

CATCH 2020

$\bullet \bullet \bullet$

Olga Beloborodova, Wout Dillen, Joshua Schäuble

ACDC | Antwerp Centre for Digital Humanities and Literary Criticism



CMG | Centre for Manuscript Genetics



We are delighted to announce the publication of our fifth genetic edition: Malone meurt / Malone Dies. The accompanying monograph on its genesis is available in Belgium through University Press Antwerp, and internationally through Bloomsbury Academic.

What is CATCH 2020?

Computer Assisted Transcription of Complex Handwriting 2020



Opening new horizons



Computer Assisted ≠ Fully Automated

Why do we not (yet) believe in a fully automated approach?

As (genetic) editors it is our task to produce accurate transcriptions

and to enrich them with annotations/mark-up

95% accuracy (5% Character Error Rate) is not good enough for editors

(not even 99.9% would be good enough to waive proof reading)

HTR + manual proof reader

VS.

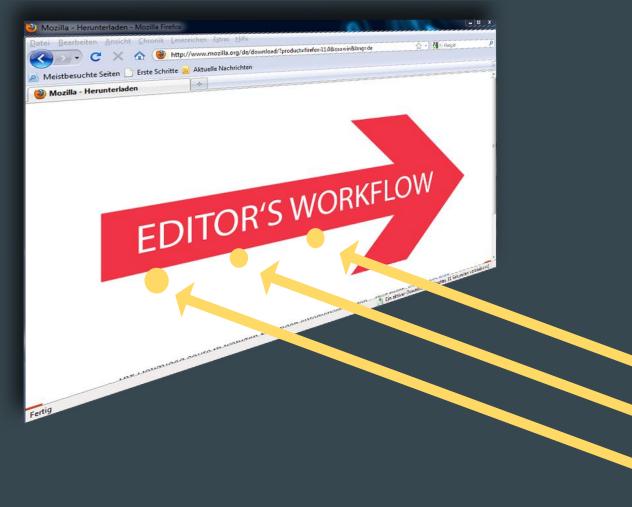
HIGHLY ACCURATE TRANSCRIPT

Manual Transcription

+ Transkribus API assistance

How can the core functions of Transkribus* best be integrated into the workflows of digital scholarly editors?

*Layout Analysis, HTR, Transcript-Facsimile Mapping



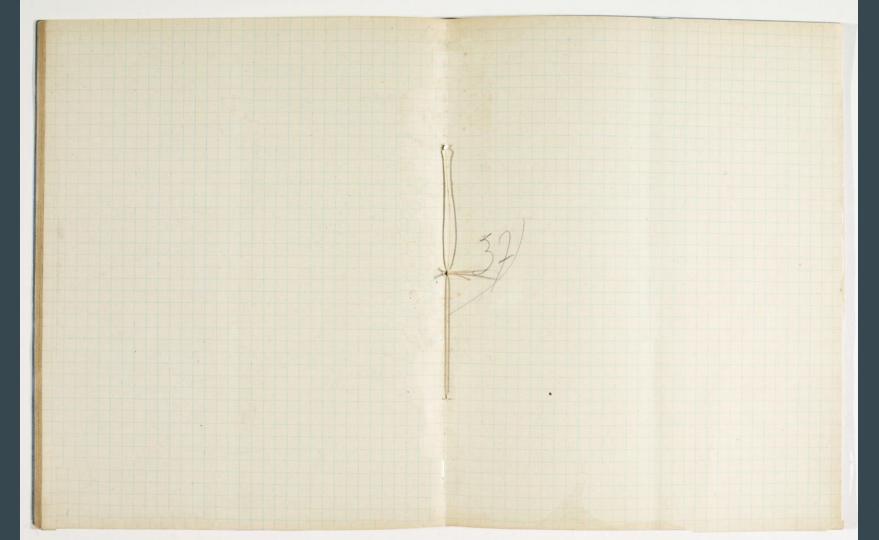


Transkribus API

Linking Revision Campaigns (TEI) & Layout Analysis

We want to capture the "writing order" (on a macro and micro level)

James Joyce – Drafts for Ulysses



A torential rain some town for to Hoot which of the mark mark which number at least for his walk mile and lifter the organd, the erubite bretake who aprimi heres the last contra of religion & the hero marky in a book of muldy water his cassook above his hoary head hupplication. Hars by the Hock stort the gring the figure of concerter this harge being of through which his greegome lyly glared furiarsh. an be awaiter the find dign he the start the est the how it is the shart we can a secapitale to shelp which has been a flack this feel but recensory office. The non sher rela y anotion was reached when the blughing bive elect but divinge, the puried ranks of the bystander and flin tonot of an incald about to sie for the fulle the line about his week and Kitled pay houster the officent pritable areas which the second of The abour & works the poore to him to the

17

A torential rain boures for for to the theat the the walk and which number at least for his walk and which and lifts the organd, the erubite brelate who atministered the last contra of religion & the hero marky in a pool of muldy water his cassort above his hoary head hupplication. Hars by the Hock stort the grin - fique of concealed in the gallon pat Knoch which his greene lyly glared furiarch. an be awaiter the find time her her No slipply this feel but accorning of elot was reached when the bluching bive elect but and the pires ranks of the bystander and flin borot of an ingeald about to sie for the they have about this week and Kitled pay houster an officient pritable areas which the second of har abour & works bergan the poore to him to the

A tarential rain poures town for bared beaut to the multitude which numbers at least for him ver A the shandary bretake who at init keep the last contr. h. of religion & The hero mertys maleasens Will in a work der. I have this in a book of muldy water, his Kuble way Cassook above his hoary head uble analet Light of structe payed out Suer Baching theirs the darian King the read Derknathing Elected and a constant Heroigh which his greesome lyly glared furiarch. an be awaiter the find time her the An an atter marken weapon by honing it efor This burger a decapitate I shelp which had been dronglyful h midely by the as i was of A This fell but recensory office. willing derkines dos the non stres when they meia blood mox - blughing bive elect but - los best they of the price and En mos work about to sie for her fulle. Steward of an an algat Uchin. The boun - out t the lung about his weigh Cuts and Dep and hited pay in the En eler,) ha an officent pritable areas conveg tuse ressels have go which the second of bergane the abour & works the work to him to the

19

By hister to cellent the affect coupting A Harenkial rain poures som for bared heart tip the malk thate which numbers at least for him the prelite who at init but the last ben prod the & hand a w the author The My maleason Will in a work der. I have this in a book of muldy water, his Kube water e. Cassock above his hoary head Myslicking Hans hayey tak stylicking Hans hayey tak stylicking the figure for the execution the part of the concealed inthe ten gallon part in our hos round eyebole Au Suen Being Hing the darian King Le van Derhalling And welling and a stand of the terosh which his green lyly glared fur iarsh, an be availes the find time her how ithe Ester the egge for ing it store this budy and a decapitules in april Juccession a flood thouse the sheet to an an in of A Free seen to bushing to reach of anothis took to the they all the price to the they took to the they all the price to the they a love to the they all the price to the they was all to the they all they about the they took to the they about the they to all the they about the they they could the they about this weight Whin fell but recensory office. Bitter Cup in the an algat Bitter Cup in the an algat Bioring ture used by ten der,) K'an bergare which the second of the abour & reach & the work to him to the

finite and & cellent the affect coup & Horardial rain boures town for bare hearts tip the multitude which myn paket number at least for him the wellike who atministered the last confr to of religion to the last unerty A. & hand a res malesgang What in a work der. s have this the aluthor by in a book of muldy water his Kuble water f. the consignation Cassook above his hoary head hupplicking that to the block the central Jerenhow his her gallon pot Suren Berling Havile the davian King Le care Derhalli Onto weplaner un man of the and the through which this grossome why cores is the ly stared for ionsh, an be that the melol. awaiter the parel figne the Hered the estil in how the Seafor way in a decapital april way way in a flak the sheets which had been When the not weccychill) extricated un anity the segend i here in a regard in the replete the trian of the solid the replete the trian of the solid of the perfect the the solid the soli This fell but recensory office. Mos was reached when they blughing bite deit but bed they af the price rank The likeful the icking work rickin. The Conse Steward of the are a with fut m the hung about his week could 2 Clets and begg and Kithed pay hometer auch ? ha way in atthe Dana an officert pritable areas B correg tuse vessels which the second of Many berare man gard permitted ther arbour & reach & the work to him to de

Anite a cellent treatfact coupting A Harenkial rain poures soon for bare heart to the mulk take which numker at least for him the melate who atministered the last confer in a religion & the laro merty A. & hand a us maleasens the author the in a book of muldy water his Kuble water for the consignation Levenhold in the payer of the state of the stand of the stock state how the stage the sol too cealed in the gallon pate the central Succession of the state the davian King Onto wepplance teroigh which his greene why coreserver augarites the faith and a that the melal the tested the egge him how ithe this Burger a decapitates in april juccession a flood when the most Trashypets to shelp which had been weccychill) extricated un Then fell but recensory office Nor was reached when the Warding to serie have usy steward of an alga 2 orgen She have about his weigh " Clefs and begg and Kitled pay someter by the lar .) has B conog tuse user the offerent pritable areas diane 5 which the second of the abour & reachs the work to him to de

Juite & cellent the affect each this A Harenkial rain poures sown for to roskey to toll a tolk in the star has had be bare heart of the malk that which numbers at least for him the melate who atministered the last control of religion & the laro menty A. & handa wi malessan the author the Kuble and the find in a book of maring water his the consuguestion Levenhold the her gallon port then central the varian hiver Le lac Derhad apphances, And any and a contraction of the second of t through which his greesome lyly stared for arch. an be awaiter the figue lie tested the egge him how ithe this Burger a decapitates in april juccession a flood Trashypets to shelp which had been Then fell but recensory office Mos the non stres when the her bushing to serie hand B corog ture verily and the the hung about his weigh and Kitled pay someter En ler ? h'and an officient pritable areas which the second of ber and the abour & washi the work to him to de

23

pealiter fite a & cellent the thirth coupting A tarenkal rain sources som for the toostake of the walk and which number at least for him be and lifty thoyand, the erabite preliate who atmimilians the last conforth of religion & the lass mentys that in a most deris han this molester A. : handa ben proved maledgang the author the in a bool of mardy wate his the consumption of Kable was the central price were as a for the the hupplickin. Hay by the block Job call non the gallon bat y the trapers Such Balling Brill the darian ting Les nace Destration How a contraction of the second secon on epphance that an ment hard through which ding grocesome lyly stared for arch. an be awaiter the fact dign the Exter the experience of deceptor when the man en hue of the Trashypets to shelp which had been weegefill Aut wor a toller i Ray wood di Aut roma toller to ray wood di this regard the most and the second of This fell but recensory office Nor was reached when the Wor and the pine hand the replensive Reia was hear man about to se for the fulle Sheward of an an a wyl 2 the hung about his weigh Cuts and dep and Kitled pay homobel ander ? hay way in atthe an officient pertable area Monie 5) convertuse visit which the second of ber and the abour & reach & the work to him to the

24

Juite of & cellent, the affre A tarenkal rain poures som for the Rootyaks while the which the which bares hears to be much much which min pareck to wating tal and the thoyand. The eradite pretake who aprimi heres the last Whiter in a melition is the line menty A. & handa w maleascing in a book of maring water his Kube us hupplication. Have by the block cer Job call non the gallon bat Succession Raise the said King ber here Denknalling One explances MAN A CONTRACTION OF THE AND A CONTRACT OF T Heroigh which his greesome eye glares for institute awaites the spirit in the lester the egge of the horithe weappend and a seception in apply juccession a flack that the m Refue to the Thought the shelp which had been weegher A Agent 1 Allem & Road De King extricated with When fell but recensory office. Mos an non stan alter of anotion blushing bive elect but They the pried rank er replenitte to that useful the in or work of the former Theward of any any they have about his weigh Clup and beg and Kitled pay houster - ofl En br.) Kin an officent pritable area 4003 which the second of in atter home be vare coive tuse used the abour & reach & the work to him to de

Further Examples

Detecting highlighted text (underlined, deleted, superscripted, overwritten)

underlined	Lish
(blue)	
	rich vocabulary because read little . Cler Henneell (1867) Col. Richard
	Burke Hubitone D'Untallither protes heart church, 12 12: ales 120
	3. 2 Millial D - landis i atom long the in Malin
	A . I FI - and a AMUMANO 'S " " Land a Ama Al
	" I la ada silada di la dana i dana pouro pouro de se a da s
	Provent and the second of the
araaad aut	Averal CRisting Blissons & man har and in the second of the
crossed out	avitie pages y population georgeon properiou : "Reltitche Prein
(blue)	
` ´	Dichter Helicht Keine nemer, Kein vionendes sin henderent
	Kline united, kline Junde. are h. The merony the religion
	Dichter heite hour a de underendes Six dengesch Dichter Heitht Reine Tremess, kein Dishendes Six dengesch Keine unword, keine Sjunde. Die hirde heidniche uligion Rennt Keine Hille (Julius Jakonny): Cattle trade len jor
	than syn wetwe wolk coming of real ship would work the
	the syn metere ? vols comby of red shifts / bours ? brops for combs) his in for those tamers fallow bores eta to manye. Hayblere meat to de bette to do whole energy of ?
	and had in a super find of the and the
	Sos to feed oxe seeper bin y Banos and Haman
crossed out	within and it nearly hateral refury rate light
(red)	then bit to many rebe . Doug tike to batterit. Killed which
	tothe informed Zheabeth Well book making dadligh
	thank a store
	JAMES JOYCE, NOTES FOR ULYSSES, NLI, MS 36,639/3

Detecting interlinear additions

Small (10 × 10) miseralle bardy furnished room (barted or tiles floor J. La sketch. It cannot be his sites "he would not kup 4 pairs of eyes and a mirror). It may be suffered that it is his mother's room, which he has not been for a number ofyrand and is to orcupy A look after the fets, the considered the horner This circumstance has no bearing on the film and is all other instated they by highto on the wall. At stands near the door, and take it in Sucurta of cinages: dog , cat htting side side, staring at him , ninor ; window , con and dog - las staring at him ; caputant, the and fish, pairot staring at him; dog + car staring Pets not distar into per and al armor the Prenare by giving couble animal not attected

SAMUEL BECKETT, FIRST DRAFT FOR FILM, UOR, MS 1227/7/6/1.

Detecting Substitutions (by combining interlinear additions with deletions in the following line)

interlinear addition + deletion (close by) = substitution Nothing that up lains ung the crud to mach. For the had done tomothing of that machical, a well as a motional who tana

- 11 Nothing That explains why she cried so much. For she
- 12 had done something of great practical, as well as
- ¹³ of emotional seriousness. ^{importance}. _{Now it is -- I suppose --}

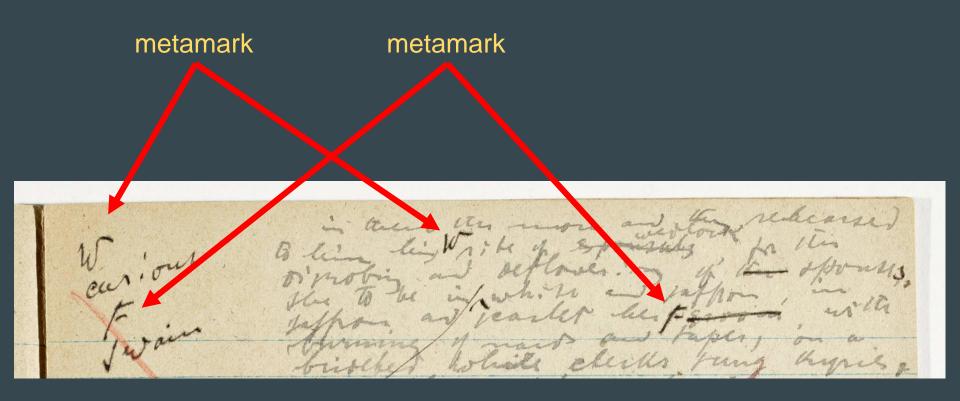
VIRGINIA WOOLF "A SKETCH OF THE PAST" AUTOBIOGRAPHICAL SKETCHES (1939),2

Genre-based layout detection

& Cadly, (Paces . Honney.) 1100 this my amand worke Pause. line of inconsecurity and manufit must, to sapphils are in course care he bother the buyenholds. (In 7 To customer, God damit, Lin, no. it's sides we dree as limits ! 14 six dags, to you tean one, mic dags, Cost made the world. Yes Sin the loss Su', Re WORLD , and you are and Morts use again & making un a fair of channes ai Me aunthe . 110000 of taylor standalised.) " But no lard, in terte loss - (distanjul jeten all donte - as do and - there a and look - Caning pare, proudly at my ango TRousers !" Pause. He terms at Made who has remained humanin, be exercised makes into a high forma lange, cirts it shouts, advance his leas towards All hand into Camilles his langh again Stence . from w May and and the has augh. Nell you lined so the but (bottom). and to are und (enafuer ed) fan ya stan 100 your, with mill mill and fait? (and under pury) another ner tit dend. (Nogg diappears a to tri, close to a o train o this. Nell doc air nears. Furgicely / May thing kingdow for a Marcusa ? O Same The ahrista . Easter blov. I than any This much I caused it as to ma?

34

Keyword spotting to trace metamarks



Version Comparison (illegible words might be legible on an earlier/later draft)

Version 1

alchnice shipt in diep 1052

Version 2 pakin'sen mandrigenand

What are your ideas? (Let's talk about it!)