

Bridging the Language Divide

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

alex@waibel.com

waibel@cs.cmu.edu

waibel@kit.edu

Languages in Europe

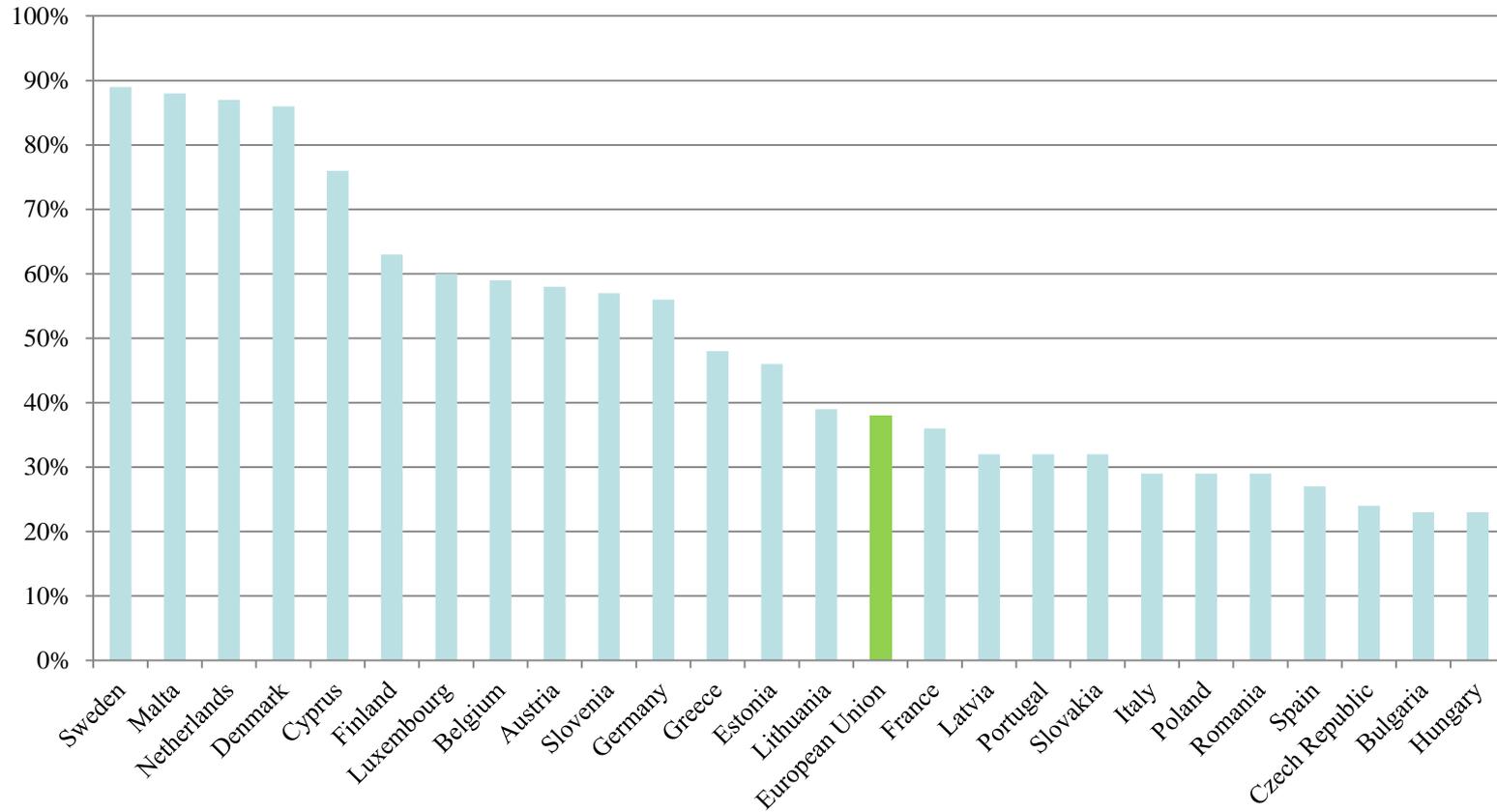


“Everyone Speaks English”... ???



In Europe:

English language knowledge (not mother tongue)







- German is the most widely-spoken first language in the EU (~100 million speakers)
- Most Germans speak at least two languages (English, French, and Russian are most common)
- Recognized minority languages:
 - Danish
 - Plattdeutsch
 - Sorbian
 - Romany
 - Frisian

Isn't Germany monolingual?



- Germany has one official language (German)
- Real life is something else:
 - Immigration
 - Tourism
 - Trade and commerce
 - Regional development, governance, and cooperation
 - Mobility and traffic
 - Energy and climate change
 - Environment and natural resources
 - Cross-border legal issues (e.g. marriage, birth, contracts)



GERMANY

A man with short dark hair and a serious expression is holding a large, hand-drawn banner. The banner is divided into two horizontal sections. The top section is filled with dense, dark scribbles. The bottom section is filled with red scribbles, and the word "GERMANY" is written across it in large, blue, block letters. The background is a blurred crowd of people, suggesting an outdoor event or protest.

Germany is the second-most popular immigration destination (after the US)

20% of residents in Germany have some roots outside Germany
6.4 million come from outside the EU



Major immigrant languages
have included:

Turkish (>2 million speakers)

Kurdish

Polish

Balkan languages

Russian

....



Major immigrant languages
have included:

Turkish (>2 million speakers)

Kurdish

Polish

Balkan languages

Russian

....





Wieder kommt das Lageso mit dem Andrang nicht klar: Am Montag warten hunderte Flüchtlinge vor der Erstaufnahmestelle. Foto: dpa

„Habibi, du wait da, guck tablet.“ Der Security-Mitarbeiter gibt sich alle Mühe, den wartenden Menschen zu erklären, was sie tun sollen.

Effective Communication is not only Text,

But:

– Speech



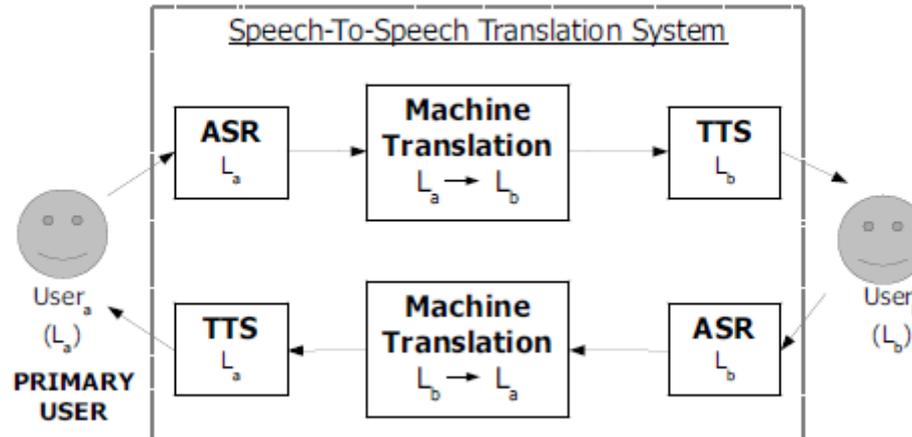
– Images



– Ill-formed Text

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

*The daunting challenge requires
innovative solutions*



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



interACT First Speech Translation VideoCall '91-92

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



First Feasibility Demo

- 1991 – First Public Demonstration of Speech
July 27, 1991 – UKA, CMU, ATR

Machine speaks three tongues

3 languages exchanged by computer

By Byron Spice
Science Editor, Post-Gazette
You speak only English. The person at the other end of the telephone conversation speaks only German.

The system has a limited vocabulary — about 400 words — and stumbles when it encounters words or sounds it does not know. Certainly no one would expect the system to carry on a conversation any time soon.

But Alex Waibel, a linguistics scientist at Carnegie Mellon University, said the system's limitations were beside the point. "You want to wait for the system," he asked, "or do you want one that handles a limited vocabulary can be easily

Named for the Roman god of beginnings, the system was designed to meet the needs of Waibel said. "conversational systems such as Janus must be registered for five years for practical meetings, order catalogs or make airline reservations."

„Janus“ als Übersetzer

Karlsruhe (dpa) - Ein Computerprogramm, das in begrenztem Umfang gesprochene Sprache erkennen, „zerlegen“ und in das Englische und Japanische übersetzen kann, hat die Karlsruher Fakultät für Informatik vorgestellt. Es wird zur Zeit über einen Wortschatz von 400 Wörtern und in zwölf Dialogen als Partner angesprochen werden.

Computer spricht drei Sprachen

Uni Karlsruhe ist Partner eines internationalen Forschungsprojekts

do. KARLSRUHE. Deutsche, amerikanische und japanische Wissenschaftler arbeiten gemeinsam an einem Übersetzungssystem, das schon in zehn Jahren die internationale Kommunikation erheblich erleichtern könnte. Dieses System „versteh“ drei Sprachen und antwortet auch in diesen Sprachen – vorerst in Deutsch, Englisch und Japanisch. Der Benutzer spricht

ligt; ein französischer Partner ist bisher noch nicht gefunden worden. Das Projekt läuft seit fünf Jahren, Ausgangspunkt war ein Auftrag des japanischen Postministeriums für ein „übersetzendes Telefon“. Andere Forschungseinrichtungen in Japan arbeiten inzwischen an ähnlichen Projekten etwa zur automatischen Untertitelung fremdsprachiger Nachrichtensendungen.

Mobile Consecutive Interpretation Technologies for Cross-Lingual Dialog

2009



JIBBIGO



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



Hola, encantado de conocerle.







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-Englis... \$4.99
- 2 European Bund... \$9.99
- 3 Chinese-Englis... \$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 with Unlimited Online Use ★ Download Offline Translations

Jibbigo LLC Web Site > Jibbigo Translator Support >

What's New in Version 2.1

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

Screenshots

iPhone iPad



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



Jibbigo:

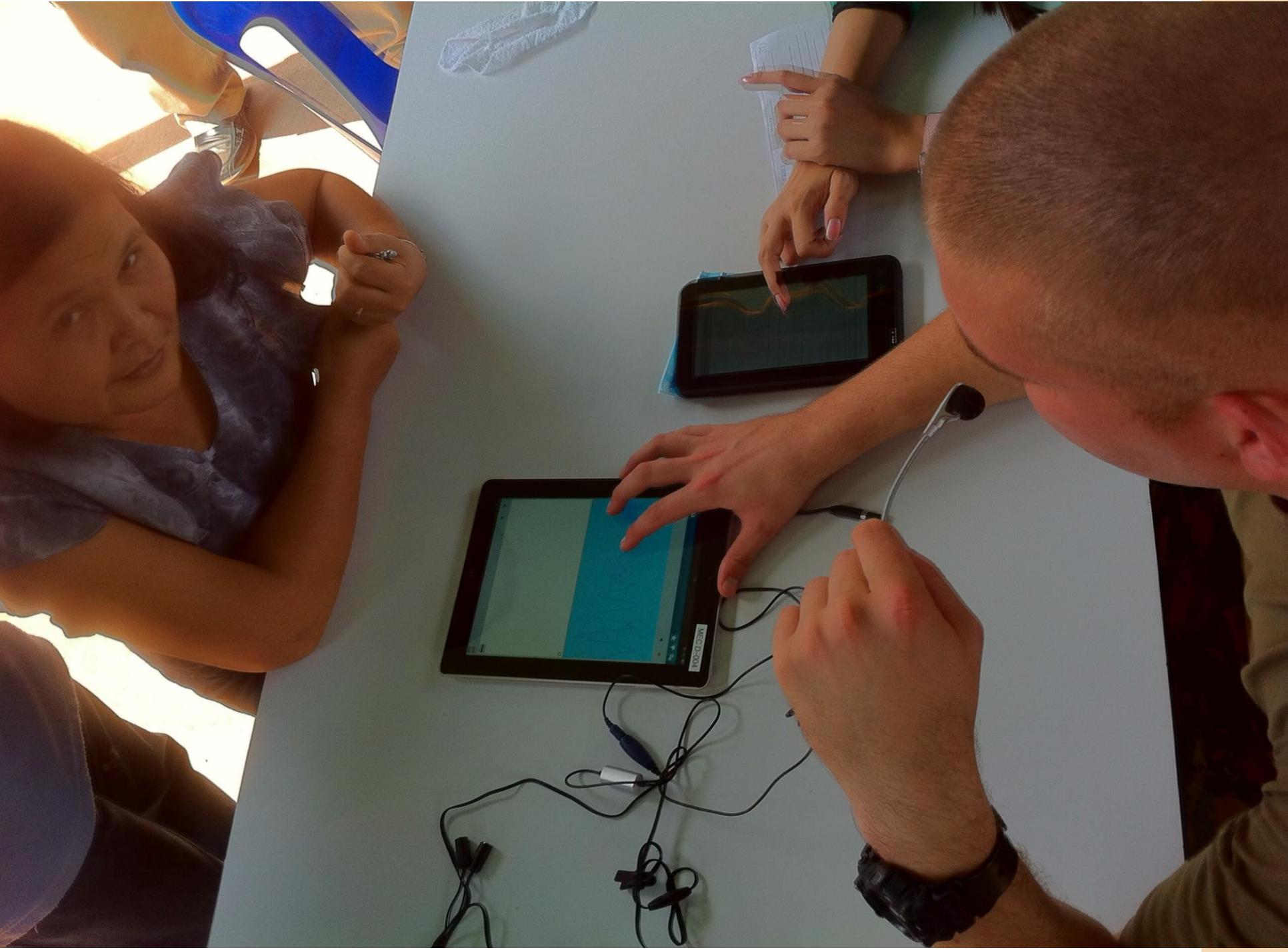
- Real-Time Translation
- 15++ languages
- 40,000 Words
- All on the Phone
- No Server Necessary!

Jibbigo on Apple Commercials



Humanitarian Deployment





Thailand









Cambodia





San Jose , Honduras



Simultaneous Interpretation

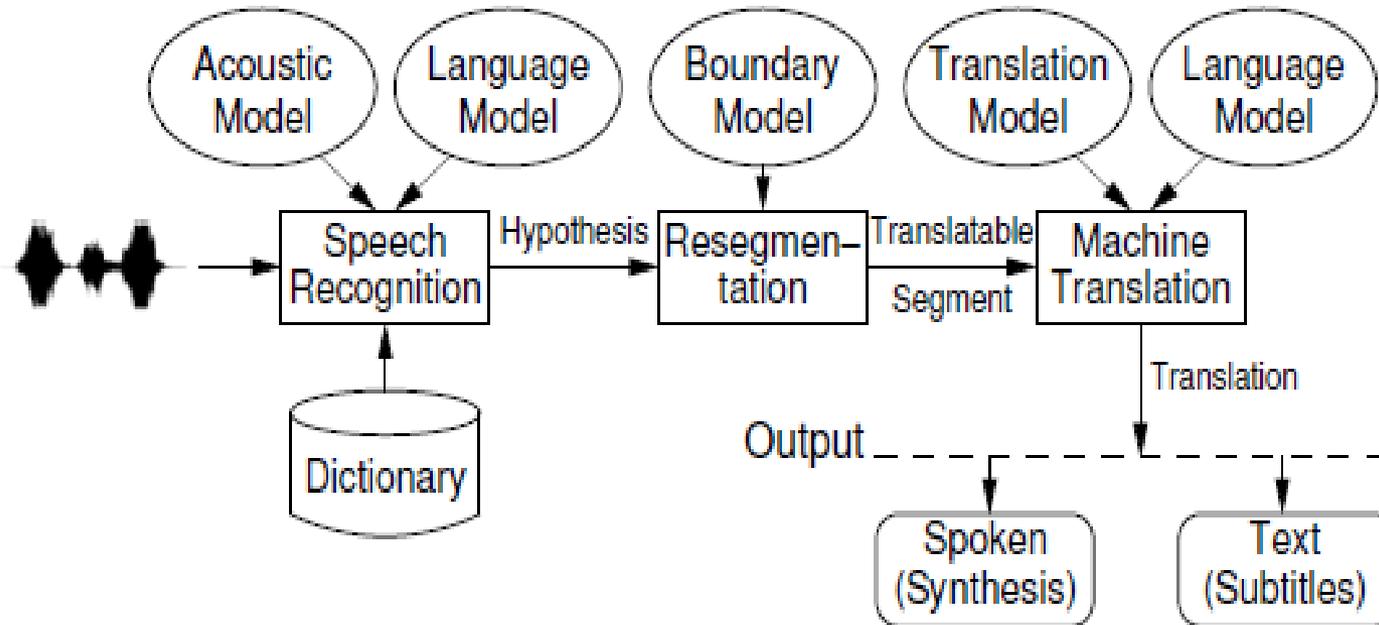
Domain Unlimited Translation of Monolingual Monologues

Domain Unlimited Translators for:

- TV/Radio Broadcast Translation
- Translation of Lectures and Speeches
- Parliamentary Speeches (UN, EU,...)
- Telephone Conversations
- Meeting Translation



你们的评估准则是什么



EU-BRIDGE – Bridges across the Language Divide



The work leading to these results has received funding from the European Union under grant agreement n° 287658



EU-BRIDGE Partners



ACCIPIO PROJECTS

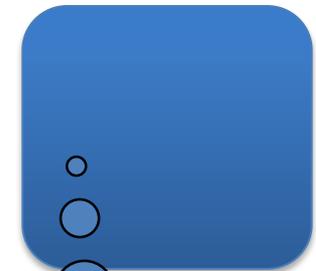
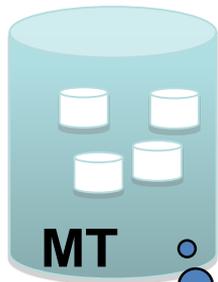
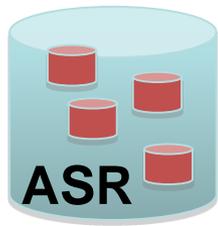


Language Service

Engines

Services

Use Cases



Use Case 1

Use Case 2

Use Case 3

Develop and Insert Improved Technology

Customization, Adaptation

Language Services for User and Developer Communities

Subtitling: BBC Weatherview



4

Subtitling & Translation: Euro-News

Euronews

Language ID + multilingual ASR + MT

8 Euronews languages

 Video Transcription [list of news]



riconquistare Tikrit l' esercito iracheno ha lanciato lunedì mattina un' operazione per riprendere la città che diede i natali a Saddam Hussein è finita a giugno sotto il controllo del sedicente Stato islamico con 30 mila uomini impegnati sul campo non solo soldati ma anche appartenenti alle forze di polizia e alle milizie sciite l' appoggio di truppe iraniane il dell' aviazione irachena e internazionale si tratta di una delle principali offensive lanciate da Baghdad e contro i gi a visti annunciata domenica dal primo ministro fa Eder ' Alaba B

Daphne

il nostro obiettivo è liberare la gente dall' oppressione dal terrorismo dell' Isil dobbiamo proteggere i cittadini e le loro proprietà

FOREIGN-SPEECH [uah uah @bg uah @bg uah nelle @bg] un tentativo di rassicurare la popolazione civile sunnita della città timorosa di ritorsioni e violenze denunciate in altre situazioni in cui sono intervenute





êß*0vúbØijBA¬pysUêÍ}hÿ5
≈fÄ<„y‡ëœkûçOF~ØΠkô#å
—
«Zeû

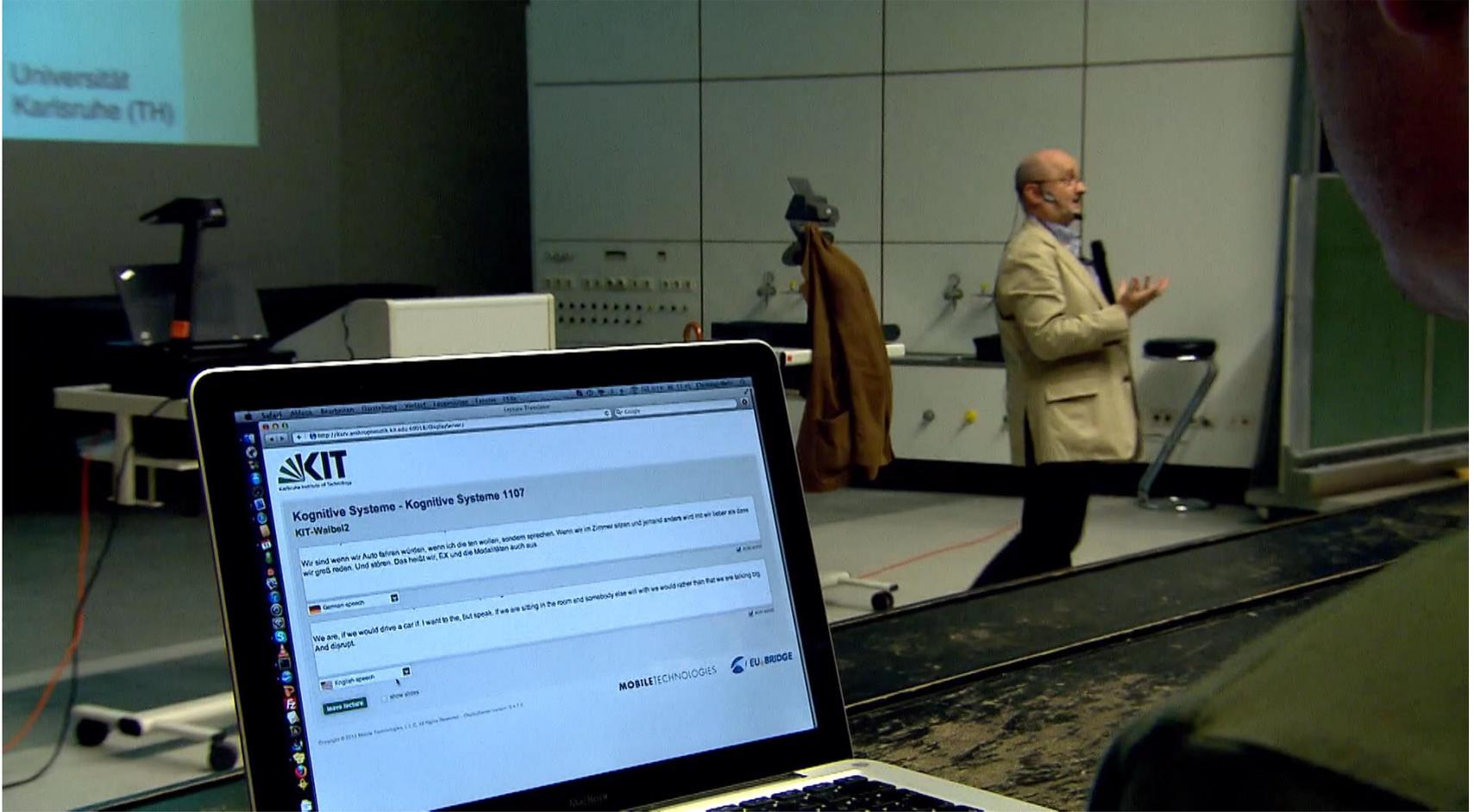
Studentin aus Tunesien

die Vorlesungen waren unverständlich



interACT

Lecture Translation



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 Inès Marta



NACHRICHTEN

Unter sich

Der Östliche Berggorilla, der in der Region der Virunga-Vulkane im Grenzgebiet zwischen der Demokratischen Republik Kongo, Ruanda und Uganda lebt, scheint Anfang der achtziger Jahre schon fast ausgestorben zu sein. Wilderei und die Zerstörung des Lebensraums hatten die Population auf 20 Exemplare schrumpfen lassen. Dank Schutzmaßnahmen hat sich diese Zahl inzwischen wieder verdoppelt. Doch es wurde befürchtet, dass das immer noch zu wenig sei, um schwere Inzuchtfolgen zu verhindern. Unter Leitung des britischen Wellcome Trust ist es nun gelungen, die Genome von sieben Berggorillas aus diesem Gebiet zu sequenzieren. Die Analyse ergab, dass die enge Verwandtschaft nicht zu genetischen Schäden geführt hat, sondern sogar nachteilige Mutationen aus dem Genpool entfernt hat. Außerdem zeigte sich, dass die Tiere offenbar schon seit Tausenden von Jahren in kleinen Gruppen überlebt haben. Es gebe keinen Grund, warum das nicht auch in Zukunft möglich sei, schreiben die Autoren in *Science*.

SPIEGEL ONLINE UNISPIEGEL

Login | Registrierung

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

Nachrichten > UnISPIEGEL > Studium > Integration > Vorlesungs-Übersetzer 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.

05.03.2015

Christoph Drösser
1 bis 32
Wissen

Jahrgang: 2015
Nummer: 10
Auflage: 631.846 (gedruckt) 510.634 (verkauft) 553.335 (verbreitet)
Reichweite: 1,62 (in Mio.)

Vochenzentung

e haben verstanden

erobern als Simultanübersetzer den und erfüllen ein großes Versprechen der Intelligenz von courtois doctus



So stellt der Künstler Jan von Holleben einen Dolmetscher-Apparat vor

von kann sich mittlerweile selbst übertragen, wer Englisch oder Spanisch spricht. Seit Dezember dürfen Nutzer des Online-Teléfono-Übersetzer-Skype, das Microsoft im Jahr 2011 gekauft hat, die Übersetzung von Englisch ins Spanische und umgekehrt testen. Im Lauf dieses Jahres sollen im Zweitsprachenwörterbuch Sprachpaare hinzukommen – auch ein Deutsch-Englisch-Dolmetscher.

Aus der Science-Fiction kennen wir das schon lange: etwa aus Douglas Adams' Roman *Der Autokrat durch die Galaxie*. Da stecken die Leute den sogenannten Rubelisch ins Ohr, der alles Fremdsprachliche in die eigene Sprache dolmetscht.

»Mir geht es gut, wie geht es dir?« wurde plötzlich zu: »Well how is it me friends?«

Der Preisvolksheld C-SPO aus Star Wars beherrschte sechs Millionen verschiedene Sprachen. Jetzt wird der jahreslange Traum Wirklichkeit, zumindest für eine breite Masse: 300 Millionen Menschen nutzen den in der Baiversion konstruierten Google-Übersetzer. Wie die marktbeherrschende Simultanübersetzer in den Alltag tritt und wie die Funktionen, die sag über ihr in Forschungsfeld aus, das wie kein zweites für die Hybrid der Intermark sehr: KI, die Künstliche Intelligenz.

Sie hat seit den fünfziger Jahren viel versprochen und wenig gehalten, auch in Sachen Digitaldolmetschen. Das Projekt: Nebenbei, in dem die Bundesregierung zwischen 1993 und 2000 umgerechnet 58 Millionen Euro verstreut, brachte außer heftiger Luft nicht viel herauf. Sein Ziel war ein elektronisches Käsechen, das zwischen zwei Sprechern als Dolmetscher fungieren sollte. Hieran kam ein rudimentäres Spracherkennungsprogramm für ein Auto, das einmündige Befehle entgegennahm. Das Scheitern der KI-Forscher löste Karrieren über intelligente Maschinen: Das schlaube Schmeißer ist für sie eher leicht, das vornehmlich Leichtes schwer. Schach spielen lernten die Computer schnell, aber erst vergangene Woche konnten Google-Forscher in Alesse vorführen, dass ihr Roboter sich simple Computerspiele wie *Space Invaders* oder *Pong* selbst beibringen konnte. Auch für die Sprachverarbeitung gilt: Komplexe wissenschaftliche Texte lassen sich leichter verstehen und übersetzen als nachgelagerte Unterhaltungen unter Fremden.

Im vergangenen Jahrzehnt aber hat die KI auf einmal auf vielen Gebieten kräftige Fortschritte gemacht. Google's Alesse gab es zuletzt durch Kalifornien, IBM Watson gewann beim Fernsehspiel.

Auch Smartphones dienen als Dolmetscher, besonders in Wildgrößen, wo Englisch als Lingua franca viele Funktionen. Eines in den Testversionen Südostasien. Dort stoßt die der Engländer nicht mündige Wirt einer kleinen Pension dem Gast karzerhand bei Sonntag-Treffen entgegen. Der kann auf Englisch insinrieren, dass er gern zwei Doppelzimmer mit Frühstück, und selbst plärrt sein Wunsch auf Koreanisch aus dem Handy-Lautesprech.

Das Scheitern der KI-Forscher löste Karrieren über intelligente Maschinen: Das schlaube Schmeißer ist für sie eher leicht, das vornehmlich Leichtes schwer. Schach spielen lernten die Computer schnell, aber erst vergangene Woche konnten Google-Forscher in Alesse vorführen, dass ihr Roboter sich simple Computerspiele wie *Space Invaders* oder *Pong* selbst beibringen konnte. Auch für die Sprachverarbeitung gilt: Komplexe wissenschaftliche Texte lassen sich leichter verstehen und übersetzen als nachgelagerte Unterhaltungen unter Fremden.

Im vergangenen Jahrzehnt aber hat die KI auf einmal auf vielen Gebieten kräftige Fortschritte gemacht. Google's Alesse gab es zuletzt durch Kalifornien, IBM Watson gewann beim Fernsehspiel. Auch Smartphones dienen als Dolmetscher, besonders in Wildgrößen, wo Englisch als Lingua franca viele Funktionen. Eines in den Testversionen Südostasien. Dort stoßt die der Engländer nicht mündige Wirt einer kleinen Pension dem Gast karzerhand bei Sonntag-Treffen entgegen. Der kann auf Englisch insinrieren, dass er gern zwei Doppelzimmer mit Frühstück, und selbst plärrt sein Wunsch auf Koreanisch aus dem Handy-Lautesprech.

Ein Geopostler übersetzen zu können, braucht eine Software drei Teilprogramme: Das erste orientiert an dem Tonklang starrer Laute und isoliert diese Buchstaben und Wörtern in. Das zweite übersetzt diese schallliche Ergebnis in Fremdsprache. Und das dritte das Ergebnis der Übersetzung von Text das in besser kein Problem mehr. Wie haben sie an Computerzimmern in Callcentern und Sit auf dem Phone großer Sprachbearbeitung und Übersetzung drängen und immer noch mit hohen Fehlerraten behaftet. Und wenn man diese zwei unvereinbaren Technologien kombiniert, dann multiplizieren sich die Fehler, sagt Wende. Anders ausgedrückt: Was in der Ausgangssprache kein sinnvolles Satz eskaliert, kann man von der Übersetzung keine Wunder erwarten.

Ein Geopostler übersetzen zu können, braucht eine Software drei Teilprogramme: Das erste orientiert an dem Tonklang starrer Laute und isoliert diese Buchstaben und Wörtern in. Das zweite übersetzt diese schallliche Ergebnis in Fremdsprache. Und das dritte das Ergebnis der Übersetzung von Text das in besser kein Problem mehr. Wie haben sie an Computerzimmern in Callcentern und Sit auf dem Phone großer Sprachbearbeitung und Übersetzung drängen und immer noch mit hohen Fehlerraten behaftet. Und wenn man diese zwei unvereinbaren Technologien kombiniert, dann multiplizieren sich die Fehler, sagt Wende. Anders ausgedrückt: Was in der Ausgangssprache kein sinnvolles Satz eskaliert, kann man von der Übersetzung keine Wunder erwarten.

Ein Geopostler übersetzen zu können, braucht eine Software drei Teilprogramme: Das erste orientiert an dem Tonklang starrer Laute und isoliert diese Buchstaben und Wörtern in. Das zweite übersetzt diese schallliche Ergebnis in Fremdsprache. Und das dritte das Ergebnis der Übersetzung von Text das in besser kein Problem mehr. Wie haben sie an Computerzimmern in Callcentern und Sit auf dem Phone großer Sprachbearbeitung und Übersetzung drängen und immer noch mit hohen Fehlerraten behaftet. Und wenn man diese zwei unvereinbaren Technologien kombiniert, dann multiplizieren sich die Fehler, sagt Wende. Anders ausgedrückt: Was in der Ausgangssprache kein sinnvolles Satz eskaliert, kann man von der Übersetzung keine Wunder erwarten.

Ein Geopostler übersetzen zu können, braucht eine Software drei Teilprogramme: Das erste orientiert an dem Tonklang starrer Laute und isoliert diese Buchstaben und Wörtern in. Das zweite übersetzt diese schallliche Ergebnis in Fremdsprache. Und das dritte das Ergebnis der Übersetzung von Text das in besser kein Problem mehr. Wie haben sie an Computerzimmern in Callcentern und Sit auf dem Phone großer Sprachbearbeitung und Übersetzung drängen und immer noch mit hohen Fehlerraten behaftet. Und wenn man diese zwei unvereinbaren Technologien kombiniert, dann multiplizieren sich die Fehler, sagt Wende. Anders ausgedrückt: Was in der Ausgangssprache kein sinnvolles Satz eskaliert, kann man von der Übersetzung keine Wunder erwarten.

Fortssetzung auf S. 32

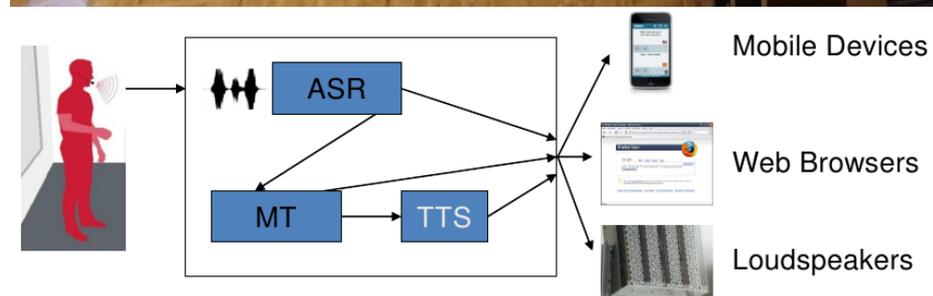


interACT

Lecture Transcription/Translation at KIT



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





English (en)

French (fr) 

German (de)

English (en) 



- Translation of Power Point Slides
- Presentation by Sub-Titles

Minimalwert

 Fakultät für Informatik – Institut für Anthropomatik
Prof. Dr. U. Hanebeck, Dr.-Ing. T. Asfour 6-8

3.1 Formale Grundlagen

Zur Untersuchung und Beschreibung der Eigenschaften und des Verhaltens von logischen Funktionen ist die **Boolesche Algebra** hervorragend geeignet.

Entwickelt wurde sie von dem Mathematiker **George Boole** (1815 –1864) als Algebra der Logik.

3.1 formal basics
For the study and description of logic functions and the behavior of digital systems, which is excellent, the Boolean algebra is well suited.
It was developed by the mathematician George Boole (1815 1864) as algebra







- Demonstrating automatic real-time lecture interpretation
- University Presidents; Interpretation Training & Services
- Promising but Controversial





Human-Machine Symbiosis

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

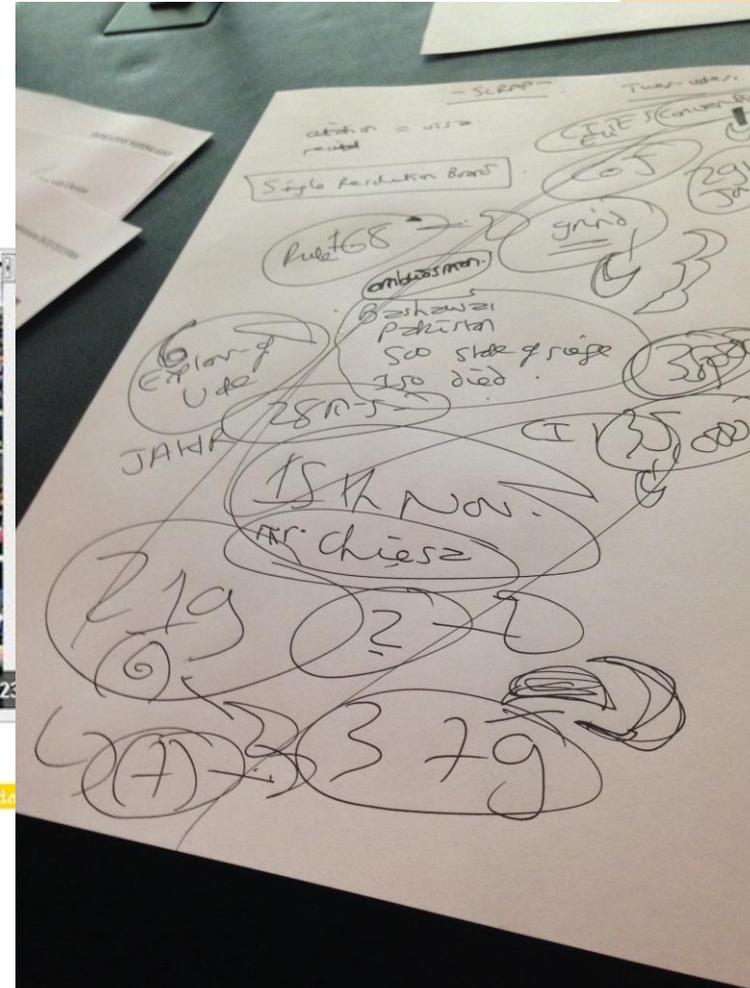
23:07 - 23:08
and that concludes that

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

23:20 - 23:22
K paragraph three

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

23:26 - 23:27



Observations:

Interpreting Voting Sessions is...

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





Language is Ambiguous at All Levels:

- Semantics:
 - The Spirit is Willing but the Flesh is Weak
 - → The Vodka is Good but the Meat is Rotten
- Syntax:
 - Time Flies Like an Arrow → 6 Different Parses
- Phonetics:
 - This Machine Can Recognize Speech →
This Machine Can Wrack a Nice Beach
 - Give me a New Display → Give me a Nudist Play



- **German** has some particularly difficult peculiarities:
 - Wordorder:
Ich schlage Ihnen einen Termin für nächste Woche in meinem Büro am Adenauerring in Karlsruhe, in dem vor.
→ *I propose [hit?]* a meeting for next week at my office in Karlsruhe on the Adenauerring...
 - Inflections and Agreement:
Zu der nächsten wichtigen interessanten Vorlesung
 - Compounds:
Worterkennungsfehlerrate
→ Word Recognition Error Rate

Die Fehlerstromschutzschalterprüfung

Die Wirtschaftsdelegationsmitglieder

Die Bankwirtschaftsfreigabeerklärung

Die Lehrverpflichtungserklärungen

Die Schiffskommunalschuldverschreibungen

Die Vorkaufsrechtverzichtserklärung

Das Mehrzweckkirschentkengerät

Die Gemeindegrundsteueranlagung

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

„Ich freue mich, dass Sie heute so zahlreich....“

→ you, she, they ?

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

→ es, sie ?



- Technical Terms & Special Usage
 - epstral-Koeffizienten, Wälzlagerungen → Roller Bearings
 - Klausur → Final Exam (not Retreat), Vorzeichen → Sign (not Omen)
- Formulas:
 - Eff von I_x → $f(x)$
- Foreign Words in a German Lecture
 - Computer Science- English Expressions
 - “Cloud”, “iPhone”, “iPad”, “Laser”
- Declinations and Compounding incl. foreign Words
 - Web-ge-casted, down-ge-loaded
 - Cloudbasierter Webcastzugriff



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

– **Ambiguity:**

Scores, Statistics, Neural Activations, ..

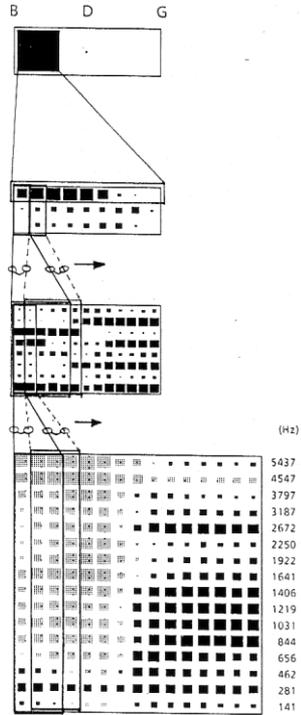
– **Learning:**

Build Models, Extract Knowledge from
Human Data & Interaction, Automatically

→ Performance Depends on Data & Computing

Neural Nets: Bigger, Deeper, Faster

(1987)



TDNN: Shift-Invariance, Waibel '87

(1989)

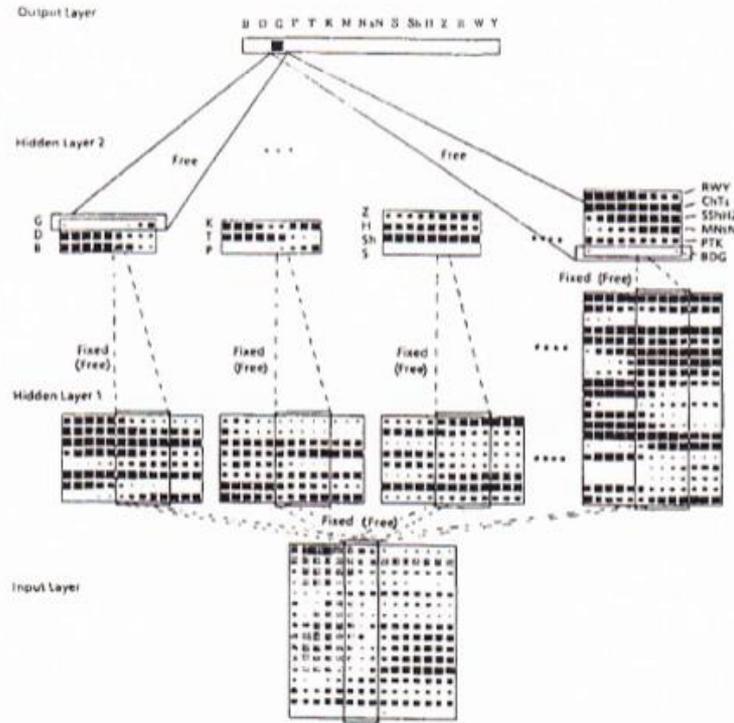
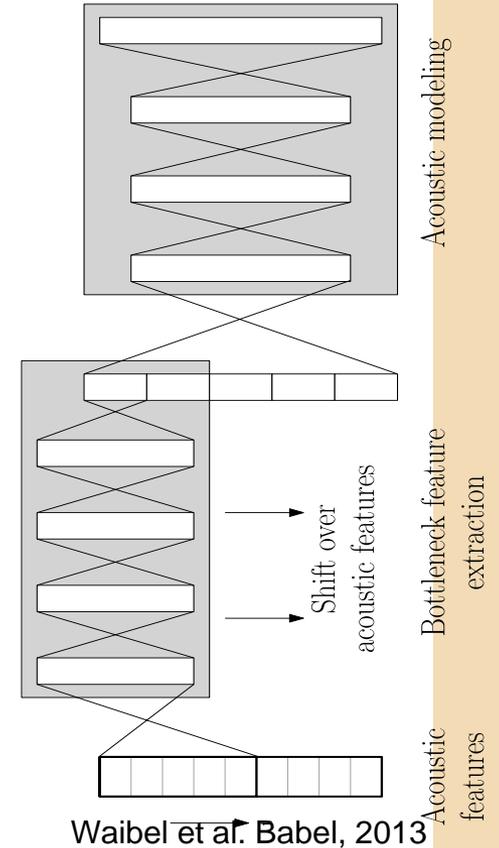


Fig. 7. Modular construction of an all consonant network.

Modular (deep) TDNN: Waibel '87

(2013)



Waibel et al. Babel, 2013

Weights: ~6,000

TrnData[hrs]: ~0.1

Time[weeks] ~1

~40,000

~1

~1

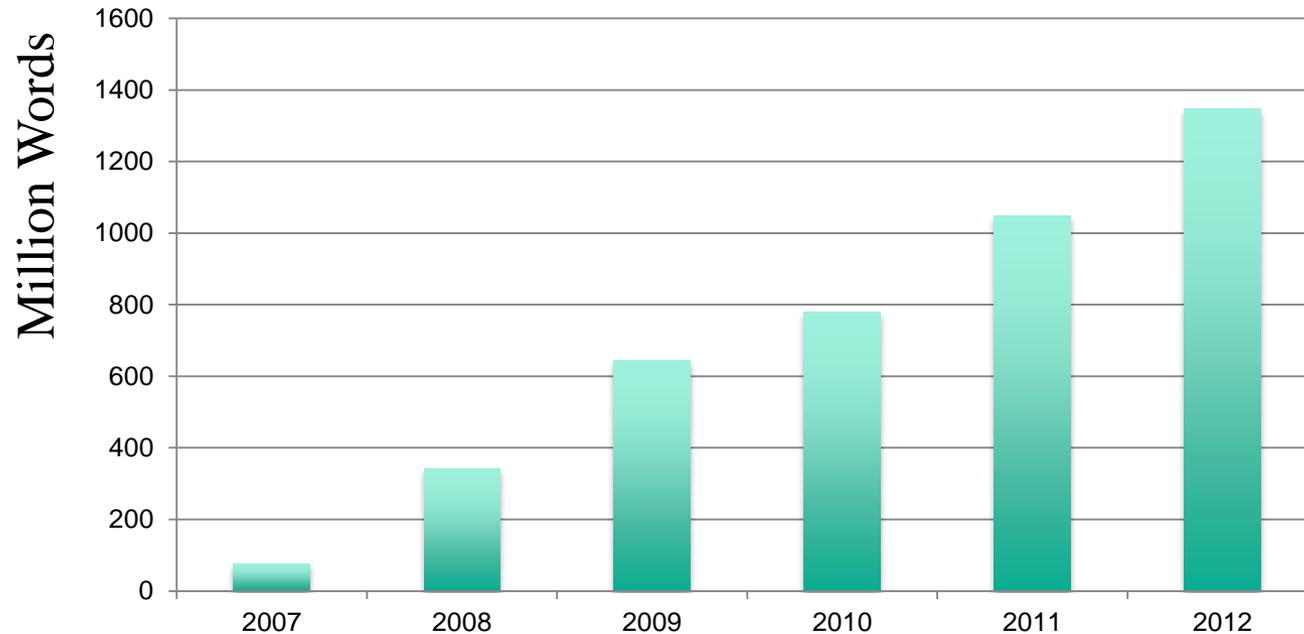
~33,000,000

~1,000

~1



News Shuffle Size



- Computer MT or ASR systems train on \gg 1GWords
 - News Shuffle, GigaWord, Europarl, VideoLectures, ...
- Human speaks 0.5 GigaWords in a Lifetime!!



interACT The Data Challenge

- Machine Learning + Massive Data
Lead to Better Performance
- Is the problem too hard? Is it too easy?
Already done? Google Translate?
- Effective Language Solutions
 - Not only from/to English, but from/to German, ...
 - Minority Languages and Regional Dialects
 - Need targeted solutions in domain/application
 - Privacy and Security
 - Dissemination, not only Assimilation
- European Language Solutions
 - Language (and technology) must be cultivated and treasured
 - Data Volume and Access Key Challenge

Communication between the people of the world

- Bridging the Linguistic Divide
- Technology can already make helpful contributions
- Methods: Machine Learning from Data
Adaptation, Error Recovery, Learning, Forgetting
- User Interaction, Appropriate Interfaces
- More Data, more Robust Performance
- Better Language Portability
- Integration into Services

