Title
Rethinking Scholarship: Data Analytics in Chinese Buddhist Philology

Permalink
https://escholarship.org/uc/item/72m0b5qx

Author
Borgman, Christine L.

Publication Date
2016-08-17
Rethinking Scholarship: Data Analytics in Chinese Buddhist Philology

Christine L. Borgman
Distinguished Professor and Presidential Chair in Information Studies
University of California, Los Angeles
http://christineborgman.info
@scitechprof

Digital Humanities and Sinology Session
Pacific Neighborhood Consortium Conference
Getty Center, Los Angeles, August 17, 2016
Big Data, Little Data, No Data: Scholarship in the Networked World

• Part I: Data and Scholarship
  – Ch 1: Provocations
  – Ch 2: What Are Data?
  – Ch 3: Data Scholarship
  – Ch 4: Data Diversity

• Part II: Case Studies in Data Scholarship
  – Ch 5: Data Scholarship in the Sciences
  – Ch 6: Data Scholarship in the Social Sciences
  – Ch 7: Data Scholarship in the Humanities

• Part III: Data Policy and Practice
  – Ch 8: Releasing, Sharing, and Reusing Data
  – Ch 9: Credit, Attribution, and Discovery
  – Ch 10: What to Keep and Why

MIT Press, 2015
Open access policies

- Open access to publications and data
  - Research Councils of the UK
  - European Union
  - Australian Research Council
  - U.S. Federal research policy
  - Chinese Academy of Sciences
  - Individual countries and funding agencies
Data
What are data?

Marie Curie's notebook

Pisa Griffin

Date: 1/2/07-75 Place: Sakàltutan
Zafor

He will grow old in his present house; new house is for sons - 5 sons. Not sure they want to live in village. He will only build another if they want him to. Es came from Germany and did the plastering. He arranged the carpentry in Kayseri Çok para gitti (much money went) Has a tractor.

Date: July 1980 Place: Sakàltutan
Zafor:

Household now Zafor and wife, Nazif Unal and wife and youngest son, still a boy. They run two dolmus; one with a driver from Sûleymani. Goes in and out once a day. He gets 8,000 a month. Zafor then said, keskin deölb { not sharp - i.e.? not profitable} I said he did very well on 8,000 TL with only two journeys a day. Nazif Unal has "bought" a Durak (dolmus stop) from Belediye and works all day in Kayseri.
Data are representations of observations, objects, or other entities used as evidence of phenomena for the purposes of research or scholarship.

Sustaining value in data

• Born digital
  – Malleable, reusable
  – Migrate to new technologies

• Digitized formats
  – Digital encoding
  – Static representations

Screen shot of searching and highlighting in CBETA, 2013. Photo by C.L.Borgman; computer of S.Zacchetti, used with permission

Chung hsi jih pao [microform] = Chung sai yat po, February 21, 1903; Alternative Title: Zhong xi ri bao; Chung Sai Yat Po Pub. Co, Publisher; February 21, 1903
Opening a box of data: Chinese Buddhist Philology

Stefano Zacchetti
Yehan Numata Professor of Buddhist Studies
Oriental Institute
University of Oxford
Fellow, Balliol College
Scholarly Inquiry

• How were Chinese Buddhist texts communicated across languages and cultures in the third to fifth centuries C.E.?
• Data sources: texts in multiple languages, scripts, and media
• Analysis:
  – Close: side by side comparisons
  – Distance: data mining of digitized texts

What are his data?

• Evidence of similar or dissimilar origins
• Observations of textual characteristics
• Units of data
  – Individual characters, punctuation marks
  – Words, phrases, sections, chapters
  – Artifacts, texts
  – Entities
  – Collections of texts

Microfilm as Data Source


Chung hsi jih pao [microform] = Chung sai yat po, February 21, 1903
Alternative Title: Zhong xi ri bao
Chung Sai Yat Po Pub. Co, Publisher; February 21, 1903
Bricks in the wall...

Brick inscribed with the Sutra on Dependent Origination *Gorakhpur district, late 5th century - early 6th century AD*. Ashmolean Museum
In order to accommodate the durative/frequency expressions, it is proposed that the latter are adjoined to the left of V (Huang, Li, Li 2009:100). Their post-verbal appearance is then the consequence of verb raising across them. An alternative to “adjunction to V” (also mentioned by Ernst, 2014) is to postulate that durative/frequency expressions in Chinese are actually in the unique SPEC position of Larsonian VP Shells. This would explain why the co-occurrence of multiple Frequency/Durative expressions is not allowed in Chinese.

Gu (1995) notes that in Chinese there are some “verbal measure phrases” that need to be analyzed in a different way from Durative/Frequentative. One of the features they show is the possibility to either follow or precede the post-verbal object. See the following examples (from Gu 1995:64 example 25):

(331) Zhangsan ti-le yi jiao nei-ge ren
   oadsgna kick-ASP one foot that-CL person
   'Zhangsan gave that man a kick.'
(332) Zhangsan ti-le nei-ge ren yi jiao
   oadsgna kick-ASP that-CL person one foot
   'Zhangsan gave that man a kick.

The analysis proposed is that yi jiao may be base-generated in COMPL of VP and raise together with V to the V head.

In general, the order of base-generation of the verb arguments is associated with a thematic hierarchy. The one proposed by Larson (1988) is Agent > Theme > Goal > Oblique (manner, location, time). For Chinese, Huang (1991) has proposed an alternative: Agent > Experiencer > referential theme > goal, indirect Object > obliques (non-referential theme, direction, duration/frequency, manner ...). The lowest role in the thematic hierarchy corresponds to the lowest position in the VP-Shell structure.

Cinque (2009, 2010) proposes that all DP/PP arguments occupy a SPEC position: he therefore eliminates the COMPL position as a base-generation position for verbal arguments. The resulting structure is consistent across languages, with word order differences to be derived through movement of the verb.

---

81 On verbal classifiers, see also Matthews & Yip (1999).
82 Non-referential theme refers to the ‘dummy’ objects such as fan in chi fan (lit. ‘eat food’ = ‘eat’) or bu in pao bu (lit. ‘run step’ = ‘run’).
83 In Chinese temporal adjuncts, place and preverbal manner need to be in higher position in order to justify the pre-verbal position.
打喷嚏
In Praise of the Light

A Critical Synoptic Edition with an Annotated Translation of Chapters 1-3 of Dharmarakṣa’s Guang zan jing 光讚經，Being the Earliest Chinese Translation of the Larger Prajñāpāramitā

Stefano ZACCHETTI

The International Research Institute for Advanced Buddhology
Soka University
Tokyo 2005
punar apraṇa śāradvatiputra
bodhisattvamahāsattvadenānānāṁ
navagṛhitacitenaṁ bhaviturkāmena
334 dāvulaḥlīkacitam/prthakkartukāmena
336 vāyāpādacitam utraṣṭukāmena
337 kauśiyācittam dāusprajñācittam
338 anutpāday[j]tukāmena
339 prajñāpāramitāyāṁ śīkṣitavyam* //

PD 29, 18-30, 2 (cf. PG 13v 6-8, 95, 20-96, 11):
punar etc. bodhisattvamahāsattvadānānāṁ
sarvāyāpādacitavastunī pratiṣṭhāpayitukāmena
śālakāyāparyāpāyasyavastunī pratiṣṭhāpayitukāmena
bhāvanāmaṇyapāṇyakriyāvastunī pratiṣṭhāpayitukāmena
vaiyāvyāśayagāte ca upadhike pratiṣṭhāpayitukāmena
punyakriyāvastunī pratiṣṭhāpayitukāmena
prajñāpāramitāyāṁ etc.

PD 30, 3-4 (PG 13v 8-10; 96, 11-20):
punar apraṇa śāriputra bodhisattvahāsattvavastunī puṇa ca ekta
utpādayitukāmena

334 PG 13v 7 & 96, 1 not as a compound: dānamayapunyakriyāvastunī etc. (so also the following terms).
335 PG 13v 7: vaiyāṣṭyeṣvahagāte (cf. BHSD 511a-b); 96, 7-8: airmāyapānahagāte (cf. BHSD 158a).
336 PG 13v 7 & 96, 10: sarvaṣadhiṣṭhagāte. Cf. Lamotte V 2246 for an exam of the various
versions of this passage; on he remarks: "corrig: suṣṭhikṣṭhagāte".
337 PG 13v 8 & sarvaṣārāṇi; 96, 12: sarvaṣārāṇ.
338 PG 13v 9 & 96, 12: nisupādayitukāmena, passim.

**INTRODUCTION – CHAPTER 1**

( )
\[
\text{restored } \text{aṣṇara(s)}
\]
\[
\text{damaged } \text{aṣṇara(s)}
\]

\[
\text{interlinear insertion}
\]

\[
\text{erased } \text{aṣṇara(s)}
\]

\[
\text{omitted } \text{aṣṇara(s)}
\]

\[
\text{superfluous } \text{aṣṇara(s)}
\]

\[
\text{virāma}
\]

\[
\text{one illegible } \text{aṣṇara}
\]

\[
\text{illegible part of an } \text{aṣṇara}
\]

\[
\text{jihvaṃūliya}
\]

\[
\text{upadhmaṇīya}
\]

\[
\text{visarga used as punctuation mark}
\]

\[
\text{dot-like punctuation mark}
\]

\[
\text{double stroke used as punctuation mark}
\]

\[
\text{avagraha, not written in the MS}
\]

\[
\text{corrupt passage}
\]

The only difference is in the beginning of roll 3, which occurs in the middle of chapter 3b in the most direct offspring of Sixi (Q and its tracing Nz1), while in the remaining editions it coincides with the beginning of chapter 4. The latter is very probably an innovation adopted by the editors of Y, and it proves in turn very helpful for establishing further, more precise subdivisions in the lower reaches of the family (see § 3.3.2 below).
Screen shot of searching and highlighting in CBETA, 2013. Photo by C.L.Borgman; computer of S.Zacchetti, used with permission
CBETA as Microscope and Telescope

Commons photo: Science Gossip, 1894

Telescope for the Sloan Digital Sky Survey, Apache Point, New Mexico
Data Scholarship: Chinese Buddhist Philology

• Data Types
  – Texts as material objects
  – Images of texts
  – Digital texts

• Data Sources
  – Libraries, archives
  – Private collections
  – Personal collections
  – Public collections

https://www.rate-watch.com/scholar/
Data Scholarship: Chinese Buddhist Philology

• Models
  – Dissemination of texts as “meritorious acts”
  – Texts are versions, not originals
  – Form and content evolved over time

• Theories
  – Claims about origins of texts
  – Claims about relationships between texts

https://www.rate-watch.com/scholar/
Data Scholarship: Chinese Buddhist Philology

- Metadata, ontologies, standards
  - Technical standards for text encoding and markup
  - Collection-specific metadata
  - Local practices for data collection and management

https://www.rate-watch.com/scholar/
Data Scholarship: Chinese Buddhist Philology

• Knowledge Infrastructure
  – Data sources are diverse and distributed
  – Technical resources are minimal
  – Standards are generic
  – Informal scholarly communication
  – Sustainability of content is decentralized
    • Personal data management
    • Religious communities
    • Libraries, archives, private collections

https://www.rate-watch.com/scholar/
Sustaining value in scholarship

• Invest in your data early and often
• Capture data in portable formats
• Migrate data as technology evolves
• Build community
  – Collections and data
  – Tools and expertise
  – Mechanisms for sharing and reuse
• Build archives and repositories
• Reuse, recombine, rethink

https://www.rate-watch.com/scholar/
Data Stewardship

- Culture
- Knowledge
- Responsibility
- Commitment
- Resources
- Actions

Community norms and goals
Knowledge about stewardship
Who takes responsibility for data
Who makes long-term commitments
Resources available for stewardship
Who takes what actions
When to invest in data?

http://www.lib.uci.edu/dss/images/lifecycle.jpg
Data Scholarship

- Data types and sources
- Models and theory
- Community practices
- Scaling factors
- Temporal factors
- Metadata, ontologies, standards
- Economics, intellectual property
- Knowledge Infrastructure

https://www.rate-watch.com/scholar/
If Data Sharing is the Answer, What is the Question?

1. How do disciplinary configurations influence the collection, use, and reuse of data and vice versa?

2. What scale factors, such as data, discipline, distribution, and duration, influence data practices, and how?

3. How does the degree of centrality of data collection influence use, reuse, curation, and project strategy and vice versa?

Grant from Alfred P. Sloan Foundation to UCLA, Christine L. Borgman, PI, Center for Knowledge Infrastructures, 2015-2018
Studying Data Practices

• Ethnography
  – Observing activities on site
  – Embedded for days or months at a time

• Interviews
  – Questions based on our research themes
  – Compare multiple sites over time

• Document analysis
  – Public and private documents and artifacts
  – Official and unofficial versions of scientific practice