



Università
Ca' Foscari
Venezia

	A	B	C	D	E	F	G	H	I	J
757							eft	eft	eft	eft
758							gewat	gewat	gewat	gewat
759							into	into	into	into
760							Galwalum	Galwalum	Galwalum	Galwalum
761	AD	AD								
762	Octauianus	Octauianus	Octauianus	Octauianus	Octauianus	Octauianus	Octauianus	Octauianus	Octauianus	Octauianus
763	ricsode	ricsode	rixode	rixode	rixode	rixode	rixade	rixade	rixade	rixade
764	lxvi	lxvi	lxvi	lxvi	lxvi	lxvi	lxvi	lxvi	lvi	lvi
765	wintra	wintra	wintra	wintra	wintra	wintra	wintra	wintra	wintra	wintra
766	on	on	on	on	on	on	on	on	on	on
767	þam	þam	þam	þam	þam	þam	þam	þam	þam	þam
768										
769			xlii	xlii	lii	lii	xlii	xlii	xlii	xlii
770	lxiiigeare	lxiiigeare	gere	gere	gere	gere	geare	geare	geare	geare
771	his	his	his	his	his	his	his	his	his	his
772	rices	rices	rices	rices	rices	rices	rices	rices	rices	rices
773	Crist	Crist	Crist	Crist	Crist	Crist	Crist	Crist	Crist	Crist
774	wæs	wæs	wæs	wæs	wæs	wæs	wæs	wæs	wæs	wæs
775	acenned	acenned	acenned	acenned	acenned	acenned	acenned	acenned	acenned	acenned
776	þa	þa								
777			Ða	Ða	Ða	Ða	Ða	Ða	Ða	Ða
778			þry	þry	ðry	ðry				
779	tungelwitgan	tungelwitgan	tungolwitigan	tungolwitigan	tungolwitigan	tungolwitigan	tungelwitegan	tungelwitegan	tungelwitegan	tungelwitegan
780	of	of	of	of	of	of	of	of	of	of
781	eastdæle	eastdæle	eastdæle	eastdæle	eastdæle	eastdæle	eastdæle	eastdæle	eastdæle	eastdæle
782	cuomon	cuomon	coman	coman	coman	coman	comon	comon	coman	coman
783										
784	to	to	to	to	to	to	to	to	to	to
785	þon	þon								

Normalisation

– a brief note

Marina Buzzoni

Bergen, 1st July 2022



Normalisation for alignment

- Alignment of the transcribed texts is a crucial step in digital stemmatology; it takes place before converting the (aligned) text into a string of characters that will be processed further.
- Alignment can be influenced by linguistic variation (orthographic, morphological, syntactic); therefore, some kind of normalisation is usually carried out by the editor.



Normalisation as an instrumental process

“Normalisation” in the context of alignment is not equivalent to normalisation that happens in transcription.

Editors can transcribe orthographic variation because they consider it important to be preserved in both the transcription and the collation output. However, [...], they may want to normalize orthographic variation because they do not want it to influence the alignment. In that case they need to normalize their tokens before inputting them in the collation software.

(Bleeker, Elli. *Mapping invention in writing: digital Infrastructure and the role of the editor*. 2017, 94)



Spadini, Elena. 2016. “Studi sul *Lancelot en prose*.” PhD
diss., Sapienza Università di Roma.

Witness	Tokens			
A	ge	-	te	conois
B	artu	je	te	conois
C	artus	je	te	conois

Before normalisation

Witness	Tokens			
A	-	ge	te	conois
B	artu	je	te	conois
C	artus	je	te	conois

After normalisation

Table 5. Correct alignment of spelling variants in *Lancelot en prose*



Open issues

Albeit only «instrumental», normalising for text alignment:

- can be extremely difficult, especially when an orthographic standard is lacking;
- is always time consuming (the process potentially involves all the linguistic layers in a text: orthographic – graphemic, spelling, numbers; punctuation; morphological; lexical; syntactical)
- requires a lot of work on the input, especially in specific, though not unusual cases such as multilingual alignment (e.g. in case of different scripts, a part of speech identifier can be used as the shadow normalisation property of the tokens).

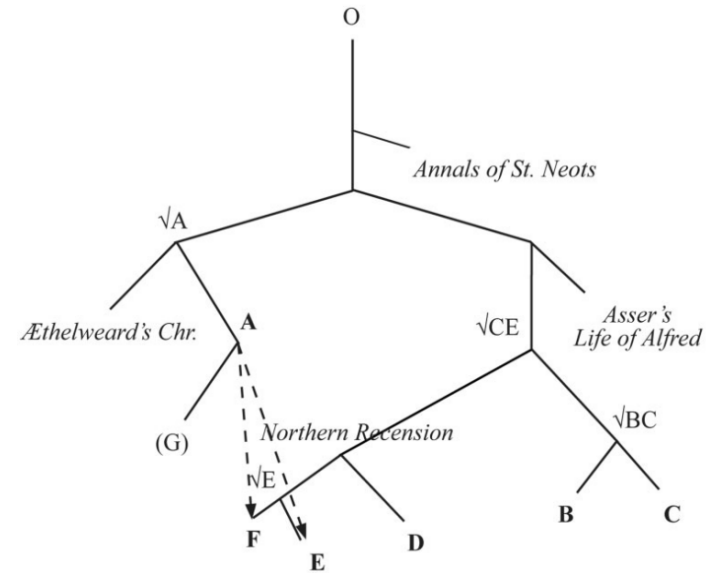
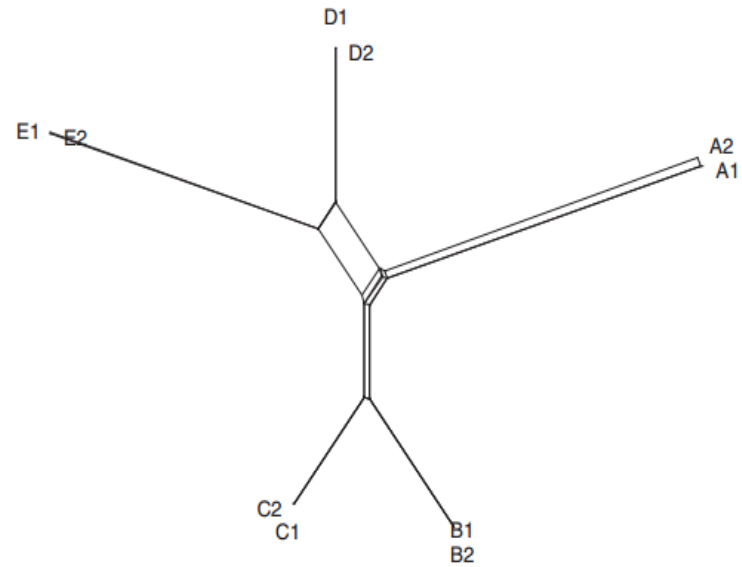


Questions

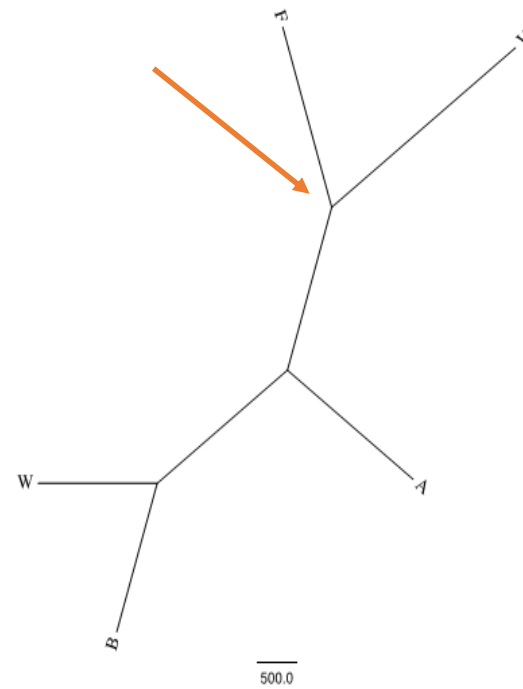
- How can we keep this instrumental normalisation process to a minimum?
- To what extent do the choices made by the editors at the alignment level affect the stemmatological output? Some (e.g. syntactic level) seem quite arbitrary.



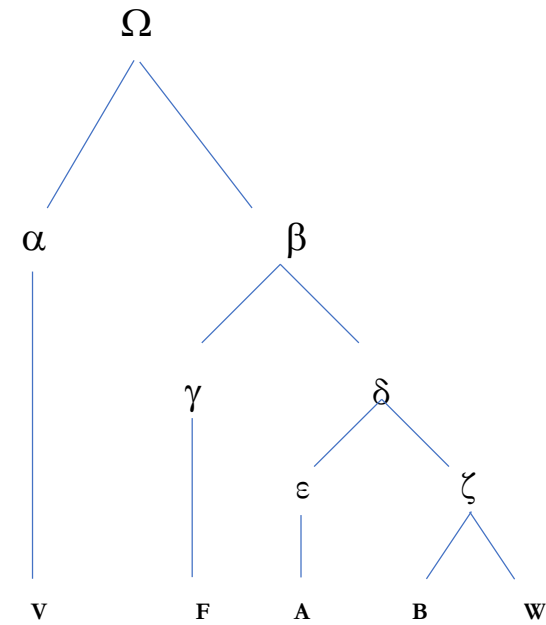
A-S Chronicle tradition: NN



Melione tradition



PAUP, PS





Università
Ca' Foscari
Venezia

<thanks to="allOfYou"/>