

As we enter the festive season of gadgety gifts and millennial fantasies, set against a background rumble of economic recession, we asked a number of language-industry luminaries to share their personal language-technology wish lists with us. We hope you agree they provide a judicious mix of lighthearted optimism, and darker concern.

So what would you put on your wish list?

# Dream On

The Language  
International  
New Year's  
Tech and Tools  
Wish List

**TIZIANA PERINOTTI**  
TGP Consulting, founder  
developer of Silicon Valley Localization  
Forum

*Dear Bobbe Natale,*

Most of all, I would like more of what there is:  
More software-localization tools for Web apps.

More use of search engines' multilingual functionality and better integration with email packages and other Internet apps: capability to query in any combination of different natural languages and retrieve full text in a selected language.

### More automation in localizing Web pages and content.

More porting/development of software localization tools for the Java environment.

More use of automatic voice recognition in a variety of languages as a standard feature in consumer applications.

**RENATO S. BENINATTO**  
president for Latin America of Language  
Management International, Inc.,  
Rio de Janeiro, Brazil

*Dear Bobbe,*

I have spent the last three days trying to look into the future and play Jules Verne for our industry. I never thought it would be so hard nor did I think about how practical constraints can affect our capacity to be creative.

Every major breakthrough that I envisaged could put me out of business, so I unconsciously boycotted it. Finally, I decided that no matter what, I would still have my language skills and I would be able to translate theater plays and poetry, act as a tour guide, or beg in the streets of Paris to practice my French.

Could I have the following:

Multilanguage speech to text. Dictate in one language and generate text in multiple languages at the same time. Translations can be read back and voice-edited.

### Translation Memory.

TMX is a reality and each company can use one tool of choice to its fullest extent. Translators can work remotely and share memories in real-time.

Machine Translation. Since I started in the translation business 16 years ago I have heard that Machine Translation will be a reality in 10 years, but there has been no real advance in the core tech-

nology in the past 30 years. I wish that by 2010 new technology allows massive volume of natural language text to be translated into and from the top 20 languages, regardless of subject.

### Subtitling tools that will allow automatic synchronization of subtitles and frames, cutting translation time in half.

Brazilian Portuguese has become the standard language for business and software development since Brazil became the leading economy in the world after winning the 2006 World Cup. Who said a wish list has to be realistic?

**MURIEL VASCONCELLOS**  
grande dame of practical machine translation  
at the Pan American Health Organization  
and tireless communicator of intelligent use  
of translation technology

*Dear Bobbe,*  
*Dear International*  
*Holiday Spirit,*

Please could I have a nice translation work station designed along the following lines:

**A piece of software I would love to have is a genre detector. In all the research prototypes we build in my lab today, we assume that the input will be of some particular genre, usually newspaper articles.**

I believe that ideally one should have (1) a good view of the source text; (2) relevant online topical word lists and glossaries, as well as a thesaurus; (3) three target-text windows to work from; and (4) online spelling and grammar checkers.

The source text should be in a column at the left (not on the upper half of the screen!) and be extensive enough to provide a good view of the overall context, including the previous sentence and two or three sentences after the current one. The word lists, etc., should be immediately accessible, and in the case of topical lists, there should be an optional switch that shows, possibly in color, when there are subject-specific entries available.

As for the three windows, the first would be the main working area, and it would start by providing a sentence from the original document (or a "machine pre-translation"), which could be overstruck or quickly deleted to allow the translator to work from scratch. The original text or pre-translation could be switched off. Of course, non-English characters and other symbols should be easy to produce (there is room for improvement in this area with Microsoft Office 97. "Drag-and-drop" is essential, and editing macros such as the ones developed by PAHO are also extremely helpful when one is overstriking or translating from scratch.

The second window would offer translation memory when such is available. The TM system should be capable of fuzzy matching, and the database should be very large, with the capability of including the organization's past texts if they are in some sort of electronic form. The third window would provide a raw machine translation. Both the translation memory and the raw MT should be easy to paste into the target document.

The spelling and grammar checkers currently available with Microsoft Office 97 are very helpful because they make the corrections while the text is being typed. The grammar checker can be tailored so that it is not so sensitive, and that's an important feature. It could stand to have a lot more grammar options. It would be ideal if one could write one's own grammar rules.

**EDUARD HOVY**

director of the Natural Language Group at the Information Sciences Institute of the University of Southern California, president of the Association of Machine Translation in the Americas (AMTA)

*What I want*

What I would like most to see from translation technology is to have it disappear altogether—into the innards of my computer. I do not want to know anything about translation technology; I simply want to have it there, ready and able, working whenever I need it. And I want to use it a lot—when I surf the Web to find new information about people and places and things; when I plan a trip to a foreign country; when I write emails to people who don't speak the languages I speak; when I download a technical paper or a newspaper article or an encyclopedia description.

Recently, I searched the Web for "hovy," and found that it means "blue" in an aboriginal language in Paraguay. There were several poems in this language, and a picture of a blue parrot. But I could not read the poems, and although I found (in Germany) a bilingual lexicon for this language and English, there was no translation engine at hand that could use this lexicon to give me even a crude translation. Useless!

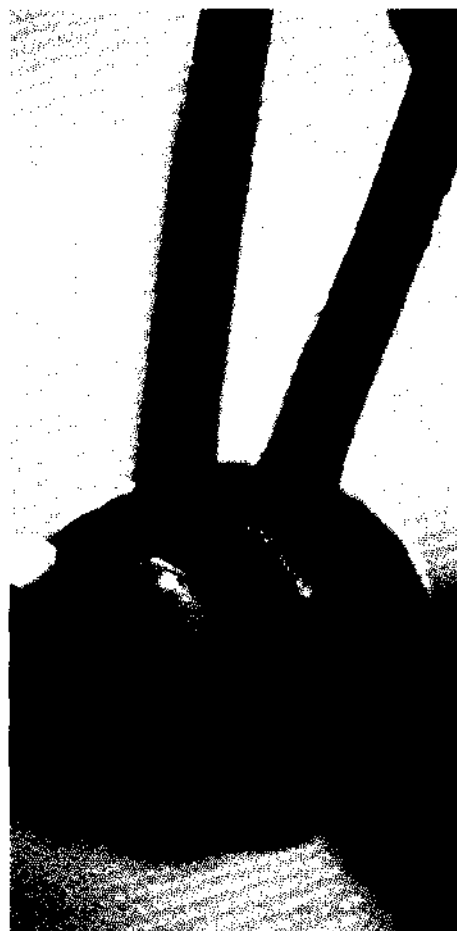
As long as I'm wishing, this translation engine must fit into my general working environment. It must tie directly into my Web browser, my document creator/editor, and my emailer. It must perform language identification correctly and must handle formatting, HTML and other markups, embedded pictures and other non-textual material, and even smiley faces :-)) without bother. In short, it must be intelligent enough to vanish.

**Next, I want a translator that I can take with me when I travel to foreign countries.**

It must be about the size of a Palm Pilot (that is, I must be able to slip it into my shirt pocket), and it must be able to scan (or even photograph) phrases from newspapers, hotel bills, road signs, and so on, and give me the translation(s). It need not be terribly accurate, but it must be comprehensive.

A piece of software I would love to have is a genre detector. In all the research prototypes we build in my lab today, we assume that the input will be of some particular genre, usually newspaper articles. When the prototype accidentally encounters a newspaper editorial, or a feature article, or— heaven forbid—a scientific article or poem, the results are hilarious (or disastrous). I know where the genre-sensitive assumptions are built into the theories and the software, and in some cases I even know how the assumptions must change for different genres, but I don't know what genres exist, and how to treat them all, or how radically they differ. We need a genre topology, a tree of more or less related types of text, and we need to have ways of recognizing and treating the different types computationally.

Finally, I would love to have a restless lexicon builder. This must be a piece of software that crawls all over the Web every night, ceaselessly collecting words, names, and phrases, and putting them into the appropriate lexicons. Wouldn't it be nice to walk in after a two-week vacation and see that in the meantime you have acquired new lexicons of 100,000 words and phrases of Tamil, Xhosa, and Twi? This may not sound like anything fun to you,



but to someone trying to build MT systems and text summarizers quickly and easily for lots of languages, this is part of heaven!

**MARK HOMNACK**

president of SimulTrans, the California-based localization and language-services company

*Dear Mr. Claus*

I'm sorry but I need to know how to translate "nothing" into 22 languages. For I have nothing on my wish list.

Why have I lost my optimism? Just like before many a Christmas eve, wonderful visions of translation technologies have danced before my eyes: translation-memory tools that suit various document formats, utilities that would recycle resource strings, developers who could support their product as well as they promised, runtime products that freelancers could use while working at home.

Alas, after reaching into many a Christmas stocking, time and time again, my hands came out with only coal, dirty with the signs of still another disappointing year. And, as if this mean ruse were not bad enough, I also felt the punishing cane of clients expressing anger that their "translation costs and turnaround times had not decreased like the miracle ads had promised." They would soon discover that all that money "invested" in building translation memory would disappear as quickly as Frosty the Snowman would melt away.

**How do you translate "nothing" into 22 languages?**

I have just about given up hope of any universal translation technology. After listening to the eternal and forever broken promises of Weidner 20 years ago, Systran 15 years ago, Logos 10 years ago, XL8 five years ago, and now Mendez's Internet ploys, I have grown cynical about the spruce hopes of Spring.

How much money has been lost in the pursuit of translation technology? Ask the developers named above, and their investors. How much benefit has been gained? Ask the clients who have bought their tools. To what extent should we believe the analysts' fantastic forecasts about translation technologies? Ask the people who have managed significant translation activities for five years or more.

**STUART SKLAIR**

document management specialist with MTL Ltd., the language services supplier, based in Luton, UK

*Dear Santa,*

Our product of the century would be:

A voice input-output translating personal organizer, including phrasebook, dictionary, and machine translation, as well as the usual personal organizer facili-

I would love to  
have a restless  
lexicon builder.  
This software  
would crawl all  
over the Web every  
night, ceaselessly  
collecting words,  
names, and  
phrases, and  
putting them into  
the appropriate  
lexicons. Wouldn't  
it be nice to walk  
in after a two-  
week vacation and  
see that in the  
meantime you have  
acquired new  
lexicons of 100,000  
words and phrases  
of Tamil, Xhosa,  
and Twi?

ties. You could also throw in internet connection and cellular videophone, though that's not specifically translation related.

You just turn it on and select the desired source and target languages. Point it at the person speaking. The microphone picks up the continuous speech. The system recognizes the source language and produces perfect spoken output in the target language. And all for under 3,000 euros. (Seiko and Systran are already working on something like this.)

As multilingual document management professionals, we'd also like to have author-memory tools, the integration of a translation and authoring toolsuite with document management and workflow, integrated TM and MT in one product, and integration of voice-activation technology with all products.

And by the way, they should not be produced by Microsoft!

**MINAKO O'HAGAN**

translator, researcher and author of *The Coming Age of Teletranslation*. She lives in Wellington, New Zealand

*Dear St. Jerome*

Please translate into reality this wish list for mainstream translator tools in the year 2005.

**Distributed Translation Memory**

Verified TM will be made available as readily accessible distributed databases which translators can tap into as required. This means that a large proportion of translation outputs would need to be linked to some kind of TM.

**Translation Memory Search Engines**

A sophisticated search engine will be an essential accompaniment to the TM databases. These will enable searches for exact matches of large amounts of text by various keywords such as phrases, authors, issuing organizations, etc. Distributed TM together with search engines will ensure translators can readily tap into previous translations when they exist.

**Software Agents**

A personalized multilingual software agent will be designed for use by a translator. The agent will cater to particular needs of its translator, interacting with other agents and information sources. It will tirelessly roam cyberspace 24 hours a day, including forums and newsgroups to pick up relevant information.

**VR encyclopedias**

These will provide the ability for the user to walk through the concept or words under study. As such they will provide an instant embodied meaning by letting the user virtually experience the meaning. Translators normally build up a gut feeling of words, phrases, concepts, etc., over time, but VR encyclopedias will cut the time the process takes.