

Computational Document Analysis: A Humanists Perspective

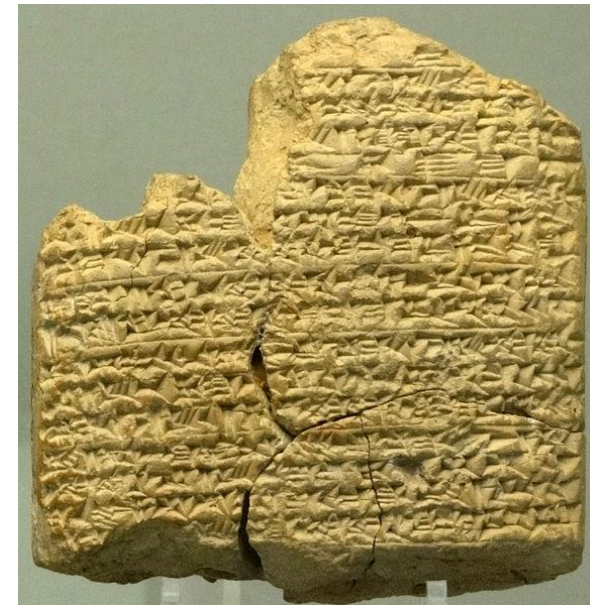
Daniel Stökl Ben Ezra

EPHE / PSL Research University Paris

daniel.stoekl@ephe.sorbonne.fr

What could we need computational humanities for?

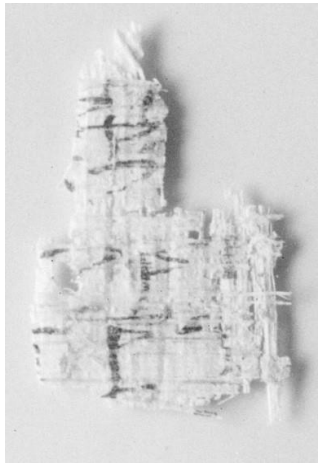
- Speed up process of cataloguing, deciphering, translating, publishing texts and understanding human culture
 - [DH (new forms of publishing, documenting, visualizing, analyzing data)]
- Large number of documents for small number of scholars (Egyptology, Akkadian, Dunhuang, Papyrology...)
 - 500.000 cuneiform tablets for ca. 100 experts



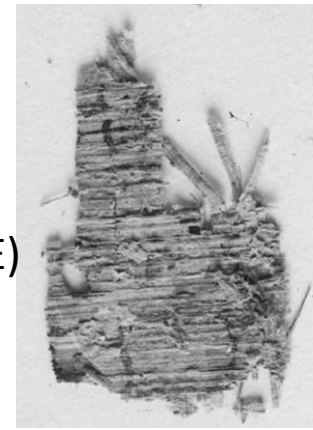
Cataloguing

Fragment spotting

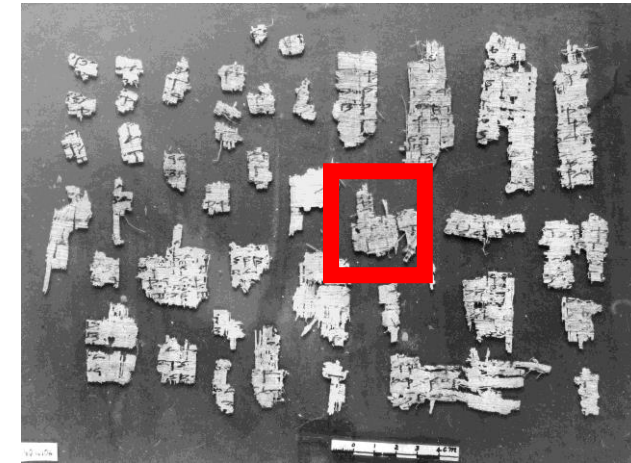
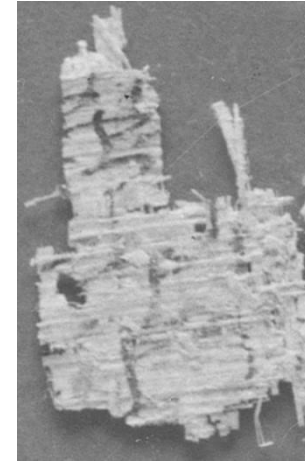
Scripta Qumranica Electronica
on the Dead Sea Scrolls (Haifa, Göttingen, IAA, Tel Aviv, EPHE)



PAM 43.410



PAM 40.633



PAM 41.990

Sample results of the Geniza project (TAU)



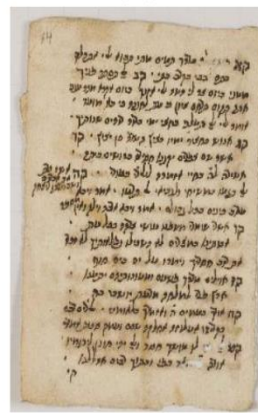
(a1)



(a2)



(b1)



(b2)



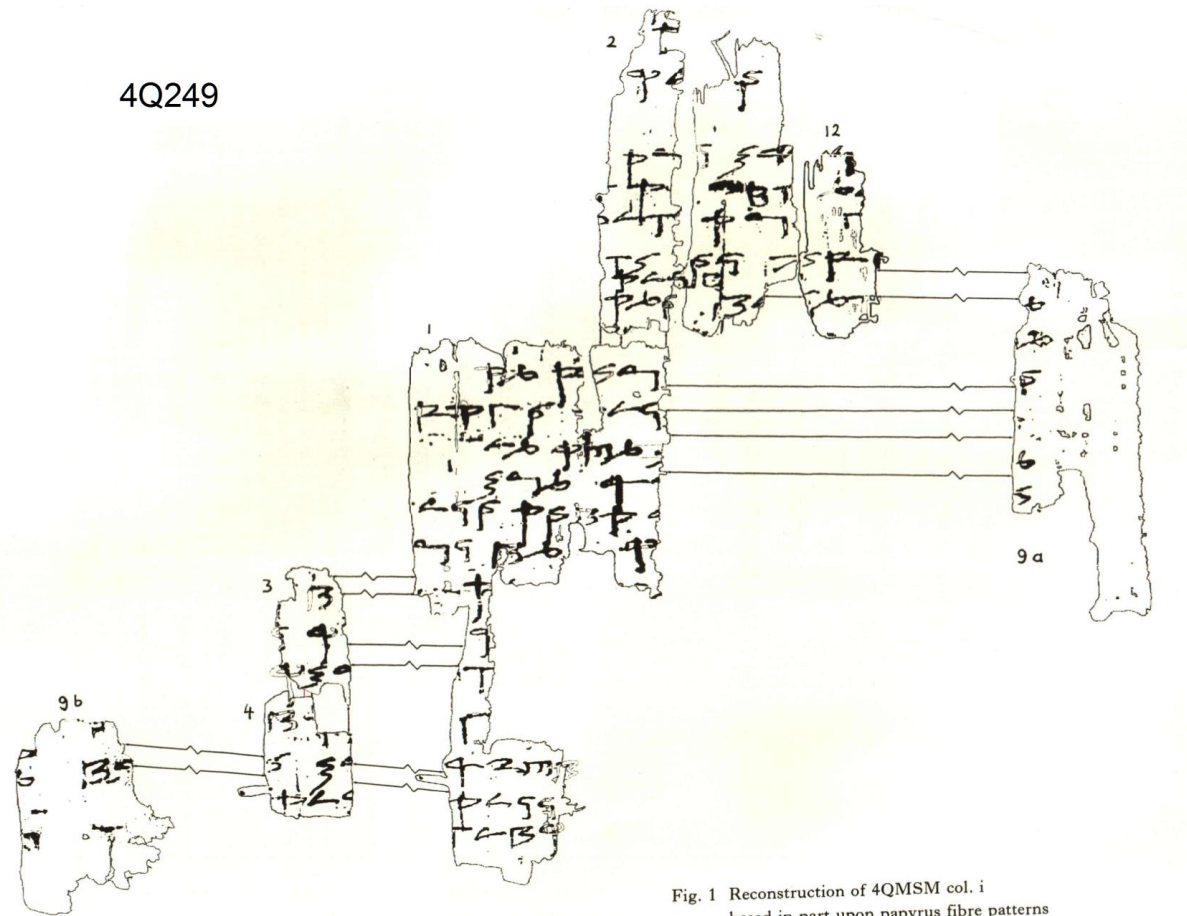
(a)



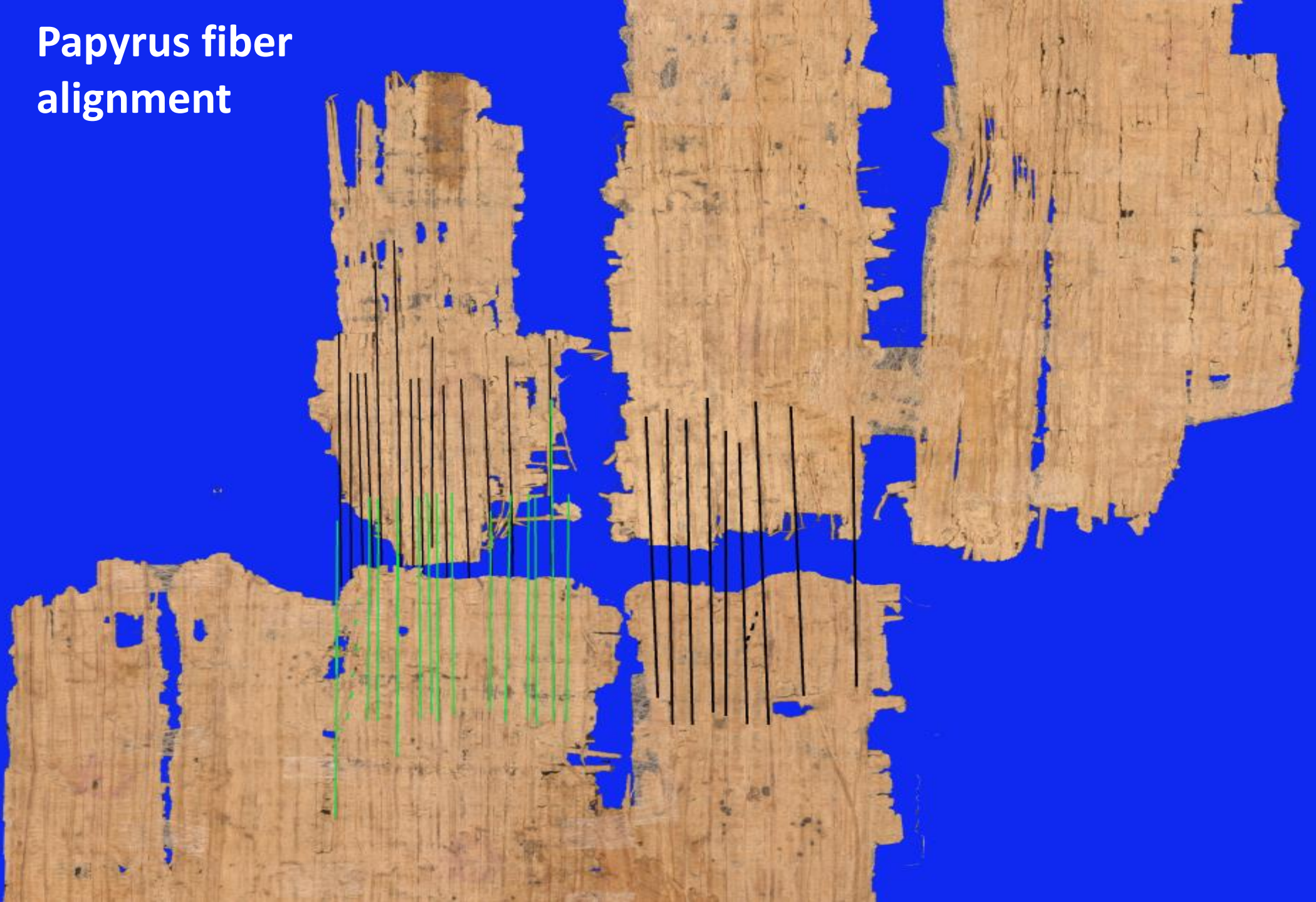
(b)

- (a) Initially no catalog data (other than material (paper)), it is found to be a philosophical work with a Naskhi script.
- (b) Initially the subject classification had three equally probable options, two cantillation options and a very broad script type categorization. The inference determines it to be a religious exegesis essay with a Tiberian semi-cursive script and partial cantillation.

Papyrus fiber alignment



Papyrus fiber alignment





TS 8102 30

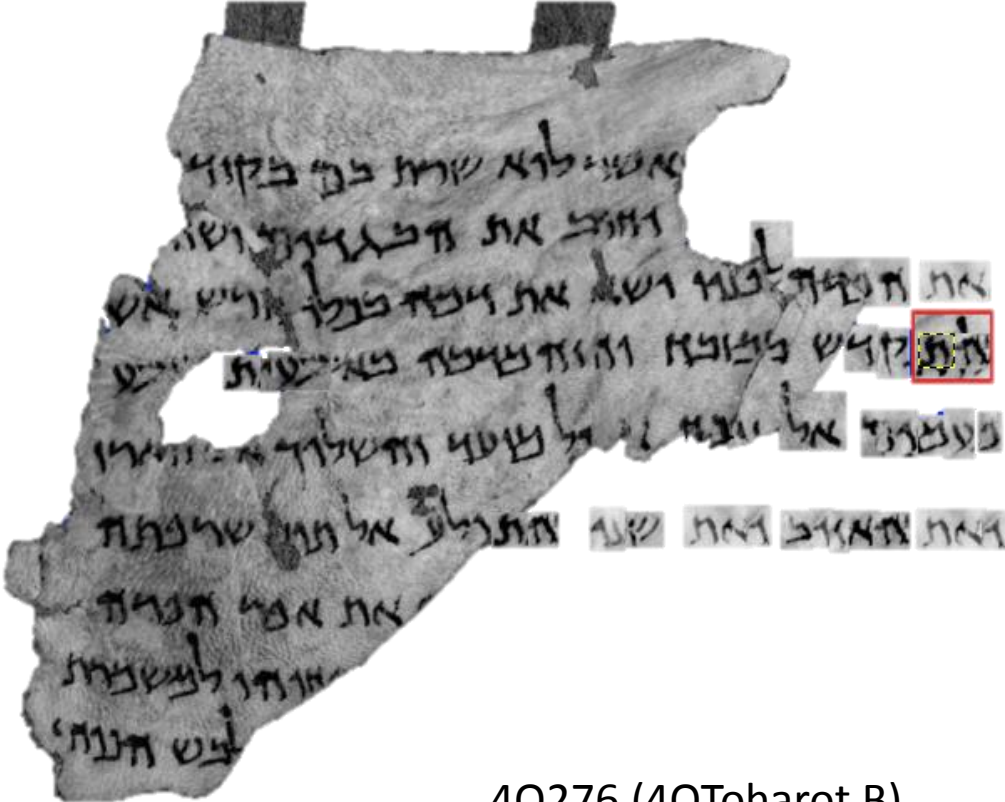


TS F10J 3D

TS F10J 3D

TS F10J 3D

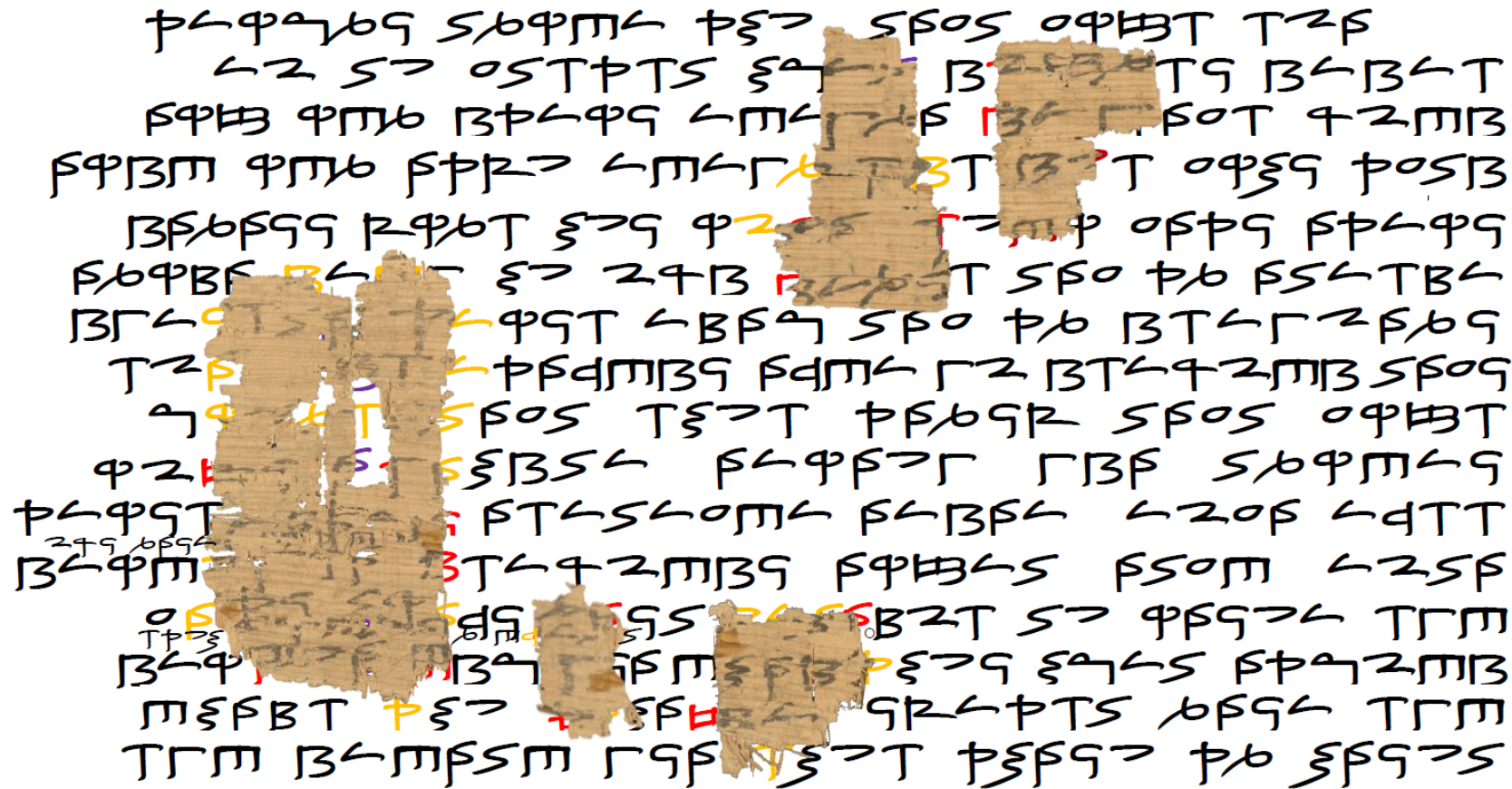
Make deciphering easier: Font creation



4Q276 (4QToharot B)

Make deciphering easier: Font creation for lacuna restoration

In planification in SQE: tool for registrating and moving r and v images of fragments linked to text reconstruction



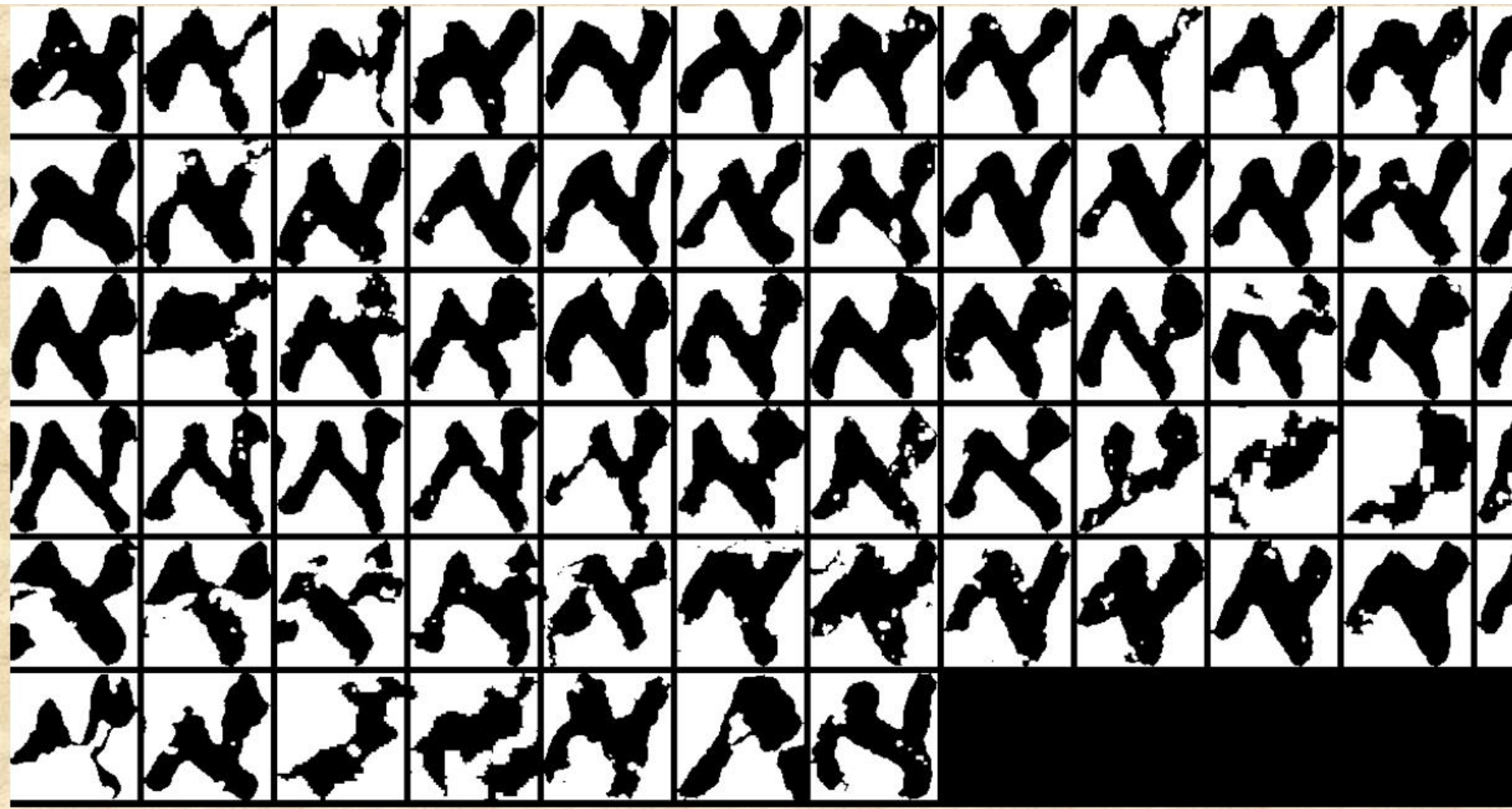
4Qcryptic A Serekh HaEda

Cryptic script from Qumran, perhaps early 1st century BCE

Jonathan Ben-Dov, Daniel Stökl Ben Ezra, and Asaf Gayer, "A Newly-reconstructed Copy of Serekh haEdah from Cave 4 in Cryptic Script," *Revue de Qumrân* (forthcoming), 80 pages

Asaf Gayer, Daniel Stökl Ben Ezra, and Jonathan Ben-Dov, "A New Join of Two Fragments of 4QcryptA Serekh haEdah and Its Implications," *Dead Sea Discoveries* 23 (2016), 139-154

Letter spotting to semiautomatically create paleographical charts, a database for lacuna restoration or font creation (Nachum Dershowitz, Lior Wolf)



Make deciphering easier: Unwrap a magic bowl



viral OFTA text image alignment tool produced in collaboration between TAU, OUI and EPHE (Icdar 2015)

אשר בחר ולשנא את נול אשר עם
רחוק פכול רע
לדבוק בכול פעשי טוב ולעשות אמת
צדקה ופעמט
ארא ולוא ללכת עוד בשרירות לב
ושפה ועיני זנות
לעשות כול רע ולהבי את כול הזדבים
עשוות חוקי אל
כבריות חסד להיחד בעצת אל ולהתהלכ
כניו ונפים כול
ונגלות לפועדי תעודותם ולאהוב כול
בני אור איש
בגורלו בעצת אל ולשנא כול בני חושך
ויש באשפות
כצקמת אל וכול הזדבים לאמתו יביאו כול
יעתם וכוחם
הונם ביחד אל לברר דעתם באמת חוקי
ול וכוחם לתנן
תם זיכיו וכול הונם בעצת צדקו ולוא
צעוד בכול אחד
פכול זרבי אל בקציהם ולוא לקדם
נתינת ולוא להתאחר
פכול ופועדיהם ולוא לסור פחוקי אמתו
לכת יפין ושפאל
כול הבאים בסרט היחד יטבורו בבריות

אשר בחר ולשנא את נול אשר עם

אשר בחר ולשנא את נול אשר עם

אשר בחר ולשנא את נול אשר עם

אשר בחר ולשנא את נול אשר עם

אשר בחר ולשנא את נול אשר עם

אשר בחר ולשנא את נול אשר עם
רחוק פכול רע
לדבוק בכול פעשי טוב ולעשות אמת
צדקה ופעמט
ארא ולוא ללכת עוד בשרירות לב
ושפה ועיני זנות
לעשות כול רע ולהבי את כול הזדבים
עשוות חוקי אל
כבריות חסד להיחד בעצת אל ולהתהלכ
כניו ונפים כול
ונגלות לפועדי תעודותם ולאהוב כול
בני אור איש
בגורלו בעצת אל ולשנא כול בני חושך
ויש באשפות
כצקמת אל וכול הזדבים לאמתו יביאו כול
יעתם וכוחם
הונם ביחד אל לברר דעתם באמת חוקי
ול וכוחם לתנן
תם זיכיו וכול הונם בעצת צדקו ולוא
צעוד בכול אחד
פכול זרבי אל בקציהם ולוא לקדם
נתינת ולוא להתאחר
פכול ופועדיהם ולוא לסור פחוקי אמתו
לכת יפין ושפאל
כול הבאים בסרט היחד יטבורו בבריות

Data bottleneck

In order to produce large quantities of – machine readable – data deep learning could be very effective.

But for the training of deep neuron networks we need large quantities of machine readable data...

Institutional bottleneck

Computer scientists are interested in publishing new algorithms in excellent CS journals.

Humanists are interested in publishing texts and analyses in excellent humanities book series and journals (and sometimes online).

Open source requirement for applications to funding organisations and promotions in universities

New French law on publications and data

Many algorithms have been developed

but they are vegetating on some forgotten hard disk of a former PhD student.

Engineers are needed to create real working applications

Standards for exchange and APIs

DivaDIA

Reciprocity

- Model ambiguity!

פנים
 ופנים
 הפנים
 ה
 פנים

המעשה ד וחוקר גזירת אמת רבן שמעון
 בן גמליאל אומ לא יחסוך לפני שהוא עב
 מפעט את פלאכתו שלבני הבית ויוסי
 או לא יחסוך מפני שהוא תני בית דין
 ה הלקוחין יבין פירותיו או שגל
 לזמן שואני עניד לפריש הריחם תרופה
 ועשרה בעשר ותשעה בעשר שיני כוחל
 ושותה ו הילול תאינים שלכול בתר
 בית והוא בבית העזריש או בשדה או בשתי
 תאנים שני עמיד דרי הם תרופה ועשר בעשר
 ותשע בעשר שיני ו הירדני או פה
 שאני עניד להפיש לחור הרי הוא בעשר
 ושאר הפעשר ספור לו זה שעשיתי בעשר
 עשוי תרופת בעשר עליו ובעשר שני בע
 בעופט ובדרכיו וכו' חלל על העשת
 ה היל לפני שתי כלכית שלטכל
 ומה בעשורת וזו הראשנה בעשורת ע
 שלה וזו שליו וזו הראשנה בעשורת בע
 בעשורותיהם בעשורת כלבלה וחברתה
 קרא שם ט טאה טבל טאה חולין
 טבל טאה ואחת טאה חולין כותקנין טאה
 בעשר טבל טאה ועשר טאה טאה טבל
 והשנים בעשר תשעים טבל ושבעים ג
 בעשר לא תפסיד כלום זה הכלל כל זמן
 שהטבל פרופה לא תפסיד כלום ו לא
 מי שהיו לו עשר שורות שלעשר עשר בני
 זין ואם שורה חמישונה אחת בעשר ואין
 ידוע אי זו היא טבל שתי חביות לו טבל
 חני שורה חמישונה אחת בעשר ואין ידוע
 אי זו היא טבל ארבע חביות כארבע חביות
 שורה אחת בעשר ואין ידוע אי זו היא ג
 טבל שורה אחת לוכסן חני שורה אחת
 בעשר ואין ידוע אי זו היא טבל שתי
 שורות לוכסן חבית אחת בעשר ואין ידוע
 אי זו היא טבל בכל חבית וחבית
 חסלת דביי פיקו ז

כלאים פרק א
 החיטים והזנון אינן כלאים זה בזה השערים
 ושכולת שועל הכוספים והשיבון הנחל
 והפסיד והפרדקן והטופח ופול הלכן והשע
 והשיעיעת אינן כלאים זה בזה
 הקישות והלפתון אינן כלאים זה בזה ריבון
 או כלאים חזרת וחזרת נלים עלשין עלשין
 שדה כרישים כרישי שדה כסבר כסבר
 שדה חרדל חרדל פיגרי ודלעת הפיגריה וה
 והרבויה ופול הפיגרי והחרוב אינן כלאים
 זה בזה ג הלפת והנפוס והכרוב והתונגור
 התרדים והלענים הוסיא ל עזיבה השום
 והשופנית הבצל והבגליעול והתרופות
 והפסולים אינן כלאים זה בזה ד
 ובאילן האנסין והקרוסטיילין והפרישין
 והעורין אינן כלאים זה בזה התפוח והחוור
 הפקקין והשקדים השימון והריקין אמ
 על פי שדוכין זה לזה כלאים זה בזה
 ה הענן והנפוס החרדל והלפסן ודלעת
 יונת עש הפיגריה והרבויה אמ על פי
 שדוכין זה לזה כלאים זה בזה ו
 הזאב והכלב כלב הכופרי והשועל העינים
 והעבאים העילים והחילים הכוס והפרד
 הכר והחפור החפור והעורד אמ על פי
 שדוכין זה לזה כלאים זה בזה ז
 אין טביאין אילן באילן ורק בירק ו לא אילן
 בירק ולא ירק באילן ר יהוד פהיר וק באילן
 ח אין טטען ודקות בתוך כנור ש
 שלשקה אין פריבין כנפס על גבי דינה
 לבנה מפני שהוא ירק באילן אין טטען
 וחרו שלמינה לתוך החיוב שיהיה בקריו
 אין תוכיין ופורה ולכן לתוך האכסות
 שהיא זורקת פיפיה לתוכי מפני שהוא
 אילן בירק אין טטען ורק דלעת לתוך
 החלפית שהיא פשערתו מפני שהוא
 ורק בירק ט הטופח אמה ועננות מה
 הנפסן אם היו בעצת העלים כנללים אינן
 חויטש לא מיטוס כלאים ולא עשוס שב

פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים

learn difficult

השעורים פנימפות עם הכל חזין מן החיטים
 ר יוח טן טרוי או נשאר תמינן בעטרפון זה ע
 עם זה ג שני קנין וקב אורז וקב
 תרופה באמצע אינן בעטרפון דבר שטעלה
 חלתו באמצע בעטרפון שכבר פהחיט חלה
 ר עכש וקב יטן טעשטו זה פנה
 ר ישמעאל או יטול טן האמציע וחכב אומ
 הטטל חלה טן קקב ר עקיב או חלה וחכב או
 אינה חלה ה שני קנים שטעלה חלה
 שניה בטני עננו ושלוח כקני עננו חזר
 ונשטן עיפה אמת ר עקיב פוטר וחכמים
 כתיבין נפצא חסרון קלו ו טטל אמס
 פני חלה בעיפה שלא הורבה חלקה לעשוטה
 כטגרה להיות פריש עליה וחלה חסת דפוי
 יע שגיבסח שחלת דפוי טעלת טן העמוד ע
 על הטטא שלא טן פונקט ו ישראל
 שהיו אריסים לגויים בסוריה ר אליע כתיב
 פירופיקן בפניטרות ובשביעית ורבן עליא
 פוסד רבן גמליאל או שתי חלות בסוריה וז
 אליע אמר חלה אחת אחיו קולו שלדכן געלו
 וקלו טלו אליע חזרו לנהוג בדברי רבן געל
 בשתי דרכים ח רבן געלי או שלוש
 ארצות אחלה בארץ ישראל ועל גיב חלה
 אחת טטיב עד העיר ועל אצטב שתי חלות
 אחת לאור ואחת לכתן שלאור זש לה שיעור
 נשלקין אין לה שיעור מהעיר ובאצטב ו ל
 למקום שתי חלות אחת לאור ואחת לכתן
 שלאור אין לה שיעור ושלפון זש לה שיע
 שיעור וטבל יום אכילה ר יוסק או אינן
 גרויה טביאן ט ואסונה לוכסן ולכות
 לטרות ולזירות נאכלת עם קוד על השלפן
 ונתפח לכל כן י אילו ניגזין לכל טון
 התרפים ותבירות מדיון טן ומזיון פטר
 חפור וקורין נקלתיים וקגיפה וראשית
 מה נשכן שריפה וקדשי קראש וקטסויים
 ר יהוד אופר כטיכורים כרשעי תרופה ר
 עקיב עגור וחכמים אופרנים י א
 נתיי איש הזונ הטיא חלות ענייטור ולא

קיבלו עליו אנשי אלכסנדריה הטיאו חלו לחוק
 אלכסנדריה ולא קיבלו טון אנשי הירבנים
 הטיאו בבורים קודם לצגרת ולא קיבלו מהם
 מפי כתוב שבהורה וכו' קקצור כפוני פישו
 אשר תזרע בשדה יב כן אטיקס העלה
 בבורות טבל ולא קיבלו טעט יוסק חכה
 הטיא כפונרי זין נשפן ולא קיבלו טעט אמ
 היא העלה את בעני נאמבני ביהו לעשות פסח
 כטן כירושלם והחזירוהו שלא יקבע הדבר
 חזקה אריסטון הטיא כפונרי טאמפנייה וקיבלו
 טעט מפני שאקרו הקוצה כפונריה כקונה
 כפרזר שבירו שלבי י
חסלת חלה פיקו ז
טבכת עליה פרק א
 הטטע לקנין ולקורות פטור טן בערלה ר יוסי
 או אמיו א הפגמי לפאכלי נכחיון לספיד
 הפגמי טנג נכחיון פטור ג עת שכמו
 אבותינו לארץ ופצאנו נטויג פטור נטעי אמ
 על פי שלא פייטשו חייב נטעי לרכים חייב
 ר יהוד פטור הטטעי פרשות הנבים והגברי
 שפטע והנצלן שטעי והטטעי כטמנה והע
 והעלה באילו טיב בערלה ג אילו ע
 שיעקר נפלע עישו שטפו עמר ופלע עיפרו
 אם יכול לחיות פטור ואם לאו חייב נעקר
 פלע פיגרו זיעענה הפחרישה או שועיעש
 נשטאו בעמר אם יכול לחיות פטור ואם לאו
 חייב ד אילן שלשתיד בו שודש פטור
 וכפה והא בשודש רבן שטע טן געלי א טטס
 ר אלעזר כן יהוד איש בדונה פקטט שלטיקן
 ה אילן שיעקר וכו' בריכה והוא חיה
 מפנה חונה וקינה לחיות ככריכה הכריכה טן
 שנה אחר שנה נטפסה עונה טיעה שנפטרה
 סיפון גפטס היטפך אה על פי שיה
 שהכריכו בארץ פוגר ר מאיר או לקיב שכוחה
 זכה פוגר ומקום שכוחה רע אסור בריכה ג
 שניטפסה נהיא עלמאה פירות אם חוסיה כג
 בפאתים אסור ו נטיעה שלעגלד
 ושלכילוי הפכס שנתענבו כטיטעית פרי וז

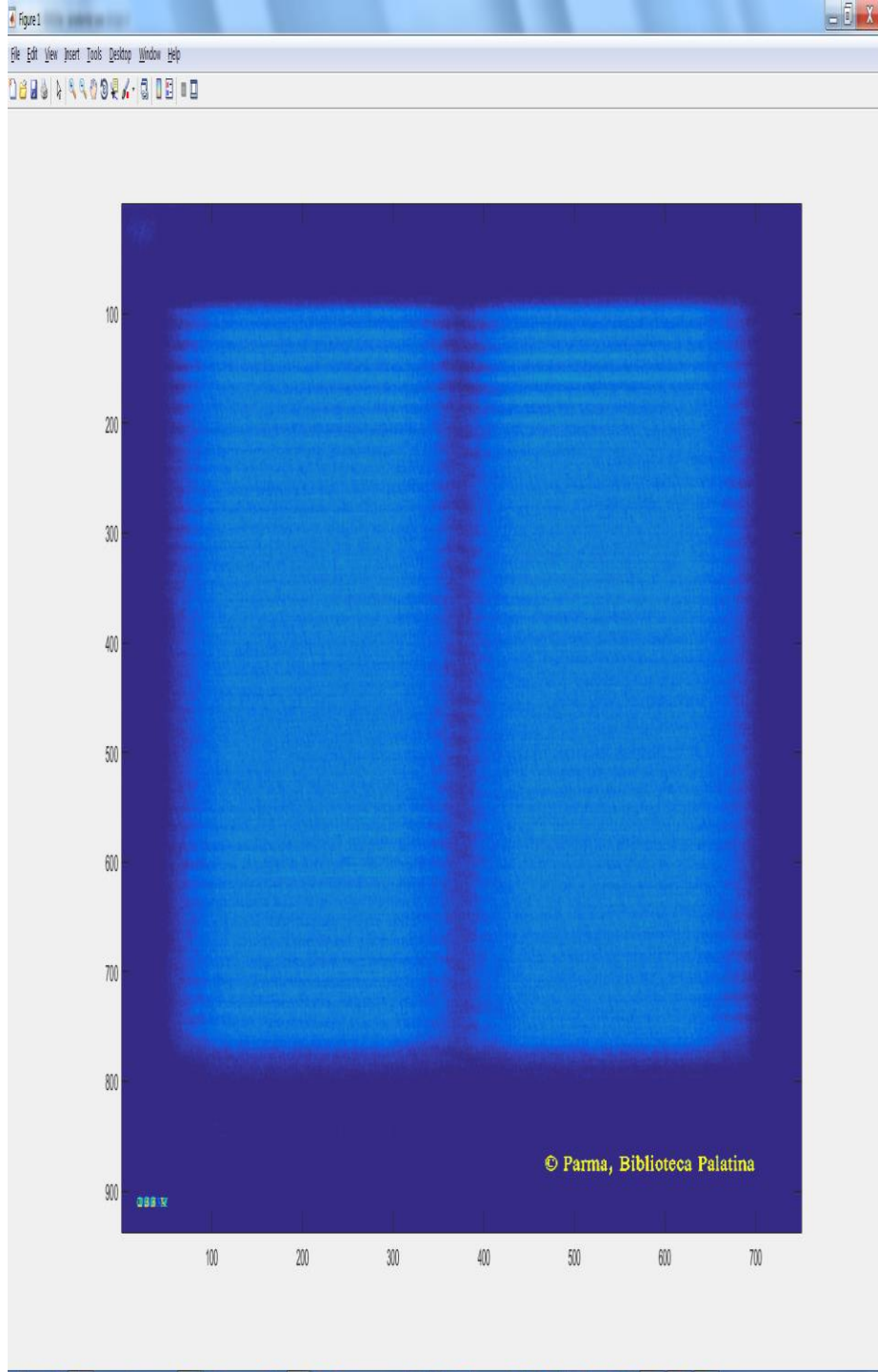
פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים
 פנים

לשנים לפנים ימי אסיל ולידליה

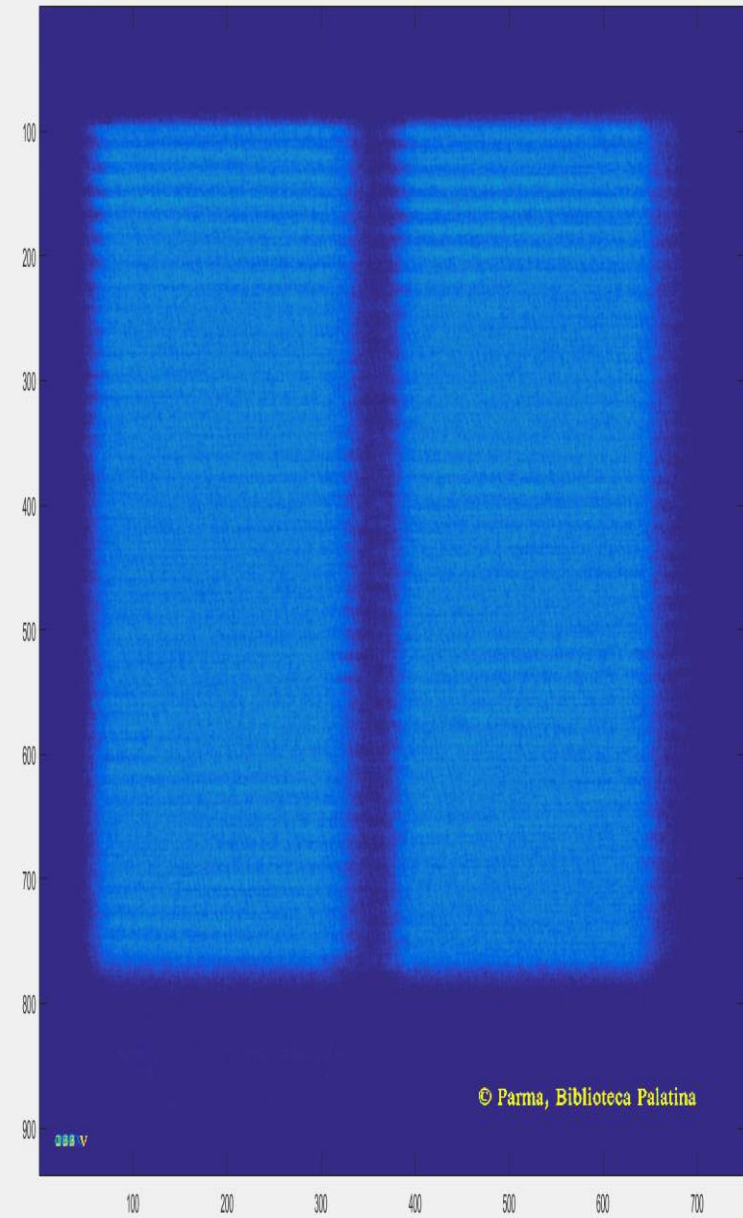
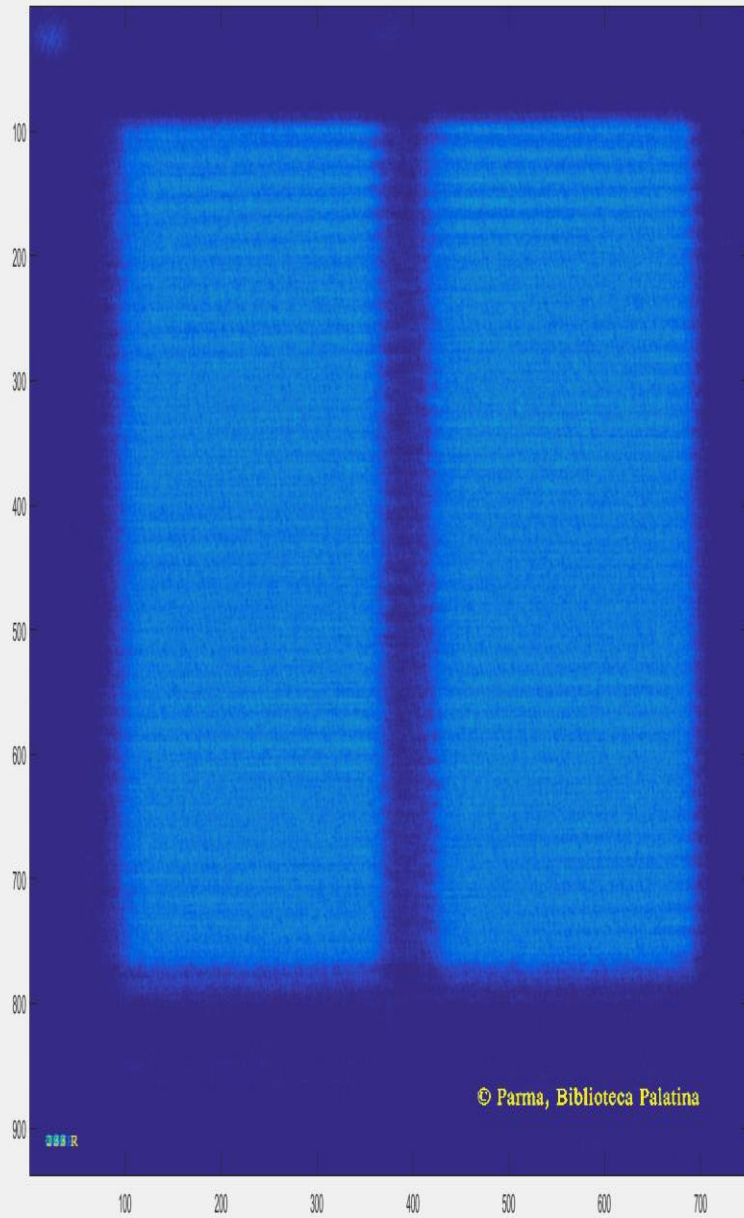
holistic approach to documents

- Literary manuscripts (often) have a constant layout that can be exploited by regarding them as 3D object.

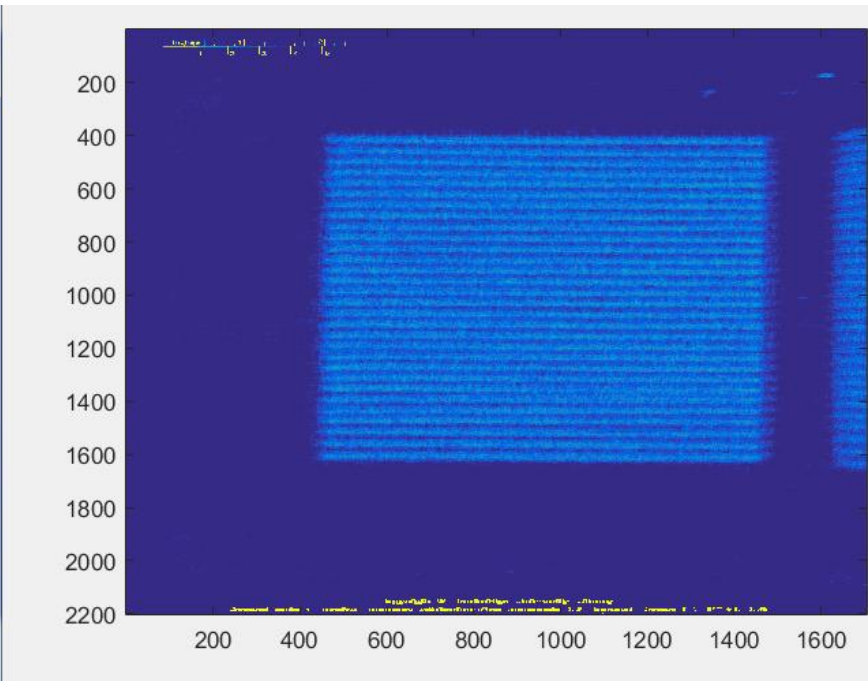
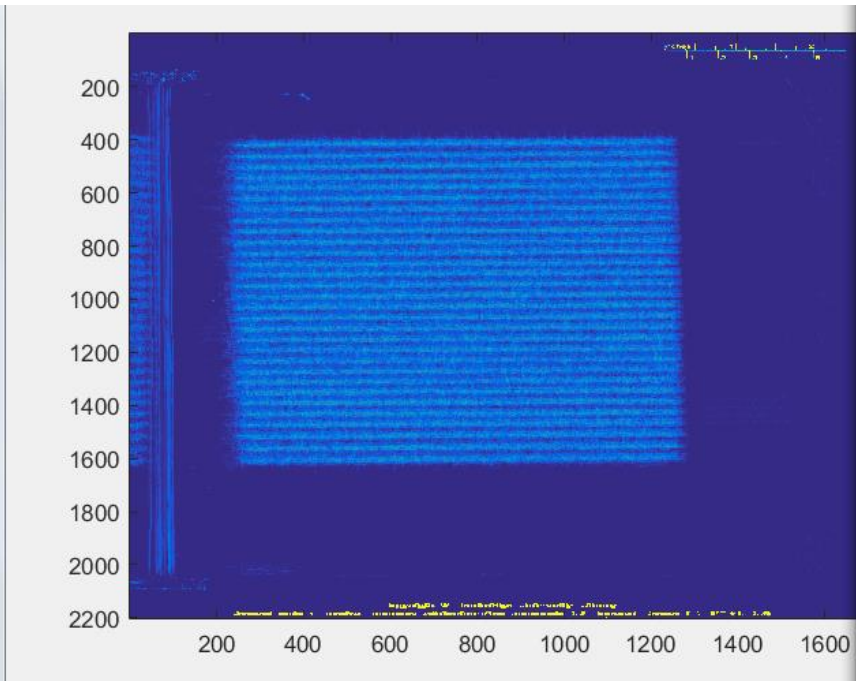
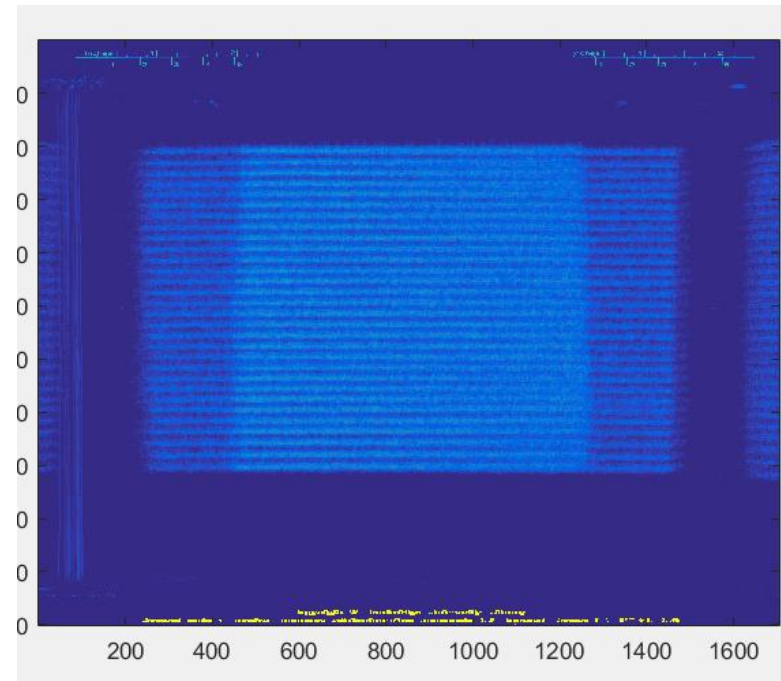
ALL



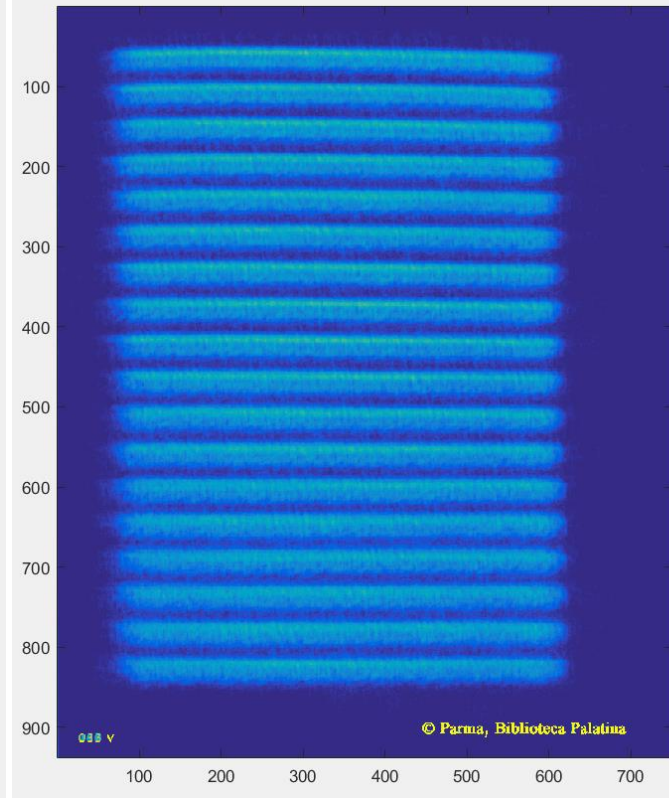
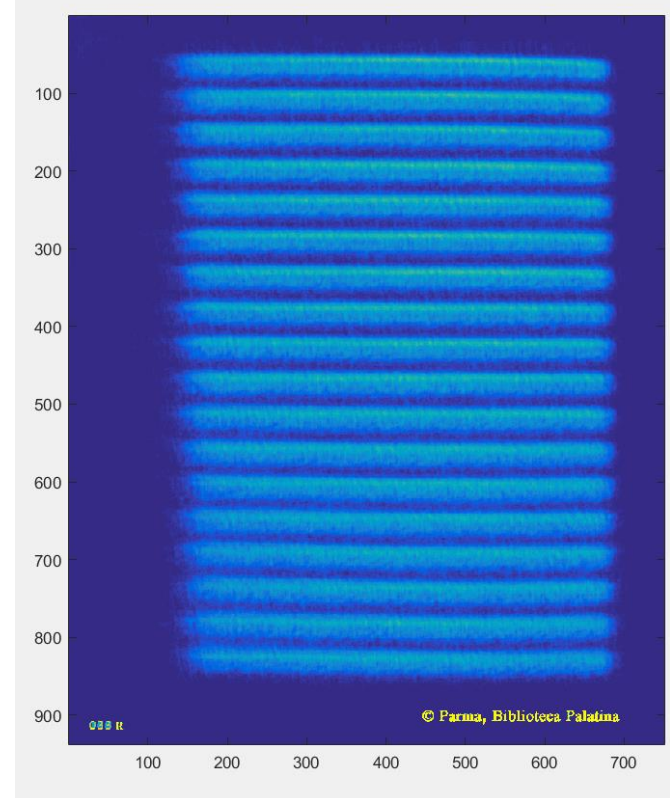
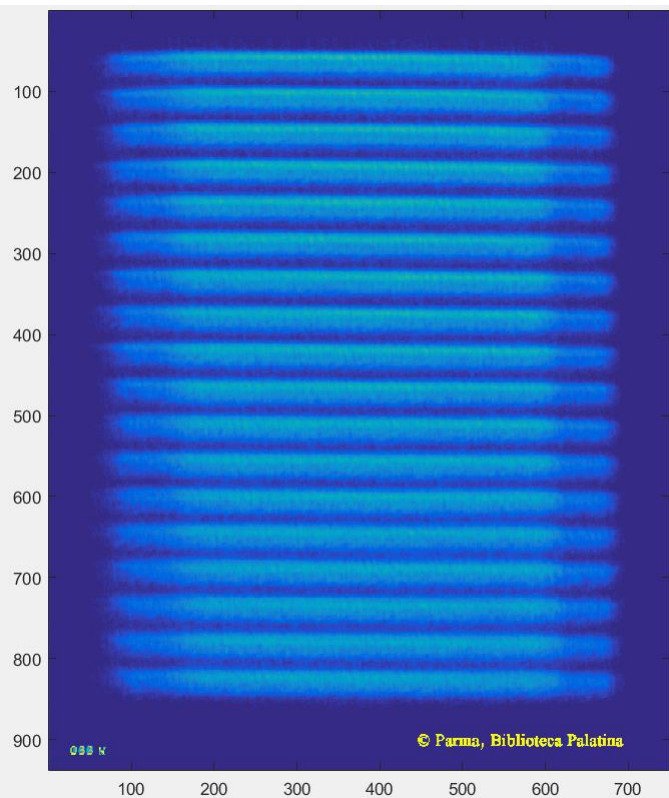
odd pages



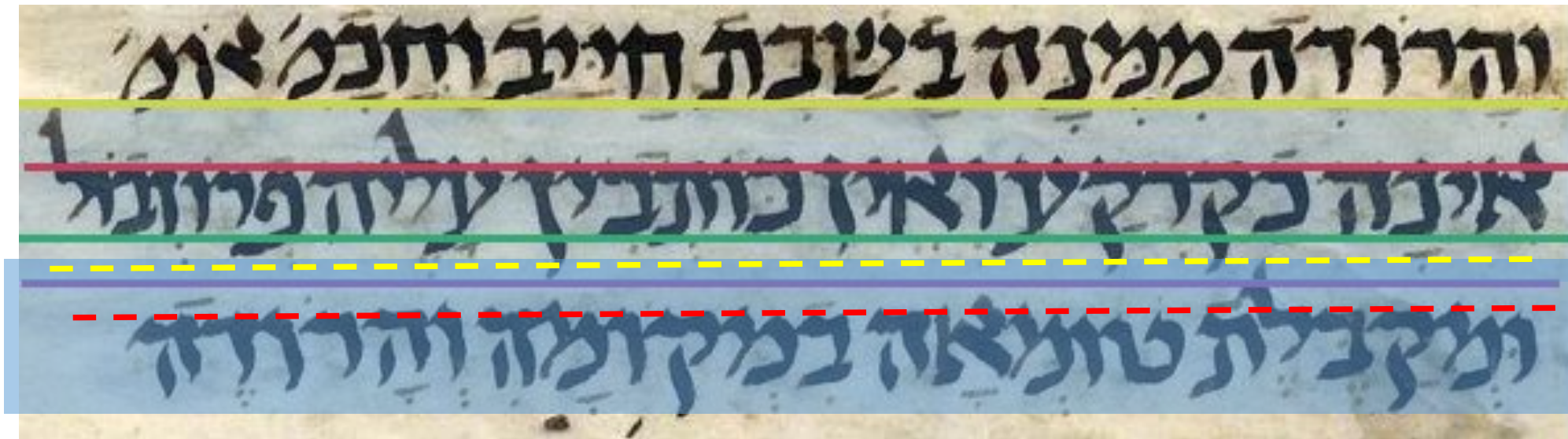
ms Cambridge 470.1



Even for lines (Parma2)



a text line is constituted by many lines



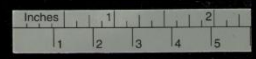
1 פתגמים צריכים שני טעמים אסור
 2 בהובין חיים בין מבושלים ר והנה
 3 או קונס צרים וחיטה שני טעמים
 4 מותר לכוס חיים
 5 היליו פרו
 6 הנודר מן הירק מותר בדיקנים
 7 וקדיבה אסור צמר גליל לר עקרון
 8 והלא אומר אדם לישלחו קט לנו
 9 ירק והוא אסר לא מעצאתי אלא ויל
 10 דילוניס אם להם וכן הדבר לא שני
 11 אלא הלא לא מעצאתי אלא קיטניות
 12 אלא שחידלושין בכלי היק אסור
 13 כפול המיני לח ומותר בכיש ב
 14 הנודר מן הדג אסור בפול המערי
 15 היבש דברי מאיר וחכם אסר אינו
 16 אסור אלא מחמת שנת המינין ר לא
 17 אסר הנודר מן הקבואה אינו אסור
 18 אלא מחמת שנת המינין אכל הנודר
 19 מן הדג אסור בכלי מותר בפירות
 20 האילן וניק 1 הנודר מן הכסו
 21 מותר בשוק דברי יונה ובהמילה אם
 22 קונס עמר עולה עלי מותר לכסות
 23 דגוי עמר פשתן עולה עלי מותר לכס
 24 באניני פשתן ר והנה אם הכל לפי
 25 הנדר טעון והנין והיה רחוק קשה
 26 אם קונס עמר ופשתן עולין עלי
 27 מותר לכסות ואסור לה פשיל אחר
 28 ד הנודר מן הבית מותר בעלייה
 29 דברי ר מאיר וחכם אין עלייה בכל

1 פתגמים צריכים שני טעמים אסור
 2 בהובין חיים בין מבושלים ר והנה
 3 או קונס צרים וחיטה שני טעמים
 4 מותר לכוס חיים
 5 היליו פרו
 6 הנודר מן הירק מותר בדיקנים
 7 וקדיבה אסור צמר גליל לר עקרון
 8 והלא אומר אדם לישלחו קט לנו
 9 ירק והוא אסר לא מעצאתי אלא ויל
 10 דילוניס אם להם וכן הדבר לא שני

Transcription alignment



טומאה לפנייהם נאמנים לומר לא נגענו וכן *בתינוקות שביניהם יונאים חוץ לפתח בית הבד ופונים לאחורי הגדר והן טהורים עד כוונת *ירחקו *ויהוא טהורין על כד' שיהא רואין * הַבְּדֵדִים וְהַבְּזוּרִים כִּיּוֹן שֶׁהֵכֵינָם לַרְשׁוֹת הַמַּעֲרָה דִּיּוּן דְּבָרֵי הַמַּאֲוֵר לְיוֹסֵי אֲחֵי זָרִיךְ לַעֲמוּד עֲלֵיהֶם עַד שֶׁיִּטְבְּלוּ ה' שִׁמְעוֹן אֲחֵי אִם טהורים להם זריך לעמוד עליהם עד שיטבולו אס טמאים להם אינו זריך לעמוד עליהם עד שיטבולו * הַנּוֹתָן מִן הַסֵּלִים וּמִן *הַמַּשְׁטָח שֶׁל אֲדָמָה בֵּית שְׁמִי אֲחֵי נוֹתָן בִּידֵים טְהוֹרוֹת וְאִם נָתַן בִּידֵים טְמֵאוֹת טִימְאֵן בֵּית הַלֵּל אֲחֵי נוֹתָן *נוֹתָן בִּידֵים טְמֵאוֹת וּמַפְרִישׁ תְּרוֹמָתוֹ בְּטַהֲרָה מִן *הַעֲבִיט וּמִן *הַמַּשְׁטָח שֶׁל עֲלִים הַכֵּל שׁוּיִם שֶׁהוּא נוֹתָן בִּידֵים טְהוֹרוֹת וְאִם נָתַן בִּידֵים טְמֵאוֹת טִימְאֵן *ה' הַאֲכֹל מִן הַסֵּלִים וּמִן הַמַּשְׁטָח שֶׁל אֲדָמָה אִף עַל פִּי *מִבּוֹקְעוֹת וּמִנְסוּפוֹת 10 לַגַּת הָרִי הַגַּת טְהוֹרָה מִן הַעֲבִט וּמִן הַמַּשְׁטָח שֶׁל עֲלִים *כֹּפֵל מוֹנֵנוּ גִרְגַּר יִסְדֵי אִם יֵשׁ לוֹ חוֹתֵם טְהוֹר *אֵין לוֹ חוֹתֵם טְמֵא כִּפְלוּ מוֹנֵנוּ עֲנָבִים וְדִרְכָן בַּמָּקוֹם *מַפְסֵה כ/ כְּבִינָה מִכוּוֵן טְהוֹר יֵתֵר מְכַבֵּינָה טְמֵא *שִׁמְכוּוֹן שִׁנְאֵת טִיפָה רֵאשׁוֹנָה נִיטְמַת כְּבִינָה * וְ מִי שֶׁהִיא עוֹמֵד וּמְדַבֵּר עַל שֶׁפֶת הַבּוֹר *נְתוּה לְנוֹרָה מַפְיו 15 אִם סֶפֶק הַגִּיעָה לְבוֹר סֶפֶק לֹא הַגִּיעָה סִפְקוֹ טְהוֹר * הַזּוֹלֵף אֶת הַבּוֹר נִמְצָא שָׂרָן בְּרֵאשׁוֹנָה כּוֹלֵם טְמֵאוֹת בְּאֶחְרוֹנָה הִיא טְמֵאָה וְכוֹלֵם טְהוֹרוֹת אֲמֵתִי בּוֹמֵן שֶׁהוּא זּוֹלֵף בְּכֵל אֶחָת וְאֶחָת אֲבָל אִם הִיא זּוֹלֵף בְּמִצַּח *נִמְצָא שָׂרָן בְּאֶחָת מֵהֶם הִיא טְמֵאָה בְּלֶבֶד *ה' אֲמֵתִי בּוֹמֵן שֶׁהוּא בּוֹדֵק וְלֹא מַכְסֵה אִו מַכְסֵה וְלֹא בּוֹדֵק הִיא בּוֹדֵק וּמַכְסֵה *נִמְצָא שָׂרָן בְּבִית הַכֵּל טְמֵא בְּבוֹר הַכֵּל 20 טְמֵא *וּבְמִצַּח הַכֵּל טְמֵא * הַ בֵּין *הַעֲיֹגֻלִים לְזוּגֵי רֶשֶׁת הַרְבִּים כֶּרֶם *לִפְנֵי בּוֹרִים רֶשֶׁת הַרְבִּים אֲמֵתִי בּוֹמֵן שֶׁהַרְבִּים כִּנְכָסִים בּוֹ וְיוֹנָאִים בּוֹ כְּלֵי בֵּית *הַבַּר וְשֶׁל גַּת וְהַעֲקַל בּוֹמֵן שֶׁן שֶׁל עֵץ מִגְבָּן וְהֵן טְהוֹרִים בּוֹמֵן שֶׁן שֶׁל גַּמִּי מִיִּשְׁכָן כֵּל שֵׁנִים עֲשָׂר חֲדָשׁ אִו חוֹלְטֵן בְּחֵמִים ה' יוֹסֵי אֲחֵי נָתַן שִׁיבּוֹלֵת הַנֶּהָר דִּיּוּן * הַל' ה' חֲסִילָה מַסְכָּתָא



232

טומאה לפנייהם נאמנים לומר לא נגענו וכן בתינוקות שביניהם יונאים חוץ לפתח בית הבד ופונים לאחורי הגדר והן טהורים עד כוונת ירחקו ויהוא טהורין על כד' שיהא רואין * הַבְּדֵדִים וְהַבְּזוּרִים כִּיּוֹן שֶׁהֵכֵינָם לַרְשׁוֹת הַמַּעֲרָה דִּיּוּן דְּבָרֵי הַמַּאֲוֵר לְיוֹסֵי אֲחֵי זָרִיךְ לַעֲמוּד עֲלֵיהֶם עַד שֶׁיִּטְבְּלוּ ה' שִׁמְעוֹן אֲחֵי אִם טהורים להם זריך לעמוד עליהם עד שיטבולו אס טמאים להם אינו זריך לעמוד עליהם עד שיטבולו * הַנּוֹתָן מִן הַסֵּלִים וּמִן *הַמַּשְׁטָח שֶׁל אֲדָמָה בֵּית שְׁמִי אֲחֵי נוֹתָן בִּידֵים טְהוֹרוֹת וְאִם נָתַן בִּידֵים טְמֵאוֹת טִימְאֵן בֵּית הַלֵּל אֲחֵי נוֹתָן *נוֹתָן בִּידֵים טְמֵאוֹת וּמַפְרִישׁ תְּרוֹמָתוֹ בְּטַהֲרָה מִן *הַעֲבִיט וּמִן *הַמַּשְׁטָח שֶׁל עֲלִים הַכֵּל שׁוּיִם שֶׁהוּא נוֹתָן בִּידֵים טְהוֹרוֹת וְאִם נָתַן בִּידֵים טְמֵאוֹת טִימְאֵן *ה' הַאֲכֹל מִן הַסֵּלִים וּמִן הַמַּשְׁטָח שֶׁל אֲדָמָה אִף עַל פִּי *מִבּוֹקְעוֹת וּמִנְסוּפוֹת 10 לַגַּת הָרִי הַגַּת טְהוֹרָה מִן הַעֲבִט וּמִן הַמַּשְׁטָח שֶׁל עֲלִים *כֹּפֵל מוֹנֵנוּ גִרְגַּר יִסְדֵי אִם יֵשׁ לוֹ חוֹתֵם טְהוֹר *אֵין לוֹ חוֹתֵם טְמֵא כִּפְלוּ מוֹנֵנוּ עֲנָבִים וְדִרְכָן בַּמָּקוֹם *מַפְסֵה כ/ כְּבִינָה מִכוּוֵן טְהוֹר יֵתֵר מְכַבֵּינָה טְמֵא *שִׁמְכוּוֹן שִׁנְאֵת טִיפָה רֵאשׁוֹנָה נִיטְמַת כְּבִינָה * וְ מִי שֶׁהִיא עוֹמֵד וּמְדַבֵּר עַל שֶׁפֶת הַבּוֹר *נְתוּה לְנוֹרָה מַפְיו 15 אִם סֶפֶק הַגִּיעָה לְבוֹר סֶפֶק לֹא הַגִּיעָה סִפְקוֹ טְהוֹר * הַזּוֹלֵף אֶת הַבּוֹר נִמְצָא שָׂרָן בְּרֵאשׁוֹנָה כּוֹלֵם טְמֵאוֹת בְּאֶחְרוֹנָה הִיא טְמֵאָה וְכוֹלֵם טְהוֹרוֹת אֲמֵתִי בּוֹמֵן שֶׁהוּא זּוֹלֵף בְּכֵל אֶחָת וְאֶחָת אֲבָל אִם הִיא זּוֹלֵף בְּמִצַּח *נִמְצָא שָׂרָן בְּאֶחָת מֵהֶם הִיא טְמֵאָה בְּלֶבֶד *ה' אֲמֵתִי בּוֹמֵן שֶׁהוּא בּוֹדֵק וְלֹא מַכְסֵה אִו מַכְסֵה וְלֹא בּוֹדֵק הִיא בּוֹדֵק וּמַכְסֵה *נִמְצָא שָׂרָן בְּבִית הַכֵּל טְמֵא בְּבוֹר הַכֵּל 20 טְמֵא *וּבְמִצַּח הַכֵּל טְמֵא * הַ בֵּין *הַעֲיֹגֻלִים לְזוּגֵי רֶשֶׁת הַרְבִּים כֶּרֶם *לִפְנֵי בּוֹרִים רֶשֶׁת הַרְבִּים אֲמֵתִי בּוֹמֵן שֶׁהַרְבִּים כִּנְכָסִים בּוֹ וְיוֹנָאִים בּוֹ כְּלֵי בֵּית *הַבַּר וְשֶׁל גַּת וְהַעֲקַל בּוֹמֵן שֶׁן שֶׁל עֵץ מִגְבָּן וְהֵן טְהוֹרִים בּוֹמֵן שֶׁן שֶׁל גַּמִּי מִיִּשְׁכָן כֵּל שֵׁנִים עֲשָׂר חֲדָשׁ אִו חוֹלְטֵן בְּחֵמִים ה' יוֹסֵי אֲחֵי נָתַן שִׁיבּוֹלֵת הַנֶּהָר דִּיּוּן * הַל' ה' חֲסִילָה מַסְכָּתָא

מסכת מקוות פרק ראשון
 שש מעלות במקוואה
 זו למעלה מזו וזו
 למעלה מזו מי גבאים שותה טמא ושתה טהור טמא שותה טמא ומילא בכלי טהור טמא שותה טמא וכפל ככר של תרומה אס *הריק טמא *אס לא *הריח טהור * מילא בכלי טמא ושתה טהור טמא מילא בכלי טמא ומילא בכלי טהור

מסכת מקוות פרק ראשון

שש מעלות במקוואה
 זו למעלה מזו וזו
 למעלה מזו מי גבאים שותה טמא ושתה טהור טמא שותה טמא ומילא בכלי טהור טמא שותה טמא וכפל ככר של תרומה אס *הריק טמא *אס לא *הריח טהור * מילא בכלי טמא ושתה טהור טמא מילא בכלי טמא ומילא בכלי טהור

* כן הוא בכתיבת היד.

[טהרות פרק י' הל' ג' - מקוות פרק א' הל' ז']

טומאה לפנייהם נאמנים לומר לא נגענו וכן בתינוקות שביניהם יוצאים חוץ לפת" בית הבד ופונים לאחורי הגדר והן טהורים עד כמה ירחקו ויהוא טהורים על כדי שיהא רואן" ה' הדדדים והבזורים כיון שהכניסן לרשות המערה דיוו דב"ר ר' מאיר ר' יוסי אולף נריך לעמוד עליהם עד שיטבולו ר' שמעון אולף אס טהורים להם צריך לעמוד עליהם עד שיטבולו אס טמאים להם אינו צריך לעמוד עליהם עד שיטבולו **ז'** הנותן מן הסלים ומן המשטח של אדמה בית שמיי אומי נותן בידים טהורות ואם נתן בידים טמאות טימאן בית הלל אומרי' נותן נותן בידים טמאות ומפריש תרומתו בטהרה מן העביט ומן המשטח של עליהם **ח'** הנותן בידים טמאות ומפריש תרומתו בטהרה מן העביט ומן המשטח של עליהם הכל שווים שהוא נותן בידים טהורות ואם נתן בידים טמאות טימאן **ט'** האוכל מן הסלים ומן המשטח של אדמה אף על פי מבוקעות ומימאן לגת הרי הגת טהורה מן העבט ומן המשטח של עליהם נפל ממנו גרגר ית"א יש לו חותם טהור אין לו חותם טמא נפלו ממנו ענבים ודרכן במקום מסת"ב כביצה מכיון טהור יתר מכביצה טמא שמכוון שיצאת טיפה ראשונה ניס"א בכביצה **י'** מי שהיה עומד ומדבר על שפת הבור נתזה צנורה מפיו ספק הגיעה לבור ספק לא הגיעה ספיקו טהור **יא'** הזולף את הבור נמצא שרץ בראשונה כולם טמאות באחרונה היא טמאה וכולם טהורות אמתי בזמן שהוא זולף בכל אחת ואחת אבל אם היה זולף במחץ נימצא שרץ באחת מהם היא טמאה בלבד **יב'** אמתי בזמן שהוא בודק ולא מכסה או מכסה ולא בודק היה בודק ומכסה נמצא שרץ בחבית הכל טמא בבור הכל טמא ובמחץ הכל טמא **יג'** בין העיגולים לזוגין רשות הרבים כרס לפני בוצרים רשות הרבים אמתי בזמן שהרבים נכנסים בזו ויוצאים בזו כלי בית הבר ושל גת והעקל בזמן שהן של עץ מנגבן והן טהורים בזמן שהן של גמין מייסגן כל שנית עשר חמש או חלפני בחמים ר' יוסי אומי אם נתנן לשיבול"ה הנהר דיו **יד'** **טו' חסילה מסכתא**

מסכת מקוות פרק ראשון שש מעלות במקוות
 זו למעלה מזו וזו ל
 למעלה מזו מי גבאים שמה טמא ושמה טהור טמא שמה טמא ומילא בכלי טהור טמא שמה טמא ונפל ככר של תרומה אם הריח טמא אם לא הריח טהור **יז'** מילא בכלי טמא ושמה טהור טמא מילא בכלי טמא ומילא בכלי טהור

26
27
28
29
30

הנהר דיו "הל' ט" חסילה מסכתא"

מסכת מקוות פרק ראשון שש מעלות במקוות

זו למעלה מזו וזו ל

למעלה מזו מי גבאים שמה טמא ושמה טהור טמא שמה טמא ומילא בכלי טהור טמא שמה טמא ונפל ככר של תרומה אם הריח טמא אם לא הריח טהור

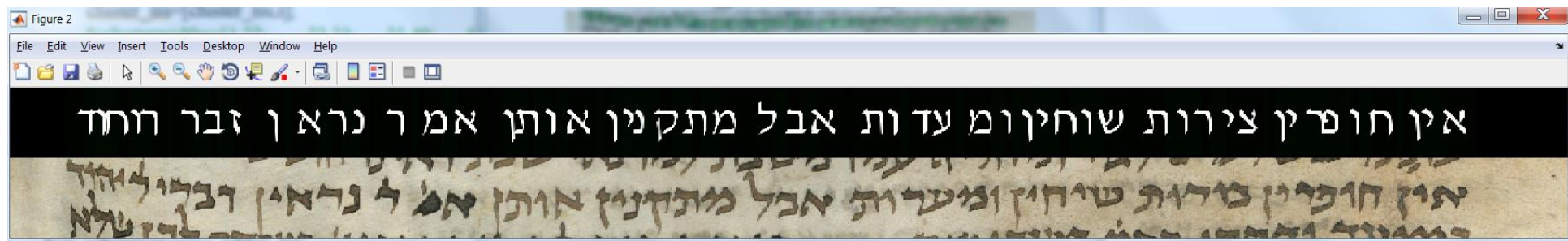
30 מילא בכלי טמא ושמה טהור טמא מילא בכלי טמא ומילא בכלי טהור

1 טומאה לפנייהם נאמנים לומר לא נגענו וכן בתינוקות שביניהם יוצאים חוץ לפת" בית הבד ופונים לאחורי הגדר והן טהורים עד כמה ירחקו ויהוא טהורים על כדי שיהא רואן" ה' הדדדים והבזורים כיון שהכניסן לרשות המערה דיוו דב"ר ר' מאיר ר' יוסי אומי צריך לעמוד עליהם עד שיטבולו ר' שמעון אומי אס טהורים להם צריך לעמוד עליהם עד שיטבולו אס טמאים להם אינו צריך לעמוד עליהם עד שיטבולו **ז'** הנותן מן הסלים ומן המשטח של אדמה בית שמיי אומי נותן בידים טהורות ואם נתן בידים טמאות טימאן בית הלל אומרי' נותן נותן בידים טמאות ומפריש תרומתו בטהרה מן העביט ומן המשטח של עליהם הכל שווים שהוא נותן בידים טהורות ואם נתן בידים טמאות טימאן **ט'** האוכל מן הסלים ומן המשטח של אדמה אף על פי מבוקעות ומימאן לגת הרי הגת טהורה מן העבט ומן המשטח של עליהם נפל ממנו גרגר ית"א יש לו חותם טהור אין לו חותם טמא נפלו ממנו ענבים ודרכן במקום מסת"ב כביצה מכיון טהור יתר מכביצה טמא שמכוון שיצאת טיפה ראשונה ניס"א בכביצה **י'** מי שהיה עומד ומדבר על שפת הבור נתזה צנורה מפיו ספק הגיעה לבור ספק לא הגיעה ספיקו טהור **יא'** הזולף את הבור נמצא שרץ בראשונה כולם טמאות באחרונה היא טמאה וכולם טהורות אמתי בזמן שהוא זולף בכל אחת ואחת אבל אם היה זולף במחץ נימצא שרץ באחת מהם היא טמאה בלבד **יב'** אמתי בזמן שהוא בודק ולא מכסה או מכסה ולא בודק היה בודק ומכסה נמצא שרץ בחבית הכל טמא בבור הכל טמא ובמחץ הכל טמא **יג'** בין העיגולים לזוגין רשות הרבים כרס לפני בוצרים רשות הרבים אמתי בזמן שהרבים נכנסים בזו ויוצאים בזו כלי בית הבר ושל גת והעקל בזמן שהן של עץ מנגבן והן טהורים בזמן שהן של גמין מייסגן כל שנית עשר חמש או חלפני בחמים ר' יוסי אומי אם נתנן לשיבול"ה הנהר דיו **יד'** **טו' חסילה מסכתא**

מסכת מקוות פרק ראשון שש מעלות במקוות
 זו למעלה מזו וזו ל
 למעלה מזו מי גבאים שמה טמא ושמה טהור טמא שמה טמא ומילא בכלי טהור טמא שמה טמא ונפל ככר של תרומה אם הריח טמא אם לא הריח טהור **יז'** מילא בכלי טמא ושמה טהור טמא מילא בכלי טמא ומילא בכלי טהור

1 טמא מילא בכלי טמא ונפל ככר של תרומה אם הריח טמא לא הידיח טהור **ג'** נפלו מים טמאים ושמה טהור טמא נפלו מים טמאים ומילא

manuscript OCR



What could we need computational humanities for?

- Speed up process of cataloguing, deciphering, translating, publishing texts and understanding human culture
[DH (new forms of publishing, documenting, visualizing, analyzing data)]
- ‘Objectify’ certain fields (e.g. paleography)

Paleographers have great pain to assess whether

- two documents written in very distant periods of activity of a scribe were written by the same scribe or whether they are the product of a scribal school.
- two documents written in two different registers (e.g. cursive / booksript) were written by the same scribe.
- two documents written in two different scripts (e.g. Syriac / Arabic) were written by the same scribe.

It would be very interesting to develop mathematical models for these and other variabilities (e.g. amateur/professional distinction)

One should be able with massive modern online ground truth to learn models for different scripts and problems that we can try to apply to historical documents.

No black boxes!

What could we need computational humanities for?

- Speed up process of cataloguing, deciphering, translating, publishing texts and understanding human culture
 - [DH (new forms of publishing, documenting, visualizing, analyzing data)]
- ‘Objectify’ certain fields (e.g. paleography)
- Connecting data and data mining to ask (and try to answer) new questions
 - e.g. interdisciplinary methodological approaches by combining analyses of spatial, textual, visual, audio, linguistic data
 - e.g. diachronic NLP
 - e.g. comparative transcultural paleography

Can we develop mathematic models to calculate the phylogenetics of long term script development (similar to genetics or linguistics)?

some elements will be absent in the pedigree

א ב ג ד ה ו ז ח ט י כ ל מ נ ס ע פ צ ק ר ש ת

Asoka inscr.
א א
ב ב
ג ג
ד ד
ה ה
ו ו
ז ז
ח ח
ט ט
י י
כ כ
ל ל
מ מ
נ נ
ס ס
ע ע
פ פ
צ צ
ק ק
ר ר
ש ש
ת ת

As55 6
א א
ב ב
ג ג
ד ד
ה ה
ו ו
ז ז
ח ח
ט ט
י י
כ כ
ל ל
מ מ
נ נ
ס ס
ע ע
פ פ
צ צ
ק ק
ר ר
ש ש
ת ת

א ב ג ד ה ו ז ח ט י כ ל מ נ ס ע פ צ ק ר ש ת

א ב ג ד ה ו ז ח ט י כ ל מ נ ס ע פ צ ק ר ש ת

What we need

Tools and infrastructures such as DivaDIA

Engineers to maintain them

reciprocal listening and teaching