Corpus-Based Research on Chinese Language and Linguistics
edited by Bianca Basciano, Franco Gatti, Anna Morbiato

Some Reflections on the *Database of Medieval Chinese Texts* as a Multi-Purpose Tool for Research, Teaching, and International Collaboration

Christoph Anderl
Ghent University, Belgium

Abstract This paper gives an introduction to a Digital Humanities project at the Department of Languages and Cultures (Ghent University), the *Database of Medieval Chinese Texts* (DMCT), a collaborative project with several international partners. The structure of the DB is multi-modular, consisting of reference modules in the form of XML marked-up medieval non-canonical Chinese Buddhist texts, as well as analytical modules such as the Variants, Syntax, and Sentence Analysis modules. The architecture is 'open' and modules can be added, modified, and interlinked based on specific research requirements. The DB is multifunctional and not only provides information on key texts and their linguistic features, but also constitutes a research tool (featuring sophisticated online input masks and analytical tools) with which researchers can input and process data. In addition to its function in a research environment, it is also used in advanced master classes, in the framework of master thesis and PhD projects, as well as for internships. The DB has also an important 'socio-institutional' function, being situated at the intersection of Buddhological and historical linguistic studies, two of the main fields of research at the department.

Keywords Digital humanities. Linguistic database. XML mark-up. Medieval Chinese. Chinese syntax. Chinese character variants.

Edizioni Ca'Foscari

Sinica venetiana 6
e-ISSN 2610-9042 | ISSN 2610-9654
ISBN [ebook] 978-88-6969-406-6 | ISBN [print] 978-88-6969-407-3

Peer review | Open access
Submitted 2020-09-29 | Accepted 2020-12-06 | Published 2020-12-21

DOI 10.30687/978-88-6969-406-6/011

1 Introduction

The digitisation of premodern Chinese texts and the availability of an increasing number of huge text corpora have revolutionised many aspects of Sinological research during the last decades. Nowadays, the tracing of the source of a specific text passage, a term, a name, or a grammatical marker can ideally be performed within a very short period, whereas previously one frequently had to consult multiple indices or dictionaries, or even read through entire texts in order to retrieve the information. In addition, statistical material concerning the frequency of semantic items or syntactic function words can be collected much more speedily as compared to pre-digitisation times.

During my participation in projects involving text corpora and databases during the last 25 years, I have been observing a variety of approaches concerning the use and integration of the swiftly developing digital collections of texts, as well as a variety of continuously changing database and programming environments, which often entailed numerous problems and often rendered certain technical frameworks obsolete after a relatively short period. Naturally, the 'fall-out' rate in this field of research is significant; on the other hand, various projects have proven to become stable digital platforms and are continuously maintained and improved, greatly facilitating the work of the targeted research community. The reasons why certain database/digitisation projects have been successful – while others have not – are manifold and will not be discussed in detail in this paper.[1]

Considering the above, initiating a new database (DB) project is a risky task, since the initial technical framework will have a great impact on the future development of the DB. Therefore, when we first started designing the Database of Medieval Chinese Texts (DMCT)[2] in 2014, we decided to take a 'hybrid' approach, i.e. a project which could

[1] Based on my experience with database projects, I have observed that successful projects seem to be often driven by the vision *of one person* or a small group of people, capable of motivating others to participate and contribute (as well as attracting the necessary funding). Among the databases I personally use most frequently, I want to mention the *Digital Dictionary of Buddhism* (DDB; ed. in chief: Charles Muller), which has developed immensely during the last years, with dozens of researchers contributing their research results, as well as the huge and ever-expanding digital collections of Buddhist texts in the form of the Chinese Buddhist Electronic Text Association (CBETA) and the SAT Daizōkyō Text databases. The collections of East Asian digital Buddhist corpora have expanded and improved at a very fast pace, one of the reasons being the work of innumerable anonymous contributors who input and proofread a vast number of texts. Another successful and innovative DB project I want to mention is *Thesaurus Linguae Sericae* (TLS, initiated more than 20 years ago by Christoph Harbsmeier), which has become an indispensable analytical tool for research on premodern Chinese texts.

[2] Concerning the editors of and contributors to the DB project, please see https://www.database-of-medieval-chinese-texts.be.

develop in a multi-functional, multi-purpose and flexible way, and a DB which could 'grow' organically according to varying research and teaching requirements (for further elaborations, please see below).

From the beginning, the DMCT has been an international and collaborative project, drawing on the expertise of specialists in various fields, the main partners being Ghent University (Department of Languages and Cultures; Ghent Centre for Buddhist Studies) and Dharma Drum Institute of Liberal Arts (DILA, New Taipei City),[3] one of the leading Asian research centres concerning the digitisation of premodern Chinese texts. In addition, we have been collaborating with specialists in digitisation and Chinese text mark-up, most importantly, with Marcus Bingenheimer (formerly DILA; now Temple University).

2 The Technical Framework

When initiating the project in 2014, we were using eXist, a platform I had used in previous projects and which is very suitable for dealing with files in XML format (i.e. the mark-up language we use for the digitised texts), but for technical reasons we migrated to MySQL ca. three years ago.[4] MySQL is a relational DB, which is organised in tables. It can use different storage engines and, depending on the specific table, we use InnoDB[5] or MyISAM. MyISAM is specifically used for all tables which are designed for full-text searches, whereas InnoDB is used for all other tables, such as the user management tables.

The programme logic is implemented in PHP,[6] using object-oriented programming (OOP) and other interfaces, like PDOs (i.e. PHP Data Objects) combined with the Open Source PHP User Management Framework UserSpice.[7]

The view of the DB is designed with Cascading Style Sheets (CSS) and further languages are HTML5 and JavaScript. Since the encoded

[3] These two institutions, in addition to the Research Foundation Flanders (FWO), have been the main sponsors of the DB. We also received financial support from the Tianzhu Foundation for the programming work. Administrative support and expert advice have been provided by members of the Dunhuang Academy, as well as by the international project *From the Ground Up. Buddhism and East Asian Religions* at the University of British Columbia.

[4] The technical work on the DB has been primarily performed by the programming specialists Christian Bell (Bell Internet Design) and Jan Schrupp.

[5] InnoDB is a product of the Oracle Corporation and is distributed under the GNU General Public Lincence. For an introduction to InnoDB storage engine, see `https://dev.mysql.com/doc/refman/8.0/en/innodb-introduction.html`. On MyIsam, see `https://dev.mysql.com/doc/refman/8.0/en/myisam-storage-engine.html`.

[6] PHP is a programming language used especially for web development.

[7] See `https://userspice.com`.

texts are XML files but the InnoDB itself is not suitable for storing XML files (unlike eXist), a XML import/export function was implemented.

Since recently, we have been using OpenProject[8] for the communication between editors/contributors and programmers, in order to improve the management of the work packages. All modules of the DB have commentary functions integrated, in order to add an interactive element in the communication with the (registered) users. The DB also features an advanced system of user management,[9] as well as sophisticated input interfaces for each module.

The DB consists of several modules whose data can be cross-referenced to each other. Currently, only some of the modules are public (the Text module, the Variant module, and the Bibliography), while some are currently for internal use only. A module for defining user rights makes it possible to assign permission to 'view' and/or 'edit' to each registered user/editor of the site, which has proven very useful in teaching environments (i.e. the students learn how to directly input data) and in the context of internships (see § 8). Unregistered visitors can fully access the public parts of the DB. By 2020, the public parts comprise all marked-up texts in two viewing modes ('diplomatic' and 'regularised'; see § 6 for more details), the module of Variant Chinese Characters ('Variant DB'; see § 7.1), and a bibliography. The internal modules are the Module of Medieval Chinese Syntactic Markers ('Syntax DB'; see § 7.2), the 'Sentence Analysis' module (see § 7.3), and the DB of 禪 *Chan* idiomatic phrases (see § 7.4). Currently, work on an additional module on Phonetic Loan Characters (通假字 *tōngjiǎzì*) is under construction.[10]

[8] OpenProject is an open-source management software which we use for the assignment and coordination of work packages in the maintenance and development of the DB (for more information on this app, see `https://www.openproject.org`). This software has proved to be very useful for enhancing the communication and workflow efficiency among the participants.

[9] I.e. 'editing'/'new entry'/'delete entry' functions can be assigned very specifically for each module of the DB. This is especially important when granting user rights to master students in the context of their internships (in order to limit the possibility of 'accidental damage' to the DB).

[10] This module will collect references concerning character substitutions in manuscript texts, including phonetic loan characters, characters exchanged based on their structural similarities in handwriting, and other types of substitutions. Since the analysis of substitutions in handwritten manuscripts is highly complex, it was not included into the standard mark-up procedures. However, substitutions were systematically marked with 'sic' in the XML files, and can thus be extracted and compiled in lists, awaiting further analysis. The editors of the DB have also initiated collaboration with Fudan University, which hosts a large project on medieval Chinese phonetic loan characters. Within the framework of a PhD project on medieval Chinese writing (main researcher: Suzanne Burdorf), we also work on the visualisation of the 'social network' of Chinese characters/variant forms, i.e. visualising the various relations a given character form has with other forms, based on phonetic substitutions and/or word family relations, graphic variations, or structural similarities (structural similarities in hand-

3 Workflow and Technical Challenges

The maintenance and development of the DB is time- and resource-intensive, since it has to be periodically updated, adjusted and programmed to include data from current research activities, and the participants of the project have to be coordinated. However, as an international project, work processes and costs are shared between several institutions, and funding has been relatively stable so far. In addition, the DB profits from the work invested in the course of specific PhD and MA projects, and a system of 3-month internships in the framework of the Ghent University MA program.

4 Stable and Flexible Aspects of the Data

Digital tools and web-based DBs are often relatively short-lived, since they have to be continuously hosted and maintained. As such, data management and preservation has become an important issue and has been addressed from the beginning of the project. The project is therefore construed so as to ensure the *long-term preservation of the raw data* in the form of digitised and high-quality marked-up texts in XML format[11] and in accordance with the guidelines of the Text Encoding Initiative (TEI). Once produced, the format of the documents allows easy storage and maintenance and can be universally decoded beyond the limitations of specific research projects.[12] In the further development of the DB we will collaborate with the Ghent Centre of Digital Humanities in order to insure long-term preservation and universal accessibility of the raw data. All textual raw data are made accessible as open-source files.

By contrast, the transformations of these raw data into specific formats and technical environments are by nature more short-lived, based on the need of continuity in the maintenance and – related to

written forms of Chinese characters are one of the main reasons of 'erroneous' substitutions in copying processes).

[11] Extensive Markup Language (XML) is an open standard for encoding documents, providing marked-up raw data (in this case textual documents) which can be conveniently transformed into a variety of applications, e.g. into XHTML for web pages, into versions suitable for printing etc. The production of XML documents is a very time-intensive process for the encoder, since the documents have to be well-formed in order to be validated. In order to facilitate the encoding to a certain degree, we use an XML editor (concretely, oXygen). The project generally follows the guidelines of the Text Encoding Initiative (the last version of the manual, TEI P5, consists of 1934 pages! For the mark-up of manuscripts, see especially pages 320-424).

[12] All marked-up manuscript texts are freely downloadable and can be used in accordance with the Creative Commons Attribution 3.0 Unported Licence (`https://creativecommons.org/licenses/by/3.0`).

that – in funding. As such, the integration and publication of the raw data as the web-based DCMT is aimed at more short-term goals, based on local research projects, publication strategies, international collaboration, and pedagogical aspects.

5 The Reference Data Collections

The core of the DB project is the collection of texts, consisting of meticulously marked-up manuscript texts, with a focus on the period between ca. 700 and 1000 CE. The late Tang (618-907), Five Dynasties (907-960) and early Song (960-1279) periods are crucial for the study of the development of grammatical markers and semantic items typical for early Mandarin/early 白話 *báihuà* literature. As such, non-canonical texts preserved in the Dunhuang corpus[13] dating from this period are of great significance for reconstructing the early phase of the development of many important features of Mandarin and other Chinese dialects. In the project, we collect a corpus of medieval Chinese texts which is relevant from *various angles of research*. Since the great majority of pre-Song Medieval Chinese texts containing colloquial elements were composed in the context of Buddhism, the DB mainly constitutes a repository of editions of non-canonical Buddhist texts. In addition, several important semi-vernacular literary genres are represented, such as early Chan doctrinal[14] and appraisal texts, 'Transformation texts' (變文 *biànwén*), Avadāna (緣起 *yuánqǐ* / 因緣 *yīnyuán*, i.e. popular versions of narratives concerning the Buddha's life), and Sūtra Lecture texts (講經文 *jiǎngjīng wén*; i.e. vernacular sermons on Buddhist scriptures).[15] All of these text types had an important impact

[13] Dunhuang texts are spread in collections around the world (for the main holdings, see Rong 2013). However, a great number of manuscripts have been made publicly available in the form of facsimiles by the International Dunhuang Project (IDP, `http://idp.bl.uk`, London, with mirror sites in Paris and Beijing).

[14] Many early Chan texts (especially those attributed to the so-called "Northern School") were contributed to the DB by Marcus Bingenheimer, based on the project *Four Early Chan Texts from Dunhuang. A TEI-Based Edition* (2014-17). The results of this project were also published in a printed form (Bingenheimer, Chang 2018). Although early Chan texts show a lesser degree of vernacularisation as compared to other late Tang genres, they are still of great importance for the study of the colloquial features of the Chinese varieties spoken during the Tang period. Some manuscripts are of special interest, e.g. S.735v, S.2503, S.7961, Beijing 1351v, S.2058, P.2270 etc., which are a treasure grove for researching the earliest predecessors of Modern Mandarin interrogative pronoun 什麼 *shénme* 'what'. In addition, some early Chan texts also show features typical for Northwestern Medieval Chinese (for an overview of scholarship on this historical dialect, see Osterkamp, Anderl 2017).

[15] For a short overview of Dunhuang popular literature, see Rong 2013, 398-412. The above genres constitute our most important sources for the study of the spoken language of the late Tang, Five Dynasties and early Song periods. Particularly the Trans-

on the development of various literary genres during the Song period. As the project progresses, we will also try to include other relevant material, such as Tang poems preserved in Dunhuang containing colloquial elements, colloquial (and sometimes bilingual) phrasebooks, schooling texts, lexicographical material etc. This corpus of texts is of great importance for research on early colloquial grammatical markers and syntactic constructions, as well as the development of lexical items. In the current version of the DB, ca. 140 texts are included (representing ca. nine years of work for an experienced encoder) with a rate of ca. fifteen new texts added every year.

6 The Digitisation of the Texts and Their Embedding in the DMCT

The manuscripts are encoded following the guidelines established by Marcus Bingenheimer (in collaboration with DILA), based on the mark-up conventions formulated by the Text Encoding Initiative (TEI). The mark-up focus is on textual features such as variant characters, loan characters and character substitutions (通假字 *tōngjiǎzì*), damaged and unclear passages, added/deleted/repeated characters, punctuation and diacritic markers, abbreviations, notes in the text etc.[16] Mark-up work is very time-consuming and difficult and one professional encoder completes in average ca. 15 manuscript texts per year, depending on the length and difficulties of the texts. After the completion of the mark-up, the texts are sent to Ghent University in XML format, transformed into HTML form and embedded in the DMCT by the project programmers. In DMCT, all texts are visualised in two ways (based on the same XML file), as a 'diplomatic' version (including references to variant characters which are projected as images on the upper right side of the screen, when the cursor moves over a character with a var-

formation texts have received considerable scholarly attention (for the genre features, see for example Mair 1983). Since recently, in the framework of a PhD project, also the variant characters of 祖堂集 *Zutang ji* (ZTJ; 10th century) are in the progress of being integrated in the DB, based on a digitised version of an original print preserved at Kyōto University (see below for more information). Currently, ca. 1,300 variants from the initial fascicles of ZTJ have been input and analysed by Laurent Van Cutsem. For a full list of marked-up texts currently publicly available in the DB, see https://www.database-of-medieval-chinese-texts.be/views/texts/mcgbd_project/showText.php and https://www.database-of-medieval-chinese-texts.be/views/texts/chan_dunhuang/showText.php.

[16] For a full list of features and how they are expressed in the mark-up, see http://wiki.dila.edu.tw/pages/%E6%95%A6%E7%85%8C%E6%BC%A2%E6%96%87%E4%BD%9B%E6%95%99%E5%AF%AB%E5%8D%B7%E9%BB%9E%E6%A0%A1%E6%9C%AC%E5%B7%A5%E4%BD%9C%E6%89%8B%E5%86%8A. Variant characters are also cross-checked with the large Taiwanese variant DB, *Dictionary of Chinese Character Variants* (https://dict.variants.moe.edu.tw/variants/rbt/home.do).

iant form, in addition to displaying other manuscript features), and a 'regularised'[17] version in which characters are represented in their standard forms and other textual features are resolved into a 'readable' text (frequently, annotations are added in the footnotes, including parallel passages from other manuscripts/texts, as well as references to dictionaries and secondary literature).

The flexibility of the XML format does not only allow various HTML transformations, but can also be used as the basis for a printed edition of a text. Below, I provide a schematic figure of the workflow from manuscript facsimile to TEI-compatible mark-up, and the transformations of the XML file to two HTML visualisations.

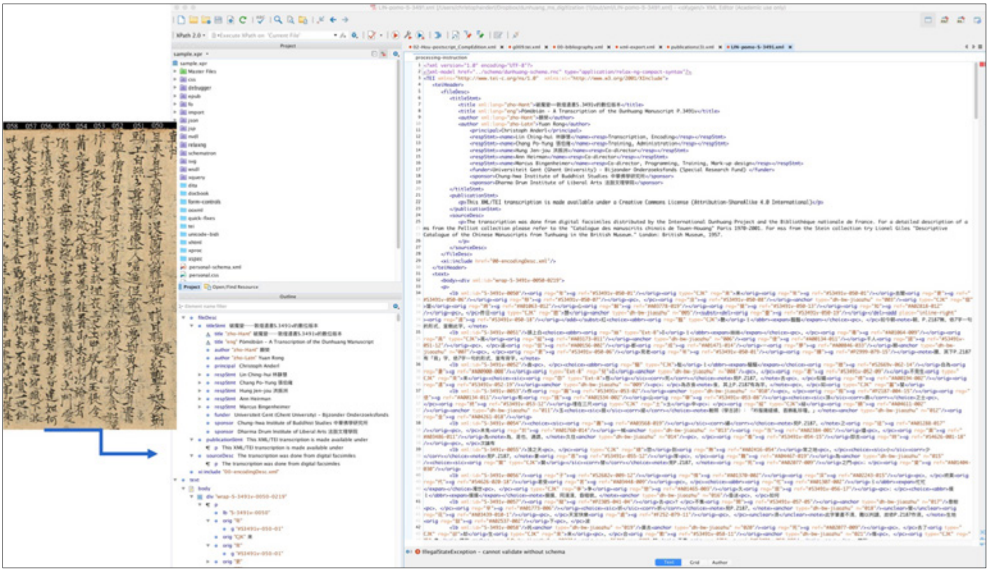
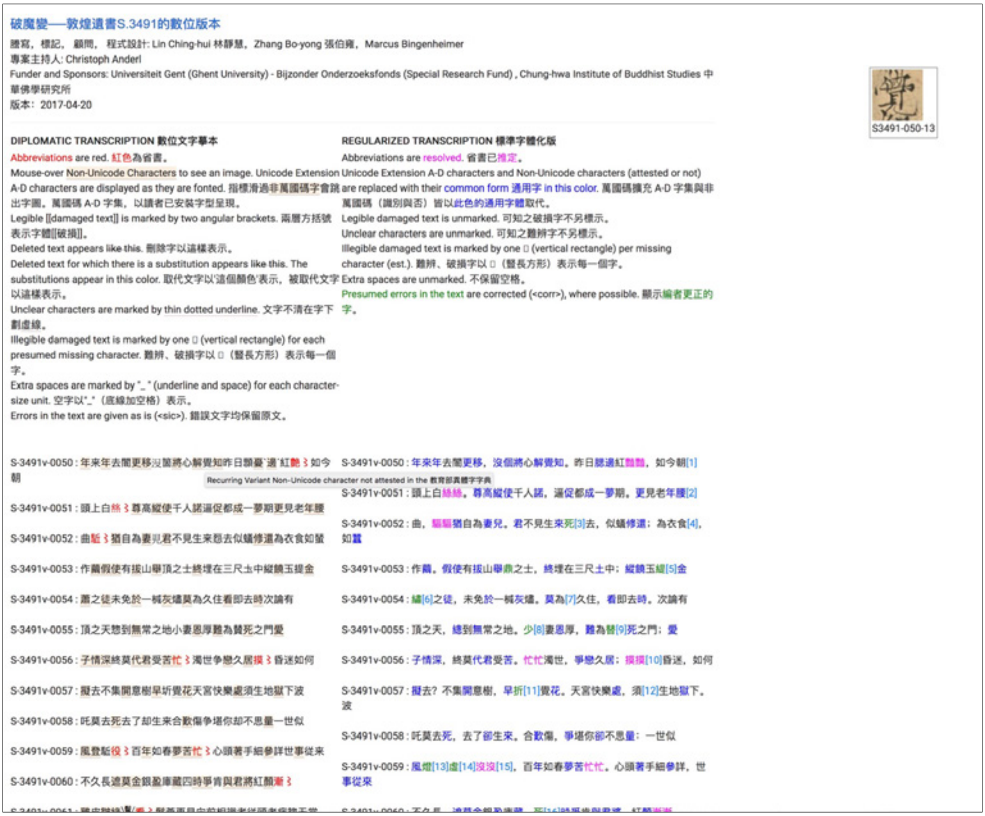


Figure 1.1 Based on the digitised facsimile of the manuscript, the text is encoded in oXygen by a specialist encoder (during the last six years, this work has been performed by Dr. Lin Ching-hui 林靜慧, DILA), following the TEI conventions for manuscript encoding with some adaptations. In addition to basic information (line number, missing/unreadable characters etc.; notes are integrated through an <anchor> element), the focus is on the identification and recording of variant characters. Phonetic loans and other substituted characters are presently only marked with <sic>, awaiting further analysis at a later date (currently, they are integrated in a <choice> element structure, 'X' being a substitution and 'Y' the assumed regularised form), the typical structure being: <choice> <sic>X</sic><corr>Y</corr></choice>. The screenshot shows the mark-up of several lines of the 破魔變 *Pò Mó Biàn* (Transformation [Text] on the Destruction of [Demon King] Māra), lines 50-58 of the manuscript Stein 3491*v.*, a Dunhuang manuscript stored at the British Library and a digitised facsimile provided by IDP

[17] On details concerning the 'regularisation' of variants, please see the link above (fn. 16).



破魔變──敦煌遺書S.3491的數位版本

謄寫，標記， 顧問， 程式設計: Lin Ching-hui 林靜慧，Zhang Bo-yong 張伯雍，Marcus Bingenheimer
專案主持人: Christoph Anderl
Funder and Sponsors: Universiteit Gent (Ghent University) - Bijzonder Onderzoeksfonds (Special Research Fund) , Chung-hwa Institute of Buddhist Studies 中華佛學研究所
版本： 2017-04-20

S3491-050-13

DIPLOMATIC TRANSCRIPTION 數位文字摹本

Abbreviations are red. 紅色為省書。

Mouse-over Non-Unicode Characters to see an image. Unicode Extension A-D characters are displayed as they are fonted. 指標滑過非萬國碼字會跳出字圖。萬國碼 A-D 字集，以讀者已安裝字型呈現。

Legible [[damaged text]] is marked by two angular brackets. 兩層方括號表示字體[[破損]]。

Deleted text appears like this. 刪除字以這樣表示。

Deleted text for which there is a substitution appears like this. The substitutions appear in this color. 取代文字以'這個顏色'表示。被取代文字以這樣表示。

Unclear characters are marked by thin dotted underline. 文字不清在字下劃虛線。

Illegible damaged text is marked by one ▯ (vertical rectangle) for each presumed missing character. 難辨、破損字以 ▯（豎長方形）表示每一個字。

Extra spaces are marked by "_ " (underline and space) for each character-size unit. 空字以"_"（底線加空格）表示。

Errors in the text are given as is (<sic>). 錯誤文字均保留原文。

S-3491v-0050 : 年來年去闇更移沒箇將心解覺知昨日顋邊'邊'紅艷〻如今朝

Recurring Variant Non-Unicode character not attested in the 教育部異體字字典

S-3491v-0051 : 頭上白絲〻尊高縱使千人諾逼促都成一夢期更見老年腰

S-3491v-0052 : 曲驅〻猶自為妻兒君不見生來怨去似蟻修還為衣食如蠶

S-3491v-0053 : 作繭假使有拔山舉頂之士終埋在三尺土中縱饒玉提金

S-3491v-0054 : 繡之徒未免於一械灰燼莫為久住看即去時次論有

S-3491v-0055 : 頂之天懸到無常之地小妻恩厚難為替死之門愛

S-3491v-0056 : 子情深終莫代君受苦忙〻濁世爭戀久居擾〻昏迷如何

S-3491v-0057 : 擬去不集閑意樹早圻覺花天宮快樂處須生地獄下波

S-3491v-0058 : 吒莫去死去了却生來合歡傷爭堪你却不思量一世似

S-3491v-0059 : 風燈駈役〻百年如春夢苦忙〻心頭著手細參詳世事從來

S-3491v-0060 : 不久長遮莫金銀盈庫藏四時爭肯與君將紅顏漸〻

REGULARIZED TRANSCRIPTION 標準字體化版

Abbreviations are resolved. 省書已推定。

Unicode Extension A-D characters and Non-Unicode characters (attested or not) are replaced with their common form 通用字 in this color. 萬國碼擴充 A-D 字集與非萬國碼（識別與否）皆以此色的通用字體取代。

Legible damaged text is unmarked. 可知之破損字不另標示。

Unclear characters are unmarked. 可知之難辨字不另標示。

Illegible damaged text is marked by one ▯ (vertical rectangle) per missing character (est.). 難辨、破損字以 ▯（豎長方形）表示每一個字。

Extra spaces are unmarked. 不保留空格。

Presumed errors in the text are corrected (<corr>), where possible. 顯示編者更正的字。

S-3491v-0050 : 年來年去闇更移，沒個將心解覺知。昨日腮邊紅豔豔，如今朝[1]

S-3491v-0051 : 頭上白絲絲。尊高縱使千人諾，逼促都成一夢期。更見老年腰[2]

S-3491v-0052 : 曲，驅驅猶自為妻兒。君不見生來死[3]去，似蟻修還；為衣食[4]，如蠶

S-3491v-0053 : 作繭。假使有拔山舉鼎之士，終埋在三尺土中；縱饒玉提[5]金

S-3491v-0054 : 繡[6]之徒，未免於一械灰燼。莫為[7]久住，看即去時。次論有

S-3491v-0055 : 頂之天，總到無常之地。少[8]妻恩厚，難為替[9]死之門：愛

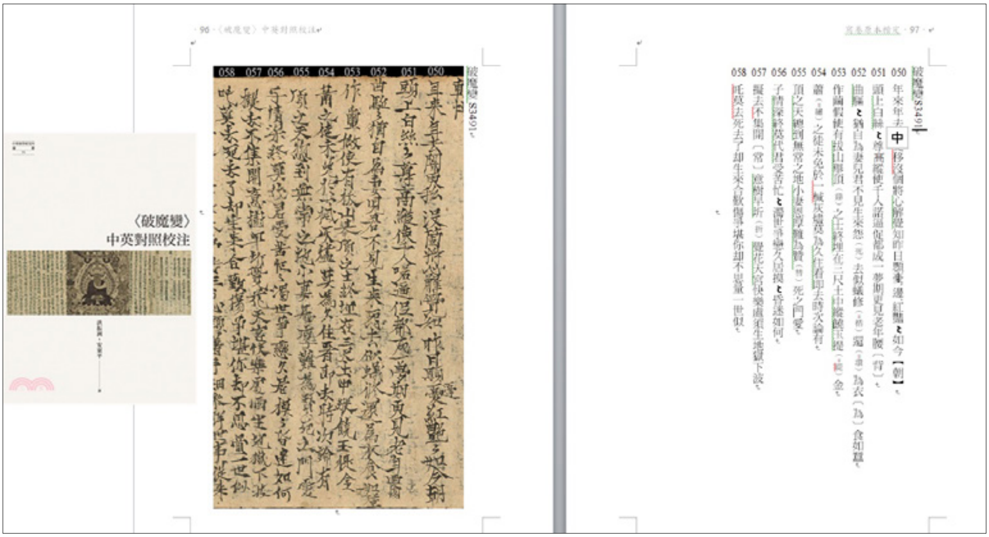
S-3491v-0056 : 子情深，終莫代君受苦。忙忙濁世，爭戀久居：擾擾[10]昏迷，如何

S-3491v-0057 : 擬去？不集閑意樹，早折[11]覺花。天宮快樂處，須[12]生地獄下。波

S-3491v-0058 : 吒莫去死，去了卻生來。合歡傷，爭堪你卻不思量：一世似

S-3491v-0059 : 風燈[13]虛[14]沒沒[15]。百年如春夢苦忙忙。心頭著手細參詳，世事從來

Figure 1.2 Screenshot exemplifying a typical workflow:the passage encoded in 1.1 is transformed into two types of HTML visualisations in the DMCT. On the left side is a 'diplomatic transcription' with information on many original features of the manuscript preserved (including the projection of variants, here referred to as "non-Unicode characters", on the right upper corner when moving the cursor over passages in light orange). To the right side, a 'regularised transcription' is visualised, with problematic passages resolved into a readable text and including annotations. Note that the ID number of the image of the variant visualised on the right corner indicates its exact positioning in the manuscript, concretely, being character 13 of the column ('line') 50 of Stein 3491 (S3491-50-13). This type of referencing helps us to interlink the graphical variants stored in the Variants Module directly with the corresponding line number of the text in which they appear



·96·《破魔變》中英對照校注

《破魔變》中英對照校注

宮本本校文·97·

破魔變S3491

050 年來年去中移沒個將心解覺知昨日腮邊紅豔〻如今【朝】

051 頭上白絲〻尊高縱使千人諾逼促都成一夢期更見老年腰（背）

052 曲驅〻猶自為妻兒君不見生來想（死）去似蟻修（循）環（環）為衣（為）食如蠶

053 作繭假使有拔山舉頂（鼎）之士終埋在三尺土中縱饒玉提（緹）金

054 繡（繡）之徒未免於一械灰燼莫為久住看即去時不論有

055 頂之天總到無常之地小妻因儻為贊（暫）死之門愛

056 子情深終莫代君受苦忙〻濁世爭戀久居摸〻昏迷如何

057 擬去不集閒（常）遮樹早折（折）擬花大宮快樂處須生地獄下波

058 吒莫去死去了卻生來合歡傷爭堪你卻不思量一世似

Figure 1.3 Occasionally, in the project, the marked-up XML file of a text will 're-materialise' in the physical form of a printed edition. As such, the circle of a text from the (physical) manuscript to a digitised facsimile, and then to digital versions in XML and HTML formats, returns to the material world in printed form. The figure here shows the same passage discussed in 1.1 and 1.2 as edited text in Lin, Anderl, Hung 2017, 97[18]

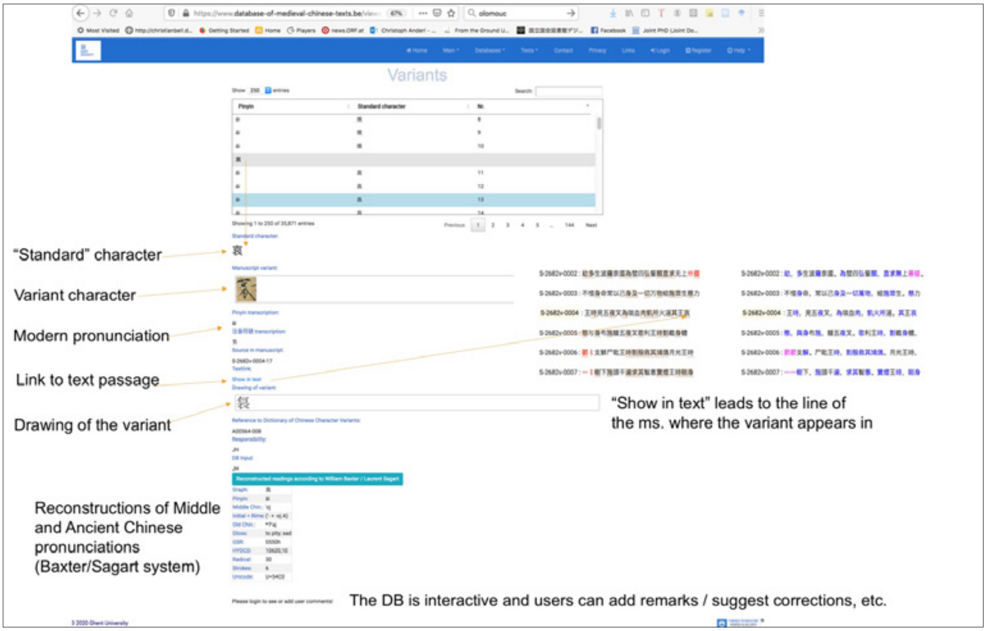
7 The Modules of the DB

7.1 The Variants DB Module

Since several research projects at the department deal with graphical variant forms of Chinese characters as encountered in medieval manuscript texts, the mark-up of the variants has become one of the priorities of the DMCT project. The mark-up is not quite homogenous in this respect, based on the fact that it combines the materials of two projects (i.e. the collaborative project with DILA, and prof. Bingenheimer's previous mark-up of early Chan texts). During the latter project, variants were, whenever possible, cross-checked with the *Dictionary of Chinese Character Variants* (DCCV), and the drawings of those graphs extracted and used in the mark-up (using the unique labels of the graphical forms in DCCV). Variants which were neither found in Unicode nor in the DCCV were newly created as drawings (many of these forms are pending to be included in future versions of Unicode fonts).

[18] For a detailed description of the process of transforming the XML file into a printed edition, please see `https://bit.ly/3sMQpPF`.

The DMCT project has continued to use those drawings whenever possible, however, every ‘new’ variant is extracted from the manuscript as an *image*, and integrated as such in the DB. In addition to the Text module, the Variants module is the most developed part of the project, currently featuring ca. 37,000 variant-text passage relations.[19]



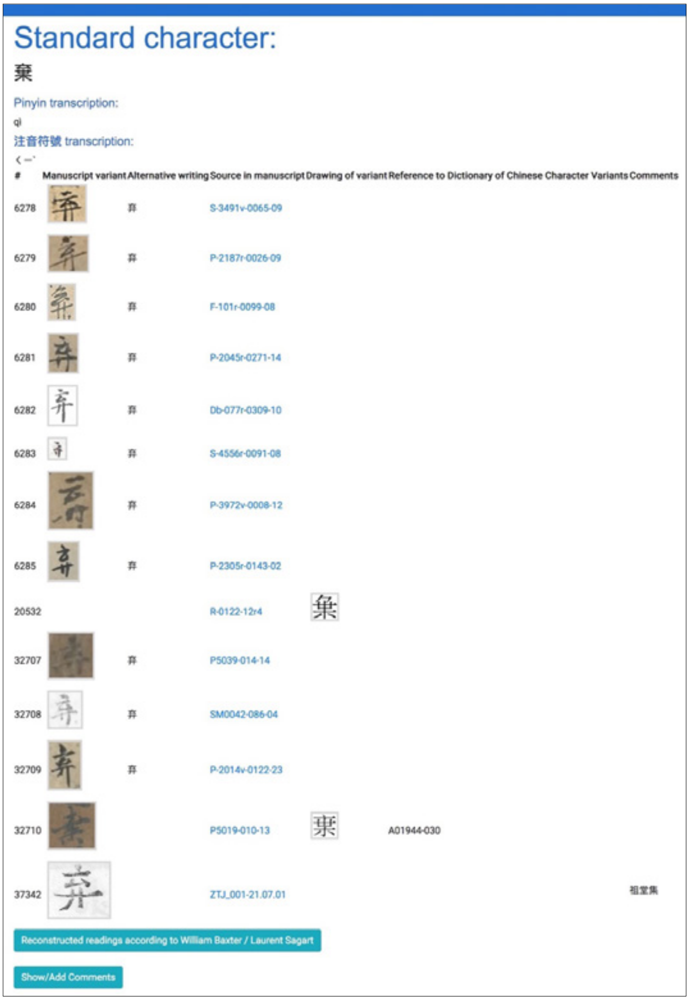
“Show in text” leads to the line of the ms. where the variant appears in

The DB is interactive and users can add remarks / suggest corrections, etc.

Figure 2.1 This is a screenshot of an entry in the Variants module (a variant of character 哀 *āi*), with explanations of the various fields. The “Source in manuscript” field leads directly to the line of the manuscript the variant appears in (exemplified by the text passage to the right). Since recently, the reconstructed readings of Old and Medieval Chinese, based on the system of Baxter and Sagart, are integrated into the Variants Module

[19] In general, we only include variants extracted from Dunhuang manuscripts. However, in the framework of a research project on the ZTJ (a text of crucial importance for studying the vernacular language of the Late Tang and Five Dynasties periods), ca. 1,300 variants were recently input by Laurent van Cutsem (covering the first fascicle of this 20-fascicle work). As a collaborative project with Kyōto University (Zinbun kenkyūjo, Research Institute for Humanistic Studies), the variants are extracted from a digitised version of a unique print of the woodblocks of ZTJ, housed at Haein-sa in Korea (as supplement to the second carving project of the Korean Buddhist Canon in the middle of the 13th century). The textual history of ZTJ – the early parts of which were probably compiled in the middle of the 10th century – is highly complicated. In addition, van Cutsem has recently produced heavily annotated marked-up versions of the two prefaces to the ZTJ (Van Cutsem 2020b, 2020c), as well as to an extensive table and visualisation in Gephi of the lineage system promoted in the text (currently integrated into the DB; see Van Cutsem 2020a).

A very useful feature that enables users to simultaneously view *all registered variants* of a given character was added recently:



Standard character:

棄

Pinyin transcription:

qi

注音符號 transcription:

ㄑㄧˋ

#	Manuscript variant	Alternative writing	Source in manuscript	Drawing of variant	Reference to Dictionary of Chinese Character Variants	Comments
6278		弃	S-3491v-0065-09			
6279		弃	P-2187r-0026-09			
6280		弃	F-101r-0099-08			
6281		弃	P-2045r-0271-14			
6282		弃	Db-077r-0309-10			
6283		弃	S-4556r-0091-08			
6284		弃	P-3972v-0008-12			
6285		弃	P-2305r-0143-02			
20532			R-0122-12r4			
32707		弃	P5039-014-14			
32708		弃	SM0042-086-04			
32709		弃	P-2014v-0122-23			
32710			P5019-010-13		A01944-030	
37342			ZTJ_001-21.07.01			祖堂集

Reconstructed readings according to William Baxter / Laurent Sagart
Show/Add Comments

Figure 2.2 Screenshot of the function of the DB to collect and visualise all the variants of a specific character registered in the Variants module, here illustrated by the variants of the character 棄 *qì* (clicking on the link in the "Source in manuscript" column, the specific variant can also be viewed as part of the text it appears in). The systematic study of variant forms is of great importance for our understanding of medieval writing practices. Whereas in more 'formal' genres (e.g. copies of canonical Confucian texts, Buddhist *sūtras*, official administrative documents etc.) the character forms are frequently adjusted to contemporary 'standard' (正 *zhèng*) forms, semi-vernacular genres are an important source for actual everyday writing practices, often using popular non-standard forms (俗字 *súzì*). From our example here, showing variants of 棄 *qì* from the 8th to the 10th century, it can be deduced that the dominant popular form for this character during that period was actually very similar to its modern abbreviated counterpart (弃)

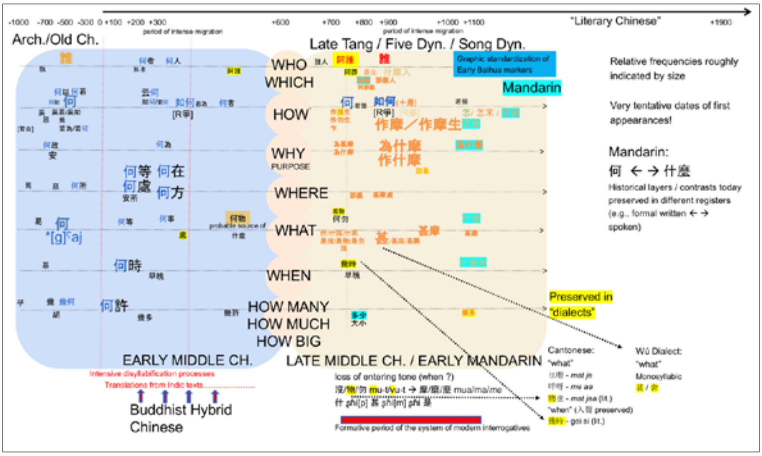
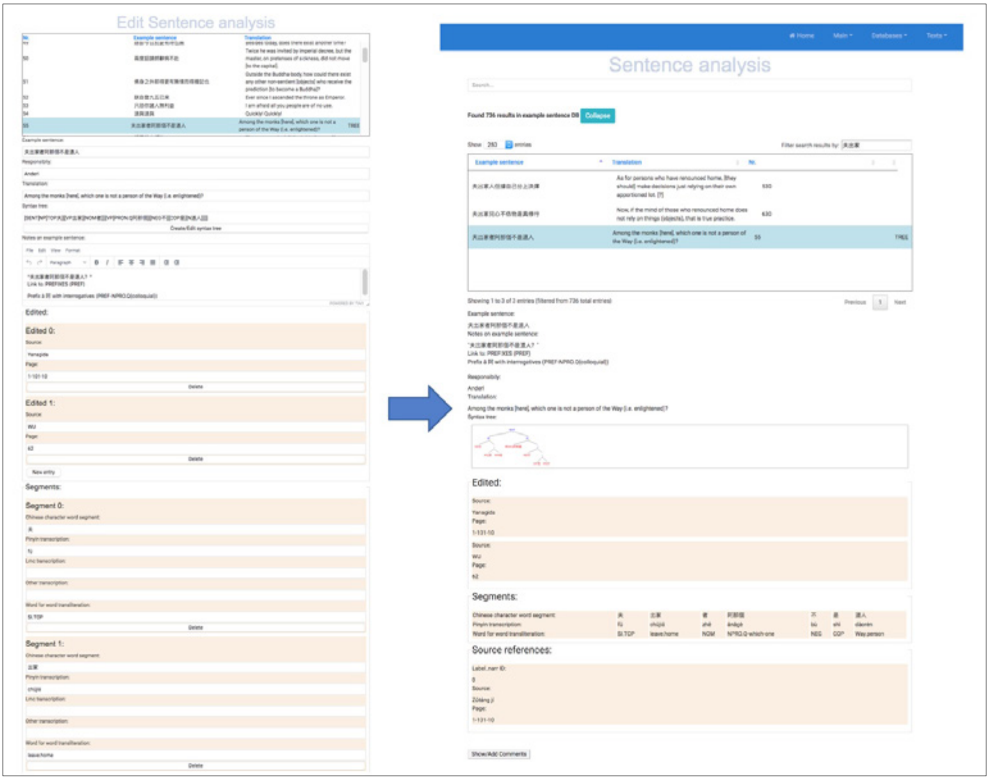

-1000 -700 -500 -300 0 +100 +200 +300
period of intense migration
+600 +700 +800 +900 +1000 +1100
period of intense migration
"Literary Chinese"
+1900
Arch./Old Ch.
Late Tang / Five Dyn. / Song Dyn.
Graphic standardization of Early Baihua markers
Mandarin
WHO
WHICH
HOW
WHY
PURPOSE
WHERE
WHAT
WHEN
HOW MANY
HOW MUCH
HOW BIG
何
*[g]ˁaj
何等
何在
何處
何方
何時
何許
如何
[R爭]
安
作摩／作摩生
為什摩
作什摩
甚摩
甚
什麼
probable source of
EARLY MIDDLE CH.
LATE MIDDLE CH. / EARLY MANDARIN
Intensive disyllabification processes
Translations from Indic texts
Buddhist Hybrid Chinese
loss of entering tone (when ?)
mu-t/vu-t → mua/ma/me
什 ʂhi[p] 甚 ʂhi[m] ʂhi
Formative period of the system of modern interrogatives
Relative frequencies roughly indicated by size
Very tentative dates of first appearances!
Mandarin:
何 ← → 什麼
Historical layers / contrasts today preserved in different registers (e.g., formal written ← → spoken)
Preserved in "dialects"
Cantonese:
"what"
- mat je
- me aa
- mat jsa (lit.)
"when" (入聲 preserved)
- gei si (lit.)
Wú Dialect:
"what"
Monosyllabic

Figure 3 Highly schematic figure of the development of the 'modern' Chinese interrogatives, many of them having their source in the period between 700-1100 (marked with orange colour). The visualisation is based on information extracted from vernacular Dunhuang manuscripts, supplemented with other primary and secondary sources. As can be deduced from the data, a new set of interrogatives started to replace the 何-type system (which appeared frequently in compound form from the early medieval period onward, as evidenced especially in Buddhist texts; the 何 interrogatives are marked with blue colour; the light blue 'box' covers the period of Ancient and Early Medieval Chinese (EMC), before the appearance of the 'modern' interrogatives). By the 10th century, the system of early Mandarin pronouns and their 'standard' orthography had been nearly completely established (marked in light green shading; the beige 'box' covers the period from ca. 700 to 1100, Late Medieval Chinese). Other pronouns evidenced by medieval manuscript material survived in other Chinese dialects (marked with yellow shading). In the figure it is also shown how external features influenced the development and spread of interrogatives, e.g. disyllabification processes since the beginning of EMC, as well as the development of 'Buddhist Hybrid Chinese', a new type of Literary Chinese mixed with vernacular elements and 'Sanscritisms' heavily influenced by translation processes from Indic languages into Chinese. Other external factors include intensive migration events between ca. the 2nd and the 4th century, and then again between the 8th and the 10th century

7.2 Syntax Module

In this part of the DB, information on syntactic markers of Late Medieval Chinese (LMC) are collected. The information on these markers is extracted from texts collected in the Text Module, external text corpora (such as SAT and CBETA), additional Dunhuang manuscript material, as well as relevant secondary literature. The module aims at functioning as a *reference tool*, providing information on the use of LMC function words, their historical development, their orthography as encountered in manuscripts, their relation to other function words etc.[20] The use of the markers is illustrated by example sen-

[20] The fields in the input interface also include information on (historical) pronunciations, notes on variants and phonetic loans used for the marker, dictionary references, as well as references of occurrences in primary and secondary sources. Since the information provided on the function words is still fragmentary, this part of the DB has not yet been opened to the public.



Edit Sentence analysis

No	Example sentence	Translation	
[illegible]	[illegible]	Besides today, does there exist another time?	
50	[illegible]	Twice he was invited by imperial decree, but the master, on pretenses of sickness, did not move [to the capital].	
51	[illegible]	Outside the Buddha body, how could there exist any other non-sentient [objects] who receive the prediction [to become a Buddha]?	
52	[illegible]	Ever since I ascended the throne as Emperor.	
53	[illegible]	I am afraid all you people are of no use.	
54	[illegible]	Quickly! Quickly!	
55	夫出家者阿那個不是道人	Among the monks [here], which one is not a person of the Way [i.e. enlightened]?	TREE

Example sentence:
夫出家者阿那個不是道人
Responsibility:
Anderl
Translation:
Among the monks [here], which one is not a person of the Way [i.e. enlightened]?
Syntax tree:
[SENT[NP[TOP夫]][VP出家][NOM者]][VP[PRON-Q阿那個][NEG不][COP是][N道人]]]
Create/Edit syntax tree
Notes on example sentence:
"夫出家者阿那個不是道人?"
Link to: PREFIXES (PREF)
Prefix ā 阿 with interrogatives (PREF-NPRO.Q(colloquial))

Edited:

Edited 0:
Source:
Yanagida
Page:
1-101-10
Delete

Edited 1:
Source:
WU
Page:
63
Delete

New entry

Segments:

Segment 0:
Chinese character word segment:
夫
Pinyin transcription:
fú
Lmc transcription:
Other transcription:
Word for word transliteration:
SI.TOP
Delete

Segment 1:
Chinese character word segment:
出家
Pinyin transcription:
chūjiā
Lmc transcription:
Other transcription:
Word for word transliteration:
leave.home
Delete

Sentence analysis

Search...

Found 736 results in example sentence DB Collapse

Show 263 entries Filter search results by: 夫出家

Example sentence	Translation	Nr.	
夫出家人[illegible]	As for persons who have renounced home, [they should] make decisions just relying on their own apportioned lot. [?]	530	
夫出家[illegible]	Now, if the mind of those who renounced home does not rely on things [objects], that is true practice.	630	
夫出家者阿那個不是道人	Among the monks [here], which one is not a person of the Way [i.e. enlightened]?	55	TREE

Showing 1 to 3 of 3 entries (filtered from 736 total entries) Previous 1 Next

Example sentence:
夫出家者阿那個不是道人
Notes on example sentence:
"夫出家者阿那個不是道人?"
Link to: PREFIXES (PREF)
Prefix ā 阿 with interrogatives (PREF-NPRO.Q(colloquial))

Responsibility:
Anderl
Translation:
Among the monks [here], which one is not a person of the Way [i.e. enlightened]?
Syntax tree:

Edited:
Source:
Yanagida
Page:
1-131-10
Source:
WU
Page:
62

Segments:

Chinese character word segment:	夫	出家	者	阿那個	不	是	道人
Pinyin transcription:	fú	chūjiā	zhě	ānàgè	bù	shì	dàorén
Word for word transliteration:	SI.TOP	leave.home	NOM	NPRO.Q-which-one	NEG	COP	Way.person

Source references:
Label_narr ID:
0
Source:
Zǔtáng jí
Page:
1-131-10

Show/Add Comments

Figure 4 The left side shows the input mask of the Sentence Analysis Module, featuring a segmentation tool (each segment has fields for the word in Chinese characters, the *pinyin* reading, reconstructed LMC readings, as well as word-for-word transliterations), a tree generator, in addition to several fields for various references (e.g. translation, notes, editions etc.). On the right side of the figure, the HTML transformation of the interface entry into a page of the Sentence Analysis Module of the DMCT is shown. The entries in this module can be linked to the respective entries in the Syntax Module (in the example above, to the entry on prefix 阿 *ā*)

tences (collected in the Example Sentence Module and linked to the respective entries in the Syntax Module), as well as links to the line where they appear in the digitised manuscripts of the DB. The individual entries (currently ca. 700) can also be arranged to form 'chapters' (e.g. on classifiers, or interrogatives etc.), and we aim at developing this feature in our future work on the DB (ideally, this module can eventually be used as a 'reference grammar'). The Syntax Module plays an important role in the department's research on Chinese historical syntax (for an example, see **fig. 3**).

7.3 Sentence Analysis Module

This module is interrelated with the Syntax Module (which is descriptive in nature and records the basic functions and the historical development of a marker) and serves the purpose of illustrating and analysing the functional realms of syntactic markers by presenting examples of their usage in phrases/sentences. The interface contains fields for the example sentence and its translation, notes on the phrase/sentence, a segmentation tool, and the possibility to include a tree analysis **[fig. 4]**.

7.4 Chan Phrases Module

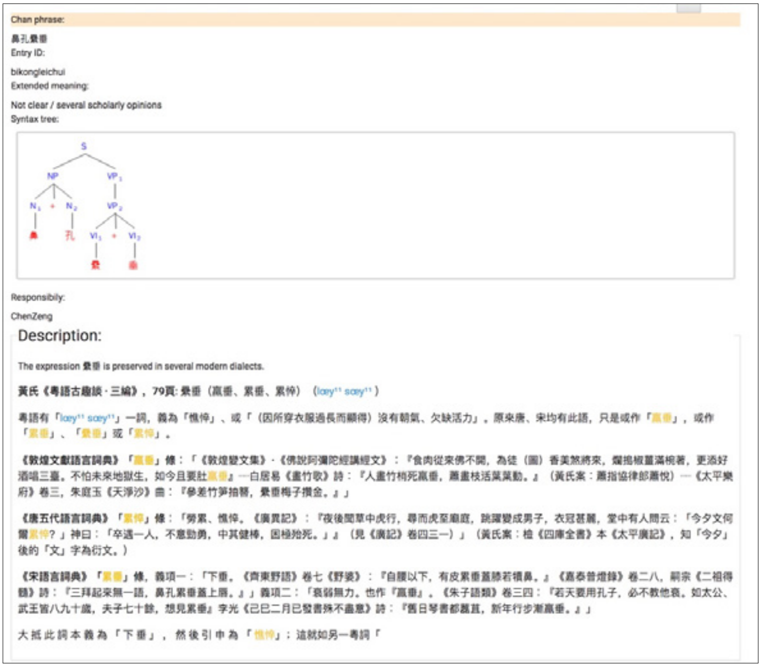
This DB module has been recently added in order to accommodate the results of an ongoing PhD project[21] on the syntax and semantics of 4-character Chan phrases of the Song dynasty, which are often contextually and pragmatically encoded, and the meaning of which is frequently very difficult to retrieve.[22] In addition, these phrases often contain dialect and local vernacular expressions (some of them still preserved in modern dialects), and are as such important sources for the historical development of lexical items.[23] The module aims at collecting these 4-character phrases which play an important role in the rhetorical structure of colloquial Chan texts of the Song and thereafter, register the source texts they appear in, collect referenc-

[21] The material of this module has been mainly collected by Zeng Chen 曾辰 (researcher of Sichuan and Ghent Universities in the framework of a Joint PhD project). Currently, most data are collected in spread sheets, including thousands of Chan phrases with references to their sources. In the further work processes, these data sets will be imported into the Chan Phrases Module. As a sub-project concerning this part of the DB and the research related to it, we will focus on the identification of dialect elements in Chan phrases, as well as try to trace their development from their historical sources to Modern Chinese dialects (the results of this work will be also presented in the form of a joint research paper, currently in production).

[22] In addition, these phrases were often alluded to and commented on in later works, as well as re-embedded in new contexts.

[23] Some of these semantic items spread even 'internationally'; a famous example is 挨拶 *āizā* 'come close and squeeze > to check; to probe' (in the Chan context, often concretely referring to engaging in an exchange of questions and answers about the Buddhist teaching), which first appeared in a Song Dynasty Chan text in the phrase 一挨一拶 *yī ái yī zā* (圓悟佛果禪師語錄 *Yuanwu Foguo chanshi yulu* 'The Recorded Sayings of Chan Master Yuanwu Foguo'; CBETA, T.47, no. 1997, p. 756, b20-5; for another example, see CBETA, T.47, no. 1998A, p. 915, b18-24). After Chan (Jap. Zen) was introduced in Japan during the 12th/13th century, the word 挨拶 *āizā* started spreading beyond the confines of the monastic communities, eventually becoming a high-frequency word with the meaning 'to greet sb. (formally)' (Jap. あいさつ *aisatsu*). In this meaning the word was probably re-introduced to China and is preserved as loanword in the Minnan dialect (*ai*^{35}*sat*^{5}*tsuh*^{3}).

es from historical and contemporary secondary material, analyse their syntactic structure and provide tentative English translations, as well as trace their path of development **[fig. 5]**.



Chan phrase:

鼻孔纍垂

Entry ID:

bikongleichui

Extended meaning:

Not clear / several scholarly opinions

Syntax tree:

Responsibily:

ChenZeng

Description:

The expression 纍垂 is preserved in several modern dialects.

黃氏《粵語古趣談·三編》，79頁: 纍垂（羸垂、累垂、累悴） （lœy^{11} sœy^{11}）

粵語有「lœy^{11} sœy^{11}」一詞，義為「憔悴」、或「（因所穿衣服過長而顯得）沒有朝氣、欠缺活力」。原來唐、宋均有此語，只是或作「羸垂」，或作「累垂」、「纍垂」或「累悴」。

《敦煌文獻語言詞典》「羸垂」條：「《敦煌變文集》·《佛說阿彌陀經講經文》：『食肉從來佛不開，為徒（圖）香美煞將來，爛搗椒薑滿椀著，更添好酒唱三臺。不怕未來地獄生，如今且要肚羸垂』—白居易《畫竹歌》詩：『人畫竹梢死羸垂，蕭畫枝活葉葉動。』（黃氏案：蕭指協律郎蕭悅）—《太平樂府》卷三，朱庭玉《天淨沙》曲：『參差竹笋抽簪，纍垂梅子攢金。』」

《唐五代語言詞典》「累悴」條：「勞累、憔悴。《廣異記》：『夜後聞草中虎行，尋而虎至廳庭，跳躍變成男子，衣冠甚麗，堂中有人問云：「今夕文何獨累悴？」神曰：「卒遇一人，不意勁勇，中其健棒，困極殆死。」』（見《廣記》卷四三一）」（黃氏案：檢《四庫全書》本《太平廣記》，知「今夕」後的「文」字為衍文。）

《宋語言詞典》「累垂」條，義項一：「下垂。《齊東野語》卷七《野婆》：『自腰以下，有皮累垂蓋膝若犢鼻。』《嘉泰普燈錄》卷二八，嗣宗《二祖得髓》詩：『三拜起來無一語，鼻孔累垂蓋上唇。』」義項二：「衰弱無力。也作『羸垂』。《朱子語類》卷三四：『若天要用孔子，必不教他衰。如太公、武王皆八九十歲，夫子七十餘，想見累垂』李光《己巳二月已營書殊不盡意》詩：『舊日琴書都蠹盡，新年行步漸羸垂。』」

大抵此詞本義為「下垂」，然後引申為「憔悴」；這就如另一粵詞「

Figure 5 Screenshot of an entry in the Chan Phrases Module (the phrase 鼻孔累垂 *bíkǒng léichuí*). The entry provides a description of the phrase, a tree analysis, sources in primary texts and references in secondary literature, links to related phrases etc. In addition, occasionally the path of development of semantic items is traced (i.e. the usage in modern Chinese dialects). Here 累垂 *léichuí* is traced to Cantonese *lœy^{11}-sœy^{11}*, which has preserved the original semantic ('to hang; dangle') of the word

8 The DB as a Pedagogical Tool

The above description focused on the DB as a tool for research on medieval Chinese texts. An additional important aspect is the integration of the DB into the teaching environment of advanced master student courses at the Department of Languages and Cultures, Ghent University. The materials provided by the DB are regularly used in classes on Chinese Buddhist texts and culture, as well as for training the students in manuscript decipherment, historical Chinese writing conventions, medieval Chinese syntax and semantics. The materials are also used to compare the Dunhuang Buddhist narratives edited in the DB to their 'canonical' versions, in order to demonstrate how key narratives have been adapted in terms of contents, language, and genre features to specific audiences (e.g. the vernacularisation processes one can observe in many manuscript versions, in order to adapt a narrative to a Chinese general audience).[24] In the master course, students also have to produce annotated translations of selected parts of the specific Dunhuang text discussed during the term. For the future development of the DB, we plan to feed the results of the master courses back into the DB, for example as revised and edited versions of the translations jointly produced by the students.

In addition to training master students in a classroom environment, the DB has also served as the basis for several master theses on Chinese Buddhist texts and/or Medieval Chinese linguistics.[25] Another aspect, which has become increasingly important during the last years, is the possibility to work on the DB in the framework of obligatory internships which master students have to perform as part of their master education (ca. 240 work hours). Most of the work is performed online (e.g. collection of materials, input of the materials into specific modules, analytical work etc.), in addition to regular meetings with the supervisor. This aspect related to the education of master students in the framework of the writing of their theses, as well as the internships,[26] have proven very promising in the development of the DB, and provides the students with an efficient training platform for working with (manuscript) texts; at the same time, it generates manpower for refining and expanding the DB.

[24] As a concrete example, the master course *Buddhism. Texts and Material Cultur*e (MA, Spring 2020) dealt with the conversion story of Nanda (who figures as one of the main disciples of Śākyamuni in Buddhist scriptures), comparing canonical versions with the 因緣 *yīnyuán* genre version preserved among the Dunhuang manuscripts. The students gained reading practice in both Buddhist Hybrid Chinese (i.e. the language of Buddhist translation literature), as well as the semi-vernacular of the Dunhuang manuscript version. In addition to the philological/linguistic aspects, the students would become familiar with various genre features and would analyse the literary structure of the various versions (which emphasise different aspects of the story).

[25] In the most recent master thesis, a student analysed the structure of prepositional phrases based on the data provided by DMCT (Dewaele 2019). Methodologically, the candidate extracted all prepositional phrases from the texts published in the DB, and analysed them comparatively and diachronically, as well as sorted by genre. Another recent master thesis dealing with vernacular Dunhuang materials is van Rentergem 2019, analysing the Buddha biographies of the so-called 八相變 *bāxiàng biàn* genre (transformation of the eight [main] events [of the Buddha's life]).

[26] Internship assignments of 2020-21 will focus on the input of character variants of the earliest period of Dunhuang manuscripts, dating from the mid-fifth and early sixth centuries (see Silk, Galambos 2017), and the comparison of several Dunhuang version of the 搜神記 *Soushen ji* (Records of the Search for the Supernatural).

9 Final Reflections

DBs and digital collections of textual materials have become indispensable tools in the field of corpus linguistics. While typical corpora are repositories of text samples reflecting natural languages, collections of premodern texts necessarily will feature a number of particularities in terms of the selection, gathering, and the preparation of texts, as well as concerning the 'mining' and analysis of linguistically meaningful data. While there are a variety of large digital DBs available for premodern Chinese texts,[27] specialised DBs on non-canonical manuscript materials (which are of paramount importance for research in the culture and language of the Late Medieval period) are still very rare and the information they provide is rather limited. Establishing the DMCT is an attempt to fill this gap, by providing high-quality digital editions of LMC key texts, and develop an analytical apparatus dealing with this type of manuscript material. As described above, the DB also has a 'socio-institutional' function, trying to address the specific research constellation at our department, and providing material for both more Buddhologically oriented, and linguistic studies.

In addition to fulfilling its main function of producing and providing high-quality marked-up medieval text versions, the DB project is driven by specific research interests and topics, and is as such in a permanent state of change and evolution. Accordingly, the DMCT is built as a system of interconnected modules, each module fulfilling a certain function and being embedded in a specific research context (predominantly PhD research projects and international collaborative projects).

In order to widen its significance – justifying the considerable investment of work power and financial resources – the DB has also become an important element in the training of advanced master students, exchange students from China, in addition to being used in the framework of internships. The work invested in the DB in the framework of these pedagogical contexts is also an important source for expanding the scope of the DB by feeding the produced data and research results back into the DB.

[27] In addition to those already mentioned, large DBs suitable for research in Chinese historical linguistics include: www.cncorpus.org, provided by Peking University and including both Chinese modern and premodern text collections; a variety of large text DBs offered by Academia Sinica, Taiwan (http://www2.ihp.sinica.edu.tw/index.php), including the Scripta Sinica Database (which comprises ancient and medieval Chinese texts, consisting of more than 700 million characters); and the huge number of premodern texts provided by CTEXT (https://ctext.org).

Bibliography

Anderl, C. (2018a). “Linking Khotan and Dunhuang. Buddhist Narratives in Text and Image”. *Entangled Religions*, 5, 250-311.

Anderl, C. (2018b). “Metaphors of ‘Sickness and Remedy’ in Early Chán Texts from Dunhuang”. Edzard, Borgland, Hüsken 2018, 27-46.

Anderl C.; Sørensen, H. (2020-21). “Northern Chán and the Siddhaṁ Songs”. Anderl, C.; Wittern, C. (eds), *Chan Buddhism in Dunhuang and Beyond. A Study of Manuscripts, Texts and Contexts in Memory of John R. McRae*. Leiden: Brill, 99-139.

Bingenheimer, M.; Chang P.-Y. (eds) (2018). *Four Early Chan Texts from Dunhuang. A TEI-Based Edition*. Taipei: Shin Wen Feng.

Chen, J. (project director). *From the Ground Up. Buddhism and East Asian Religions*. Vancouver: University of British Columbia. `https://frogbear.org`.

CBETA = *Zhonghua dianzi fodian xiehui* 中華電子佛典協會 (Chinese Buddhist Electronic Text Association). `https://www.cbeta.org`.

DDB = *Digital Dictionary of Buddhism*. Ed. in chief: Charles Muller. Tokyo University. `http://www.buddhism-dict.net/ddb/`.

DCCV = *Dictionary of Chinese Character Variants* (2017). Taipei: Ministry of Education. `https://dict.variants.moe.edu.tw/variants/rbt/home.do`.

DMCT = *Database of Medieval Chinese Texts*. Ghent University, Belgium and Dharma Drum Institute of Liberal Arts, Taiwan. `https://www.database-of-medieval-chinese-texts.be`.

Dewaele, J. (2019). *On Coverbs and Prepositions in Late Medieval Chinese. A ‘Field’ Study and Diachronic Perspective Based on Early Chan and Dunhuang Avadāna Texts* [MA thesis]. Ghent: Ghent University.

Edzard, L.; Borgland, J.W.; Hüsken, U. (eds) (2018). *Reading Slowly. A Festschrift for Jens E. Braarvig*. Wiesbaden: Harrassowitz.

IDP = *The International Dunhuang Project*. `http://idp.bl.uk`.

Lin C.-H. 林靜慧; Anderl, C.; Hung C.-C. 洪振洲 (2017). “*Po Mo bian*” *zhong-ying duizhao jiaozhu* 《破魔變》中英對照校注 [“Po Mo Bian” Critical Edition with Annotated Translations into Modern Chinese and English]. Taipei: Fagu wenhua.

Mair, V. (1983). *Tun-huang Popular Narratives*. Cambridge: Cambridge University Press.

Osterkamp, S.; Anderl, C. (2017). “Northwestern Medieval Chinese”. Sybesma, R. et al. (eds), *The Encyclopedia of Chinese Language and Linguistics*, vol. 3. Leiden: Brill, 218-29.

Rong X. (2013). *Eighteen Lectures on Dunhuang*. Translated by I. Galambos. Leiden: Brill.

SAT = *The SAT Daizōkyō Text* Database. `https://21dzk.l.u-tokyo.ac.jp/SAT/index_en.html`.

Silk, J.; Galambos, I. (2017). “An Early Manuscript Fragment of Dharmarakṣa’s Translation of the **Ajātaśatrukaukṛtyavinodana*”. Edzard, Borgland, Hüsken 2018, 409-31.

Sūn C.-W. 孫昌武; Kinugawa K. 衣川賢次; Nishiguchi Y. 西口芳男 (eds) (2007). *Zutang ji* 祖堂集 (Collection from the Patriarchs’ Hall). 2 vols. Beijing: Zhonghua shuju.

TEI P5 = *Guidelines for Electronic Text Encoding and Interchange*. Edited by the Technical Council of the TEI Consortium. Text Encoding Initiative Consortium, July 2019. `https://tei-c.org/guidelines/p5`.

The Dunhuang Research Academy. Dunhuang yanjiuyuan 敦煌研究院. `http://public.dha.ac.cn/index.html`.

TLS = *Thesaurus Linguae Sericae*. `https://hxwd.org/index.html`.

Van Cutsem, L. (2020a). *The Zutang ji* 祖堂集 *(K. 1503; B25, No. 0144). A Comprehensive.xlsx Table on its Contents and Structure*. Draft Version. Ghent: Ghent University and Database of Medieval Chinese Texts.

van Cutsem, L. (2020b) "Chán Master Jìngxiū's 淨修禪師 Preface to the Zǔtáng jí 祖堂集 (K.1503): A TEI/XML-Based Edition". Database of Medieval Chinese Texts. Ghent University and Dharma Drum Institute of Liberal Arts 法鼓文理學院. `https://www.database-of-medieval-chinese-texts.be/views/texts/zutang_ji/showText.php`.

van Cutsem, L. (2020c). "The Goryeo 高麗 Preface to the Zǔtáng jí 祖堂集 (K.1503): A TEI/XML-Based Edition". Database of Medieval Chinese Texts. Ghent University and Dharma Drum Institute of Liberal Arts 法鼓文理學院. `https://www.database-of-medieval-chinese-texts.be/views/texts/zutang_ji/showText.php`.

Van Rentergem, S. (2019). *A Study of the Dunhuang Baxiangbian. With Annotated Translations* [MA thesis]. Ghent: Ghent University.

Zutang ji 祖堂集 (Collection from the Patriarchs' Hall). Scanned Copy of an Original Print of the second *Goryeo Daejanggyeong* 高麗大藏經 Woodblock Edition (1245) of the *Zutang ji* Stored at the Library of the Institute for Research in Humanities of Kyōto University, Japan.

Zutang ji 祖堂集 (Collection from the Patriarchs' Hall). Digital versions of the text: `https://raw.githubusercontent.com/cbeta-org/xml-p5/master/B/B25/B25n0144.xml` and `https://cbetaonline.dila.edu.tw/zh/B0144`.