



Universiteit
Leiden

The Netherlands

Teaching handwritten text recognition: Can new technologies save old skills?

Caers, B.J.M.

Citation

Caers, B. J. M. (2024). Teaching handwritten text recognition: Can new technologies save old skills? *Quaerendo. A Journal Devoted To Manuscripts And Printed Books* , 54(2-3), 198-209.

doi:10.1163/15700690-bja10024

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/4303999>

Note: To cite this publication please use the final published version (if applicable).



BRILL

QUÆRENDO 54 (2024) 198–209

Quaerendo

brill.com/qua

Teaching Handwritten Text Recognition

Can New Technologies Save Old skills?

Bram Caers | ORCID: 0000-0001-7482-3198

Assistant Professor of Middle Dutch Literature and Postdoctoral Research
Fellow (NWO-veni, 016.Veni.195.371), Leiden University, Leiden,
The Netherlands

b.j.m.caers@hum.leidenuniv.nl

Received 6 September 2023 | Accepted 15 March 2024 |

Published online 18 June 2024

Abstract

In the past decades, the development of Handwritten Text Recognition (HTR) has made considerable progress. One of the most user-friendly platforms providing access to this technology to scholars is Transkribus. Since the launch of this platform not over eight years ago, it has appealed to a steadily growing community of users. From a scholarly point of view, the platform and its technology have predominantly been put to practice in historical research and in providing access to material kept in archives. Philologists, codicologists and book historians have been slower in embracing the technology, but are now discovering its advantages. This contribution discusses how HTR technology and the Transkribus platform can be implemented in teaching, within the broader context of the evolution towards digital scholarly text editions.

Keywords

Transkribus – HTR – paleography – Medieval manuscripts – text editions

For some time now, libraries and archives across the world have been stepping up their digitisation initiatives.¹ The covid pandemic only boldened these efforts, further pushing institutions to provide digital, remote access to (some of) their holdings at a time in which physical access was restricted. With the improved availability of manuscript material online come new challenges: leading researchers to relevant content requires the material to be searchable, not only on a level of metadata or through the (manual) attribution of certain keywords, but also in terms of *full text*.² With the rapid development and implementation of optical character recognition (OCR) technologies, printed texts are becoming ever more accessible, for example through initiatives such as Google Books (even if the OCR result for printed text, especially from the handpress period and in other type fonts than those commonly used today, continues to be far from perfect).³ Manuscript material, however, was lagging behind. In order to provide easier, semi-automated access to text preserved only in manuscript form, digital humanities scholars and computer scientists have been developing handwritten text recognition (HTR), a technology that takes the basic principle of OCR to a new level of machine learning.⁴

While HTR technology is being developed in different places, it is fair to say that one of the most successful platforms offering easy access to it for a wide audience of scholars, is Transkribus.⁵ Initially developed within an EU-funded

-
- 1 This introduction builds on Guenter Muehlberger et al., 'Transforming Scholarship in the Archives Through Handwritten Text Recognition. Transkribus as a Case Study', *Journal of Documentation*, 75.5 (2019), pp. 954–76, pp. 955 ff. The authors of this contribution could of course not take into account the influence of the pandemic on digitisation efforts. According to Annemieke Romein et al., 'State of the Field. Digital History', *History. the Journal of the Historical Association*, 105 (2020), pp. 291–312, 'mass-digitisation of libraries and archives [has been] under way since the 1990s' (p. 292).
 - 2 Muehlberger et al., 'Transforming Scholarship in the Archives Through Handwritten Text Recognition', pp. 955–56.
 - 3 The increased availability of machine-readable text material on Google Books also has a downside. For a wrap-up of scholarly discussion on this issue, see Romein et al., 'State of the Field. Digital History', pp. 304–06.
 - 4 For a short overview of the impact of both technologies, see for example Romein et al., 'State of the Field. Digital History', pp. 294 ff. For a more technical discussion about the similarities and differences between OCR and HTR, see Muehlberger et al., 'Transforming Scholarship in the Archives Through Handwritten Text Recognition'.
 - 5 Compare Romein et al., 'State of the Field: Digital History'; Joe Nockels et al., 'Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts. A Systematic Review of Transkribus in Published Research', *Archival Science*, 22 (2022), pp. 367–92, pp. 368–69 and Muehlberger et al., 'Transforming Scholarship in the Archives Through Handwritten Text Recognition', pp. 956–57 and *passim*. The latter contribution is not entirely unprejudiced, as it stems from the research group that developed the programme. The

(Horizon 2020) research project (READ) headed by the University of Innsbruck, Austria, the programme now is a paid service within a cooperative structure, which uses the revenue it generates to innovate its platform and technologies.⁶ It features a desktop ‘Expert Client’ to be installed on your computer as a separate application, as well as a browser version that is much easier to navigate, but does not as yet feature the full potential of the programme. Both versions, however, are relatively easy to master, and they are splendid examples of how the most advanced technology that digital humanities scholars are producing, is made accessible for scholars who are not programmers or computer scientists themselves. Over the years, Transkribus has grown extensively, and now counts over 100,000 users worldwide. Their input helps make the technology more precise, as thousands of users are working with all kinds of handwritten as well as printed material – from Medieval to Modern – in very different languages and scripts. Based on an input of human transcriptions – a few dozen pages will suffice – the computer develops a reading model that can be more precise, and that will certainly be quicker, than handing the transcription work out to humans. These HTR-models can then be shared with the scholarly community, a practice that makes HTR technology readily available to all users, even those who do not train their own models.

With an ever-growing community of users, both individual users and institutional ‘members’, Transkribus is leading the way in the field of handwritten text recognition. Scholarly interest for the technology, and its usage in research, keep increasing at a steady pace. While the technology initially was developed for archives, and was picked up enthusiastically by historians and archivists, literary historians are increasingly seeing its potential.⁷ Gradually, Transkribus

authors do raise a valuable point in stressing the community-based development and finality of Transkribus.

6 For an outline, see Muehlberger et al., ‘Transforming Scholarship in the Archives Through Handwritten Text Recognition’, pp. 957–58. The authors of this paper could not foresee the present cooperative structure. For more information on this, see the READ COOP website (<https://readcoop.eu/>), last accessed 15 March 2024.

7 There is no lack of inspiring and successful projects implementing HTR technology through the Transkribus platform. By way of example, for the Dutch context I refer to the *Chronicling Novelty* project, working on hundreds of unknown early modern chronicles to discover perceptions of novelty (headed by Erika Kuijper and Judith Pollmann); and the *Alle Amsterdamse Akten* project, an initiative of herculean proportions, to digitise and make accessible some 3,5 kilometres of Amsterdam notarial deeds. For a comparison of research from a historical or literary historical perspective, see Nockels et al., ‘Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts’, p. 380. A recent example from Middle Dutch scholarship that showcases the potential of HTR technology for philological research, is Wouter Haverals and Mike Kestemont, ‘From Exemplar to Copy. The Scribal

(and HTR more generally) are being integrated into new research projects, which are bearing ever more fruit. Scholarly work addressing Transkribus itself, rather than using the programme as a means to an end, has focused very much on the technology behind the platform and its direct implementation in research. Contributions focusing on a user perspective, especially in an educational context, are few and far between.⁸ Instead of taking such a technological or research-oriented viewpoint, this paper addresses the potential of the programme in teaching, from the perspective of producing scholarly text editions. In the following pages, I will briefly outline the context of HTR and Transkribus, after which I will report from a course taught at Leiden University, where students have been working with Transkribus for the past four years.⁹ Discussing the advantages and disadvantages of this process and the lessons learned over the years, I hope, will inspire fellow teachers to engage with HTR in their classes.

Will HTR Save Text Editions?

When in the nineteenth century, the Royal Library of Belgium (KBR) acquired the book collection of the important collector Charles van Hulthem, not all of the books listed at auction were transferred immediately to the Royal Library's holdings. It would take some time before the entire collection was in Brussels, as some of Van Hulthem's manuscripts were dispersed within a network of interested friends and colleagues. Among these 'loans' was a manuscript that is nowadays considered one of the most important books in the history of Dutch literature, and that has taken the name of its last private owner in scholar-

Appropriation of a Hadewijch Manuscript Computationally Explored', *Journal of Data Mining and Digital Humanities* (2023).

- 8 Nockels et al., 'Understanding the Application of Handwritten Text Recognition Technology in Heritage Contexts', pp. 376–80. While writing this article, the READ COOP behind Transkribus published an interesting update on educational use of the platform, see 'Transkribus in Education', 7 June 2023 (<https://readcoop.eu/transkribus-in-education/>), last accessed 22 February 2024. For the fourth iteration of the Digital Editing class, the author was granted a stipend in the form of Transkribus credits within the Transkribus Scholarship Programme.
- 9 This text is an elaborated version of a blog related to this subject, see Bram Caers, 'Automated Reading of Medieval Manuscripts. An Alternative for Palaeography Classes?', *Leiden Medievalists Blog*, 20 July 2022 (<https://www.leidenmedievalistsblog.nl/articles/automated-reading-of-medieval-manuscripts-an-alternative-for-palaeography-classes>), last accessed 22 February 2024. The approach here is broader and aimed more at scholarship generally, while the blog focused on the educational use of Transkribus.

ship: the ‘Hulthem Manuscript’. Renowned mainly for containing the earliest worldly plays in any European language, this book is a veritable treasure trove for Middle Dutch literature, because it contains a myriad of different texts that are known only from this text witness.¹⁰ Philologists in the course of the nineteenth century were well aware of the importance of this manuscript. In 1828, Van Hulthem gave it to his close friend C.A. Serrure, who intended to produce a scholarly edition. In 1832, J.F. Willems had the chance to copy and publish some of its contents, but it appears that the manuscript remained in the possession of Serrure. When Hulthem’s collection was auctioned off to the Royal Library, in 1836, the precious book remained there, much to the annoyance of other scholars interested in its contents. In the meantime, material from the manuscript was published here and there, but based mainly on loose copies made by various scholars. The manuscript was finally transferred to Brussels only after pleas by J.F. Willems to the Minister of Internal Affairs to force Serrure to hand it over. And subsequently, Willems was allowed to study it in the privacy of his own home in Ghent.

To our modern eyes, anecdotes such as these seem rather quaint. Manuscripts nowadays are cherished heritage objects that can be consulted only by trained professionals, and that leave custodian institutions only rarely, and if so, heavily insured and accompanied by staff to make sure that they are handled with care and kept in precisely the right conditions. The very idea that for the purpose of making an edition, one would simply loan a medieval book from a custodian collection, is ludicrous. Or is it? Advanced digitisation in a way means a return to the relatively free access to manuscripts for scholars in the past, allowing researchers to work on the manuscripts of their choice from the comfort of their own desks. Even more so: the increased ‘interoperability’ of digitised collections, for example through the IIF framework, makes for the fact that scholars can visually align manuscripts from different collections on their screens for their research purposes, as if they would have the books lying side by side on their desks. And while Serrure and Willems quarrelled over having the manuscript for themselves, scholars now of course can work on the same material simultaneously, all across the world. Whereas

10 See the introduction to the authoritative edition of this manuscript by *Het handschrift-Van Hulthem. Hs. Brussel, Koninklijke Bibliotheek van België, 15,589–623*, eds. Herman Brinkman and Janny Schenkel (Hilversum: Verloren, 1999). Most of the material from this paragraph is based on their history of the manuscript, pp. 16 ff., but with nuances and corrections based on correspondence on the subject held in the Royal Library, which is being studied by Jan Pauwels (KBR, Brussels). I thank him for sharing his notes on this correspondence, on which he is preparing a more extensive publication.

the nineteenth-century scholars had to activate personal networks and access was often dependent on their scholarly reputations, digital access nowadays is available to anyone who is able to find what they are looking for, and is anonymous. Modern scholars are way ahead of their nineteenth-century predecessors, one would say, if it were not for the great leveller: medieval handwriting.

Generations of students in the footsteps of their nineteenth-century predecessors, have sat grudgingly peering over small medieval letters, trying to decipher the writing into a legible transcription. While the skill of palaeography is associated most with historians, students of languages too, have to learn how to read medieval handwriting and early printed type, in order to study the material sources of their discipline directly, instead of through curated intermediaries such as text editions or source books. Seemingly evident a skill for a trained (literary) historian, palaeography as a discipline has been under pressure in organising university curricula. Like other auxiliary sciences that have a whiff of the past about them – which student is still deftly introduced into the noble art of heuristics? – palaeography is likely to lose its status as a compulsory course, becoming a subject that is optional at best, if offered at all.¹¹ It would seem that planning committees estimate that reading medieval and early modern handwriting is a skill that is becoming superfluous to the modern academic, as digital tools and increased availability of primary text material are rapidly taking away the necessity of training students to transcribe handwritten texts themselves.

At the same time, certainly in literary historical scholarship, researchers have again turned to the primary sources of the literary texts they study, in the wake of ‘Material Philology’ and the various guises this renewed attention for the manuscript as a material carrier of meaning took on in the last decades of the twentieth century. Scholars realised that some of the nineteenth-century text editions they had been working with did not sufficiently reflect the individuality of the medieval material sources they were based on, ignoring the historical contexts of specific manuscripts and their makers and readers. The renewed attention for the material context of medieval literature of course requires scholars to master relatively old-school skills, such as codicology, palaeography, and philology. In Middle Dutch Studies, the ‘material turn’

11 This evolution has even been picked up by news media outlets such as *NRC* or *The Guardian*, see Theo Toebosch, ‘De middeleeuwen zijn onleesbaar geworden’, *NRC*, 13 May 2022 (<https://www.nrc.nl/nieuws/2022/05/13/de-middeleeuwen-zijn-onleesbaar-geworden-a4-124327>), last accessed 22 February 2024; and compare John Grace, ‘Writing off the UK’s last palaeographer’, *The Guardian*, 9 February 2010 (<https://www.theguardian.com/education/2010/feb/09/writing-off-last-palaeographer-university>), last accessed 22 February 2024.

heralded a new edition series in which the focus was on miscellanies or *Sammelbände: Middeleeuwse Verzamelhandschriften uit de Nederlanden (MVN)*. The intention was to do justice to the individuality of a text collection within one manuscript, rather than artificially putting the spotlight on a single text, without acknowledging the fact that it is accompanied by others and takes its meaning partly from that specific material context.¹² The rigid focus on one manuscript miscellany required lucid editorial decisions to reflect the unity of the text witness that was being edited: in principle, all editions of the (*MVN*) series were diplomatic, following an extensive and authoritative set of guidelines that continues to weigh in on text editing initiatives in Middle Dutch Studies.¹³

While the *MVN* series definitely set a standard for text editing in Middle Dutch Studies, its editors recently announced the end of the series.¹⁴ The steady pace at which the series produced new volumes in the past had slowed down considerably, and the last four volumes were published online exclusively. Digital text edition of course heralds new possibilities, but at the same time, many digital editions published in the recent past turn out to become obsolete or ephemeral relatively quickly, because standards change so rapidly.¹⁵ There are of course various reasons why the editors of *MVN* chose to pursue a new course, and this is not the place to go into them in any detail, but it is fair to say that one of the challenges facing the series is a general problem for literary historical scholarship. Text editions as a form of publication are not valued enough as a product of scholarly activities, and scholars are pushed towards

12 See the proceedings of a paradigmatic conference on this topic: *Middeleeuwse verzamelhandschriften uit de Nederlanden: congres Nijmegen, 14 oktober 1994*, ed. Gerard Sonnemans (Hilversum: Verloren, 1996); and the introduction to the first volume in the series: *Het Geraardsbergse handschrift. Hs. Brussel, Koninklijke Bibliotheek Albert I 837-845*, ed. Marie-Josée Govers et al. (Hilversum: Verloren, 1994).

13 Thom Mertens et al., 'Richtlijnen voor de uitgave van Middeleeuwse Verzamelhandschriften uit de Nederlanden', in *Het Geraardsbergse handschrift*, ed. Govers et al., pp. 173–191.

14 The festive presentation of four new digital volumes, on 2 October 2023 in Amsterdam, also featured the announcement of an end to the series. Reflections on this topic are expected to be published in *Queeste* and in *Spiegel der Letteren*.

15 See from an introductory point of view: Matthew J. Driscoll and Elena Pierazzo, 'Old Wine in New Bottles?', *Digital Scholarly Editing. Theories and Practices*, eds. Matthew J. Driscoll and Elena Pierazzo (Cambridge: Open Book Publishers, 2016), pp. 1–18; and the contribution by Greta Franzini, Melissa Tarras and Simon Mahony in this volume; and compare Steven W. May, 'All of the Above. The Importance of Multiple Editions of Renaissance Manuscripts', *Literature Compass*, 7.2 (2010), pp. 95–101.

other, more rewarding forms of output.¹⁶ Editions of course take considerable time, and scholars struggle to get edition projects funded, as financing bodies are looking for novelty, for fresh ideas and groundbreaking, paradigm-shifting research projects. While text edition as a scholarly pursuit certainly witnesses its own innovation – especially in a digital sense – the harsh reality is that the current climate blocks scholars from taking on ambitious edition projects to replace the hallmark text editions of canonical texts that were made in the nineteenth century.

Back to Transkribus. Could it be that this platform, with its easy access to the technology of Handwritten Text Recognition (HTR), solves two problems in one strike? While compensating for the relatively weak(ening) palaeographical skills of its users, it would allow them to transcribe huge amounts of text material, saving considerable time in the first step of producing a text edition. Automated transcripts would only need to be collated against a material source to become a valuable basis for a diplomatic edition. In principle, users need to transcribe a relatively small set of pages (a few dozen will suffice) as training material to produce an HTR model that is specific to a single hand, for example for one scribe working in a single book hand in the middle of the fourteenth century. An effective model, with a *Character Error Rate* (CER – the amount of mistakes per 100 characters) under 5%, takes a few hours to train but will automatically transcribe a new manuscript page in a matter of seconds. Collating for mistakes of course takes time, but arguably less than making a full transcription from scratch.

In a following step, HTR models trained on specific hands can be combined into more general models that in theory will tackle any kind of handwriting that is fed to them. These will of course be more accurate when they combine similar handwriting (e.g. several hands from the second half of the fourteenth century). But as the technology progresses, and more users make their HTR models available to other scholars, it will increasingly become possible to produce powerful models that can read any kind of handwriting from several centuries, and in different languages. Transkribus itself, incidentally, is leading the way by producing so-called ‘Transformer models’, HTR models that have been trained on stupendous amounts of data from different periods and languages, harvested from the accounts of their growing community of users, and that in theory will become generally applicable to any material that is fed to

16 Elena Pierazzo has noted the lack of status even for digital editions, which ‘have not won recognition as an authoritative expression of scholarship’, ‘What Future for Digital Scholarly Editions? From Haute Couture to Prêt-à-Porter’, *International Journal for Digital Humanities*, 1 (2019), pp. 209–20, p. 210.

them in the near future.¹⁷ Powerful HTR models such as these will make it altogether unnecessary for scholars to produce training data themselves. When editing a text, it will simply suffice to feed a digitised version of a manuscript to Transkribus, and then move on to collating and correcting mistakes.

In the very near future, producing text editions will therefore become much more time-efficient, as scholars can skip the entire transcription phase. A next step would be to offer an automatically generated diplomatic transcription (i.e. without the intervention of scholars) as an extra information layer in manuscript databases and inventories that provide curated access to digitised images and metadata. New projects that are being developed, such as eCodicesNL or MMFC (to restrict myself to the Low Countries) should definitely keep an open mind to these possibilities (and are doing so). Imagine the prospect of generating a machine-readable library of *all* digitised manuscript content, and the research possibilities that such a database would open. This would revolutionise the way in which scholars access material kept in manuscript form, quite in the same way that Google Books revolutionised searching through printed material. While automated transcripts at this point are perhaps too faulty, I am fairly confident that their quality will increase at a steady pace, so that a ‘Google Manuscripts’ may well be materialising on the horizon.¹⁸ A glimmer of this future can be seen in the announcement that Transkribus teamed up with Wikimedia, making HTR technology available for the automated transcription of handwritten sources on Wikisource, a global collection of primary sources in the public domain.¹⁹ An integration of Transkribus’ powerful HTR technology into a globally available and open platform such as Wikisource, shows the contours of a future in which automated transcriptions of any handwritten text material available in digital form are at users’ fingertips.

17 ‘Introducing Transkribus Super Models – Get Access to “The Text Titan I”’, 10 July 2023 (<https://readcoop.eu/introducing-transkribus-super-models-get-access-to-the-text-titan-i/>), last accessed 22 February 2024. After completion of this contribution, Transkribus launched a new powerful transformer model for Dutch material, ‘Dutch Demeter’.

18 See Bram Caers, ‘Middel nederlandse en historische letterkunde na de ‘turbo-mediëvis-tiek’. Dobberen in het kielzog of een nieuwe koers?’, *Queste*, 30.1 (2023), pp. 75–99, pp. 89–90, where this idea is developed from the point of view of Middle Dutch studies.

19 ‘Preserving Cultural Heritage. Transkribus Integration with Wikimedia Projects’, 5 October 2023 (<https://readcoop.eu/preserving-cultural-heritage-transkribus-integration-with-wikimedia-projects/>), last accessed 22 February 2024.

Teaching HTR: Can New Technologies Save Old Skills?

For four subsequent years now, in the context of a class on ‘Digital Editing of Literary Texts’ at the University of Leiden, students have been introduced to *Transkribus*.²⁰ In an introductory session, they acquire the basic skills to navigate the programme and to produce basic transcriptions. Working with manuscript and print material, students learn how to transcribe the text, tag the transcription with relevant information (on material conditions, the use of abbreviations, etc.), working towards a set of transcriptions that can be used as a basis for training a new HTR model. Their *Transkribus* work is embedded in a more theoretical discussion of scholarly literature on the history of text criticism and editing throughout the ages, with a focus on how recent scholarship deals with increased digital availability of primary text material, including the development of metadata databases, and readymade tools such as *Transkribus*.

For students, the theoretical context is helpful to see *Transkribus* and HTR technology as a next step in a long line of scholarship dealing with primary text material. Moreover, they are motivated by the fact that they are actively taking part in training new HTR models that subsequently can be shared with other scholars and as such be helpful in future research. Students were eager to produce a sufficiently large set of precise transcriptions to help train the computer models. Combining transcriptions of print and manuscript sources stimulated discussions on how both types of script varied, in the eyes of humans and in the eyes of computers, and how this would affect the results of the HTR model, in a similar way to how it affected the time they spent transcribing themselves. While going over the transcriptions during group feedback sessions, students became aware of how important it is to work in a ‘hyper-diplomatic;’ way, so as not to confuse the computer model when it is comparing transcriptions to the image of the manuscript. The various choices that the group had to make, are not dissimilar to questions related to classical text edition (how to deal with abbreviations, material damage, poorly readable text, scribal errors, etc.). In this way, students become familiar with relatively ‘old-school’ skills in transcribing primary text material and progressing towards a diplomatic edition, while working in an environment that is state of the art, and that will form the basis of an HTR model based on machine learning. This combination of old-school skills and modern technology strongly appeals to students.

20 See Caers, ‘Automated Reading of Medieval Manuscripts’.

At the end of the class, students are confronted with the HTR models that were trained based on their own transcriptions, and are baffled at the results. The group's efforts in the past four years each time led to models performing at well below 4% CER for manuscript sources, and even under 2% CER for print. While at that rate, the manuscript models make about the same amount of errors as a trained student, the profit is less in accuracy than it is in time. The computer model takes a few dozen seconds to transcribe a full page of manuscript material, while even the more talented students would need anywhere between 15 minutes and several hours to produce a transcription at par with what the computer model delivers. The fact that Transkribus *works* of course is no new information – what I would like to stress here is that it not only *works* in a technical sense, but that it *works wonders* in teaching. It makes students aware of the (growing) power of HTR technology, while also lending them a sense of pride that the computer model was developed based on their concentrated labour in painstakingly producing detailed transcriptions.

Another interesting observation is the difference in the mistakes made by students on the one hand, and the computer on the other. While trained students and the computer would be close rivals in terms of character error rate quantitatively speaking, their respective errors are very revealing. Students would unknowingly 'modernise' the specifically Middle Dutch spelling of a medieval source, for example: transcribing modern Dutch *daar* for Middle Dutch *daer* (there). This produces a perfectly legible text, albeit in a spelling that would have seemed strange in medieval eyes. The computer, on the other hand, much more likely produced nonsensical readings, unhindered as it is by any knowledge of modern (or indeed medieval) Dutch. This comparison too, turned out to provide food for discussion, especially in terms of suggestions to improve the computer's performance (e.g. by linking it up with online dictionaries and lexicographical sources for Middle Dutch).

All in all, working with Transkribus turned out to be a stimulating way of revitalising discussion on the process of text editing, through the refreshing new perspective of HTR technology. Students learned to understand the principles behind HTR, and the relation between the human input needed for the model to 'learn', and the very unhuman mistakes it keeps making. Students realise that the technology can reach impressive results, but that it relies on their own precision, and therefore requires them to master the skill of palaeography themselves. While becoming acquainted with a specific tool that can be useful for their research and future professional careers, students therefore also train 'old-school' skills associated with the age-old science of text editing: palaeography, collation, understanding how text variants between different

manuscripts work, etc. And on top of that, there are less tangible skills at play that are difficult to measure: patience, precision, collaboration ...

The more difficult step is to frame the very practical work that students are doing in producing an HTR model, within the general process of working from a primary manuscript to a text edition. Ideally, students would produce a small text edition based partly on HTR-generated transcriptions that they have collated themselves. While these follow-up steps were not part of the Leiden course in its current set-up, working with Transkribus turned out to be a valuable experience to students, especially when discussing the topic of text editions. And what is abundantly clear, is that HTR technology, when used correctly, will save large amounts of time for future scholars producing text editions. As more models become available to the public, the necessity of producing basic transcriptions for the purpose of training an HTR model will dwindle. And who knows that in a few years' time, it will not become possible to browse computer-generated diplomatic transcripts of any manuscript that has been made available online. If that goal is ever reached, the Leiden students can be proud to have been part of the process, by sharing their transcription with Transkribus, and the world.