



## SRILM - The SRI Language Modeling Toolkit

SRILM is a toolkit for building and applying statistical language models (LMs), primarily for use in speech recognition, statistical tagging and segmentation, and machine translation. It has been under development in the [SRI Speech Technology and Research Laboratory](#) since 1995. The toolkit has also greatly benefitted from its use and enhancements during the [Johns Hopkins University/CLSP summer workshops](#) in 1995, 1996, 1997, and 2002 (see [history](#)).

These pages and the software itself assume that you know what statistical language modeling is. To learn about language modeling we recommend the textbooks

- [Speech and Language Processing](#) by Dan Jurafsky and Jim Martin (chapter 6 in the 1st edition, chapter 4 in the 2nd edition)
- [Foundations of Statistical Natural Language Processing](#), by Chris Manning and Hinrich Schütze (chapter 6).

Either book gives an excellent introduction to N-gram language modeling, which is the main type of LM supported by SRILM.

SRILM consists of the following components:

- A set of C++ class libraries implementing language models, supporting data structures and miscellaneous utility functions.
- A set of executable programs built on top of these libraries to perform standard tasks such as training LMs and testing them on data, tagging or segmenting text, etc.
- A collection of miscellaneous scripts facilitating minor related tasks.


SRILM runs on UNIX and Windows platforms.

SRILM has been used in a great variety of statistical modeling [applications](#).

Others have published [extensions](#) to SRILM that add new functionality.

### Documentation

SRILM is still under development. The documentation in particular is work in progress. Best documented are the executable programs, scripts, and file formats, in the form of UNIX-style [manual pages](#). The libraries are documented mostly in the source code. An overview of what the software can do and its design philosophy can be found in the paper "SRILM - An Extensible Language Modeling Toolkit", in Proc. Intl. Conf. Spoken Language Processing, Denver, Colorado, September 2002 ([postscript](#), [PDF](#)). Links to other papers and tutorials, as well as frequently asked questions, are also given [here](#).

 NEW! A [recent paper](#) summarizes updates to SRILM since the 2002 paper.

### Terms of Use

Government agencies, and schools, universities, and non-profit organizations can [download](#) SRILM free of charge under SRI's "[Research Community License](#)", for use in projects that do not receive external funding other than government research grants and contracts. For other uses please inquire about [commercial licensing](#).

### Mailing List

Exchange of information among SRILM users, as well as some level of technical support, is provided through the mailing list [srilm-user@speech.sri.com](mailto:srilm-user@speech.sri.com). Check the [user mailing list archive](#) or [announcement mailing list archive](#) for past contributions.

To subscribe and obtain more information, follow [this link](#).

[About Us](#) † [R&D Divisions](#) † [Careers](#) † [Newsroom](#) † [Contact Us](#)

©2011 SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025-3493

SRI International is an independent, nonprofit corporation. [Privacy policy](#)

Last modified Aug 23, 2022