從main.py的檔案開始trace code，接著要看的是train.py中trainIters的部份

再來是到 loadPrepareData處理Input，原本是一行Q，一行A，在建立Q&A的對話資料
ptt資料集則是把Q A放在同一行用tab（'/t'）隔開，這邊取pair的方式要去修改。

還有要修改的地方是中文和英文在轉word vector的差別，英文一個字一個字間會是用空格隔開，所以會看到它用_.split(' ')去拆字串，但中文每個字元就是一個字，所以直接去統計和轉換成vector就行。除了在load的部份要改外，train.py裡有去分每個batch資料的地方也要改。

最後雖然跟這次作業無關，但如果要自己寫個txt檔看他test出來的結果時有些地方要注意下，可以在上面載到pretrained好的model，而他的code會從這個檔案的路徑名字去建網路架構再load參數進來。RNN(這邊是用GRU)的input size是92391，是movie_subtitles資料集的word vector維度(也就是總共有幾個不同的字)，但在把input data轉換成word vector時是用save/training_data/<dataset name>/ 下的voc.tar檔，所以如果用你新建的txt來跑他會對這個txt建一個新的voc.tar，然後轉換的word vector維度就會不一樣，就會發生與網路架構size mismatch的錯誤，可以去看issues裡Pretrained model: size mismatch那篇。解決辦法是先跑movie_subtitles dataset取得他的voc.tar，再跑自己建的txt創出資料夾與pair.tar，然後把voc.tar換掉即可，唯一就是要盡量確保新的txt沒有出現前面movie_subtitles沒出現過的字。