**MIMIC CXR Report Analysis**

This is the analysis of fist 2000 text files of reports present in the MIMIC CXR Data:

1. **40 most commom words:**

[('pleural', 896), ('effusion', 863), ('right', 850), ('pneumothorax', 845), ('normal', 713), ('left', 587), ('seen', 540), ('lungs', 490), ('lung', 476), ('consolidation', 454), ('chest', 449), ('pulmonary', 445), ('clear', 440), ('mediastinal', 418), ('focal', 410), ('unchanged', 390), ('silhouette', 375), ('contours', 375), ('heart', 364), ('cardiomediastinal', 360), ('size', 347), ('hilar', 326), ('lateral', 310), ('atelectasis', 295), ('cardiac', 259), ('edema', 256), ('prior', 253), ('mild', 249), ('noted', 231), ('unremarkable', 228), ('lower', 223), ('opacity', 217), ('tube', 213), ('stable', 208), ('osseous', 201), ('\_\_\_', 200), ('upper', 196), ('views', 191), ('within', 190), ('without', 175)]

1. **40 most common 2-gram phrases:**

[(('pleural', 'effusion'), 677), (('effusion', 'pneumothorax'), 539), (('focal', 'consolidation'), 324), (('cardiomediastinal', 'silhouette'), 261), (('hilar', 'contours'), 244), (('lungs', 'clear'), 233), (('heart', 'size'), 214), (('pulmonary', 'edema'), 195), (('lateral', 'views'), 181), (('views', 'chest'), 176), (('mediastinal', 'hilar'), 160), (('lung', 'volumes'), 153), (('consolidation', 'pleural'), 150), (('contours', 'normal'), 132), (('pneumothorax', 'seen'), 128), (('within', 'normal'), 122), (('acute', 'osseous'), 122), (('normal', 'limits'), 121), (('size', 'normal'), 119), (('pleural', 'effusions'), 118), (('cardiac', 'mediastinal'), 117), (('pa', 'lateral'), 113), (('clear', 'without'), 110), (('cardiac', 'silhouette'), 107), (('without', 'focal'), 100), (('lower', 'lobe'), 99), (('osseous', 'abnormalities'), 95), (('frontal', 'lateral'), 95), (('right', 'hemidiaphragm'), 93), (('pulmonary', 'vasculature'), 92), (('pneumothorax', 'cardiomediastinal'), 90), (('consolidation', 'effusion'), 89), (('mediastinal', 'contours'), 89), (('silhouette', 'normal'), 86), (('mediastinal', 'silhouettes'), 81), (('right', 'lung'), 81), (('clear', 'pleural'), 77), (('free', 'air'), 73), (('lung', 'base'), 73), (('vascular', 'congestion'), 73)]

1. **20 most common 3-gram phrases:**

[(('pleural', 'effusion', 'pneumothorax'), 432), (('lateral', 'views', 'chest'), 175), (('mediastinal', 'hilar', 'contours'), 150), (('consolidation', 'pleural', 'effusion'), 149), (('focal', 'consolidation', 'pleural'), 138), (('within', 'normal', 'limits'), 120), (('heart', 'size', 'normal'), 106), (('acute', 'osseous', 'abnormalities'), 95), (('effusion', 'pneumothorax', 'seen'), 94), (('without', 'focal', 'consolidation'), 92), (('pa', 'lateral', 'views'), 89), (('effusion', 'pneumothorax', 'cardiomediastinal'), 84), (('lungs', 'clear', 'without'), 80), (('clear', 'without', 'focal'), 77), (('pneumothorax', 'cardiomediastinal', 'silhouette'), 73), (('hilar', 'contours', 'normal'), 72), (('cardiomediastinal', 'silhouette', 'normal'), 70), (('focal', 'consolidation', 'effusion'), 69), (('cardiac', 'mediastinal', 'silhouettes'), 68), (('clear', 'pleural', 'effusion'), 66)]