Spoken Language Translation:

Interactivity as the Key to Real-world Development

November, 2003

overview

- Automatic speech-to-speech translation: an age-old dream
- *Practical* Spoken Language Translation (SLT) systems are now possible
  - … *but* users must cooperate and compromise
- Past, present, …
  - technologies
  - research
  - commercial
- … and future
  - challenges
  - opportunities

Star Trek? *Not!*

- The dream: speak as usual
  - freely shift topics
  - full range of vocabulary, idioms, structures
  - spontaneous language: fragments, false starts, hesitations
  - mumble
  - converse in noisy environments
  - ignore the translation program
- Needed: realistic expectations!

scientific/technical issues:

component integration

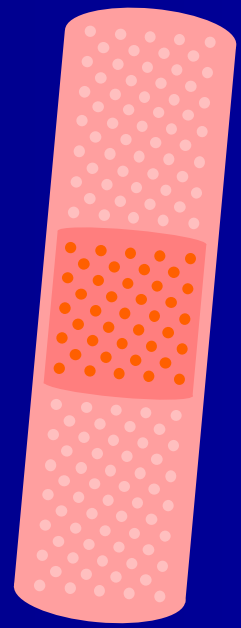
- Component technologies imperfect, hard to integrate
  - Speech recognition
  - Machine translation
  - Text-to-speech
- SR, MT both introduce ambiguities
- Usable separately, but error rates combine, compound
- Spontaneous speech
  - hesitations, fragments, repetitions
  - Even perfect SR would give noisy MT input!

solutions: med/long term

- For ambiguity:
  - integrate multiple knowledge sources
    - phonological, prosodic, morphological, syntactic, semantic, discourse, domain ...
- For syntactic noise:
  - robust parsing
    - ignore noisy parts
    - extract important stuff
    - patch together fragments

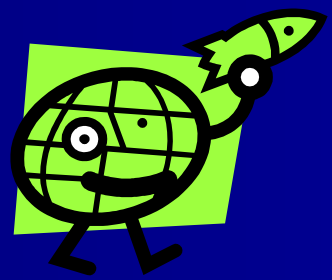
solutions: shorter term
user cooperation

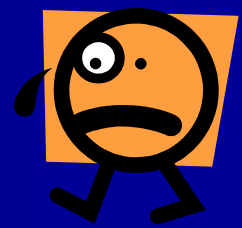
- Speak loudly and clearly, in quiet environments
- Restrict domain, e.g. hotel reservations
- Restrict use of audio input equipment, networks, etc.
- Correct speech recognition errors, by voice or by typing
- Train speaker-dependent acoustic models
- Provide only well-formed input
- Provide extra information in the input to aid analysis, e.g. word separations or brackets
- Resolve lexical or structural ambiguities
- Provide missing information, e.g. references for zero pronouns
- Tolerate rough or incomplete translations
- Spell or type out words that prove hard to recognize
- Use richer or more complex interfaces, e.g. including GUIs as opposed to voice-only

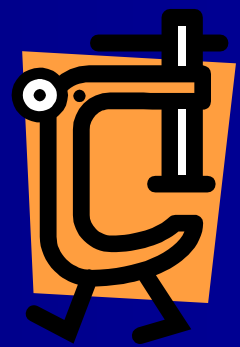


3 classes of cooperative SLT

- Class One: voice-driven phrase book
- Class Two: robust speech translation within very narrow domains
- Class Three: highly interactive speech translation with broad linguistic and topical coverage







class one: voice-driven phrase book

- **Coverage vs. cooperation**
  - *linguistic coverage: narrow*
  - *topical coverage: narrow*
  - *cooperation required: low*
- **Technology**
  - Speech recognition: built-in or IVR
  - MT: flat lookup, template or example-based
  - Engineering exercise: low risk
- **Fixed expressions or templates only**
  - "It's a pleasure to meet you." "Show me where it hurts." "Put your hands up!"
- **Advantages for user**
  - no need to carry a phrasebook
  - selection of phrase by voice – alternative to mouse, finger
  - translation output pronounced by native
- **PDA**
  - Phraselator - military
  - NEC Travel Interpreter - tourist
- **Phone (IVR)**
  - Interface: tree descent using VXML
  - Retain dialog info for responses

Phraselator
by VoxTec

A division of Marine Acoustics, Inc.

class two: robust SLT within narrow domains

- **Coverage vs. cooperation**
  - *linguistic coverage: broad*
  - *topical coverage: narrow*
  - *cooperation required: medium*
- **Examples**
  - *Uh, could I reserve a double room for next Tuesday, please?*
  - *I need to, um, I need a double room please. That's for next Tuesday.*
  - *Hello, I'm calling about reserving a room. I'd be arriving next week on Tuesday.*
- **Robust analysis (like info extraction)**
- **Advantages**
  - Lots of experience
  - Can optimize SR, MT: special grammars (patterns)
  - Interlingua possible for MT
- **Challenges**
  - Interactive SR unusual, so MT input is dirty
  - Robust parsing still imperfect
  - Some user frustration inevitable, but balanced by freedom
  - Risk: medium

class two: representative research

SYSTEM	DEVELOPER	TIME	DOMAINS	LANGUAGES	MT	VOCAB
Head Transducers	AT&T Labs (USA)	1996	Travel information accessing	English-Chinese / English-Spanish	Statistical	1200/1300
JANUS-III	CMU (USA)	1997-	Hotel reservation flight / train ticket booking etc.	English-German, Japanese, Spanish, etc.	Multi-engine	open
ATR-MATRIX	ATR-SLT (Japan)	1998-2001	Hotel reservation	Japanese-English German etc.	Pattern-based	2000
Verbmobil	Univ. of Karlsruhe, DFKI etc. (Ger.)	1993-2000	Meeting appointment	German, English, Japanese	Multi-engine	10000/2500
Lodestar	CAS-NLPR (China)	1999	Hotel reservation, travel information accessing	Chinese-Japanese, English	Multi-engine	2000

class three:
highly interactive, broad coverage, speech translation

- Coverage vs. cooperation
  - *linguistic coverage: broad*
  - *topical coverage: broad*
  - *cooperation required: extensive*
- User pays for broad coverage by supervising
- SR: dictation for broad coverage
  - Philips Speech
- MT: broad coverage, good quality
  - modifiable for interactive correction
  - Linguatec, Word Magic
- Challenges
  - All components should be server-based: SR is hard
  - Interactive translation: SELECT database
  - Burden of interaction: need usability testing
  - Risk: medium to high

Spoken's innovations

Interactive correction of …

- translation
  - SELECT™ technology to gather, align Meaning Cues™ (definitions, examples, …)
  - users can select intended word meanings
  - supports meaningful back-translations
- dictation
  - first real-time correction of server-based dictation (Speech Cues™)

Commercialization of highly interactive, broad coverage SLT system (Class Three)

broad coverage …

French: Qu'est-ce que vous étudiez?
(What do you study?)
English: Computer science.
(L'informatique.)

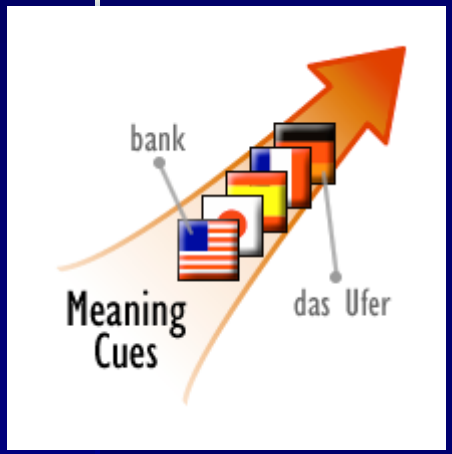
French: Qu'est-ce que vous faites plus tard?
(What are you doing later?)
English: I'm going skiing.
(Je vais faire du ski.)
French: Vous n'avez pas besoin de travailler?
(You don't need to work?)
English: I'll take my computer with me.
(Je prendrai mon ordinateur avec moi.)
French: Où est-ce que vous mettrez l'ordinateur pendant que vous skiez?
(Where will you put the computer while you ski?)
English: In my pocket.
(Dans ma poche.)

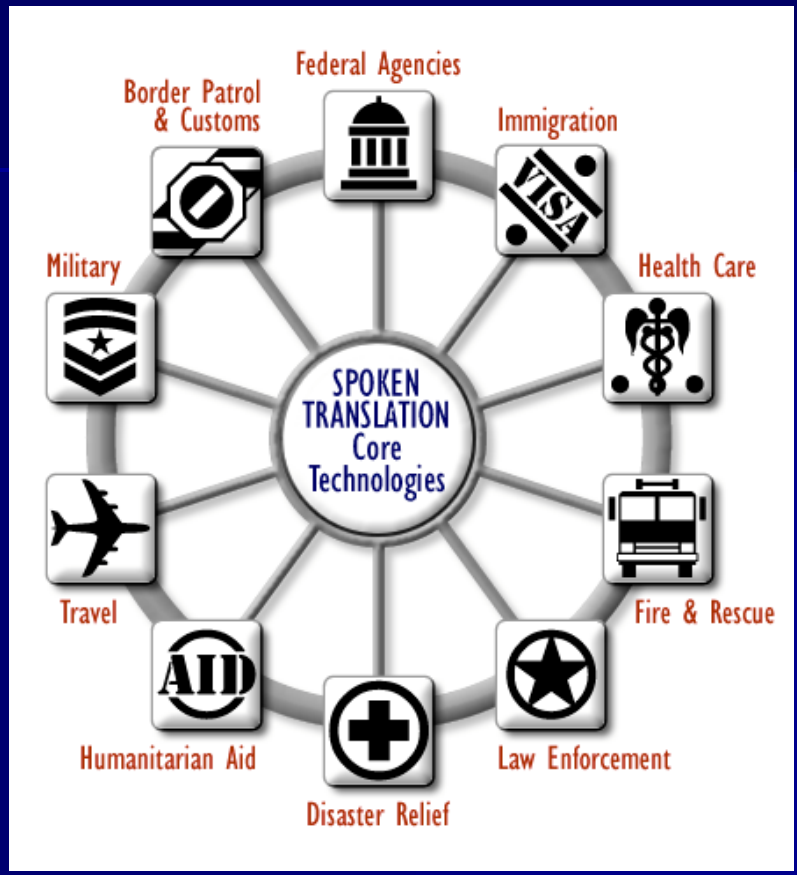
early adopter markets

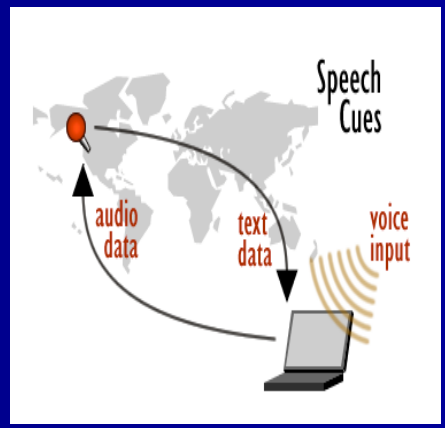
- Class one (fixed phrase translation)
  - Military
  - Travel
- Class two (robust in narrow domain)
  - Hotel reservations
  - Health care
- Class three (highly interactive broad coverage)
  - Health care
  - Military, security, intelligence
  - Customer service for technology companies
  - B2B, intra-business in high tech sector
  - Family/social communications
- All three?
  - border patrol & customs; disaster relief; federal agencies; fire & rescue; humanitarian aid; immigration; law enforcement

about family communications …

the Spoken Wheel







www.spokentranslation.com

mark.seligman@spokentranslation.com