

vocab size factor: 3				ds1: 4 4 4		ds2: 4 4 4 4 4		ds3: 16 16 16				
batch size	learning rate	hidden size	temp update	train	val	train	val	train	val			
32	0,0005	128	0,97	0,640	0,610	0,460	0,440	0,440	0,410			
			0,99	0,780	0,650	0,950	0,930	0,710	0,680			
		256	0,97	0,630	0,590	0,960	0,970	0,920	0,900			
			0,99	0,970	0,920	0,990	0,980	0,990	0,960			
	0,0010	128	0,97	0,740	0,690	0,500	0,460	0,650	0,590			
			0,99	0,960	0,890	0,990	0,990	0,720	0,670			
		256	0,97	0,120	0,140	0,090	0,080	0,860	0,820			
			0,99	0,990	0,990	0,980	0,970	0,960	0,940			
64	0,0005	128	0,97	0,770	0,720	0,380	0,330	0,380	0,360			
			0,99	0,870	0,840	0,990	0,980	0,480	0,410			
		256	0,97	0,710	0,660	0,840	0,830	0,600	0,570			
			0,99	0,960	0,910	1,000	0,990	0,830	0,780			
	0,0010	128	0,97	0,610	0,600	0,680	0,670	0,470	0,440			
			0,99	0,900	0,820	0,990	0,980	0,480	0,440			
		256	0,97	0,720	0,680	0,930	0,920	0,630	0,580			
			0,99	0,900	0,840	0,630	0,610	0,960	0,940			
128	0,0005	128	0,97	0,620	0,580	0,460	0,450	0,340	0,320			
			0,99	0,560	0,520	0,650	0,590	0,510	0,470			
		256	0,97	0,890	0,880	0,590	0,580	0,380	0,360			
			0,99	0,650	0,580	0,990	0,980	0,570	0,480			
	0,0010	128	0,97	0,410	0,370	0,530	0,520	0,500	0,480			
			0,99	0,860	0,800	0,990	0,980	0,470	0,390			
		256	0,97	0,910	0,870	0,810	0,840	0,720	0,680			
			0,99	0,270	0,230	1,000	0,990	0,950	0,910			