Multidimensional analysis instructions

1) The tool expects input in the following format (only a small part of the dataset is shown).

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>J</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Filename</td>
<td>Register</td>
<td>PAST</td>
<td>PERF</td>
<td>PRES</td>
<td>PLACE</td>
<td>TIME</td>
<td>1PRON</td>
<td>2PRON</td>
<td>3PRON</td>
</tr>
<tr>
<td>2</td>
<td>AmE06_A01</td>
<td>News_reportage</td>
<td>4.75</td>
<td>0.49</td>
<td>4.45</td>
<td>0.49</td>
<td>0.49</td>
<td>1.43</td>
<td>0.3</td>
<td>3.02</td>
</tr>
<tr>
<td>3</td>
<td>AmE06_A02</td>
<td>News_reportage</td>
<td>3.92</td>
<td>0.83</td>
<td>4.51</td>
<td>0.29</td>
<td>0.74</td>
<td>1.13</td>
<td>0.2</td>
<td>3.19</td>
</tr>
<tr>
<td>4</td>
<td>AmE06_A03</td>
<td>News_reportage</td>
<td>4.64</td>
<td>0.55</td>
<td>3.54</td>
<td>0.2</td>
<td>0.85</td>
<td>0.3</td>
<td>0.05</td>
<td>1.85</td>
</tr>
<tr>
<td>5</td>
<td>AmE06_A04</td>
<td>News_reportage</td>
<td>4.18</td>
<td>0.4</td>
<td>4.85</td>
<td>0.25</td>
<td>0.2</td>
<td>1.04</td>
<td>0.1</td>
<td>0.54</td>
</tr>
<tr>
<td>6</td>
<td>AmE06_A05</td>
<td>News_reportage</td>
<td>3.13</td>
<td>0.65</td>
<td>5.91</td>
<td>0.1</td>
<td>0.4</td>
<td>2.43</td>
<td>0.4</td>
<td>1.54</td>
</tr>
<tr>
<td>7</td>
<td>AmE06_A06</td>
<td>News_reportage</td>
<td>1.97</td>
<td>1.03</td>
<td>6.25</td>
<td>0.25</td>
<td>0.44</td>
<td>0.44</td>
<td>0.0</td>
<td>1.72</td>
</tr>
<tr>
<td>8</td>
<td>AmE06_A07</td>
<td>News_reportage</td>
<td>2.21</td>
<td>0.8</td>
<td>4.47</td>
<td>0.4</td>
<td>0.3</td>
<td>0.55</td>
<td>0.05</td>
<td>1.56</td>
</tr>
<tr>
<td>9</td>
<td>AmE06_A08</td>
<td>News_reportage</td>
<td>2.83</td>
<td>1.07</td>
<td>5.85</td>
<td>0.24</td>
<td>0.78</td>
<td>1.17</td>
<td>0.68</td>
<td>3.22</td>
</tr>
<tr>
<td>10</td>
<td>AmE06_A09</td>
<td>News_reportage</td>
<td>5.84</td>
<td>0.45</td>
<td>3.46</td>
<td>0.25</td>
<td>0.34</td>
<td>1.14</td>
<td>0.1</td>
<td>1.14</td>
</tr>
<tr>
<td>11</td>
<td>AmE06_A10</td>
<td>News_reportage</td>
<td>3.97</td>
<td>0.69</td>
<td>4.31</td>
<td>0.25</td>
<td>0.35</td>
<td>0.59</td>
<td>0.05</td>
<td>2.97</td>
</tr>
<tr>
<td>12</td>
<td>AmE06_A11</td>
<td>News_reportage</td>
<td>2.92</td>
<td>0.68</td>
<td>4.19</td>
<td>0.44</td>
<td>1.17</td>
<td>0.24</td>
<td>0.81</td>
<td>0.68</td>
</tr>
<tr>
<td>13</td>
<td>AmE06_A12</td>
<td>News_reportage</td>
<td>2.89</td>
<td>0.8</td>
<td>4.28</td>
<td>0.44</td>
<td>0.05</td>
<td>0.35</td>
<td>0.2</td>
<td>2.24</td>
</tr>
<tr>
<td>14</td>
<td>AmE06_A13</td>
<td>News_reportage</td>
<td>2.7</td>
<td>0.95</td>
<td>5.8</td>
<td>0.4</td>
<td>0.2</td>
<td>0.8</td>
<td>0.1</td>
<td>2.25</td>
</tr>
<tr>
<td>15</td>
<td>AmE06_A14</td>
<td>News_reportage</td>
<td>3.67</td>
<td>1.04</td>
<td>3.18</td>
<td>0.35</td>
<td>0.45</td>
<td>0.7</td>
<td>0.1</td>
<td>1.69</td>
</tr>
<tr>
<td>16</td>
<td>AmE06_A15</td>
<td>News_reportage</td>
<td>5.54</td>
<td>1.1</td>
<td>3.94</td>
<td>0.3</td>
<td>0.7</td>
<td>0.5</td>
<td>0.45</td>
<td>2.49</td>
</tr>
<tr>
<td>17</td>
<td>AmE06_A16</td>
<td>News_reportage</td>
<td>4.77</td>
<td>0.98</td>
<td>4.97</td>
<td>0.44</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
<td>0.39</td>
</tr>
<tr>
<td>18</td>
<td>AmE06_A17</td>
<td>News_reportage</td>
<td>1.78</td>
<td>0.64</td>
<td>5.88</td>
<td>0.49</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>19</td>
<td>AmE06_A18</td>
<td>News_reportage</td>
<td>3.8</td>
<td>0.49</td>
<td>6.18</td>
<td>0.74</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Relative frequencies of linguistic variables in the individual files.

2) If the comparison is to be made with Biber’s (1988) dimensions (see comparison with Biber’s (1988) dimensions), the linguistic variables and their order need to correspond to those in Biber’s work (see the end of this document).

Biber variables (abbreviated):

1  PAST   15  GER   29  TSUB   43  TTR   57  SUAV
2  PERF   16  NN   30  TOBJ   44  MWL   58  SMAP
3  PRES   17  PASS   31  WHSUB   45  CONJ   59  CONT
4  PLACE   18  BYPASS   32  WHOBJ   46  DWNT   60  THATD
5  TIME   19  BE   33  PIRE   47  HDG   61  STPR
6  1PRON   20  EXIST   34  SERE   48  AMP   62  SPINF
7  2PRON   21  THVC   35  CAUS   49  EMPH   63  SPAU
8  3PRON   22  THAC   36  CONC   50  DPAR   64  PHC
9  IT   23  WHCL   37  COND   51  DEMO   65  ANDC
10  DEMPRON   24  INF   38  OSUB   52  POSS   66  SYNEG
11  INDPRON   25  PRESP   39  PP   53  NECESS   67  ANNEG
12  DO   26  PASTP   40  JJATR   54  PRED
13  WHQU   27  WZPAST   41  JJPRED   55  PUBV
14  NOMZ   28  WZPRES   42  ADV   56  PRIV
3) The output for the tool can be obtained from any corpus tool which is able to provide a breakdown of frequencies for individual files (e.g. CQPweb, SketchEngine, MonoConc Pro). For full comparability with Biber’s (1988) dimensions, the output from Biber tagger would be required.

Currently, a freeware tool MAT is available (https://sites.google.com/site/multidimensionaltagger/) that replicates Biber’s tagging. The tool was created by Andrea Nini. The output of the tool can be directly copied into the Multidimensional tool. However, some minor adjustment is required (see the expected data format above).

DIRECT COMPARISON OF THE DATASET WITH BIBER’S DIMENSIONS

4a) Paste the data into the text box, select ‘Comparison with Biber’s (1988)’ dimension and press ‘Perform MD analysis’.
5a) The expected output is a series of six dimensions with individual registers based on the dataset (left) and Biber’s (1988) dimensions (right) for comparison.

**Dimension 1: Involved vs. Informational**

Example of the output (only first dimension is shown)

**FULL MULTIDIMENSIONAL ANALYSIS**

4b) Paste the data into the text box, select ‘Full MD’ and press ‘Perform MD analysis’.
5b) Based on the Scree plot, decide how many factors you want to extract. Press ‘Select’

The red line indicates a cut-off based on Parallel analysis (PA).

Eigen value = 1 -> possible cut-off point

Eigen values level off -> possible cut-off point

6) Both factor loadings and new dimensions based on the extracted factors are displayed. Factor loadings that are larger than 0.35 or smaller than -0.35 are considered.

<table>
<thead>
<tr>
<th>Features</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAST</td>
<td>0.370</td>
<td>-0.076</td>
<td>0.169</td>
<td>0.871</td>
</tr>
<tr>
<td>PERF</td>
<td>0.137</td>
<td>0.105</td>
<td>0.155</td>
<td>0.411</td>
</tr>
<tr>
<td>PRES</td>
<td>0.303</td>
<td>0.004</td>
<td>-0.091</td>
<td>-0.905</td>
</tr>
<tr>
<td>PLACE</td>
<td>0.235</td>
<td>-0.400</td>
<td>0.015</td>
<td>0.236</td>
</tr>
<tr>
<td>TIME</td>
<td>0.303</td>
<td>-0.321</td>
<td>0.024</td>
<td>0.101</td>
</tr>
<tr>
<td>X1PRON</td>
<td>0.600</td>
<td>-0.113</td>
<td>0.032</td>
<td>-0.061</td>
</tr>
<tr>
<td>X2PRON</td>
<td>0.624</td>
<td>-0.111</td>
<td>-0.053</td>
<td>-0.241</td>
</tr>
<tr>
<td>X3PRON</td>
<td>0.480</td>
<td>-0.171</td>
<td>0.267</td>
<td>0.471</td>
</tr>
<tr>
<td>IT</td>
<td>0.577</td>
<td>-0.188</td>
<td>0.285</td>
<td>-0.139</td>
</tr>
<tr>
<td>DEPRN</td>
<td>0.167</td>
<td>0.182</td>
<td>0.375</td>
<td>-0.121</td>
</tr>
<tr>
<td>INDPRN</td>
<td>0.368</td>
<td>-0.085</td>
<td>0.269</td>
<td>0.180</td>
</tr>
<tr>
<td>DO</td>
<td>0.536</td>
<td>-0.119</td>
<td>0.165</td>
<td>-0.046</td>
</tr>
<tr>
<td>WHQU</td>
<td>0.458</td>
<td>0.003</td>
<td>0.188</td>
<td>0.096</td>
</tr>
<tr>
<td>NOMZ</td>
<td>-0.653</td>
<td>0.538</td>
<td>0.058</td>
<td>-0.058</td>
</tr>
<tr>
<td>GER</td>
<td>-0.238</td>
<td>0.000</td>
<td>-0.092</td>
<td>-0.135</td>
</tr>
<tr>
<td>NN</td>
<td>-0.467</td>
<td>-0.153</td>
<td>-0.665</td>
<td>-0.178</td>
</tr>
</tbody>
</table>

Dimension 1:
- Fiction_general
- Romance
- Action_mystery
- Humour
- Fiction_science
- Fiction_adventure

- News_editorial
- News_reportage
- Biography
- News_reviews
- Skill
- Popular_non
- Religion
- Misc_non
- Academic
The full set of Biber's (1988) features:

A. Tense and aspect markers
   1. past tense
   2. perfect aspect
   3. present tense

B. Place and time adverbials
   4. place adverbials (e.g., behind, downstairs, locally)
   5. time adverbials (e.g., eventually, immediately, nowadays)

C. Pronouns and pro-verbs
   6. first-person pronouns
   7. second-person pronouns
   8. third-person personal pronouns (excluding it)
   9. pronoun it
   10. demonstrative pronouns (that, this, these, those as pronouns)
   11. indefinite pronouns (e.g., anyone, everybody, nothing)
   12. pro-verb do

D. Questions
   13. direct WH questions

E. Nominal forms
   14. nominalizations (ending in -tion, -ment, -ness, -ity)
   15. gerunds (participial forms functioning as nouns)
   16. total other nouns

F. Passives
   17. agentless passives
   18. by-passives

G. Stative forms
   19. be as main verb
   20. existential there

H. Subordination features
   21. that verb complements (e.g., We felt that we needed a financial base.)
   22. that adjective complements (e.g., It's quite obvious that certain things can be
      separated.)
   23. WH clauses (e.g., I wondered what to do.)
   24. infinitives
   25. present participial adverbial clauses (e.g., Screaming with rage, he ran up the stairs.)
   26. past participial adverbial clauses (e.g., Given these characteristics, it is not
      surprising that . . .)
   27. past participial postnominal (reduced relative) clauses (e.g., the exhaust air
      volume required by the 6-ft. × 4-ft. grid)
   28. present participial postnominal (reduced relative) clauses (e.g., the currents of
      dissent swirling beneath the surface)
   29. that relative clauses on subject position (e.g., the papers that are on the table)
   30. that relative clauses on object position (e.g., the papers that she thought would be
      interesting)

33. pied-piping relative clauses (e.g., the way in which food is distributed)
34. sentence relatives (e.g., We waited for six hours, which was ridiculous)
35. causative adverbial subordinators (because)
36. concessive adverbial subordinators (although, though)
37. conditional adverbial subordinators (if, unless)
38. other adverbial subordinators (e.g., inasmuch as, such that, although)

I. Prepositional phrases, adjectives, and adverbs
   39. total prepositional phrases
40. attributive adjectives (e.g., the small room)
41. predicative adjectives (e.g., the room is small)
42. total adverbs

J. Lexical specificity
43. type/token ratio
44. mean word length

K. Lexical classes
45. conjunctions (e.g., alternatively, nevertheless, therefore)
46. downtoners (e.g., mildly, partially, somewhat)
47. hedges (e.g., almost, maybe, sort of [except as true noun])
48. amplifiers (e.g., completely, totally, utterly)
49. emphatics (e.g., a lot, for sure, really)
50. discourse particles (e.g., sentence initial expletive, now, well)
51. demonstratives

L. Modals
52. possibility modals (can, could, may, might)
53. necessity modals (must, ought, should)
54. predictive modals (shall, will, would)

M. Specialized verb classes
55. public verbs (e.g., complain, explain, promise)
56. private verbs (e.g., believe, think, know)
57. suasive verbs (e.g., command, propose, recommend)
58. seem and appear

N. Reduced forms and dispreferred structures
59. contractions
60. complementizers that deletion (e.g., I think [if he's gone already])
61. stranded prepositions (e.g., the person that I was talking to)
62. split infinitives (e.g., I want to completely convince you that . . .)
63. split auxiliaries (e.g., they have apparently sold it all . . .)

O. Coordination
64. phrasal coordination (NOUN and NOUN; ADJ and ADV; VERB and VERB; ADV
   and ADV)

65. independent clause coordination (clause initial and)

P. Negation
