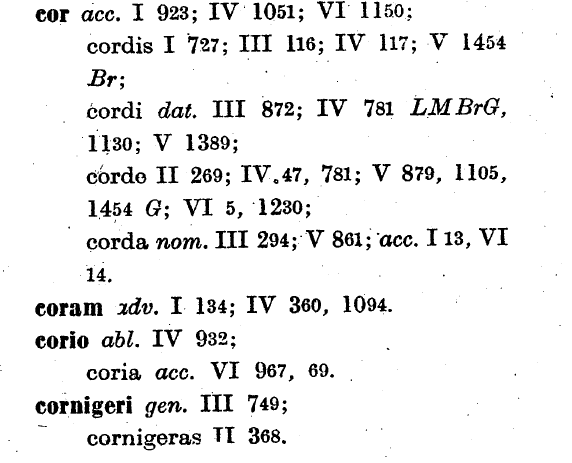
Here is a sketch of the workflow for converting Paulson's Index Lucretianus to a lemmatized text of Lucretius:

Scan of the Book



Text as digitized by NewGen

<B>cor</B> <I>acc.</I> I 923; IV 1051; VI 1150;

cordis I 727; III 116; IV 117; V 1454 <I>Br</I>;

cordi <I>dat.</I> III 872; IV 781 <I>LM BrG,</I> 1130; V 1389;

corde II 269; IV 47, 781; V 879, 1105, 1454 <I>G</I>; VI 5, 1230;

corda <I>nom.</I> III 294; V 861; <I>acc.</I> 113, VI 14.

<B>coram</B> <I>adv.</I> I 134; IV 360, 1094.

<B>corio</B> <I>abl.</I> IV 932;

coria <I>acc.</I> VI 967, 69.

<B>cornigeri</B> <I>gen.</I> III 749;

cornigeras II 368.

<pn,31>

Stage 1: remove all letters not between B tags, but preserve all Roman and Arabic numerals; remove all <I> and </I> tags and the text between them; remove all <pn ....> tags:

<B>cor</B> ~~<I>acc.</I>~~ I 923; IV 1051; VI 1150;

~~cordis~~ I 727; III 116; IV 117; V 1454 ~~<I>Br</I>~~;

~~cordi~~ ~~<I>dat.</I>~~ III 872; IV 781 ~~<I>LM BrG~~,~~</I>~~ 1130; V 1389;

~~corde~~ II 269; IV 47, 781; V 879, 1105, 1454 ~~<I>G</I>~~; VI 5, 1230;

~~corda~~ ~~<I>nom.</I>~~ III 294; V 861; ~~<I>acc.</I>~~ 113, VI 14.

<B>coram</B> ~~<I>adv.</I>~~ I 134; IV 360, 1094.

<B>corio</B> ~~<I>abl.</I>~~ IV 932;

~~coria <I>acc.</I>~~ VI 967, 69.

<B>cornigeri</B> ~~<I>gen.</I>~~ III 749;

~~cornigeras~~ II 368.

~~<pn,31>~~

Stage 2: whenever Roman numerals do not occur after a comma or semi-colon and before Arabic numerals, add the most recently occurring Roman numeral

<B>cor</B> I 923; IV 1051; VI 1150;

I 727; III 116; IV 117; V 1454;

III 872; IV 781; IV 1130; V 1389;

II 269; IV 47, IV 781; V 879, V 1105, V 1454; VI 5, VI 1230;

III 294; V 861; V 113, VI 14.

<B>coram</B> I 134; IV 360, IV 1094.

<B>corio</B> IV 932;

VI 967,VI 69.

<B>cornigeri</B> III 749;

II 368.

Stage 3: convert Roman numeral (= book numbers) to Arabic numerals, and remove <B> and </B> tags

cor 3 923; 4 1051; 4 1150;

1 727; 3 116; 4 117; 5 1454;

3 872; 4 781; 4 1130; 5 1389;

2 269; 4 47, 4 781; 5 879, 5 1105, 5 1454; 6 5, 6 1230;

3 294; 5 861; 5 113, VI 14.

coram 1 134; 4 360, 4 1094.

corio 4 932;

6 967,6 69.

cornigeri 3 749;

2 368.

Stage 4: convert to spreadsheet as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A | B | C | D | E |
| Lemma | Note | Form | Book | Verse |
| COR |  |  | 1 | 923 |
|  |  |  | 4 | 1051 |
|  |  |  | 6 | 1150 |
|  |  |  | ETC | ETC |
| CORAM |  |  | 1 | 134 |
|  |  |  | 4 | 360 |
|  |  |  | 4 | 1094 |
| CORIO |  |  | 4 | 932 |
|  |  |  | 6 | 967 |

Stage 5:

Run lemmatizer on column A to get correct LASLA lemmata.