

<b>Data Set</b>	<b>Sentence Pairs</b>	<b>Arabic Tokens</b>	<b>English Tokens</b>
<b>MSA-150MW</b>	8.0M	151.4M	204.4M
<b>Dialect-1500KW</b>	180k	1,545,053	2,257,041
<b>MSA-1300KW</b>	71k	1,292,384	1,752,724
<b>MSA-Web-Tune</b>	6,163	145,260	184,185
<b>MSA-Web-Test</b>	5,454	136,396	172,357
<b>Lev-Web-Tune</b>	2,600	20,940	27,399
<b>Lev-Web-Test</b>	2,600	21,092	27,793
<b>Egy-Web-Test</b>	2,600	23,671	33,565
<b>E-Facebook-Tune</b>	3,351	25,130	34,753
<b>E-Facebook-Test</b>	3,188	25,011	34,244