To make dictionary search in Pashto, Urdu, and Bengali easier by:
- Building a tool—a morphological parser—to help users find forms of words that don’t appear in the dictionary
- Finding and repairing structural anomalies in electronic dictionaries
- Building a spell correction tool

**Objective**

**Definition**

A **morphological parser** is a computer program that analyzes a word into its meaningful parts and relates this form of the word to the parts that make up its meaning. For example, a morphological parser for English would analyze the word *cats* into *cat* plus the plural suffix –*s*.

**Findings**

Words can be hard to find in the dictionary for several reasons, including:
- Complex word formation rules that make it hard to know what form to look up
- Hard-to-hear sounds, variable spellings or pronunciations, and differing transcription systems
- Anomalies in the electronic dictionary

Some languages are more challenging than others because of dialectal variation and few linguistic resources.

We have written two kinds of grammars for Pashto, Urdu, and Bengali:
- A grammar of word formation that a linguist can use
- A formal grammar, which can be converted into a morphological parser

**Relevance**

**Language tools can help analysts . . .**

Language analysts must often work with languages in which they are not experts, and they frequently rely on dictionaries for interpretation. Tools can help analysts—particularly analysts with limited knowledge of the language—look up words in languages where word formation, hard-to-hear sounds, or spelling variations cause problems.

. . . understand strategic languages.

Since Pashto, Urdu, and Bengali are strategic languages, we need to provide tools that can enhance language analysis by government professionals with less-than-expert knowledge of these languages. Such tools must be available quickly and economically.

**Reports & Products**

- **Validation of accuracy of Bengali and Urdu parsers** (February 2009)
- **Methodology and costs for creation of dual-use grammars** (January 2009)
- Revised Urdu dictionary and Urdu spell corrector
- Morphological parsers and documentation for Urdu and Bengali
- Morphological grammars for Pashto, Urdu, and Bengali

**For More Information**

**U.S. Government Point of Contact**
David Cox, Government Technical Director at CASL
University of Maryland Center for Advanced Study of Language
(301) 226-8970 | dcox@casl.umd.edu | www.casl.umd.edu

**CASL Principal Investigator**
Anne David, PhD, Assistant Research Scientist
University of Maryland Center for Advanced Study of Language
(301) 226-8969 | adavid@casl.umd.edu | www.casl.umd.edu