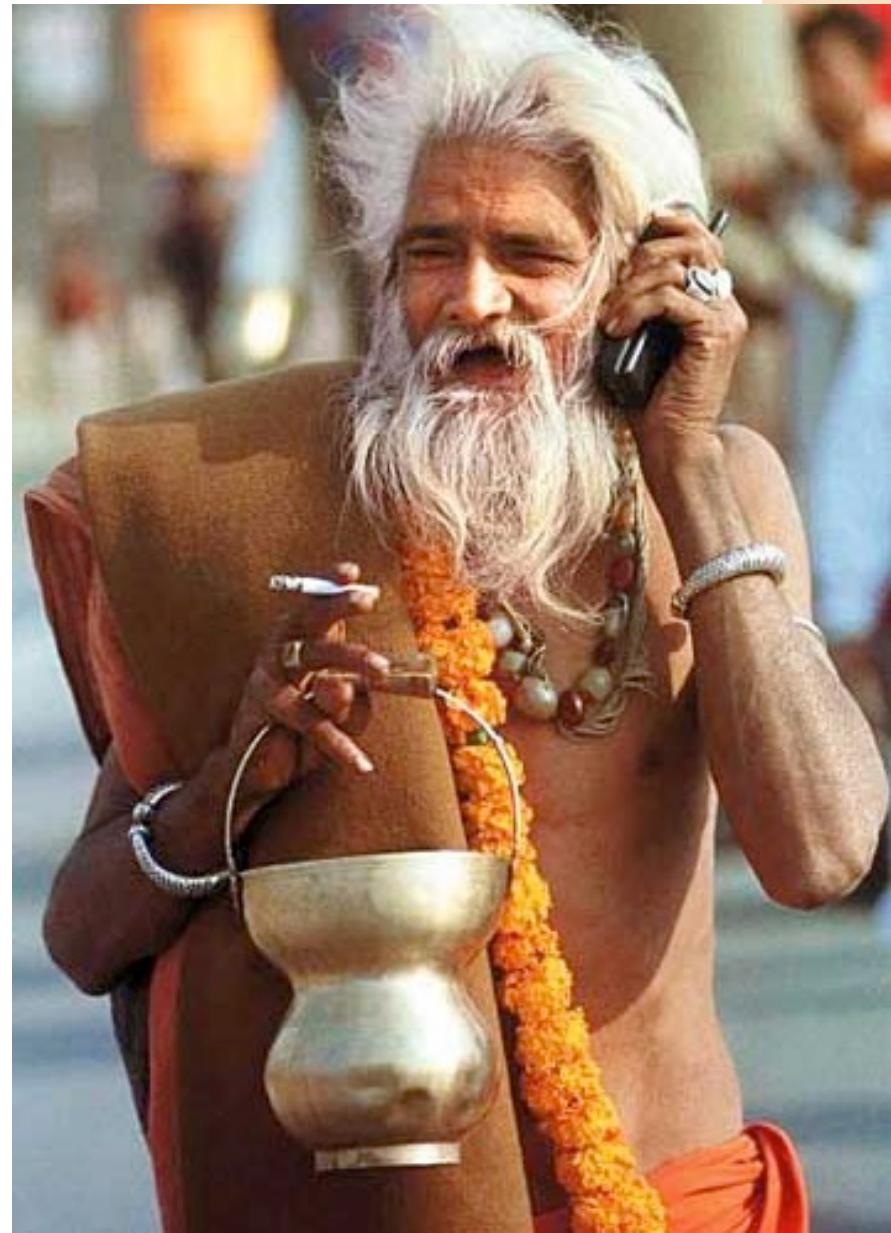


Technologies for Multilingual Human Communication

Alex Waibel and the InterACT Team
Carnegie Mellon University
Karlsruhe Institute of Technology

alex@waibel.com
waibel@cs.cmu.edu
waibel@kit.edu

- Living in the Global Village
 - **6 Billion Phones** (1 Billn PC's)
 - Unprecedented Exchange and Communication
 - Global/European Integration
- Cultural Diversity:
 - **6000 Languages**
 - Identity, Language, Culture,..
 - Pride and Individualism
- Challenge:
 - Access and Integration
↔ Diversity/Identity



1.) Communication is Multimodal

Communication is not only by Text, but:

- Speech, Gesture, Gaze..



- Images



你们的评估准则是什么

- Ill-formed Text

“lol-jah I want hr to be like dat...”, Hppyyyy BD, CU, LMK



2.) Communication is Multilingual



3.) Language is Ambiguous

„Give me a new display“

„Time flies like an arrow“

„The Spirit is willing but flesh is weak“

„Zentraleuropa“

„If the baby does not like the milk, boil it“

3.) Language is Ambiguous

„Give me a new display“ → „Give me a nudist play“

„Time flies like an arrow“ → 6 Parses

„The Spirit is willing but flesh is weak“ →

„The vodka is good but the meat is rotten“

„Zentraleuropa“ → „Headquarter Great Grandpa“

„If the baby does not like the milk, boil it“
→ the milk? the baby?

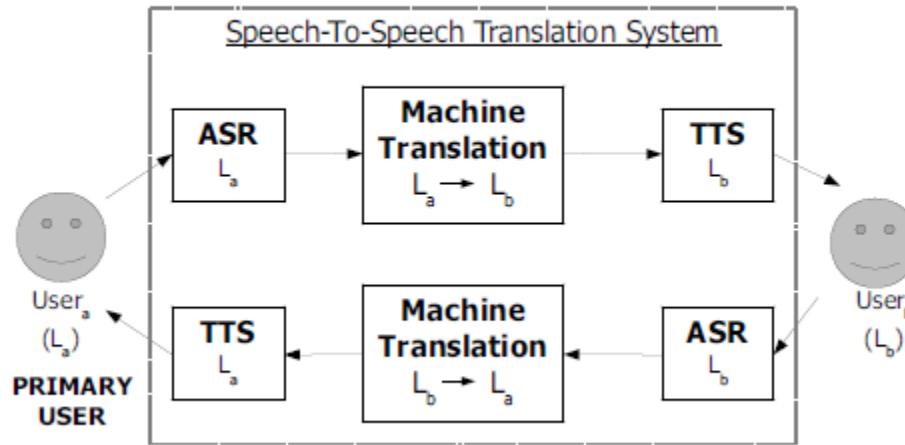


3.) Language is Ambiguous

Language Problems can only be Conquered,
if Machines Embrace, Represent, Process:

- **Ambiguity:**
Scores, Statistics, Neural Activations, ..
- **Machine Learning:**
Build Models, Extract Knowledge from
Human Data & Interaction, Automatically

→ Performance Depends on Data & Computing

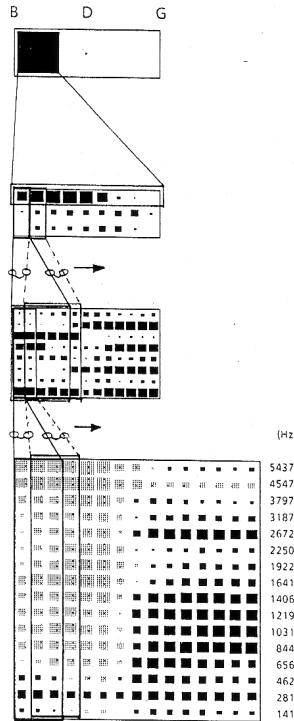


To Build a Language Communicator

- 6 Component-Engines: Automatic Speech Recognition, Machine Translation, and Text-to-Speech Synthesis
- Each is in Principle Language Independent, but Requires Language Dependent Models
- Models are Automatically Trained but Require Large Corpora
- Certain Language Dependent Challenges still Persist

Neural Nets 3.0: Bigger, Deeper, Faster

(1987)



TDNN: Shift-Invariance, Waibel '87

(1989)

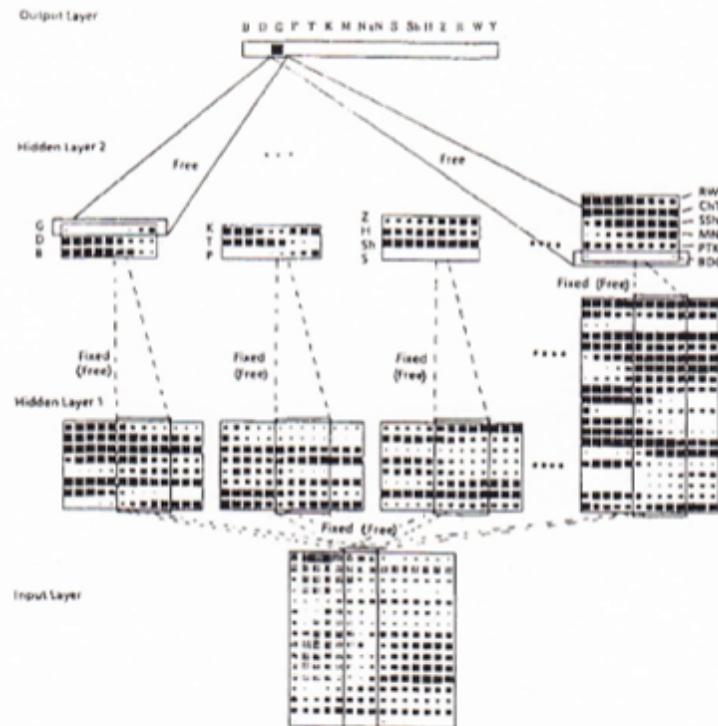
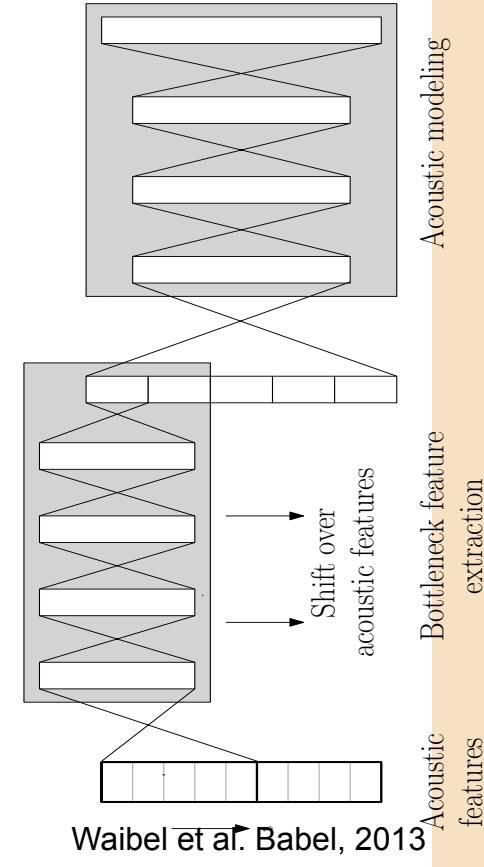


Fig. 7. Modular construction of an all consonant network.

Modular (deep) TDNN: Waibel '87

(2013)



Waibel et al. Babel, 2013

Weights: ~6,000

~40,0000

~33,000,000

TrnData[hrs]: ~0.1

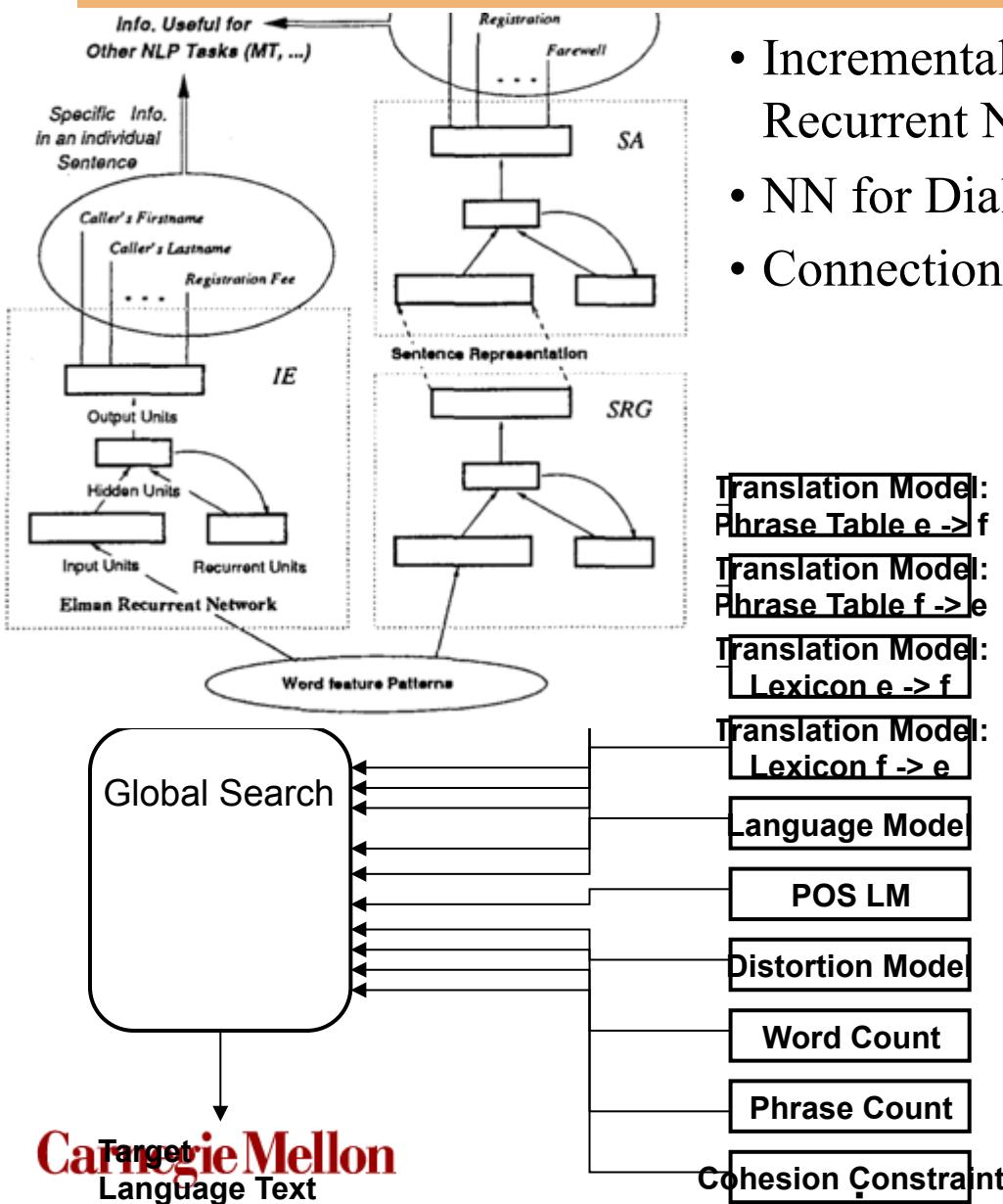
~1

~1,000

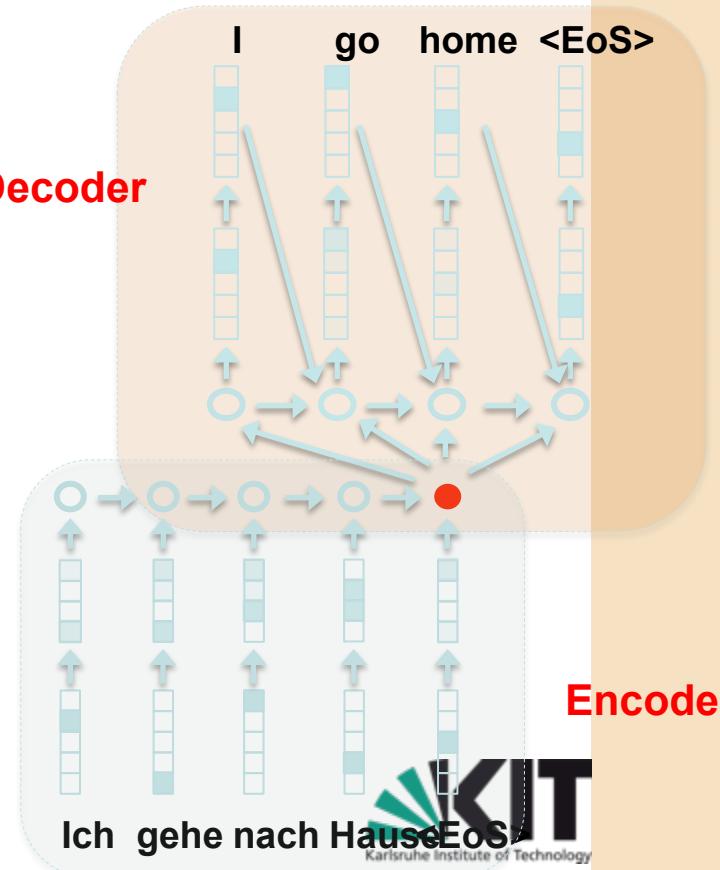
Time[weeks] ~1

~1

~1



- Incremental Parsing by Modular Recurrent Nets (Jain, Waibel '89-92)
- NN for Dialog Processing (Wang, Waibel '89-'91)
- Connectionist Transfer in MT (Wang, Waibel '95)



- 1992 – C-STAR Consortium for Speech Translation Advanced Research
- 1993 – Public C-STAR Demo, ATR-CMU-UKA-Siemens



2009



JIBBIGO



Hello. Nice to meet you!
(Hello, nice to meet you.)



Hola, encantado de conocerle.



*Pocket Translators for Travelers
offline embedded on iPhone - Jibbigo*



App Store > Travel > Jibbigo LLC



Free App

This app is designed for both iPhone and iPad

Category: Travel

Updated: Oct 12, 2012

Version: 2.1

Size: 15.0 MB

Languages: English, Arabic, Chinese, French, German, Italian, Japanese, Korean, Spanish, Thai

Seller: Jibbigo LLC

© 2012

Rated 4+

Requirements: Compatible with iPhone, iPod touch, and iPad. Requires iOS 5.0 or later. This app is optimized for iPhone 5.

Top In-App Purchases

1 Spanish-English...	\$4.99
2 European Bund...	\$9.99
3 Chinese-English...	\$4.99
4 French-English...	\$4.99
5 Asian Bundle fo...	\$9.99
6 Japanese-Engli...	\$4.99
7 Italian-English (...)	\$2.99
8 German-Englis...	\$4.99
9 Thai-English (I...)	\$4.99
10 Filipino (Tagalo...	\$4.99

Jibbigo Translator

Description

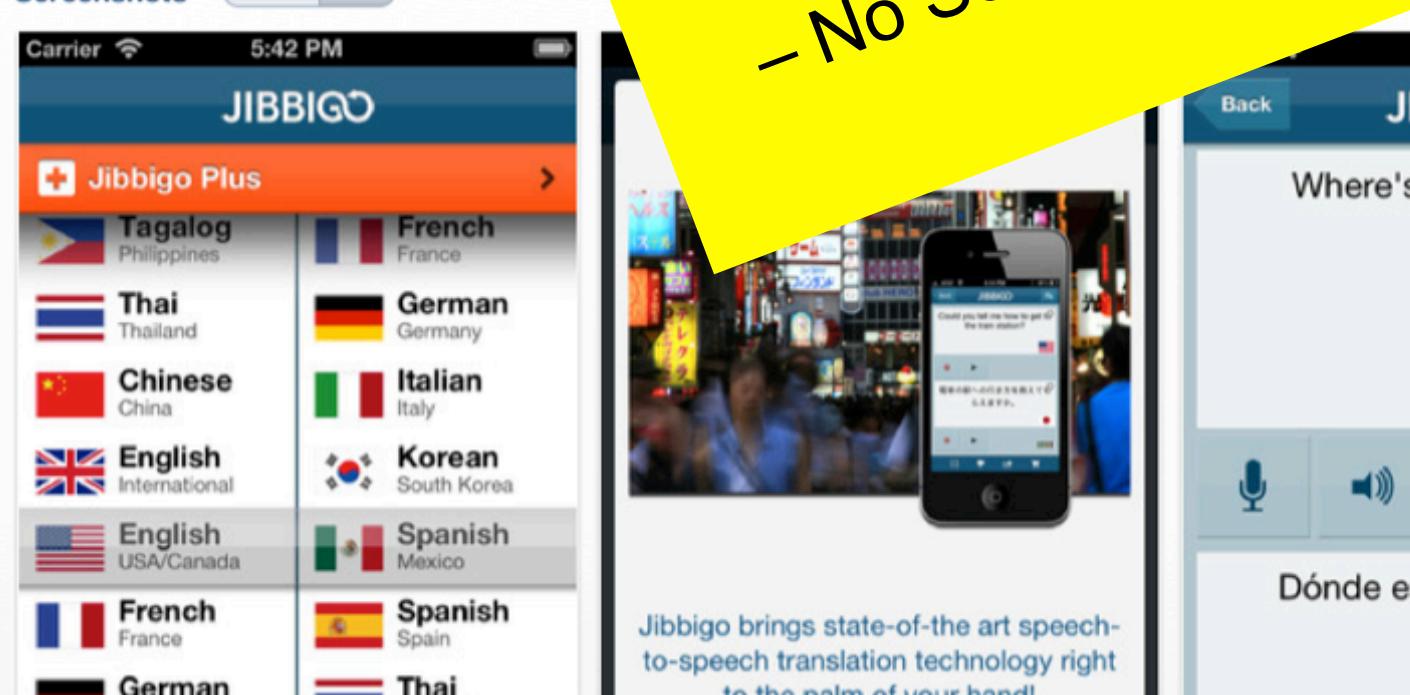
★★★★★ Finally! A FREE online voice translator. Unlimited Online Use ★Download Offline Translat...

[Jibbigo LLC Web Site](#) > [Jibbigo Translator S...](#)

What's New in Version 2.1

- * Support for iPhone 5 Resolution
- * Better Download UI
- * Bugfixes / Improved Error Handling...

Screenshots



Jibbigo:

- Real-Time Translation
- 15++ languages
- 40,000 Words
- All on the Phone
- No Server Necessary!

Jibbigo brings state-of-the art speech-to-speech translation technology right to the palm of your hand!



The Economist Jan. 2013. "...The most sophisticated technology currently belongs to Jibbigo....."





Jibbigo Joins Facebook, Sept. 2013

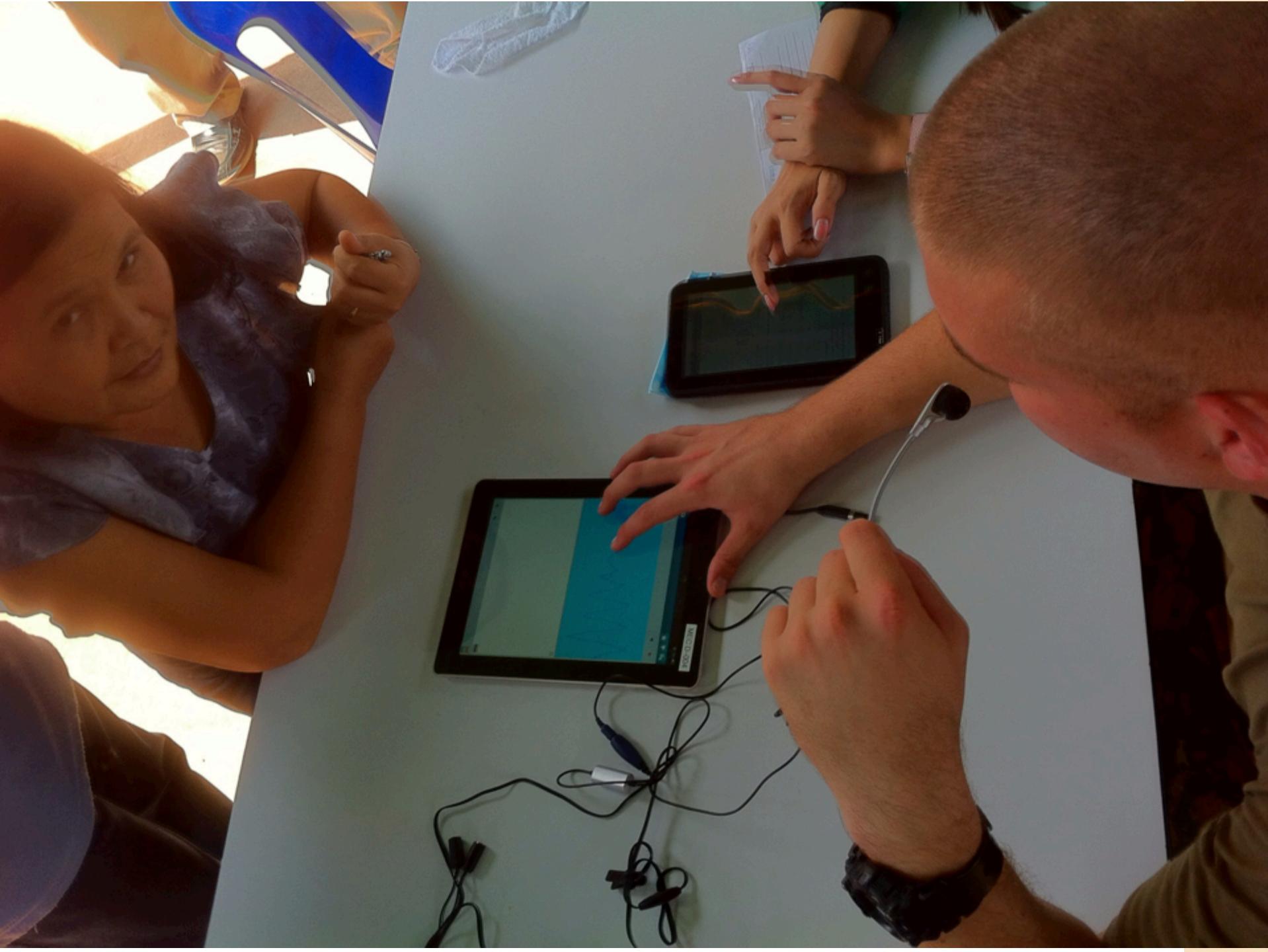


Humanitarian Relief



Thailand









Cambodia



San Jose , Honduras

Refugee Crisis





Unlimited Domain Simultaneous Speech Translation Technologies



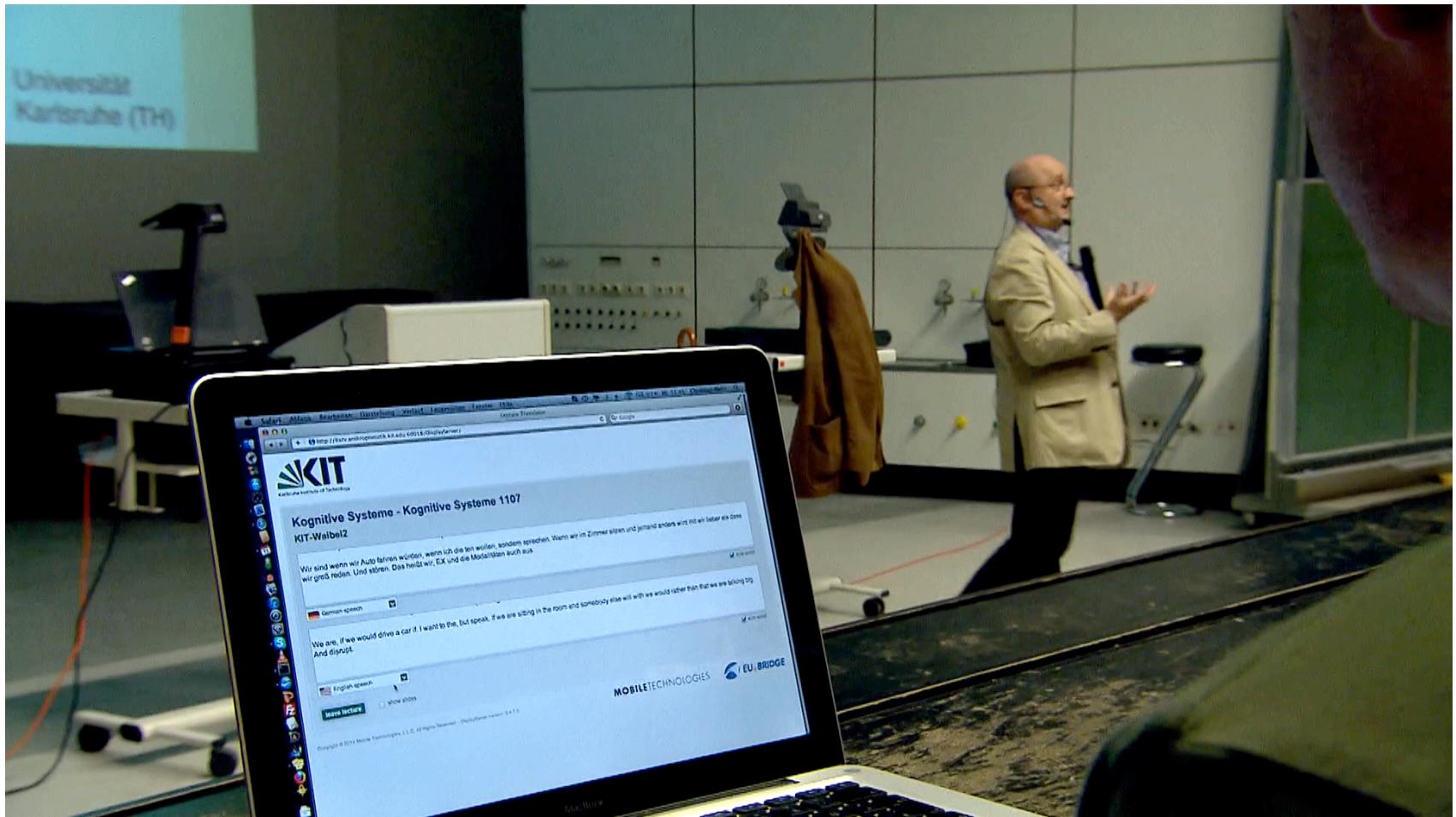
- Speech more Spontaneous than TED
- Real-Time Requirement
- Specialist Vocabularies

Studentin aus Tunesien



die Vorlesungen waren unverständlich

A close-up photograph of a young woman with long, dark brown hair. She is looking directly at the camera with a neutral expression. The background is blurred, showing what appears to be an indoor setting with bookshelves.

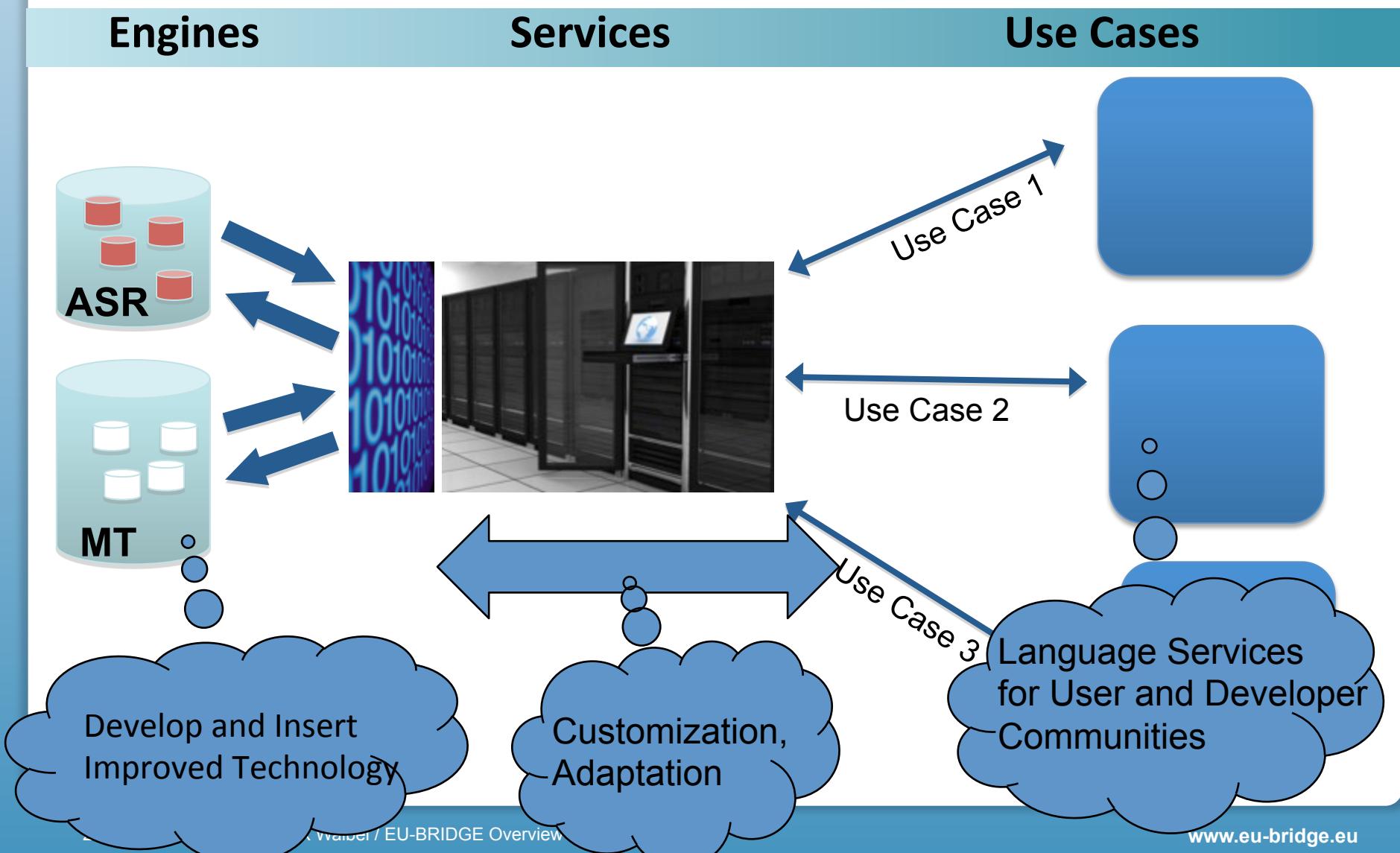


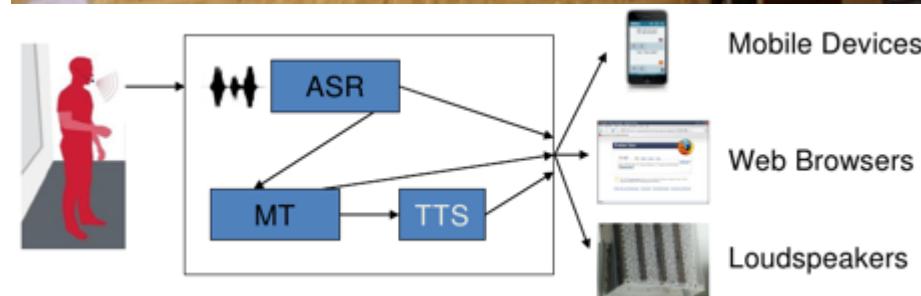
- Students bring their own Devices
- Transcription/Translation Output is Delivered via Web Page
- Interpretation Done on Server
- User Can Select Languages
- Launched 2012 as KIT Student Service
- Data Collected in Use and Evaluate



Prof. Alex Waibel

Language Services





English (en)

French (fr)

Wissenschaft

Was sagt der Professor da nur?

Wer kein Deutsch spricht, wird sich kaum für ein Auslandssemester in Deutschland entscheiden. Ein automatisches Übersetzungsprogramm soll Abhilfe schaffen. Von Joana Inés Marta



SPIEGEL ONLINE UNISPiegel

Login | Registrierung



Politik | Wirtschaft | Panorama | Sport | Kultur | Netzwerk | Wissenschaft | Gesundheit | einestages | Karriere | Uni | Reise | Auto | Stil

Nachrichten > UniSPIEGEL > Studium > Integration > Vorlesungsübersetzer am KIT: Translator-Programm

Vorlesungsübersetzer: "Very very, very many possible word follow"



Studentin Yanglingzi Wang am KIT: Mindestens 80 Prozent verstanden

In einer Vorlesung sitzen und alles verstehen? Das wünscht sich sicher jeder ausländische Student in Deutschland. In Karlsruhe haben Wissenschaftler nun ein Programm in Betrieb genommen, das in Echtzeit übersetzt, was der Dozent erzählt.

5.03.2015

christoph Drösser

1 bis 32

Wissen

Wochenzeitung

Jahrgang:

2015

10

631.846 (gedruckt) 510.634 (verkauft)
553.335 (verbreitet)

1.62 (in Mio.)

NACHRICHTEN

Unter sich

Der Ostliche Berggorilla, der in der Region des Virunga-Vulkane im Grenzgebiet zwischen dem Demokratischen Republik Kongo, Ruanda und Uganda lebt, scheint Anfang der achtziger Jahre schon fast ausgestorben zu sein. Wilderer und das Zerstören des Lebensraums hielten die Population auf so niedrig wie nie wieder. Dank Schutzaufnahmen hat sich diese Zahl inzwischen wieder verdoppelt. Doch es wurde befürchtet, dass das immer noch zu wenig sei, um schwere Erneuerungen zu verhindern. Unter dem Namen 'berühmtes Wilderntier' tritt es an man geln, die Genome von vielen Berggorillas zu diesem Gedanken zu superponieren. Die Analyse ergab, dass die enge Verwandtschaft nicht so geringe Schäden generiert wie man dachte. Eine Mischung aus Mutationen aus dem Genpool entsteht. Außerdem zeigte sich, dass die Tiere offenbar schon seit Tausenden von Jahren in kleinen Gruppen überlebt haben. Es gibt also keinen Grund, warum sie nicht auch in Zukunft möglich ist, solvielen die Ameisen in Sizien.

buch Wales

Manuskript komplett in Sprache ist das sogenannte Buch von Carnes-Semmelrohrschrift aus einem Jahrhundert, deutscher ist sie, entziffern kann es derzeit niemand. Dieses Antikenschriftenstücke. Eine Doktor-Candidatur entdeckte im Manuscript im Jahr 2010 von vier britischen Prinzipalitäten. Sie sind jetzt wieder zurück geworden, nach der Fertigstellung eines gleichen Codexes, teilte er mit Cambridge mit.

t, Papaya

sitzt in China und im Hintergrund des Computers, der nach dem deutschen ist als englischen Jargon, der in und unverständige, jeden zweitend. „Mir geht es gut, aber ich kann mich nicht mehr mit dem Computer unterhalten“, schreibt der Dozent.

Man kann sich vorstellen, wie Menschen Pfeile und Bogen und Waffen und Rüstungen und Körbe in die Zukunft schicken. Eigentlich hätten sie es vorstellen müssen, wie Computer einfach die Internet-Netzwerke überwachen und verstehen. Nachdem das Menschenkind global per Telefon und Internet versteht, was es will, werden die Grenzen zwischen Mensch und Maschine verschwimmen. Stattdessen gehen die Themen zu einer Demonstration der Schreibgeschicktheit, die dieses Menschenkind ist.

Zugriffen muss man den Computer, der mit seinen Hörer und Melodien versteht, das Kommando geben, um es zu benutzen. Und dann ist eine Interaktionstechnik, die man Vokabeln oder Videobildschirme im Prinzip benötigt. Die

e haben verstanden

Foto: © Universität Karlsruhe des und erfüllen ein großes Versprechen der in endgültige von kleinen raus.



So sieht sich der Künstler Jan von Holzen einen Diomede-Aapparat vor

will mit einer neuen Übersetzungstechnik, und dann seien es zwei, und eigentlich, Menschen, die an einer Sprache arbeiten. Sie sind dann nicht mehr alleine, sondern sie haben zusammen mit anderen kleinen Menschen, die eine entsprechende Sprachmaschine konstruiert – auch eine Diomede-Aapparatur.

Bei der Science-Fiction kennen wir das schon: Eine Gruppe von Menschen, die das selbstkritische System der Diomede-Aapparatur schafft, aber mit ungängiger Weise kommt ein Google-Forscher in dieser Welt und bringt die Diomede-Aapparatur in die Welt der Menschen. Das scheint eine Art von sehr leichter, die versteckte Sprachmaschine zu sein, die auf einer einzigen Biobasis ist. Obwohl es sich um eine sehr einfache Biobasis handelt, kann sie die Sprache des Menschen leichter verstehen und kann sie leichter und schneller übersetzen.

„Mir geht es gut, wie geht es dir?“ wurde plötzlich zu „Well how is it me friend?“

Parallelenbildung: CIBIO am KIT. Der Wissenschaftler Michael Beetz ist überzeugt, dass in zehn Jahren wird des jetzigenwurde. Tausen Wissenschaftler umsetzen und für eine hohe Masse 500 Millionen verschiedene Sprachen. Das ist eine riesige Anzahl. „Wir die staunende Beobachter, werden uns darüber freuen, dass wir bald aus, die wir kein Interesse für die Worte der anderen Sprachen haben“, sagt Michael Beetz. „Wir werden uns dann auf den Kurs konzentrieren, der Samsung Edition streng.“ Der Kurs der English-Sprachmaschine ist eine Art von technologischer Entwicklung, die nicht nur technologisch, sondern auch sozialen und kulturellen Veränderungen folgt, auch in Sachen Digitalisierung. Das Projekt Vorlesender, in dem die

Beteiligungen zwischen 1990 und 2000 umgehen. In Tübingen ist Michael Beetz und seine Tochter auch sehr froh darüber, was die neuzeitlichen Sprachmaschinen gebracht haben. „Die Sprache ist etwas, das wir nicht verstehen“, sagt Michael Beetz. „Wir können sie nicht verstehen, weil sie nicht versteht.“ Michael Beetz ist überzeugt, dass die Sprachmaschine die Sprache besser verstehen kann, als wir es tun. „Ich denke, dass es eine Art von Sprachmaschine ist, die die Sprache besser verstehen kann, als wir es tun.“

Auch Sprachmaschinen können die Biobasis, bei denen es sich um eine Biobasis handelt, leichter und schneller, als einer Konversation in das ganze andere, weiß Chris Wrede, bei Michael Beetz. „Wir können die Biobasis leichter und schneller verstehen, als wir es tun.“

„Mir geht es gut, wie geht es dir?“ wurde plötzlich zu „Well how is it me friend?“

Übergangsphänomene übernehmen, braucht es nicht viel. „Wir können die Biobasis leichter und schneller verstehen, als einer Konversation in das ganze andere, weiß Chris Wrede, bei Michael Beetz. „Wir können die Biobasis leichter und schneller verstehen, als wir es tun.“

Und die dann hat die Ergebnisse der Überprüfung von „Well how is it me friend?“

WISSEN





Three Use Cases:

- Terminology Support
- Named Entity Support
- Interpreter's 'Cruise Control'



22:41 - 22:43
three by roll call the vote is op

22:48 - 22:49
vote is closed

22:53 - 22:54
and that you rejected

22:55 - 22:59
I put the joint motion for resolu

23:00 - 23:00
those in favour

23:02 - 23:05
those against abstentions that is

23:07 - 23:08
and that concludes that

23:11 - 23:18
commission work programme but we now move to the vote on the report by **Nicolas Yannakou**
tissues and cells

23:20 - 23:22
X paragraph **three**

23:24 - 23:25
by split vote in **two parts**



Voting Sessions

Observations:

Interpreting Voting Sessions is...

- Boring and Repetitive
- Still Stressful, and Demanding
- Many Numbers
and Named Entities



Continuing Research Issues

- Neural MT
- All the usual issues:
 - Morphology, Pronoun Resolution, Word Order
- Learning in Practice
 - Incremental Learning
 - Crowd Source
- Improving Quality of Speech Translation
 - Low Latency
 - Dealing with Errors and Disfluencies
 - Put the Drama back
 - Suitable Intonation or Textual Realization
 - Emotion and Emphasis
 - Contextually Aware MT
 - Formality/Informality in Translation
 - Male Female
 - Plausibility in MT

Die Fehlerstromschutzschalterprüfung

Die Wirtschaftsdelegationsmitglieder

Die Bankwirtschaftsfreigabeerklärung

Die Lehrverpflichtungserklärungen

Die Schiffskommunalschuldverschreibungen

Die Vorkaufsrechtverzichtserklärung

Das Mehrzweckkirschtentkerngerät

Die Gemeindegrundsteuerveranlagung

Die Nummernschildbedruckungsmaschine

Der Mehrkornroggenvollkornbrotmehlzulieferer

Die Verkehrsinfrastrukturfinanzierungsgesellschaft

Die Feuerwehrrettungshubschraubernotlandeplatzaufseherin

Das Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz

Zentraleuropa:

Zentral-Europa → Central Europe

Zentrale-Ur-Opa → Headquarter-Great-Grandpa

Dramatisch:

drama-t-isch → dramatic

drama-tisch → drama table

Asiatisch:

asia-t-isch → asian

asia-tich → asia table

- Foreign Words in German Language
 - Computer Science, English Expressions
 - Political Speeches, Latin Proverbs
- Accent
 - “Würfelkalkül” (Asfour)
- Foreign Words in German Language
 - “Cloud”, “iPhone”, “iPad”, “Laser”
- Inflections & Declinations of these Words
 - Web-ge-casted, down-ge-loaded
- Formation of Compounds:
 - Cloudbasierter Webcastzugriff

1.) Communication is Multimodal

Communication is not only by Text, but:

- Speech, Gesture, Gaze..



- Images



你们的评估准则是什么

- Ill-formed Text

“lol-jah I want hr to be like dat...”, Hppyyyy BD, CU, LMK

- Robust Text Translation for Posts, Texting at Facebook:
Use posts data to expand “vocabulary” and language use

[!\[\]\(4be02dac9417b9e8d32253d4bd08acc3_img.jpg\) Like](#)[!\[\]\(e2d6a0f151413f4a76086198b0b68027_img.jpg\) Comment](#)[!\[\]\(6428903e136f53fa8dd5d5c5ed1ecaa7_img.jpg\) Share](#)

Nicoletta Calzolari Zamorani and 36 others

[View 16 more comments](#)**Claudio Bertuccelli** Enjoy your new location! I lived in Boston 8 years...[Like](#) · [Reply](#) ·  1 · June 7 at 7:27am**Stella Ilchenko** Nice space.[Like](#) · [Reply](#) ·  1 · June 7 at 8:09am**Valentina Cortopassi** Anch'io sono esperta di traslochi e arredi... Se vuoi ti vengo a dare una mano... Ahahahahaha!

Scherzi a parte ora è il momento del caos, poi arriverà la soddisfazione!!! Baci

I'm an expert of moving house and furniture... If you want I'll to lend a hand...
Ahahahahaha!

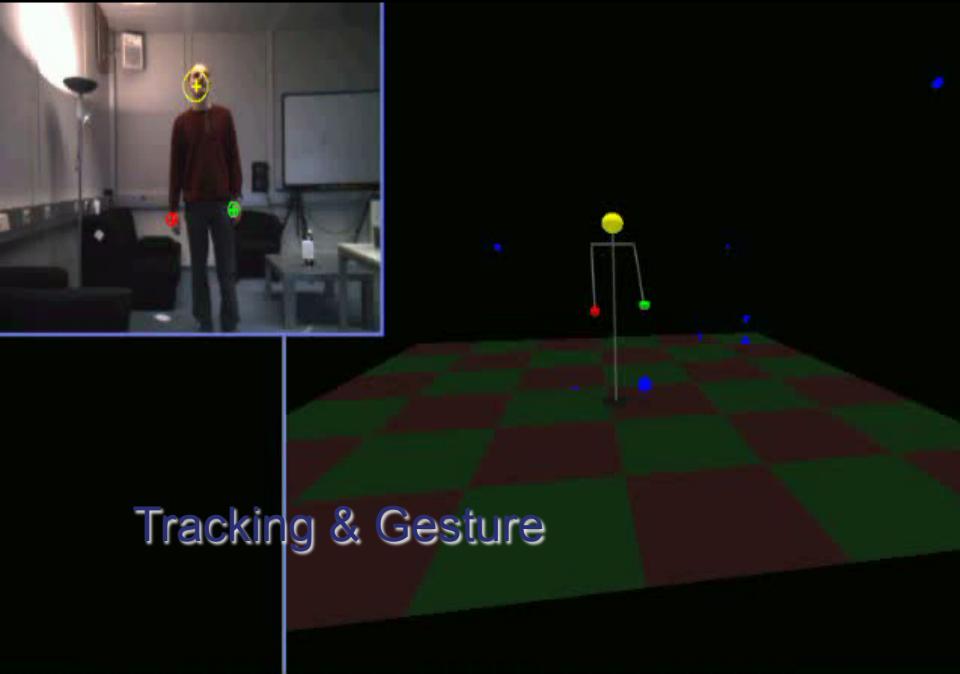
Seriously now is the time of chaos, then will come the satisfaction!!! Kisses

Road Sign Translator (2001)



A PDA-Based Sign Translator, ICMI 2002, Zhang, Yang, Chen, Waibel

Localization

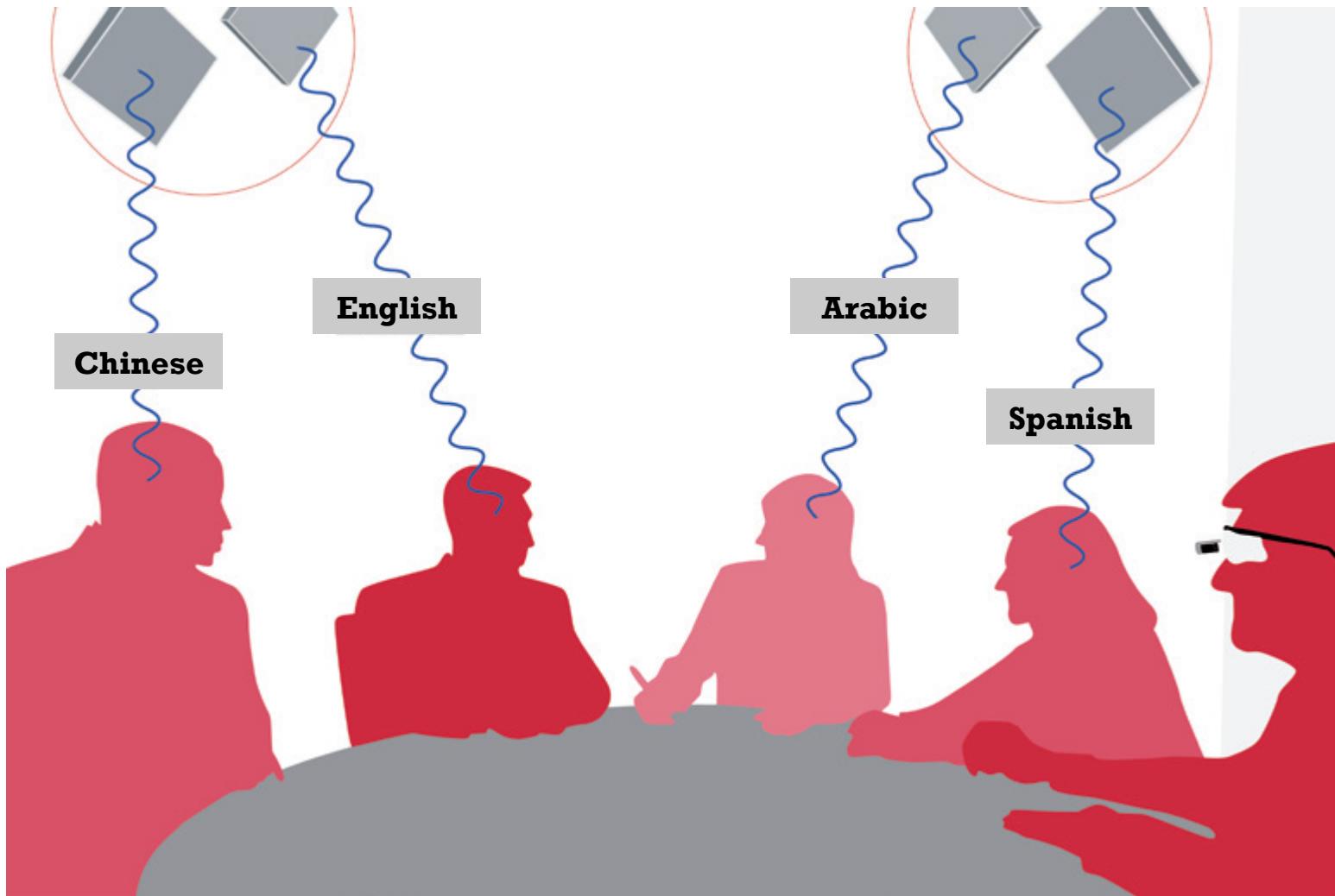


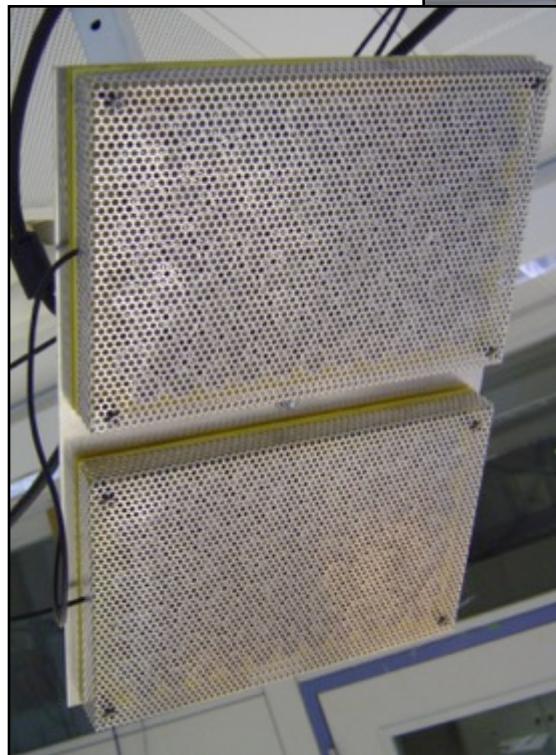
Tracking & Gesture

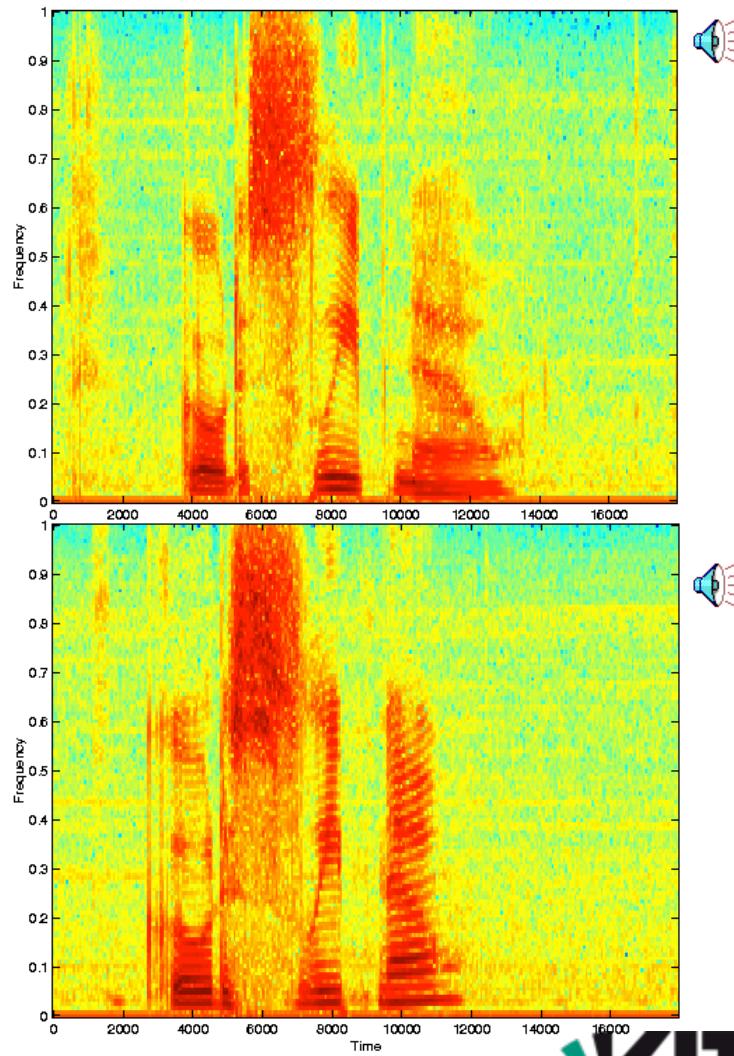
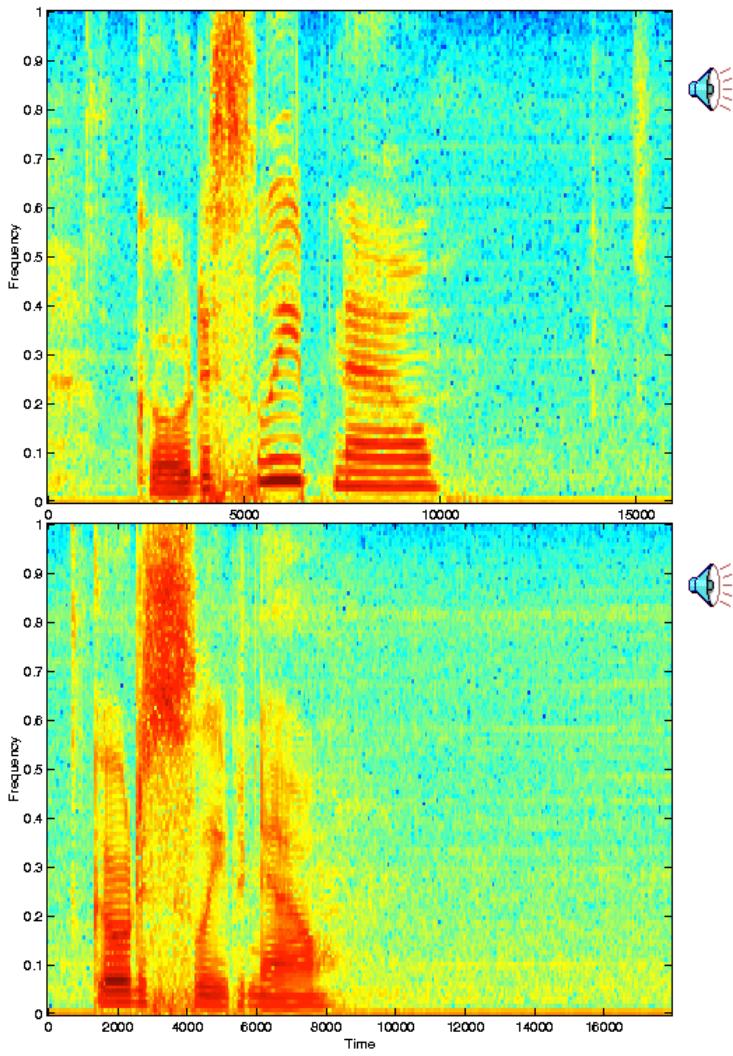


Identification

Focus of Attention



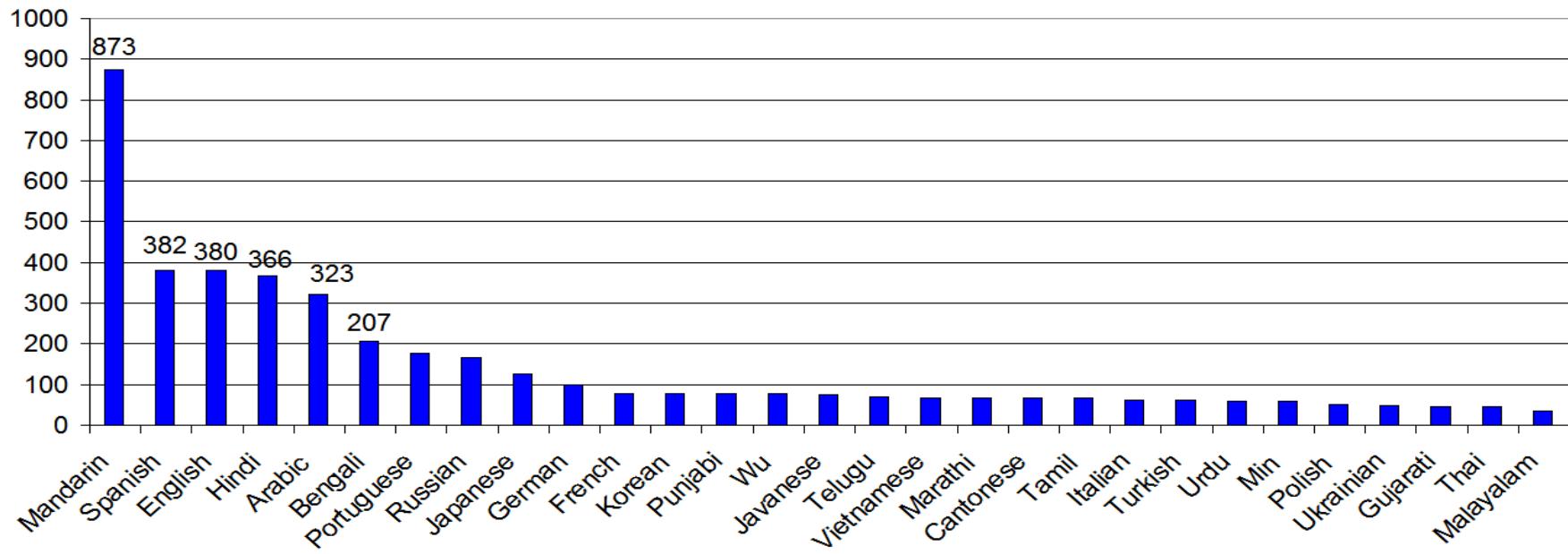




The Long Tail of Language

- Languages:
 - Only a Few Languages are Currently Addressed (~20)
 - Development of Technology Takes Long & Is Expensive
 - Cost ~.5M \$ and DevTime more than 1 Year per Language
 - 6, 000 languages, 36 M potential language pairs, Plus Dialects
 - Technology is Always a Step (or Two!) Behind Deployment

Languages by Million Native Speakers



From Babylonian
Confusion...





