Split	train	valid	test
#Words	20,298	1,127	1,129
#Entries	146,486	8,087	8,352
#Tokens	1,406,440	77,948	79,699
Avg length	6.60	6.64	6.54