

Appendix 2 Stopword List and Discussion

之	是	于	元	后	哉	還	甚	求	氏	焉
不	與	在	正	作	難	絕	本	說	外	我
也	夫	非	多	因	稱	往	止	左	同	復
以	可	六	西	雖	屬	己	興	起	受	千
而	五	諸	足	始	宜	邪	耳	會	反	亦
其	將	必	又	里	聽	固	廣	定	少	九
為	使	然	高	請	終	首	益	通	常	七
曰	何	若	內	女	遠	由	應	對	過	方
者	至	及	當	右	盡	共	十	所	此	乃
子	四	未	去	敢	異	徒	則	故	太	百
有	矣	萬	北	前	進	任	無	三	謂	皆
於	自	吾	來	易	初	更	一	二	如	乎

In retrospect, the stop word list used for the current study, presented above, may have been overly aggressive in certain respects (removing, for instance, words such as “woman” (*nu* 女) or “listen/obey” (*ting* 聽) that could have been useful to include) and failing to eliminate common words such as *ren* 人 (person, human) that may have distorted some of the topic modeling or other analyses presented below. The creation of properly calibrated stop-word lists is obviously an area where careful thought and expertise are required, and what constitutes a well-chosen list very much depends on the purpose of the analysis. For instance, although pronouns and grammatical particles are typically included in stop lists, a scholar interesting in problems of textual dating may very well want to include them. A planned follow-up study will explore the variable effects of different types of stop-word lists in more detail.