JADH Conference 2014

—Bridging GLAM and Humanities through Digital Humanities—

ABSTRACTS

19–21 September 2014

Union of Library and Information-media Studios
(ULIS Bldg., Kasuga Campus)
University of Tsukuba

http://conf2014.jadh.org/

Hosted by:
JADH2014 Organizing Committe
under the auspices of the Japanese Association for Digital Humanities

Co-hosted by:
Faculty of Humanities and Social Sciences, University of Tsukuba
Faculty of Library, Information and Media Science, University of Tsukuba
Centre of Research for Development in
 Japanese Language, Culture and Society, University of Tsukuba
Digital Humanities Research Unit, University of Tsukuba
Digital Humanities Initiative, Center for Evolving Humanities,
 Graduate School of Humanities and Sociology, The University of Tokyo
International Institute for Digital Humanities

Co-sponsored by:
The Mathematical Linguistic Society of Japan
IPSJ SIG Computers and the Humanities
Japan Association for East Asian Text Processing (JAET)
Japan Association for the Contemporary and Applied Philosophy
Japan Art Documentation Society (JADS)
Japan Society of Library and Information Science (JSLIS)
Japan Society for Information and Media Studies (JSIMS)

Programme Committee:
Tabata, Tomoji (Osaka University, Japan), Chair
Akama, Hiroyuki (Tokyo Institute of Technology, Japan)
Arthur, Paul (Australian National University, Australia)
Cummings, James (University of Oxford, UK)
Fraistat, Neil (University of Maryland, USA)
Hara, Shoichiro (Kyoto University, Japan)
Hsiang, Jieh (National Taiwan University, Taiwan)
Inaba, Mitsuyuki (Ritsumeikan University, Japan)
Muller, Charles (University of Tokyo, Japan)
Murai, Hajime (Tokyo Institute of Technology, Japan)
Miyake, Maki (Osaka University, Japan)
Nagasaki, Kiyonori (International Institute for Digital Humanities, Japan)
Nerbonne, John (University of Groningen, Netherlands)
Ore, Espen (University of Oslo, Norway)
Pierazzo, Elena (King's College London, UK)
Rockwell, Geoffrey (University of Alberta, Canada)
Schreibman, Susan (National University of Ireland Maynooth, Ireland)
Shimoda, Masahiro (University of Tokyo, Japan)
Siemens, Raymond (University of Victoria, Canada)
Suzuki, Takafumi (Toyo University, Japan)
Uda, Norihiko (University of Tsukuba, Japan)
Wittern, Christian (Kyoto University, Japan)
Yamada, Taizo (University of Tokyo, Japan)

Organizing Committee:
Ikeda, Jun (Faculty of Humanities and Social Sciences, University of Tsukuba), Chair
Doi, Hiroto (Faculty of Humanities and Social Sciences, University of Tsukuba)
Ikeuchi, Atsushi (Faculty of Library, Information and Media Science, University of Tsukuba)
Ishida, Takeru (Faculty of Humanities and Social Sciences, University of Tsukuba)
Matsumura, Atsushi (Faculty of Library, Information and Media Science, University of Tsukuba)
Miyake, Yutaka (Faculty of Humanities and Social Sciences, University of Tsukuba)
Nagai, Masakatsu (Faculty of Humanities and Social Sciences, University of Tsukuba)
Nagasaki, Kiyonori (International Institute for Digital Humanities)
Ono, Motoi (Faculty of Humanities and Social Sciences, University of Tsukuba)
Sakaguchi, Tetsuo (Faculty of Library, Information and Media Science, University of Tsukuba)
Suzuki, Nobutaka (Faculty of Library, Information and Media Science, University of Tsukuba)
Takahashi, Yona (Faculty of Humanities and Social Sciences, University of Tsukuba)
Tokii, Maki (Faculty of Library, Information and Media Science, University of Tsukuba)
Uda, Norihiko (Faculty of Library, Information and Media Science, University of Tsukuba)
Waki, Toshihito (Faculty of Humanities and Social Sciences, University of Tsukuba)
Watanuki, Toyoaki (Faculty of Library, Information and Media Science, University of Tsukuba)
Yazawa, Makoto (Faculty of Humanities and Social Sciences, University of Tsukuba)

Programme and Contents

Sunday, 21 September

16:30 Break

16:40 Closing Plenary Lecture (Harold Short) (Media Hall, 2F)

Lecturer: *Harold Short*

17:30 Closing (Media Hall, 2F)

Poster/Demo Presentation

Poster Session (Lecture Room, 1F)

Additional Demonstration (Lecture Room, 1F)

PLENARY

Life on the Outside:
Collections, Contexts, and the Wild, Wild Web

Tim Sherratt (National Library of Australia / University of Canberra)

More and more cultural collections are being released online. Such acts of liberation open collections to discovery and reuse, but carry the risk that something integral to their interpretation, understanding and authenticity might be lost. Context, it seems, can be fickle and fragile.

But are our collections more robust than we think? This paper will explore some of the new lives and meanings they can find in the wilds of the web. Contexts can be created and enriched. New ways of seeing can emerge, bringing both understanding and unease. Many small, capricious, joyous, angry acts of meaning-making can challenge ideas of authority and privilege inherent in fixed conceptions of context. Things can change.

Life on the outside isn't always pleasant or predictable, but by letting our collections run free we allow them to grow. And we challenge ourselves to think differently about the nature of collections themselves.

Developing and Sustaining Digital Humanities Partnerships

Paul Arthur (University of Western Sydney)

In this talk I refer to three ways in which digital humanities collaborations can be developed and sustained. The first is around building networks and associations of people. The second is the role that very large or long-term projects can play in building research communities. The third is how institutional frameworks within universities or research institutions can promote digital humanities to wider constituencies. I give the example of initiatives from Australasia including the Digital Humanities 2015 international conference, to be hosted by the University of Western Sydney.

A Visualization and Analysis System for Japanese Language Change: Quantifying Lexical Change and Variation using the Serial Comparison Model

Bor Hodošček (Meiji University)   *Makiro Tanaka (Meiji University)*
Hilofumi Yamamoto (Tokyo Institute of Technology / University of California)

This paper introduces an online system for the visualization and analysis of Japanese language change spanning a little over a century (1874–2008). The scale and depth of possible analyses of diachronic linguistic change has been greatly increased through the efforts of large book digitalization projects such as the Google Books corpus (Michel et al., 2011) as well as more curated historical corpora such as the Corpus of Historical American English (COHA) (Davies, 2010). Furthermore, these resources are supplemented with online search interfaces offering relatively complex query capabilities, including POS and syntactic annotation based search queries (Davies, 2011). The present situation for the diachronic analysis of Japanese language corpora is not developed to the same extent: while corpora representing different time periods and genres are available, they are isolated and encoded in incompatible formats. Furthermore, while some of them provide search interfaces (Chūnagon and NINJAL-LWP are two such examples), the incomplete coverage of corpora and lack of focus on diachronic analysis means that the analysis workflow is not as streamlined and automated as it is for English. Having provided some of the motivations for developing the system, we detail the corpora and processing methodologies to be used in constructing the system, as well as introduce the role of the Serial Comparison Model in quantifying lexical language change.

The following five corpora are used as a sample of modern and contemporary written Japanese:

- The Balanced Corpus of Contemporary Written Japanese (c. 1975–2008)
- The Sun corpus (c. 1895–1925)
- The Meiroku Zasshi corpus (c. 1874–1875)
- The Kindai Josei Zasshi corpus (c. 1894–1925)
- A subset of the Aozora Bunko (c. 1890s–)

The present collection of corpora does not evenly cover the different genres of Japanese as is the case in some other diachronic corpora such as COHA (Davies, 2013). However, through the use of metadata such as the Nippon Decimal Classification (NDC) system, date of publication, and other available metadata on the authors, well-specified subsets of the collection can nevertheless provide insights into language change in the context of the user's query.

All text is processed into morpheme tokens using the morphological analyzer MeCab and, depending on the time period, the modern or contemporary version of the UniDic morphological dictionary. A unique property of both variants of UniDic is their organization of word tokens under lemma that cover the many orthographic variants observed in Japanese writing. Taking the basic lemma-word orthography pairs as a base, we construct cooccurrence networks between all words occurring in the same sentence or paragraph. This cooccurrence network is constructed so that we are able to generate sub-networks that match some metadata query, such as year and NDC code, which can then be used to compare with other sub-networks.

The Serial Comparison Model (SCM) introduced in Yamamoto, Tanaka, and Kondo (2012) models lexical changes occurring between two reference frames. Taking as an example a diachronic analysis between two corpora A and A' sampled from different time periods t and t', the model provides five classification codes representing the different possible transitions of a word between two time periods (Figure 1). Using this model and its categorization scheme, we will visualize changes within a lemma's

SESSION 1

cooccurrence network in the transition between two or more time periods.
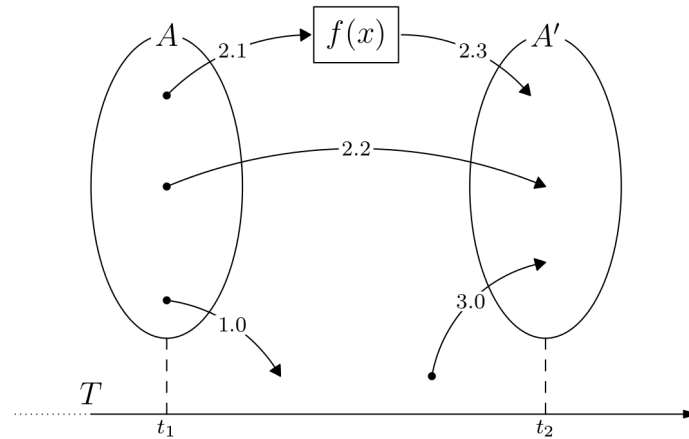
Figure 1: Diachronic lexical transitions as represented in the Serial Comparison Model: 2.2 represents no change in form; 1.0 and 3.0 respectively represent the discarding and introduction of a form; 2.1 and 2.3 represent a change in form; $f(x)$ is the transition function behind the change from 2.1 to 2.3.

Reference

[1] Davies, M.: *The Corpus of Historical American English: 400 million words, 1810–2009,* (2010), `<http://corpus.byu.edu/coha/>` (visited on 05/07/2014).

[2] Davies, M.: *Google Books (American English) Corpus (155 billion words, 1810–2009),* (2011). `<http://googlebooks.byu.edu/>` (visited on 05/07/2014).

[3] Davies, M.: "Google Scholar and COCA-Academic: Two Very Different Approaches to Examining Academic English," *Journal of English for Academic Purposes,* Vol. 12, No. 3, pp. 155–165 (2013).

[4] Michel, J.-B. et al.: "Quantitative Analysis of Culture using Millions of Digitized Books," *Science,* Vol. 331, No. 6014, pp. 176–182 (2011).

[5] Yamamoto, H., Tanaka, M. and Kondo, Y.: "Diachronic Corpus and Linguistic Space: New Methods for the Analysis of Language Change," *The 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing (SNPD),* IEEE, pp. 381–384 (2012).

Defining Literary Data:
An Information Theoretical Approach

Devin Higgins (Michigan State University)   *Thomas Padilla (Michigan State University)*
Arend Hintze (Michigan State University)

In addition to forming a piece of the lasting and living embodiment of the cultural heritage of humanity, literature also comprises a form of data. The features of this data are precisely what define the "literary" as such. In order to "understand the structural continuity of the step from information to literature and back again...[and] to grasp the nonuniqueness of literature in an absolute structural sense," that is, to specify a difference of degree rather than kind between literature and other forms of data, it will be necessary to isolate and define the features of the literary band of the data spectrum in as nuanced a way as possible.[1]

To isolate and define features of literary data, the authors will employ a range of information-theoretical techniques to analyze literary text and find distinguishing patterns. An algorithm developed at Michigan State University to study the information novelty in DNA sequences can equally be applied to strings of arbitrary text. This algorithm has been used to quantify how much information is generated by Twitter users on a daily basis, and will be adapted to measure information novelty patterns over the duration of texts. Do literary and non-literary texts give rise at similar rates to new tokens of length n, to new words, and to new concepts? How does the frequency with which new n-long character strings appear vary by genre? Comparing corpora of literary texts (defined in published bibliographies) against random subsets of non-canonized texts published in the same era and location, will potentially reveal distinctive patterns of information.

In addition to literary and non-literary corpora, these algorithms will also be used to assess highly formalist works which employ diverse lexical and structural strategies to achieve new literary effects, such as works originating from the OuLiPo, and Christian Bök's Eunoia. How will radical approaches to form reveal themselves within the text as data? Are algorithmic literatures particularly amenable to computational analysis?

A scientific theory of literature existed as a premise for literary analysis throughout the 20th century. Formalists, structuralists, semioticians, and others utilized scientific techniques to describe and formalize aspects of language particular to the literary domain. Advances in computing and the advent of information theory in the middle of the century led to perhaps the first conference on literary data in 1964.[2] Broadly construed, these approaches to literary studies have been characterized as reductive yet the impulse to structure and categorize the text was not intended to ascribe unquestionable stability to literature. Rather, every text "manifests to some degree features of a work of art."[3] The "literary" exists on a spectrum and the motive of the information theoretical approach is not to categorize, but to trace the signal.

In order to perform this analysis, appropriate datasets will need to be collected. The proposed project will describe methods of extracting relevant texts from the portion of the Google books dataset that MSU hosts locally—approximately 3 million volumes. Using linked data to sort and extract texts in the dataset according to non-bibliographic forms of description will allow the authors to create textual datasets based on an expanded list of parameters, and with greater granularity, than is currently possible using existing search and retrieval functions.

The presentation at JADH will share techniques for the creation of these worksets, in addition to describing the results of new investigations into the informational content and form of literary text construed as data.

Reference

[1] Parker, Terence quoted in Hayot: "What is Data in Literary Studies?" `http://erichayot.org/ephemera/mla-what-is-data-in-literary-studies/`

[2] Lotman, Yurij M.: "The Future for Structural Poetics," *Poetics,* Vol. 8, No. 6 (1979).

[3] Literary Data Processing Conference proceedings, September 9, 10, 11 (1964). IBM, Data Processing Division (1964).

SESSION 1

Extracting Factors of Small Stories from the Synoptic Gospels

Hajime Murai (Tokyo Institute of Technology)

SESSION 1

In the Bible, the four Gospels of the New Testament record the words and deeds of Jesus Christ, putting them at the center of Christian thought over the centuries. Among these Gospels, various theological concepts are presented in the form of parables or stories told by Jesus, it is difficult to objectively identify the structure of concept relationships within them. One of the reasons is that the Bible includes abstract representation. It is difficult to interpret the meaning of each representation. Further, it is believed that the Gospels were written through complicated editing processes, therefore similar expressions and contents are found in multiple places. This is also one of the reasons why it is difficult to understand the entire structure of the theological concepts of the Gospel.

However difficult it may be to interpret the meaning of the abstract representation, it is possible to extract a word or phrase which, used characteristically, refers to a particular theological concept. Further, it might be possible to analyze the relationship of the words and phrases that indicate the various concepts scattered by the complex editing processes, and such an analysis may contribute to estimating the intention of the editor or editors of the Gospels.

In this study, the Synoptic Gospels (that is, Matthew, Mark, and Luke), which are considered to have a high degree of similarity, are selected as the target. Based on an analysis of the co-occurrence words and phrases, the relationships between them are extracted in order to estimate the overall theological structure of the Gospels.

In general, similar contents are often expressed by different words. Even in the original Greek Gospels, similar content is represented by multiple words. In the exegesis of the Bible, it is important to read the nuances by analyzing in detail the use of such words. But this study focuses on estimating the rough structure of the theological concept of the Gospels; therefore, similar words are treated as a group of synonyms. The method for grouping words is based on the interpretation of the Bible scholars of later generations. Two different words are included in the same group if the two different words of the original text are translated into the same word in the translation. Four Japanese Bible translations and four English Bible translations were used as source data, and words were grouped together as synonyms of the original text if such grouping is common in the translation of five or more versions. As a result, synonym group of 192 species were obtained.

In order to extract the theological concept, which is composed of the co-occurring synonym groups and words, the factor analysis method was applied, based on the small story segments of the Gospels (referred to as "pericope" in biblical studies). To extract from major words and synonym groups a factor that related to the gospel as a whole, the variables of factor analysis were selected from synonym groups and words that appear in more than ten percent of all small story segments. As a result, 119 words and synonym groups were analyzed in the factor analysis. As a result of the factor analysis, 42 factors whose eigenvalues are more than 1 were extracted. The structure and relationship of the theological concepts in the Synoptic Gospels was expressed as a network based on the correlation coefficient of these factors. Also, the differences of factors in Matthew, Mark, and Luke were extracted based on the factor scores.

Keywords Bible, Goepel, structure, factor analysis, synonym

SESSION 2

Visualization of the Practices of the Theravadins in Mainland Southeast Asia on Google Earth

Makiko Harada (Tokyo Metropolitan University)
Julien Bourdon-Miyamoto (Kyoto University)
Hidenori Watanave (Tokyo Metropolitan University)

Figure 1: Screenshot of the content

Geo-spatial and temporal data constitute the heart of Area Informatics[1]. Beside pure research objectives, some projects require specific visualizations in order to appeal to the general public.

In this paper, we propose to construct visual web content with Google Earth (Figure 1), and provide a system and interface to easily insert new data to which new information is added day by day. In the study, we visualize the practices of the Theravadins in Mainland Southeast Asia. There are two classes of data. "Temple data" includes pictures, GPS data, years of consecration defined by the date that temple was allowed to host monks, and facilities. "Monk data" includes the id of the monks, home villages, and their whereabouts from 2001 to 2010. We extract data as expanded XML.

We describe the interface of the system below.

Operation Screen

The operation screen contains information required to control particular data on Google Earth. It contains four types of information.

- Demonstration of tracking: This shows the movement of a specific set of monks.
- List of temples / monks: When one is selected, it flies to the temple or makes a tracking line of the monk. The detailed data of the temple and the monk is inserted into the information table.
- Information table: The detailed text data is inserted here.
- Google Map: This moves in conjunction with Google Earth. It is a bird's eye view.

Interface of Google Earth

- Icon: The icon image is the picture of the temple. The icon is expanded by a mouse over. The tracking line of a monk who visits the temple is drawn, and the detailed data of the temple is inserted into the information table by a single click. Users can see the detailed high resolution pictures of the temple by a double-clicking.
- Time line: This can handle the visible information based on temporal data. It enables users to see the icon and the line within the parameters of time.

Data uploading system

We construct the PHP code which converts the expanded XML into KML for Google Earth and generates javascript automatically. To reflect them on web content, all we have to do is to add some text to "index.html". Users input the text to call up javascript in the header, and set up an array of data to read.

This content contains a multitude of visual elements and some text data. Since the earth, which is three-dimensional, can be handled by simple mouse operation, the content is able to show research data to the general public more intuitively and easily than the previous visual content for experts in analysis[2]. Additionally, this data uploading system enables Area Informatics researchers to publish current research results easily. This content can be accessed by the following URL[3].

Keywords visualization, Area Informatics, Theravadins, map

Reference

[1] Hara, Shoichiro: "Area Informatics: Concept and Status," *Culture and computing,* Springer Berlin Heidelberg (2010), pp. 214–228.

[2] `http://gaia.net.cias.kyoto-u.ac.jp/mapping/`

[3] `http://templemapping.jp/`

SESSION 2

Landscape Documentation: Collecting "Personal" Landscapes for Sharing within Communities

Tsuyoshi Tamura (Akita University of Art)   *Shoko Sumida (Ritsumeikan University)*
Mariko Kaname (Osaka University)

The aim of this study is to contribute to the awareness of the community landscape among residents so that they will regard maintaining the local landscape as an urgent matter for regional development. Additionally, drawing one's internal image of the landscape should be encouraged to ensure that adequate documentation exists.

First, we determined whether members of a community recognized their local landscape based on their actual experiences; we referred to a "landscape philosophy" espoused by Professor Nobuo Kioka of Kansai University. According to his theory, a characteristic of landscape structure is its susceptibility to social effects. This model has a trilaminar, triangular structure from bottom to top, a "fundamental landscape" experienced by individuals, a "collective landscape" shared within a defined community, and an "expressive landscape" typified by visual images on a postcard or famous painting and so on. These parts must be able to influence one another, but we doubt that individual fundamental landscapes have been sufficiently shared within the community. Therefore, our research focuses on the images, which are supposed to be immersed in the fundamental landscape, tentatively referencing the personal landscape here.

Second, our research explores the means of collecting personal landscapes for documentation; we incorporated a unique interview method in this effort. Data reflecting personal landscapes contain visual images (e.g., photos) and texts (e.g., comments). Thus, through the process of sorting these images and texts, personal landscapes were transformed into visible and readable data; ultimately, the data became recordable as digital data. Because one's "original landscape" is present in one's mind, it tends to be unobserved. Therefore, it is necessary to examine how to collect personal images or original landscapes from one's memory or association through observation. In this study, we conducted interviews (surveys) in combination with environmental autobiographies, spending about 90 minutes in an unstructured (semi-structured) interview and about 30 minutes in a semi-structured interview. The resulting research had two implications: 1) our interpretation of data (photos and comments) offered by participants enabled testing of Professor Kioka's hypothesis (mentioned previously), and 2) documentation of photos and comments provided opportunities to make community members sufficiently aware that their original landscapes had been shared locally. Thus far, we have tried to offer the collected data to a limited number of people representing local communities or to people who have personal connections with "weak ties" via magazines (e.g., Noshukufukei Saishu Shimpo has a small circulation and an Internet following at `http://scapebear.jimdo.com/`).

Indeed, people understand the individual nature of the personal landscape, but they are unclear about the root of this concept. Thus, this study attempts to create awareness of "shareability" within society, which has been overlooked. Collecting documents is not indicative merely of one's visibility of landscapes; rather, it reflects individual and aesthetic materials. The future challenge is to examine more accessible routes such as digital social networking site (SNSs) for the public.

In this study, we focused on a semi-open/semi-closed network such as Facebook (`https://www.facebook.com/scapebear`). At this SNS, information suppliers and users have a relationship. Through this digital SNS, we can expect that personal landscapes as individual aesthetic experiences will be shared loosely but definitely.

Keywords 'personal' landscapes, documentation, interview, community, shareablity

Aceh *Paleotsunami* Reconstruction for
Disaster Risk Reduction and Global Information

Nurjanah Jane (Tokyo Metropolitan University)
Hidenori Watanabe (Tokyo Metropolitan University)

Indonesia is a disaster-prone area. Indian Ocean Tsunami 2004 is one of the biggest catastrophe incidents over the last 100 years .The Indian Ocean tsunami contributed to the damaged of infrastructure, individual properties and environment in many coastal areas around the Indian Ocean. The tsunami was estimated to cause more than 200,000 dead casualties. Based on some reviews, the tsunami disaster in Aceh is not new and it has happened several times. Although the tsunami is not new and it had happened several times in Aceh, that disregards lessons from the past. Discontinuation of knowledge about earthquake and tsunami has led to information gap in the majority region of Aceh. Today, there are several ways to explore and to track tsunami record in the past. One of them is called as *paleotsunami. Paleotsunami* is a theory to track the record of tsunami events in the past, either through scientific or historical approaches. The track record of tsunami event in Aceh or Aceh *paleotsunami* has been traced since thousands years ago, i.e. the scientific research by means of tsunami deposit carbon dating has been recorded since prehistoric period around 5000 years ago. This method firstly develop systems analysis of tsunami deposit records from tsunami in the past, where the tsunami caused sea water flooded into the mainland, creating deposits typical horizontal, and white sand. The tsunami deposits is geological evidence that can be analyzed using carbon dating to determine the time tsunami events that occurred in the past. Coastal geomorphology records can also detect tsunami events in the past from the evolution of the beach to be an indication of the tsunami earlier. Another way to track *paleotsunami* by historical approach. The reconstruction of Aceh *paleotsunami* based on history approach could be performed using prose and manuscripts literatures. The history recording through ancient manuscripts since 1000 years ago have noted earthquake occurances that related to tsunami events in Aceh. By integrating and synchronizing *paleotsunami* Aceh into science and historical approach, the validity of Aceh tsunami records from 1000 years ago will be more precise. *Aceh Paleotsunami Digital Archive* is one of the ways to fill the gap of information from past disaster for global information.

Keywords *paleotsunami*, gap information, DRR, global information.

Reference

[1] Monecke, Katrin: *Coastal Progradation Patterns as a Potential Tool in Seismic Hazard Assessment,* Wellesley College: USA (2006).

[2] Candace, A., Grand, P., Benjamin P. Norton, Charles M. Rubin, Andrea D. Hawkes, Mudrik R. Daryono, Gary Rosenberg and Stephen J. Culver: "Stratigraphic Evidence for an Early Holocene Earthquake in Aceh, Indonesia," Elsever, Vol. 54, No. 142–151 (2009).

[3] Kesley, Ismail dkk.: *Great Earthquakes and Tsunamis of the Northwestern Aceh Coast Sumatra,* (2011).

[4] Ella, Meilianda: *Past, Present and Future Morphologycal Development of a Tsunami Affected Coast,* Netherlands (2009).

[5] Hermansyah: *Mitigasi Bencana dan Kearifan Lokal Aceh,* ADIC International Proceeding, Malaysia (2011).

[6] Krueng Raja and Sjeh Rih: *Geumpa di Atjeh,* 2(08)00236-X ANZYB LAMJONG, Lorong puloot 533/C, Banda Aceh (1964).

[7] Watanabe, Hidenori: *Aceh Tsunami Archive,* Industrial Art of Tokyo Metropolitan University, Japan (2013).

[8] Nurjanah, Nazli Ismail, Husaini Ibrahim, Hermansyah: *Aceh Paleotsunami for Disaster Risk Reduction,* E-Journal Program Pasca Sarjana Universitas Syiah Kuala, Banda Aceh (2013).

SESSION 2

A Visualization Method of Field Notes based on Locations and Topic Models

Yurina Takata (Tokyo Metropolitan University)
Hidenori Watanave (Tokyo Metropolitan University)
Masayuki Yanagisawa (Kyoto University)   *Taizo Yamada (University of Tokyo)*

1 Introduction

The data recorded by researchers at field research are archived and gathered up as field notes. We can only derive a result of field research by qualitative analysis of the data, because it is just possible to view these data in writing until now.

Therefore, this study is to visualize the field notes data to be able to find local features.

2 Methods

We developed a system to visualize the place to have been recorded and information of the place by mapping the text data, which is based on topic models created by text mining of field notes documents, at that point using Google Earth. This system was implemented as web contents using Google Earth API. We state details below.

2.1 Mapping Data

We classified the text into 6 categories—landscape, land use, hearing survey, location, summary, and the others. We created icons for each category, and mapped the text data at a point where we could guess from the text on Google Earth. If you click an icon, the text of that point is displayed. Then, the data has time information because the data was numbered each ID in order. So, it is possible to visualize the context.

2.2 Retrieval System by Topic Models

First, topic modelling is method to guess a latent topic of a text by a combination of appearance probability of a word. In this study, we set 30 topics and each color. The color was settled so that the higher the relation of between topics becomes, the more close the hue becomes. The words of the text were marked up into each color.

Next, we implemented a retrieval system to visualize topics on the map. When a word from a topic list or from a text is selected, circle pictures of its topic color are displayed on points including the same words. The retrieval system selecting a word from a text can be classified into 2 functions below.

- Where selected same word appears is displayed at the topic colors which the word belongs to at each place. It is possible to visualize how to use the same word at other places by this function because even same word is sometimes used as other topics at other places.

- Where selected same topic as well as word appears is displayed at the main topic colors at each place which can be judged from the context. It is possible to visualize what kind of other places the same word which has same meaning belongs to.

3 Results

By searching several words using this system, we could find a tendency of the colors. Topics of a same word at other places tended to have various hues by regions. On the other hand main Topics where same topic as well as word appears tended to be collected same hue by several regions.

In the future, we will find the meaning of several topics, and we want to research matters being able to read by using this system in more various patterns. So, we expect that this system is useful for understanding local features.

Keywords field notes, topic model, visualize, mapping, Google Earth

Language Processing Pipeline for Narrative Emergence: Digging into Human Rights Violations*

Ben Miller (Georgia State University)   *Jennifer Olive (Georgia State University)*
Ayush Shrestha (Georgia State University)   *Nicolas Subtirelu (Georgia State University)*
Jin Zhao (Georgia State University)   *Yanjun Zhao (Georgia State University)*

Narratives of survival begin, according to the Russian Formalists, with the basic elements of event, person, location, and time. Computationally enabling this argument by automatically identifying these elements within a narrative then automatically correlating their correspondences across a corpus of such stories can algorithmically elicit new narratives. These new narratives run transversely through a corpus and can stitch together stories spanning many hundreds to many tens of thousands of individual narratives. Emerging from the cross-document connections made among the narratological elements, these new narratives can reveal the stories of victims and perpetrators who had not previously offered their testimony of an event. Our paper presents a new language processing pipeline integrating various low-level natural language processing (NLP) tools for the recognition of entities, spatiality, and temporality; a machine learning component for fuzzy matching their outputs; and a data visualization component for cleaning the resulting correlations to explore higher-level humanities problematics.

Mieke Bal in *Narratology: Introduction to the Theory of Narrative*[1] terms the Formalist's essential components of narrative as *fabula*, which precedes and facilitates the organizational work of narrative, or *syuzhet*. NLP tools, when combined, provide the processes for encoding the various fabulaic elements for extraction and incorporation into other analyses. Although the *fabula/syuzhet* schema has been well criticized, it serves as a good foundation for testing the possibilities of algorithmically produced transversal narratives and experimenting in methods for cross-document co-reference.

Our paper will demonstrate this methodology, developed for "Digging into Human Rights Violations," with selected data from that project: corpora of heterogeneous documents related to the specific traumatic events surrounding the 9/11 attacks in New York, consisting of 511 first responder witness statements collected in the months following the attacks, and the longer historical trauma of apartheid in South Africa, consisting of documentation from a three-year Truth and Reconciliation Commission that heard from 22,000 witnesses and received 7,400 amnesty petitions in 12 languages. Our pipeline consists of paralleling existing modules, which reinforces the sequencing for the various textual markup tasks necessary to constitute a fabula. By parallel, we mean that multiple modules process the text for each fabular element. For example, both Stanford CoreNLP's SUTime and Brandeis' TARSQI will identify lexical indicators of temporality. The pipeline marks up language indicative of named and unnamed entities such as "Chief Ganci" or "The Lieutenant"; anaphora resolution of entities; classification of named and unnamed entities into traditional NLP categories of Person, Location, Organization, and Unknown; location entities for look up in external GIS resources and internal gazetteers; and absolute and relative temporal indicators such as "the morning of May 7" or "after that." Following multiple layers of parallel markup, machine learning algorithms chunk the marked text into fabulaic units of person at place at time. These units are visualized using techniques developed for this project, called StoryGraph and StoryGram, showing the movements of actors across the spatiotemporal geography of a corpus and providing a mechanism for the end-user to clean the data.

Reference

[1] Bal, Mieke: *Narratology: Introduction to the Theory of Narrative,* Trans. Christine van Boheemen, Buffalo, NY: U of Toronto P (1985). Print.

SESSION 3

Kenneth Rexroth's Syllabism Studied by Statistical and Text Analysis

Takeo Yamamoto (The University of Tokyo)

American poet Kenneth Rexroth (1905–1982) is known to have written syllabic poems in his middle and later years. However, it has not been clear when he started to write syllabic poetry, and the details of the development of his poetic style have not been studied. What led him to syllabism has never been clarified, either. Some have assumed that he adopted syllabism under the influence of Japanese poetry, as many of the imagists did. It has also been noted that he rarely kept to a single number of syllables per line throughout a poem. For this reason Rexroth's syllabism has sometimes been called "loose" but discussion of this aspect has progressed almost not at all.

In this study, I used a perl program to estimate the number of syllables in each verse (line) of all of his poems in The Complete Poems of Kenneth Rexroth, and another perl program to detect a potential accentual-syllabic meter in a poem. The results were then used, together with direct reading, to classify most of Rexroth's poems into three categories: syllabic, completely free, and accentual-syllabic. The above results together with the change in time of the most frequent values and the distribution of the number of syllables were used, together with textual analysis, in tracing the development of Rexroth's poetical style during his career of over sixty years. Similar computer-assisted methods have been used for quite some time to compare the styles of different poets and to scan a large number of accentual-syllabic poems, but to my knowledge such methods have not been used to analyze the development of the style of one poet who composed predominantly in syllabic meter over the course of his poetical career.

From the above analysis, the following results were obtained:

(1) *The Homestead Called Damascus*, which is Rexroth's first long poem and which he wrote mostly in his teens is, even this early, in syllabic meter. He experimented with more avant-garde styles of free verse twice in his lifetime, once in his youth (teens to about thirty) and later (in his late fifties), but both times came back to syllabism. This means that his commitment to syllabism was almost lifelong, and that Rexroth may be the link between the two generations of modern poets whose poetry demonstrates interest in syllabics: the imagists and the beat poets.

(2) The very first poem in *The Complete Poems of Kenneth Rexroth* is accentual-syllabic, but there are few other accentual-syllabic poems. Most of these are addressed to other poets who compose in traditional meter, whose styles Rexroth mimics in his poems.

(3) The direct influence of Japanese poetical style in his syllabic poetry is seen only in his late poetry (his late sixties to seventies.)

(4) Rexroth wrote only a few poems with an identical number of syllables per line of verse. This is rather remarkable for a poetical career largely spent in syllabism. The possible reasons for this are discussed.

Keywords text processing, statistics on syllable-per-line, poetical style

SESSION 3

SESSION 3

The Changing Appellations of "Japan" in Russian Magazine *Rubesh* in Harbin

Mao Sugiyama (Osaka University)

This presentation aims to illustrate how Russian people looked at Japan, by investigating the Russian magazine *Rubesh* (Рубеж), published during the period 1926–1944 in Harbin, Manchukuo. As Sawada (2005) points out, *Rubesh* would refer to Japan as two different appellations, "Iaponiia" (Япония) and "Nippon" (Ниппон), but Sawada does not mention any details about how they were used. The occurrences of the words, however, seem to have some patterns. Thus it is worth looking at how they were distinguished in *Rubesh* in order to elucidate how the Russians saw Japan those days.

At that time, a large number of population migrated to Harbin from the Czarist Rus- sia. One of the main reasons of the migration was to escape the Russian Revolution in 1917. The refugees were called "White Russians". After Japan built Manchukuo in China in 1932, the Japanese government gradually controlled not only Chinese in- habitants but also the White Russians. Particularly, the Japanese government took advantage of their anti-Soviet sentiment. Since publications in Manchukuo were, under strict control of the overseeing government, editorials or articles raising any criticism of Japan were hardly found in *Rubesh*.

The two appellations of Japan in Russian, "Iaponiia" and "Nippon", evoke very different images. This is supported from a Google and a corpus search. Let me show an example of the corpus search: The national corpus of Russian language shows a stark contrast with the use of the two appellations. "Iaponiia" is related to military contents, nationality and just as the calling of the country, while "Nippon" is used almost aways in the string "Dai-Nippon" (Empire of Japan). This distinction should have arisen since "Nippon" in Russian is a borrowing word from the Japanese language. Though these results are based on contemporary linguistic use, it should not be surprising to say that "Nippon" in Russian is related to military contend, whereas "Iaponiia" does not have such a connotation from the perspective of a "digital humanistic" study.

In 1934 the NHK (Japan Broadcasting Corp.) decided to change the appellation of Japan from "Nihon" to "Nippon" . Moreover prints on stamps used in Japan were also changed from "Japan" to "Nippon". The change of Japanese appellations in *Rubesh* might have been triggered by the change of policy that the Japanese gorvenment made, namely, the change of appellations from "Nihon" to "Nippon". In Russian magazine *Rubesh*, the words "Iaponiia" and "Nippon" co-occur with words of positive senti- ment. A close examination of articles in *Rubesh* shows us how Japanese government manipulated Russian to have a particular image of Japan, which is quite different from what Russians have today in terms of concreteness of words evoked in "Iaponiia" and "Nippon".

In order to investigate the distinction in the use of the words, I went through all articles that include either or both words. By using a digitized corpus of *Rubesh*, I will illustrate how spectacularly the usage pattern of the appellations of "Japan" shifts by means of "the paragraph-analysis method" which I used in this study.

Scope of Cultural Resources Studies: Text-Mining of a Newly Created Interdisciplinary Graduate Program with MIMA Search

Yusuke Nakamura (University of Tokyo)　　*Hideki Mima (University of Tokyo)*
Katsuya Masuda (University of Tokyo)　　*Chikahiko Suzuki (University of Tokyo)*

How do interdisciplinary research/education programs develop to form their own style and find their own niche in the midst of a cluster of various specialized competing disciplines in higher education institutions? In this short paper, we report the preliminary results of mining of texts related to a new graduate program created in the Graduate School of Humanities and Sociology, University of Tokyo in 2000. For the text mining and visualizations, we used MIMA (Mining Information for Management and Acquisition) Search, a search engine developed by Mima for extracting and visualizing relationships among texts using Natural Language Processing.

The program is named "Cultural Resources Studies" (hereafter, CRS), and characterized by humanities-based interdisciplinarity and the linkage with practitioners of cultural activities. As the main target of the analysis, we focus on the creation of M.A. theses for the following reasons. First, M.A. students tend to pursue the themes not covered by other disciplines (one major reason why they chose the CRS). The seminars offered also tend to cover a variety of topics in ways that are more experimental as compared to traditional humanities curricula. Second, many M.A. students are writing a truly academic paper for the first time, and yet there has not been an established model to follow as a CRS M.A. thesis. Thus, when examined as a group, the M.A. theses of the first decade of the CRS program reveal us an interesting case of collective efforts to create a new form of interdisciplinary studies from scratch.

We have already analyzed two categories of texts related to the CRS program since its foundation: (1) syllabi of lectures and seminars (a total of about 600), (2) summaries (less than 40,000 characters) attached to M.A. theses (a total of 76). The cluster analysis of syllabi and thesis summaries reveals the emergence of a few stable clusters of related terms that can be represented by the high-ranking terms such as "management" and "exhibition." However, for the rest they are so diverse that it is difficult to choose a particular term as a representative label. We interpret this result as a delicate balance between an attractive force toward practical applications and a repulsive force toward a search for diverse cultural resources hitherto understudied.

Now we are extending our analysis to the institutions referred to in the M.A. theses. There are a total of about 2,500 institutions in our database, from universities to private companies. We have selected 257 institutions that (i) have issued periodicals, (ii) are referred to twice or more, and (iii) have web pages. We are analyzing their attributes and basic documents (such as mission statements and manifestos) to study what kinds of institutions our M.A. students have considered as the key organizations for the study of cultural resources in modern Japan.

In the presentation, we will show the results of analysis of the up-to-date data to examine the present scope of the CRS program, and discuss the methods for monitoring the scope of ongoing interdisciplinary research/education programs.

Keywords interdisciplinary research/education program, text mining, visualization

SESSION 3

Digitization of a Catalogue of Oracle Bones

Tomohiko Morioka (Kyoto University)

This report describes a digitization of "Catalogue of the Oracle Bones in the Kyoto University Research Institute for Humanistic Studies"（京都大學人文科學研究所所藏甲骨文字; ZOB）. Oracle Bone script is an ancient Chinese script. Oracle Bone characters were engraved on animal bones or turtle shells (in this report, "Oracle Bone" is used as a general term for these materials). ZOB is a catalogue of Oracle Bone collection in Institute for Research in Humanities, Kyoto University. This catalogue consists of "PLATES PART 1"（圖版册 上; published in 1959）, "PLATES PART 2"（圖版册 下; published in 1959）, "TEXT" （本文篇; published in 1960） and "INDEX"（索引; published in 1968）. ZOB is an early work of Oracle Bone studies, therefore some interpretations may be obsoleted and it lacks recent results of studies. On the other hand, the Oracle Bones collection is a rare heritage, so the PLATE and INDEX parts are usable resources. In addition, information of old work may have academic value in the history of Oracle Bone studies.

At the first step of the digitization, we scanned every page of ZOB, and made a Web service to view these pages. Then we extracted images of rubbings and photos of Oracle Bone pieces from the PLATES parts. Each file name of extracted rubbing image is "`rubbings/<ID>(.<ext>)`" (e.g. `rubbings/B1234.png`). `<ID>` consists of `<MATERIAL-TYPE>` and `<NUMBER>`. `<MATERIAL-TYPE>` is a symbol to indicate material of Oracle Bone: "B"= animal bone; "S"= turtle shell. Likewise, each file name of extracted photo image is "`photos/<ID>(.<ext>)`" (e.g. `photos/S0123.tif`).

We also digitized the INDEX part as an Oracle Bone character database. Currently, Oracle Bone characters are not included in Unicode, so we use CHISE technology[1] to represent them. We defined these character features to represent information of the INDEX part:

`zinbun-oracle-page`	page number of the INDEX part
`=zinbun-oracle`	glyph ID of Oracle Bone character
`<-denotational`	link for abstract Oracle Bone character
`shuowen-radical`	number of shuowen radical（説文部首）
`<-Oracle-Bones`	corresponding modern Chinese characters
`sources`	identifiers of Oracle Bones

We also added character feature '`ideographic-structure`' to represent visible structure of character. It is the same format to represent IDS (Ideographic Description Sequence)[2] in CHISE. In original IDS, each component must be modern Chinese characters included in UCS, however our extended IDS accepts Oracle Bone characters (or modern Chinese characters not included in UCS) as components. This feature is basically not depended on interpretation of Oracle Bone character. It is usable to search Oracle Bone characters.

This Oracle Bone character database is integrated with the CHISE character ontology. The source code of it is available at [3] as a part of XEmacs CHISE. Likewise for modern Chinese characters (or other various characters), information of Oracle Bone character can be viewed by CHISE-wiki, such as: `http://www.chise.org/chisewiki/view.cgi?character=rep.zinbun-oracle:339`

In CHISE-wiki, character feature '`sources`' works as links for Oracle Bone images (rubbings) which include the displayed Oracle Bone character. Likewise, character feature '`zinbun-oracle-page`' works as a link for a page of the INDEX. "CHISE IDS Find"[4] is also available to search Oracle Bone characters by their components. In addition, each Oracle Bone character which has

character feature '`<-Oracle-Bones`' is linked from the corresponding modern Chinese character (e.g. `http://www.chise.org/chisewiki/view.cgi?character=`吉).

Keywords Oracle Bones, ancient Chinese characters, character database, rubbings, Linked Open Data

Reference

[1] Morioka, Tomohiko: "CHISE: Character Processing based on Character Ontology", Large-scale Knowledge Resources (LKR 2008), pp. 148–162, LNAI 4938.

[2] Defined in ISO/IEC 10646 (UCS)

[3] `<http://git.chise.org/gitweb/?p=chise/xemacs-chise.git;a=blob;f=lisp/utf-2000/Oracle-Bones.el>`

[4] `<http://www.chise.org/ids-find>`

SESSION 4

Toward to the Definition of Safe Character Set of Nushu in ISO/IEC 10646

Toshiya Suzuki (Hiroshima University)

Nushu (Chinese word meaning "Women Writing") is a writing system which was used by women in area. Although its origin is still unidentified because the reliable historical evidence is limited, widely accepted assumption is that the most part of remaining Nushu is derived from Hanzi. But the identities of the source Hanzi (shape, pronunciation and meaning) are remarkably simplified; single Nushu character can be used for different Hanzi with different meanings, as far as their phonetic values are similar. Sometimes same character could be used for different phonetic values. From the survey of the dialectic sound system in the area, the number of the distinctive phonetic values for Nushu users is estimated to be from 800 to 1100, but the number of the characters that an author could use distinctively is estimated to be from 500 to 1000. Thus, Nushu could be understood as a transitional script evolving from the ideographic script to the syllablic script. The typical usage of Nushu was the one-to-one communication between the adopted sisters, and the one-to-many communication (like epigraphis, sutras, open letters, newspapers etc) via Nushu is rare. As a result, the identifications of Nushu characters have not been stabilized yet. Furthermore, the syllablicalization of Nushu characters is not uniform; some authors find a semantic and phonetic differences in a pair, others find no difference and take the pair as an interchangeable variant. As a result, there is no character list for the elementary study (like "Thousand Character Classic"（千字文） or "Cangjie Wordbook"（倉頡篇） for Hanzi). There are several Nushu-Chinese dictionaries compiled for the scholars, but the identifications of Nushu characters are quite varied.

Since 2003, a discussion to include Nushu in ISO/IEC 10646 (the basis of Unicode) has begun; the proposal is primarily submitted by Professor Zhao Liming from China. However, the ballot to fix the scripts to be included in ISO/IEC 10646:2012 /Amd.1:2014 decided to postpone Nushu to the next amendment, because several technical questions are raised to the proposed character set. The important problem is that most existing Nushu dictionaries have their own mutually-incompatible collation methods and the number of indexing characters are quite varied too, therefore it is difficult to compile a stable Nushu character set by combining the existing dictionaries; according to the research by Sun Qi (2005), the number of the indexing characters in 8 Nushu dictionaries published in 1986–2002 are varied from 470 to 1800. In addition, the combinations of the representative glyphs and its descriptions are often inverted for the similar-but-different characters (e.g. in a dictionary, glyph A and meaning X, glyph B and meaning Y are listed, but another dictionary lists glyph A with meaning Y). Because the machinery computation of the cross sections of the existing Nushu dictionaries is impossible, the standardization of Nushu for ISO/IEC 10646 have to start from the definition how Nushu characters in the Unicode should be identified.

Zhao tried to exclude the confused relationship between the glyph and its usage by making the per-author statistics (2006). But the glyph distinction rule is not clarified, and the proposed charset mixes the statistic results without the normalization. As a result, the definition of the entity to be enumerated is still unclear. In this report, the detailed discussion in ISO/IEC JTC1/SC2/WG2 (the working group for the development of ISO/IEC 10646) and the future tasks are summarized.

Keywords Nushu, glyph, character set, character encoding, Unicode

Textual Encoding for Government-Designated Textbooks (*Kokutei Tokuhon*)

Akihito Kawase (National Institute for Japanese Language and Linguistics)
Toshinobu Ogiso (National Institute for Japanese Language and Linguistics)

This study examines possibilities for converting several textbooks to XML format, which will provide fundamental documents for analyzing reading books written in the late Meiji period to the early Showa period. The documents selected for the analysis are the *Kokutei Tokuhon*, government-designated textbooks that were used in elementary schools through the late 1940s.

The *Kokutei Tokuhon* are the textbooks which the Japanese government (Ministry of Education) designated for elementary school national language education, and have been repeatedly published in revised and altered versions since the first edition in 1903 (Meiji-36) to the last sixth edition in 1947 (Showa-22).

It is suspected that the alterations of editions mainly reflect social influences and cultural awareness with the unique and remarkable characteristics as described below: (Edition 1) sudden change of social life after the victory in the First Sino-Japanese War, (Edition 2) the change in social conditions and the spread of nationalism after the victory in the Russo-Japanese War, (Edition 3) the development of the education campaign and the child-centered educational philosophy after World War I, (Edition 4) nationalism after the Manchurian Incident, (Edition 5) strain under the Pacific War, (Edition 6) reformation of the educational system with the establishment of the Fundamental Law of Education after World War II.

Therefore, describing the *Kokutei Tokuhon* in a machine-readable manner has profound significance, not only for providing resources for studying the fundamental lifestyle changes during the social modernization through the Meiji, Taisho, and Showa eras, but also for providing indispensable documents to clarify the flow and tendencies of national language education. It is expected that the output of this study will create a fundamental resource for supporting studies on the background of the establishment of the modern Japanese language.

Furthermore, since the National Institute for Japanese Language and Linguistics (NINJAL) has been conducting digitization and morphological analysis of Japanese classics, these studies will contribute to capturing the features of diachronic changes in the Japanese language by comparing morphological data from the *Kokutei Tokuhon* to those previous studies from a corpus-linguistic viewpoint.

We specifically describe the *Kokutei Tokuhon* in detail in the following manner:

1. The fundamental text structure (elements from `<text>` to `<s>` level)

In the case of *Kokutei Tokuhon*, the number of texts and publication content vary based on the edition, but most of the structure is common. The fundamental text structure was common across different editions, although publication numbers and content varied. Here, we interpret each textbook as the smallest unit of the file, and convert total 71 textbooks with `<text>` elements. The structure of each `<text>` element is a combination of three parts: the front matter, the body part that contains text with an illustration, and the back matter.

Since this structure is well aligned with the composition of an orthodox Western manuscript, we can mark up most of the elements in reference to TEI P5: Guidelines. Table 1 shows the list of elements used to mark up the textbook.

SESSION 4

Element	Role	Element	Role
`<text>`	whole text	`<q>`	quote
`<front>`	front matter	`<s>`	sentence unit
`<body>`	body matter	`<rs>`	referencing string
`<back>`	back matter	`<pb>`	page preak
`<div>`	divisions	`<cb>`	column break
`<head>`	title	`<lb>`	line break
`<p>`	paragraph	`#PCDATA`	character data

Table 1: List of Elements With Explanations of Their Roles in Marking up the *Kokutei Tokuhon*

2. Elements from `<s>` to `#PCDATA` level

The following two issues remain in the process of encoding the *Kokutei Tokuhon*: (A) structuring ruby annotations and (B) editing prolonged sound marks, a Japanese symbol which indicates a long vowel (e.g. the symbol "ー" in おとーさん, which is pronounced otoh-san).

If we give priority to morphological analysis or constructing a dictionary for modern Japanese, revising the original text (below the sentence level) with the TEI element `<seg>` (arbitrary segment), representing any segmentation of text below the 'chunk' level might be one of the solutions to issues (A) and (B).

However, to conduct string matching or structural comparisons based on the original text, the corrected text (the optimal characters for morphological analysis) and the original appearance information must be preserved. Throughout the study, we would like to have constructive discussions and come up with practical suggestions for resolving these issues.

Keywords *Kokutei Tokuhon*, modern Japanese, XML

SESSION 4

NeCTAR Virtual Laboratories and the Humanities Networked Infrastructure (HuNI) Project

Paul Arthur (University of Western Sydney)   *Deb Verhoeven (Deakin University)*

This paper discusses the complexities and challenges of aggregating data in one of the most ambitious multidisciplinary 'virtual laboratory' projects yet to be undertaken in digital humanities worldwide. The Humanities Networked Infrastructure Project (HuNI) is one of the first large-scale eResearch infrastructure projects for the humanities in Australia, funded through the government's National eResearch Collaboration Tools and Resources (NeCTAR) scheme. The broad goal of the virtual laboratories program is goal is to create interoperable databases and tools that utilise high-speed data networks, cloud and super-computing. Initially using a Linked Open Data approach, HuNI set out to align and disambiguate information originally stored across 28 national cultural datasets, including the Australian Dictionary of Biography, the Auslit database and the biographical Dictionary of Australian Artists, making them discoverable through a single interface. The datasets brought together comprise more than 2 million authoritative records relating to the people, objects and events that make up the country's rich heritage. This paper also comments more widely on the implications of such large infrastructure projects for research across disciplines.

The National eResearch Collaboration Tools and Resources project (NeCTAR) is a $47 million Australian Government, Super Science project, financed by the Education Investment Fund. The University of Melbourne is the lead agent appointed by the Commonwealth Government. The Australian Research sector has committed $54 million as co-investment to the NeCTAR project, resulting in NeCTAR injecting $101 million to Australia's research infrastructure. The presenter is a member of the NeCTAR Project Board, and was the Australian National University lead researcher for the HuNI project in 2012–13.

Keywords e-research, interoperable databases, cultural datasets

SESSION 5

SMART-GS Web: A HTML5-Powered, Collaborative Manuscript Transcription Platform

Yuta Hashimoto (Kyoto University)

Some historically important manuscripts, especially those written in the modern age, are hard to read due to their authors' unclear handwriting. Transcription processes for these manuscripts tend to be more time-consuming, eventually decreasing historians' productivity. When manuscripts are written in East-Asian languages such as Japanese, which have a vast number of characters, transcriptions are even harder.

SMART-GS, a desktop application for image-based historical studies, has been developed by Japanese historians and developers since 2006 to help historians work on such manuscripts. The system supports a variety of features to help historians work with illegible manuscripts such as image-and-text markup, search for handwritten text, and perform other research tasks. SMART-GS has been successfully applied to several historical research projects, including the transcription project of Baron Yuzaburo Kuratomi's diary. However, due to its lack of a web interface, its actual use has been limited to somewhat small circles. SMART-GS Web, a web version of SMART-GS has been developed to offer SMART-GS' various features to a wider community of users (Figure 1).

With the development and diffusion of HTML5 technologies and modern web browsers such as Firefox and Google Chrome, we can implement rich features such as image editing that were previously only possible on desktop applications. In particular, newly introduced protocols like WebSocket and WebRTC enable users to communicate with each other in real-time. SMART-GS Web makes use of these technologies and offers following new features that the original SMART-GS doesn't have:

Real-time Collaboration

Transcriptions of huge amount of manuscripts are often done by teamwork. SMARTGS Web offers groupware features for this purpose. Every change made by each user will be immediately reflected in other users' workspaces without page reloading. In addition, the mouse cursor and text cursor of each member will be shared with other users in real-time so that they can see what the others are currently focusing. It's also easy to share resources on SMART-GS Web among users: manuscript images imported into a project will be stored in Amazon S3 storage and made available to every project member. Metadata added to images as well as transcription texts will be similarly stored in cloud storage and made searchable by an indexing server.

TEI Support

SMART-GS Web has an embedded HTML editor for transcription text. But it's also possible to export these transcriptions to TEI documents. Markups on images will include TEI's `<sourceDoc>` elements. SMART-GS Web's TEI support is not yet sophisticated, so I am planning further development.

Vertical Text Editing

The editor embedded in SMART-GS Web supports vertical text editing, which is enabled by the writing-mode property introduced in CSS3. This feature will be especially useful when transcribing manuscripts written in East-Asian languages such as Japanese.

SMART-GS Web is not yet a mature project, and needs much more improvement. However it can make significant contributions to historians who work on historical manuscripts.
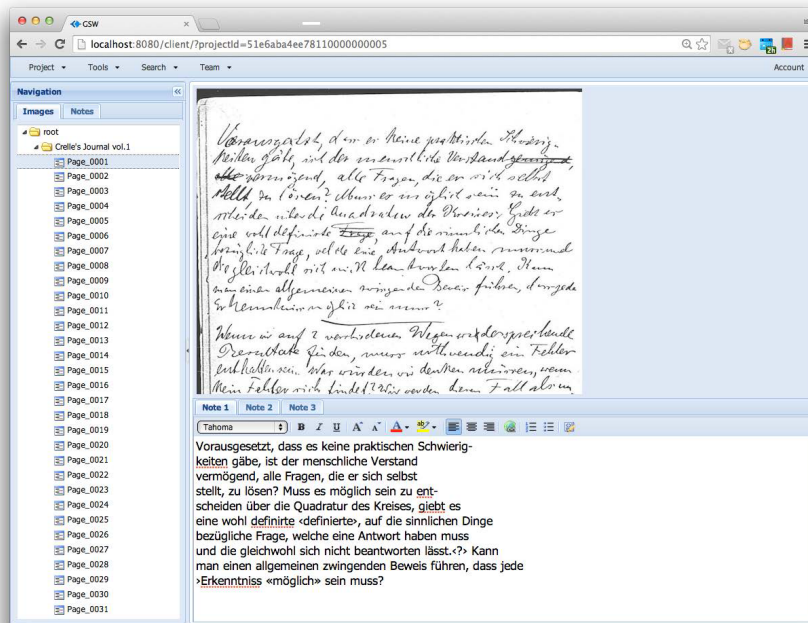
Figure 1: SMART-GS Web running on Google Chrome

Keywords historical research, transcription, TEI, collaboration

Participatory Agency and Technological Literacy via Museum Collection:
The Case of Peter Mitterhofer Museum

Ya-Ju Yeh (Aletheia University)

One of the greatest achievements for the world of information technology was undoubtedly the invention of the typewriter. After their invention in the 1860s, typewriters quickly became indispensable tools widely used by professional writers in private homes or for business correspondence in offices. As the very first pioneer of writing technology, the typewriter challenges human's cognition and practice of machinery and launches a new age of technological culture. From the nineteenth century onwards, this machine exerts profound influences upon traditional ways of personal writing as well as social and cultural transformations. The typewriter has been popular for over one hundred years, yet by the end of the 1980s, personal computers had largely displaced typewriters. The start of the computer era marks the end of the mechanical typewriter era, and typewriters now become museum artifacts of significance or individual collection for preservation or display other than everyday practical objects.

Peter Mitterhofer, a carpenter from Austria, developed several models and a fully functioning prototype typewriter in 1867. Peter Mitterhofer Museum has gathered a wide ranging and formidable collection of typewriters dedicated to him since 1998. The museum brings visitors or participants into contact with the culture of writing in general and with the various kinds of typewriters. The museum, aside from categorized showcases of exhibitions, provides inspiring forms of demonstration, such as dioramas, virtual tours and video clips to present the historical development of the typewriter. Interactive devices and thought provoking programs are also used in order to reinforce visitors' memory and experiences of remarkable inventions. This paper aims to explore participatory approaches inspired by museum collection and emphasize visitors' sensory responses to typing devices with which human has inaugurated mechanical writing. The central aspects are to be further discussed: what are features of human's technological literacy derived from the evolutionary design of the typewriter? What are effects of visitors' sensory interaction with museum objects? Given material perspectives of collecting and interacting, this paper reasserts the significance of participants' agency and technological literacy with the case of Peter Mitterhofer Museum.

Keywords typewriter, museum, agency, technological literacy, Peter Mitterhofer

A Multilingual Digital Humanities Project for Asia

Nirmala Menon (Indian Institute of Technology Indore)

This paper will look at the borders of languages and disciplines and re-imagine Digital Humanities for a multilingual Asian context from the point of publishing and translations.

Both Digital Humanities and Publishing will have to be radically reconceptualized in multiple languages for its success. I believe that Postcolonial DH should consciously facilitate projects in multiple postcolonial languages so as to preempt any canonical/linguistic hierarchy of languages. There is exciting potential to be able to do just that with the aid of technology that facilitates ventures such as publishing and translating with a lot more ease than ever before. The said tools are still nascent and often unsophisticated (especially the ones for translations) but that is precisely the challenging space of critical evaluation that calls for the Humanities in Digital Humanities. In other words, it is not about simply reengineering computing excellence but a conscious attempt at reimagining postcolonial publishing paradigms. This paper examines the critical work being done in various language literatures in India and its non-representation in postcolonial and other theoretical frameworks. I argue that a Digital Humanities project that provides a platform for conversations between these various critical works in literature will enrich and deepen our understanding of postcolonial conditions beyond the Anglophone writings. My project focuses on one aspect of this complexity—the huge amount of research and criticism being done in different language literatures in India. I envision a digital database that: 1) compels visibility and conversation between scholars of these different languages and 2) enable translations between the languages and into English of academic publishing in various languages. This will require a two step strategy: 1) Develop a comprehensive collaborative project that will provide the infrastructure for a multilingual digital project and 2) expand the project to include translations from and between the various discourses. While I envisage this project to begin with academic works in the discipline of literature, I am convinced it will be important for the sciences and technology publishing too. It will truly be an interdisciplinary project envisioned and initiated by the Humanities but that crosses the borders of liminality across languages.

Keywords digital publishing, multilingual scholarship, literature, translation, database development

SESSION 5

Sōseki's Worlds: A Digital Representation

Dawn Lawson (New York University)

The goal of this project is to create a robust digital environment centered on the life and works of Natsume Sōseki that will further research and teaching about him throughout the world. In April 2014, in part to coincide with the one hundredth anniversary of the serialization of Kokoro, a symposium was held in the United States at the University of Michigan, Ann Arbor. Entitled Sōseki's Diversity, this gathering of some 150 scholars from three continents featured nine panels over three days and keynote addresses by John Nathan and Tawada Yōko. Plans for a follow-up conference in 2015 are under way. Our project will support and reinforce the widespread interest in these gatherings and serve as a virtual home for Sōseki students and scholars everywhere.

At JADH 2014 we will be presenting about the first stage of the project, which will consist of the creation of a bibliography of Sōseki works in English translation and the implementation of a tool for the side-by-side reading of his works in the original and English translation. (We look forward to expanding to translations in other languages as we attract collaborators who can assist us with them.) Zotero was used for the initial data entry of the bibliography. The comparison tool under consideration is called the Versioning Machine (see v-machine.org), which is a framework and an interface used to display multiple versions of text encoded according to the Text Encoding Initiative (TEI) Guidelines. Using the Versioning Machine poses numerous challenges. Even if the texts we use are already encoded in TEI—and not all of them will be—the Versioning Machine requires the use of the critical apparatus tag set (tei.textcrit), which is unlikely to have been applied to those texts. In addition, we are not aware of a use of the Versioning Machine with Japanese text to date, and although we don't anticipate any particular problems, something might arise. We hope to have proof of concept by the time of this conference. As for evaluation of the project when it is mature, the number of visitors to the site and number of downloads (if applicable) will be potential assessment tools, and we will also conduct a survey about the usefulness of the materials in teaching and research.

This project will be a collaboration involving both librarians and scholars, including (among many others) myself, J. Keith Vincent of Boston University, Alan Tansman of the University of California at Berkeley, and Molly Des Jardin of the University of Pennsylvania. In addition to the bibliography and text comparison components mentioned above, Sōseki's Worlds might include geographical visualizations of the travels of both the author and his characters; a timeline; and other creative uses of technology to enhance both pedagogical and research use of his writings.

Sōseki and his works deserve as diverse and complex a digital environment as our technologies and imaginations can construct. Careful planning at the early stages, widespread participation and collaboration, and close cooperation between librarians and teaching faculty will make that possible.

Keywords Natsume Soseki, Soseki's Diversity, DH

SESSION 5

Quantitative Analysis of Books about How to Write Narrative Text: Extracting Characteristics of Screenwriting, Playwriting, and Fiction Writing

Ryoichi Takahashi (Tokyo Institute of Technology)
Hajime Murai (Tokyo Institute of Technology)
Takehiro Inohara (Tokyo Institute of Technology)

The purpose of this research is to build a basis for the scientific analysis of how to write narrative text. Utilizing quantitative analysis of books about how to write narrative text written by skillful professional writers, we attempt to investigate the similarities and differences between screenwriting, playwriting, and fiction writing.

Research in the fields of psychology and cognitive science explores the creative narrative writing process as a manifestation of creativity (Kaufman & Kaufman, 2009). However, with regard to writing strategies used by skillful professional writers, much of the research has taken the form of case studies and only a little bit of it has been sample-based (quantitative), because collecting a lot of psychological data from skillful professional writers is difficult. Moreover, little is known about differences in writing strategies between narrative genres. According to expertise theory, an expert's knowledge is not domain general, but domain specific. Hence, it can be assumed that skillful writers' knowledge differs between narrative genres. However, most studies have not focused on this difference.

In this study, to attempt to understand writers' strategies and knowledge, we analyze books about how to write narrative text written by skillful professional authors. "Books about how to write narrative text" here implies reference works on how to write a screenplay, playscript, or novel. These books do not directly reflect the cognition undergone by the writers during their actual production process. However, they are nevertheless the most suitable form of data for investigating writers' strategies key ideas in narrative writing, since their content reflects the matters or aspects of the narrative writing process to which the writers attach importance. Applying data mining techniques to these books, we enumerate their content.

We analyze thirteen screenwriting books, ten playwriting books, and ten fiction-writing books, all selected based on their place in Amazon.com's bestseller rankings, the top 100 in each category. All the books were written in English and published in the United States, since it can be assumed that American books about how to write narrative text are of high quality given the commercial success of American narrative content (for example, Hollywood movies, Broadway plays, and popular American fiction). We convert books to plain text data and then use Tree-Tagger (Schmid, 1994) to assign part of speech and lemma tags with the corpus.

Analysis 1: Using the Network Centrality Analysis function of by NetworkX (Hagberg, Swart, & Chult, 2008), we mathematically identify key nouns referring to concepts related to writing narrative text. Six common key concepts overall genre—character, story, scene, time, action, and audience/reader—are extracted as a result.

Analysis 2: Employing Dependency Analysis using MaltParser (Nivre, 2007), we analyze verbs closely related to the six key concepts (nouns). The results show that verbs that have a dependency relation to key concepts differ by the narrative genre under discussion. This result suggests that screenwriters, playwrights, and novelists approach these key concepts from different points of view.

Analysis 3: We build manually a sentiment lexicon based on the model of Plutchik's (1980) emotional categories. Extracting words expressing sentiments from the corpus, we count word frequencies for each category and analyze them using a chi-squared test. It is found that the frequency of sentiment words

differs significantly by genre. This result suggests that affective characteristics qualifying screenwriting, playwriting, and fiction writing differ across these genres.

Keywords quantitative analysis, creative writing, screenplay, playscript, novel.

Reference

[1] Hagberg, A., Swart, P. and S Chult, D.: "Exploring Network Structure, Dynamics, and Function using Network," *Proceedings of the 7th Python in Science conference,* (2008).

[2] Kaufman, S. B. and Kaufman, J. C. (eds.): *The Psychology of Creative Writing,* Cambridge University Press (2009).

[3] Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kubler, S., Marinov, S. and Marsi, E.: "Malt-Parser: A Language-Independent System for Data-Driven Dependency Parsing," *Natural Language Engineering,* Vol. 13, pp. 95–135 (2007).

[4] Plutchik, R.: *Emotion: A Psychoevolutionary Synthesis,* New York: Harper and Row (1980).

[5] Schmid, H.: "Probabilistic Part-of-Speech Tagging using Decision Trees," *Proceedings of the International Conference on New Methods in Language Processing,* (1994).

POSTER

Construction of 21st Century Archives for the Research Foundation of the Library and Information Professions and Education: Towards a Methodological Synthesis of Library and Information Science and Archival Science

Norihiko Uda (University of Tsukuba)   *Yuko Yoshida (University of Tsukuba)*
Tetsuya Shirai (University of Tsukuba)

We launched a project aimed to organize and create digital archives for library and information professions that will contribute to both research and education. The archive will comprise historical materials owned by the University of Tsukuba's Graduate School of Library, Information and Media Studies, which shaped the history of the education of library and information professions for most of the 20th century.

This project includes the formation of a comprehensive system for collecting and conserving materials and the implementation of a system for their utilization. The aims of this project are twofold: (1) to examine a methodological synthesis of library and information science and archival science and (2) to re-examine the history of the library and information professions and education.

Major users of the archives are library and information science researchers specialized in library and information professions and the history of librarianship. This system will establish a multilingual interface for international researchers who are interested in comparing professional development models of education for librarianship. Museum materials will be used for educational content on the history of media technology.

There are no precedents for a comprehensive archival system for the education of library and information professions that includes archival documents and museum documents. This archive thus has the potential to serve as an international model for knowledge-sharing on library and information science.

Six subordinate cross-disciplinary studies are in progress under the project. They comprise (1) A detailed examination of the collection and oral histories investigation, (2) Information architecture for digital archives, (3) Historical research on library and information professions and education, using archival materials, (4) Modeling, implementation and evaluation of the archive, (5) Modeling of the exhibition space, and (6) Analysis of the research project (museums, libraries and archives collaboration). This poster presentation focuses on studies in progress on (2) Information Architecture for digital archives and (3) Historical research on library and information professions and education, using archival materials.

The first study reconsidered "digital archives" to drive forward the project entitled "21st Century Archives for the Research Foundation of the Library and Information Professions and Education." As a result, digital archives are defined as mapping of the original archives, holding to the principles of provenance, respect for the original order, preservation of the original forms of the archives, and permanent preservation. A real digital archive must be constructed based on an information architecture methodology.

Information architecture is the structural design of shared information environments; the art and science of organizing and labeling websites, intranets, online communities and software to support usability and findability. It involves analysis of information context and organization, labeling, navigation and retrieval using a consistent design philosophy. Information design prioritizes how information is presented on a website. Information architecture, on the other hand, is a website construction method that focuses on how to organize information elements. This study aims to achieve faithful mapping between the context of the original resources and electronic resources through the application of information architecture

POSTER

methodology to the establishment of digital archives.

In the second study, we adopted the methodology of digital humanities as our analytical scheme to clarify the historical development, current conditions, and future vision of the library and information professions and education. The materials in our archive collections for historical research of librarianship are periodicals, annual reports, minutes of meetings, syllabuses, business diaries, photographs, audio materials, video materials and films. They include unpublished and handwritten materials. Whatever their form, these materials are an aggregation of the memory of the philosophy of an institution and records of past activities. Previous studies have not included these materials due to their uncatalogued nature and often poor state of preservation.

We used two illustrative examples that utilize digital archive collections to demonstrate the progress of research. The first attempt was to trace the history of library and information professions focusing on changes in nominal designations of librarians/information professionals. The other was a consideration of the gender issue concerning female librarians, using digital images from the archive. These primary investigations have suggested the potential for reconstructing a discourse on the library and information professions and education.

Keywords library and information science, archival science.

POSTER

A Case Study of a Contemporary Fiction Writer's Revision Process for Creative Writing

Akira Kudo (University of Tokyo)   *Takeshi Okada (University of Tokyo)*
Dominick Chen (Dividual Inc.)

To date, there has been little attention to the creative processes of expert writers, such as fiction writers, poets, and script writers, because creative writing has generally been seen as shaped by practice and education. In general, previous studies of writing have used the methodology of creative experiments and protocol analysis, although there have been few studies of the authentic writing situation due to difficulties with the approach. Moreover, there have been insufficient investigations into the characteristics of writing and revision using word processors that are generally ubiquitous. Although traditional literary research have tracked handwritten additions and deletions to manuscripts for insights into the creative writing process, as Yoshida (2007) points out, with contemporary writers using word processors, prior versions of a manuscript are generally not available.

The purpose of this case study is to describe the revision processes of a contemporary expert fiction writer from a cognitive science perspective. More specifically, we utilize Type Trace which is a text editor devised by Dividual Inc. as an analysis tool for observing the writing process（Chen, 2013）. Type Trace makes it possible to record and reproduce the processes involved in the usual inputting of text and preparing drafts. This study focuses on a Type Trace record for "Maijo Novel Powder Snowstorm" (2007), written by Otaro Maijo, who has been a writer of fiction covering many genres since 2001. In particular recognition of his achievement in literature, he was awarded the Mishima Prize for "Ashura Girl" (2003). More importantly, it's also unusual that Maijo is an anonymous author who has not revealed a private information before, therefore it can be saying that selecting Maijo as a target is extremely significant in 'research for fiction writer'.

First, revisions are categorized as either additions, deletions, substitutions, permutations, distributions, or consolidations based on observations of Maijo's writing process (cf. Faigley & Witte, 1981). Second, revisions are further defined "generative revisions", where a new text is produced, or as "retroactive revisions", where the text is corrected or modified by substitution or addition. Comparisons of these revisions indicated that generative revisions were three times more frequent than retroactive revisions, suggesting that many revisions simultaneously involved the generation of new text. Substitutions and additions constituted 80 and 10 percent, respectively, of the generative revisions. In contrast, substitutions and additions comprised 30 and 60 percent, respectively, of the retroactive revisions. These results indicate that writing during periods of generation proceeds with minor substitution-based adjustments, while writing during revision periods proceeds by enriching the rhetoric and expressions through additions. The distributions that separate a sentence into two sentences and the consolidations that combine two sentences into a sentence have a high ratio of generative revisions to retroactive revisions, indicating that the rhythm from sentence length is mostly determined at the text production stage.

Keywords creativity, literature, creative writing, revision process

POSTER

Reference

[1] Chen, D.: Internet wo seimeika suru prochronism no siso to jissen [Vitalizing the Internet: Theory and Practice of Prochronism], Tokyo: Seidosha Inc. (2013).

[2] Faigley, L. and Witte, S.: Analyzing Revision. College Composition and Communication, pp. 400–414 (1981).

[3] Maijo, O.: Ashura girl [Ashura girl], Tokyo: Shinchosha Co., Ltd. (2003).

[4] Maijo, O.: Maijo shosetsu kona fubuki [Maijo Novel Powder Snowstorm], *Gunzo,* Vol. 63, No. 1, pp. 36–57, Tokyo: Kodansha Co., Ltd. (2007).

[5] Yoshida, J.: Shosetsu no shinso wo meguru tabi: Proust to Ryunosuke Akutagawa [The travel over deep layer of novel: Proust and Akutagawa], Tokyo: Iwanami Publishing Co., Ltd. (2007).

POSTER

Quantitative Analysis of the Musical Style of Mozart: How Many Persons Should Play *Eine Kleine Nachtmusik*?

Michiru Hirano (Tokyo Institute of Technology)
Hajime Murai (Tokyo Institute of Technology)
Takehiro Inohara (Tokyo Institute of Technology)

Eine kleine Nachtmusik K.525, one of Mozart's most popular works, holds many mysteries, one of which has to do with how many players are intended to play it. The work indicates five string parts, but there is no instruction on whether it should be played like chamber music (wherein each string part is played by only one person) or as orchestral music (i.e., with each string part played by more than one person). For most of his other instrumental works, Mozart specified a form that would indicate how many players were intended. For example, works titled "string quartet" are always played as chamber music, and works titled "symphony" are always played in orchestral form. Therefore, it seems likely that Mozart would also imagine *Eine kleine Nachtmusik* as played with a certain form. Unfortunately, there is little information about this work: the purpose of composition is unknown, and no document remains from the time of its composition concerning its intended form. Four other Mozart works—*Divertimentos K.113, 136, 137, 138*—have the same problem.

Lacking infallible evidence, an analysis and investigation of the musical style of these works might reveal valuable clues for solving the problem. That is, if Mozart composed chamber music differently from orchestral music, and the clear difference between the two forms could be objectively recognized, then a work lacking a clear form might be assessed by looking for these differences and considering its similarity with other form-clear works. This study uses a quantitative approach to measure features of musical works in the scores to avoid the subjectivity of analysts on which conventional musical analysis is usually based.

It is often said that in chamber works, each string part tends to have more independence, more of a solo quality, than in orchestral works[1]. Therefore, at the beginning of this study, three types of "uniqueness" are identified and calculated for each part in the strings. "Rhythmic uniqueness" is an index of the frequency at which a note given to a part is unique in terms of its duration among notes of all parts placed at the same time. For example, when a half note is given to the first violin part and eighth notes are simultaneously assigned to the other parts, only the first violin part is unique; thus, a value corresponding to the duration of a half note is added to the "uniqueness" of the first violin part. The concepts of "melodic uniqueness" and "harmonic uniqueness" are similar to that of "rhythmic uniqueness," but they focus on the direction of pitch changes (upward, downward, or horizontally) and on the kind of steps used to construct the harmony, respectively.

As subjects of this research, 23 string quartets (chamber form) and 39 symphonies (orchestral form) are under investigation. The process of the analysis is as follows.

1. Data from the score (the New Mozart Edition by Bärenreiter-Verlag) are imported into the computer and exported into MusicXML format.
2. The features are measured using the program designed for this research.
3. Cluster analysis is performed in order to confirm whether the features are suited for distinguishing between chamber form and orchestral form.

POSTER

The progress of the research to date shows that the rhythmic uniqueness of the first violin part of early symphonies is clearly lower than that of string quartets.

Keywords music, Mozart, quantitative analysis, instrumentation, the number of string players

Reference

[1] Milliot, Sylvette: *Le quatuor,* Presses Universitaires de France (1986).

POSTER

Quantitative Analysis of Dissonance in Solo Piano Works by Claude Debussy

Aya Kanzawa (Tokyo Institute of Technology)
Akihiro Kawase (National Institute for Japanese Language and Linguistics)
Hajime Murai (Tokyo Institute of Technology)
Takehiro Inohara (Tokyo Institute of Technology)

The goal of this study is to highlight how dissonance came to be used in modern Classical music, which was typically used less often before the Romantic School. The basic research that helped achieve this goal included a quantitative analysis of consonance and dissonance to highlight history of harmony in solo piano works composed by Claude Debussy.

Today, it is generally agreed that dissonance came to be used frequently in Classical music from the latter half of the nineteenth century. Around this time, the collapse of tonal music, which lasted for about 300 years from the Baroque School to the Romantic School, gave way for modern music. Presently, it is generally agreed that Debussy introduced a new era. Debussy was not caught in the existing concept that is functional harmony and formed impressionist music by using a unique harmony including unresolved dissonance. Impressionist music is characterized by heavy use of dissonance in music. Regarding history of harmony, including dissonance, Classical music has been analyzed as a case study using the humanities method. However, it lacks objectivity and empirical analysis, since large-scale music groups have not been targeted.

In order to highlight how dissonance came to be used in modern Classical music, given that it was used less frequently in music before the Romantic School, pieces composed by Debussy in which history of harmony were apparent were analyzed.

In this paper, 66 of Debussy's solo piano works were targeted. Solo piano works were particularly useful since they included only one instrument, which minimized noise, and they provided limited variables. Musical scores were analyzed rather than recorded sounds, allowing for comparison among composers from various times, as many musical scores written during that time used unified rules. Musical scores of the original text version, which were published by Durand, France, were used. Mr. Roy Howat, an internationally renowned pianist and Debussy scholar, supervised this musical score.

Dissonance has multiple definitions; however a quantitative approach will allow for a more comprehensive analysis.

The specific steps of analysis are as follows:

1. Build a music corpus by collecting Debussy's solo piano pieces as data.
2. Extract any sounds that can be heard simultaneously from MusicXML.
3. Determine each interval by noting the combination of the two sounds being heard simultaneously.
4. Classify into consonance or dissonance, based on the definition.
5. Extract statistics regarding the frequency of use of dissonance, its density in music, and any repeated patterns.

Keywords classical music, piano, dissonance, Claude Debussy, quantitative analysis

POSTER

Development of an Asymptotic Word Correspondence System between Classical Japanese Poems and their Modern Translations

Hilofumi Yamamoto (Tokyo Institute of Technology / University of California)

Bor Hodošček (Meiji University)   Hajime Murai (Tokyo Institute of Technology)

Objectives

This project will develop an automatic word concordance system for parallel texts comprising of Classical Japanese poem texts and their associated modern translations. By using these parallel texts, we will clarify the details of language change within Japanese in an objective procedural manner that is not influenced by human observations.

Problem

Many scholars of classical Japanese poetry have tried to explain the constructions of poetic vocabulary using their intuition or the experiences they accumulated during their studies. Thus, they often produce modern Japanese translations through means of holistic explanations of each poem since they, even as specialist in classical Japanese poetry, cannot adequately explain the precise meanings of all words. Even if they could clearly explain all the words, their translations could include contradictions with their own explanations. This shows that translations include possibilities of non-literal elements, which are not expressible using ordinary word explanations.

To find the non-literal elements, we cannot manually match classical words in poem texts with modern words in translation texts. The problem of manual matchings of word for word from a parallel corpus, especially one comprised of classical texts, is that the act of judgment in identifying correspondences can lead to a loss of the original meaning of a word, since our present knowledge of classical words is conjectured and classified based on our knowledge of modern language.

To this end, it is necessary to employ computer assisted correspondence methods without relying on this human knowledge. We therefore use the asymp- totic correspondence vocabulary presumption method (Murai, 2012) to estimate corresponding pairs of classical Japanese words and their modern Japanese translations.

Methods

Using the asymptotic correspondence vocabulary presumption method to classic texts and those modern translations, the proposed method allows the extraction of corresponding vocabulary pairs. A word in a poem text will be paired with every word in the corresponding translation text of the poem. This process is repeated for all of the words in the poem using our program. Our system will generate poem-word and translation-word patterns from Kokinshū poem #1 to #1000 respectively. Based on the frequency data of poem-word and translation-word patterns, we calculate mutual information (MI) scores for each pair. Then, for each iteration, we determine the best-scoring corresponding pair. After determining the best-scoring pair words, we remove all occurrences of it from both poem texts and translation, recalculate the MI scores with the re- maining pairs, and finally determine the second best-scoring pair. This process is repeated until the MI score goes below a certain preset threshold value.

POSTER

Materials

We will use the Kokinshū with ten corresponding sets of modern translations. The Kokinshū is an anthology compiled under imperial orders (ca. 905). The Kokinshū consists of 1,111 poems including long poems (chōka) and head-repeating poems (sedōka) which are not short poems (tanka; 5/7/5/7/7 syllable style). We will only use the short poem form, which amounts to 1,000 poems, for stylistic consistency. The ten sets of modern translations of the Kokinshū are translated from 1927 to 1998 by ten Japanese poetry scholars.

This project has already begun: the parallel corpus of the Kokinshū has been constructed. We are now working on the development of computer software and the optimization of the calculation methods. As a result of our development and experimentation, we expect that literal words and non-literal elements will be extracted using our system.

Keywords corpus linguistics, word correspondences, classical Japanese poetry, modern translations, asymptotic algorithm

Reference

[1] Murai, Hajime: "Semantic Networks between Texts in Different Language Based on the Asymptotic Correspondence Vocabulary Presumption Method," *The Computers and the Humanities Symposium,* Vol. 2012, No. 7, pp. 61–68, Nov (2012).

POSTER

Memory Hunting: A Mobile App for Collecting the Location Metadata of Old Photographs

Asanobu Kitamoto (National Institute of Informatics)

The minimum requirement of using old photographs for reconstructing historical landscape is the availability of proper metadata about locations, such as the location that the photograph was created (the location of the camera), or the location shown in the photograph (the location of the target). A photograph sometimes comes with a caption that suggests the content of the photograph, but textual description usually describes a place name whose vagueness limits the precision of geographic analysis in comparison to today's geotags measured by devices such as GPS. To find the precise location of the photograph, a basic strategy is to go to the site and compare the present landscape with the past landscape recorded in the photograph. The best match between two landscapes suggests the place that the photograph was taken, but this task requires a "mental transformation of landscape" in our brain so that a 2D old photograph is mapped into a 3D landscape by virtually rotating, scaling, and projecting the image, which is not easy to perform.

Instead, we propose a simple solution; namely, creating a mobile app to show an old photograph as an overlay on the viewfinder of the camera so that two landscapes can be directly compared. At the best match, pressing the shutter button triggers the recording of location metadata through the GPS of the device with a capture of the present landscape. Photographs and metadata are then uploaded to the server to allow the comparison of landscapes between the present and the past. This workflow is realized on our mobile app, "Memory Hunting," to be used on either smart phones or tablets.

Here, comparison with "augmented reality" (AR) helps understanding the feature of the app. First, AR shows information (such as old photographs) as an overlay in the real space to enhance the interpretation of the present landscape. On the other hand, our app uses the present landscape to enhance the interpretation of old photographs, which can be regarded as "augmented photography" (AP). Second, AP can be used as a preparation step for AR, because the app can record metadata about the location and the direction of the camera. This metadata can be directly incorporated into AR, so Memory Hunting can be considered as a complementary tool to AR.

A by-product of this app is the natural introduction of a "gamification" mechanism. The app transforms the task of finding the location into the task of finding the best match between two landscapes. The latter has a clearly-defined goal that can be achieved by iteratively moving the camera to improve the goodness-of-match, and this task can give the sense of fun for users. It is expected that the sense of game may help expand the style and domain of activities based on this app. First, the style of activities can be expanded from team work to crowd-sourcing, where people enjoy taking pictures and providing location metadata at the same time. Second, the domain of activities can be expanded from historical studies to other domains such as tourism, where people enjoy walking around the town for scouting the location that photographs were taken, like a quiz.

The first version of the app was tested by the author at Tokyo and Kyoto in April 2014, using an Android smart phone. The result was promising. First, the comparison of two landscapes was significantly easier on the viewfinder than in the brain. Second, the author felt that finding the best match is not a given task but a task he would like to perform well. Moving back and forth at the site to find the best match was a trial worth spending some time. On the other hand, the author felt frustration on the limitation of the camera, especially when the best match could not be achieved due to a different angle of the lens. This problem is difficult to solve due to the limited feature of the camera on smart devices, and the lack of standard

libraries to control various types of camera on Android OS. The author suggests that these problems are the largest obstacles to expand the user base of the app.

We will release the second version of the app on Google Play in the near future so that people can download it for free and use it for their own projects. Our goal is to make a convenient tool so that digital humanists can use it in their field study using their own smart devices.

Keywords mobile app, old photograph, location metadata, augmented photography, gamification

POSTER

The Construction of Accurate Old Landform Data based on the Collaborative Edit Scientifically

Yoichi Seino (National Institutes for the Humanities)
Mamiko Mataza (Ritsumeikan University)   *Takafusa Iizuka (Ritsumeikan University)*

The old geographical information is necessary for not only the academic purpose but also disaster management or social education etc. So, to create the old geographical information is very important but that is not suitable for the business because it is not able to use for the routing or geocoding in our contemporary life. Therefore this information should be created according to the academic purpose at first and then it will be published. On the other hand, to create the large and wide geographical information requires the much time or cost as for the personal activity. In this paper, we consider the possibility of achievement about the above issues by setting up the place to create the old geographical data collaboratively and to discuss the authenticity of that data through the Internet and the Web.

We mentioned the importance of the creation about the digital old terrain database before (Seino 2013.) At that time, we refer to the importance of the raster digital elevation data. But the raster data tend to be the large file size, so it is difficult to prepare the server resources or to overcome some technical issues.

Therefore, we attempt to create the shoreline vector data as a fundamental data for creating the old terrain data. To digitize the area under the water at that time and now emerged is usable to consider various analyses on the GIS. The vector data have many advantages. One is the vector data are light and small file size, and the other is easy to convert the human-readable text data, so it is easy to control on the version control system (VCS.)

OpenStreetMap (OSM) is the most popular Volunteered Geographic Information (VGI) system in the world. OSM build a large community and a lot of people join it. OSM's number of participants is an advantage rather than other projects. However, OSM is the map that is hoped to use by ordinary people living now and the project focuses on the current geographic information, so we are not able to use this system to create the old geographical data. In addition, OSM corresponds the revision control, showing the existing data (XML format) or downloading the data but it is difficult to view the raw data because OSM uses the Relational Database Management System (RDBMS) and OSM only supports vector data so we are not able to use raster data if we want to manage the raster data in the future. In the aspect of user management, OSM does not avoid the authentic issue or edit wars because OSM is open to everyone and not only to the researchers.

The data that we created are the old bankline of the Lake Biwa based on the maps published in the Meiji and Taisho period. However, at least, we recognize that it is necessary to digitize the oldest maps surveyed by the modern style and this information is the fundamental data to restore or to simulate the premodern terrain.

We use the "GitHub" which is the source code repository and hosting facility based on the VCS software "Git" and "geojson.io" which is hosted by the MapBox as the extension service of the GitHub to create, change, and publish the geographic information.

The significant feature of the VCS is the strong revision control function. This function records who committed the shared data and if you want, you are able to rollback. In the many VCS softwares, the Git has an advantage that it is easy to "fork" and "merge" than other VCS softwares. This advantage makes us to set the Master and the branch line easily and each member can work at their own pace, and then if they need, they are able to merge to the master. These functions of the VCS is very efficient in the academic collaboration. Each member is able to work keeping their individual opinion.

POSTER

These data that we created help restoring the premodern bankline of the Lake Biwa. We try to gather the other old historical resources about the premodern reclamation or sedimentation and consider the older bankline as the future tasks. These tasks must to be discussed from various aspects. At that time, this platform will be efficient for the discussions. The substantial discussions make the data more authentic like the peer-reviewed research production in the academic society.

Keywords GIS (Geographic Information System), collaborative work, vector data, VCS (Version Control System), repository

Reference

[1] Seino, Yoichi: "The possibility about the collaborative creation of wide area ancient digital terrain database," The Second International Conference of Asian Network for GIS-based Historical Studies (ANGIS), held at Clock Tower Centennial Hall, Kyoto University, Kyoto, Japan (2013).

POSTER

List of Contributors

Arthur, Paul `<p.arthur@uws.edu.au>`
> Professor of Digital Humanities, University of Western Sydney

> Paul Arthur is Professor of Digital Humanities at the University of Western Sydney, where he leads the UWS Digital Humanities Research Group. He was previously Deputy Director of the ANU Centre for European Studies, a joint-funded special initiative of the European Commission and the Australian National University. From 2010–13 he was Deputy Director of the National Centre of Biography, ANU, and Deputy General Editor of the Australian Dictionary of Biography. A Fellow of the Royal Historical Society, he has published widely in fields of history, literature, communication and cultural studies, and has held fellowships in Australia, Europe and North America. See `paul.arthur.com`.

Bourdon-Miyamoto, Julien
> Area Informatics of the Center for Integrated Area Studies (CIAS), Kyoto University

> Researcher for the Area Informatics of the Center for Integrated Area Studies (CIAS), Kyoto University. His current research interests are pattern-based analysis and visualization of large multilingual documents corpus as well as geo-spatial processing of Area Studies data. He completed his BSc in computer science in 2005 at the University of Caen (France), studied for MSc of Advanced Computer Science in the University of Sheffield (United Kingdom). He got his MSc in Web Intelligence in 2007 from the University of Saint-Etienne (France) where his thesis title was "Multi-agent systems for the automatic composition of semantic web services in a dynamic environment".

Chen, Dominick `<dominick@dividual.jp>`
> Director, Dividual Inc.

Harada, Makiko `<sumakokima@gmail.com>`
> Graduate School of Tokyo Metropolitan University

> She is Ph.D. student of System Design at Graduate School of Tokyo Metropolitan University.

Hashimoto, Yuta `<yhashimoto1984@gmail.com>`
> Graduate School of Letters, Kyoto University

> Yuta Hashimoto is a doctoral student at Graduate School of Letters, Kyoto University. His major is history of modern science and digital history.

Higgins, Devin `<higgi135@msu.edu>`
> Digital Library Programmer, Michigan State University

> Devin Higgins is Digital Library Programmer at Michigan State University. His current professional interests include text analysis, data visualization, and building digital collections that promote user exploration. One current project uses network visualization to model interdisciplinary collaborations among academic programs. At the Center for Black Music Research in Chicago, he text-mined jazz interview transcripts to extract descriptive metadata and build an online exhibit about genre and musician biography. He holds an MA in English from the University of Illinois at Chicago and an MS in Library and Information Science from the University of Illinois at Urbana-Champaign.

Hintze, Arend

Computer Science / Microbiology & Molecular Genetics, Michigan State University

Arend Hintze is by training a Developmental Biologist (University of Braunschweig, Germany), now working on computational biology, artificial intelligence, and evolution first at the Keck Graduate Institute for Applied Biosciences, and now at Michigan State University BEACON Center of "studying evolution in action." He is one of the developers of the novelty search software and published various papers using complex data analysis tools, graph theory, and information theory.

Hirano, Michiru `<jumann@valdes.titech.ac.jp>`

Graduate Student, Tokyo Institute of Technology

Career: (2013) B.S. in Earth Sciences, Tokyo Institute of Technology. Majors: Game theory, Mathematical decision making. Research Theme: Scientific analysis of musical style.

Hodošček, Bor `<bor.hodoscek@gmail.com>`

Meiji University

Bor Hodošček is a JSPS Postdoctoral Foreign Fellow at Meiji University. He earned a Ph.D. in Engineering from the Tokyo Institute of Technology and is currently working on the quantitative modeling of register in Japanese as well as exploring its role in writing assistance systems. His interests include quantitative linguistics, natural language processing, and educational technology.

Iizuka, Takafusa `<takafusaiizuka@gmail.com>`

Graduate Student, Graduate School of Letters, Ritsumeikan University

Takafusa Iizuka graduated the Geography major, the University of Komazawa and got B.A. and completed the master's course of Graduate School of Arts and Sciences, Komazawa University and got M.A. He is now interested in the Changing of River Transportation in Modern Japan. He usually use Historical GIS, historical geographic data and some of historical features (old maps, historical statistics and historical documents).

Inohara, Takehiro

Professor, Tokyo Institute of Technology

Takehiro Inohara received his B.S. (Mathematics, 1992), M.S. (Systems science, 1994), and Ph.D. (Systems science, 1997), all from Tokyo Institute of Technology, Tokyo, Japan. He is currently a professor in Department of Value and Decision Science in Tokyo Institute of Technology, and doing researches in the fields of social systems modeling, mathematical decision making, and evaluation of systems.

Kaname, Mariko `<mariko@cscd.osaka-u.ac.jp>`

Visiting Associate Professor, Osaka University

Education: PhD (March 2001) in Literature, Osaka University. MA (March 1996) in Art, Musashino Art University. BA (March 1994) in Philosophy, Waseda University. Academic Positions: (2011–present) Visiting Associate Professor of Aesthetics, Centre for Study of Communication-Design (CSCD), Osaka University. (2009–2011) Assosiate Professor, Graduate School of Architecture and Design, Aichi Sangyo University. (2005–2009) Lecturer, CSCD, Osaka University. (2001–2004) Research Associate, Graduate School of Letters, Osaka University. Main Books: *Image: The Image in Society*, Nakanishiya, 2012. *Image: The Image in Itself*, Nakanishiya, 2011. *Roger Fry's Critical Theories: the Logic of Sensibility*, Toshindo, 2005.

Kanzawa, Aya `<aknz@valdes.titech.ac.jp>`

Graduate Student, Tokyo Institute of Technology

Career: (2013) B.A. in Mathematics, Tokyo University of Science. Majors: Cognitive Science and Music Information Processing. Research Theme: scientific analysis of piano works about classical music.

Kawase, Akihiro `<a_kawase@ninjal.ac.jp>`

National Institute for Japanese Language and Linguistics

Akihiro Kawase is a researcher at the Department of Corpus Studies, National Institute for Japanese Language and Linguistics, Japan. He received Ph.D. in cognitive science from the Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Japan in 2011. He is engaged in structuring classical Japanese texts. He won the Best Poster Award Gold Prize from the PNC/ECAI & Jinmoncom (IPSJ SIG-CH) Joint Meeting in 2013, and the ADHO Bursary Award at the Digital Humanities 2014 conference.

Kitamoto, Asanobu `<kitamoto@nii.ac.jp>`

Associate Professor, National Institute of Informatics

Asanobu Kitamoto is associate professor of National Institute of Informatics. His research focuses on data-centric approaches to humanities, earth science and biological imaging with the help of state-of-the-art computing technology such as databases, data mining and visualization. He is now leading the digital humanities project "Digital Silk Road" with the development of information infrastructure to study visual and spatial sources for humanities research.

Kudo, Akira `<ak-kudo@p.u-tokyo.ac.jp>`

Project Research Professor, University of Tokyo

Education: (2012 December) Ph.D. in Arts and Science [Tokyo Institute of Tokyo]. (2010 September) M.A. (Master of Arts) [Tokyo Institute of Tokyo]. Job: (2013 January–2014 March) Research Fellowship for Young Scientists at University of Tokyo. (2013 April–2013 September) Part-time instructor at Tokyo Institute of Technology. (2012 June–2012 August) Research Assistant at Tokyo Institute of Technology. (2012 April–2013 December) Research Fellowship for Young Scientists at Tokyo Institute of Technology. (2010 October-2012 March) Teaching Research Assistant at Tokyo Institute of Technology. (2008 December-2011 December) Research Assistant at Tokyo Institute of Technology.

Lawson, Dawn `<dawn.lawson@nyu.edu>`

Librarian for East Asian Studies (Associate Curator), New York University

Dawn Lawson has been the Librarian for East Asian Studies at New York University since 2004. This month she will receive a PhD in East Asian Studies from New York University. Her dissertation is entitled "Women, Creativity, and Translation in Mid-Meiji Japan: The Literature of Nakajima Shoen (1861–1901)". Having received a Fulbright Graduate Student Fellowship in 2011–2012, she spent that year at Josai Kokusai Daigaku studying with Mizuta Noriko and Kitada Sachie. She holds an MLS from Long Island University, an MA in Japanese literature from Harvard University and a BA in East Asian Studies from Oberlin College.

Masuda, Katsuya

Project Research Associate, University of Tokyo

Mataza, Mamiko <lt047074@ed.ritsumei.ac.jp>

Graduate Student, Graduate School of Letters, Ritsumeikan University

Mamiko Mataza is a graduate student of the Graduate School of Letters, Ritsumeikan University, and a bar association functionary. She specializes in Japanese history after the seventeenth century.

Menon, Nirmala <nimmenon@gmail.com>, <nmenon@iiti.ac.in>

Assistant Professor of Literature, Humanities and Social Sciences Department, Media and Publishing Studies Research Group, Indian Institute of Technology (IIT) Indore

Dr. Nirmala Menon is an Assistant Professor, Indian Institute of Technology (IIT) Indore India. She is the author of *Creole Cosmopolitanism: Narratives of Migrant Postcoloniality* Peter Lang, NY 2014 and also essays as *Re-Mapping the Postcolonial Canon and The Hullabaloo about Hybridity: Kiran Desai's Inheritance of Loss*. Her current research explores Digital Humanities projects in multilingual scholarship. She is an Editor on *Open Library for Humanities (OLH)*. She is reader and reviewer for various postcolonial journals like the *Journal for Postcolonial Writing (JPW) and Postcolonial Text*. Earlier, she was an Asstt Professor of Postcolonial Literature in St Anselm College, Dept of English, Manchester, NH USA.

Miller, Ben <miller@gsu.edu>

Assistant Professor, Georgia State University

Dr. Ben Miller (Ph.D. in Comparative Literature, Emory, 2009) works in critical cultural studies of computational and networked media around collective memory of traumatic events. Currently assistant professor of English and Communication at Georgia State University where he co-directs the New and Emerging Media Initiative, Dr. Miller has held postdoctoral and faculty positions at the Georgia Institute of Technology, M.I.T., and the University of North Florida.

Mima, Hideki

Associate Professor, University of Tokyo

Morioka, Tomohiko <tomo@zinbun.kyoto-u.ac.jp>

Assistant Professor, Center for Informatics in East Asian Studies, Institute for Research in Humanities, Kyoto University

Assistant professor of the Center for Informatics in East Asian Studies, Institute for Research in Humanities, Kyoto University. Ph.D (information science) from Japan Advanced Institute of Science and Technology in 1999.

Murai, Hajime <h_murai@valdes.titech.ac.jp>

Assistant Professor, Tokyo Institute of Technology

Hajime Murai is an assistant professor at the Graduate School of Decision Science and Technology, Tokyo Institute of Technology. He earned a Ph.D.in Engineering from the Tokyo Institute of Technology. His majors are natural language processing, bibliometrics, and bible studies. He is currently working on the scientific analysis of text interpretation, and quantitative analysis within the humanities.

Nagasaki, Kiyonori

Senior Fellow, International Institute for Digital Humanities / Project Associate Professor, Interfaculty Initiative in Information Studies, University of Tokyo

Nakamura, Yusuke <y-nkmr@l.u-tokyo.ac.jp>

Associate Professor, University of Tokyo

Education: (1995) Ph.D. Graduate School of Arts and Sciences, The University of Tokyo. Work Experience: (2004–present) Associate Professor at the Department of Cultural Resources Studies, Graduate School of Humanities and Sociology, The University of Tokyo. Research Areas: Development Studies, Latin America, Cognitive Sciences, Cultural Resources Studies.

Nurjanah, Jane

Department of Industrial Art, Graduate School of System Design, Tokyo Metropolitan University

Ogiso, Toshinobu

National Institute for Japanese Language and Linguistics

Okada, Takeshi <okadatak@p.u-tokyo.ac.jp>

Professor, University of Tokyo

Olive, Jennifer <jolive1@gsu.edu>

Ph.D Student, Georgia State University

Padilla, Thomas

Digital Humanities, Linguistics, and Philosophy Librarian, Michigan State University

Thomas Padilla is Digital Humanities, Linguistics, and Philosophy Librarian at Michigan State University, where he works to augment digital collections, foster Digital Humanities instruction and collaboration, and expand data curation services. While earning his MSLIS, he worked in the Preservation Unit and the Scholarly Commons of the University of Illinois at Urbana-Champaign Libraries. Before attending UIUC he worked at the Library of Congress helping to extend digital preservation professional development opportunities throughout the United States. He holds an MSLIS from the University of Illinois at Urbana-Champaign and an MA in World History from San Francisco State University.

Seino, Yoichi <seinoyoichi@gmail.com>

National Institutes for the Humanities

Yoichi Seino graduated the Archaeology major, the University of Tsukuba and got B.A., and he completed the master's course of Graduate School of Humanities and Social Sciences, Mie University and got M.A. He finished the doctoral course of Graduate School of Human and Environmental Studies, Kyoto University. He is now interested in the local administration system of the Japanese Ancient State. He usually use the GIS, archaeological data and some of historical features (ancient road, ancient topography and historical documents). He works at the National Institutes for the Humanities now and maintains the Resource Sharing System.

Sherratt, Tim

Trove Manager, National Library of Australia

He's currently the manager of Trove at the National Library of Australia and an Adjunct Associate Professor in the Centre for Creative and Cultural Research at the University of Canberra.

Shirai, Tetsuya

Faculty of Library, Information and Media Science, University of Tsukuba

Tetsuya Shirai is a professor of University of Tsukuba. He is interested in theories and practical issues to preserve and use the records and memories in local communities in the future, and consider all the various factors together while learning from the experiences of archival science or museology.

Short, Harold

Professor, King's College London

Shrestha, Ayush

Ph.D Candidate, Georgia State University

Subtirelu, Nicolas

Ph.D Student, Georgia State University

Sugiyama, Mao <sugiyama.mao0420@gmail.com>

Master Student, Graduate School of Language and Culture, Osaka University

Sumida, Shoko

Part-time Lecturer, Ritsumeikan University

Education: (2011) PhD at Ritsumeikan University, Kyoto. Doctoral dissertation entitled: "A Study about the Relationship between Paul Gauguin's Works and His Nostalgia". Academic Positions: (2012–2013) Visiting Reserach Fellow at the International Insitute of Language and Culture Studies, Ritsumeikan University. (2012–present) Part-time Lecturer at the Faculty of Social Sciences, Ritsumeikan University. Main articles: "The Reception of Paul Gauguin by Japanese Painters in the 1910s: A Study about Bakusen Tsuchida's Shima no Onna (1912) and Ama (1913)," *The Journal of Asian Arts & Aesthetics*, No. 3, pp. 85–92 (2009–12).

Suzuki, Chikahiko

Doctoral Student, University of Tokyo

Suzuki, Toshiya `<mpsuzuki@hiroshima-u.ac.jp>`

Research Assistant, Hiroshima University

Suzuki Toshiya earned his Ph.D. degree in Dept. of Physics from Tohoku University, Sendai, 1998. Since 2000, he has been working around printing, fonts and textlayout related softwares in open source softwares, especially around PostScript technology: ghostscript, gs-cjk and FreeType. Since 2007, he has been an official member of ISO/IEC JTC1/SC34 (the subcommittee for font and document standard) and also participates ISO/IEC JTC1/SC2 (the subcommittee for coded character set).

Takahashi, Ryoichi `<ryoichit@valdes.titech.ac.jp>`

Graduate Student, Tokyo Institute of Technology

Career: (2009) B.A. in Pychology, Aoyama Gakuin University. Majors: Cognitive Psychology and Natural Language Processing. Research Theme: scientific analysis of creative writing and narrative text comprehension.

Takata, Yurina `<yurina.tkt@gmail.com>`

Graduate School of System Design, Tokyo Metropolitan University

Yurina Takata is currently a Ph.D. student in Department of Industrial Art, Graduate School of System Design, Tokyo Metropolitan University.

Tamura, Tsuyoshi Research Assistant, Akita University of Art

Education: M.A. (March 2007) and B.A. (March 2005) in Sociology, Ritsumeikan University. Academic Position: (April 2013–present) Research assistant, Landscape design course, Akita university of Art, Akita, Japan. Main articles: "Designing Kyoto's scenery plan: Kyo-machiya developing as a design standard for the entire townscape," *The Journal of Asian Arts & Aesthetics*, Vol. 4, pp. 113–122 (2010). "The Potential of 'My Scenes' in the Formation of Landscape Notions," *Ritsumeikan Studies in Language and Culture*, Vol. 25, No. 1, pp. 75–89 (2013).

Tanaka, Makiro `<makiro@meiji.ac.jp>`

Meiji University

Makiro Tanaka is a professor at Meiji University. He earned a Ph.D. in Linguistics from the Tokyo Institute of Technology and is currently working on the construction of a diachronic corpus covering Early to Middle Japanese as well as examining the relationship between Yamato and Sino-Japanese vocabulary during the Heian period. His interests include Japanese linguistics and tackling language barriers caused by complicated terminology.

Uda, Norihiko

Faculty of Library, Information and Media Science, University of Tsukuba

Watanave, Hidenori

Department of Industrial Art, Graduate School of System Design, Tokyo Metropolitan University

He is Ph.D. Information Architect. Associate professor at Faculty of System Design, Tokyo Metropolitan University Visiting associate professor at The Center for Integrated Area Studies (CIAS), Kyoto University.

Yamada, Taizo

Historiographical Institute, University of Tokyo

Yamamoto, Hilofumi `<yamagen@ryu.titech.ac.jp>`
> Tokyo Institute of Technology / University of California, San Diego

> Hilofumi Yamamoto is an associate professor at the Tokyo Institute of Technology. He earned a Ph.D. in Linguistics at Australian National University and is currently working on the mathematical modeling of vocabulary, linguistic change, and language complexity.

Yamamoto, Takeo `<ty@tylab.org>`
> Ph.D. Candidate, English Department, Graduate School of Literature and Sociology, The University of Tokyo

> Career: (1967) D. Sc. (in Chemistry) from University of Tokyo. (1967–1972) Assistant Professor, Faculty of Science, University of Tokyo (Chemistry). (1972–1981) Associate Professor, Computer Centre, University of Tokyo (Computer Systems). (1981–2001) Professor, Faculty of Library and Information Science, University of Library and Information Science (Information Systems). (1995–1999) Vice President, University of Library and Information Science. (2001 to present) Professor Emeritus, University of Library and Information Science. (2001–2005) Professor and Director, Division of Information and Media Research Division, National Institute of Informatics. (2005 to present) Professor Emeritus, National Institute of Informatics. (2005 to present) Research Fellow, The Oriental Library.

Yanagisawa, Masayuki
> Center for Integrated Area Studies, Kyoto University

Yeh, Ya-Ju `<au4469@mail.au.edu.tw>`
> Assistant Professor, Aletheia University, Taiwan

> Ya-Ju Yeh is currently Assistant Professor in the department of English at Aletheia University in Taiwan. Her Ph.D. dissertation explores the relation between body and public space in Henry James, Grant Allen, and Dorothy Richardson's novels. Her interests in research include urban studies, women studies, material culture studies, and digital humanities. She is working on the research project entitled "The Popularization of Typing Technology" which explores issues about digital literacy, modern office culture and the transformation of body experiences in relation to digital products.

Yoshida, Yuko
> Faculty of Library, Information and Media Science, University of Tsukuba

> Yuko Yoshida is a professor of University of Tsukuba. She works on historical research on the philosophy and practices of American public libraries.

Zhao, Jin
> Ph.D Candidate, Georgia State University

Zhao, Yanjun
> Ph.D Candidate, Georgia State University