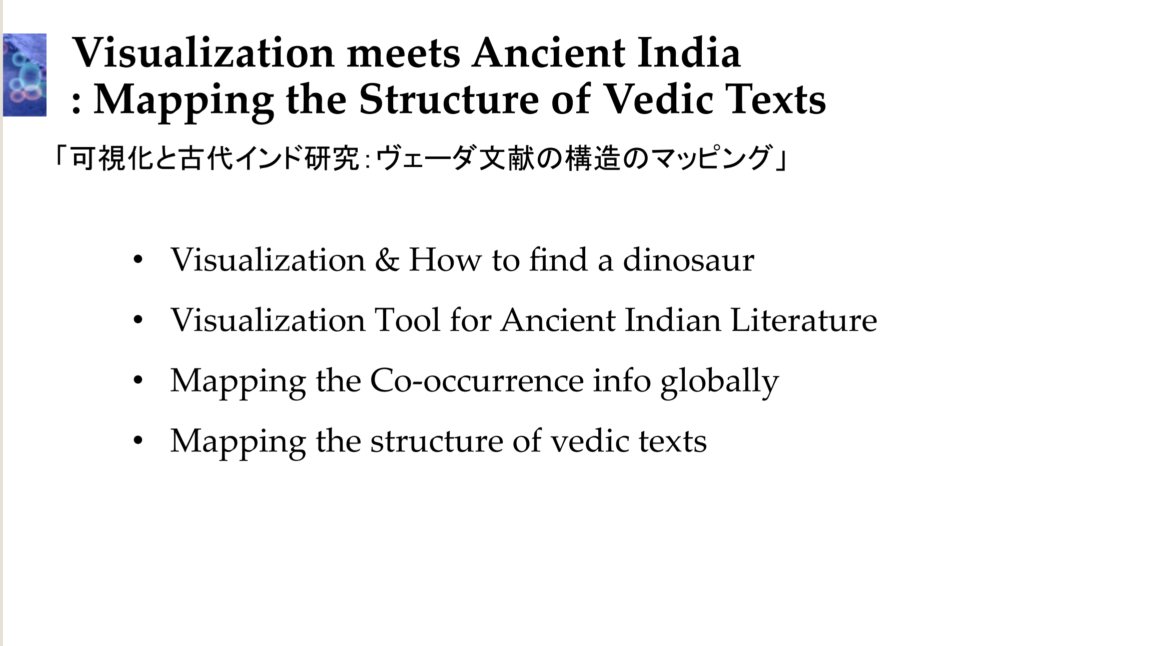
Visualization meets Ancient India
: Mapping the Structure of Vedic Texts

「可視化と古代インド研究：ヴェーダ文献の構造のマッピング」

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

京都大学
KYOTO UNIVERSITY
SPIRITS

2



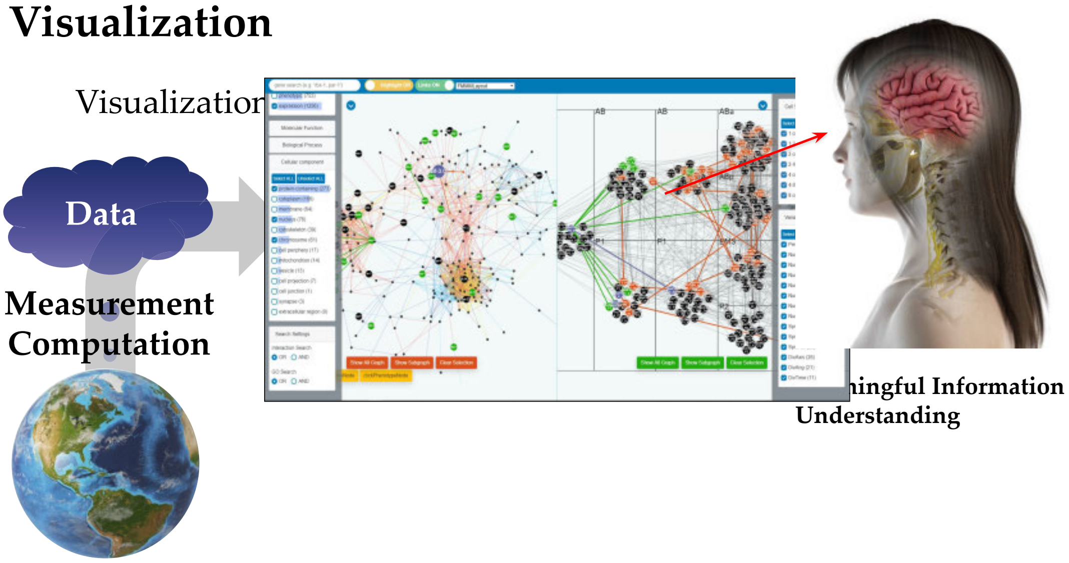
Visualization meets Ancient India
: Mapping the Structure of Vedic Texts

「可視化と古代インド研究：ヴェーダ文献の構造のマッピング」

- Visualization & How to find a dinosaur
- Visualization Tool for Ancient Indian Literature
- Mapping the Co-occurrence info globally
- Mapping the structure of vedic texts

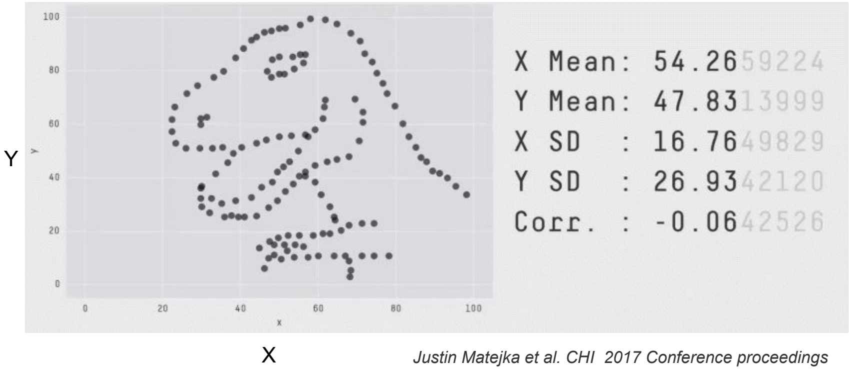
3

Visualization



4

Visualization & Visual Analytics



Justin Matejka et al. CHI 2017 Conference proceedings

Importance of looking at the data itself

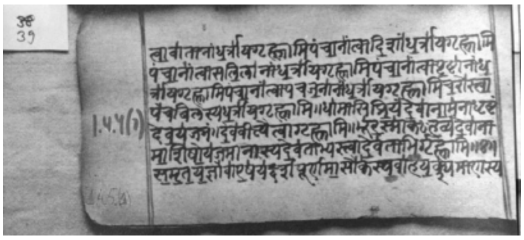
5

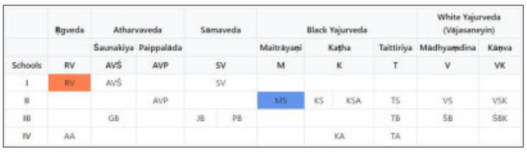
Visualization Contributing to the Analysis of Ancient Indian Literature

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

Ancient Indian ritual texts BC1200-500

- Mantras (祝詞)
- 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	Mādhyaṃdina	Kāṇva
Schools	RV	AVŚ	AVP	SV		M	K		T	V	VK
I	RV	AVŚ		SV							
II			AVP			MS	KS	KSA	TS	VS	VSK
III		GB		JB	PB				TB	ŚB	ŚBK
IV	AA						KA		TA		

6

Database

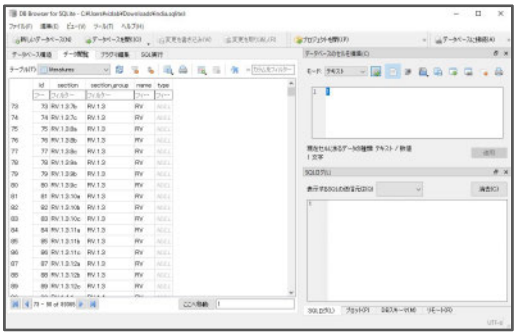
Ancient Indian ritual texts BC1200-500

•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

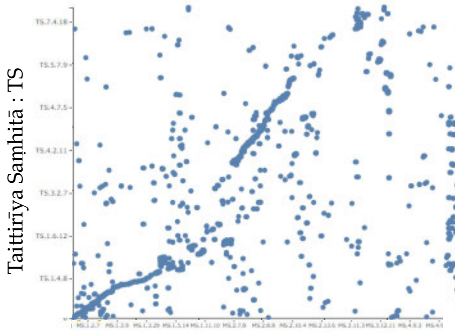
→Relationships among about 150 sets of documents


DB Browser for SQLite - C:¥Users¥clab¥Downloads¥india.sqlite3
ファイル(F) 編集(E) ビュー(V) ツール(T) ヘルプ(H)
新しいデータベース(N) データベースを開く(O) 変更を書き込み(W) 変更を取り消し(R) プロジェクトを開く(P) データベースに接続(A)
データベース構造 データ閲覧 プラグマ編集 SQL実行
テーブル(T) literatures
データベースのセルを編集(C)
モード テキスト
現在セルにあるデータの種類 テキスト / 数値
1 文字
適用
SQLログ(L)
表示するSQLの送信元(O)
消去(C)
SQLログ(L) プロット(P) DBスキーマ(M) リモート(R)
73 - 90 of 81085
ここへ移動
UTF-8

	id	section	section_group	name	type
	フィ	フィルター	フィルター	フィ	フィ
73	73	RV.1.2.7b	RV.1.2	RV	NULL
74	74	RV.1.2.7c	RV.1.2	RV	NULL
75	75	RV.1.2.8a	RV.1.2	RV	NULL
76	76	RV.1.2.8b	RV.1.2	RV	NULL
77	77	RV.1.2.8c	RV.1.2	RV	NULL
78	78	RV.1.2.9a	RV.1.2	RV	NULL
79	79	RV.1.2.9b	RV.1.2	RV	NULL
80	80	RV.1.2.9c	RV.1.2	RV	NULL
81	81	RV.1.2.10a	RV.1.2	RV	NULL
82	82	RV.1.2.10b	RV.1.2	RV	NULL
83	83	RV.1.2.10c	RV.1.2	RV	NULL
84	84	RV.1.2.11a	RV.1.2	RV	NULL
85	85	RV.1.2.11b	RV.1.2	RV	NULL
86	86	RV.1.2.11c	RV.1.2	RV	NULL
87	87	RV.1.2.12a	RV.1.2	RV	NULL
88	88	RV.1.2.12b	RV.1.2	RV	NULL
89	89	RV.1.2.12c	RV.1.2	RV	NULL

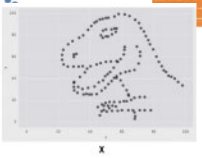
7

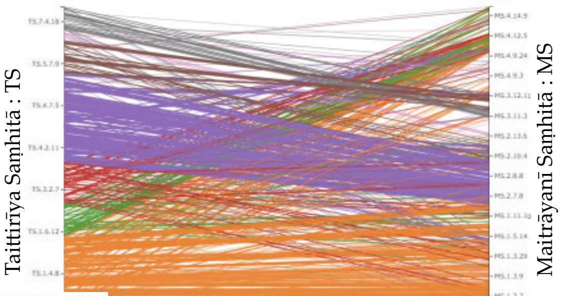
Visualizing the Co-occurrence of Mantras in Ancient Indian Literature



Maitrāyaṇī Saṃhitā : MS

Scatter plot

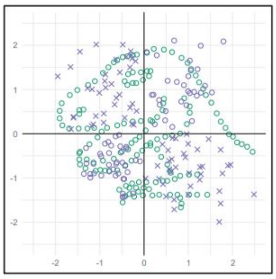




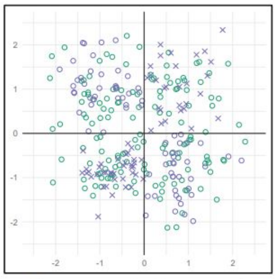
Parallel coordinate plot

8

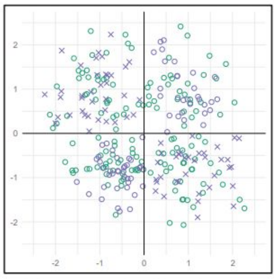
How to find a dinosaur: Inattentional Blindness



An example dinosaur



Plot type 1: Forward-slash ("x"s only in quadrants 1 and 3)



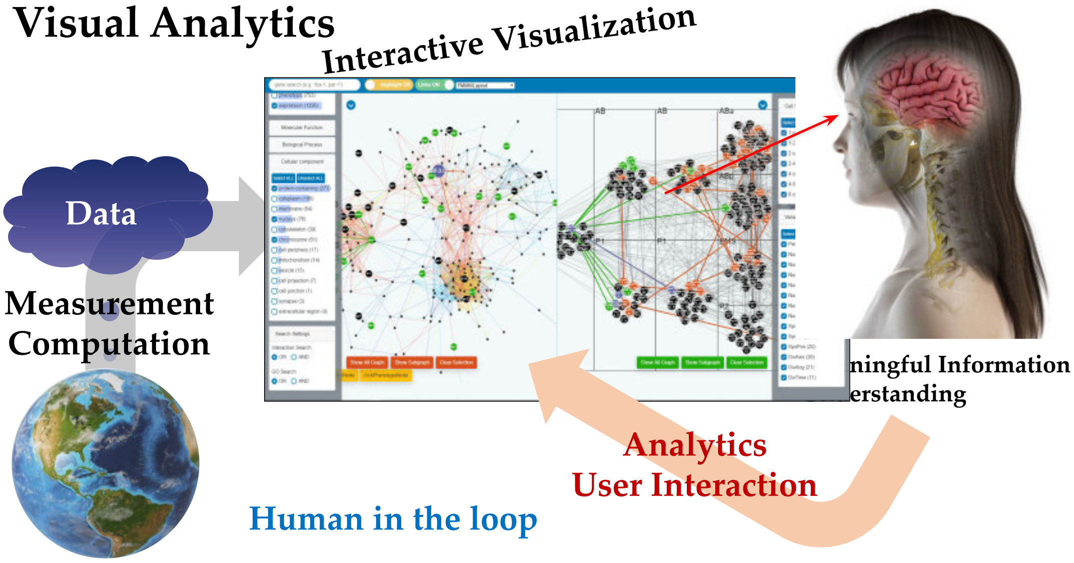
Plot type 2: Back-slash ("x"s only in quadrants 2 and 4)

Tal Boger et al. IEEE VIS 2021 Conference short paper

Importance of iterative process of exploration

9

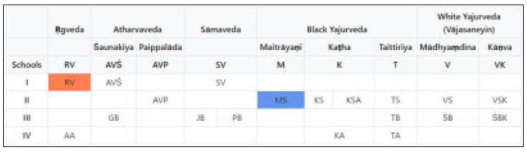
Visual Analytics



10

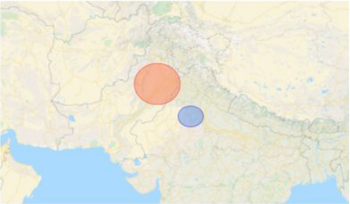
A tool for visualizing the co-occurrence of Mantras

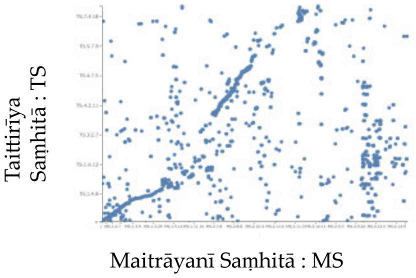
Schools and establishment period

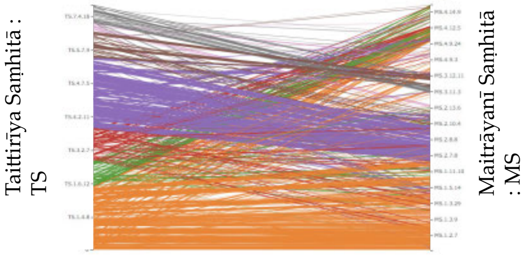


	Ṛgveda	Atharvaveda		Sāmaveda		Black Yajurveda				White Yajurveda (Vājasaneyin)	
		Śaunakiya	Paippalāda			Maitrāyaṇi	Kaṭha		Taittirīya	Mādhyaṃdina	Kāṇva
Schools	RV	AVŚ	AVP	SV		M	K		T	V	VK
I	RV	AVŚ		SV							
II			AVP			MS	KS	KSA	TS	VS	VSK
III		GB		JB	PB				TB	ŚB	ŚBK
IV	AA						KA		TA		

Geographical information





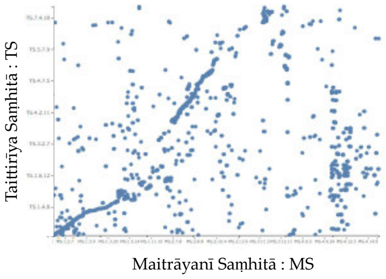


11

A tool for visualizing the co-occurrence of Mantras

Challenges

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



12

A tool for visualizing the co-occurrence of Mantras

User goal: To examine the characteristics and similarities of mantra co-occurrence in different literature. To understand the co-occurrence of mantras by considering the ritual information.

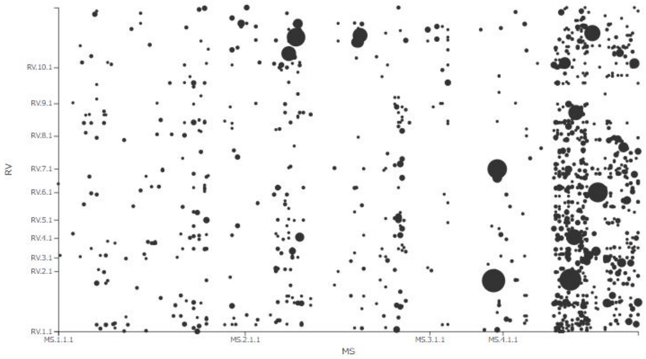
Analysis tasks:
1. to identify the patterns of co-occurrence of mantras in the literature
2. to present the correspondence with the schools and geographical information in the literature
3. link mantra co-occurrences with ritual and chapter information for interpretation.

30 types of ritual event such as Agnicayana(AG).

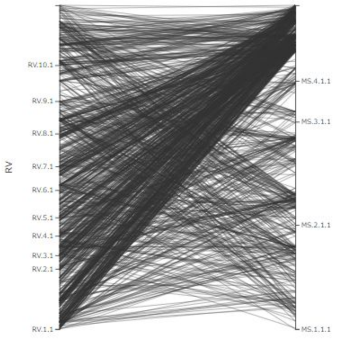
13

Analysis tasks

1. to identify the patterns of co-occurrence of mantras in the literature



Scatter plot

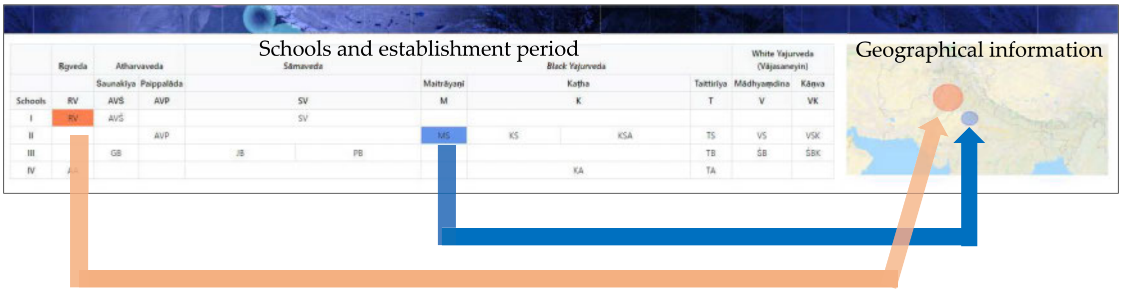


Parallel coordinate plot

14

Analysis tasks

2. to present the correspondence with the schools and geographical information in the literature



Schools and establishment period

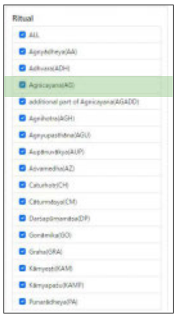
	Ṛgveda	Atharvaveda		Sāmaveda		Black Yajurveda				White Yajurveda (Vājasaneyin)	
		Śaunakīya	Paippalāda			Maitrāyaṇī	Kaṭha		Taittirīya	Mādhyaṃdina	Kāṇva
Schools	RV	AVŚ	AVP	SV		M	K		T	V	VK
I	RV	AVŚ		SV							
II			AVP			MS	KS	KSA	TS	VS	VSK
III		GB		JB	PB				TB	ŚB	ŚBK
IV	AA						KA		TA		

Geographical information

15

Analysis tasks

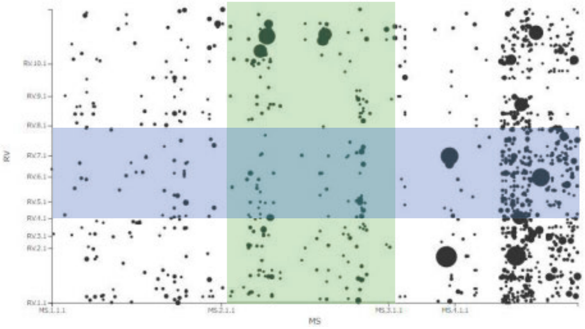
3. link mantra co-occurrences with ritual and chapter information for interpretation.

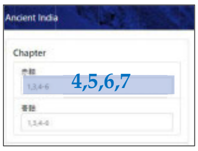


Ritual

- ☑ ALL
- ☑ Agnyādheya(AA)
- ☑ Adhvara(ADH)
- ☑ Agnicayana(AG)
- ☑ additional part of Agnicayana(AGADD)
- ☑ Agnihotra(AGH)
- ☑ Agnyupasthāna(AGU)
- ☑ Aupānuvākya(AUP)
- ☑ Aśvamedha(AZ)
- ☑ Caturhotṛ(CH)
- ☑ Cāturmāsya(CM)
- ☑ Darśapūrṇamāsa(DP)
- ☑ Gonāmika(GO)
- ☑ Graha(GRA)
- ☑ Kāmyeṣṭi(KAM)
- ☑ Kāmyapaśu(KAMP)
- ☑ Punarādheya(PA)

Ritual filtering





Ancient India

Chapter

4,5,6,7

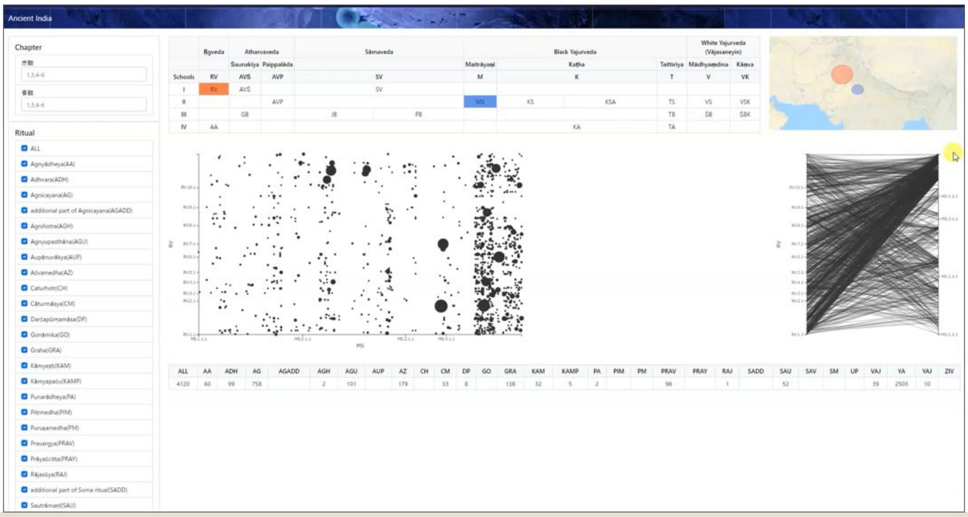
1,3,4-6

1,3,4-6

Chapter filtering

16

A tool for visualizing the co-occurrence of Mantras



Ancient India

Chapter

1,3,4-6

1,3,4-6

Ritual

- ☑ ALL
- ☑ Agnyādheya(AA)
- ☑ Adhvara(ADH)
- ☑ Agnicayana(AG)
- ☑ additional part of Agnicayana(AGADD)
- ☑ Agnihotra(AGH)
- ☑ Agnyupasthāna(AGU)
- ☑ Aupānuvākya(AUP)
- ☑ Aśvamedha(AZ)
- ☑ Caturhotṛ(CH)
- ☑ Cāturmāsya(CM)
- ☑ Darśapūrṇamāsa(DP)
- ☑ Gonāmika(GO)
- ☑ Graha(GRA)
- ☑ Kāmyeṣṭi(KAM)
- ☑ Kāmyapaśu(KAMP)
- ☑ Punarādheya(PA)
- ☑ Pitṛmedha(PIM)
- ☑ Puruṣamedha(PM)
- ☑ Pravargya(PRAV)
- ☑ Prāyaścitta(PRAY)
- ☑ Rājasūya(RAJ)
- ☑ additional part of Soma ritual(SADD)
- ☑ Sautrāmaṇī(SAU)

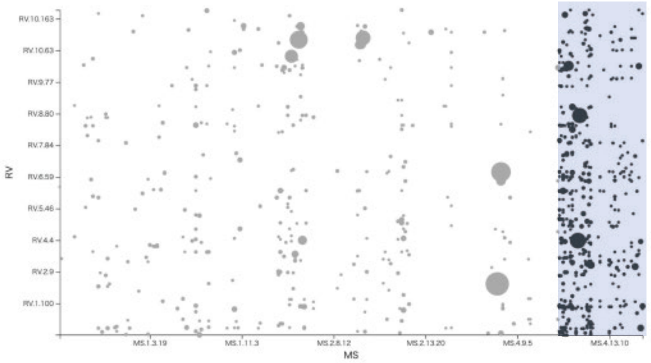
	Ṛgveda	Atharvaveda		Sāmaveda		Black Yajurveda				White Yajurveda (Vājasaneyin)	
		Śaunakīya	Paippalāda			Maitrāyaṇī	Kaṭha		Taittirīya	Mādhyaṃdina	Kāṇva
Schools	RV	AVŚ	AVP	SV		M	K		T	V	VK
I	RV	AVŚ		SV							
II			AVP			MS	KS	KSA	TS	VS	VSK
III		GB		JB	PB				TB	ŚB	ŚBK
IV	AA						KA		TA		

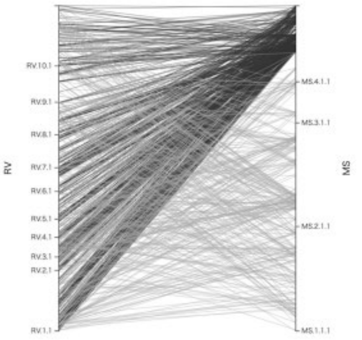
ALL	AA	ADH	AG	AGADD	AGH	AGU	AUP	AZ	CH	CM	DP	GO	GRA	KAM	KAMP	PA	PIM	PM	PRAV	PRAY	RAJ	SADD	SAU	SAV	SM	UP	VAJ	YA	YAJ	ZIV
4120	60	99	758		2	101		179		33	8		138	32	5	2			98		1		52				39	2503	10	

17

A tool for visualizing the co-occurrence of Mantras

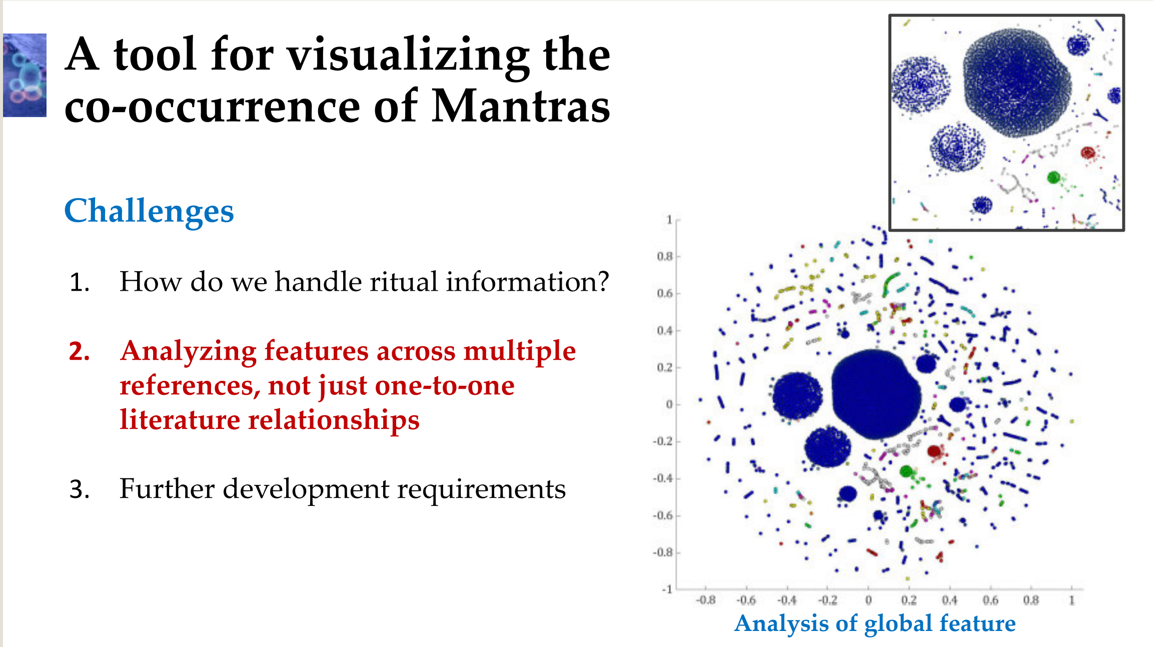
YA: Yājyānuvākyā "Verses to invite and worship gods"





In MS, mantras related to YA are in chapter 4, but they are distributed over a whole chapter in RV.

18



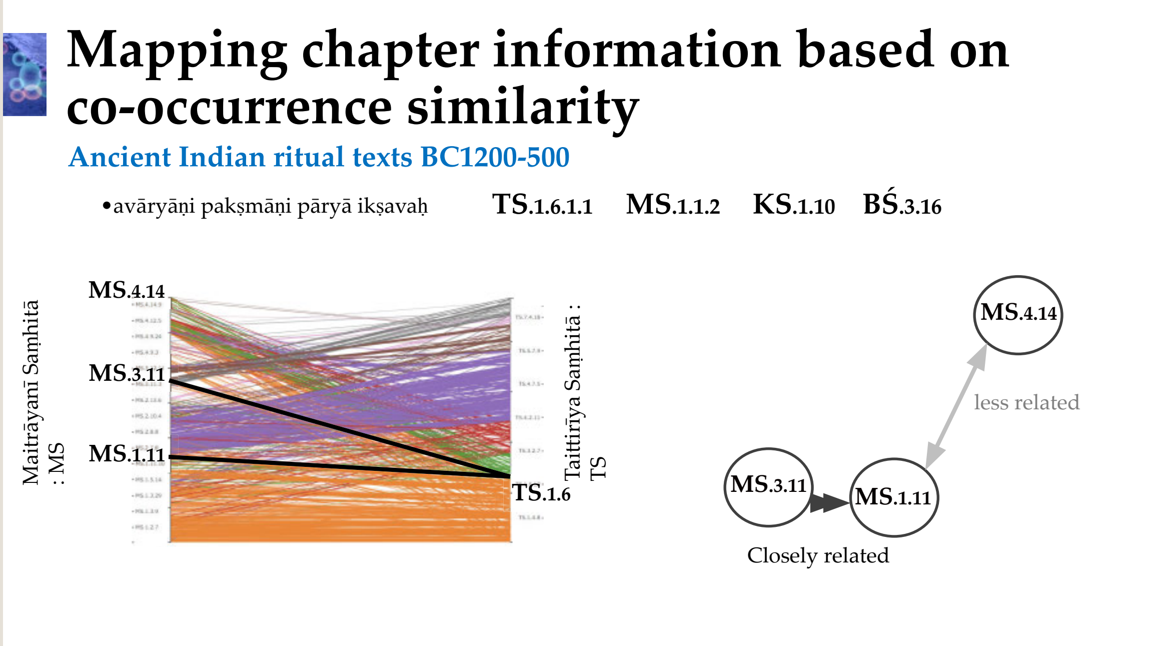
A tool for visualizing the co-occurrence of Mantras

Challenges

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

Analysis of global feature

19

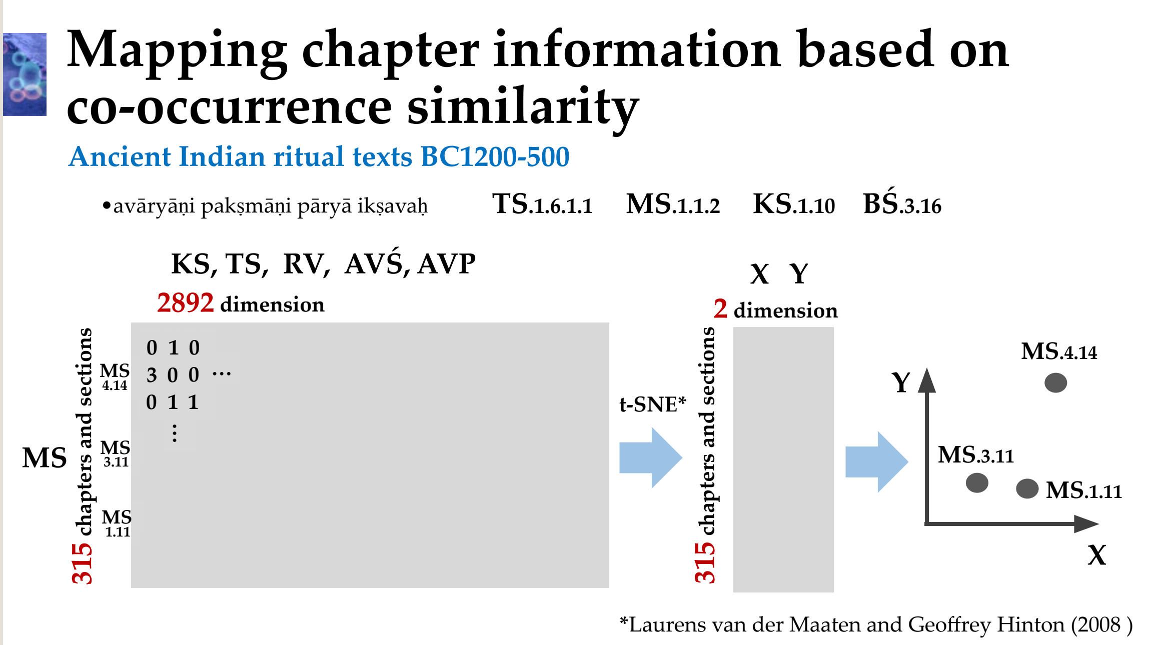


Mapping chapter information based on co-occurrence similarity

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

20



Mapping chapter information based on co-occurrence similarity

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

*Laurens van der Maaten and Geoffrey Hinton (2008)

21

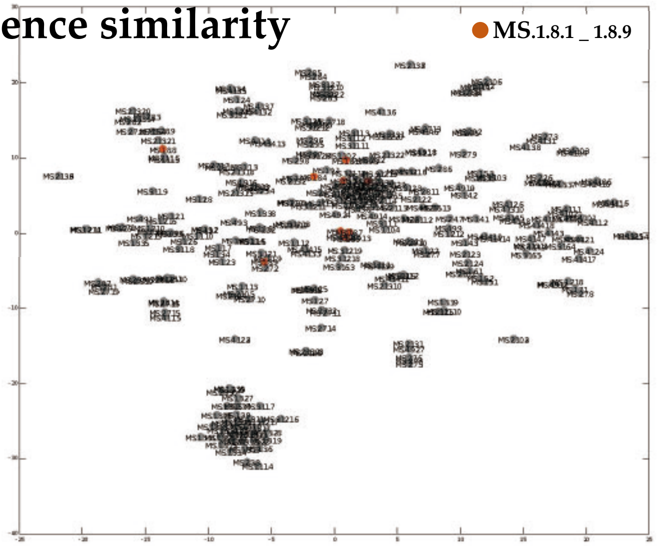
Mapping chapter information based on co-occurrence similarity

Maitrāyanī Saṃhitā : MS

2892 dimensions

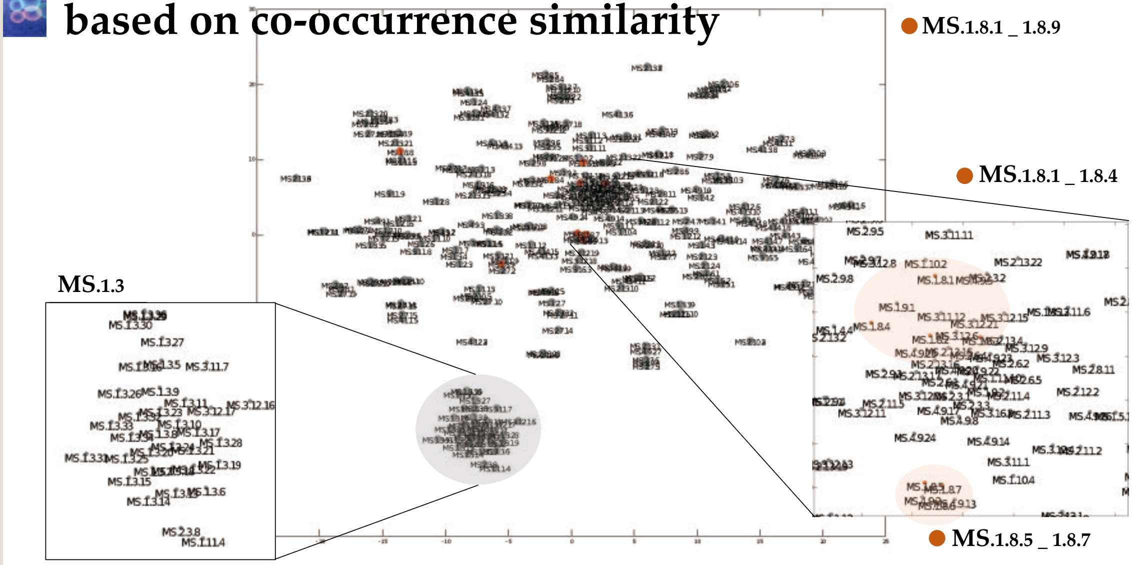
315 chapters and sections

Some clusters can be found based on co-occurrence similarity.



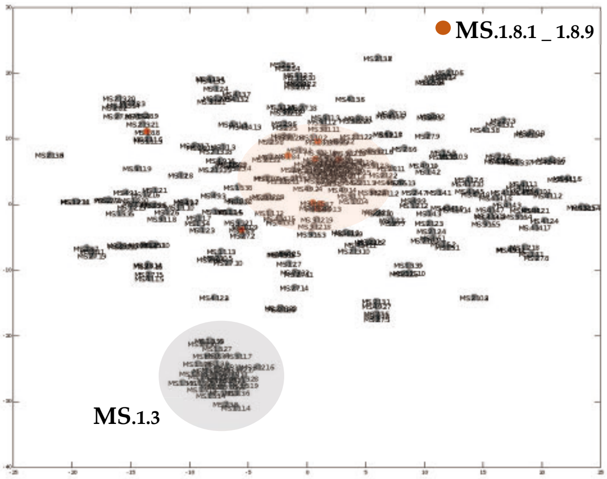
22

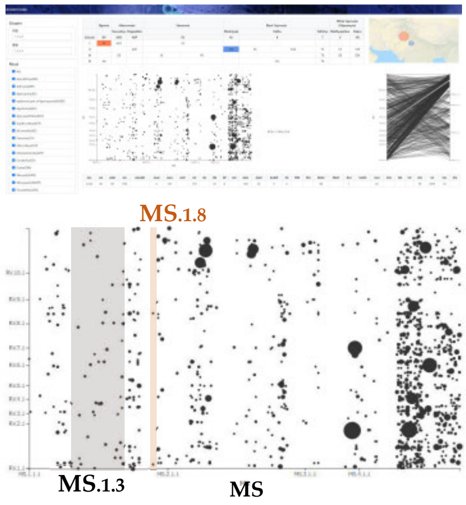
Mapping chapter information based on co-occurrence similarity



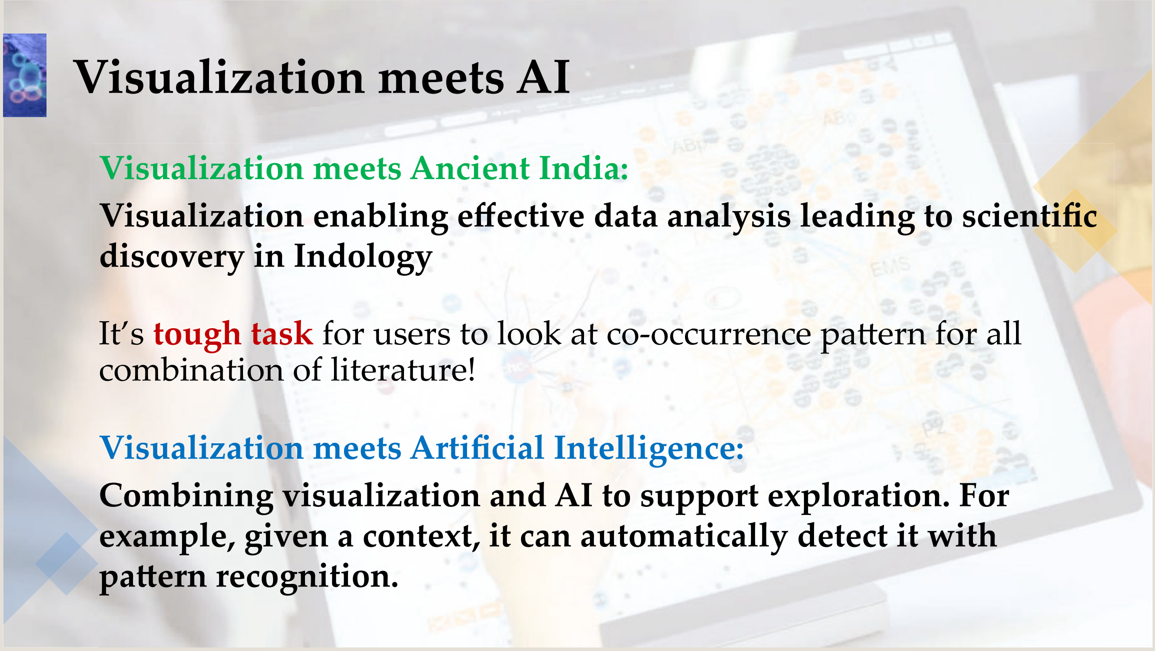
23

Mapping chapter information based on co-occurrence similarity





24



Visualization meets AI

Visualization meets Ancient India:

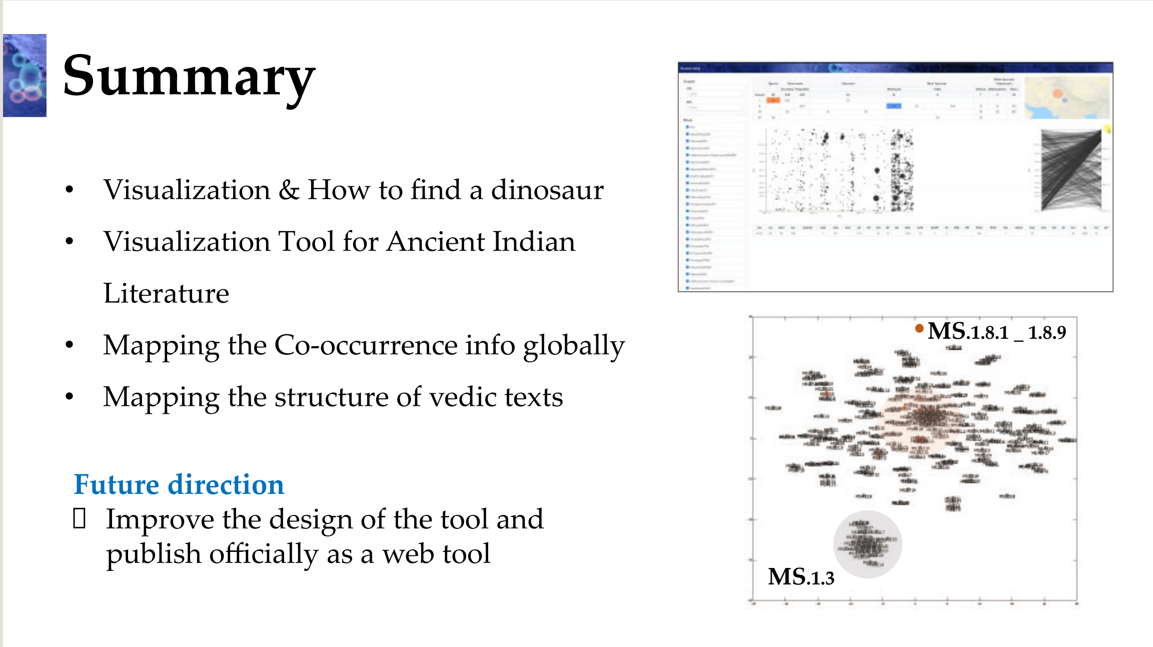
Visualization enabling effective data analysis leading to scientific discovery in Indology

It's **tough task** for users to look at co-occurrence pattern for all combination of literature!

Visualization meets Artificial Intelligence:

Combining visualization and AI to support exploration. For example, given a context, it can automatically detect it with pattern recognition.

25



Summary

- Visualization & How to find a dinosaur
- Visualization Tool for Ancient Indian Literature
- Mapping the Co-occurrence info globally
- Mapping the structure of vedic texts

Future direction

□ Improve the design of the tool and publish officially as a web tool

MS.1.8.1 _ 1.8.9
MS.1.3

26

Thank you for your attention

京都大学
KYOTO UNIVERSITY
SPIRITS

Presentation 3 ｜報告 3

One Step Further: Assessing Semantic Similarity in Sanskrit Using Word Embeddings with a Weighting Factor

検証の次なる段階へ：重み付けを伴う単語分散表現によるサンスクリット文献の類似度推定

Yuki Kyogoku (Leipzig University, Indology)

京極祐希（Leipzig University, Indology）

1

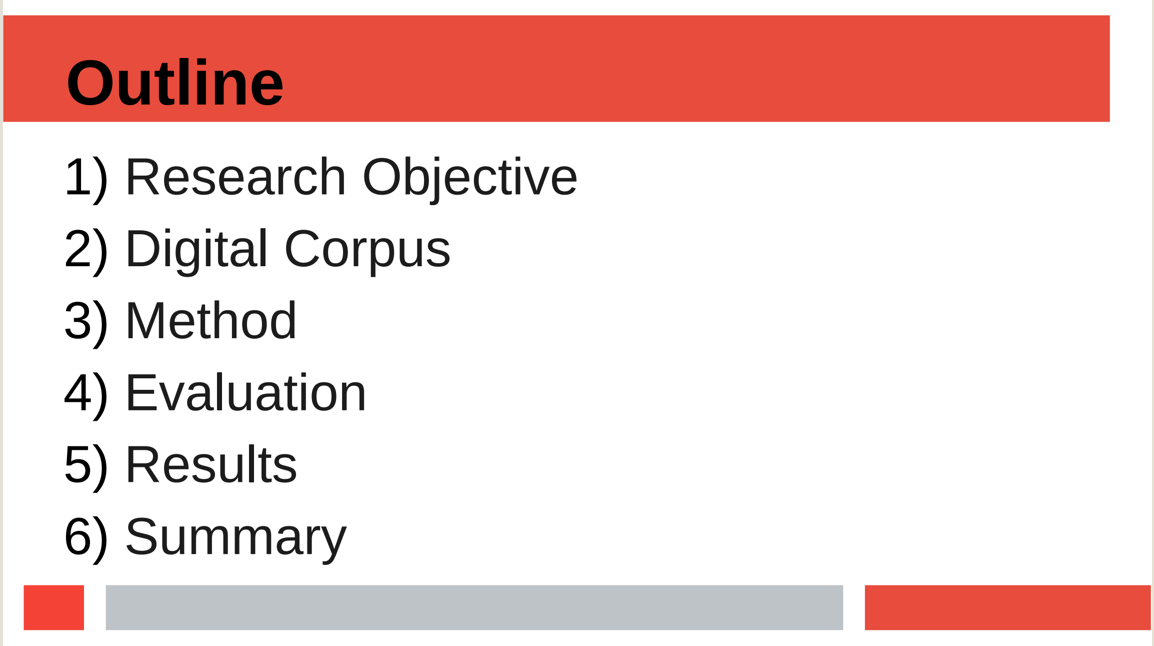
One Step Further:
Assessing Semantic Similarity in Sanskrit
Using Word Embeddings with a Weighting Factor

検証の次なる段階へ：重み付けを伴う単語分散表現によるサンスクリット文献の類似度推定

Yuki Kyogoku
Doctoral Candidate, Institute of Indology

UNIVERSITÄT
LEIPZIG

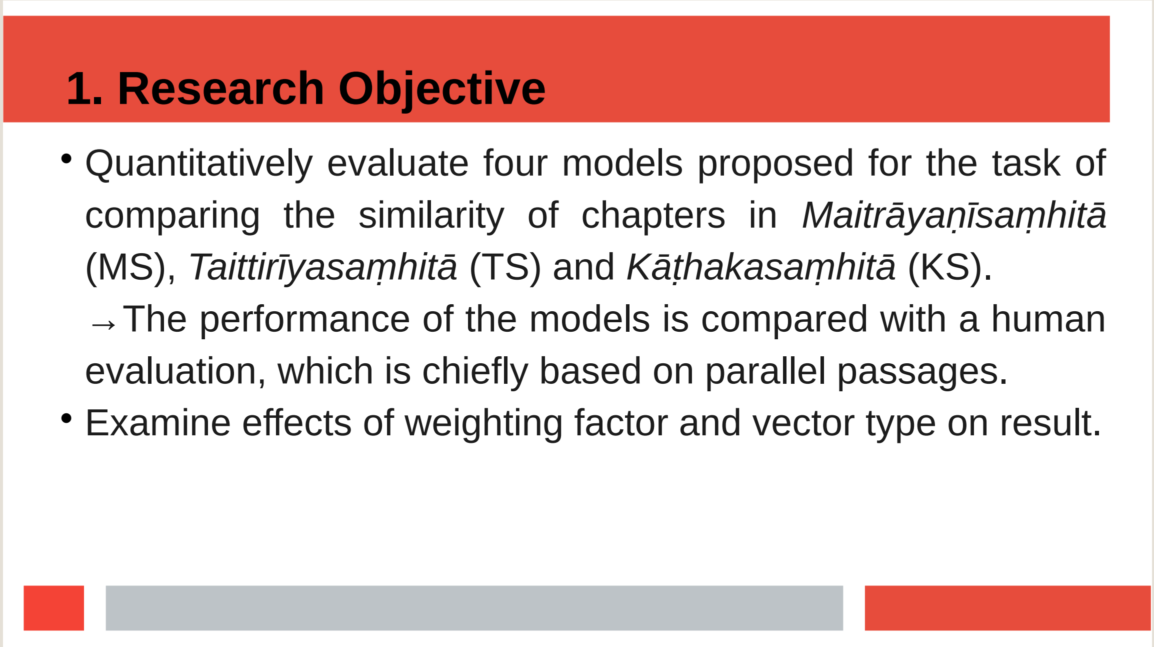
2



Outline

1) Research Objective
2) Digital Corpus
3) Method
4) Evaluation
5) Results
6) Summary

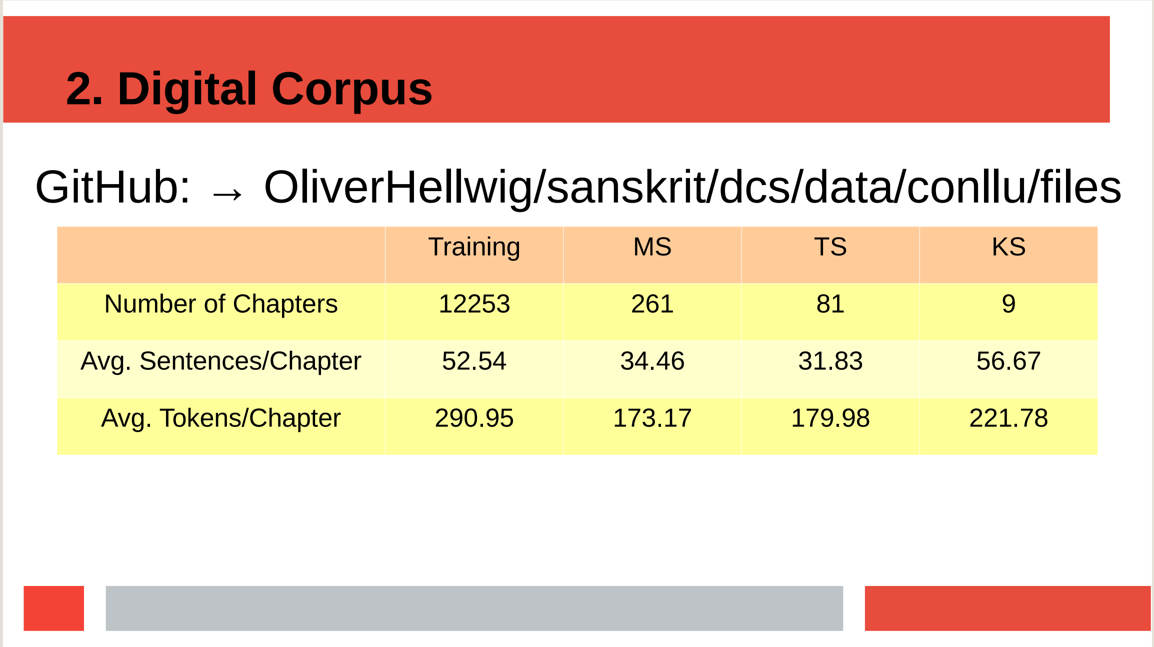
3



1. Research Objective

- Quantitatively evaluate four models proposed for the task of comparing the similarity of chapters in *Maitrāyaṇīsaṃhitā* (MS), *Taittirīyasaṃhitā* (TS) and *Kāṭhakasaṃhitā* (KS).
  → The performance of the models is compared with a human evaluation, which is chiefly based on parallel passages.
- Examine effects of weighting factor and vector type on result.

4

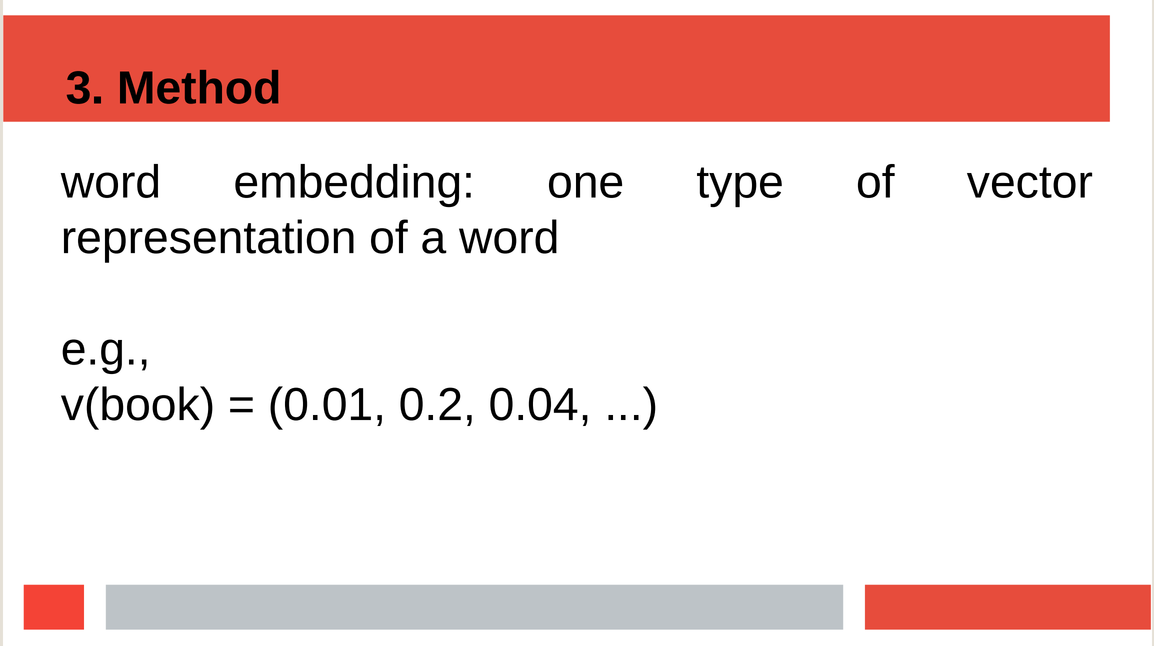


2. Digital Corpus

GitHub: → OliverHellwig/sanskrit/dcs/data/conllu/files

	Training	MS	TS	KS
Number of Chapters	12253	261	81	9
Avg. Sentences/Chapter	52.54	34.46	31.83	56.67
Avg. Tokens/Chapter	290.95	173.17	179.98	221.78

5

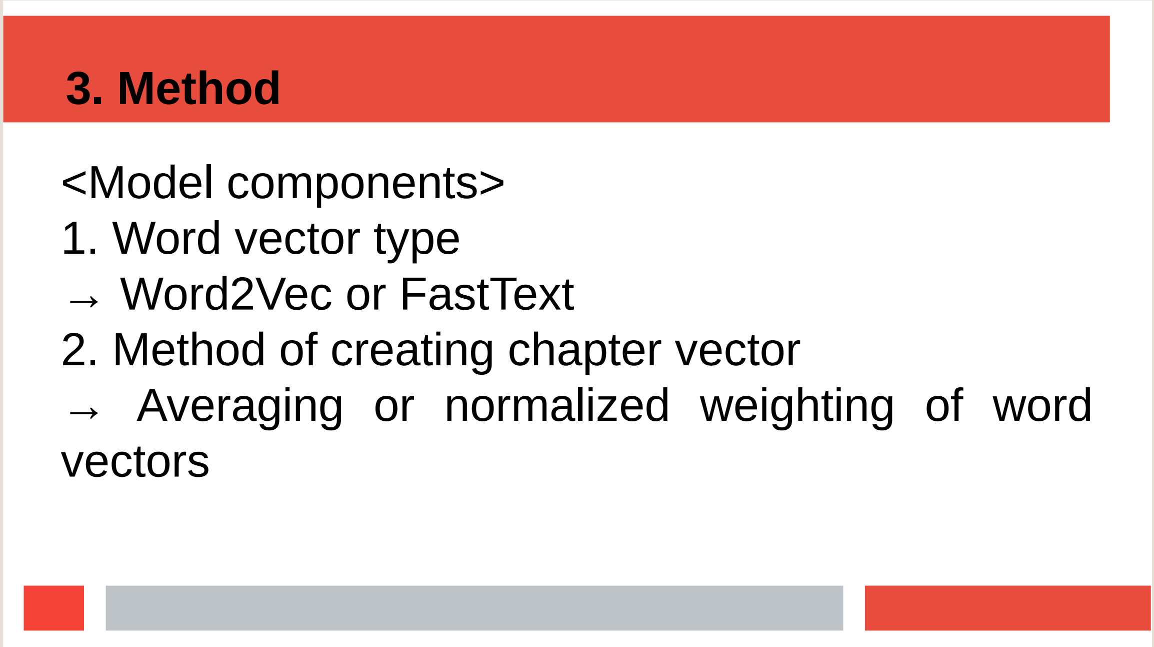


3. Method

word embedding: one type of vector representation of a word

e.g.,
v(book) = (0.01, 0.2, 0.04, ...)

6



3. Method

<Model components>

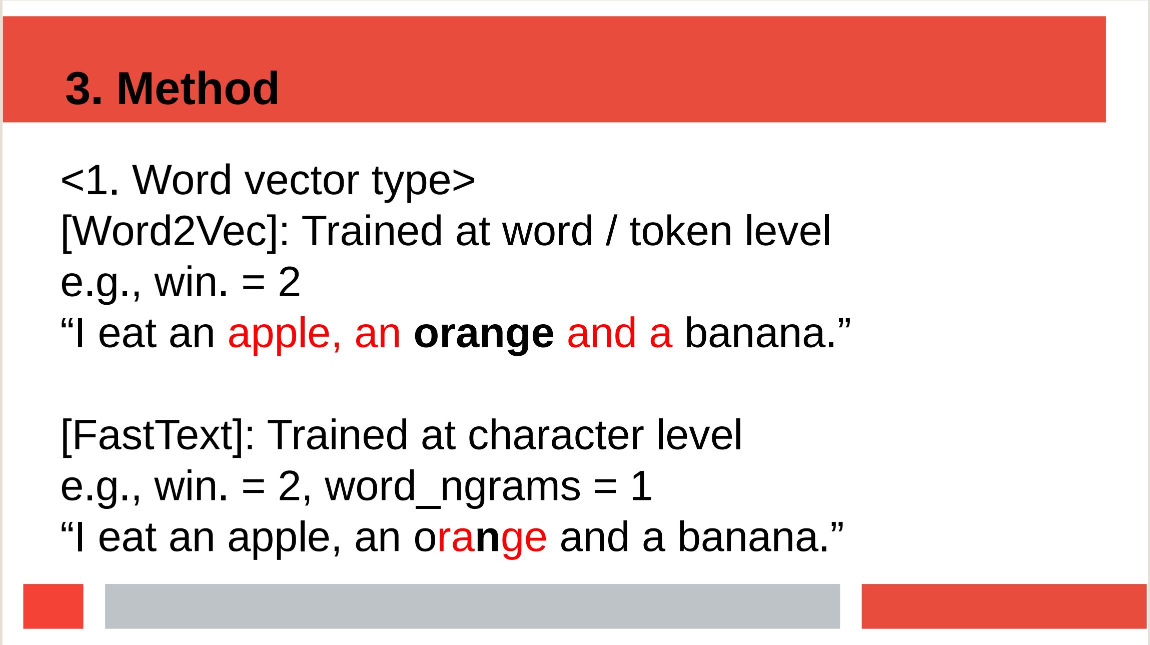
1. Word vector type

→ Word2Vec or FastText

2. Method of creating chapter vector

→ Averaging or normalized weighting of word vectors

7



3. Method

<1. Word vector type>

[Word2Vec]: Trained at word / token level

e.g., win. = 2

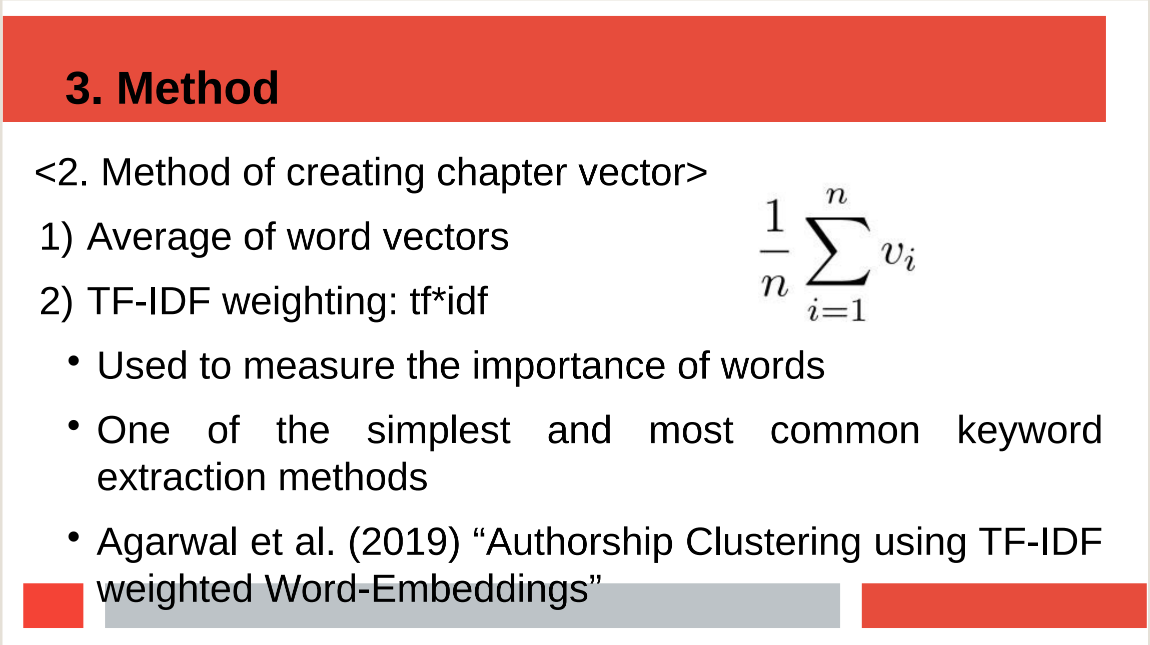
“I eat an apple, an **orange** and a banana.”

[FastText]: Trained at character level

e.g., win. = 2, word_ngrams = 1

“I eat an apple, an ora**n**ge and a banana.”

8



3. Method

<2. Method of creating chapter vector>

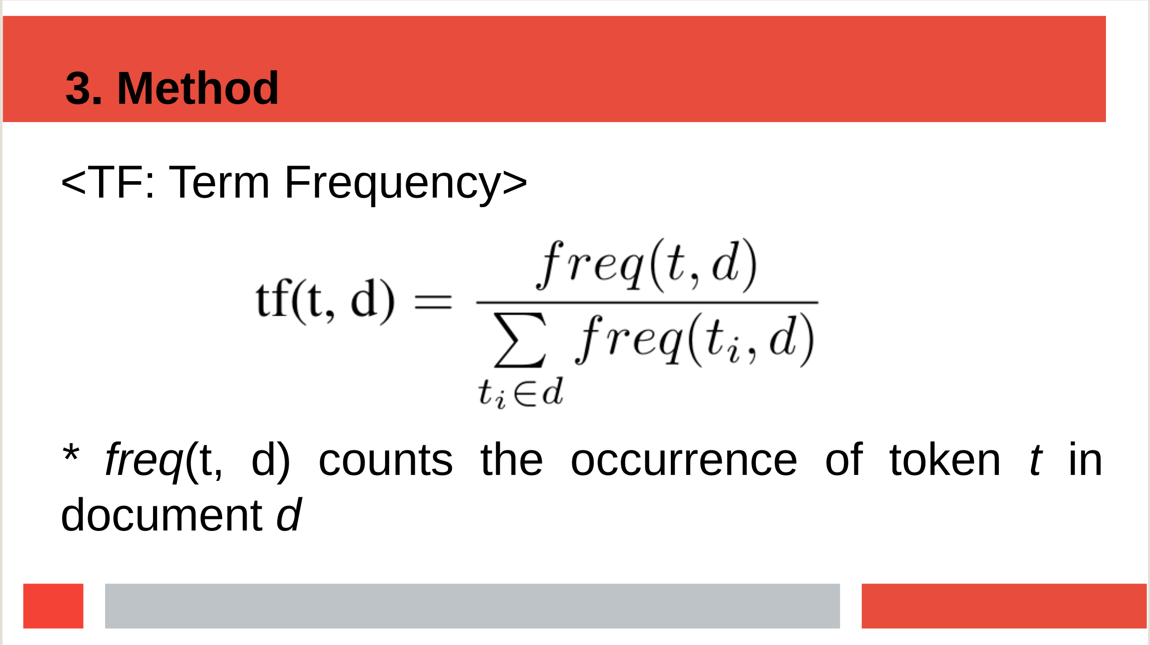
1) Average of word vectors

$$\frac{1}{n}\sum_{i=1}^{n} v_i$$

2) TF-IDF weighting: tf*idf

- Used to measure the importance of words
- One of the simplest and most common keyword extraction methods
- Agarwal et al. (2019) “Authorship Clustering using TF-IDF weighted Word-Embeddings”

9



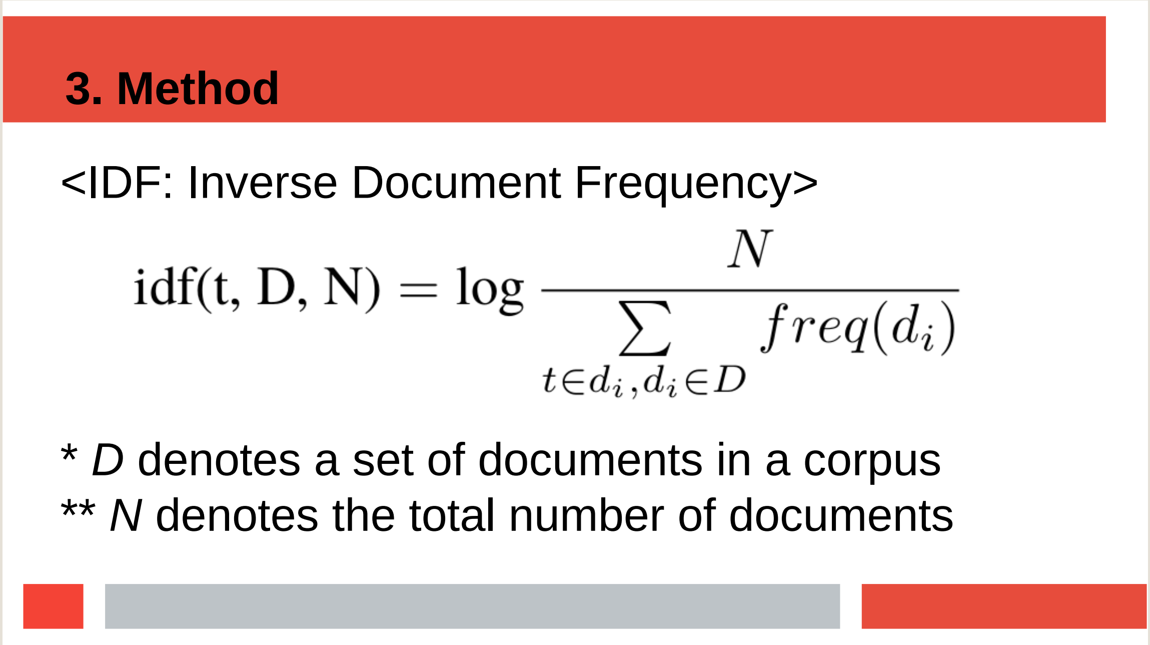
3. Method

<TF: Term Frequency>

$$\mathrm{tf(t, d)} = \frac{freq(t,d)}{\sum_{t_i \in d} freq(t_i,d)}$$

* *freq*(t, d) counts the occurrence of token *t* in document *d*

10



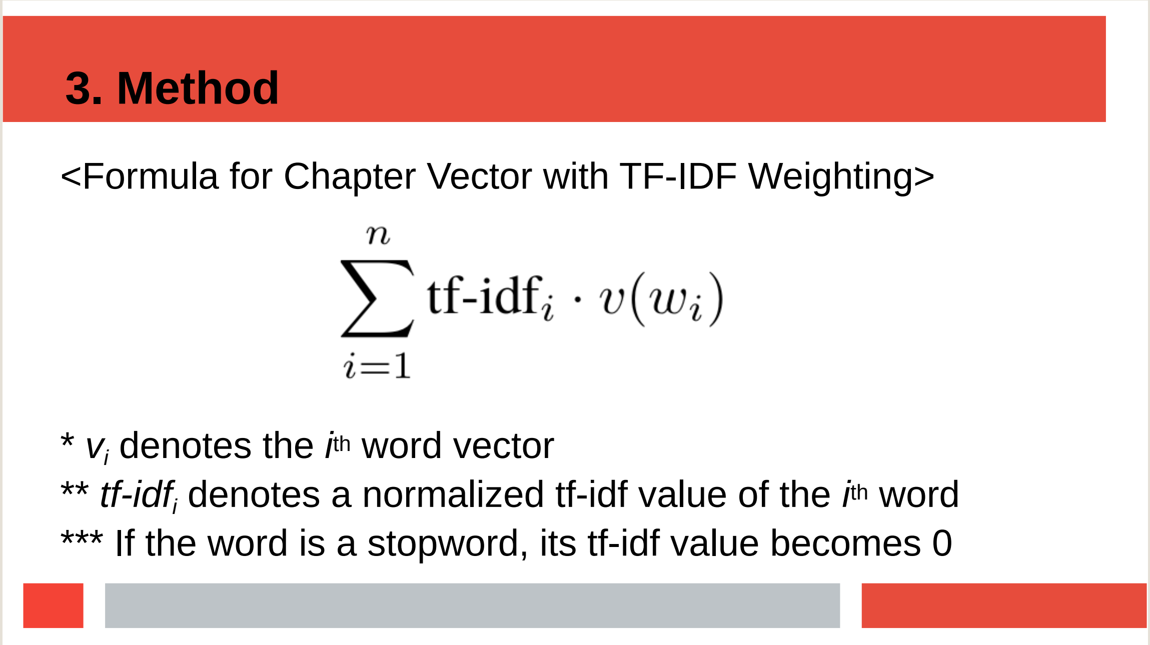
3. Method

<IDF: Inverse Document Frequency>

$$\mathrm{idf(t, D, N)} = \log \frac{N}{\sum_{t \in d_i, d_i \in D} freq(d_i)}$$

* *D* denotes a set of documents in a corpus
** *N* denotes the total number of documents

11



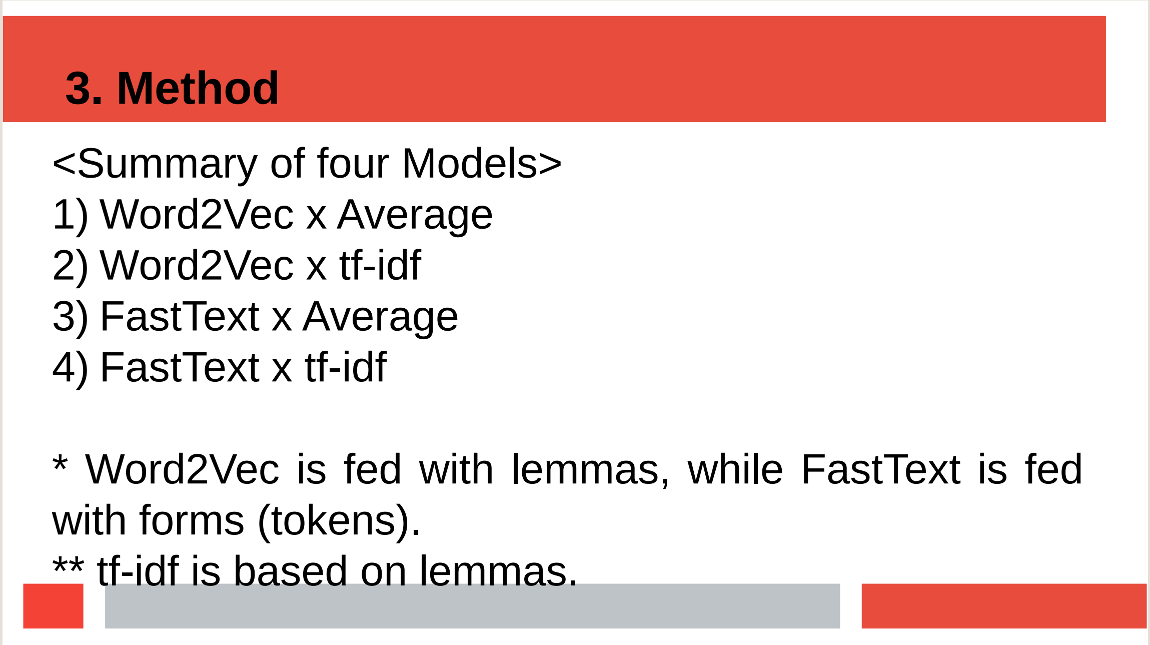
3. Method

<Formula for Chapter Vector with TF-IDF Weighting>

$$\sum_{i=1}^{n} \text{tf-idf}_i \cdot v(w_i)$$

* v_i denotes the ith word vector
** *tf-idf$_i$* denotes a normalized tf-idf value of the ith word
*** If the word is a stopword, its tf-idf value becomes 0

12

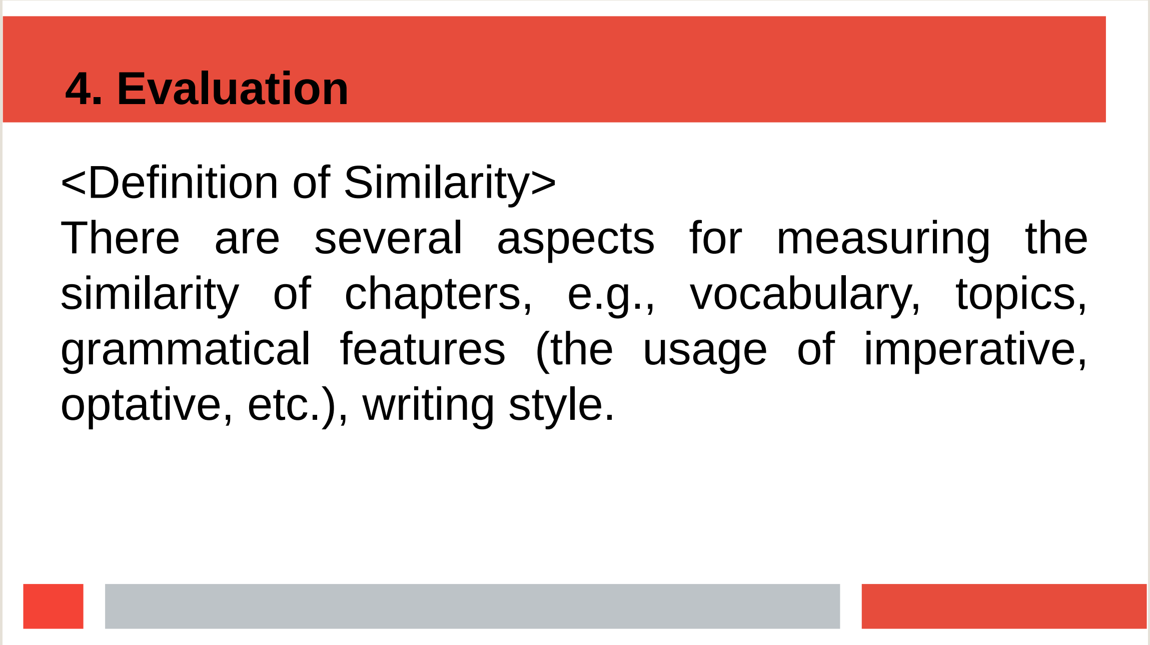


3. Method

<Summary of four Models>
1) Word2Vec x Average
2) Word2Vec x tf-idf
3) FastText x Average
4) FastText x tf-idf

* Word2Vec is fed with lemmas, while FastText is fed with forms (tokens).
** tf-idf is based on lemmas.

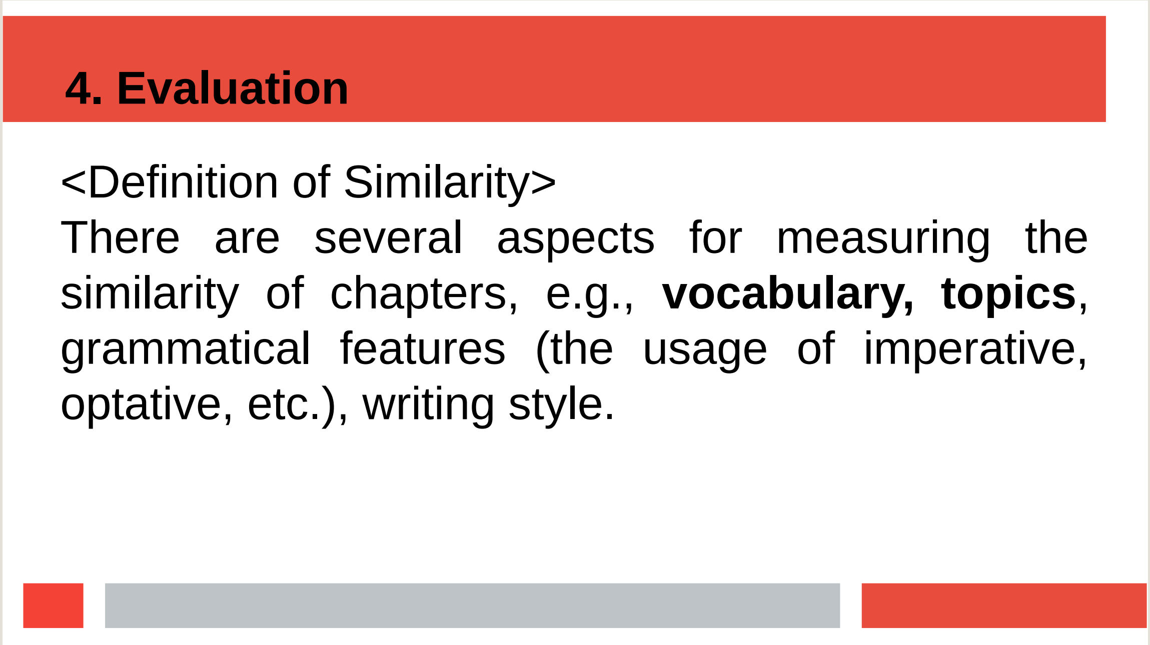
13



4. Evaluation

<Definition of Similarity>
There are several aspects for measuring the similarity of chapters, e.g., vocabulary, topics, grammatical features (the usage of imperative, optative, etc.), writing style.

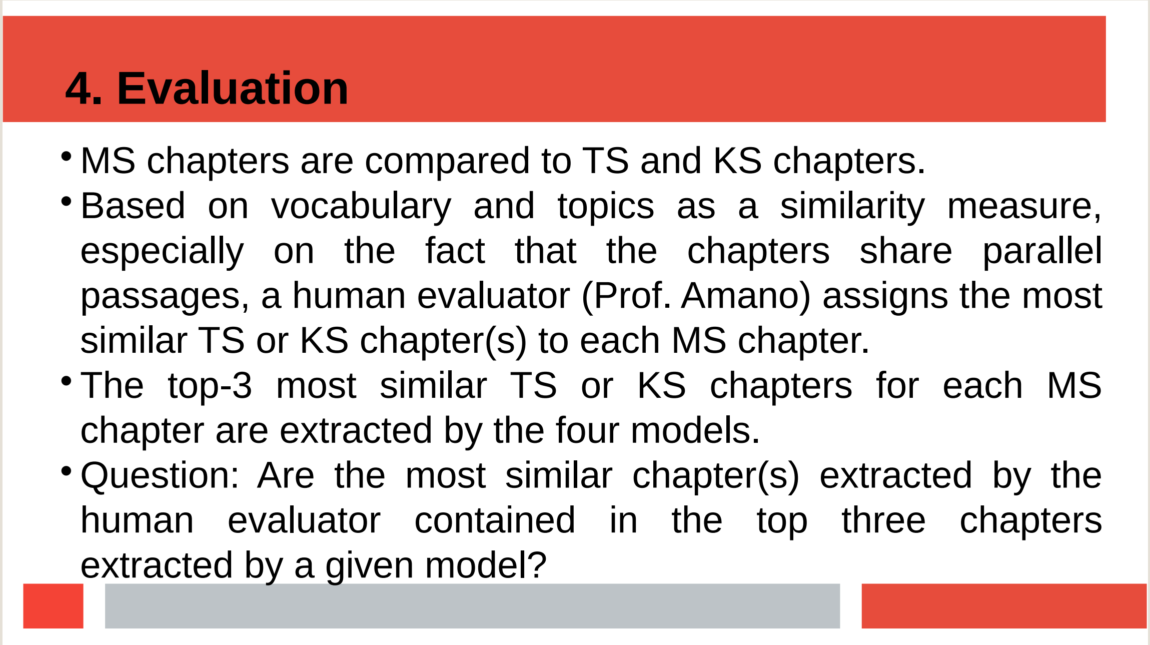
14



4. Evaluation

<Definition of Similarity>
There are several aspects for measuring the similarity of chapters, e.g., **vocabulary, topics**, grammatical features (the usage of imperative, optative, etc.), writing style.

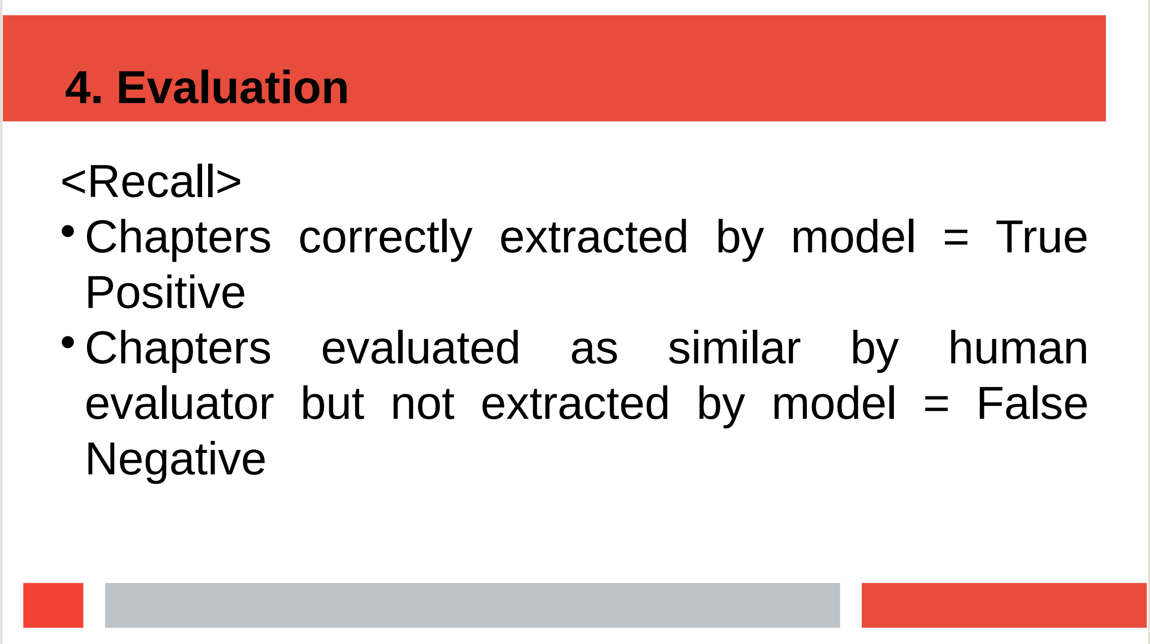
15



4. Evaluation

- MS chapters are compared to TS and KS chapters.
- Based on vocabulary and topics as a similarity measure, especially on the fact that the chapters share parallel passages, a human evaluator (Prof. Amano) assigns the most similar TS or KS chapter(s) to each MS chapter.
- The top-3 most similar TS or KS chapters for each MS chapter are extracted by the four models.
- Question: Are the most similar chapter(s) extracted by the human evaluator contained in the top three chapters extracted by a given model?

16

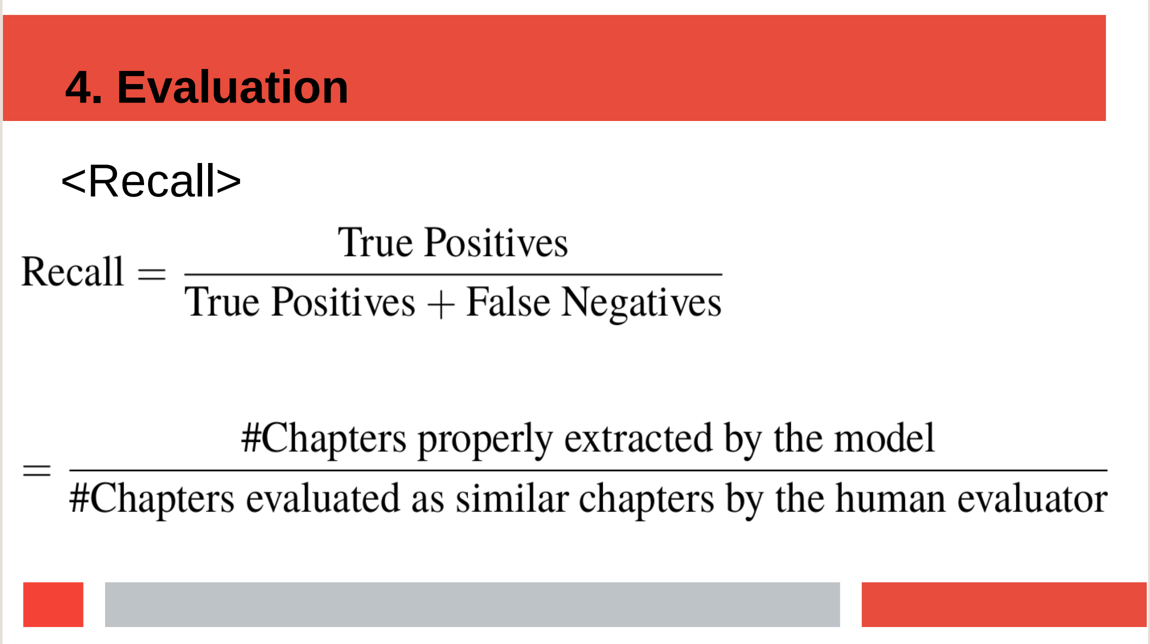


4. Evaluation

<Recall>

- Chapters correctly extracted by model = True Positive
- Chapters evaluated as similar by human evaluator but not extracted by model = False Negative

17



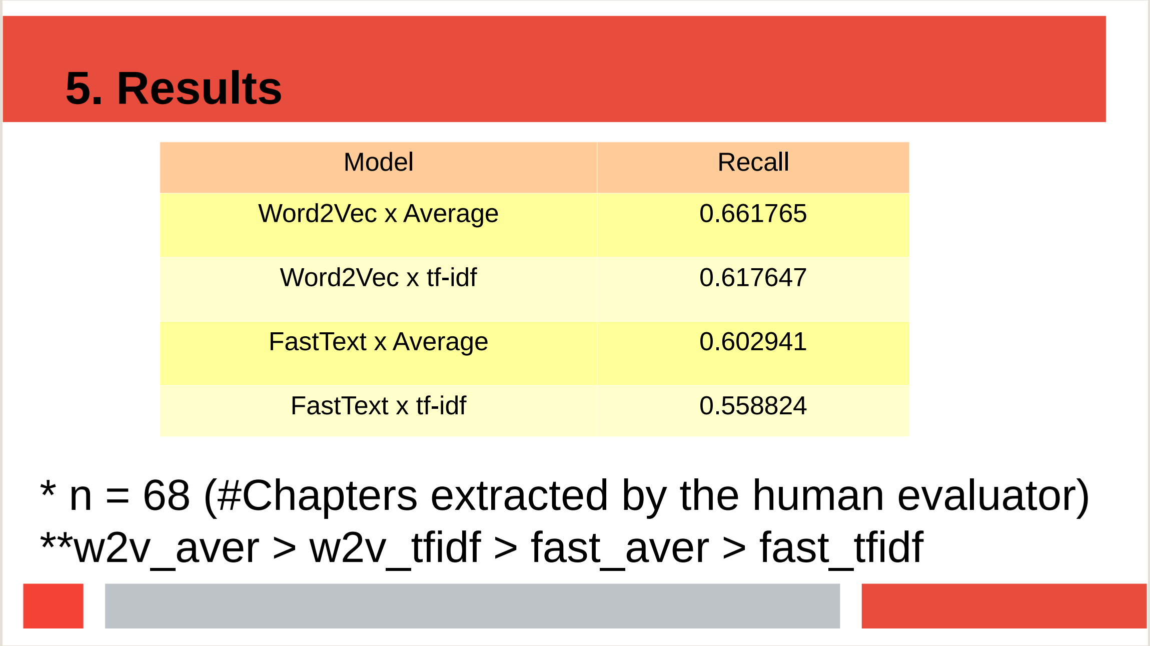
4. Evaluation

<Recall>

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$= \frac{\text{\#Chapters properly extracted by the model}}{\text{\#Chapters evaluated as similar chapters by the human evaluator}}$$

18

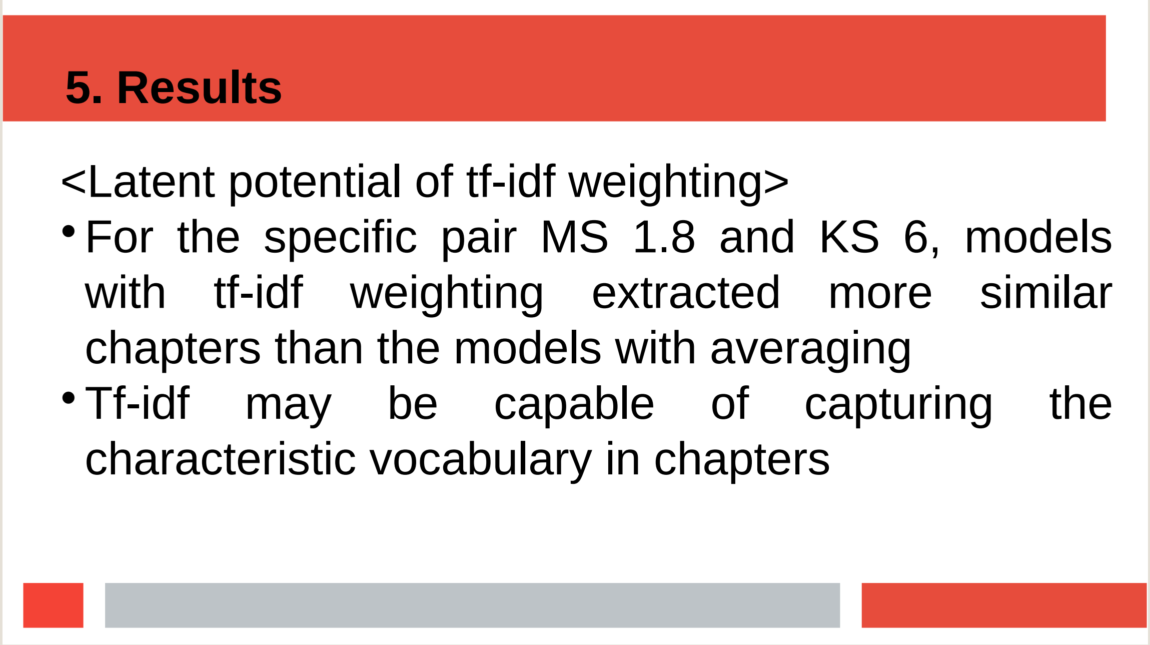


5. Results

Model	Recall
Word2Vec x Average	0.661765
Word2Vec x tf-idf	0.617647
FastText x Average	0.602941
FastText x tf-idf	0.558824

* n = 68 (#Chapters extracted by the human evaluator)
**w2v_aver > w2v_tfidf > fast_aver > fast_tfidf

19

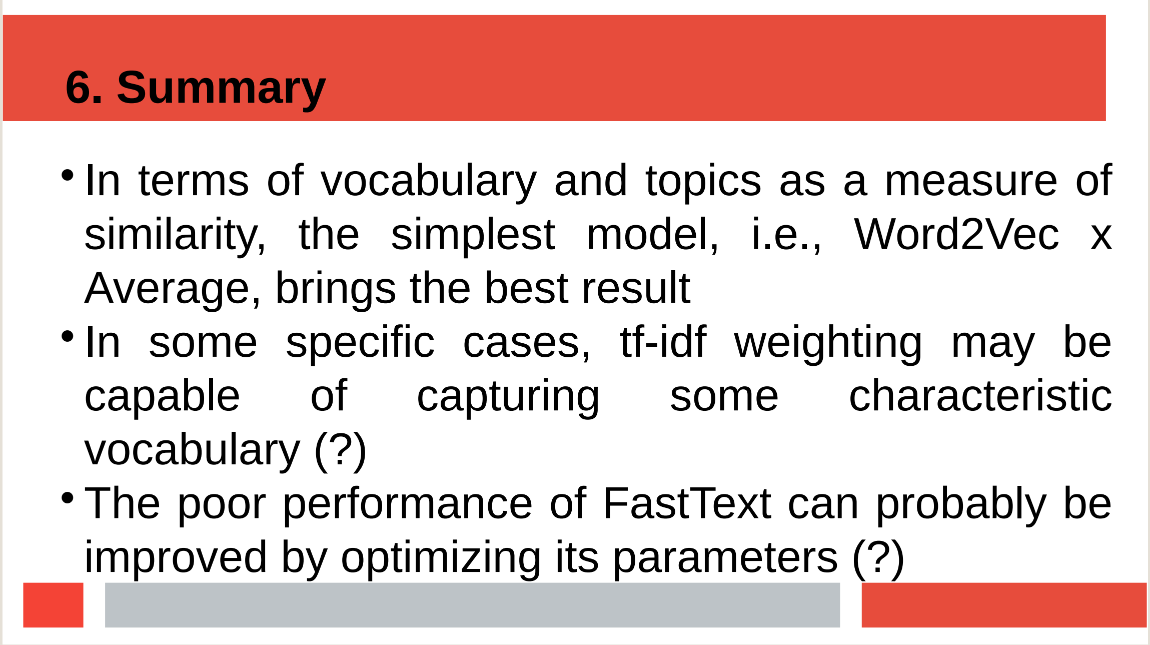


5. Results

<Latent potential of tf-idf weighting>

- For the specific pair MS 1.8 and KS 6, models with tf-idf weighting extracted more similar chapters than the models with averaging
- Tf-idf may be capable of capturing the characteristic vocabulary in chapters

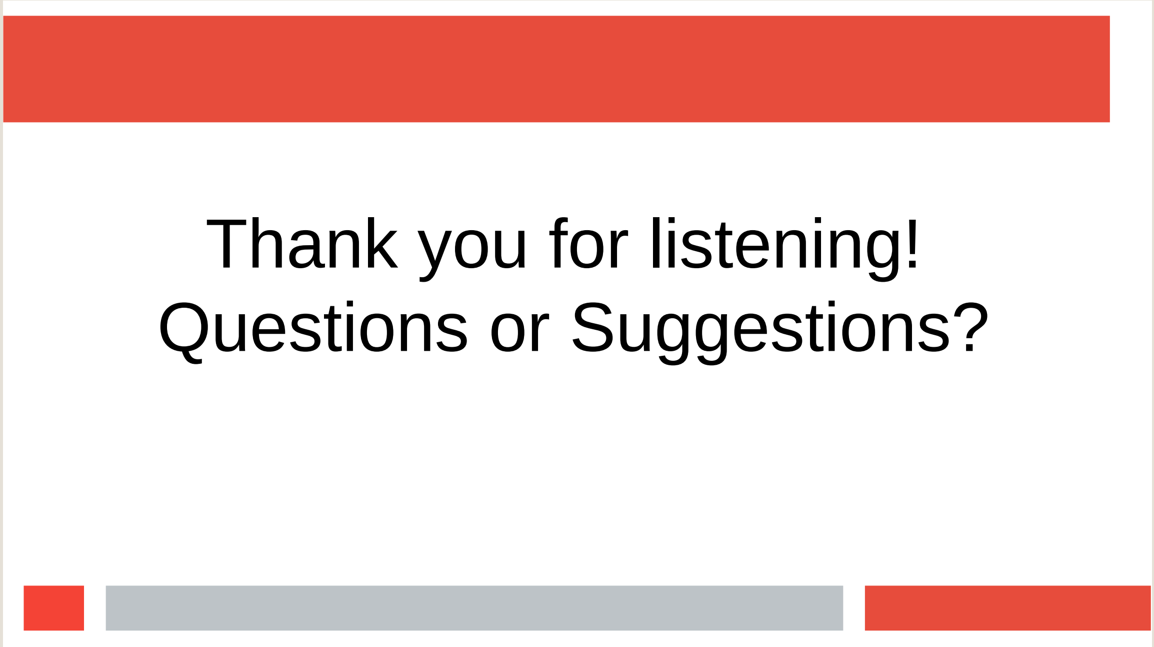
20



6. Summary

- In terms of vocabulary and topics as a measure of similarity, the simplest model, i.e., Word2Vec x Average, brings the best result
- In some specific cases, tf-idf weighting may be capable of capturing some characteristic vocabulary (?)
- The poor performance of FastText can probably be improved by optimizing its parameters (?)

21



Thank you for listening!
Questions or Suggestions?

Presentation 4 | 報告 4

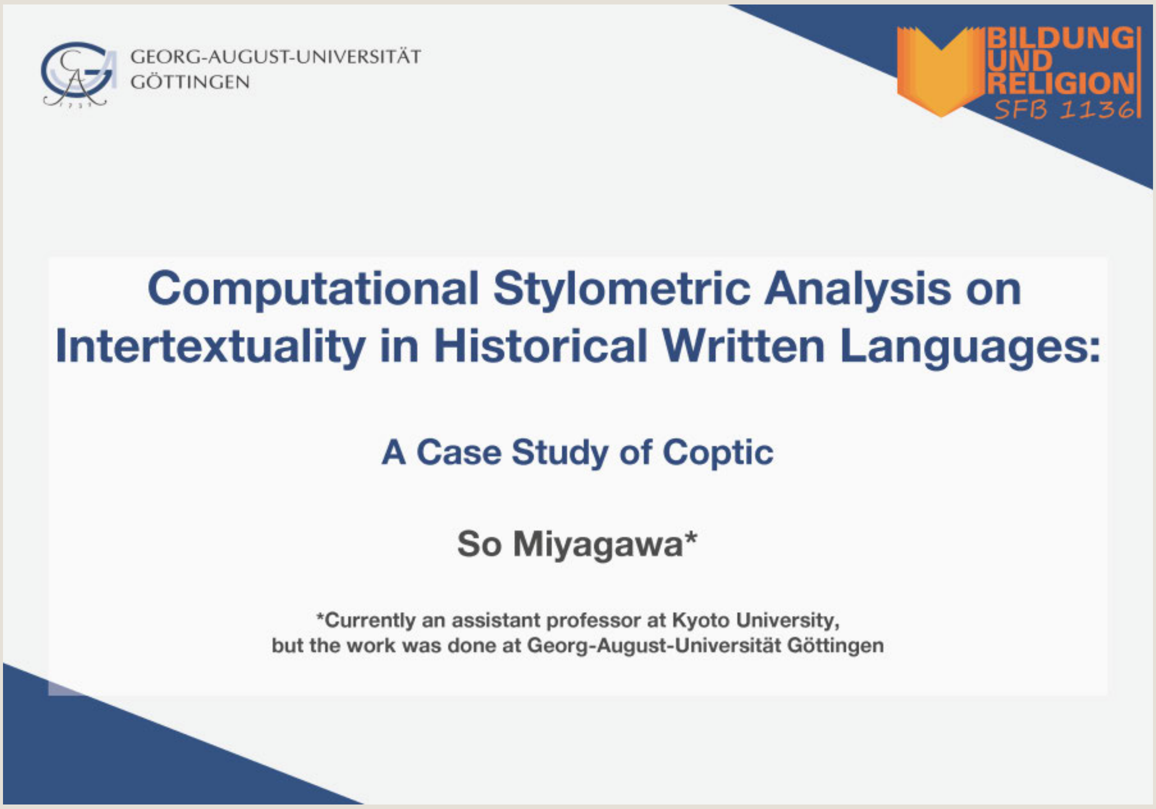
Computational Stylometric Analysis on Intertextuality in Historical Written Languages: A Case Study of Coptic

文献言語における間テクスト性の計算言語学的・計量文献学的分析：コプト語における事例研究

So Miyagawa (Kyoto University, Graduate School of Letters / Center for Cultural Heritage Studies and Inter Humanities)

宮川創（京都大学 文学研究科 / 文学研究科附属文化遺産学・人文知連携センター）

1



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

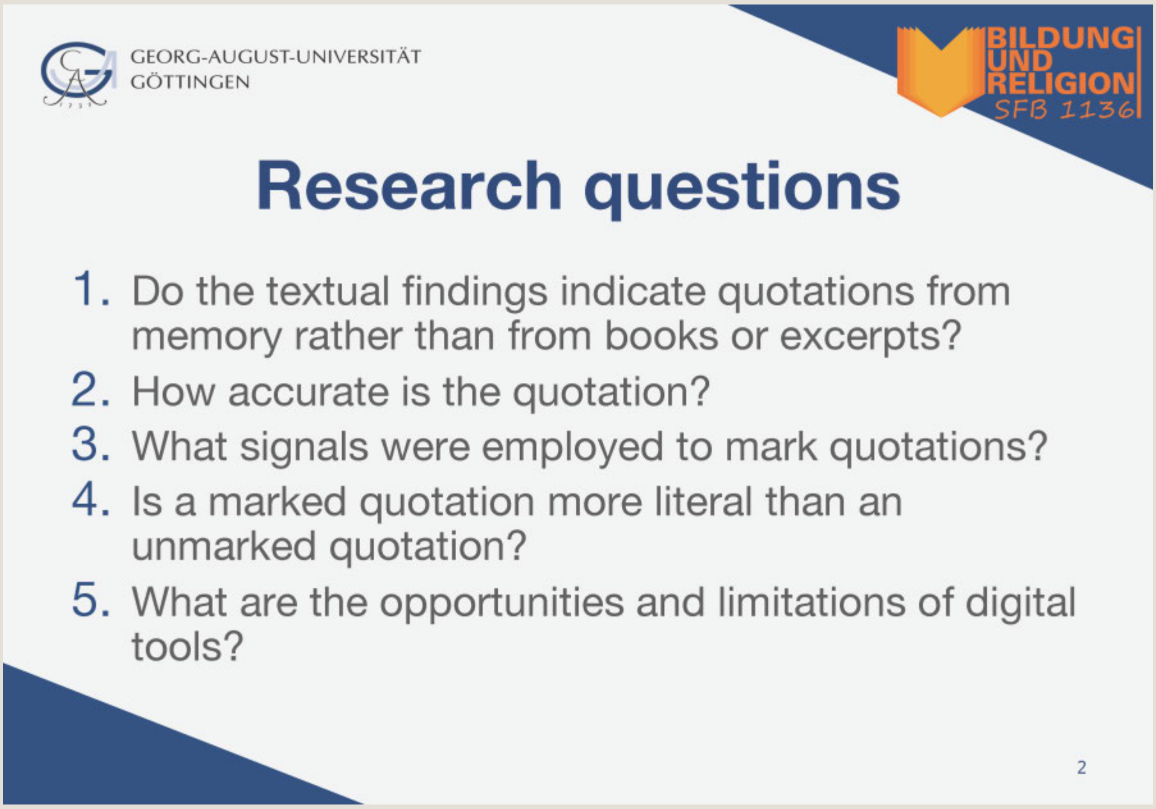
Computational Stylometric Analysis on Intertextuality in Historical Written Languages:

A Case Study of Coptic

So Miyagawa*

*Currently an assistant professor at Kyoto University, but the work was done at Georg-August-Universität Göttingen

2



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

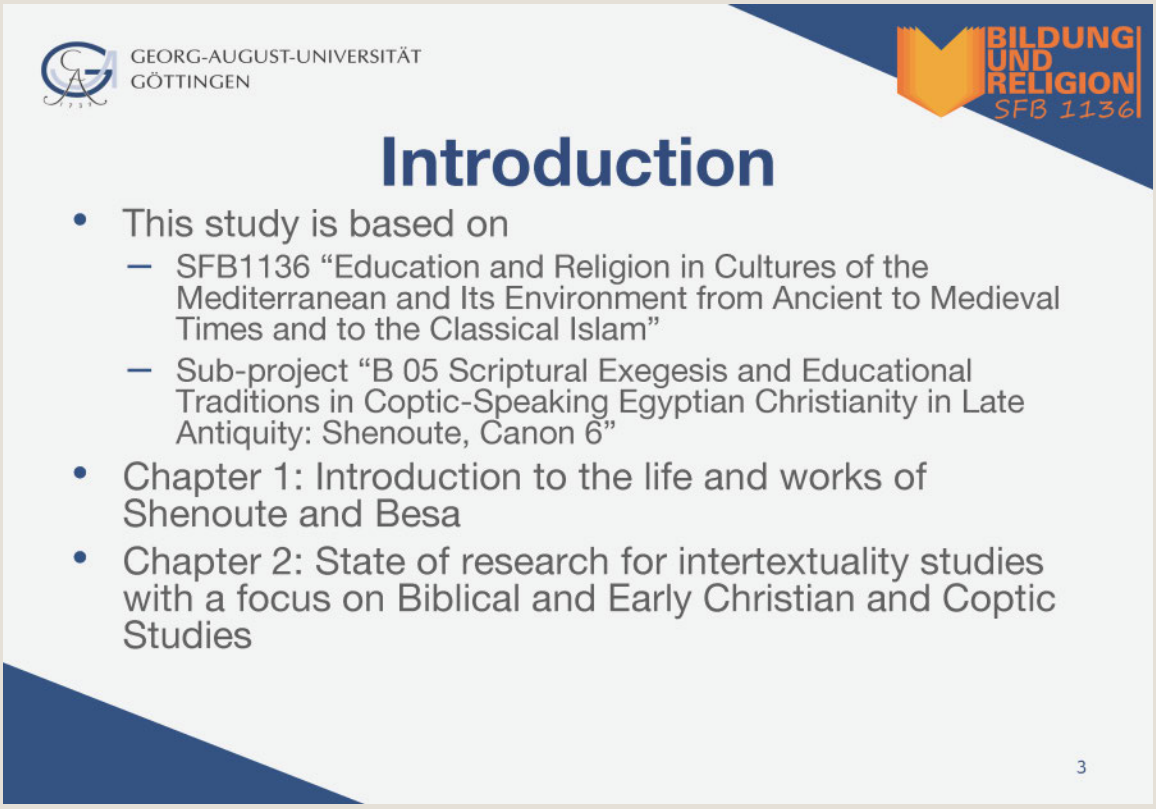
BILDUNG
UND
RELIGION
SFB 1136

Research questions

1. Do the textual findings indicate quotations from memory rather than from books or excerpts?
2. How accurate is the quotation?
3. What signals were employed to mark quotations?
4. Is a marked quotation more literal than an unmarked quotation?
5. What are the opportunities and limitations of digital tools?

2

3



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

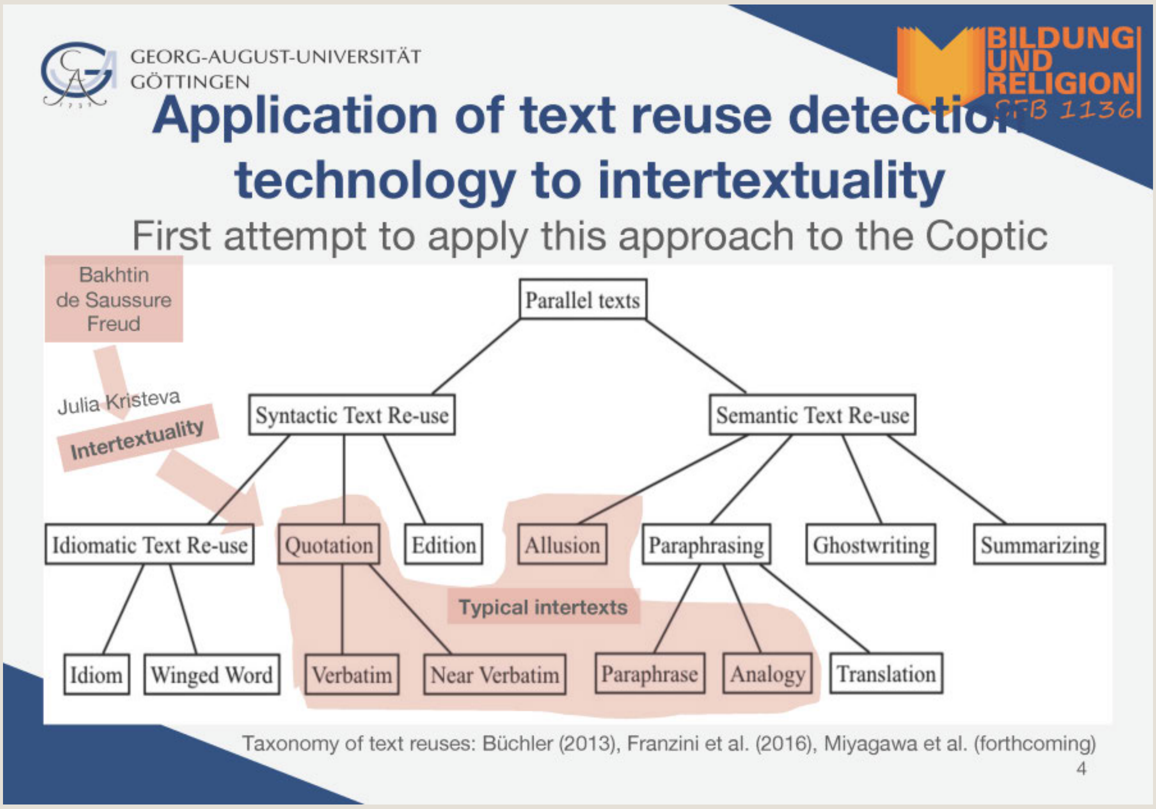
BILDUNG
UND
RELIGION
SFB 1136

Introduction

- This study is based on
  - SFB1136 “Education and Religion in Cultures of the Mediterranean and Its Environment from Ancient to Medieval Times and to the Classical Islam”
  - Sub-project “B 05 Scriptural Exegesis and Educational Traditions in Coptic-Speaking Egyptian Christianity in Late Antiquity: Shenoute, Canon 6”
- Chapter 1: Introduction to the life and works of Shenoute and Besa
- Chapter 2: State of research for intertextuality studies with a focus on Biblical and Early Christian and Coptic Studies

3

4



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

Application of text reuse detection technology to intertextuality

First attempt to apply this approach to the Coptic

Bakhtin
de Saussure
Freud

Julia Kristeva

Intertextuality

Parallel texts

Syntactic Text Re-use

Semantic Text Re-use

Idiomatic Text Re-use

Quotation

Edition

Allusion

Paraphrasing

Ghostwriting

Summarizing

Typical intertexts

Idiom

Winged Word

Verbatim

Near Verbatim

Paraphrase

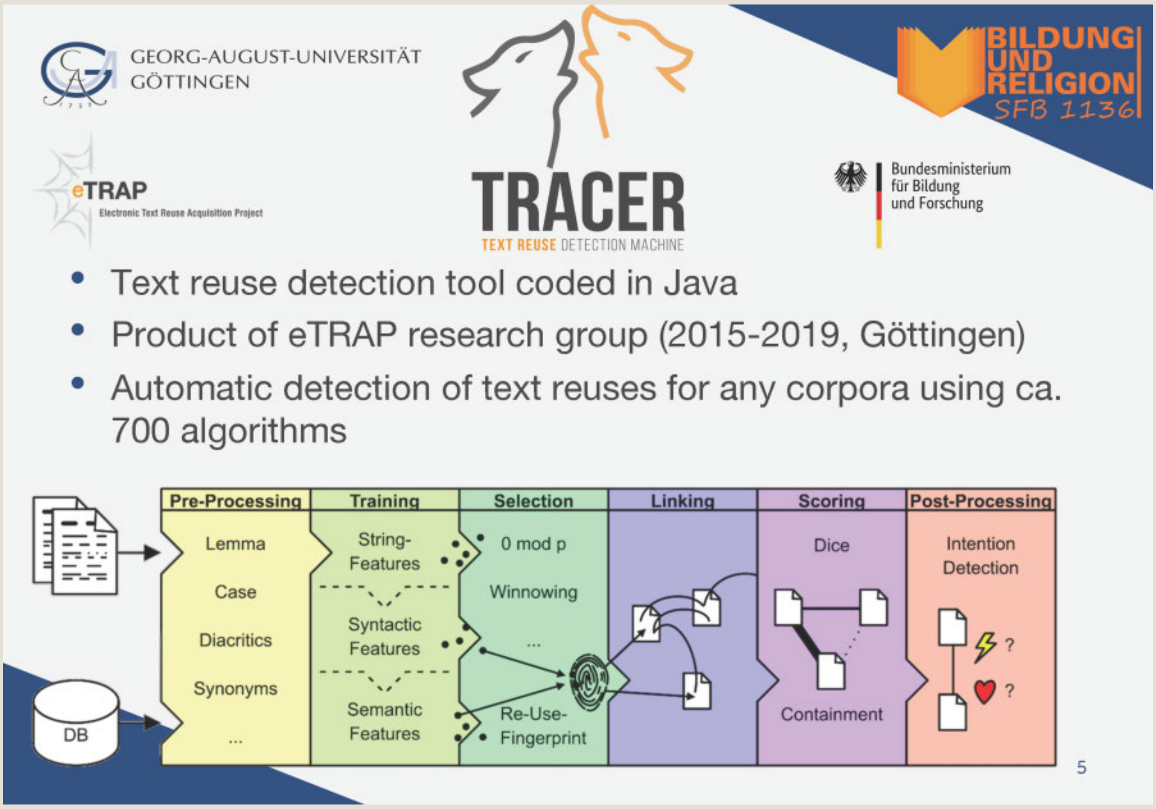
Analogy

Translation

Taxonomy of text reuses: Büchler (2013), Franzini et al. (2016), Miyagawa et al. (forthcoming)

4

5



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

eTRAP Electronic Text Reuse Acquisition Project

TRACER
TEXT REUSE DETECTION MACHINE

Bundesministerium für Bildung und Forschung

- Text reuse detection tool coded in Java
- Product of eTRAP research group (2015-2019, Göttingen)
- Automatic detection of text reuses for any corpora using ca. 700 algorithms

5

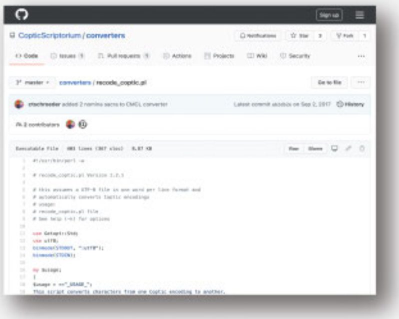
6

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

BILDUNG
UND
RELIGION
SFB 1136

Tools for pre-processing 1

Unicode converter



CopticScriptorium / converters

converters / recode_coptic.pl

ctschroeder added 2 nomina sacra to CMCL converter — Latest commit ab0d5cb on Sep 2, 2017

2 contributors

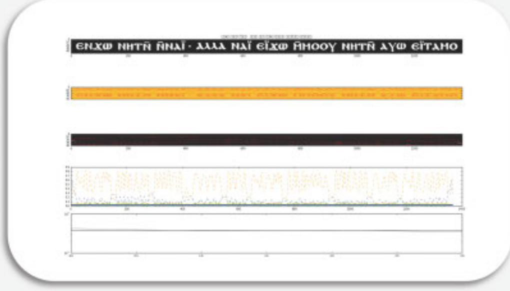
Executable File | 483 lines (367 sloc) | 8.87 KB

```
#!/usr/bin/perl -w

# recode_coptic.pl Version 1.2.1

# this assumes a UTF-8 file in one word per line format and
# automatically converts Coptic encodings
# usage:
# recode_coptic.pl file
# See help (-h) for options

use Getopt::Std;
use utf8;
binmode(STDOUT, ":utf8");
binmode(STDIN);

my $usage;
{
$usage = <<"_USAGE_";
This script converts characters from one Coptic encoding to another.
```

Zeldes, Schroeder and Miyagawa (2013-2016)

Old transcription into Unicode

Optical Character Recognition



Lincke et al. (2019) and Miyagawa et al. (2019)

Automatic transcription

6

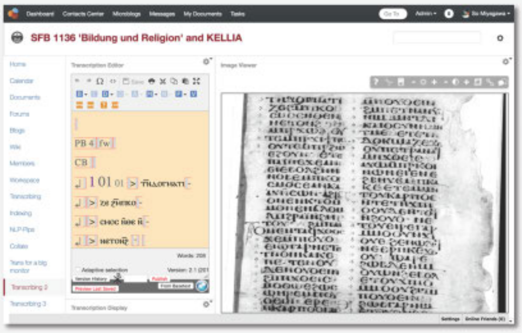
7

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

BILDUNG
UND
RELIGION
SFB 1136

Tools for pre-processing 2

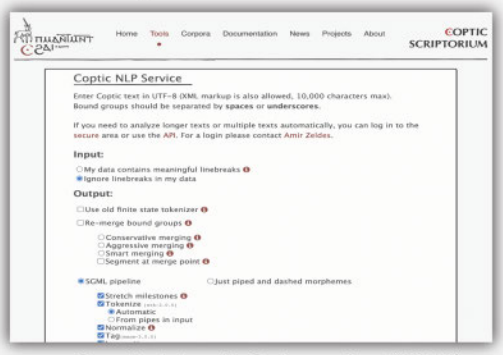
Virtual Manuscript Room



See Behlmer & Feder (2017)

Digital transcription

Coptic NLP Service



Coptic NLP Service

Enter Coptic text in UTF-8 (XML markup is also allowed, 10,000 characters max).
Bound groups should be separated by **spaces** or **underscores**.

If you need to analyze longer texts or multiple texts automatically, you can log in to the secure area or use the API. For a login please contact Amir Zeldes.

Input:

○ My data contains meaningful linebreaks
● Ignore linebreaks in my data

Output:

□ Use old finite state tokenizer
□ Re-merge bound groups
 ○ Conservative merging
 ○ Aggressive merging
 ○ Smart merging
 □ Segment at merge point

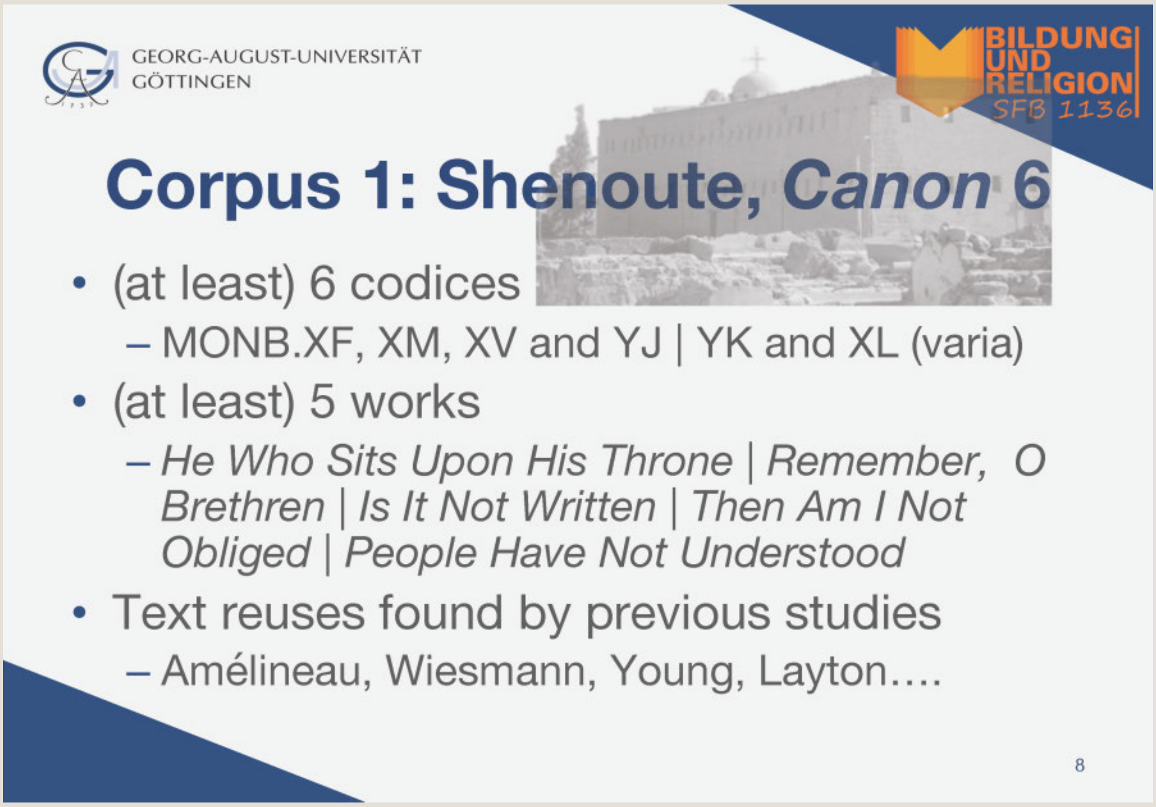
● SGML pipeline ○ Just piped and dashed morphemes
 ☑ Stretch milestones
 ☑ Tokenize
  ● Automatic
  ○ From pipes in input
 ☑ Normalize
 ☑ Tag

See Zeldes & Schroeder (2016)

Linguistic tagging

7

8



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

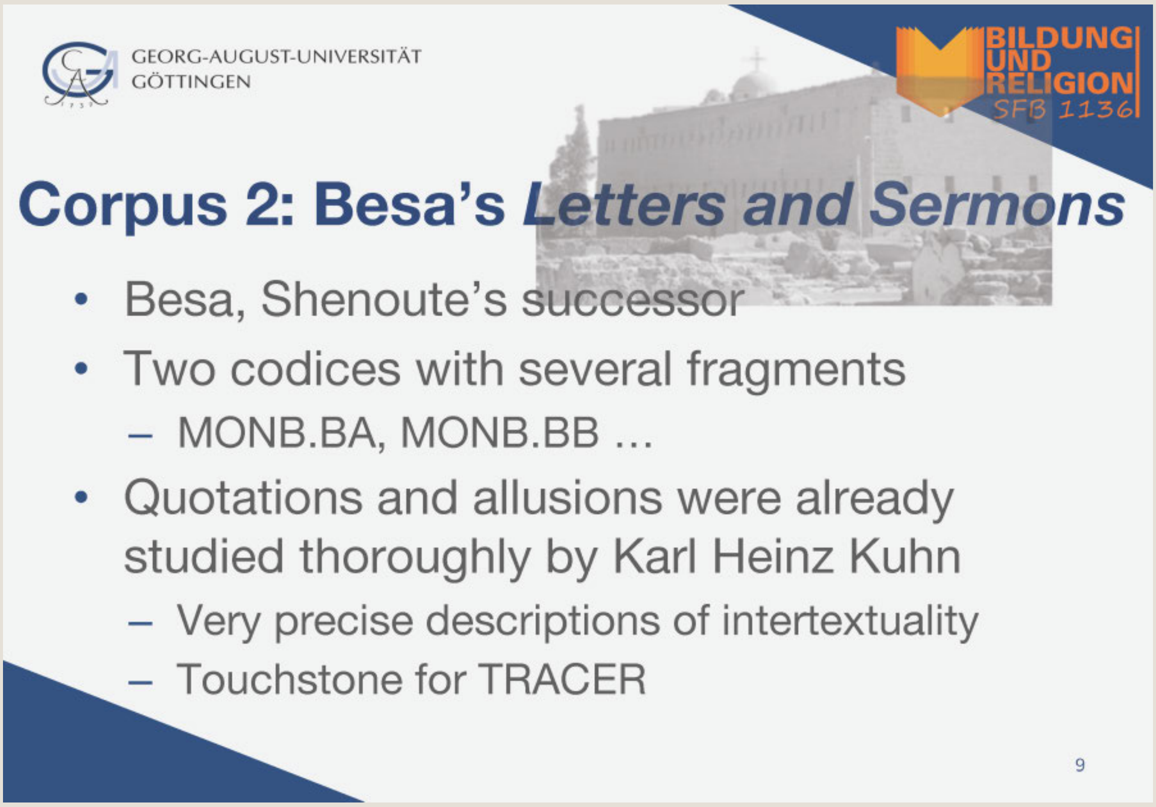
BILDUNG UND RELIGION SFB 1136

Corpus 1: Shenoute, *Canon* 6

- (at least) 6 codices
  - MONB.XF, XM, XV and YJ | YK and XL (varia)
- (at least) 5 works
  - *He Who Sits Upon His Throne* | *Remember, O Brethren* | *Is It Not Written* | *Then Am I Not Obliged* | *People Have Not Understood*
- Text reuses found by previous studies
  - Amélineau, Wiesmann, Young, Layton….

8

9



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

Corpus 2: Besa's *Letters and Sermons*

- Besa, Shenoute's successor
- Two codices with several fragments
  - MONB.BA, MONB.BB …
- Quotations and allusions were already studied thoroughly by Karl Heinz Kuhn
  - Very precise descriptions of intertextuality
  - Touchstone for TRACER

9

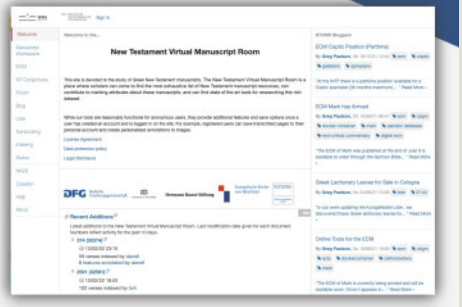
10

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

BILDUNG
UND
RELIGION
SFB 1136

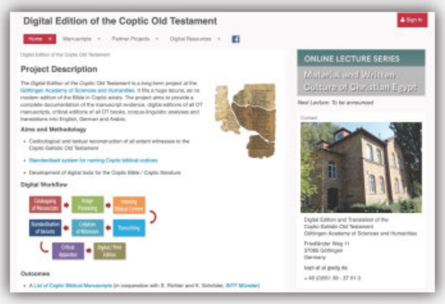
Comparison corpus

- Sahidic Bible 2.0
- Digital “base text” of Sahidic NT & OT
  - Product of INTF Münster and CoptOT Göttingen



New Testament Virtual Manuscript Room

Recent Additions



Digital Edition of the Coptic Old Testament

Project Description

Aims and Methodology

Digital Workflow

Outcomes

ONLINE LECTURE SERIES

Material and Written Culture of Christian Egypt

11

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

BILDUNG
UND
RELIGION
SFB 1136

First attempt (2016)

Source Text	Target Text	TR Cand.
Sahidic Bible	Besa	13,835
Sahidic Bible	Canon 6	8414

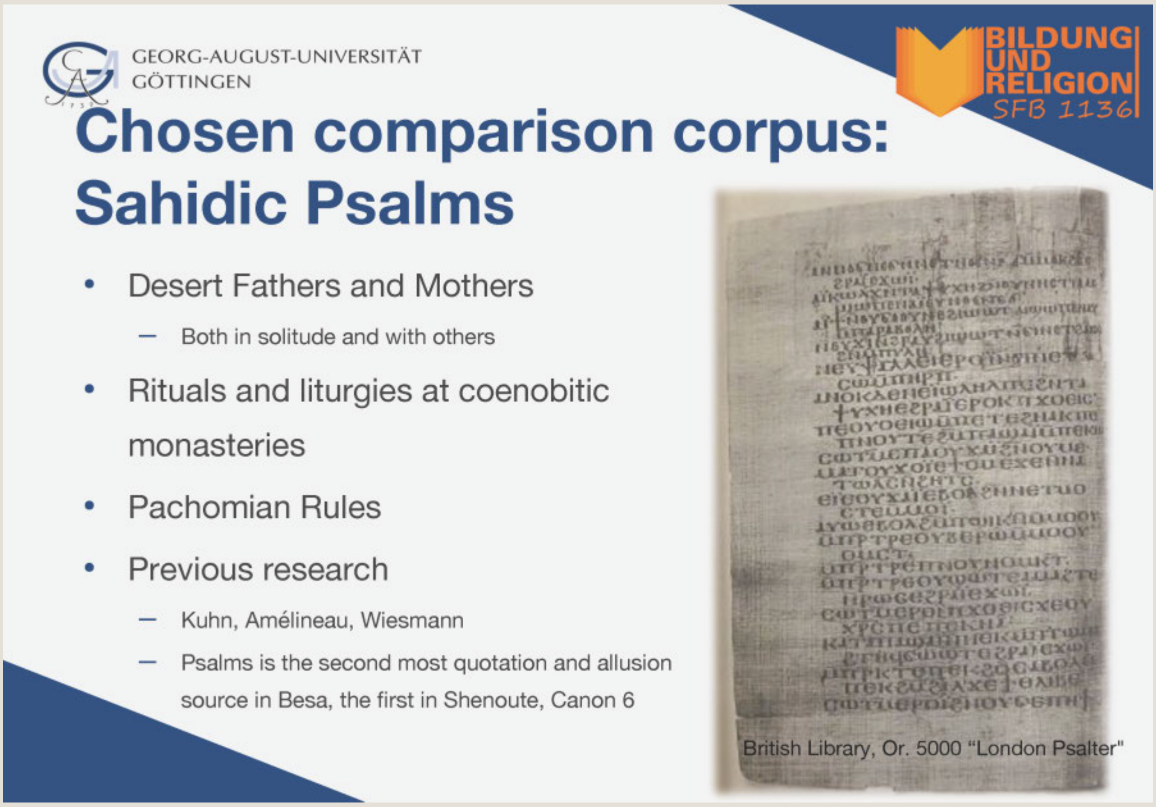
↓

Too many text reuse candidates!

↓

I chose the Sahidic Psalms

12



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

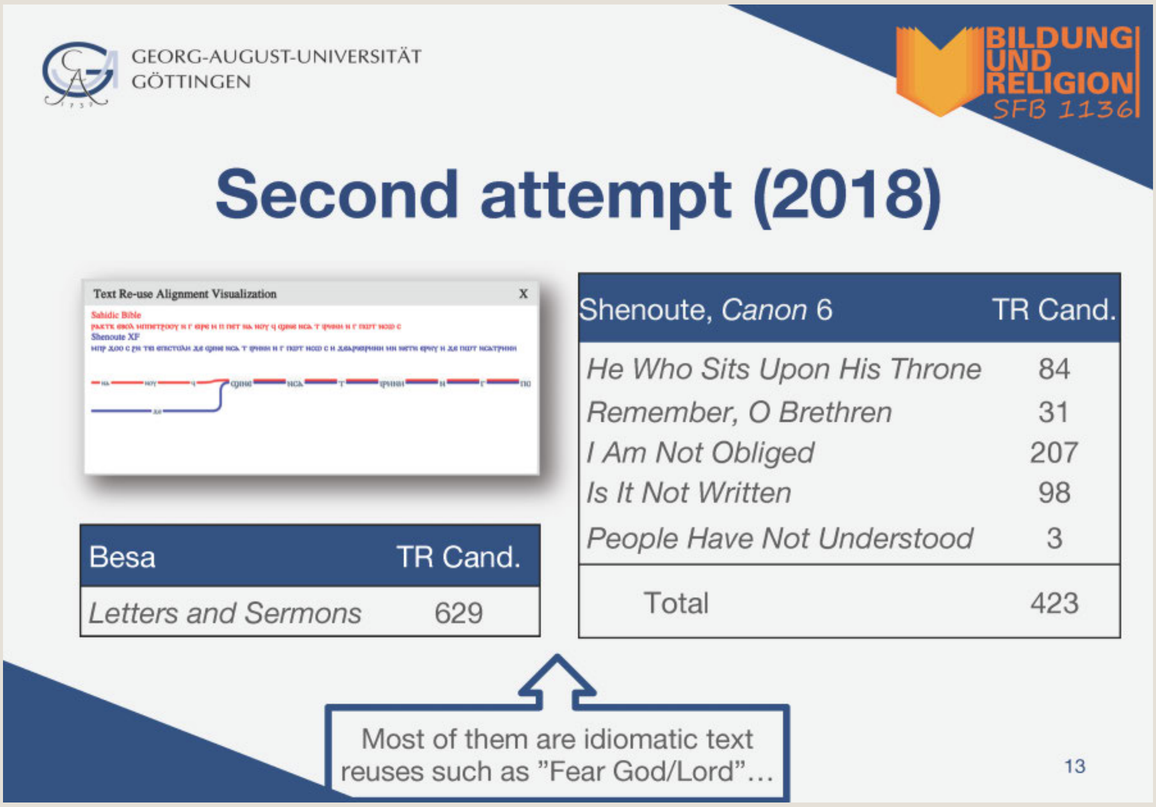
BILDUNG UND RELIGION SFB 1136

Chosen comparison corpus: Sahidic Psalms

- Desert Fathers and Mothers
  - Both in solitude and with others
- Rituals and liturgies at coenobitic monasteries
- Pachomian Rules
- Previous research
  - Kuhn, Amélineau, Wiesmann
  - Psalms is the second most quotation and allusion source in Besa, the first in Shenoute, Canon 6

British Library, Or. 5000 "London Psalter"

13



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

Second attempt (2018)

Text Re-use Alignment Visualization X

Sahidic Bible

ⲣⲁⲕⲧⲕ ⲉⲃⲟⲗ ⲙⲡⲡⲉⲧϩⲟⲟⲩ ⲛ ⲅ ⲉⲓⲣⲉ ⲙ ⲡ ⲡⲉⲧ ⲛⲁ ⲛⲟⲩ ϥ ϣⲓⲛⲉ ⲛⲥⲁ ⲧ ⲉⲓⲣⲏⲛⲏ ⲛ ⲅ ⲡⲱⲧ ⲛⲥⲱ ⲥ

Shenoute XF

ⲙⲡⲣ ϫⲟⲟ ⲥ ϩⲛ ⲧⲉⲓ ⲉⲡⲓⲥⲧⲟⲗⲏ ϫⲉ ϣⲓⲛⲉ ⲛⲥⲁ ⲧ ⲉⲓⲣⲏⲛⲏ ⲛ ⲅ ⲡⲱⲧ ⲛⲥⲱ ⲥ ⲛ [illegible] ϫⲉ ⲡⲱⲧ [illegible]

ⲛⲁ
ⲛⲟⲩ
ϥ
ϫⲉ
ϣⲓⲛⲉ
ⲛⲥⲁ
ⲧ
ⲉⲓⲣⲏⲛⲏ
ⲛ
ⲅ
ⲡⲱ

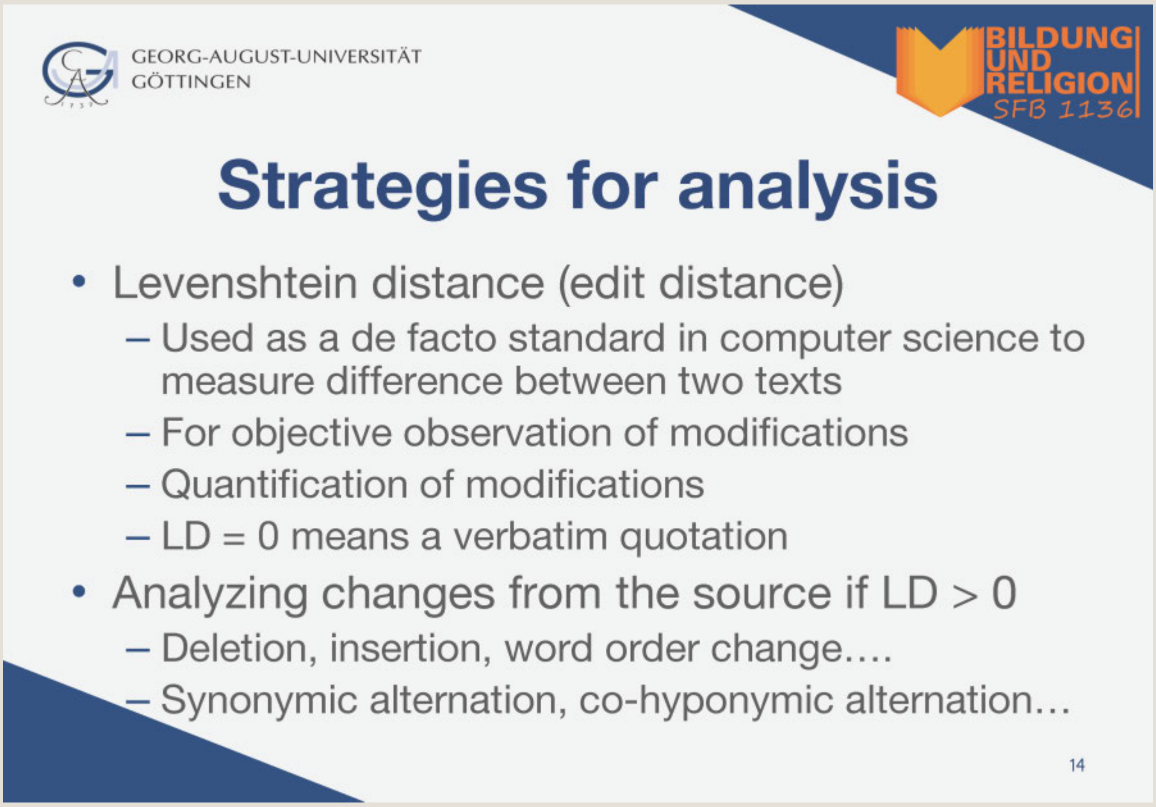
Besa	TR Cand.
Letters and Sermons	629

Shenoute, *Canon* 6	TR Cand.
He Who Sits Upon His Throne	84
Remember, O Brethren	31
I Am Not Obliged	207
Is It Not Written	98
People Have Not Understood	3
Total	423

Most of them are idiomatic text reuses such as "Fear God/Lord"…

13

14



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

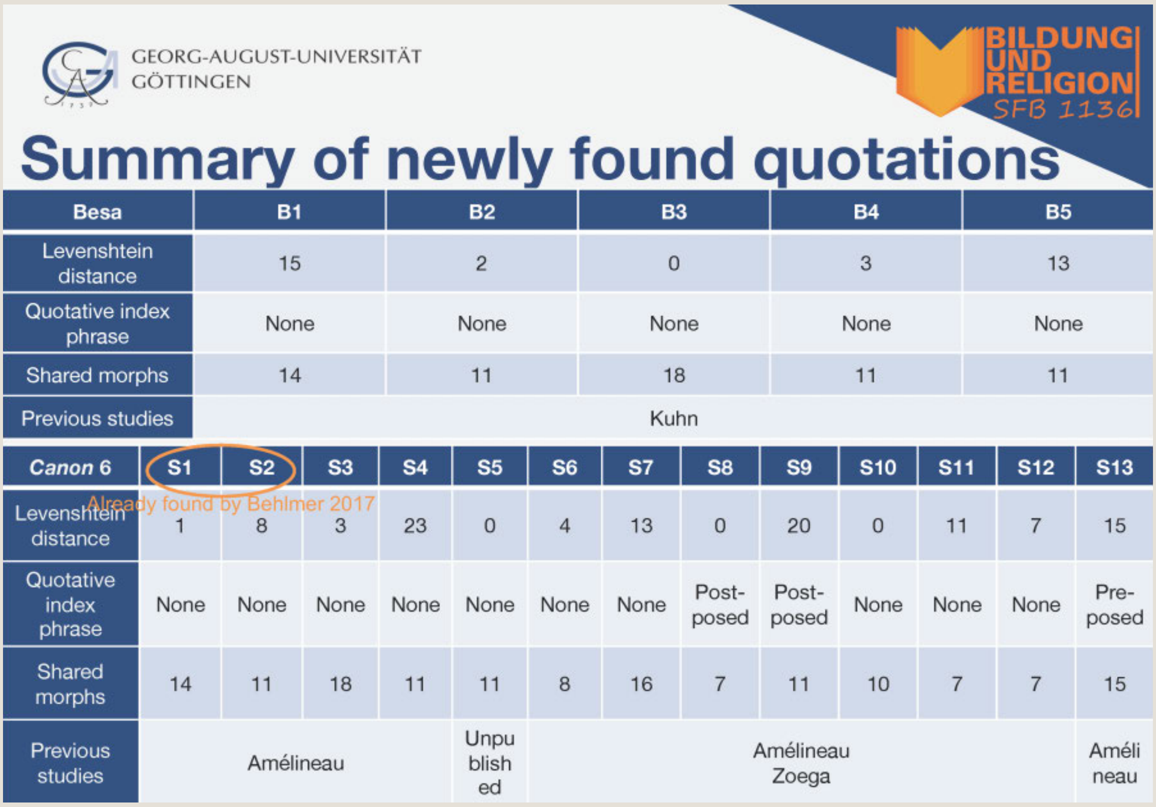
BILDUNG UND RELIGION SFB 1136

Strategies for analysis

- Levenshtein distance (edit distance)
  - Used as a de facto standard in computer science to measure difference between two texts
  - For objective observation of modifications
  - Quantification of modifications
  - LD = 0 means a verbatim quotation
- Analyzing changes from the source if LD > 0
  - Deletion, insertion, word order change….
  - Synonymic alternation, co-hyponymic alternation…

14

15



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

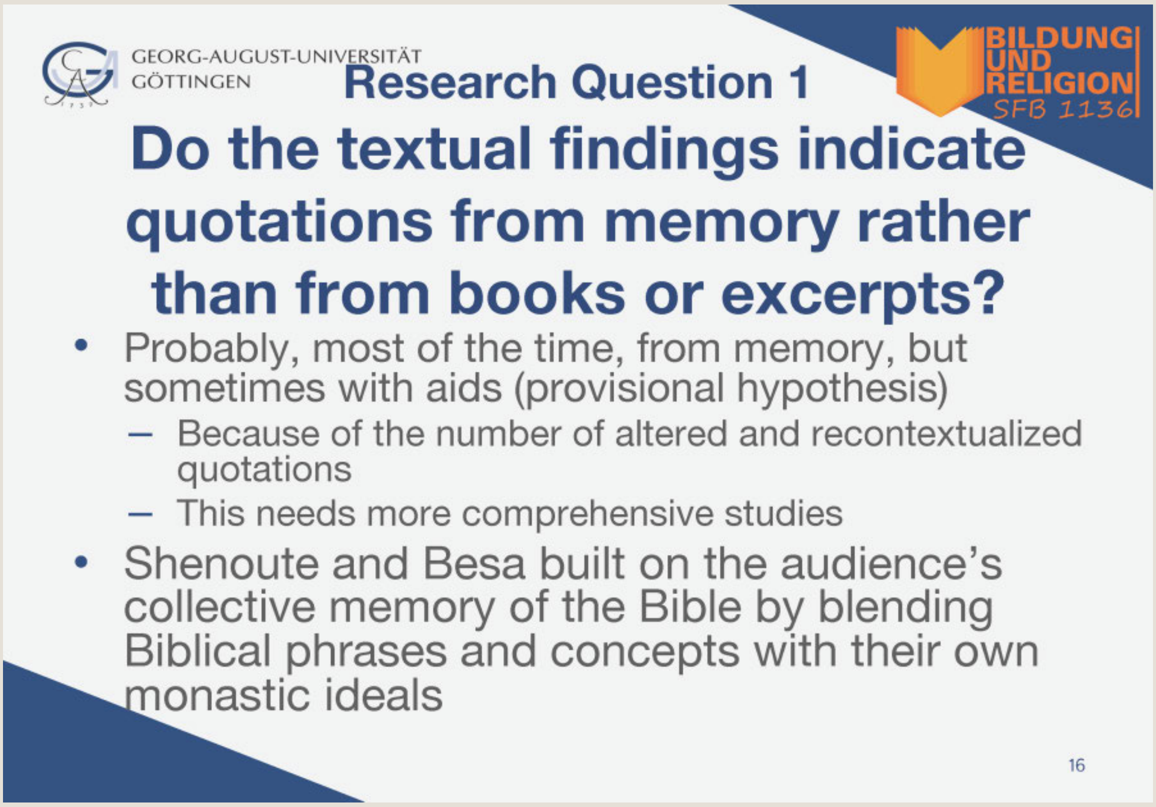
Summary of newly found quotations

Besa	B1	B2	B3	B4	B5
Levenshtein distance	15	2	0	3	13
Quotative index phrase	None	None	None	None	None
Shared morphs	14	11	18	11	11
Previous studies	Kuhn				

Already found by Behlmer 2017

Canon 6	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Levenshtein distance	1	8	3	23	0	4	13	0	20	0	11	7	15
Quotative index phrase	None	None	None	None	None	None	None	Post-posed	Post-posed	None	None	None	Pre-posed
Shared morphs	14	11	18	11	11	8	16	7	11	10	7	7	15
Previous studies	Amélineau				Unpublished	Amélineau Zoega							Améli neau

16



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

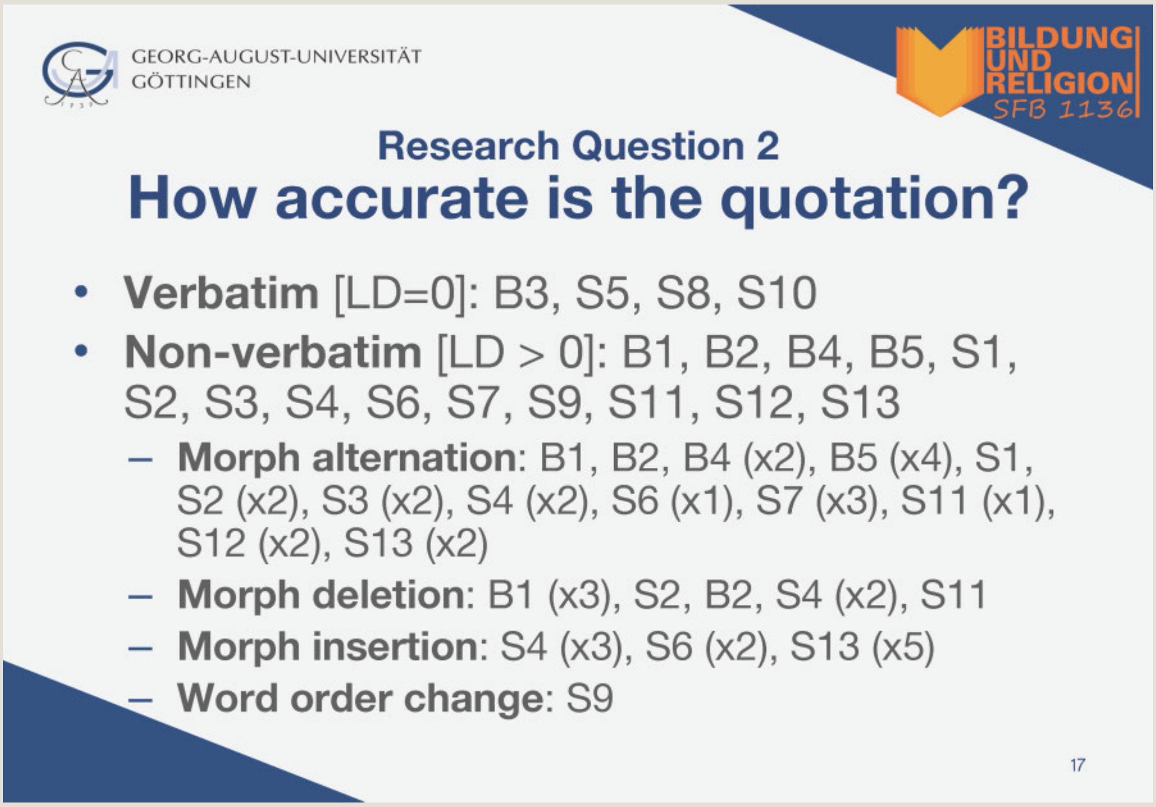
Research Question 1

Do the textual findings indicate quotations from memory rather than from books or excerpts?

- Probably, most of the time, from memory, but sometimes with aids (provisional hypothesis)
  - Because of the number of altered and recontextualized quotations
  - This needs more comprehensive studies
- Shenoute and Besa built on the audience's collective memory of the Bible by blending Biblical phrases and concepts with their own monastic ideals

16

17



GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

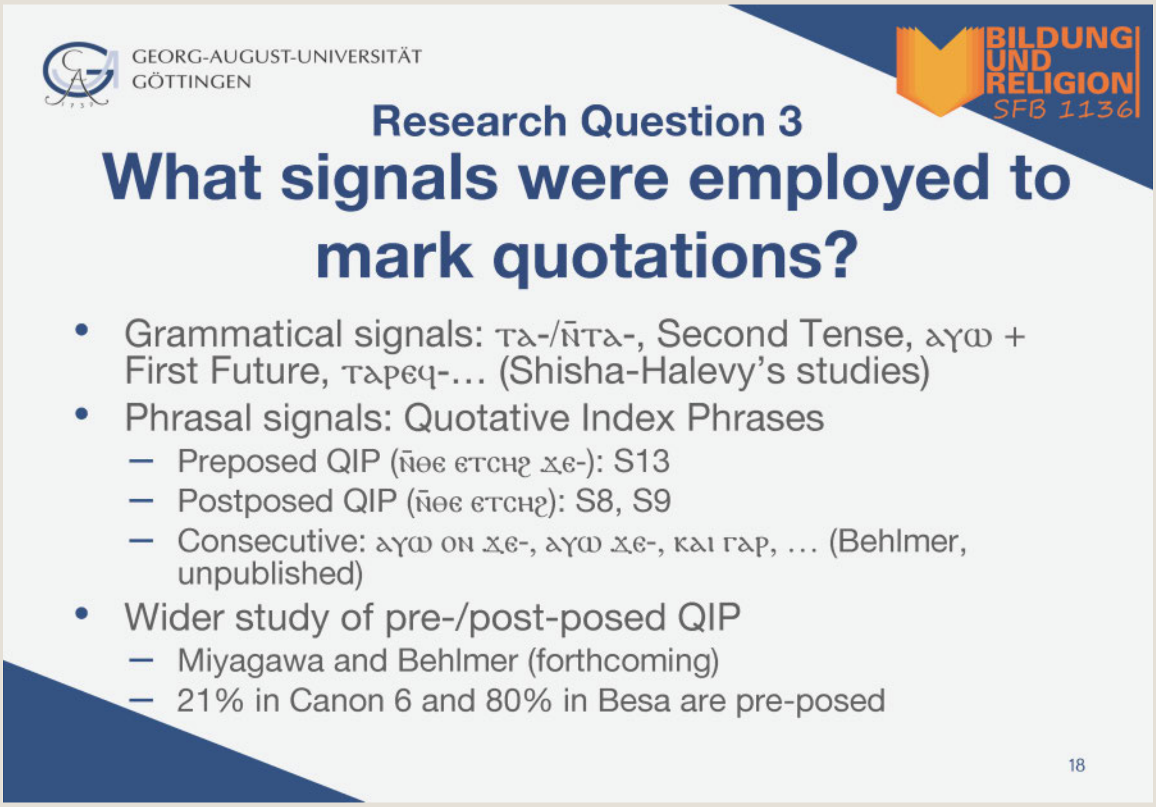
Research Question 2

How accurate is the quotation?

- **Verbatim** [LD=0]: B3, S5, S8, S10
- **Non-verbatim** [LD > 0]: B1, B2, B4, B5, S1, S2, S3, S4, S6, S7, S9, S11, S12, S13
  - **Morph alternation**: B1, B2, B4 (x2), B5 (x4), S1, S2 (x2), S3 (x2), S4 (x2), S6 (x1), S7 (x3), S11 (x1), S12 (x2), S13 (x2)
  - **Morph deletion**: B1 (x3), S2, B2, S4 (x2), S11
  - **Morph insertion**: S4 (x3), S6 (x2), S13 (x5)
  - **Word order change**: S9

17

18



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

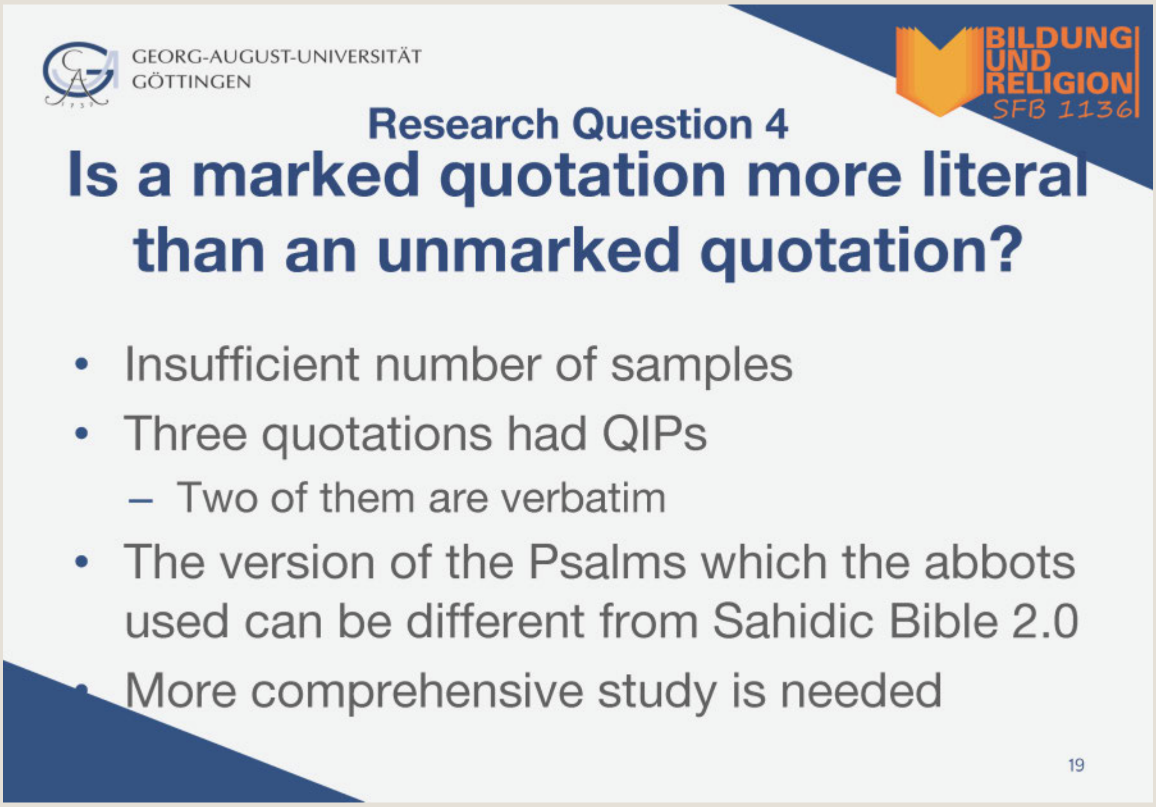
BILDUNG
UND
RELIGION
SFB 1136

Research Question 3
What signals were employed to mark quotations?

- Grammatical signals: ⲧⲁ-/ⲛ̄ⲧⲁ-, Second Tense, ⲁⲩⲱ + First Future, ⲧⲁⲣⲉϥ-… (Shisha-Halevy's studies)
- Phrasal signals: Quotative Index Phrases
  - Preposed QIP (ⲛ̄ⲑⲉ ⲉⲧⲥⲏϩ ϫⲉ-): S13
  - Postposed QIP (ⲛ̄ⲑⲉ ⲉⲧⲥⲏϩ): S8, S9
  - Consecutive: ⲁⲩⲱ ⲟⲛ ϫⲉ-, ⲁⲩⲱ ϫⲉ-, ⲕⲁⲓ ⲅⲁⲣ, … (Behlmer, unpublished)
- Wider study of pre-/post-posed QIP
  - Miyagawa and Behlmer (forthcoming)
  - 21% in Canon 6 and 80% in Besa are pre-posed

18

19



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

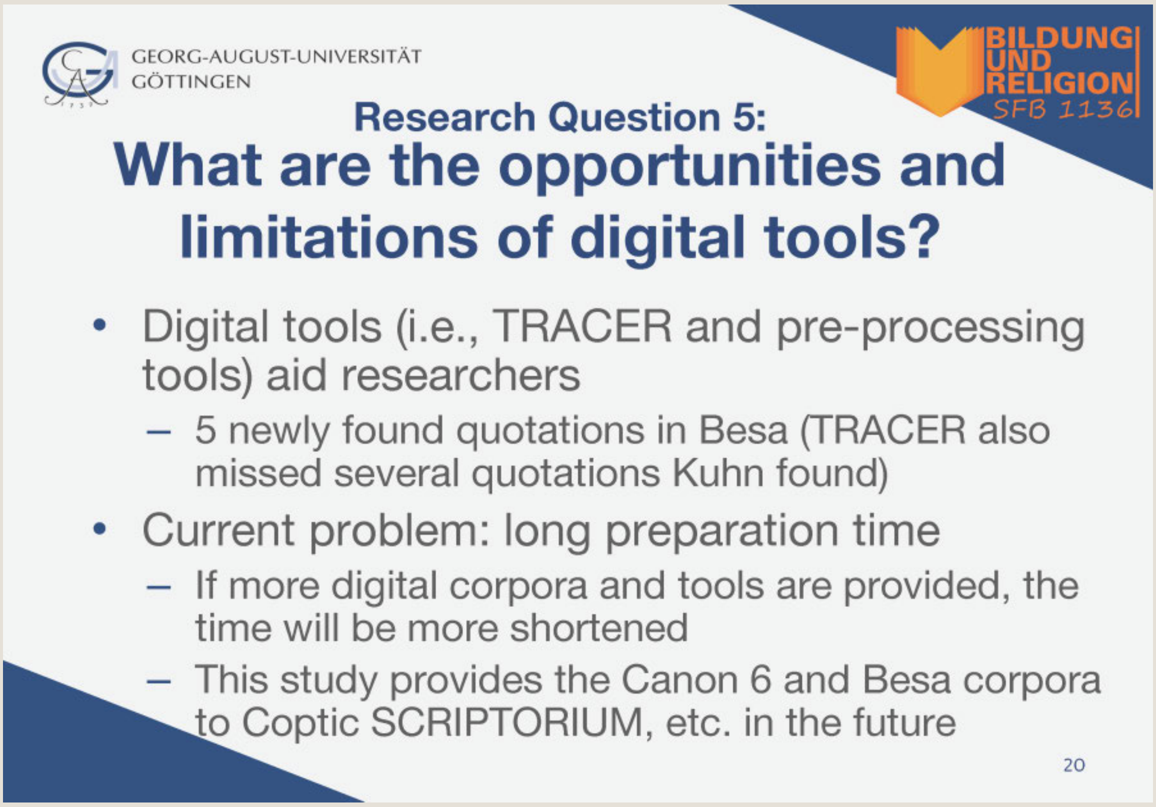
BILDUNG
UND
RELIGION
SFB 1136

Research Question 4
Is a marked quotation more literal than an unmarked quotation?

- Insufficient number of samples
- Three quotations had QIPs
  - Two of them are verbatim
- The version of the Psalms which the abbots used can be different from Sahidic Bible 2.0
- More comprehensive study is needed

19

20



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

Research Question 5: What are the opportunities and limitations of digital tools?

- Digital tools (i.e., TRACER and pre-processing tools) aid researchers
  - 5 newly found quotations in Besa (TRACER also missed several quotations Kuhn found)
- Current problem: long preparation time
  - If more digital corpora and tools are provided, the time will be more shortened
  - This study provides the Canon 6 and Besa corpora to Coptic SCRIPTORIUM, etc. in the future

20

21

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

BILDUNG UND RELIGION SFB 1136

Thank you!

So Miyagawa

- Contact: miyagawa.so.36u@kyoto-u.jp
- Website: http://somiyagawa.com/

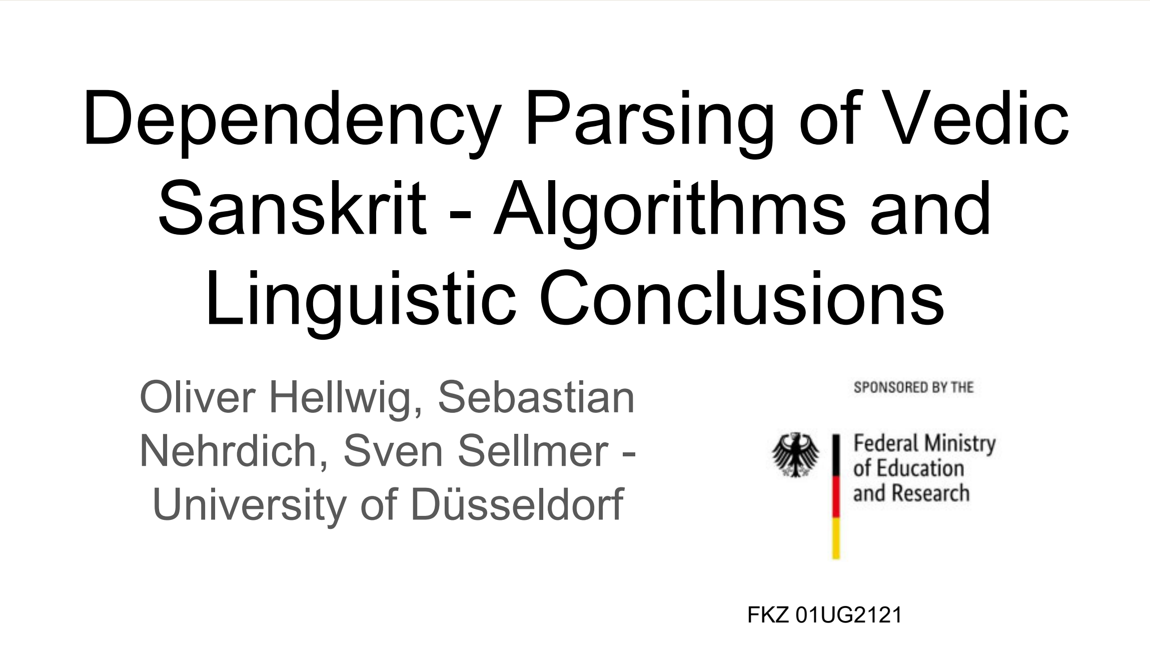
21

Presentation 5 ｜報告5

Dependency parsing of Vedic Sanskrit — Algorithms and linguistic conclusions

Oliver Hellwig (Dusseldorf University, Institute for Language and Information)
Sebastian Nehrdich (Dusseldorf University, Institute for Language and Information)
Sven Selllmer (Dusseldorf University, Institute for Language and Information)

1



Dependency Parsing of Vedic Sanskrit - Algorithms and Linguistic Conclusions

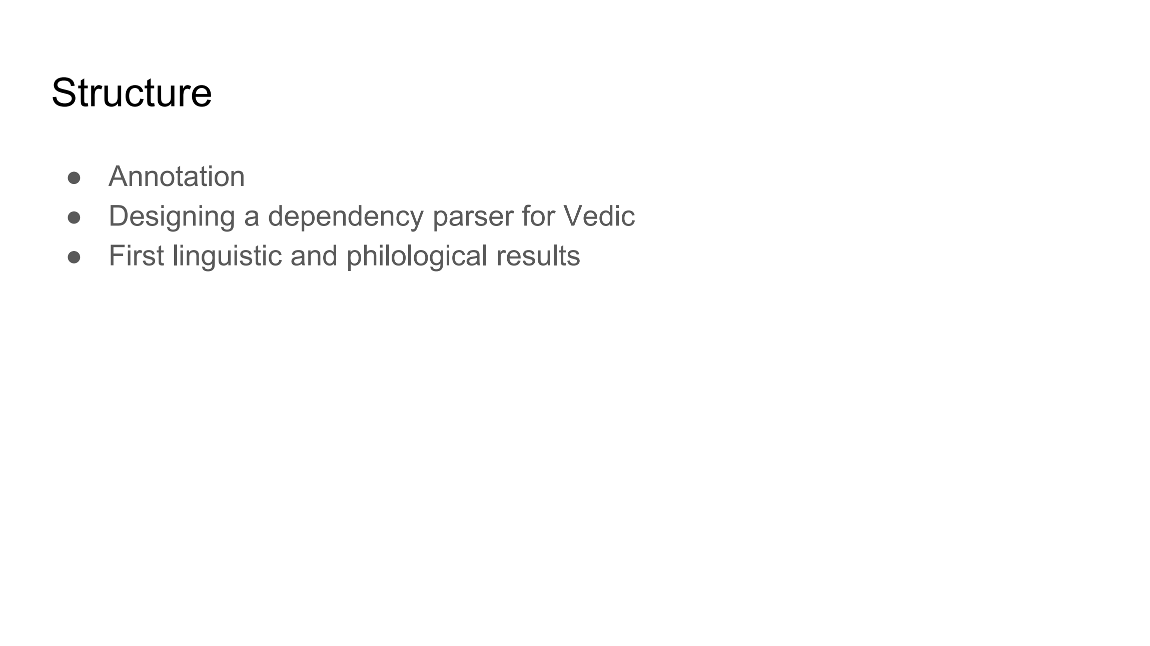
Oliver Hellwig, Sebastian Nehrdich, Sven Sellmer - University of Düsseldorf

SPONSORED BY THE

Federal Ministry of Education and Research

FKZ 01UG2121

2



Structure

- Annotation
- Designing a dependency parser for Vedic
- First linguistic and philological results

3

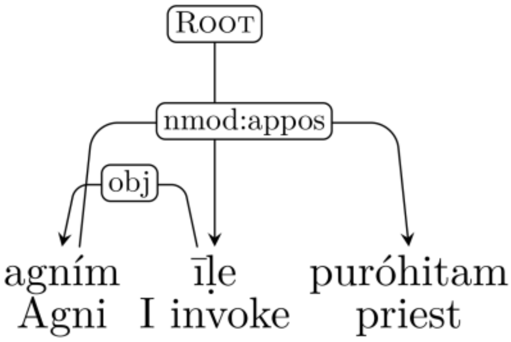
Syntactic dependencies

Aim: Labelling a sentence with syntactic roles and arcs connecting the constituents

Universal Dependencies (UD) standard

Our guidelines:

https://doi.org/10.5167/uzh-212699

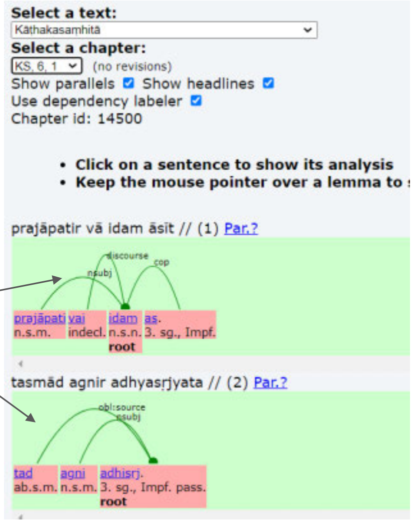


4

Annotation

Performed directly in the DCS interface

Dependency annotations


Select a text:
Kāthakasaṃhitā
Select a chapter:
KS, 6, 1 (no revisions)
Show parallels ☑ Show headlines ☑
Use dependency labeler ☑
Chapter id: 14500

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

prajāpatir vā idam āsīt // (1) Par.?

prajāpati n.s.m.
vai indecl.
idam n.s.n. root
as. 3. sg., Impf.

tasmād agnir adhyasṛjyata // (2) Par.?

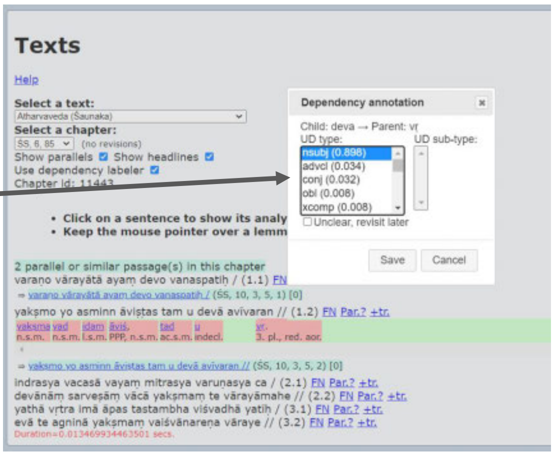
tad ab.s.m.
agni n.s.m.
adhisṛj. 3. sg., Impf. pass. root

5

Annotation

Labeler based on neural networks

See: Hellwig, O., Scarlata, S., Ackermann, E. and Widmer, P. (2020): The Treebank of Vedic Sanskrit. In: Proceedings of the LREC


Texts

Help

Select a text:
Atharvaveda (Śaunaka)
Select a chapter:
ŚS, 6, 85 (no revisions)
Show parallels ☑ Show headlines ☑
Use dependency labeler ☑
Chapter id: 11443

- Click on a sentence to show its analy
- Keep the mouse pointer over a lemm

Dependency annotation

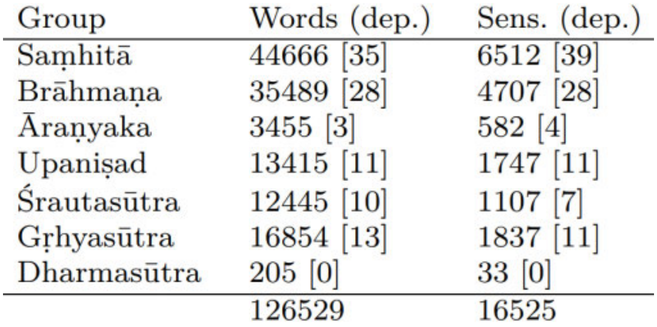
Child: deva → Parent: vṛ
UD type:
nsubj (0.898)
advcl (0.034)
conj (0.032)
obl (0.008)
xcomp (0.008)
UD sub-type:
☐ Unclear, revisit later

Save Cancel

2 parallel or similar passage(s) in this chapter
varaṇo vārayātā ayaṃ devo vanaspatiḥ / (1.1) FN
⇒ varaṇo vārayātā ayaṃ devo vanaspatiḥ / (ŚS, 10, 3, 5, 1) [0]
yakṣmo yo asminn āviṣṭas tam u devā avīvaran // (1.2) FN Par.? +tr.
yakṣma n.s.m. yad n.s.m. idam l.s.m. āviś. PPP, n.s.m. tad ac.s.m. u indecl. vṛ. 3. pl., red. aor.
⇒ yakṣmo yo asminn āviṣṭas tam u devā avīvaran // (ŚS, 10, 3, 5, 2) [0]
indrasya vacasā vayaṃ mitrasya varuṇasya ca / (2.1) FN Par.? +tr.
devānāṃ sarveṣāṃ vācā yakṣmaṃ te vārayāmahe // (2.2) FN Par.? +tr.
yathā vṛtra imā āpas tastambha viśvadhā yatiḥ / (3.1) FN Par.? +tr.
evā te agninā yakṣmaṃ vaiśvānareṇa vāraye // (3.2) FN Par.? +tr.
Duration=0.013469934463501 secs.

6

Annotation



Group	Words (dep.)	Sens. (dep.)
Saṃhitā	44666 [35]	6512 [39]
Brāhmaṇa	35489 [28]	4707 [28]
Āraṇyaka	3455 [3]	582 [4]
Upaniṣad	13415 [11]	1747 [11]
Śrautasūtra	12445 [10]	1107 [7]
Gṛhyasūtra	16854 [13]	1837 [11]
Dharmasūtra	205 [0]	33 [0]
	126529	16525

Current dump of the syntactic data:
https://github.com/OliverHellwig/sanskrit/tree/master/papers/2020lrec/treebank

7

Background of Dependency Parsing

•Dependency parsing is the process of analyzing the grammatical structure of a sentence by determining the relationship between “head” words and the words that modify those heads

•Two parsing methods are widely used: transition-based and graph-based

•graph-based parsers show better performance for morphologically rich IE languages and for those of the SOV-type; Vedic has both characteristics

8

Parser design

•We adapt the biaffine parser of Dozat and Manning, 2017 since it achieved state of the art UAS on all CoNLL 09 languages and because it has better performance on non-projective languages

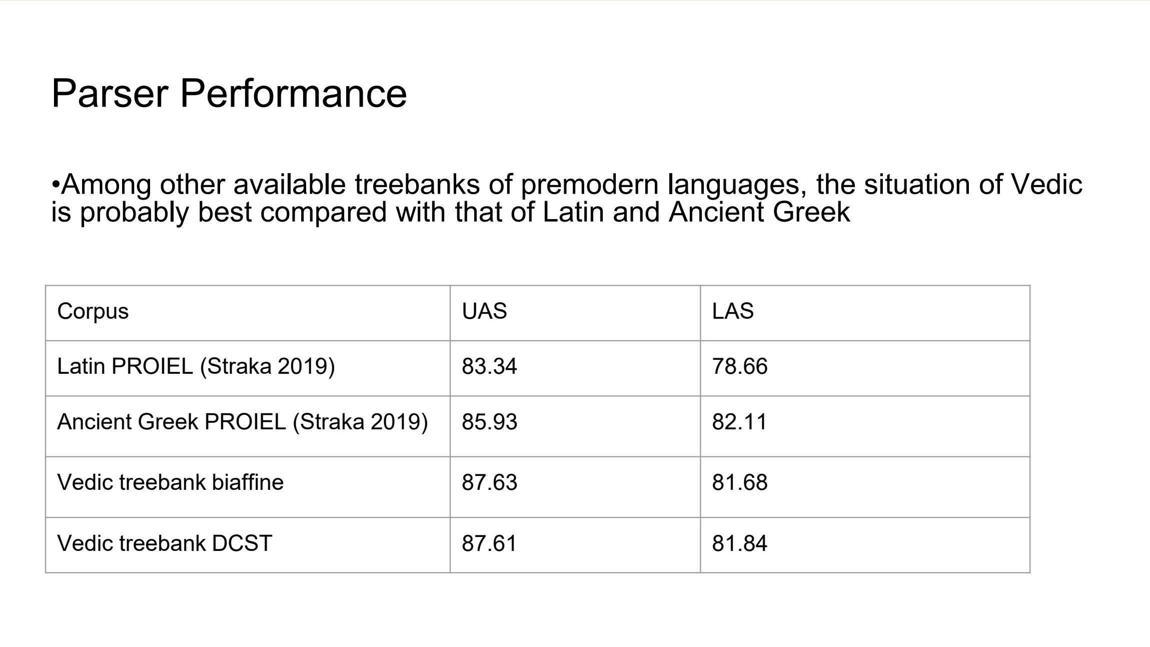
•We add a character-based CNN on the inflected form and integrate a larger number of categorical input features: Morpho-syntax, Verbal nouns, inflected word forms, punctuation and text-historical layers

•We augment the training data by randomly concatenating up to four, not necessarily subsequent sentences from the training set

• We evaluate the effect of deep contextual embedding models such as ELMo and BERT on the biaffine parser

•Other parsers for Sanskrit such as Kulkarni 2021 or Krishna et al. 2020/2021 exist, but their application to Vedic is not promising since they either rely on the Pāṇinian system of grammar (Kulkarni) or are trained on a rather small corpus of “Neo-Sanskrit” (Krishna et al. 2020/2021) and thus don’t perform satisfyingly on cross-domain tasks

9

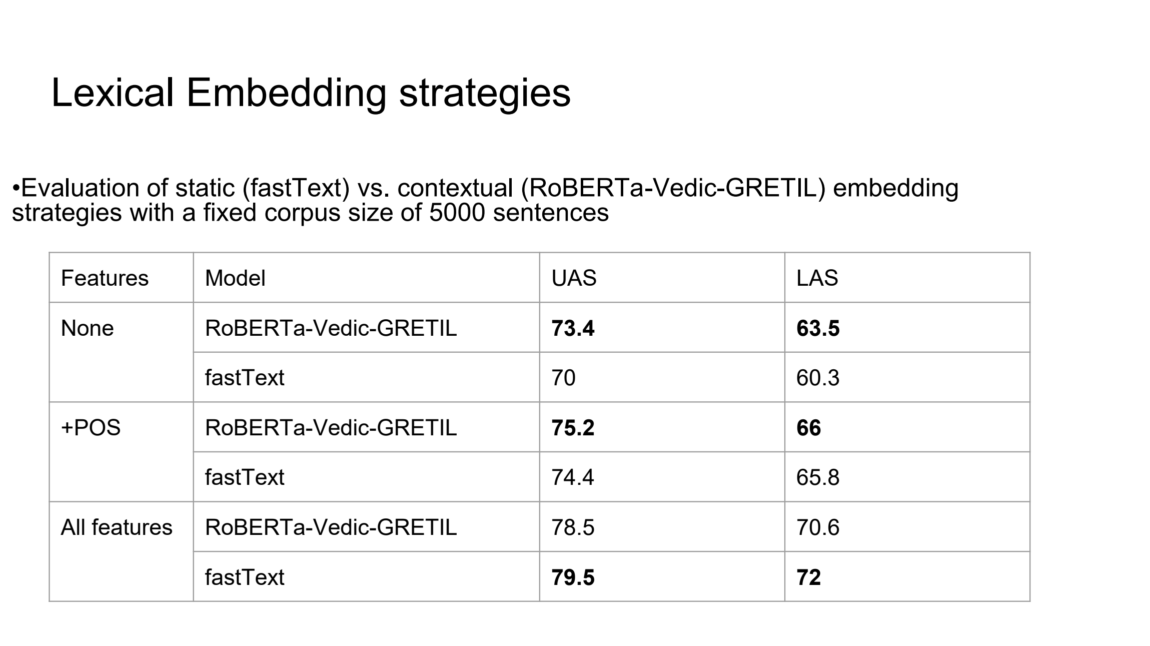


Parser Performance

- Among other available treebanks of premodern languages, the situation of Vedic is probably best compared with that of Latin and Ancient Greek

Corpus	UAS	LAS
Latin PROIEL (Straka 2019)	83.34	78.66
Ancient Greek PROIEL (Straka 2019)	85.93	82.11
Vedic treebank biaffine	87.63	81.68
Vedic treebank DCST	87.61	81.84

10

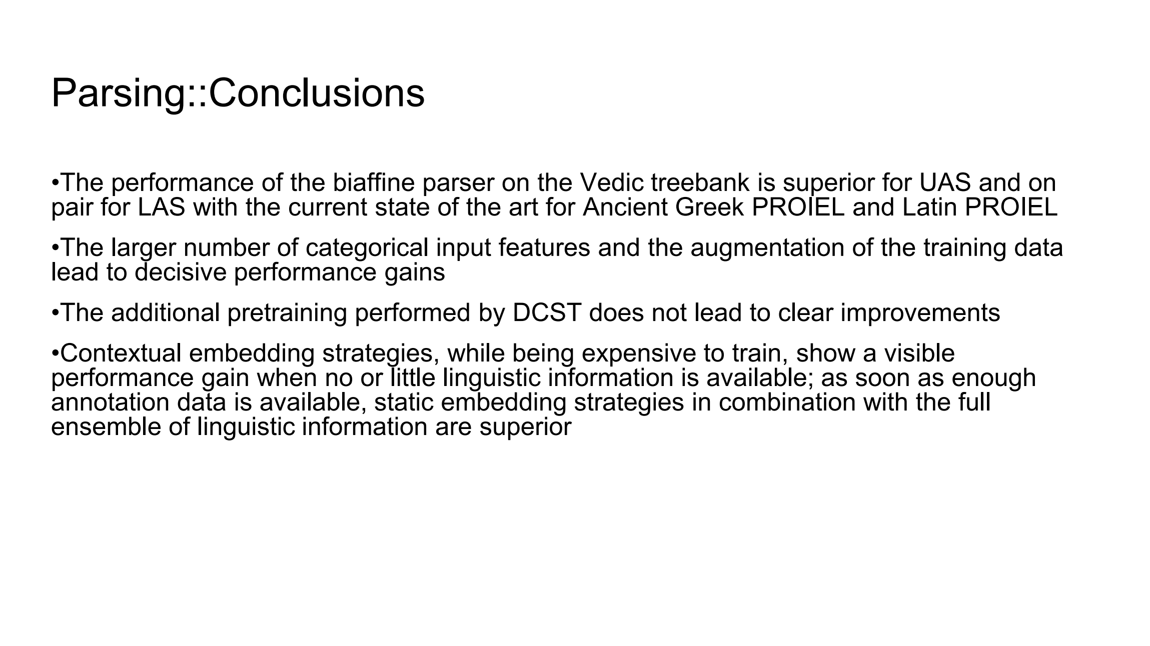


Lexical Embedding strategies

- Evaluation of static (fastText) vs. contextual (RoBERTa-Vedic-GRETIL) embedding strategies with a fixed corpus size of 5000 sentences

Features	Model	UAS	LAS
None	RoBERTa-Vedic-GRETIL	**73.4**	**63.5**
	fastText	70	60.3
+POS	RoBERTa-Vedic-GRETIL	**75.2**	**66**
	fastText	74.4	65.8
All features	RoBERTa-Vedic-GRETIL	78.5	70.6
	fastText	**79.5**	**72**

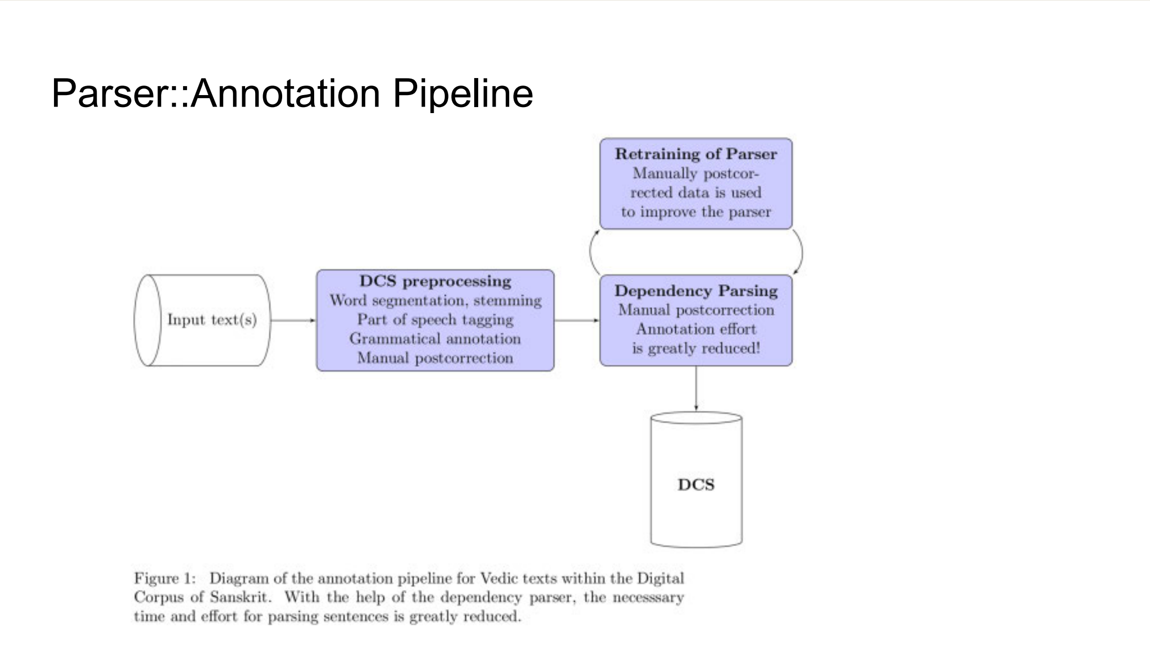
11



Parsing::Conclusions

- The performance of the biaffine parser on the Vedic treebank is superior for UAS and on pair for LAS with the current state of the art for Ancient Greek PROIEL and Latin PROIEL
- The larger number of categorical input features and the augmentation of the training data lead to decisive performance gains
- The additional pretraining performed by DCST does not lead to clear improvements
- Contextual embedding strategies, while being expensive to train, show a visible performance gain when no or little linguistic information is available; as soon as enough annotation data is available, static embedding strategies in combination with the full ensemble of linguistic information are superior

12



Parser::Annotation Pipeline

Input text(s)

DCS preprocessing
Word segmentation, stemming
Part of speech tagging
Grammatical annotation
Manual postcorrection

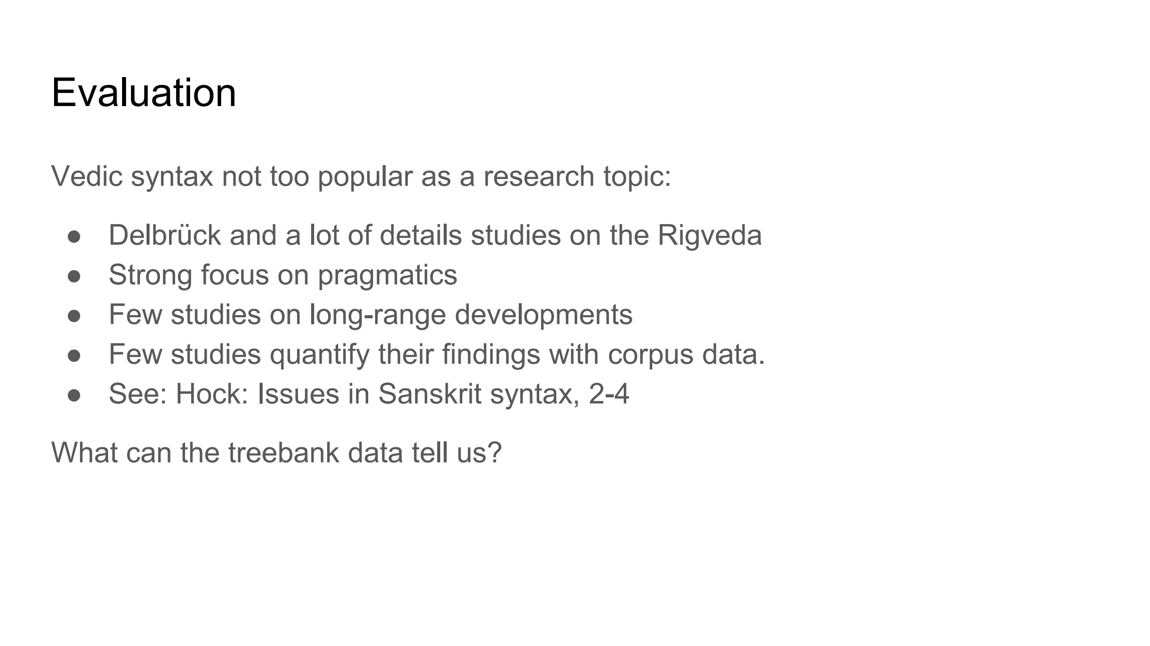
Dependency Parsing
Manual postcorrection
Annotation effort
is greatly reduced!

Retraining of Parser
Manually postcorrected data is used to improve the parser

DCS

Figure 1: Diagram of the annotation pipeline for Vedic texts within the Digital Corpus of Sanskrit. With the help of the dependency parser, the necesssary time and effort for parsing sentences is greatly reduced.

13



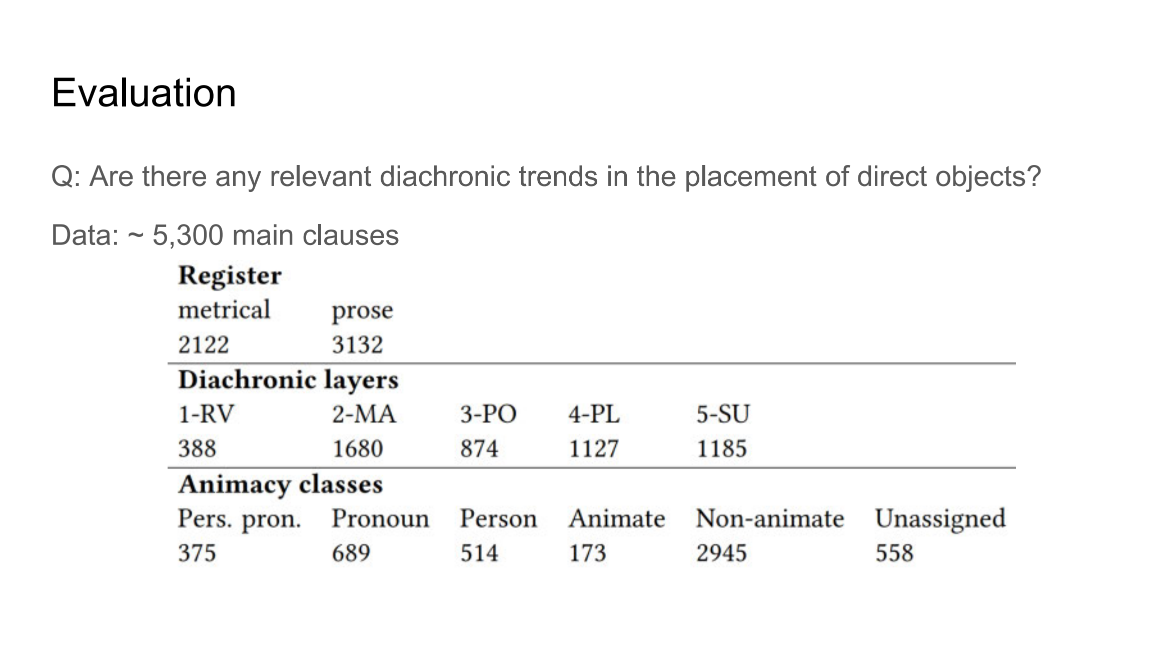
Evaluation

Vedic syntax not too popular as a research topic:

- Delbrück and a lot of details studies on the Rigveda
- Strong focus on pragmatics
- Few studies on long-range developments
- Few studies quantify their findings with corpus data.
- See: Hock: Issues in Sanskrit syntax, 2-4

What can the treebank data tell us?

14



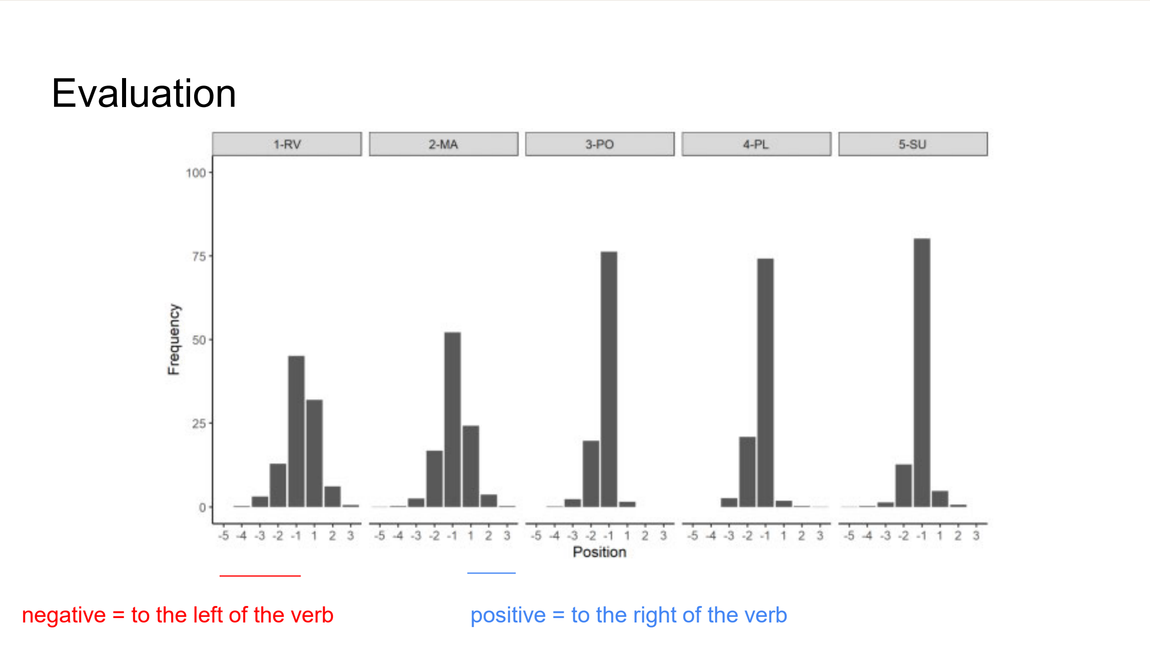
Evaluation

Q: Are there any relevant diachronic trends in the placement of direct objects?

Data: ~ 5,300 main clauses

Register					
metrical	prose				
2122	3132				
Diachronic layers					
1-RV	2-MA	3-PO	4-PL	5-SU	
388	1680	874	1127	1185	
Animacy classes					
Pers. pron.	Pronoun	Person	Animate	Non-animate	Unassigned
375	689	514	173	2945	558

15

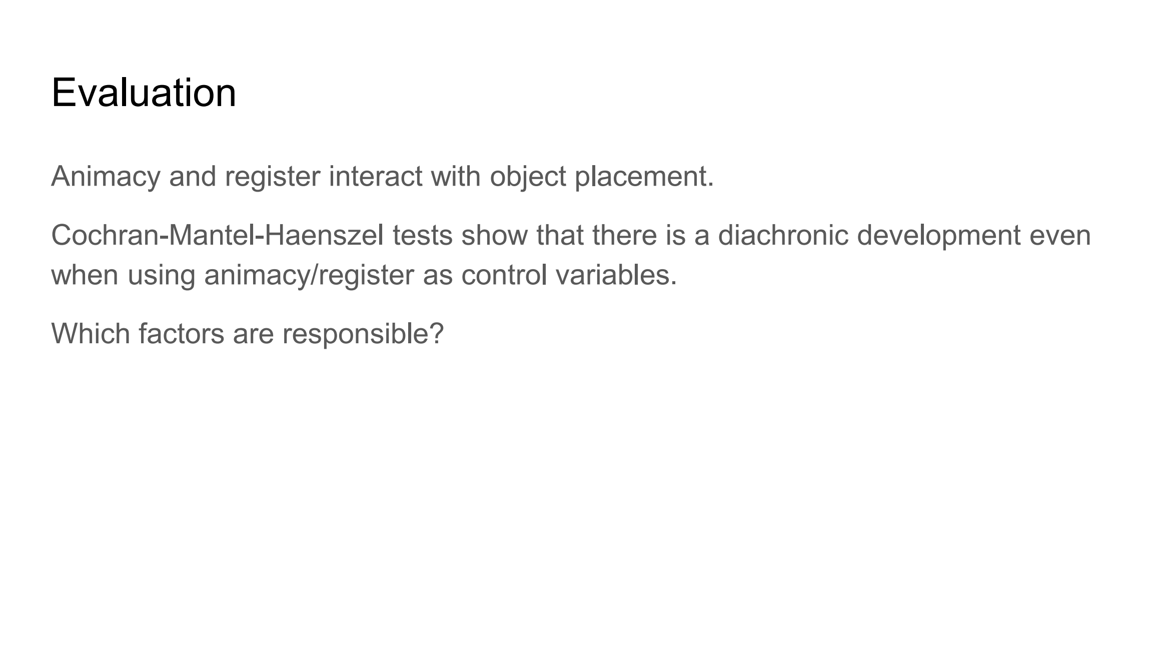


Evaluation

negative = to the left of the verb

positive = to the right of the verb

16



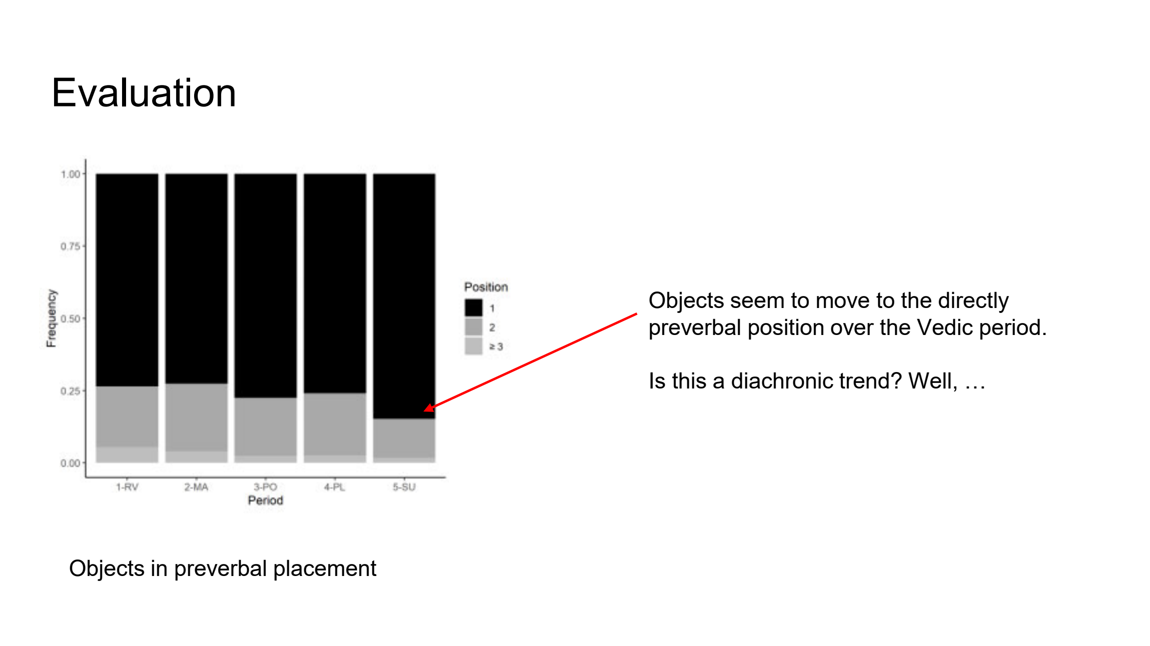
Evaluation

Animacy and register interact with object placement.

Cochran-Mantel-Haenszel tests show that there is a diachronic development even when using animacy/register as control variables.

Which factors are responsible?

17



Evaluation

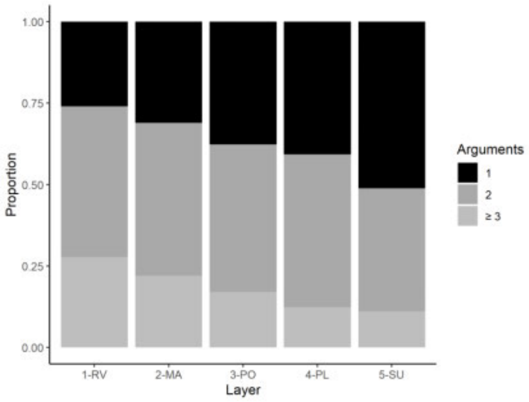
Objects seem to move to the directly preverbal position over the Vedic period.

Is this a diachronic trend? Well, …

Objects in preverbal placement

18

Evaluation



> The number of verbal arguments (nsubj, obj, obl, iobj) decreases significantly during the Vedic period.

19

Summary

- Syntactic data open new perspectives on the Vedic literature.
- Annotation interface and parsers are available. Collaborations are most welcome!
- Current focus: Detecting chronological syntactic trends and markers in an unsupervised manner.

Ancient India meets

Ancient India meets Data-Science

The 2nd and concluding Workshop of SPIRITS project
"Chronological and Geographical Features of Ancient Indian Literature Explored by Data-Driven Science",

2022
02/11
FRI
JP National Day
16:00-19:00 JST
(8:00-11:00 CET)

it's also
A Kick-off for Joint International Research
"A Study of Language Layers in Vedic Literature for the Development of a Program for Age-Estimation"

16:00-16:30 JST (= 8:00-8:30 CET)
The Result of the Two-Year SPIRITS Project and Our Vision for the Next Research
Kyoko Amano (Kyoto University, Hakubi Center / Institute for Research in Humanities)

16:30-17:00 JST (= 8:30-9:00 CET)
Visualization meets Ancient India: Mapping the Structure of Vedic Texts
Hiroaki Natsukawa (Kyoto University, Academic Center for Computing and Media Studies)

17:00-17:30 JST (= 9:00-9:30 CET)
"One Step Further: Assessing Semantic Similarity in Sanskrit Using Word Embeddings with a Weighting Factor"
Yuki Kyogoku (Leipzig University, Indology)

17:30-17:45 JST (= 9:30-9:45 CET)
Break

17:45-18:15 JST (= 9:45-10:15 CET)
"Computational Stylometric Analysis on Intertextuality in Historical Written Languages: A Case Study of Coptic"
So Miyagawa (Kyoto University, Graduate School of Letters / Center for Cultural Heritage Studies and Inter Humanities)

18:15-18:45 JST (= 10:15-10:45 CET)
Dependency parsing of Vedic Sanskrit - Algorithms and linguistic conclusions
Oliver Hellwig, Sebastian Nehrdich, Sven Selllmer (Dusseldorf University, Institute for Language and Information)

18:45-19:00 JST (= 10:45-11:00 CET)
Discussion and Concluding remark: Oliver Hellwig

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 until 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); Fostering Joint International Research (B) of KAKENHI "A Study of Language Layers in Vedic Literature for the Development of a Program for Age-Estimation" (Representativ researcher: Kyoko Amano)

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)

SPIRITS
SUPPORTING PROGRAM FOR INTERACTION-BASED INITIATIVE TEAM STUDIES

KYOTO UNIVERSITY KYOTO JAPAN FOUNDED 1897

Hakubi
Kyoto Univ.

Ancient India meets Data Science

Ancient India meets Data-Science
古代インドとデータサイエンス

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」
第２回（最終）ワークショップ

2022
02/11
金・祝
オンラインにて開催
16:00-19:00 JST
(8:00-11:00 CET)

国際共同研究
「ヴェーダ文献における言語層の考察と
それを利用した文献年代推定プログラムの開発」
のキックオフを兼ねて。

16:00-16:30 JST (= 8:00-8:30 CET)
The Result of the Two-Year SPIRITS Project and Our Vision for the Next Research
「２年間の SPIRITS プロジェクトの成果と今後の研究への展望」
天野恭子（京都大学　白眉センター / 人文科学研究所）

16:30-17:00 JST (= 8:30-9:00 CET)
Visualization meets Ancient India: Mapping the Structure of Vedic Texts
「可視化と古代インド研究：ヴェーダ文献の構造のマッピング」夏川浩明（京都大学 学術情報メディアセンター）

17:00-17:30 JST (= 9:00-9:30 CET)
"One Step Further: Assessing Semantic Similarity in Sanskrit Using Word Embeddings with a Weighting Factor"
「検証の次なる段階へ：重み付けを伴う単語分散表現によるサンスクリット文献の類似度推定」
京極祐希（Leipzig University, Indology)

17:30-17:45 JST (= 9:30-9:45 CET)
Break 休憩

17:45-18:15 JST (= 9:45-10:15 CET)
"Computational Stylometric Analysis on Intertextuality in Historical Written Languages: A Case Study of Coptic"
「文献言語における間テクスト性の計算言語学的・計量文献学的分析：コプト語における事例研究」
宮川創（京都大学　文学研究科 / 文学研究科附属文化遺産学・人文知連携センター）

18:15-18:45 JST (= 10:15-10:45 CET)
Dependency parsing of Vedic Sanskrit - Algorithms and linguistic conclusions
Oliver Hellwig, Sebastian Nehrdich, Sven Selllmer (Dusseldorf University, Institute for Language and Information)

18:45-19:00 JST (= 10:45-11:00 CET)
Discussion and Concluding remark
ディスカッションおよび総括 : Oliver Hellwig

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

定員なし、参加費無料
ワークショップ終了までご登録いただけます。

主催：京都大学研究支援 SPIRITS　2021「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)、国際共同研究強化 (B)2021-2026「ヴェーダ文献における言語層の考察とそれを利用した文献年代推定プログラムの開発」(研究代表者：天野恭子、課題番号：21KK0004)

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）2020-2022「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)

SPIRITS
SUPPORTING PROGRAM FOR INTERACTION-BASED INITIATIVE TEAM STUDIES

KYOTO UNIVERSITY FOUNDED 1897 KYOTO JAPAN

Hakubi 白眉 Kyoto Univ.