

The Web offers to breath new life into the fantasy of artificial-intelligence fans —machine translation.

by Rose Lockwood



Rose Lockwood

Machine Translation (MT) has become a complex industry. Gone are the days when the only available options were either \$99 PC packages or \$50,000 plus main-frame systems. As most Web users are aware,

the Internet has proved a ripe market for fully automatic machine translation, both as an online tool for quick “gisting” translations of short documents, and as a tool for surfing Web sites in unfamiliar languages.

There are now dozens of sites advertising “free translation” or various options for online translation services. The MT engines behind such services are often the same, highlighting the fact that there is now a clear value chain in the MT market (see illustration). The number of developers of the basic technology (the MT engine) tends to grow slowly—this is not an easy startup business to enter. But the businesses that add value to MT, by providing it conveniently within portals, or as part of other applications (messaging, browsing, etc.), are growing fast.

Developers of MT for online delivery can offer MT services directly to consumers, or as software to corporate customers for implementation on intranets. Increasingly, however, they are supplying various kinds of resellers such as general or specialized portals, integrators of language tools and services, and e-business solutions developers. These companies, in turn, develop value-added versions of the software which they sell to their own customers—or, in the case of portals, provide as a service enhancement, free to their users.

MT Engine Developers

Like other areas of the language industries, the MT market has seen significant consolidation, though in this case principally through the activities of one vendor (Lernout & Hauspie) which has acquired a number of different MT products over the last few years. Of the older established MT developers working with European languages, the main players are now IBM, L&H, Logos, Transparent Language, and Systran. (IBM’s MT development is almost exclusively for internal use, and for some OEM resellers such as Alis.) There are, of course, many other companies working on MT, for European and Asian languages as well. A few of these are actively targeting

the online market for European languages—notably Softissimo in France, Prompt in Russia, and CyterTrans in South Africa (combining European and African languages). (See the table at the end of this article for a summary of developers active in online MT.)

Beyond these developers of what we might call “mainstream MT,” a number of

MT is now unquestionably mainstream, and with the massive growth of multiple-language content on the Web, the future for MT is promising—about time, as some of this technology has been in development for 30 years!

vendors are also actively supplying word-translation engines on the Web. These include Language Dynamics (WinDi), Language Force (providing free single-word translations in many languages on the gotoworld.com site), and Translation Experts (who provide dictionary translations between some 25 different languages via InterTran). While word translators are not an adequate substitute for gisting MT, they potentially provide a useful vehicle for up-to-date online dictionaries in many languages, and are especially important for their coverage of languages, and language pairs, for which MT is not available.

Portal Aggregators

Portal aggregators are the general search-engine-based sites that currently provide the center of gravity for most Web users. Aside from an early experiment by CompuServe (offering MT within its online Forums more than five years ago), the AltaVista implementation of Systran is probably the earliest and best known portal offering MT. The same service is available on the Infoseek/Go portal as well.

The French portal Voila (owned by France Telecom) provides two different MT engines on its sites. Reverso (developed by Softissimo) is provided on www.voila.fr

(for French-speaking users), while the English-language site (www.voila.com) provides the Systran engine. A number of Spanish-language portals are now MT-enabled, several using e-Lingo's combination of engines (outlined below).

The main advantage of portal-based MT is that it provides browsing-translation within the browsing environment—making it the simplest and most convenient application of MT for this most basic Web function.

Language-Technology Integrators

Language-technology integrators combine language-technology products, and package them for resale to customers for specific applications. One clear trend appears to be for integrators to provide access to more than one MT engine—to offer both a wider range of language options, and possibly (though this has not been demonstrated) to select engines better suited to a particular type of text, domain, etc. Given the difficulty of developing a large number of language pairs by a single vendor, this is probably a strategy for the long term.

e-Lingo, formerly known as WorldBlaze, is a privately held company based in San Mateo, California, which offers MT ser-

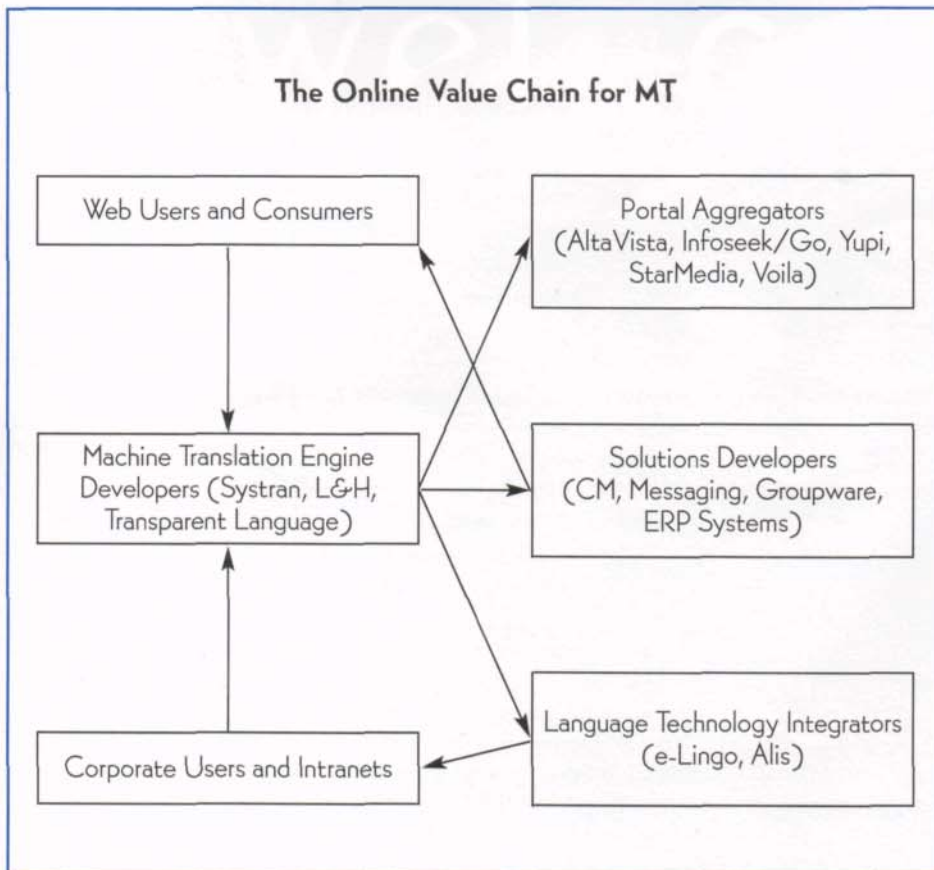
In preparing this article, we ran a "translation test" on the main engines for English, French, German, and Spanish, with general business, news, and information text. The results were, without exception, useful. Aside from the occasional unrecognized word, the meaning was generally clear.

vice integration currently based on MT products from L&H and Transparent Language. The company claims to be the "leading provider of outsourced Internet Translation Solutions," and positions its service as a "global platform for human communication, commerce, and information exchange." Languages currently supported include English, French, German, Italian, Portuguese, and Spanish, with Arabic, Chinese, Japanese, and Korean in development (presumably by L&H).

e-Lingo's Internet MT solutions are currently used by a number of Spanish-language portals, including yupi.com (US, Mexico, Argentina, Spain), starmedia.com (global Spanish/Portuguese portal, highly geographically targeted), periscopio.com (New York based Hispanic), to2.com (global Spanish-language portal), guay.com (Iberian Spanish portal), and amazescape.com (a US-based "multilingual" portal which offers content in English, French, Italian, Mandarin, Portuguese, and Spanish, plus MT). Solutions developers that license e-Lingo's integrated platform include messaging outsource suppliers usa.net and commtouch.com, the translation portal uniscape.com, and Q&A content developer Broad Daylight.

Alis Technologies, based in Montreal, has been in the language-technology market longer than most, having cut its teeth solving UI problems for the Arabic copier market. Alis now concentrates on language-technology integration, especially

The Online Value Chain for MT



through the “Gist-in-Time” platform which integrates technologies supplied by IBM, InXight, Neocortech, Systran, and Transparent Language, as well as tools developed at RALI at the University of Montreal. In addition to MT, Gist-in-Time uses InXight’s “pre-editing” tool (developed by Xerox), and a language identifier from RALI.

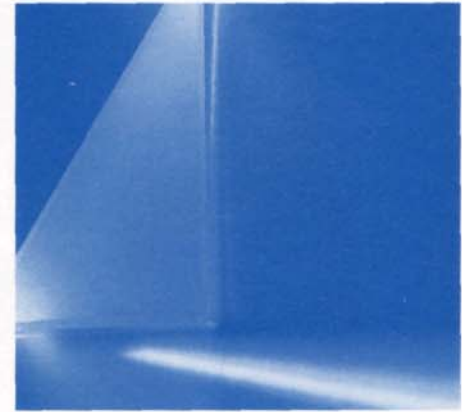
Getting the Gist

MT is now unquestionably mainstream, and with the massive growth of multiple-language content on the Web, the future for MT is promising—about time, as some of this technology has been in development for 30 years! In preparing this article, we ran a “translation test” on the main engines for English, French, German, and Spanish, with general business, news, and

information text. The results were, without exception, useful. Aside from the occasional unrecognized word (some engines couldn’t handle *balbutiante* in French, *Neuwagenkäufer* in German, or *asociación estratégica* in Spanish), the meaning was generally clear. In one or two cases, the translation was even fairly well formed.

But there were differences. Where a key word or phrase was unrecognized, this reduced the value of the translation. Different engines performed better on different texts, for no apparent reason other than vocabulary. This highlights the fact that tuning MT to the text to be translated is vital for high-quality results. This and the need for a massive expansion in the number of available languages are the biggest challenges for the next generation of fully automatic translation.

Rose Lockwood is director of e-business research for Berlitz GlobalNET, where she provides market and technology analysis, and consulting to GlobalNET customers worldwide. Contact her at rose.lockwood@berlitz.ie



Online MT Developers*

CyberTrans

Web site: <http://lexica.epiuse.co.za>
 Languages: English to Afrikaans, French, German, Zulu
 Afrikaans, French, German, Portuguese, Swahili, Tswana to English
 Comments: Up to 20 words for a free sample. This is basically a translation portal (MT cost: 1 cent/word)
 Text: Yes
 URLs: No

Lernout & Hauspie (iTranslator)

Web site: www.lhs.com/itranslator/enterprise.asp
 Languages: English to/from French, Italian, German, Spanish, Portuguese
 French to/from German, Italian
 German to/from Italian
 Italian to/from Spanish
 (and many others in development)
 Comments: Currently available only via channels, such as e-Lingo (www.elingo.com/text_version/texthome.html) and the Microsoft Office Update site (<http://officeupdate.lhsl.com>); L&H also sells personal, professional and enterprise MT software products, and human translation services

Prompt (PITS)

Web site: www.translate.ru
 Languages: Available from the English language site:
 English to/from Russian, French
 German to/from French
 German to English
 English to Spanish
 Available only from the Russian language site:
 Russian to/from German
 Russian to/from French
 Italian to Russian
 Comments: Free service; also sells professional MT software products
 Text: Yes

Softissimo (Reverso)

Web site: www.softissimo.com/index.htm
 Languages: English to/from French
 French to/from German
 German to English
 English to Spanish
 Comments: Free service; also sells personal, professional, and enterprise MT software products
 Text: Yes
 URLs: Yes

Systran

Web site: www.systransoft.com
 Languages: English to/from French, German, Italian, Portuguese, and Spanish
 French to/from German
 Comments: Free service, also available on portals (AltaVista, Infoseek/Go) and via Alis (Gist-in-Time); also sells personal, professional, and enterprise MT software products
 Text: Yes
 URLs: Yes

Transparent Language (TranscendRT)

Web site: www.freetranslation.com
 Languages: English to/from French, German, Spanish
 English to Italian, Portuguese, Norwegian
 Comments: Free service (the engine is also used by e-Lingo); also sells personal, professional and enterprise MT software products, and human translation services
 Text: Yes
 URLs: Yes

*Note: This is a summary of active developers of online MT engines for European languages; there are many companies developing technology for Asian languages (especially Japanese) which are not included, as well as other European language developers who are not active in the Web market.