

# Innovation

**SBIR Topic Number:**  
AF071-041

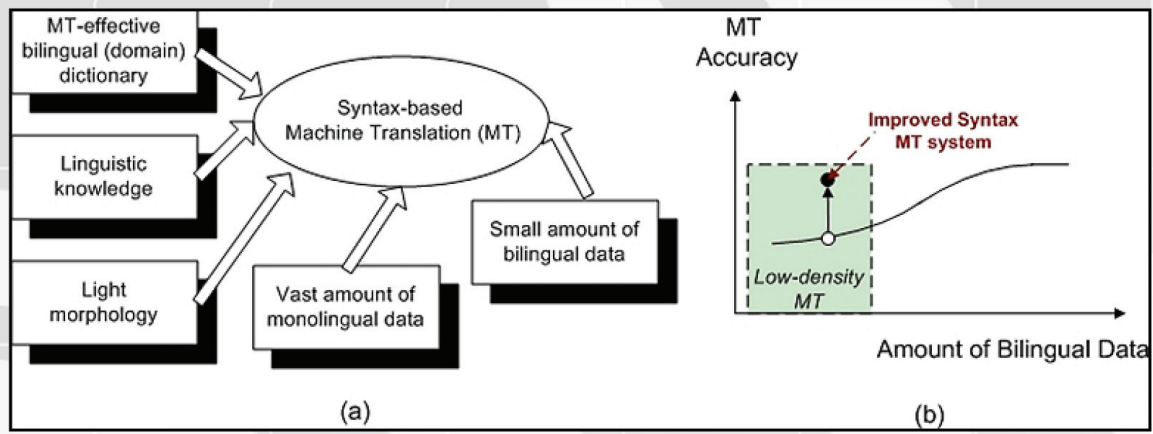
**SBIR Title:**  
Rapid Development Tools and Techniques for Spoken Language Translation Systems

**Contract Number:**  
FA8650-08-C-6863

**SBIR Company Name:**  
SDL Language Weaver (formerly Language Weaver, Inc.), Los Angeles, CA

**Technical Project Office:**  
AFRL Human Effectiveness Directorate, Wright-Patterson AFB, OH

This Air Force SBIR/STTR Innovation Story is an example of Air Force supported SBIR/STTR technology that met topic requirements and has outstanding potential for Air Force and DoD.



**Further improvements to SDL Language Weaver’s Syntax-based Machine Translation framework with the components listed in (a) resulting in additional Machine Translation Accuracy as seen in (b).**

## New Approach to Innovative Techniques in Language Technologies

- Much of the information needed to effectively operate in the Global War on Terror, humanitarian relief, and coalition operations is found in foreign speech and text; however, there is a critical lack of linguists to translate these sources
- The technologies developed by SDL Language Weaver consist of the use of the SDL Language Weaver syntax-based machine translation (MT) approach; the construction of better, MT-effective dictionaries; the exploitation of smart, effective morphology analysis; and the integration of large-scale language models
- SDL Language Weaver produced data-efficient learning techniques that significantly reduce development costs for the construction of high-accuracy machine translation systems, particularly for translating from/into morphologically-rich, data-poor languages and domains
- The MT technology developed by SDL Language Weaver can be used with a variety of human language technologies wherever a translation component is needed

88ABW-2011-2596

**A**

DISTRIBUTION A:  
Approved for public release; distribution unlimited.

## Air Force Requirement

Department of Defense (DoD) personnel are operating all over the world with the Global War on Terror, humanitarian relief, and coalition operations. Much of the information needed to effectively operate in these situations is found in foreign speech and text; however, there is a critical lack of linguists to translate these sources. To address this problem, the DoD has funded the development of translators for both speech and text.

Currently, the development process has been slow and costly as the standard method to improve performance has been to collect, transcribe, and translate ever larger amounts of training data. While this methodology involves low technical risk, it will not be able to address rapid turn-around requirements in new languages and domains of interest to the DoD. Innovative techniques are sought that can perform at least on a par with standard techniques but that do not require substantial amounts of new training data when being adapted for use in a new language and/or domain.

## SBIR Technology

Under this SBIR project, SDL Language Weaver produced data-efficient learning techniques that significantly reduce development costs for the construction of high-accuracy machine translation systems, particularly for translating from/into morphologically-rich, data-poor languages and domains. In simple terms, morphologically-rich languages are those languages that use a large number of prefixes and/or suffixes that attach to word stems to affect meaning. Many such languages are of importance to the DoD.

The technologies developed by SDL Language Weaver consist of the use of the SDL Language Weaver syntax-based machine translation (MT) approach; the construction of better, MT-effective dictionaries; the exploitation of smart, effective morphology analysis; and the integration of large-scale language models. These technologies led to accurate syntax-based machine translation products.

Rather than using the phrase-based machine translation approach that relies heavily on lexicalized translational phrase pairs as the translation knowledge, this project uses the SDL Language Weaver syntax-based approach that translates with syntax translation rules. A translation rule can either

capture lexical translation information or encode abstract (or generalized) syntactic translation knowledge. This makes the syntax approach a better fit to this project because it conveniently allows it to:

- plug in domain-specific dictionaries
- integrate human linguistic knowledge
- generalize from the smaller amount of training examples

## Potential Application

This technology exploits higher-level knowledge sources to not only create an MT system that surpasses the performance of traditional phrase-based systems, but does so on morphologically-rich languages. By leveraging additional data sources, and developing training techniques specific to handling the challenges of morphology-rich languages, SDL Language Weaver can apply this knowledge to additional languages and domains. The MT technology developed by SDL Language Weaver can be used with a variety of human language technologies wherever a translation component is needed.

Efficient use of limited data resources provides for both rapid and cost efficient model development. SDL Language Weaver MT products are currently used by multiple DoD customers. The potential for SDL Language Weaver to either increase the performance of existing languages and/or domains, or to add new language capabilities will be a great asset to the existing (or future) customer base.

## Company Impact

"This Air Force Research Laboratory (AFRL)-funded SBIR program allowed SDL Language Weaver to explore and acquire more expertise in data-efficient development of machine translation systems to significantly reduce MT development costs," states Dr. Daniel Marcu, Chief Technology Officer of SDL Language Technologies. "The exciting technological capabilities developed under this program have led to fast, memory efficient and accurate MT products that are released to the market, delivered to customers and are driving more customer requests. These technologies also set the groundwork for fast developments of new MT systems that can provide better quality and a satisfactory experience to our customers."



# SBIR/STTR

Air Force SBIR Program  
AFRL/XP  
1864 4th Street  
Wright-Patterson AFB OH 45433

AF SBIR/STTR Program Manager: Augustine Vu  
Website: [www.afsbirsttr.com](http://www.afsbirsttr.com)  
Comm: (800) 222-0336  
Fax: (937) 255-2219  
e-mail: [afrl.xppn.dl.sbir.hq@wpafb.af.mil](mailto:afrl.xppn.dl.sbir.hq@wpafb.af.mil)

