



THE GOTHAM TRANSLATOR

March/April 2003

A Publication of The New York Circle of Translators

Conflict of Interests?



by Helene Bergman

This past January I joined a group of 20 "anglos" for a ten-day English-immersion experience for 20 Spanish executives in Barco de Ávila, Spain. The program was run by Vaughan Systems, one of the leading providers of English courses in Spain. As part of its extensive program of language courses, Vaughan Systems offers an immersion course in any of several locations in Spain. What each location has in common with the others is its almost total isolation from the rest of Spain. Indeed, one is almost unaware of the country they are in.

The Barco program is based in the Hotel Puerta de Gredos, a modern (actually still under construction at the time) hotel at the foot of the Gredos Mountains, about 2 hours due west of Madrid. The nearest town, Barco de Ávila, is 2 kilometers away. Dairy farms and an ostrich ranch surround the hotel. The views are spectacular.

The basic premise of the program is that once one has reached a certain level of proficiency in a second language, it is imperative to be totally immersed in it so as to begin to acquire that fluency we all so desperately desire. A fluency which, in addition, is necessary to the international businessman. By speaking and hearing English 24/7 for ten days the participants should feel more comfortable with the language that, for many of them, had previously been heard only in the context of the language classroom.

The Spanish participants came from all regions of Spain and from many different fields of work. Their levels of proficiency, too, differed considerably. They were a very bright, motivated group of people whose ages ranged from 28 to 45. Almost a third of the group was women.

The "anglos", as we were called, hailed from the USA, Canada, Wales, Australia, and Great Britain. We were students, teachers, HR people, and retirees all with one common link: a love of the spoken word! We ranged in age from about 25 to 70.

The group convened on a street corner in Madrid to board the bus for the trip to Barco. From that moment until we left Barco ten days later there was nothing but English spoken. We lived, breathed, ate, and slept ENGLISH. There was a daily schedule which paired people up, usually in two's, for one-hour meetings. We met for four hours before lunch, and three after siesta. All meals

were taken together. In the evenings we took turns making presentations, and performing in impromptu skits. Many participants rarely went to sleep before 3 am. Breakfast was at 9, NO MATTER WHAT. There was a one-day excursion to Ávila and two special evening events.

The one-on-one meetings were held outside: we walked, and walked, and walked. There was nothing else to do! Sometimes we took 2 hours and walked into town, had a coffee, and walked back. The weather was in the 30's and 40's and sunny.

I believe the program was successful despite certain inconveniences in the scheduling. Each Spaniard with whom I spoke thought he/she had gained a tremendous amount of self-confidence. Many spoke noticeably better, making fewer errors and demonstrating much better comprehension. The exposure to so many English accents was rather difficult at times but did sensitize them so that as the days wore on they were able to understand all of us much better. The common complaints were that there were not enough activities (it did get a

(continued on page 3)

INSIDE THIS ISSUE

ATA Announcement	2
Letter from the President-Elect	3
Language by the Numbers	4
Circle News	10
Monthly Meetings	11

THE GOTHAM TRANSLATOR

Editor

Helene Bergman
editor@nyctranslators.org

Editorial Board

Trudy Balch
Marian Greenfield
Nancy Wright

Design and Layout

Bahl Graphics
kbahlmann@austin.rr.com

Printing

Mail Boxes, Etc. South Orange NJ
mailboxso@mindspring.com

The Gotham Translator is published six times a year by the New York Circle of Translators, Inc., a chapter of the American Translators Association. The deadline for acceptance of contributions is the first day of the month preceding the month of publication. Articles and other materials should be e-mailed to the Editor at editor@nyctranslators.org. The opinions and views expressed are those of the authors, and do not necessarily reflect the opinions or the policies of the NYCT or the ATA. All material submitted is subject to editing and becomes the property of the New York Circle of Translators unless accompanied by a copyright statement.

© 2003 by the New York Circle of Translators. All rights reserved. Nothing contained in this issue may be reprinted without explicit prior permission of the editor or, in the case of copyrighted material, the copyright holder.

AMERICAN TRANSLATORS ASSOCIATION (ATA) PROFESSIONAL DEVELOPMENT COMMITTEE

CALL FOR VOLUNTEERS

LEGAL TRANSLATION CONFERENCE

May 2-4, 2003
Hyatt Regency
2 Exchange Place
Jersey City, New Jersey

Come and meet fellow translators, network and take advantage of an opportunity for learning as a volunteer at the Legal Translation Conference presented by the ATA's Professional Development Committee.

To find out more about how to get involved in assisting with seminar registration and logistics, either for a few hours or for a full day, please contact Memuna Williams at memuna.williams@verizon.net.



Geneva Worldwide, Inc. Translations & Interpreting Services

DIRECTOR OF BUSINESS DEVELOPMENT

We are an established NYC based translation/interpreting company looking for a key individual for our organization.

Exciting growth opportunity for entrepreneurial individual with following.
Proven track record a must. Possible vesting opportunity. Salary open.

Fax or mail resume to:

GENEVA WORLDWIDE, INC.
62 WEST 22 STREET
NEW YORK CITY, NY 10010
212-255-8409 (FAX)

For further information or to schedule an appointment, contact in confidence:

Ms. Patricia Coates, Vice President
212-255-8400

Letter from the President-Elect

Dear NYCT members:

I wanted to say hello and share some news with you. Due to health reasons, I am resigning from the NYCT Board as President-Elect, effective March 14, 2003.

I am okay, but have decided that I need to reprioritize my life a bit. All of the current Board members are doing such a great job this year (as

always) that there is little I really have to do for the Circle at the moment. But I did have things I wanted to accomplish for the Circle this year, and I definitely don't feel in a position to be President next year, so I believe the best thing I can do for the Circle is resign and allow the appointment of a new President Elect for 2003.

I enjoyed being NYCT Treasurer and a member of the Board for the past two years and thank you for your

confidence in electing me President - Elect of the 2003 Board. I want to thank all of you—both Board and Circle members—for being so great to work with during my tenure as an officer. I will still be at meetings from time to time, and am, of course, available for questions, coffees, hellos and anything else.

Sincerely,

Shena Crane

*Former NYCT Treasurer, 2001-2003;
NYCT President-Elect, 2003*

(continued from front page)

little boring after ten days; one week might have been better) and that there should have been grammar classes (the program specifically does NOT include any type of traditional class; we were asked NOT to "teach" but to converse). In addition, many participants believed that it would be better to have more time in small groups. One-on-one has its advantages but small groups provide a different type of experience, one which is probably more realistic in the business world. It is also much easier to come up with topics when there are more people involved.

As for the Anglos, most were there because they loved to teach, they loved to talk, and they wanted to learn more about Spain and the Spaniards. Only one or two of us actually spoke Spanish and were "under oath" not to speak so as much as one word of it for the duration of the program.

I joined the program for the above-stated reasons. I enjoyed meeting the participants and the former teacher in me loved helping them learn English. At times I questioned my sanity, as teaching people English certainly could not help my career as an interpreter/translator! But one of the wonderful things about translation/interpretation is that it brings us all closer together. Translators and interpreters enable people to communicate with one and other. This program did the same. And, although the participants did gain more self-confidence and were able to converse a little more easily, those who will be doing business in an English-speaking country will most likely still require or want an interpreter by their side to help them understand the nuances of the language and the many cultural notes which can be picked up only by someone fluent in the language and culture of the country. As I said goodbye to my new friends, I did not hesitate to hand out my business cards nor did they refuse them! I hope to see them and their friends here either as clients or as friends come to visit!

If you are interested in participating in this program contact Vaughan Systems at Vaughan@vausys.com. ■

NEW YORK CIRCLE OF TRANSLATORS (212) 334-3060

2003 Board of Directors

Teresa S. Waldes, President

(212) 769-9411

president@nyctranslators.org

Consuelo Corretjer-Lee, Secretary

(212) 627-4506

secretary@nyctranslators.org

Maureen Gonzalez, Treasurer

(718) 884-2952

treasurer@nyctranslators.org

Deborah Folaron, Program Director

(718) 381-4303

programdirector@nyctranslators.org

Committee Chairs

Teresa Waldes, ATA Accreditation

(212) 769-9411

twaldes@ix.netcom.com

Milena Savova, Continuing Education

(212) 998-7033

milena.savova@nyu.edu

Anne Witt-Greenberg, Mentoring

(516) 487-6128

trlegdoc@aol.com

Renée A. Borio-Román, Programs

(718) 823-8325

rborio@peoplepc.com

David Mintz, Webmaster

dmintz@panix.com

Language by the Numbers

By Joanne Cavanaugh Simpson

Reprinted with permission of Johns Hopkins Magazine, ©2002

By tapping the increasingly sophisticated "intelligence" of computers, Hopkins researchers aim to make possible the translation of nearly every written language in the world.

At the dawn of modern computing, the first "machines" weighed 30 tons, used bulky vacuum tubes, and had memories smaller than today's pocket calculators. Yet scientists were already looking to the day when this 20th-century invention would catch up with millions of years of evolution — and match the spark of intelligent life that fuels the human brain.

British mathematician Alan Turing, who designed a protoccomputer to break the German Enigma code during World War II, also proposed a test in 1950 that he believed would demonstrate when computers reached this level of artificial intelligence, or AI.

In the Turing Test, as it has become known, an observer or "judge" initiates a question-and-answer session via a keyboard with two entities: one a computer, one a person. If the judge can't tell the difference in the majority of cases, the machine could be described as effectively "thinking."

Despite Turing's prediction that by the year 2000 a human judge would not have more than a 70 percent chance of making the correct identification — and even with a number of contests, including one offering \$100,000 in prize money — no program has met the Turing Test to any degree of general acceptance.

Sure, by the mid-1990s, IBM's Deep Blue computer

defeated chess world champion Garry Kasparov in 19 moves, and a computer at the Argonne National Laboratory in Illinois developed a proof for the Robbins Conjecture, a difficult problem that had stumped human mathematicians for more than 60 years. But the question remains: When exactly does a computer understand? What is "understanding" or intelligence anyway?

Here at Hopkins, unique research into the areas of language and computer programming has been probing such questions. As part of the Center for Language and Speech Processing (CLSP) at the Whiting School of Engineering, researchers are training computer programs to "understand," translate, and cull information from texts in Chinese, Basque, Tagalog, Czech, and dozens of other sometimes obscure languages around the world.

David Yarowsky, associate professor of computer science, co-leads the Natural Language Processing, or NLP, research group. "A lot of people in computer science don't worry about whether computers think, or what qualifies as intelligence," says Yarowsky. "That is a philosophical question in the realm of Sartre or Kierkegaard, up there with the question of 'What

is the meaning of life?' After a while, what does it matter? If the computer gets so good at something that it looks like it's intelligence, maybe you can just call it that.

"Computers now play chess so well, and some of the questions answered by a machine can be quite sophisticated," Yarowsky adds. "Maybe the computer is just doing sophisticated pattern matching. But if you get back the right answer, does it matter if the computer understood?"

Yarowsky is sitting on a worn brown couch in his department's stripped-down lounge in Homewood's New Engineering Building. He has long been intrigued by foreign languages — having lived abroad, he speaks Spanish, Japanese, Nepali, and Ladakhi, a Tibetan dialect. And he

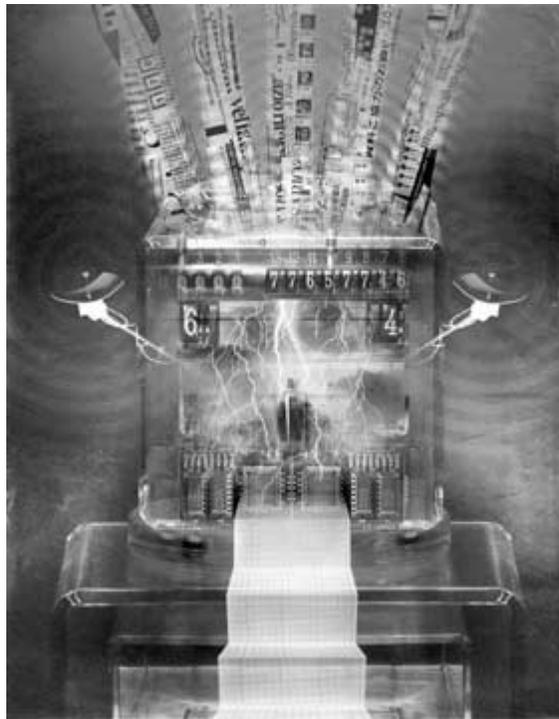


Illustration by Stuart Bradford

touts the potential for computer translation of human languages, also known as machine translation, in an ever-shrinking world where what's whispered in a mountain cave in Afghanistan is of interest to the U.S. Pentagon.

Automatic translation technology is useful outside national security circles as well. With vast and ever-growing information sources worldwide, today's scholars and researchers, for example, can't access all the archived texts or published papers — especially in foreign languages they don't understand. So the ability to use computers to scan texts in various languages for a piece of information, a trend, or a link between disease symptoms, for example, would be invaluable. There are potential benefits, too, to international commerce, where e-mail and Web sites could be more accurately translated, as could manuals, legal documents, and even phone conversations. "The goal is the universality of information," Yarowsky notes.

To help accomplish this goal, NLP researchers are tapping the vast memory capability, processing power, and increasingly sophisticated "intelligence" of computers to make machine translation, as well as information extraction, possible for nearly every written language in the world. As Yarowsky explains: "We want to make humans able to understand foreign languages, and computers able to understand any human language."

There's that word again. Understanding.

"The notion of a Universal Translator is a very real concept," says Yarowsky. "I believe that in my lifetime we will have computers that can roughly translate all the written languages in the world."

A computer program named Brutus can now translate simple Latin into English, helping students learn the ancient Roman language. An IBM scientist and world traveler recently used a digital camera and cell phone to send pictures of Chinese grocery signs to a server, where software translated the text and flashed the words in English on his cell phone screen.

American soldiers in Afghanistan this year used a hand-held device

called the Phraselator to translate up to 1,000 phrases, including, "I am here to help you" and "Show me your identification," into Pashto, Urdu, Arabic, or Dari. And in Croatia last year, conversation between Croatian and

English speakers became possible using a portable computer translator and speech synthesizer.

Such computer-aided language translation seems like science fiction. And Universal Translators like those used by Star Trek's Captain Kirk and crew won't be on Circuit City store shelves any time soon. Nevertheless, says Yarowsky, "the notion of a Universal Translator is a very real concept. I believe that in my lifetime we will have computers that can roughly translate all the written languages in the world."

Yarowsky, who earned his PhD in computer and information science from the University of Pennsylvania, says he was drawn to this field after Harvard University computer science professors showed him how computers could analyze language. "Human languages have so many different interesting properties and complexities," says Yarowsky. An adventurer by nature, he did volunteer work through a Rockefeller fellowship in Nepal and Tibet in the late 1980s, after completing his undergraduate degree at Harvard. In the early 1990s, he worked with speech synthesizers and language analysis at Bell Labs.

About eight years ago, Yarowsky decided to take the academic approach to improving computer-based speech understanding and translation by joining Hopkins' interdisciplinary CLSP, of which the NLP group is the computer science wing. CLSP, which was set up at Hopkins in 1992 with support from the National Science Foundation (NSF), the Department of Defense, and other federal agencies, brings together researchers from six Hopkins departments, including Biomedical Engineering, Cognitive Science, and Computer Science. Through collaborations, researchers focus on such areas as language modeling and acoustic processing (how humans hear language), as well as on how language is acquired. The center, considered one of the best in the world, draws top guest lecturers in the field and hosts an annual international research workshop.

In one project at CLSP, for example, researchers are working on speech recognition technology to help transcribe more than 117,000 hours of interviews with Holocaust survivors videotaped by the Shoah Visual History Foundation. With that much material, it's dauntingly time-consuming — and exorbitantly expensive — to have humans transcribe or index every tape. So, as part of a \$7.5 million NSF research grant, Hopkins computer scientists are developing software to recognize several languages, including Czech, Russian, and Polish. It's a challenging endeavor. As Bill Byrne, associate research professor in the Department of Electrical and Computer Engineering, has noted, such speech is heavily accented and highly

charged. "When people get emotional, the [speech] recognizers have a hard time. But that is the sort of spontaneous speech we want to record."

Various universities have built large research groups dedicated to computers and language — including Carnegie Mellon, the University of Pennsylvania, and the University of Southern California's Information Sciences Institute. Hopkins, through the NLP lab, has found a cutting-edge niche, specializing in what's known as statistically based machine translation and text analysis.

Currently, most machine translation technology, including consumer-oriented programs such as Systran's Babel Fish, have been "taught" the rules of language, such as verb tenses and when to use parts of speech. Programmers painstakingly hand-build systems based on such rules. "The computer is told, if you see this thing in Russian, replace it with this thing in English," explains Yarowsky.

While somewhat effective, such systems are time-consuming to build (consider how long it takes most humans to learn a language and all its rules), and resulting translations are still marred by grammatical and other errors. Those that do work fairly well usually tackle popular Western languages, such as French, German, and Spanish; there are few translation programs developed for other important tongues, such as Chinese, Turkish, or Arabic, let alone for more obscure languages like Tajik.

To tackle a broader range of the world's languages, and to improve on the quality of machine translation, Yarowsky and his Hopkins colleagues are developing computer programs that can be trained to figure out any language using statistical analysis, i.e., looking at the probabilities of language patterns. In what's known as automatic knowledge acquisition, the computer could "learn" Serbian well enough to translate future documents or conversation, or at the least pick out pertinent words like "bomb."

As Yarowsky explains: "Say you want to teach a computer how to translate Chinese: You give the computer 100,000



Photo by Jefferson Jackson Steele

Yarowsky has long been intrigued by foreign languages. Having lived abroad, he speaks Spanish, Japanese, Nepali, and Ladakhi, a Tibetan dialect.

words are likely translations.

"It's all just observation," Yarowsky adds. "Children do the same thing, but they also do it through visual stimulation and feedback. They see a book and hear the word 'book,' and eventually they learn that it's a book. They see a bird with its wings flapping around and learn that is called a bird. It's the same with machines, only they have much better memories. Computers could remember exactly when and where they saw the words bird and book."

So, instead of telling a computer how to do something — conjugate the verb 'to be' in Spanish, for example (I am = soy) — researchers give it tens of thousands of examples and program the computer to find repeated patterns that the computer can use to conjugate new verbs. Trained this way, the program could potentially "learn" phrase structure and the rules of translation.

As Yarowsky notes in his 100,000-sentence example, one way to accomplish automatic knowledge acquisition is to use bilingual or parallel text. The program "reads" a document in English and then a version in a second language. Such texts used by Hopkins researchers include the Bible, which is available on the Web in more than 60 languages, the Book of Mormon (over 60 languages), and the United Nations Declaration of Human Rights (240 languages).

Aiding the computer is the fact that the English version of such texts can be annotated by hand or using another computer program — essentially marked up to show, for example, that Jesus is a noun and pray is a verb. The translation program-in-training needs such information because it cannot translate future text just by substituting individual words in each language; it must also be able to analyze how sentences work. To do so, the computer program uses pattern recognition templates and other tools to understand

sentences on a syntactic level. Simply put, the program is essentially given clues to know what to look for, notes Yarowsky: "It should figure out the subject, figure out the object, and other elements of sentence structure."

Other tools used by Hopkins researchers to train computer programs to translate languages include bilingual dictionaries or lexicons that can be fed into the program, as well as WordNet, a thesaurus of sorts that shows links between words like pain, headache, and migraine. The end result: A computer program will be "trained" to translate Pashto or Basque or Hindi into English, even though it doesn't actually understand them. Or does it?



Photo by Jefferson Jackson Steele

"It sort of understands," says Yarowsky. "It partially understands some of the ambiguities, some of the meanings when words can mean multiple things. It can understand a lot of the structures of language, but it won't understand deeper subtleties. Some languages, for example Chinese, don't distinguish the male and female pronoun. He or she is the same word, so it can be ambiguous who something refers to. And sometimes there's a subtle metaphor."

So far, statistically based translation is faster to develop and more flexible, though often more plagued by grammatical or translation subtlety errors than the rule-based approach. Hopkins researchers have trained a program for Chinese, as well as one for Czech and French, that could roughly translate nearly any text. They are pursuing other projects with data from 240 languages. "It's intense work," Yarowsky notes. In some languages, like Turkish, a whole sentence can be represented by a single word and with Chinese, there are no spaces between words. A Chinese translation program created by Yarowsky and his colleagues already has outperformed current commercially available programs at recent machine translation competitions. "It's much more accurate on news text, which is what it was trained on, but it probably won't do very well on poetry," Yarowsky says. "Its accuracy depends on how many training sentences it has seen."

A famous anecdote in the machine translation field centers on the biblical saying "The spirit is willing, but the flesh is weak." When the phrase was translated into Russian by an

early computer translation program in the 1950s, the story goes, the answer came back: "The whiskey is strong, but the meat is rotten." Over the years, that story has been debunked as myth.

Yet enter the same phrase into Babel Fish Translation online today and translate it into, say, Spanish, and the answer comes back, "The alcohol is ready, but the meat is

Solving different pieces of the puzzle: (l to r) Eisner, Florian, and Schafer

weak." For some real fun, translate that back into English. The resulting phrase harkens to that game known as "Telephone"

where a phrase is passed down the line and misinterpreted along the way. The next Spanish-to-English version reads: "The ready alcohol this, but the meat is debil." And that's for two of the most commonly spoken and computer-translated human languages.

In the 1950s, during the infancy of machine translation, hopes were high that systems would soon be developed to rival high-quality human translation. The United States government poured millions of dollars into projects, fueled by an interest in Cold War-era translations and language analysis of Russian documents and radio transmissions.

With all the early limitations in hardware, software, and computer memory, the first machine translation researchers relied almost solely on bilingual dictionaries, and word-for-word translation. But researchers quickly realized that "perfect translation" was more difficult than they imagined. A federally commissioned report by the Automatic Language Processing Advisory Committee (ALPAC) found that machine translation had failed to reach its goal of adequate-quality translation by the 1960s, and likely would never be cost-effective. Generous funding sources soon dried up.

The Holy Grail question then and now remains: Will a computer ever be as good as a human translator? In many ways, not even close, at least until AI reaches the level of *Star Trek's* android character, Data. That's because language, in its many forms, is complicated and nuanced, ambiguous and contradictory, illogical and artistic — much like humans themselves. "Language is an incredibly complex, multifaceted puzzle, too big for any one person to solve," Yarowsky says.

Nonetheless, advances are being made today. And researchers are finding that machine translation doesn't

need to be "perfect" to be useful. Computers, in some cases, can do much of the heavy lifting in translation, with post-editing being done by humans. Partly to minimize such clean-up measures, Hopkins NLP researchers are tackling theoretical research in language acquisition and creating practical tools to improve translation.

Gideon Mann, now starting the fourth year of his PhD in computer science at Hopkins, says he was a fan of science fiction who hoped someday to converse with Asimov-style robots: "When I grew up, I was really upset that there weren't any computers I could talk to, so I thought, 'I guess I'll just have to build them,'" says Mann.

So far, Mann is developing software that can answer simple questions by analyzing sentences. Say, for example, that one has the question: "When did Hitler's armies invade France?" Mann's programs can search the Internet, looking for Web pages where a date or year is found near words from the question (i.e., invade, France, Hitler). In this case, "1940" would be the program's most confident answer based on statistical analysis relative to the syntactic context. In general, "the Web has a nearly limitless supply of information, and the more we understand about language structure, the more effectively we can harness this information," says Mann.

While such approaches are incremental and highly specific, these are the building blocks on which language "understanding" works — for humans as well as computers. Yarowsky, and the other researchers in his lab, are, in a way, engineers and architects and general contractors figuring out how to make each piece of the computer-language edifice fit together.

Linguistics is at the cornerstone of their endeavors.

The NLP lab's co-leader, Jason Eisner, assistant professor in computer science, uses a familiar computer science tool known as "finite-state machines" to program computers to analyze sentences on a highly syntactic level known as parsing — much like how English students look at the logical structure of sentences when diagramming parts of speech.

Richard Wicentowski, a linguist and computer scientist who has just finished his PhD at Hopkins, has been working with morphology, or the study and description of word formation. "Basically, it's the way that new words are built up from old words," Wicentowski says.

To provide a clearer picture of this linguistics-computer science link, Wicentowski explains how he trains computer programs to discern whether one word is related to another, such as drink and drank. "What you are trying to do is find ways for the computer to automatically discover the relationship between drink and drank," Wicentowski says. One way is to recognize that the words are nearly the same, except for one letter. Or, the program could scan nearby words, such as Coke or milk, for clues.

In a unique demonstration of how this technique could be used in any language, Wicentowski trained a program to conjugate Klingon, a language made up by particularly avid *Star Trek* fans. "It turns out Klingon is a very easy language for computers to learn because although it is complex morphologically, it was designed very consistently by one person," he says. Though Yarowsky's office boasts a copy of Shakespeare's *Hamlet* translated into Klingon by the Klingon Language Institute, neither he nor Wicentowski speaks Klingon. (The obvious question? If you want to ask, "To be or not to be?" in Klingon, simply utter "taH pagh taHbe!".)

Wicentowski says that using Klingon in translation and language research emphasizes how a computer program doesn't, in his opinion, actually "understand" the text: "The computer couldn't possibly understand what it is doing because I'm the one who told it what to do, and I don't understand."

For researchers like Wicentowski, it's the ambiguous meaning of words that remains — as was shown by the spirit-is-willing example — one of the primary hurdles. The word "plant," for example, could refer to a biological

The ambiguous meaning of words remains a primary hurdle. The word "plant," for example, could refer to a biological organism, a factory, a police "plant," or a ringer in the audience. How's a computer program to know?

organism, a factory, a police "plant," or a ringer in the audience. How's a computer program to know? The process to clarify the meaning of such words in various languages is known as "word sense disambiguation."

Radu Florian, also finishing his PhD, has been working on algorithms, or sequences of instructions, that teach computer programs to assign a specific sense to a word by giving it a large number of examples for when each meaning of the word is used. Through statistics, the program will know there's a 70 percent chance that when it sees the word worker near plant, plant will likely refer to a factory. "The program is given different parameters for different words," Florian says. "If the word leaf is near the plant, it would know that it's a living plant, not a manufacturing plant."

Yarowsky envisions how advances like those being pursued by himself, Eisner, Florian, and others will inevitably propel statistical machine translation to the next plateau. "With each [researcher] tackling a different piece of this puzzle, he says, "they can help provide an end-to-end solution."

A database residing at the NLP lab holds two terabytes of memory — that's 2,000 *billion* bytes or characters of text. And lab researchers have filled most of that memory up with stored text from over 100 languages, mostly news stories pulled off the Web. On a daily basis, a computer robot that acts like a super search engine accesses the Internet and automatically visits many of the newspapers and news sites in those 100 languages and downloads information.

"It takes the pages and strips the images and the ads and what's left is a news story about the events of the day. We try to line those stories up across languages," Yarowsky says. "If there is an earthquake in Chile, for example, a story on the earthquake might run in Poland, and China, and in Bangladesh."

Though the stories won't necessarily match word for word, much of the content, including the use of the word "earthquake" in various languages, will be similar. Through a process known as "iterative alignment," a computer program, given enough text, will start to pick out such similarities and translate key words.

Before the advent of the Web, and the subsequent explosion of sites in hundreds of languages, the availability of bilingual text was limited — especially in such languages as Azeri, Icelandic, or Punjab. Today's researchers, howev-

er, can in most cases find the comparable documents they need to train translation programs, whatever the language.

Hopkins graduate student Charles Schafer does research in information projection across languages. He uses bilingual texts to take NLP programs that analyze English and automatically develop the same analysis skill for a different language.

"Say you have a program that reads English sentences and identifies where people are claiming responsibility for bombings — people have spent lots of effort creating this capability over the years," Schafer says. "We can then run our existing programs on the English text, and use statistical techniques to figure out what kinds of clues in the Arabic translation indicate people claiming responsibility for a bombing. So you get the Arabic NLP program for free — as long as you can find the translated texts you need for this technique."

Schafer, now in the fifth year of his PhD, also was drawn to this focus in computer science because of a fascination with language, but in his case it was the history of English and the origins of words. He doesn't own an Oxford English Dictionary, though he points to a tattered expanded Random House sitting on a shelf by his desk. "The OED is on my wish list," says the graduate student.

Schafer's wish list includes perfecting the area of science he intends to make his career. It's a long shot, he knows. "We can make estimates that in several decades we will have one million times the processing capability," he says. "For the time being, we can improve. But it won't be human-quality anytime soon."

But that doesn't mean computer scientists can't dream about what Turing himself envisioned as the spark of non-biological intelligence that could someday lead to a deeper level of understanding, perhaps even surpassing that of humans.

Hopkins PhD student Florian tells a well-known joke about scientists building a computer as large as a planet. Once they build it, they try to figure out what to ask it. Eventually, they decide on the most central question plaguing humankind since the dawn of civilization: Is there a God?

The answer: "There is now." ■

Joanne Cavanaugh Simpson is a senior writer at Johns Hopkins Magazine.

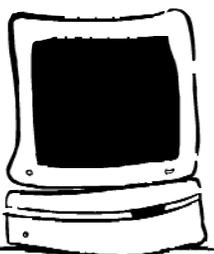
CIRCLE NEWS

The 2003 Directory is out!

The new Directory was printed and mailed (bulk mail) to all paid-up members in late March. If you haven't received your copy by April 15, please drop an email to the Secretary: secretary@nyctranslators.org



Gotham Translator online



The "Gotham Translator" page of our website has been totally revamped and it currently offers links to the Oct, Nov, and Dec 2002 issues. To view the new webpage, go to the website and click on "Gotham Translator." Providing access to an online archive of past newsletters is simply an added convenience for members and visitors to our website. The Gotham will still be mailed first-class to all current members 6 times a year.

Do we have your current email address?

Broadcast emails, a valuable membership benefit, inform you of time-sensitive job offers, upcoming courses and seminars, or NYCT events. We have improved the method for compiling the email address list, so that from now on NYCT emails will be reaching all current members (and only current members). However, some members are missing out on this benefit because they have not provided an email address as part of their online profile. Other members' addresses are outdated (after each broadcast at least 6 emails bounce back as undeliverable). Please make sure your profile includes your current email address.



Changes at SpanSIG

SpanSIG, which began life as the Spanish Special Interest Group for members of NYCT, has in recent years been operating independently with its own membership and board. In the near future SpanSIG's independent status will be formalized by its becoming a separate association with a new name, Intrades (Intérpretes y Traductores de Español), and a new website address, www.intrades.org. On another note, an interview with Leticia Molinero, President of SpanSIG, was recently featured in *El Diario - La Prensa*, a New York Spanish-language daily.

New United Nations multilingual glossary online



Long-time NYCT member Elizabeth Scott Andrews reports that the language service at the UN has finally made a permanent home for its six-language database on the UN Internet site. The database can be accessed through the following link: <http://www.unterm.un.org>. Liz warns that there are still some kinks in this link. Sometimes the site is not accesible. If this happens to you, just try again later.

Published NYCT Author

Consuelo Corretjer, Secretary of NYCT translated a short story *The Properties of Magic* for "Tameme," an annual bilingual journal of new writing from North America, published in California. The 2003 issue titled "Reconquest/Reconquista" will be released in March. Copies may be obtained by request to: Tameme, 199 First Street, Los Altos, CA 94022 or through the website: www.tameme.org. The journal's editor is C. M. Mayo.

Upcoming Monthly Meetings on April 22 and May 20

Remember that, because of our new publication schedule, you will not receive the next issue of the GT until June 1st. So **mark your calendars now** for the April and May meetings.

Note that for security purposes The Foundation Center requires a list of all attendees 24 hours in advance. From now on all members who wish to attend a meeting must preregister by calling or emailing Debbie, even if they are not going to the dinner.

April Meeting

Our April presentation entitled "Interpreting Interpreted" will touch among others on the following questions: What does it feel like to interpret? What is the state of mind an interpreter must achieve to work effectively? What happens when an interpreter cannot come up with the right word? What actions by clients and interpreting partners are the most likely to drive an interpreter out of her mind? What are some of the hazards which can result from working as an interpreter over a long period of time? What do famous writers have to say about the act of interpretation? What are some popular images of the interpreter? Why do laypeople always ask if interpreters 'change' what they are interpreting and some ways to answer this question..

Our presenter, Laura Esther Wolfson is a freelance Russian-English interpreter who has worked for a decade and a half in the fields of disarmament, diplomacy, human rights, cultural exchange, court interpreting and business, among others. She authored the popular "Miss Interpreter" columns in the ATA Chronicle, and recently was elected to the ATA Board of Directors. Her book translation, "Stalin's Secret Pogrom," published by Yale University Press, recently won the National Jewish Book Award prize for a work of Eastern European history.

The meeting will be held on Tuesday April 22, at 6:30 p.m. at the newly-renovated training annex of The Foundation Center, 79 Fifth Avenue, 2nd floor (between 15th and 16th Streets). Afterwards join us for

dinner at 8:00 p.m. at Cal's, 55 West 21st Street (between 5th and 6th Streets). Dinner is \$30 per plate (tax and tips included) with individual drinks paid for separately. Please preregister for the meeting and dinner or both by e-mailing Debbie at dfolaron1@nyc.rr.com or calling (212) 334-3060.

May Meeting

Guylaine Laperrière will give a talk on "How to get into voice-overs." She will give an overview of the type of voice that clients look for, the audition process, basic microphone techniques and studio etiquette.

Guylaine came to translation and voice-overs through acting. For the past 10 years, she has translated and recorded various TV and radio ads, corporate videos, training CDs, phone systems and language tapes (spoken and sung). Aside from the ATA and the Circle, she is also a member of the American Federation of Television and Radio Artists (AFTRA), Screen Actors Guild and Actors' Equity Association.

The meeting will be held on Tuesday May 20, at 6:30 p.m. at the newly-renovated training annex of The Foundation Center, 79 Fifth Avenue, 2nd floor (between 15th and 16th Streets). Afterwards join us for dinner at 8:00 p.m. at L'Annam, 121 University Place (13th Street). Dinner is \$25 per plate (tax and tips included) with individual drinks paid for separately. Please preregister for the meeting and dinner or both by e-mailing Debbie at dfolaron1@nyc.rr.com or calling (212) 334-3060.

Localization workshop June 2-3, in Montreal, Quebec, Canada

Specialized 2-day workshops on all facets of localization projects, with the internationally renowned Bert Esselink of Lionbridge Technologies and Marie-Claude L'Homme, expert Canadian specialist in CAT tools, of the University of Montreal.

For translators, localizers and project managers.

Ateliers de localisation et de traductique
Université Concordia - Études françaises
1400, boul. de Maisonneuve Ouest
Montréal (Québec) H3G 1M8
Canada

Information available at:
<http://www.vdn.ca/~nataliag/ad-com/L10N/index.php>
Program will be posted soon!

THE GOTHAM TRANSLATOR



The New York Circle of Translators (NYCT) is a New York State not-for-profit corporation regrouping independent translators and interpreters as well as companies and organizations. It is a chapter of the American Translators Association (ATA) which is, in turn, an affiliate of the International Federation of Translators (FIT).

NYCT members work in a variety of languages and specialties. Our members are committed to the exchange of ideas and mutual support. One of our goals is to educate the general public about the professional nature of interpreting and translating.

NYCT members enjoy the following benefits:

- Free monthly meetings featuring speakers on all aspects of the translation profession
- Networking opportunities at monthly dinners and annual holiday party
- Professional development workshops and seminars
- Subscription to our newsletter, *The Gotham Translator*
- Listings in the NYCT Online Membership Directory and the annual NYCT printed Membership Directory
- Referrals (if you indicate that you accept them, existing members may direct work requests to you or clients may contact you directly)

Visit us at www.nyctranslators.org and join on-line!

THE GOTHAM TRANSLATOR

PO Box 4051
Grand Central Station
New York, NY 10163-4051