



Ground Truth Metadata

Project Acronym	TibSchol
Project Number	101001002
Authors	Rachael Griffiths, Pascale Hugon
Document version	1.0
Date	20.12.2023
License	CC0-1.0

Title	Ground truth data for handwritten Tibetan cursive scripts (Drutsa and Betsug)
Alternative Title	Tibetan Cursive Ground Truth
Publication Date	20/12/2023
Version	1.0
URL	https://github.com/ERC-TibSchol/Tibetan-cursive-GT
Author	Rachael Griffiths
Point of Contact	Rachaelgriffiths1@gmail.com
Description	<p>This ground truth data was originally used to train two HTR models on Transkribus, using PyLaia. One model (Transkribus ID:54525) transcribes the handwritten Tibetan cursive script known as Drutsa ('<i>bru tsha</i>), the other (Transkribus ID:54935) transcribes Betsug (<i>dpe tshugs</i>).</p> <p>The Drutsa model was trained on ground truth data derived from 466 folios. These folios are selected from a diverse range of 19 Tibetan treatises under exploration in the TibSchol project (see Ground Truth Overview.csv). The training set comprises 422 folios, while an additional 44 folios form the validation set.</p> <p>Built upon the foundation of the Drutsa model, the Betsug model was trained on 93 folios drawn from a curated selection of 8 Tibetan treatises. The training set comprises 85 folios, with an additional 8 folios forming the validation set.</p> <p>The transcriptions adhere to the extended Wylie transliteration system in the Roman alphabet.</p> <p>Abbreviations are faithfully transcribed in line with their appearance in the manuscripts. Refer to the list of abbreviations for details.</p> <p>Interlinear and marginal notes are mostly excluded from the ground truth data. The model excludes the recognition of markers and signs, with the exception of <i>yig mgo</i> (ལྷོ), <i>shad</i> (།), and <i>sbrul shad</i> (།). The <i>yig mgo</i> is only included in transcriptions when a new text begins, not at the start of each new folio.</p>
Data type	.txt files. Images are not included. A table is provided to match transcriptions with corresponding images.
Keyword	Handwritten Text Recognition; Tibetan language; <i>dbu med</i>

Related Publications	Griffiths, Rachael (2022) Transkribus in Practice: Improving CER, <i>The Digital Orientalist</i> . doi: 10.1553/TibSchol_ERC_Cog_101001002_Griffiths_CER Griffiths, Rachael (2022) Transkribus in Practice: Abbreviations, <i>The Digital Orientalist</i> . doi: 10.1553/TibSchol_ERC_Cog_101001002_Griffiths_Abbreviations
Language	Tibetan
Production Date	21/08/2023 (Drutsa) and 08/09/2023 (Betsug)
Production Location	Vienna, Austria
Contributor	TibSchol PI: Pascale Hugon
Funding Information	ERC H2020 CoG No. 101001002. Project TibSchol “The Dawn of Tibetan Buddhist Scholasticism (11 th -13 th c.).” This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 101001002).
Data Type	Ground Truth
License	CC0-1.0
Software	Transkribus, Version: 1.26.0
Data Source	<p>The ground truth data for these models encompass a rich compilation of transcriptions from Tibetan treatises (see Ground Truth Overview.csv). The transcriptions in the ground truth data consist of original transcriptions by Rachael Griffiths and of portions of pre-existing available transcriptions that have been adapted, completed, and proofread. The latter include the following, as well as transcriptions resulting from a combination of the following:</p> <ul style="list-style-type: none"> • transcriptions from Pascale Hugon and Kazuo Kano’s project “A Gateway to Early Tibetan Scholasticism” • transcriptions from Kyeongjin Choi’s project “Construction of Basic Materials for the Study of the Kadam School of Tibetan Buddhism” • transcriptions from Yoichi Fukuda’s project “Basic research for the history of early Tibetan logic” previously available at http://tibetan-studies.net/ (last accessed 2.9.2020) • unpublished draft transcriptions by Pascale Hugon, Kazuo Kano, Ritsu Akahane • diplomatic version of a critical edition by Kazuo Kano and Kazuhiro Kawasaki in “リンチェンサン ポ著 『チャクラサンヴァラ・アビサマヤ注』 —蔵文校訂テキスト—,” 高野山大学論叢 49 (2014): 1–36.