1. 前言：當人類遇上數位…

身為數位人文學家，常遇到有人問我：數位人文學與傳統人文學是否存在差異，以及有何差異？對於此問題，我的答案很直白：我認為無論是在數位環境還是非數位環境裡進行研究，其文學研究的本質以及證據/發現和見解之間的詮釋關係是不變的。然而，有了數位工具，我們閱讀文本的方式可以產生很大的改變：電腦使我們能以非線性的方式閱讀文本。語言的本質是線性的，畢竟文本即是字詞的線性排序，而字詞則即是字母的線性排序。人不可能同時說出兩個或兩個以上的字；資訊的流通是單向的。因此，人類乃是以線性的方法處理語言，並以線性的方法閱讀文本。由於文學是語言的藝術，因此文學也無可避免受到語言線性本質的約束。然而，數位人文學卻能跳脫線性的束縛。

那麼，以非線性的方法閱讀文本有何用處？傳統上，文本可以精讀，能以美學/文學的角度欣賞，但一旦經過數位化，文本便轉化為「一袋字詞」，能進行解構，能進行非線性處理。數位化將文本轉化成一種掙脫語言線性束縛的實體。如同人體細胞染色體可以定序成一串基因序列（ACGTs[1]），文本也能簡化成字詞、字母、頻率等向量，然後我們便能將統計數據以視覺化呈現，並進行「遠讀」（distant reading）。如同基因組可以定序並接受生物資訊分析，然後將結果繪成樹枝狀圖以表達不同物種的共同祖先趨異演化的過程，在數位人文學領域中我們也能將古典小說語料轉化成表達字詞頻率的高維度向量，並能透過分群樹狀結構圖和多維度散佈圖將文本之間的複雜關係以視覺化呈現。

非線性文本閱讀法有以下若干功能：

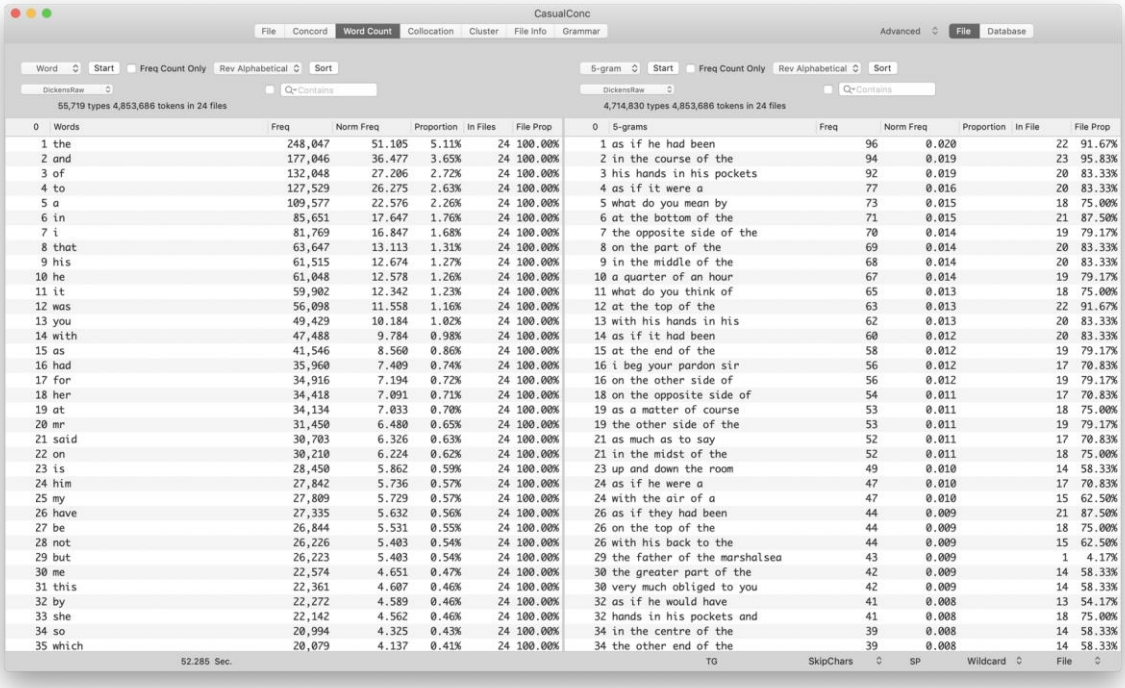
- 計算、分類、索引、解構、去脈絡、自助抽樣
- 將文本轉化為表達字詞、數字、頻率等之向量
- 資料/文本採礦、視覺化、「遠讀」（Moretti, 2013）

2. 非線性閱讀作為宏觀/遠閱讀

表 1 顯示高頻字詞與五元語法（五個字的語塊序列）。字頻統計是數位人文學領域中最基本的操作。經過訓練的分析師可以透過字平頻統計排名前十、二十、三十項來準確判斷該文本的種類和語域，但是若要猜測某語料中最常出現的五元語法為何，其難度非同小可。頻率 n 元語法（n 即一序列的字數）能顯示語料的措辭、指涉與風格上的特色。以「his hands in his pockets」這個五元語法為例，我們可以透過關鍵字索引（KWIC）來「垂直」（vertically）閱讀文本。透過這樣的垂直閱讀，我們便能發現這個片語其實屬於更大的措辭架構，例如「with his hands in his pockets」，或是搭配其他身體部位的描寫。

局部文本功能

[1] 染色體可以解構成四種鹼基，分別是腺嘌呤（Adenine, 縮寫為 A）、鳥嘌呤（Guanine, G）、胞嘧啶（Cytosine, C）與胸腺嘧啶（Thymine, T）



CasualConc

File | Concord | Word Count | Collocation | Cluster | File Info | Grammar

Advanced | File | Database

Word | Start | Freq Count Only | Rev Alphabetical | Sort

DickensRaw | Contains

55,719 types 4,853,686 tokens in 24 files

0	Words	Freq	Norm Freq	Proportion	In Files	File Prop
1	the	248,047	51.105	5.11%	24	100.00%
2	and	177,046	36.477	3.65%	24	100.00%
3	of	132,048	27.206	2.72%	24	100.00%
4	to	127,529	26.275	2.63%	24	100.00%
5	a	109,577	22.576	2.26%	24	100.00%
6	in	85,651	17.647	1.76%	24	100.00%
7	i	81,769	16.847	1.68%	24	100.00%
8	that	63,647	13.113	1.31%	24	100.00%
9	his	61,515	12.674	1.27%	24	100.00%
10	he	61,048	12.578	1.26%	24	100.00%
11	it	59,902	12.342	1.23%	24	100.00%
12	was	56,098	11.558	1.16%	24	100.00%
13	you	49,429	10.184	1.02%	24	100.00%
14	with	47,488	9.784	0.98%	24	100.00%
15	as	41,546	8.560	0.86%	24	100.00%
16	had	35,960	7.409	0.74%	24	100.00%
17	for	34,916	7.194	0.72%	24	100.00%
18	her	34,418	7.091	0.71%	24	100.00%
19	at	34,134	7.033	0.70%	24	100.00%
20	mr	31,450	6.480	0.65%	24	100.00%
21	said	30,703	6.326	0.63%	24	100.00%
22	on	30,210	6.224	0.62%	24	100.00%
23	is	28,450	5.862	0.59%	24	100.00%
24	him	27,842	5.736	0.57%	24	100.00%
25	my	27,809	5.729	0.57%	24	100.00%
26	have	27,335	5.632	0.56%	24	100.00%
27	be	26,844	5.531	0.55%	24	100.00%
28	not	26,226	5.403	0.54%	24	100.00%
29	but	26,223	5.403	0.54%	24	100.00%
30	me	22,574	4.651	0.47%	24	100.00%
31	this	22,361	4.607	0.46%	24	100.00%
32	by	22,272	4.589	0.46%	24	100.00%
33	she	22,142	4.562	0.46%	24	100.00%
34	so	20,994	4.325	0.43%	24	100.00%
35	which	20,079	4.137	0.41%	24	100.00%

52.285 Sec.

5-gram | Start | Freq Count Only | Rev Alphabetical | Sort

DickensRaw | Contains

4,714,830 types 4,853,686 tokens in 24 files

0	5-grams	Freq	Norm Freq	Proportion	In File	File Prop
1	as if he had been	96	0.020		22	91.67%
2	in the course of the	94	0.019		23	95.83%
3	his hands in his pockets	92	0.019		20	83.33%
4	as if it were a	77	0.016		20	83.33%
5	what do you mean by	73	0.015		18	75.00%
6	at the bottom of the	71	0.015		21	87.50%
7	the opposite side of the	70	0.014		19	79.17%
8	on the part of the	69	0.014		20	83.33%
9	in the middle of the	68	0.014		20	83.33%
10	a quarter of an hour	67	0.014		19	79.17%
11	what do you think of	65	0.013		18	75.00%
12	at the top of the	63	0.013		22	91.67%
13	with his hands in his	62	0.013		20	83.33%
14	as if it had been	60	0.012		20	83.33%
15	at the end of the	58	0.012		19	79.17%
16	i beg your pardon sir	56	0.012		17	70.83%
16	on the other side of	56	0.012		19	79.17%
18	on the opposite side of	54	0.011		17	70.83%
19	as a matter of course	53	0.011		18	75.00%
19	the other side of the	53	0.011		19	79.17%
21	as much as to say	52	0.011		17	70.83%
21	in the midst of the	52	0.011		18	75.00%
23	up and down the room	49	0.010		14	58.33%
24	as if he were a	47	0.010		17	70.83%
24	with the air of a	47	0.010		15	62.50%
26	as if they had been	44	0.009		21	87.50%
26	on the top of the	44	0.009		18	75.00%
26	with his back to the	44	0.009		15	62.50%
29	the father of the marshalsea	43	0.009		1	4.17%
30	the greater part of the	42	0.009		14	58.33%
30	very much obliged to you	42	0.009		14	58.33%
32	as if he would have	41	0.008		13	54.17%
32	hands in his pockets and	41	0.008		18	75.00%
34	in the centre of the	39	0.008		14	58.33%
34	the other end of the	39	0.008		14	58.33%

TG | SkipChars | SP | Wildcard | File

表 1：單字頻率計算與五元語法索引