Dataset	# Training files	# Testing PDF files	# Training Lines	# Testing Lines
ICDAR 2013	50	17	804	224
ACL Anthology	9,280	186	357,892	346