C SC 620
Advanced Topics in Natural Language Processing

Lecture 22

4/15

Reading List

- *Readings in Machine Translation*, Eds. Nirenburg, S. *et al*. MIT Press 2003.
  - 19. Montague Grammar and Machine Translation. Landsbergen, J.
  - 20. Dialogue Translation vs. Text Translation – Interpretation Based Approach. Tsujii, J.-I. And M. Nagao
  - 21. Translation by Structural Correspondences. Kaplan, R. et al.
  - **22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.**
  - 31. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. Nagao, M.
  - 32. A Statistical Approach to Machine Translation. Brown, P. F. et al.

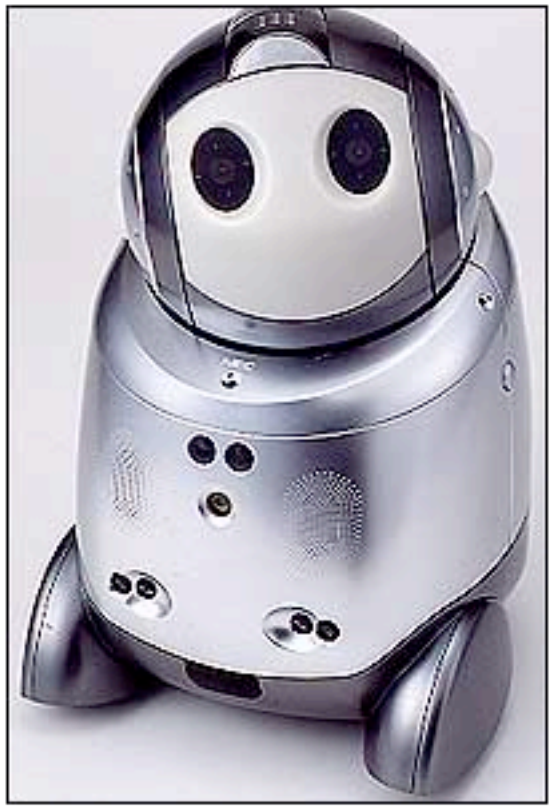
BBC NEWS UK EDITION

Last Updated: Thursday, 4 March, 2004, 09:15 GMT

Robo-talk helps pocket translator

By Jo Twist
BBC News Online technology reporter

Small robots with friendly faces have helped out in the development of handheld translation gadgets to be tried out by travellers in Japan.



Papero has lent its translation ability to tourists

Visitors landing at Tokyo's Narita Airport will be able to hire a device which can translate the local lingo.

The speech-to-speech technology was developed by NEC, tested in Papero robots and then put in PDAs.

Papero is the first all-hearing, all-seeing robot to be able to talk in conversational colloquialisms.

The PDA hire scheme is part of a wider project, e-Airport, to make Japan's main international airport the most hi-tech in the world.

Lend me your brain?

As well as being able to understand and imitate human behaviour, Papero (Partner-Type Personal Robot), is the first robot to translate verbally between two languages in colloquial tongue.

It can cope, in other words, with slang and local chatter, and has a vocabulary of 50,000 Japanese and 25,000 English travel and tourism related words.

After Papero demonstrated its translation ability, the PDAs borrowed its brain and tongue. Users can talk into the device and it will talk back in almost-perfect Japanese in a second.

The e-Navi is meant to be a travelling companion for tourists

It has voice recognition, digital voice translation and a voice synthesiser to talk to users, explained Chris Shimizu, NEC's corporate relations manager, and the quality of the voice spoken back to users is much more human than robotic.

The devices also serve as mobile phones, and have airport and local guides, as well as unlimited wireless net access.

FUNCTIONS

∴ PDA

MAIN MENU

VOICE TRANSLATOR

The following functions are provided in English

English-Japanese Voice Translator

This function provides handy phrases for traveling, translating from American English to Japanese and from Japanese to American English.

Voice sample

English
Japanese
*Voice samples are in Windows Media format.
Click here to download Windows Media Player 9 Series FREE

PDA Phone (outward domestic calls only)

The PDAs are also equipped with a telephone function. However, this cannot be used for international calls or for incoming calls. The total connection time that a participant can spend during the participation period will be limited. (Click here for further details).

NARITA Airport Guide

This is a guide to the facilities at Narita Airport, hotels in the immediate area, local tourism, traffic and transport information, currency exchange and flight information from the Narita Airport website in a searchable format.

- Similar to the Phraselator

You are in: **Sci/Tech**

Tuesday, 12 February, 2002, 09:16 GMT

Languages in the palm of your hand

The UT-103 translator is small and light

By BBC News Online's Alfred Hermida

If you travel a lot but languages are not your strong point, then a Russian company might just have the answer.

It has developed a handy pocket-sized gadget that translates English phrases into French, German or Spanish.

Instead of thumbing through a dictionary, you just say a phrase in English, the device translates it and then repeats out loud in the foreign language in a robotic voice.

"This is the first translator in the world that understands voice and it was primarily designed for travellers," said Arkady Davydov of Ectaco which developed the product.

"It is more than an electronic phrasebook because it recognises any phrase you say.

"In the future we will have models for all the other languages," he told the BBC programme, Go Digital.

The next steps are adding more languages, including English to Chinese by the end of the year.

"Two speakers, English and Chinese, will be able to communicate live without having to use the phrasebook or dictionary," said Mr Davydov. "It is going to be really amazing."

Accurate 90% of the time

The Universal Translator UT-103 was developed by the Ectaco company based in St Petersburg, Russia.

The device fits in a pocket with ease. It uses AA batteries and costs $249.95, which could pay for a few bulky paper dictionaries.

The specially-developed speech recognition software allows it to recognise and translate 3,000 phrases commonly used in all kinds of travel.

They include categories such as eating, shopping and driving.

Davydov: Designed for travellers in mind

During a
demonstration, the translator did make a few mistakes, failing to recognise some phrases.

It translated a question about a charge for extra luggage as "when is the train going to depart".

"Unfortunately there is a lot of background noise here," explained Mr Davydov. "Usually it works 90% of the time."

Specialist uses

Ectaco say the translator will understand what you say, regardless of your accent.

Developing the device, they recorded more than 700 native English and foreign speakers to create a phonetic bank of all recorded phrases.

BBC NEWS UK EDITION

Last Updated: Monday, 13 October, 2003, 13:43 GMT 14:43 UK

Getting lost in the translation

Dot.life - where technology meets life, every Monday

By Brendan O'Neill

Relying on online translation tools can be a risky business, especially if you expect too much of it. For the time being, might translation be something best left to the humans?

Earlier this month the small German town of Homberg-an-der-Efze, north of Frankfurt, had to pulp an entire print run of its English-language tourism brochure - after officials used an internet translating tool to translate the German text.

Not everyone can have a human translator on hand

According to one report, the brochure was "rendered meaningless" by the online tool. Tourists were promised "casual value", the literal translation of the German word for "leisure potential", at venues such as the "free bath" - better known as an "open-air swimming pool".

Martin Wagner, mayor of Homberg-an-der-Efze, admits that the town made a "blunder". As a result of officials trying to save money by getting the internet to do a translator's job, a total of 7500 brochures had to be binned.

Why is foreign text "rendered meaningless" in this way, when passed through an online translation tool? According to Sabine Reul, who runs the Frankfurt-based translation company Textburo Reul, translation tools have limited uses - and problems arise when web users expect too much from them.

"A translation tool works for some things," says Reul. "Say a British company wants to order a box of screws from a German supplier. A sentence like 'We need one box of a certain type of screw' is something that a machine could translate reasonably accurately - though primitively."

Using the internet may be a lot quicker than "human input"

Yet when it comes to translating blocks of text - words and sentences that convey thoughts and sentiments - online tools are bound to fail, she adds. "Beyond simple sentences, the online process simply doesn't work because machines don't understand grammar and semantics, never mind idiom and style."

"Language is not a system of signs in the mechanical sense of the word", says Reul. "It is a living medium that is used to convey thought. And that is where machines fail. Human input is indispensable as long as computers cannot think."

Reul and other translators look forward to the day when clever computers might help to ease their workload - but that time has not arrived yet.

"It would be nice if computers could do the job. And certainly the quest for machine translation has prompted a lot of linguistic research that may prove valuable in unforeseen ways. But experience to date confirms that even the most subtle computer program doesn't think - and you need to be able to think in order to translate."

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- Time: 90s
- Introduction: Why is the Pivot Approach Not Universally Used?
  - Pivot (interlingua): O(n) parsers/analyzers
  - Transfer: $O(n^2)$ parsers/analyzers
  - n = number of languages
  - Pivot dictionaries: monolingual
  - Transfer dictionaries: bilingual

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- Pure Pivot Approaches
  - Independent pivot lexicon
  - Universal notation for determination, quantification, actualization (time/modality/aspect), thematization, etc.
- I.1 Pure Pivot Lexicons are Challenging …
- 1.1 … But Specific of a Domain (Interpretation Language)
  - May be possible to define a completely artificial language for a fixed and restricted domain
  - TITUS system: textile domain
- 1.2 … Or Specific of a Language Group (Standard Language)
  - Standard Language: e.g. English
    - Double translations for all pairs of languages not containing the pivot
    - No implementation known
    - “Idiosyncratic gap” between language families

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- 1.2 … Or Specific of a Language Group (Standard Language)
  - Artificial Language: e.g. Esperanto
    - BSO project
    - Double translations for all pairs of languages
    - Lack of sufficient technical vocabulary
      - need about 50,000 terms in any typical technical domain
      - Esperanto too small
    - "Idiosyncratic gap" still exists
      - Esperanto borrows from several language families
      - but unavoidable that many distinctions and ways of expression are left out
      - mur (French) - wall
      - muro (Italian, seen from outside), parete (seen from inside)

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- 1.3 … And Always Very Difficult to Construct (Conceptual Decomposition/Enumeration)
  - Define small number of conceptual primitives and decompose all lexical items in terms of them
  - Conceptual dependency graphs will be huge
  - Use “subroutines” - conceptual enumeration
  - Japanese CICC project: 250,000 concepts
  - Construction process is non-montonic
    - new concept, revise dictionary for all languages
  - Difficult to see if concept already exists if its name is difficult to guess
    - “pros and cons” translated into another language

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- I.2 Pure Pivot Structure Loses Information …
  - Extremely rare that two different terms or constructions of a language are completely synonymous
  - Unavoidable information useful for quality translation will be lost
- 2.1 … At the Lexical Level
  - wall -> wall seen from outside -> muro
  - wall (seen from outside) -> ???
  - muro -> wall
  - parete -> wall (distinction lost)
- 2.2 … At the Lower Interpretation Levels (Style)
  - One obtains paraphrases
    - Impossible to parallel styles as all trace of the source expression is lost
- 2.3 … At Non-Universal Grammatical Levels
  - “All or nothing” problem

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- II. Transfer Approaches
  - Avoid Pivot difficulties
  - 1 -> many or many -> 1 situations
- II.1 The Hybrid Approaches May Be Worse, Because the Square Problem Remains …
  - Lexical language-specific
  - Grammatical and relational symbols are universal
  - Big transfer dictionary needed
- 1.1 … If the Lexicons are Only Monolingual (CETA)
  - Grenoble group (CETA)
  - Hybrid pivot approach

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- 1.2 … And Even If Some Part Becomes Universal (EUROTRA)
  - EUROTRA (1983)
  - 9 languages
  - linguistic development scattered across 11 countries
  - transfer approach
  - part number approach for technical terms
- II.2 Transfer Architectures Using *m*-Structures
  - Sequential or
  - Integrated approach using a multilevel structural descriptor
- 2.1 … Allow to Reach a Higher Quality
  - no universal notation for tense/aspect/modality
  - source language specific
- 2.2 … May be Preferable in 1->m Situations
  - Big firms - documentation produced in one language

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- III. Both Approaches for the Future?
- III.1 Pivot
- 1.1 Domain-Specific Pivots: New Applications?
  - CAD/CAM and expert systems: generation from knowledge base
- 1.2 Conceptual Decomposition/Enumeration a Challenge
  - EDR
  - Multilingual conceptual database (EuroWordNet?)

Paper 22. Pros and Cons of the Pivot and Transfer Approaches in Multilingual Machine Translation. Boitet, C.

- III.2 Transfer
- 2.1 Conversion from First to Second Generation
  - SYSTRAN (used in babelfish.altavista)
  - 1G to 2G (?), see comments on CETA (pg.276)
  - Concepts dictionaries
- 2.2 Composition in n<->n Situations: The Structured Language Approach
  - Relay translation
    - 4 Romance languages
    - 4 Germanic languages
    - Greek