Journal of Data Mining and Digital Humanities

EPI sciences overlay journals

Machine transliteration of long text with error detection and correction

Mohamed Abdellatif[1,2], **Joel U. Bretheim**[1], **Marina Rustow**[2]

[1]Research Computing, Princeton University
[2]Department of Near Eastern Studies, Princeton University

Corresponding author: Mohamed Abdellatif , m.abdellatif@princeton.edu

Abstract
Different writing systems have been (historically and contemporarily) used to write out the same language. This is typically done by substituting letters (or symbols, in the case of non-alphanumeric systems). However, depending on the language and the involved writing systems, the process may not be purely deterministic. Quoting Becker and Becker [2000]

> *even such basic acts as* ***transliteration*** *involve interpretation– to the extent that there is meaning in the medium itself*
>
> .
>
> .
>
> *In* ***transliteration*** *itself there is exuberance (that is, meaning is added) and deficiency (meaning is lost).*

This gives significance to the problem of Machine Translation in the intersection of Digital Humanities and Natural Language Understanding.

Transformer-based models achieved success modeling human languages. However, many of them have the limitation of handling an input of maximum length of 512 tokens. To reuse a pre-trained model with this limitation for downstream tasks (e.g., Machine Transliteration) on input of sequences longer than 512 tokens, we propose a method to segment the input into interleaving (not mutually exclusive) pieces, invoke the model in a piecewise manner and construct the result. To consolidate the result, we propose a method to detect and correct potential (duplication and elimination) errors that reduces Word Error Rate from 0.0985 to 0.0.

Keywords
document-level; machine transliteration; bert

I INTRODUCTION

Transliteration is the process of mapping a source script (a given language expressed in a grapheme) to a target one (the same language expressed in a different grapheme). A language is different from a writing system. A language is a form a communication that is defined by verbal words (vocabulary) and rules to properly build sentences (grammar). On the other hand, a writing system is a system to express a language in writing. This system has symbols to type up sounds and verbal words (script) and rules to type up bigger units properly (orthography). Since transliteration is the process to map from one writing system to another (both exist to express the same language), that makes it different from translation. Some examples of languages that use different writing systems:

- Sanskrit-based language + Arabic-based writing system $\rightarrow$ Urdu
- Turkish-based language + Arabic-based writing system $\rightarrow$ Ottoman Turkish
- Arabic-based language + Hebrew-based writing system $\rightarrow$ Judaeo-Arabic
- German-based language + Hebrew-based writing system $\rightarrow$ Yiddish
- Spanish-based language + Hebrew-based writing system $\rightarrow$ Ladino

Machine Transliteration is the NLP task of automatically transliterating documents using a computer model/system. A rule-based solution may be good enough if the chances of ambiguity are small. If the target script has a higher number of characters and combinations, that increases the chances of ambiguity and makes the task more challenging. To machine-transliterate long texts using a model that is trained based on BERT (Devlin [2018], which has length limit of 512 tokens), we propose a three-stage approach. First, slice the input text using three adjacent sliding windows: leading context, core, and trailing context. Second, transliterate the generated pieces (complying with the maximum length limit, i.e., $\leq$ 512 tokens). Finally, construct the long text in the domain of the target grapheme. To consolidate the result, we investigate various techniques to detect and correct potential duplication and elimination errors in the stitching process to construct the output.

The Cairo Geniza is a cache of written material in the Ben Ezra synagogue in Cairo, Egypt. The room housed documents dating back to the 10^{th} century. The Jewish community of that era did not throw out any written material. The Princeton Geniza Project studies documentary material (e.g., letters, legal deeds, lists, accounts, state documents) dating back to the 10^{th} through 12^{th} centuries. In addition to the Jewish community in medieval (Fatimid) Egypt, the documents give us details about how the society functioned under the Fatimid rule.

Many of the documents are written in Judaeo-Arabic which is Arabic expressed in Hebrew letters. To recover the Arabic communication, the digitized Judaeo-Arabic needs to be transliterated. However, this transliteration direction is more challenging than Arabic to Judaeo-Arabic. Since Hebrew script has only 22 letters while Arabic has 26, some letters would have to do double duty. That is one of the reasons why recovering Arabic from Judaeo-Arabic is too challenging to be done with acceptable accuracy using a rule-based approach.

As a consequence of the scarcity of parallel data in digitized format, the number of machine-learning models trained on it is scarce too. Since we don't have the parallel data, we work with a model that is trained on off of BERT on parallel data. Although the more recent Transformer models have much larger context windows, tried on the less computationally-demanding BERT, the proposed method would generalize to larger documents with larger models. We release the code [1] on GitHub as open source.

II RELATED WORK

Transformer-based models have achieved success in modeling human languages. However, the $O(n^2)$ of the attention step and positional embedding impose an upper limit of 512 tokens on input text. Recent works attempt to address this limitation by modifying the $O(n^2)$ attention into a less computationally demanding one: local (e.g., Beltagy et al. [2020]), hierarchical (e.g., Yang et al. [2016]), sparse (e.g., Child et al. [2019]), approximate (e.g., Wang et al. [2020]), and IO-aware (e.g., Dao et al. [2022]) attention. Another direction addresses the memory to allow for the high computational demand (e.g., Dai [2019]). A third line of research is manipulating the context/text itself to comply with the limitation combined with a workaround to process long texts: selection (e.g., Ding et al. [2020]), aggregation (Izacard and Grave [2020]) and compression (e.g., Wingate et al. [2022]). One of the advantages to the context/text manipulation technique is the ability to work with model architectures that have the 512-tokens limitation since it does not require modifying the transformer architecture; this is the category wherein our contribution lies.

III METHOD

3.1 Dataset

Princeton University's Geniza collection[2] has 2910 documents with Judaeo-Arabic transcribed but without the corresponding Arabic transliteration. Only 9% of these have lengths of less than 512 characters. The lengths of 10% of them range from 500 to 1000 characters. The lengths of 7% ranges from 1000 to 1500 characters. The full histogram of character lengths is shown in Figure 1. To pre-process it, we used the document id (PGP ID) to select the contents of the untranslated documents, then we save the result in an <id_text> format. We exclude documents that are fewer than five characters in length.

3.2 Models

Rom and Bar [2024] pre-trained a BERT model (Devlin [2018]) on the Arabic and Hebrew portions of the OSCAR dataset (Suárez et al. [2020]). The goal was to build a bilingual (Hebrew-Arabic) model. Although the goal was not to transliterate Judaeo-Arabic, they mapped the Arabic OSCAR dataset into Judaeo-Arabic for the two languages to be represented in the same writing system (Hebrew script). That motivated Mitelman et al. [2024] to build on top of it. However,

[1] https://github.com/princetongenizalab/pgp_transliteration/tree/develop
[2] https://github.com/Princeton-CDH/test-geniza-metadatad

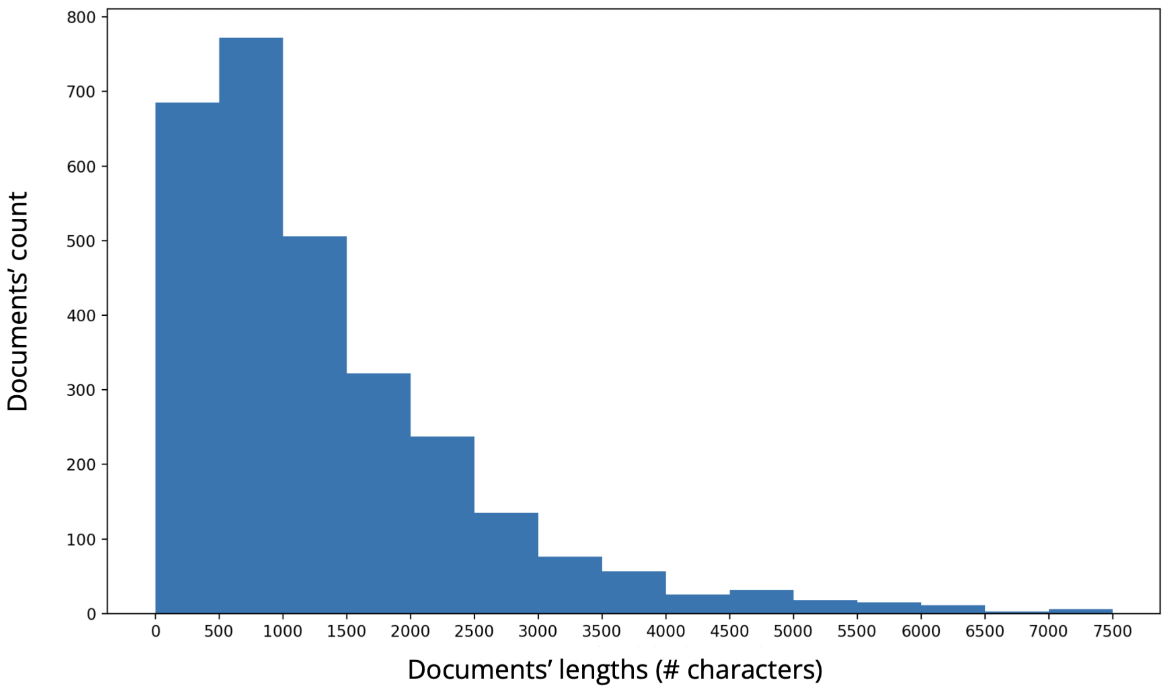

Figure 1: Histogram of document transcription lengths of the Geniza collection's untranslated documents, the vast majority of which have lengths greater than 512 characters.

the goal of the latter work was to transliterate from Judaeo-Arabic to Arabic. Since it is built on top of BERT, that implies it has the inherent limitation of supporting sequences of at most 512 tokens in length.

3.3 Context usage vs overhead

A straightforward approach to segmenting a long (i.e., > 512 tokens) source script is to divide it into non-interleaving pieces of 512 tokens, transliterate each one, and then stitch them back together. However, this approach implies 0% context sharing across successive model invocations.

A 100% context sharing approach implies invoking the model once on all the text without division. This is not possible for long texts since the model used has the inherent 512-token limitation. Instead of the 0% context sharing approach, we implement a solution that represents a compromise between the 0% and 100%: as shown in Figure 2, we further divide the 512-window into three sub-windows: leading context (100 tokens), core (300 tokens, highlighted in green) and trailing context (100 tokens), transliterate the whole 500 tokens window but consider only the core (300 tokens) sub-window when reconstructing the target long document. To stitch together the short core windows to construct the full (long) target script text, we need to locate those core windows in the transliterated pieces of 512-tokens. By stripping off the leading and trailing contexts from each generated 512-token unit and concatenating the middle 300-token core windows, the target script is constructed. This implies more invocations to the model to span the long text since the core sub-window of 300 tokens is shorter than the 512 token limit.

Figure 2: Input segmentation and output construction of long sequences. Only the core (green highlighted) text on the left side of the figure is considered when reconstructing the output.

This approach results in (200/512)*(100) = 39% context sharing.

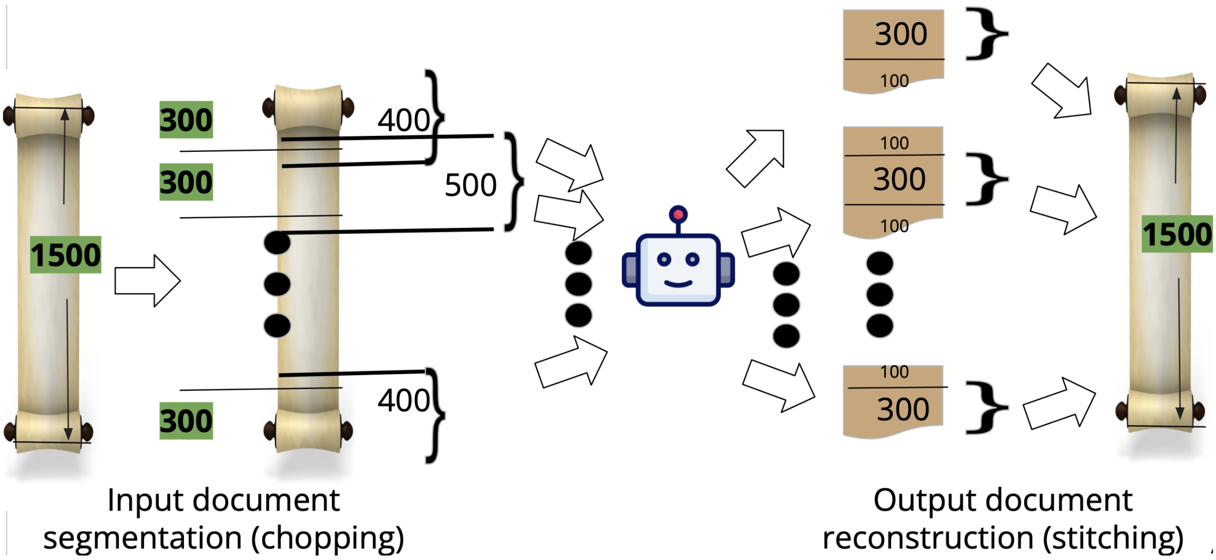

Figure 3: System architecture of long sequence transliteration[3]

IV ERROR ANALYSIS

Possible errors in the process of constructing the output are duplication and elimination. Appendix A lists examples of these errors while this section details why they happen and presents an approach to automate the detection and correction.

To invoke a model with a 512-token limit, we must pre-process the text (remove stop words and punctuation, tokenize, stem, and lemmatize). As shown in Figure 3, the borders set during the segmentation of the three sub-windows on the source (before) script may not align correctly

[3]The robot icon is public domain. Designed by Hilmy Abiyyu A.

with those on the target (after) script. Depending on those borders to strip the leading and trailing context from each single invocation-output and stitch back the core ones would result in duplication and/or elimination since the borders now differ slightly due to text pre-processing. However, if the text is non-repetitive, those errors can be detected.

When running the model on the data in question and applying the stitching techniques to reconstruct the long sequence in the target script domain, 48419 words were duplicated and 10693 were dropped, resulting in 0.0985 Word Error Rate.

	Leading context	**Core**	**Trailing context**		
1	We need	*to buy*	more groceries		
2	to buy	*more groceries*	before dinner	*to buy more groceries*	✔
3	to	*buy more groceries*	before dinner	*to buy buy more groceries*	✖
4	to buy more	*groceries*	before dinner	*to buy groceries*	✖

Table 1: Three test-cases of merging two outputs of successive invocations of the machine transliteration model (the basic operation to construct the long output). Different columns represent different pieces of text that were cut off by static borders. Row 1 represents the result of the first invocation. Rows 2 through 4 represent different cases of the second invocation (row 2: no error, row 3: duplicated-text error and row 4: missing-text error). Recall: to construct the output, the core window (middle column) from the first invocation is concatenated with that of the second invocation.

As illustrated by the red circles in Figure 3, the misalignment of barriers between input and output sequences is the root cause of duplications and missing transliteration errors. To recover the long-sequence output, relevant pieces of text (core windows) need to be extracted from successive text pieces outputted from the model, and then merged together. However, a given piece is not mutually exclusive from the next one (recall that there is interleaving text to increase context sharing between successive model invocations). Given an invocation-output text piece, the decision where to set the border on the following (model invocation) piece is the one that may introduce duplication or dropping of text.

Table 1 illustrates three cases: no error, duplication, and elimination errors in its last three rows, respectively. The goal is to recover the long sequence (i.e., *to buy more groceries*) as would be the case in row 2. Row 3 illustrates the result when the border is set earlier than it should (duplication) while row 3 shows the result when it is set later than it should (elimination). Finding the repetition between row 1's and row 3's middle cells, the duplicated-text error (too early border) can be spotted (and removed, i.e., corrected). Finding the repetition between row 1's trailing context and row 4's leading context, the missing-text error (too late border) can be spotted (and re-inserted, i.e., corrected). This can be done by incrementally listing substrings, reversing, finding the intersection, and making sure it is greater than a certain threshold. The pseudo-code 2 (and 1) illustrate further.

The illustrated methods to detect and correct errors successfully reduce the Word Error Rate to zero. However, this method assumes that the text does not contain any repetition. If it does, this

Algorithm 1: list_substrings

Input: A uni-code string in_str

Output: A list of uni-code strings out_lst

$n \leftarrow length(in\_str)$;
$out\_lst \leftarrow [\,]$;
for $i = 0$ **to** $(n-1)$ **do**
  $crt\_substring \leftarrow slice(in\_str, 0, i+1)$;
  $append(out\_lst, crt\_substring)$
end

method will not work as accurately in error detection. We leave alternative approaches to error detection and correction for future work.

V CONCLUSION AND FUTURE WORK

The proposed method successfully transliterates long sequences with controllable sliding windows and with a low error rate. Furthermore, the ability to control the size of the context and core windows allows us to compromise between the accuracy of the model (context sharing) and the computational expense (number of model invocations). To transliterate long sequences more effectively, future work includes enabling additional ways to support long sequences (e.g., starting off the training process on a more suitable architecture), investigating safer, more secure, and efficient ways to deploy the model to allow for minimum delay inference (towards real-time transliteration), and alternative ways to detect and correct errors assuming the text may have inherent repetitions.

ACKNOWLEDGMENT

Many thanks to Rebecca Sutton Koeser from the Center for Digital and Humanities at Princeton University for her input on the manuscript.

References

Alton L Becker and Alton L Becker. *Beyond translation: Essays toward a modern philology*. University of Michigan Press, 2000.

Iz Beltagy, Matthew E Peters, and Arman Cohan. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*, 2020.

Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating long sequences with sparse transformers. *arXiv preprint arXiv:1904.10509*, 2019.

Zihang Dai. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv preprint arXiv:1901.02860*, 2019.

Tri Dao, Dan Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in Neural Information Processing Systems*, 35:16344–16359, 2022.

Jacob Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Algorithm 2: intersect_strings

Input: A uni-code string in_str1
Input: A uni-code string in_str2
Input: A boolean flag $which\_reversed$
Output: The common occurrence of uni-code sequence out_seq

if $which\_reversed = False$ **then**
  $subs1 \leftarrow list\_substrings(reverse(in\_str1))$;
  $subs2 \leftarrow list\_substrings(in\_str2)$;
else
  $subs1 \leftarrow list\_substrings(in\_str1)$;
  $subs \leftarrow list\_substrings(reverse(in\_str2))$;
end
$out\_seq \leftarrow intersect(subs1, subs2)$

Ming Ding, Chang Zhou, Hongxia Yang, and Jie Tang. Cogltx: Applying bert to long texts. *Advances in Neural Information Processing Systems*, 33:12792–12804, 2020.

Gautier Izacard and Edouard Grave. Leveraging passage retrieval with generative models for open domain question answering. *arXiv preprint arXiv:2007.01282*, 2020.

Daniel Weisberg Mitelman, Nachum Dershowitz, and Kfir Bar. Code-switching and back-transliteration using a bilingual model. *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1501–1511, 2024.

Aviad Rom and Kfir Bar. Training a bilingual language model by mapping tokens onto a shared character space. *arXiv preprint arXiv:2402.16065*, 2024.

Pedro Javier Ortiz Suárez, Laurent Romary, and Benoît Sagot. A monolingual approach to contextualized word embeddings for mid-resource languages. *arXiv preprint arXiv:2006.06202*, 2020.

Sinong Wang, Belinda Z Li, Madian Khabsa, Han Fang, and Hao Ma. Linformer: Self-attention with linear complexity. *arXiv preprint arXiv:2006.04768*, 2020.

David Wingate, Mohammad Shoeybi, and Taylor Sorensen. Prompt compression and contrastive conditioning for controllability and toxicity reduction in language models. *arXiv preprint arXiv:2210.03162*, 2022.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, pages 1480–1489, 2016.

A OUTPUT CONSTRUCTION ERRORS

The appendix shows two documents. In each of them, the machine-transliterated Arabic content is shown before and after error handling. There are two kinds of errors: duplicated text is shown in bold with a bar above in the $\overline{\boldsymbol{before}}$ version (constructed) while the eliminated text is shown shown with the same format, but in the $\overline{\boldsymbol{after}}$ version (constructed and corrected).

1.1 Document 451

A letter[4] from Marduk bin Musa (Alexandria) to Nahray bin Nissim (Fustat). The transliterated Arabic had three errors (two duplications and one elimination).

[4] https://geniza.princeton.edu/en/documents/451/

1.1.1 Constructed

كتابي اطال الله بقاه مولئي اشيك وادام عزهو وتائدهو وسعادتهو وسلامتهو من اسخندرية يوم حد لح بقى كتمو اءلة علينا وعليك باحسن كاتمه عن حال سلامة وعافية قرب الله اشتماع علا افضل ال احوال اعلمتك ياموﻻ ان قد تقدم كتابي اليك واعلمتك حال ال يج ضرة الذي كانت معى انها خلصت لله التمر منها صورة موخططة عليها اسم يوسپ وثانية خلطتك منه عليها اسمك واسموه وثالثة عليها مكتب من قبل موسى ابن يحيا وما ندري يامولاي لمنهي فكتابك يعلمني لمن اصوره الثاليتة وقد كان سالت مولاي في كتابي المتقدم ان يشترى لي تاسومه بزمامين ولابي عمران ولابو الفصل وقد كان اعلمني مولاي قبل سفري ان استعملة مداسات ريفية فنريد وتشترى لي رطل فلفل ورطل قرفة ودستين ورق وذوا بيان ذواتى تكون مدورة تصلوح للخريطة ونحيب نستانس يا مولاي حاجة حاجة الا ان يفريج الله فلما كان في يومى هذه وصل قاريب بن موليكه من لوق وذكر ان الامير جوباره رد المركب وجميع التقل الذي انزلهو الا برقة ارجو ان يكون ذليك صحيح اسال الله ان يجمع علا من كان لهو **لهو** فيه شي وقد وصيت عن عقبان قبل خروجهو عن كرا توجده سدس وقد بلغني عن الساق الذي انفدتهو معوة الشيك ابي سعد لي بين قد فعل ما يشكهو وقد بالله يامولئي شق على ذلك كثير فنحيب منك يا مولئي لالا توريد عنى هذه ال حوائج فاني محتاج اليها وما المشي الا حافي وتعلمني ايش الذي صح في ال ميزر وما كان لمولئي من **وما كان لمولئي** من حاجة يشرفني بقضاها وتخوص عنى روحك بافضل السلام وتخوص مولاي ارب افضل اسلام وتعلمه شوقى اليه وكداليك مولاي اشيك ابي زكري وتعلمه حديث المركب ارجو ان يكون ذليك صحيح واسلام لمولاي اشيك ابي يحيا نهرئ بن نسيم شاخيروه مردوك اطال الله بقاه وادام عزه بن موسى

1.1.2 Constructed and corrected

كتابي اطال الله بقاه مولئي اشيك وادام عزهو وتائدهو وسعادتهو وسلامتهو من اسخندرية يوم حد لح بقى كتمو اءلة علينا وعليك باحسن كاتمه عن حال سلامة وعافية قرب الله اشتماع علا افضل ال احوال اعلمتك ياموﻻ ان قد تقدم كتابي اليك واعلمتك حال ال يج ضرة الذي كانت معى انها خلصت لله التمر منها صورة موخططة عليها اسم يوسپ وثانية خلطتك منه عليها اسمك واسموه وثالثة عليها مكتب من قبل موسى ابن يحيا وما ندري يامولاي لمنهي فكتابك يعلمني لمن اصوره الثاليتة وقد كان سالت مولاي في كتابي المتقدم ان يشترى لي تاسومه بزمامين ولابي عمران ولابو الفصل وقد كان اعلمني مولاي قبل سفري ان استعملة مداسات ريفية فنريد **انفادها** وتشترى لي رطل فلفل ورطل قرفة ودستين ورق وذوا بيان ذواتى تكون مدورة تصلوح للخريطة ونحيب نستانس يا مولاي حاجة حاجة الا ان يفريج الله فلما كان في يومى هذه وصل قاريب بن موليكه من لوق وذكر ان الامير جوباره رد المركب وجميع التقل الذي انزلهو الا برقة ارجو ان يكون ذليك صحيح اسال الله ان يجمع علا من كان لهو فيه شي وقد وصيت عن عقبان قبل خروجهو عن كرا توجده سدس وقد بلغني عن الساق الذي انفدتهو معوة الشيك ابي سعد لي بين قد فعل ما يشكهو وقد بالله يامولئي شق على ذلك كثير فنحيب منك يا مولئي لالا توريد عنى هذه ال حوائج فاني محتاج اليها وما المشي الا حافي وتعلمني ايش الذي صح في ال ميزر وما كان لمولئي من حاجة يشرفني بقضاها وتخوص عنى روحك بافضل السلام وتخوص مولاي ارب افضل اسلام وتعلمه شوقى اليه وكداليك مولاي اشيك ابي زكري وتعلمه حديث المركب ارجو ان يكون ذليك صحيح واسلام لمولاي اشيك ابي يحيا نهرئ بن نسيم شاخيروه مردوك اطال الله بقاه وادام عزه بن موسى

1.2 Document 4011

A court order[5] listing items to be returned to the divorced.

[5] https://geniza.princeton.edu/en/documents/4011/

1.2.1 Constructed

الذي تسلم الشيك ابو السرور رب پرخيه هلوي هزقن بن رب بنيمين هزقن شص بمئر هدرت هنگيدوت ه يروم
ومٔر بتي دينيم من الندونيا المختصة بست الحسب ابنة ترنو تر هلويم رحمة الله عصابة ذهب مكملة عدتها سبعة
قطع زوج حلق خبار مخلل وزوج حلق صغار كوس زوج بُر ذهب مكلل بتسمٔيك لولو وزوج عخاخيز بُر حنك لولو
دستينج مينا ميمون عنبر وذهب خ حبوب عنبر وستة ذهب حديدة الازورد مراة محلاة وخمسة خراسى دبل مداف ود
اغطية فضة درج فضة وخاتمين ذهب وخاتم فضة وزوگ بُر ذهب عوض الخاتم الذي في ال كتوبه ومكخلتين بلور
بافمام ذهب مداف بلور ميلة بلور دستينج ومرود الملبوس ثوب بياض ذبيقى مذهب ومعجر حريري ابيض حله
دعوى **دعوى** ثوب ومعجر وعصابة مني خلعة رماني الثوب والمعجر خلعة سمنجوني الثوب والمعجر وعصابة مذهبة
ملاة ذبيقى ثوب معجر ازرق حريري ونقاب درى مذهب مختومة عمل الدار معرقة لالس معرقة لالس مذهب وخمار
مغربي منديل ي ووسط منديل ذبيقي ازرق بياض منديل قفض اخصر معرقة متخت ونصف ردا جسيل ردتين زرق
ومشفع ونقاب اسود **اسود** ملحفة جديدة ونقاب وملحفة جسيل د جواخين منديلين للصواني منديلين للبرادة ونمط
ديباج و وسبعة مناديل بقي منديل ميذهب بين ديباغ ود مخذ ذيباغ طراحة طبرستان بمخدها مخذتين سوسنجرد مخذ
زرق وزوج مخذ شرب خصر وزوج مخذ ذبيقي وتكت ديبقي وبردعة لحاف خلافي وملحفة سقلاطون طراحة النحاس
منارة عرائسي ومنارة **سقلاطون طراحة النحاس منارة عرائسي ومنارة** شمعي وطستين وابريق ل وصدر وشرابيه وحسكة
وتور اشنان مكمل وسطل وخوز زيت وسراج وحقه مكملة مقدمة دبل وعاج وخرسيها وسفط بخرسية ومنشفة برنية
مطئيب حصل على جميع ما في الكتوبة وتسلمه الشيك ابو السرور المذكور ولم يعدم منه شي غير زوج اسوره مخرز
ذكر انها تحت يد الشيك ابو ولم يعدم منه شي غير زوج سٔوره مخرز ذكر انها تحت يد الشيك ابو ال فرج ع وطراحة
طبرستان في دار ل عند ربنو خيیـד نظروهي من شميا سقلاطون ذكر انهم عندهم والشرابية عندهم وايضا غارية تدعا
شعب وكان ذلك في العشر الاوسط من تتري

1.2.2 Constructed and corrected

الذي تسلم الشيك ابو السرور رب پرخيه هلوي هزقن بن رب بنيمين هزقن شص بمئر هدرت هنگيدوت ه يروم
ومٔر بتي دينيم من الندونيا المختصة بست الحسب ابنة ترنو تر هلويم رحمة الله عصابة ذهب مكملة عدتها سبعة
قطع زوج حلق خبار مخلل وزوج حلق صغار كوس زوج بُر ذهب مكلل بتسمٔيك لولو وزوج عخاخيز بُر حنك لولو
دستينج مينا ميمون عنبر وذهب خ حبوب عنبر وستة ذهب حديدة الازورد مراة محلاة وخمسة خراسى دبل مداف ود
اغطية فضة درج فضة وخاتمين ذهب وخاتم فضة وزوگ بُر ذهب عوض الخاتم الذي في ال كتوبه ومكخلتين بلور
بافمام ذهب مداف بلور ميلة بلور دستينج ومرود الملبوس ثوب بياض ذبيقى مذهب ومعجر حريري ابيض حله
دعوى ثوب ومعجر وعصابة مني خلعة رماني الثوب والمعجر خلعة سمنجوني الثوب والمعجر وعصابة مذهبة ملاة
ذبيقى ثوب معجر ازرق حريري ونقاب درى مذهب مختومة عمل الدار معرقة لالس معرقة لالس مذهب وخمار
مغربي منديل ي ووسط منديل ذبيقي ازرق بياض منديل قفض اخصر معرقة متخت ونصف ردا جسيل ردتين زرق
ومشفع ونقاب اسود ملحفة جديدة ونقاب وملحفة جسيل د جواخين منديلين للصواني منديلين للبرادة ونمط ديباج و
وسبعة مناديل بقي منديل ميذهب بين ديباغ ود مخذ ذيباغ طراحة طبرستان بمخدها مخذتين سوسنجرد مخذ زرق
وزوج مخذ شرب خصر وزوج مخذ ذبيقي وتكت ديبقي وبردعة لحاف خلافي وملحفة سقلاطون طراحة النحاس منارة
عرائسي ومنارة شمعي وطستين وابريق ل وصدر وشرابيه وحسكة وتور اشنان مكمل وسطل وخوز زيت وسراج وحقه
مكملة مقدمة دبل وعاج وخرسيها وسفط بخرسية ومنشفة برنية مطئيب حصل على جميع ما في الكتوبة وتسلمه
الشيك ابو السرور المذكور ولم يعدم منه شي غير زوج اسوره مخرز ذكر انها تحت يد الشيك ابو ولم يعدم منه شي
غير زوج سٔوره مخرز ذكر انها تحت يد الشيك ابو ال فرج ع وطراحة طبرستان في دار ل عند ربنو خيیـד نظروهي
من شميا سقلاطون ذكر انهم عندهم والشرابية عندهم وايضا غارية تدعا شعب وكان ذلك في العشر الاوسط من

تتري