

## Translation

Speech-to-speech systems translate lectures electronically in real-time. Ultra-directional speakers (background) enable the translation to be directed to a specific area in the auditorium.

# Catching the Babelfish

*Wouldn't it be nice to have a small device that translates your speech into any other language? Travelling would be a lot easier. If you think that's science fiction, like the Hitchhiker's Babelfish, think again! Researchers are close to fulfilling that dream – without the use of a fish.*

"Aiteru heya arimasuka?" I'm speaking Japanese! And it took me less than a second to learn it. All I had to do was push a button on the small device my *host* handed me and ask in English for a hotel room. Only seconds later, the device *squeaks* my question in fluent Japanese while the display shows my *transcribed* words and the translation in Japanese characters. "How much is a room for one person?" "Hitoriyo no heya wa ikura desuka?" This is fun. I go through the whole booking procedure, even *ranting* about missing towels and the smell of smoke in my room, while my host answers me in Japanese. Before I know it, I'm happily chatting away in a foreign language without even registering that I'm talking to, or better through, a machine.

The scene isn't taking place in a Japanese cyberhotel in the year 2050, but today at the University of Karlsruhe. The device is an *off-the-shelf* PDA, and my host is Prof. Alexander Waibel, director of Interact, a *joint* centre between the University of Karlsruhe and the Carnegie Mellon University in Pittsburgh, USA. The centre's *aim* is to research technologies and processes that *facilitate* and improve cross-cultural understanding and cross-lingual communication: A slight *understatement* for a system that, in Star Trek fashion, translates

natural speech in real-time into spoken words of another language.

How does this work? Waibel smiles at the *stunned* look in my eyes, a look that must be all-too familiar to him. "Speech-to-speech translation is basically a three-stage process," he says, this time in his native German. "It consists of automatic speech *recognition*, machine translation, and text-to-speech synthesis." Recognition and synthesis are, although *custom-made* and tightly integrated, more or less standard applications. The interesting part is the machine translation. Here, Waibel and his colleagues →

Vokabeldownload: [www.engine-magazin.de/extras](http://www.engine-magazin.de/extras)

<i>aim</i>	Zielsetzung
<i>consist of, to</i>	bestehen aus
<i>custom-made</i>	maßgeschneidert
<i>device</i>	Gerät, Apparat
<i>facilitate, to</i> • <i>fäßilität</i>	erleichtern, fördern
<i>host</i> • <i>houßt</i>	Gastgeber
<i>joint</i>	gemeinsam
<i>off-the-shelf</i>	Standard-, serienmäßig
<i>rant, to</i>	schimpfen, schwadronieren
<i>recognition</i>	Erkennung
<i>squeak, to</i>	quieken
<i>stunned</i>	fassungslos
<i>transcribe, to</i>	übertragen, abschreiben
<i>understatement</i>	Untertreibung

*Wäre es nicht schön, ein Gerät zu haben, das gesprochene Worte in jede andere Sprache übersetzen kann? Das Reisen wäre auf einmal viel einfacher. Wer denkt, das sei Science Fiction wie der berühmte Babelfish, der muss jetzt umdenken. Forscher lassen diesen alten Traum langsam wahr werden – ganz ohne Fisch.*



<http://interact.ira.uka.de>  
(engl.)

Interact ist ein gemeinsames Forschungszentrum der Universität Karlsruhe und der Carnegie Mellon University. Unter anderem entstehen hier maschinelle Übersetzungssysteme.

[www.mobytrans.com](http://www.mobytrans.com)  
(engl.)

Mobile Technologies ist eine Ausgliederung zur Kommerzialisierung der beschriebenen Technologien.

<http://chil.server.de> (engl.)

Ein weiteres Projekt, das die Interaktion zwischen Mensch und Maschine untersucht und diese zugunsten des Menschen verbessern will.

[www.ira.uka.de](http://www.ira.uka.de) (dt.)

Fakultät für Informatik an der Universität Karlsruhe.

take a radically different *approach* to the commercial translation software we know from our desktops.

"These systems are mainly rule-based while we *rely on* statistical methods," explains Waibel. Rule-based systems try to analyze the structure and grammar of a sentence. They locate subject, verb and all the other parts that make a sentence, recognize *tense*, active, passive or indirect speech, and translate them into their equivalent in the other language. This is an approach that *requires* a lot of linguistic *expertise* and computational *teach-in*. The resulting *rigid* framework generates good results on texts with an equally rigid structure like manuals, operating procedures or business letters. But it produces the well known *slips* on more 'lyrical' texts. While the famous 'The Whiskey is good, but the steaks cannot be *recommended*' for 'Der Geist ist willig, doch das Fleisch ist schwach' might be a joke, the results we get from rule-based systems are often just as funny.

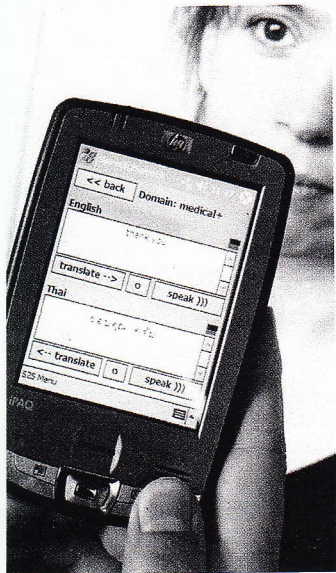
Statistical machine translation, in contrast, doesn't care about rules and grammar but instead, tries to figure out the *probability* of a word's meaning in a certain context. The method is not unlike the way we learn our own *mother tongue*. As a baby, we can't *consult* grammar books and dictionaries. Instead, we *estimate* the correct meaning of a word or the right tense from the way our parents and the people around us call things. This statistical approach enables us to speak fluently long before

we even know what a *gerund*, a *relative clause* or the past perfect is.

A method which Waibel's team is trying to *emulate* with their computers! All they have to do is feed the computer with words and sentences and their equivalents in another language, the more the better. Almost no teach-in is required since the algorithms and the statistical groundwork are universal and independent from the language set. "In theory, we can teach our computers a language that we might not even know," says Waibel, describing the advantage of the statistical method, "although speaking the language and knowing if something is right or wrong helps to *debug* the system. We also use a very limited number of rules which, for example, improve the recognition of numbers and times."

## Just Speak Freely

*Advances* in computing power and storage capacity allow Waibel to run these systems on a laptop or even a good standard PDA. Storage is especially important since, as with any statistics, the more data you have, the better the results are. To train the little PDA, which knows about 10,000 *terms* about tourism and *medical consultation*, several hundred thousand words in different contexts and from different speakers were necessary. A laptop translator might be based on several million words. "A good source



If the topic is restricted to a certain domain, e. g. tourism or medical consultation, speech-to-speech translation can be handled by a standard PDA.

<i>advance</i>	Fortschritt
<i>approach</i>	Ansatz, Herangehensweise
<i>array</i>	Reihe, Feld
<i>attend, to</i>	teilnehmen, besuchen
<i>audible * ʔdɔbl</i>	hörbar
<i>audience * ʔdiənʃ</i>	Publikum, Zuhörerschaft
<i>beam</i>	Strahl
<i>bother, to</i>	sich die Mühe machen
<i>consult, to</i>	hier: nachschlagen
<i>debug, to</i>	Fehler beseitigen, austesten
<i>dish</i>	Gericht
<i>domain</i>	Bereich, Gebiet
<i>don, to</i>	anziehen, anlegen
<i>eloquent</i>	beredt, redegewandt
<i>emulate, to</i>	nachbilden, nachahmen
<i>encounter</i>	Begegnung
<i>estimate, to</i>	schätzen, abschätzen
<i>expertise * ekʃpo'tiʃs</i>	Fachwissen
<i>gerund</i>	Gerundium, Verlaufsform
<i>guild * gild</i>	Zunft, Berufsverein
<i>headset microphone</i>	Kopfbügelmikrofon
<i>interpreter</i>	Übersetzer, Dolmetscher
<i>lecture</i>	Vortrag
<i>lecture, to</i>	vortragen
<i>medical consultation</i>	Arztbesuch

<i>mental arithmetic</i>	Kopfrechnen
<i>mock-up</i>	Attrappe, Übungstest
<i>mother tongue * mʔðə tʔŋg</i>	Muttersprache
<i>negotiate, to</i>	verhandeln
<i>newscast</i>	Nachrichtensendung
<i>odd</i>	seltsam, merkwürdig
<i>pick up, to</i>	merken, aufnehmen
<i>pick-up line</i>	Anmachspruch
<i>probability</i>	Wahrscheinlichkeit
<i>range</i>	Reichweite, Grenze
<i>recommend, to</i>	empfehlen
<i>relative clause</i>	Relativsatz
<i>rely on, to</i>	sich verlassen auf
<i>remote</i>	entfernt, entlegen
<i>require, to</i>	benötigen
<i>rigid * ridschid</i>	steif, starr
<i>scroll, to</i>	rollen
<i>skip, to</i>	überspringen, auslassen
<i>slip</i>	Ausrutscher, Fehler
<i>sonic</i>	Schall-, akustisch
<i>teach-in</i>	Einlernen
<i>tense</i>	Tempus, Zeit (grammat.)
<i>term * tɔm</i>	Begriff, Ausdruck
<i>topic</i>	Thema, Inhalt
<i>ultrasonic transducer</i>	Ultraschallwandler



for language sets we can employ to optimise our underlying processes are EU parliamentary speeches," says Waibel. Hours and hours of debates and thousands of documents were accurately translated into several European languages. A mechanical translator's dream! "Such large data sets are a perfect base to create open domain systems," raves Waibel.

Open domain? To give an explanation, he dons a headset microphone and starts lecturing in English. Usually I don't have any problems following an English conversation, but this time I did. While listening, I try to confirm the words he's using, which are continuously transcribed by the computer. I don't even bother to check the real-time translations of the lecture scrolling along on the screen in Portuguese and Arabic. But what really irritates me is an assistant who is trying to focus a sonic beam on my head which is whispering the lecture to me in Spanish. I, for my part, am speechless.

Sensing my linguistic overload, Waibel stops the demonstration and explains: "In contrast to domain-based systems for tourism or health care, open domain systems are not limited to a specific topic. They also work continuously without the need to press a button for every sentence. Typical applications would be monologue-like texts such as newscasts, meetings or lectures." The little demonstration was a mock-up for a system that could be applied at a conference. While the speaker talks freely in his mother tongue, the international audience can follow the lecture either by reading the 'subtitles' or listening to the translation.

To free the listeners from headphones, the set-up uses ultra-directional speakers. Unlike normal speakers, which emit sound in all directions, these are capable of focusing a sound into a very narrow beam, just like an acoustic searchlight. The odd sensation of a sound disappearing when you step out of focus is generated by an array of ultrasonic transducers. Only where their signals overlap do they create an audible interference. With such ultra-directional speakers a lecture room can easily be divided into a French, German, Chinese or Russian section.

But it still might be some time before we meet for a conference in Babel. The translation quality of the current open domain systems is comparable to an online translator for websites – it's not perfect, but you get the idea. Still, it might be good enough when you want to attend a conference and happen to speak an exotic language.

"Human translators are far more eloquent," declares Waibel, "but in real-life situations, they tend to skip sentences or parts of sentences when they can't follow the speaker. While our

system might not be that accurate, it works continuously without skipping." An advantage not to be underestimated: After following a translated speech test, people were asked questions about the content. With a human translation, 72 percent of the answers were right; with a machine translation, a close 65 percent of the answers were correct.

It seems that soon enough Waibel will become the public enemy of the translator's guild. "I don't think so," he smiles. "There are many situations with sensitive content, for example a political debate, where the correct interpretation is far too important to leave it to a machine. We want to facilitate communication and, therefore, concentrate on applications where the alternative would be to have no translation at all." A typical use would be a small technical conference with no budget for several conference interpreters, or a humanitarian organisation operating in a remote area with small, exotic languages. Naturally, the military is very interested in speech-to-speech translation systems. Addressing the locals in Iraq or Afghanistan in their own language and being able to understand them, even through a machine, could de-escalate an encounter dramatically.

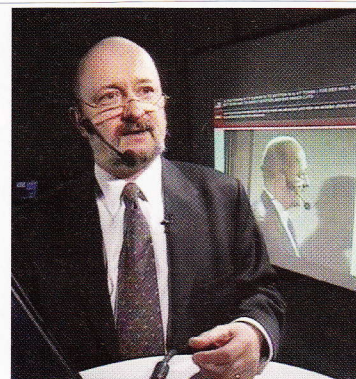
## Never Again Speechless

And, of course, mechanical translation also has a huge commercial value. Who wouldn't want such a PDA translation software package for his next trip to Spain, Thailand or China? Imagine being able to finally order the right dish in the restaurant, to extensively complain about the non-existent sea-view or to flirt with the beautiful blonde at the bar – although she might not find a pick-up line from a squeaky PDA overly romantic. "We are already negotiating with some companies," is all Waibel will say. Negotiating? I want one!

But won't we lose our interest in learning languages? Worse even, won't we stop reading 'engine'? Probably not! We can still do mental arithmetic in spite of pocket calculators; we still use maps in spite of GPS. But all these systems are extending our range, are getting us there faster. "Most people using our domain-based system pick up standard phrases fairly quickly and use them on their own without the PDA. Language learning is another possible application area for our technology," adds Waibel.

As if for confirmation, he smiles and says: "Itte rasshai." When I stare at him blankly, he picks up the PDA, pushes a button and repeats: "Itte rasshai." "So long." I just learned my first Japanese word. ■

Matthias Meier



Prof. Alexander Waibel loves to lecture about Interact's translation software in any language – especially the ones he doesn't know.