

Serbian NGrams (SrpNGrams) represent set of N-grams extracted from Serbian Lemmatized and PoS Annotated Corpus (SrpLemKor) for N from 1 to 5. Each unigram is maximum continuous chunk of non-whitespace lower-case characters. The resource contains all unique N-grams preceded by number of occurrences (Table 1). It also contains n-gram language models (1-5) in the standard ARPA text and binary format, created by IRST Language Modeling Toolkit.

SrpLemKor texts consist of fiction written by Serbian authors in 20th and 21th century, various scientific texts from various domains (both humanities and sciences), legislative texts and general texts.

General texts represent daily news published in newspaper "Politika" 2000-2002 and 2005-2010, texts in journals and magazines 1991-2002 ("Danica", "Ebit", "Ekonomist", "Glasnik", "NIN", "Ilustrovana politika", "Kalibar", "Moje srce", "Mostovi", "Pravoslavlje", "Svet", "Teološki pogledi", "Trn", "Viva", "Republika"), internet portal texts 2011-2012 (Peščanik), TANJUG agency news 1995-96, newspaper feuilletons published in newspapers "Politika" (2001-2003), "Večernje novosti" (2008-2011) and "Danas" (2002-2006).

<b>n-gram type</b>	<b>quantity</b>
unigrams	402395
bigrams	2148570
trigrams	3309528
4-grams	3664760
5-grams	3740032

**Table 1**