

Training set	Parallel sentences	Words per lang
In-domain SMS data	17,192	35k
Medical domain	1,619	10k
News wire domain	13,517	30k
Glossary	35,728	85k
Wikipedia parallel sentence	8,476	90k
Wikipedia named entities	10,499	25k
The bible	30,715	850k
Haitisurf dictionary	3,763	4k
Krengle dictionary	1,687	3k
Krengle sentences	658	3k