Machine Translation Systems

Gerald Penn

CS224N / Ling 284

[Based on slides by Kevin Knight, Dan Klein, Dan Jurafsky and Chris Manning]

MT Evaluation

(left over to 2011/01/24)

Illustrative translation results

- *la politique de la haine .* *(Foreign Original)*
- politics of hate . (Reference Translation)
- the policy of the hatred . (IBM4+N-grams+Stack)

- *nous avons signé le protocole .* *(Foreign Original)*
- we did sign the memorandum of agreement . (Reference Translation)
- we have signed the protocol . (IBM4+N-grams+Stack)

- *où était le plan solide ?* *(Foreign Original)*
- but where was the solid plan ? (Reference Translation)
- where was the economic base ? (IBM4+N-grams+Stack)

对外经济贸易合作部今天提供的数据表明，今年至十一月中国实际利用外资四百六十九点五九亿美元，其中包括外商直接投资四百点零七亿美元。

the Ministry of Foreign Trade and Economic Cooperation, including foreign
direct investment 40.007 billion US dollars today provide data include
that year to November china actually using foreign 46.959 billion US dollars and

MT Evaluation

- Manual (the best!?):
  - SSER (subjective sentence error rate)
  - Correct/Incorrect
  - **Adequacy and Fluency** (5 or 7 point scales)
  - Error categorization
  - **Comparative ranking of translations**
- Testing in an application that uses MT as one sub-component
  - Question answering from foreign language documents
- Automatic metric:
  - WER (word error rate) – why problematic?
  - **BLEU (Bilingual Evaluation Understudy)**

BLEU Evaluation Metric

(Papineni et al, ACL-2002)

Reference (human) translation:

The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/ chemical attack against public places such as the airport .

Machine translation:

The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

- N-gram precision (score is between 0 & 1)
  - What percentage of machine n-grams can be found in the reference translation?
    - An n-gram is an sequence of n words
  - Not allowed to match same portion of reference translation twice at a certain n-gram level (two MT words *airport* are only correct if two reference words *airport;* can't cheat by typing out "the the the the the")
  - Do count unigrams also in a bigram for unigram precision, etc.
- Brevity Penalty
  - Can't just type out single word "the" (precision 1.0!)
- It was thought quite hard to "game" the system (i.e., to find a way to change machine output so that BLEU goes up, but quality doesn't)

BLEU Evaluation Metric

(Papineni et al, ACL-2002)

Reference (human) translation:
The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/ chemical attack against public places such as the airport .

Machine translation:
The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

- BLEU is a weighted geometric mean, with a brevity penalty factor added.
  - Note that it's precision-oriented
- BLEU4 formula
  (counts n-grams up to length 4)

exp (1.0 * log p1 +
0.5 * log p2 +
0.25 * log p3 +
0.125 * log p4 –
max(words-in-reference / words-in-machine – 1, 0)

p1 = 1-gram precision
P2 = 2-gram precision
P3 = 3-gram precision
P4 = 4-gram precision

Note: only works at corpus level (zeroes kill it);
there's a smoothed variant for sentence-level

BLEU in Action

枪手被警方击毙。 (Foreign Original)

the gunman was shot to death by the police . **(Reference Translation)**

the gunman was police kill .	#1
wounded police jaya of	#2
the gunman was shot dead by the police .	#3
the gunman arrested by police kill .	#4
the gunmen were killed .	#5
the gunman was shot to death by the police .	#6
gunmen were killed by police ?SUB>0 ?SUB>0	#7
al by the police .	#8
the ringer is killed by the police .	#9
police killed the gunman .	#10

green	= 4-gram match	(good!)
red	= word not matched	(bad!)

Multiple Reference Translations

Reference translation 1:
The U.S. island of Guam is maintaining a high state of alert after the Guam airport and its offices both received an e-mail from someone calling himself the Saudi Arabian Osama bin Laden and threatening a biological/chemical attack against public places such as the airport .

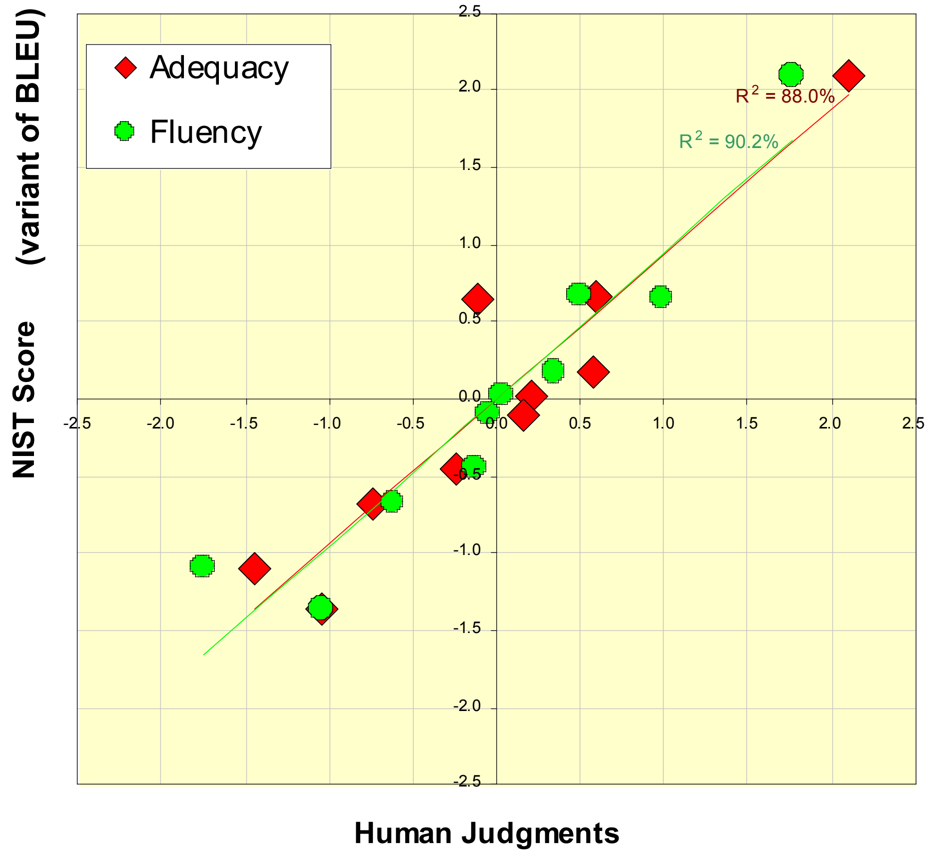
Reference translation 2:
Guam International Airport and its offices are maintaining a high state of alert after receiving an e-mail that was from a person claiming to be the wealthy Saudi Arabian businessman Bin Laden and that threatened to launch a biological and chemical attack on the airport and other public places .

Machine translation:
The American [?] international airport and its the office all receives one calls self the sand Arab rich business [?] and so on electronic mail , which sends out ; The threat will be able after public place and so on the airport to start the biochemistry attack , [?] highly alerts after the maintenance.

Reference translation 3:
The US International Airport of Guam and its office has received an email from a self-claimed Arabian millionaire named Laden , which threatens to launch a biochemical attack on such public places as airport . Guam authority has been on alert .

Reference translation 4:
US Guam International Airport and its office received an email from Mr. Bin Laden and other rich businessman from Saudi Arabia . They said there would be biochemistry air raid to Guam Airport and other public places . Guam needs to be in high precaution about this matter .

Initial results showed that BLEU predicts human judgments well



slide from G. Doddington (NIST)

Automatic evaluation of MT

- People started optimizing their systems to maximize BLEU score
  - BLEU scores improved rapidly
  - The correlation between BLEU and human judgments of quality went way, way down
  - StatMT BLEU scores now approach those of human translations but their true quality remains far below human translations
- Coming up with automatic MT evaluations has become its own research field
  - There are many proposals: TER, METEOR, MaxSim, SEPIA, our own RTE-MT
  - TERpA is a representative good one that handles some word choice variation.
- MT research really requires *some* automatic metric to allow a rapid development and evaluation cycle.

Quiz question!

FOR MONDAY JANUARY 24TH

Hyp: The gunman was shot dead by police .
Ref1: The gunman was shot to death by the police .
Ref2: The cops shot the gunman dead .

Compute the unigram precision P_1 and the trigram precision P_3.

(Note: punctuation tokens are counted, but not sentence boundary tokens.)

(a) $P_1 = 1.0$ $P_3 = 0.5$
(b) $P_1 = 1.0$ $P_3 = 0.333$
(c) $P_1 = 0.875$ $P_3 = 0.333$
(d) $P_1 = 0.875$ $P_3 = 0.167$
(e) $P_1 = 0.8$ $P_3 = 0.167$

A complete translation system

Decoding for IBM Models

- Of all conceivable English word strings, find the one maximizing P(e) x P(f | e)

- Decoding is NP hard
  - (Knight, 1999)
- Several search strategies are available
  - Usually a beam search where we keep multiple stacks for candidates covering the same number of source words
- Each potential English output is called a *hypothesis*.

Search for Best Translation

voulez – vous vous taire !

Search for Best Translation

voulez – vous vous taire !

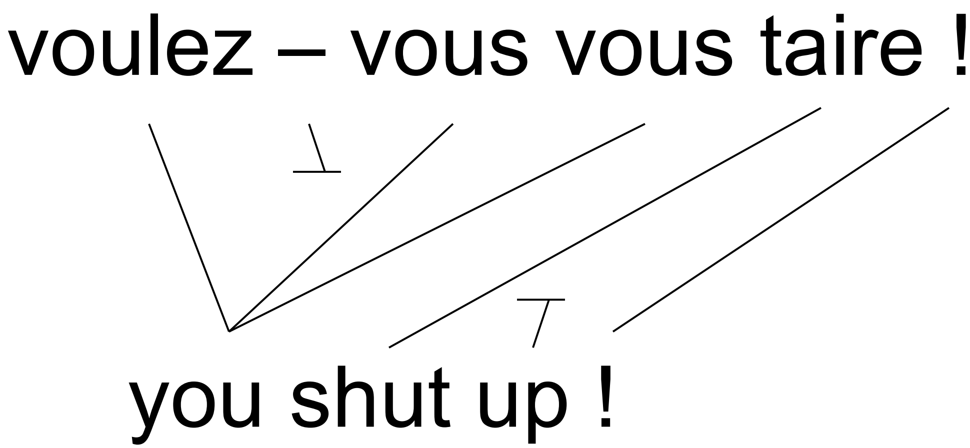
you – you you quiet !

Search for Best Translation

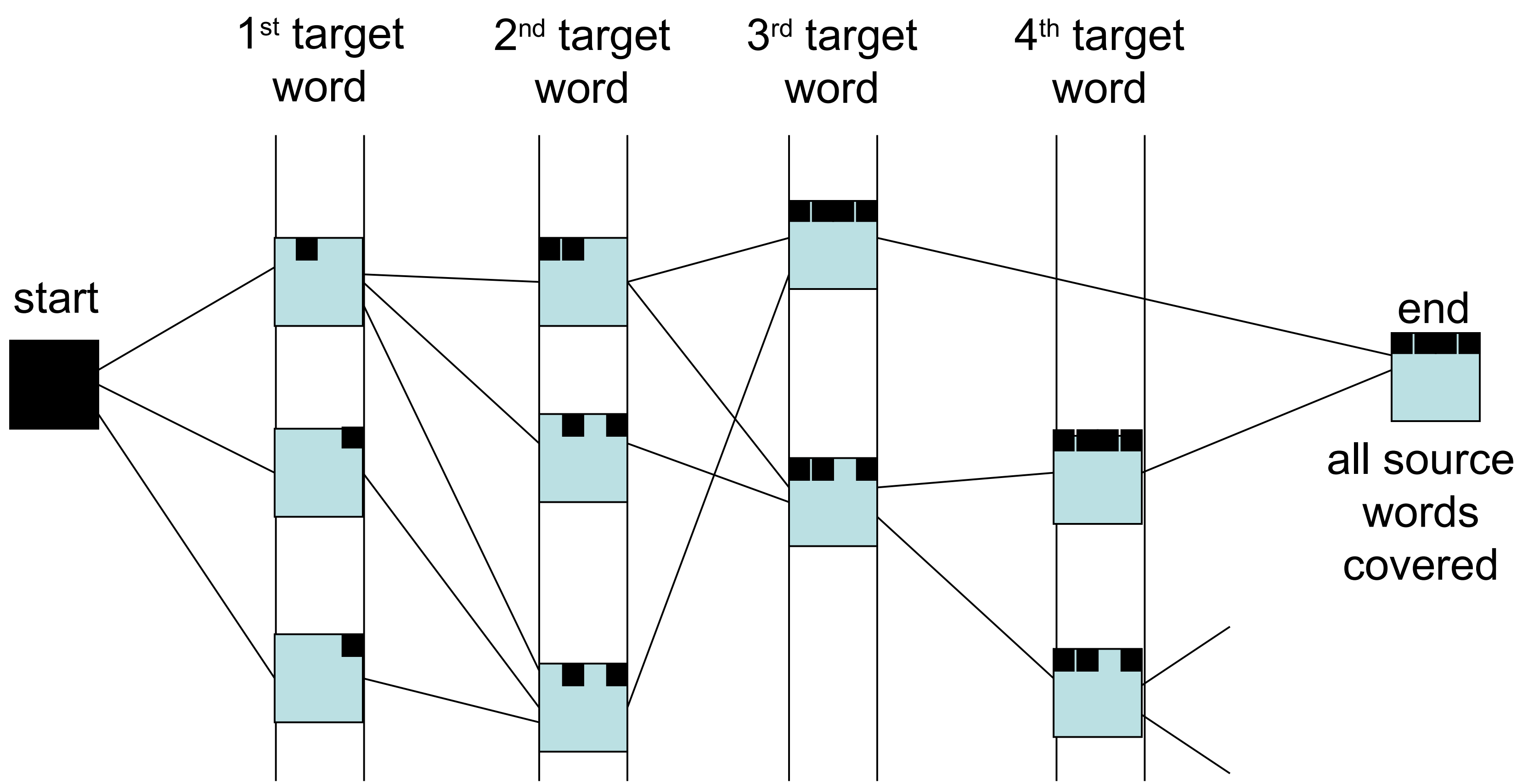
voulez – vous vous taire !

quiet you – you you !

Search for Best Translation



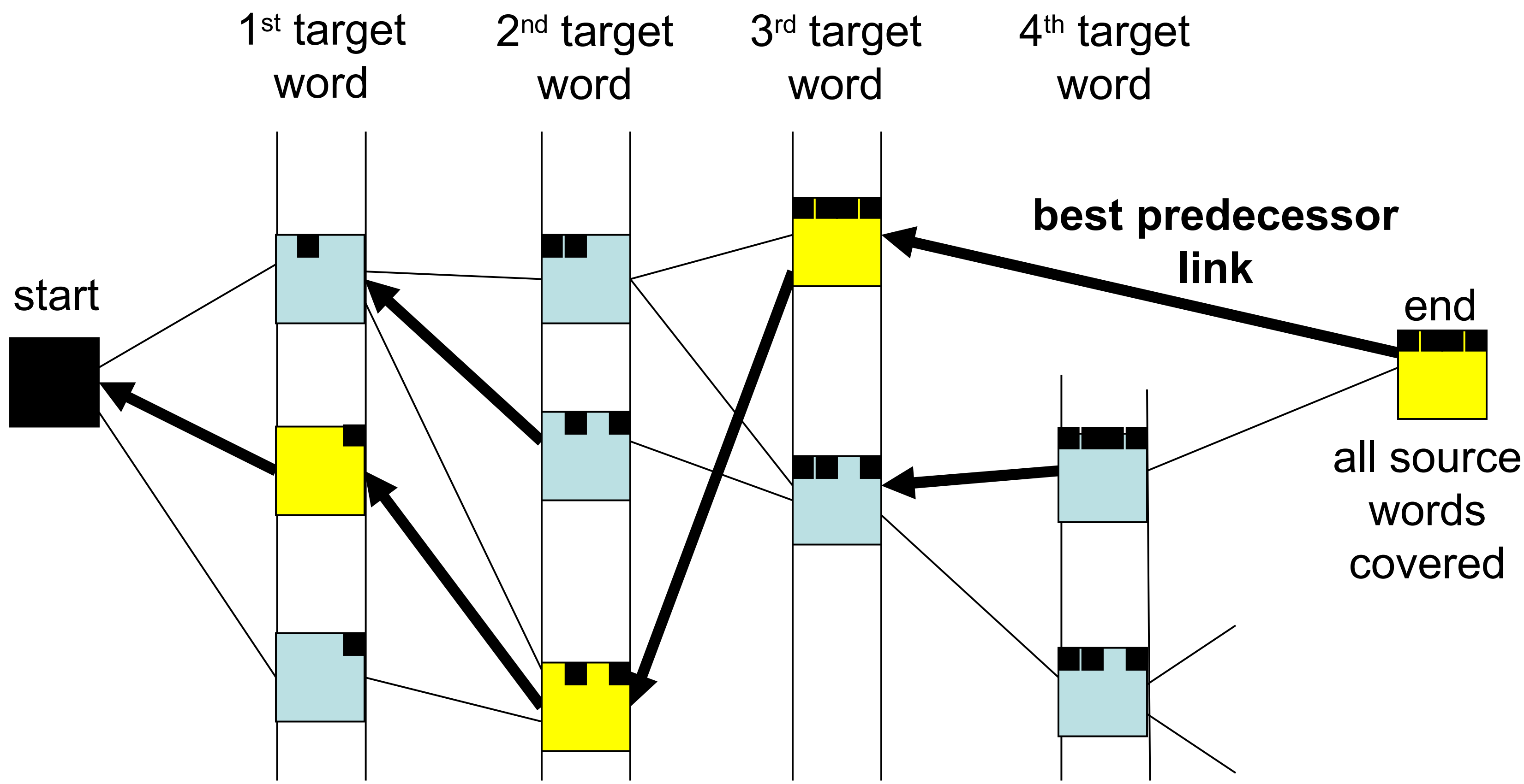
Dynamic Programming Beam Search



Each partial translation hypothesis contains:
- Last English word chosen + source words covered by it
- Next-to-last English word chosen
- Entire coverage vector (so far) of source sentence
- Language model and translation model scores (so far)

[Jelinek, 1969;
Brown et al, 1996 US Patent;
(Och, Ueffing, and Ney, 2001]

Dynamic Programming Beam Search



Each partial translation hypothesis contains:
- Last English word chosen + source words covered by it
- Next-to-last English word chosen
- Entire coverage vector (so far) of source sentence
- Language model and translation model scores (so far)

[Jelinek, 1969;
Brown et al, 1996 US Patent;
(Och, Ueffing, and Ney, 2001]

The "Fundamental Equation of Machine Translation" (Brown et al. 1993)

$$\hat{e} = \underset{e}{\operatorname{argmax}}\ P(e \mid f)$$

$$= \underset{e}{\operatorname{argmax}}\ P(e) \times P(f \mid e) / P(f)$$

$$= \underset{e}{\operatorname{argmax}}\ P(e) \times P(f \mid e)$$

What StatMT people do in the privacy of their own homes

$$\underset{e}{\operatorname{argmax}} \; P(e \mid f) \; =$$

$$\underset{e}{\operatorname{argmax}} \; P(e) \times P(f \mid e) \,/\, P(f) \quad \neq$$

$$\underset{e}{\operatorname{argmax}} \; P(e)^{1.9} \times P(f \mid e)$$ … works better!

Which model are you now paying more attention to?

What StatMT people do in the privacy of their own homes

$$\underset{e}{\operatorname{argmax}} \; P(e \mid f) =$$

$$\underset{e}{\operatorname{argmax}} \; P(e) \times P(f \mid e) \,/\, P(f)$$

$$\underset{e}{\operatorname{argmax}} \; P(e)^{1.9} \times P(f \mid e) \times 1.1^{\text{length}(e)}$$

Rewards longer hypotheses, since these are 'unfairly' punished by P(e)

What StatMT people do in the privacy of their own homes

$$\underset{e}{\text{argmax}} \; \underbrace{P(e)^{1.9} \times P(f \mid e) \times 1.1^{\text{length}(e)} \times KS^{3.7} \ldots}$$

Lots of knowledge sources vote on any given hypothesis.

"Knowledge source" = "feature function" = "score component".

Feature function simply scores a hypothesis with a real value.

(May be binary, as in "e has a verb").

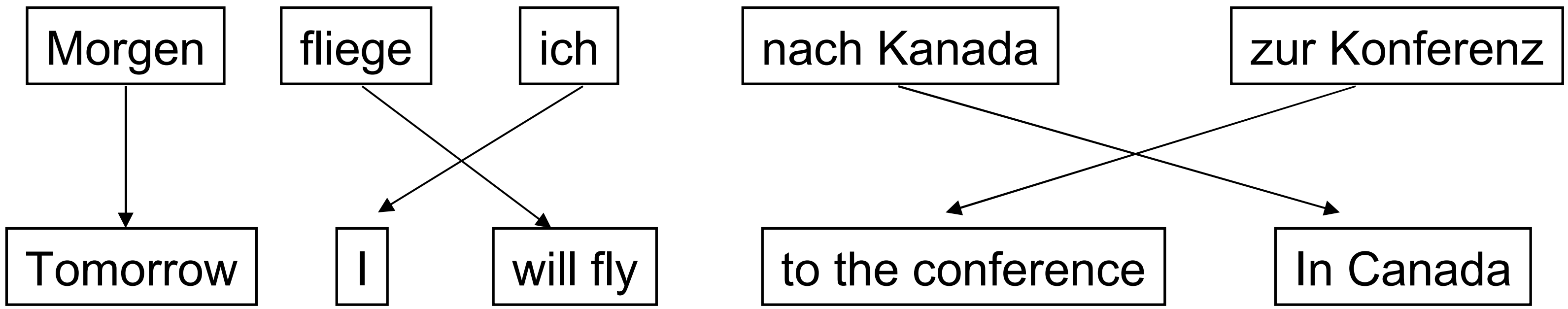
Problem: How to set the weights?
(We look at one way later: maxent models.)

Flaws of Word-Based MT

- Multiple English words for one French word
  - IBM models can do one-to-many (fertility) but not many-to-one
- Phrasal Translation
  - “real estate”, “note that”, “interested in”
- Syntactic Transformations
  - Verb at the beginning in Arabic
  - Translation model penalizes any proposed re-ordering
  - Language model not strong enough to force the verb to move to the right place

Phrase-Based Statistical MT

Phrase-Based Statistical MT



- Foreign input segmented into phrases
  - "phrase" is any sequence of words
- Each phrase is probabilistically translated into English
  - P(to the conference | zur Konferenz)
  - P(into the meeting | zur Konferenz)
- Phrases are probabilistically re-ordered

See J&M or Lopez 2008 for an intro.

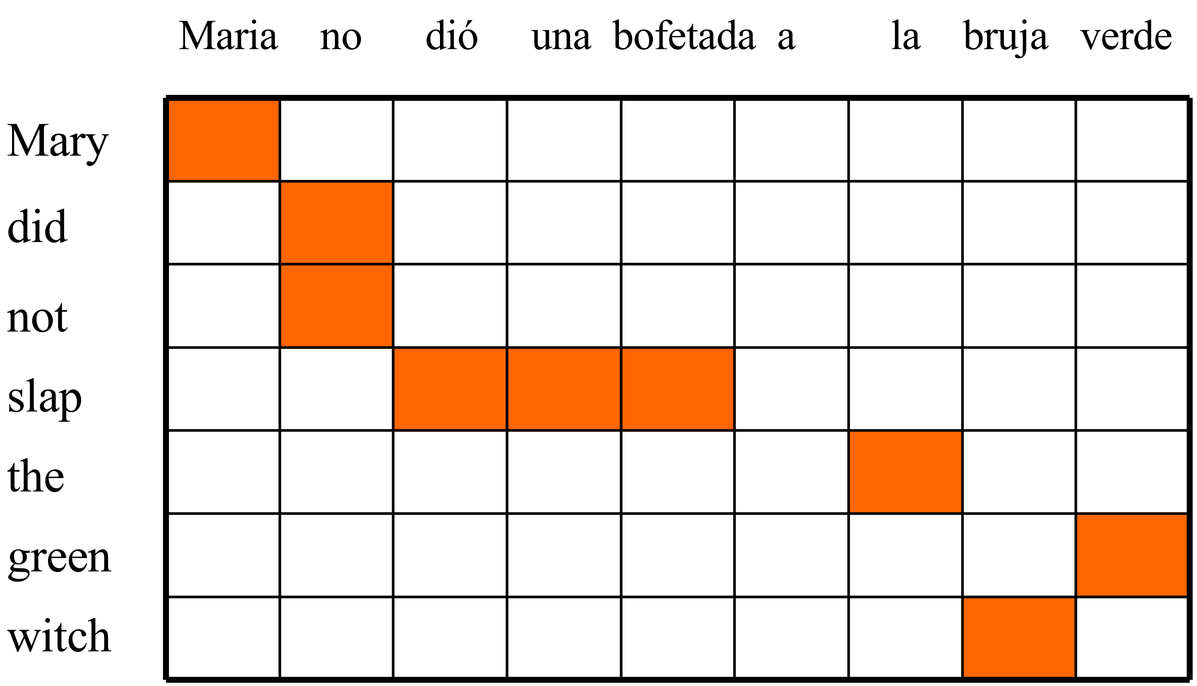
This is still pretty much the state-of-the-art!

Advantages of Phrase-Based

- Many-to-many mappings can handle non-compositional phrases
- Local context is very useful for disambiguating
  - "interest rate" → …
  - "interest in" → …
- The more data, the longer the learned phrases
  - Sometimes whole sentences

How to Learn the Phrase Translation Table?

- Main method: “alignment templates” (Och et al, 1999)
- Start with word alignment, build phrases from that.

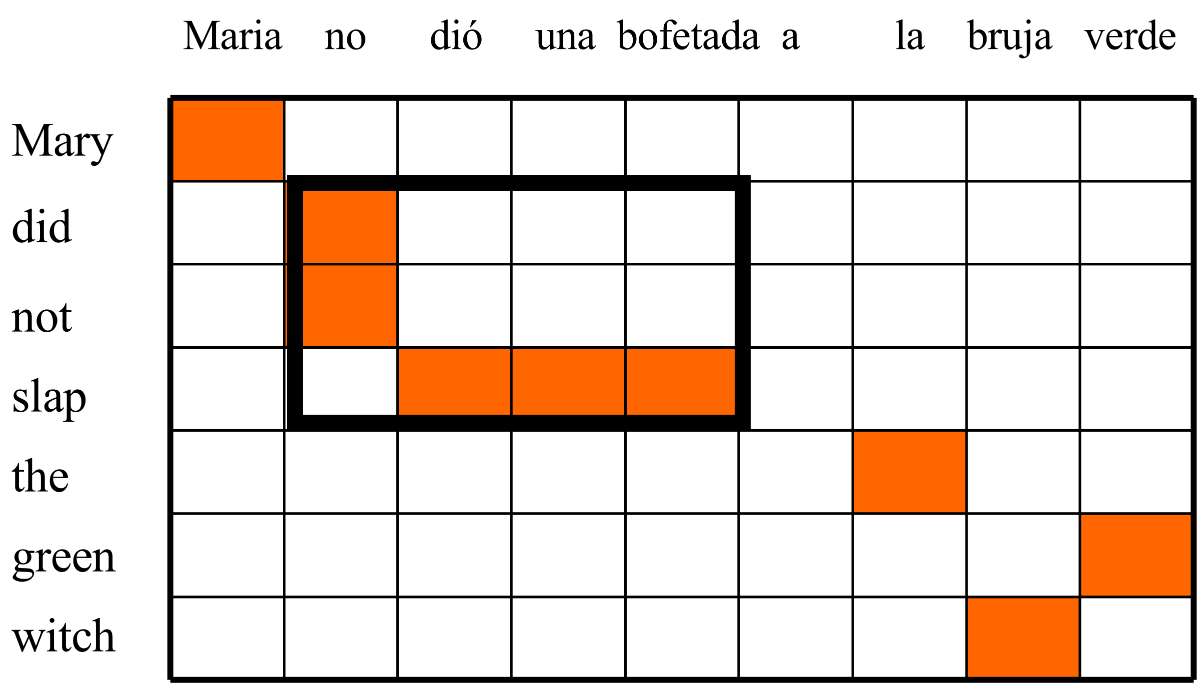


This word-to-word alignment is a by-product of training a translation model like IBM-Model-3.

This is the best (or “Viterbi”) alignment.

How to Learn the Phrase Translation Table?

- Main method: “alignment templates” (Och et al, 1999)
- Start with word alignment, build phrases from that.



This word-to-word alignment is a by-product of training a translation model like IBM-Model-3.

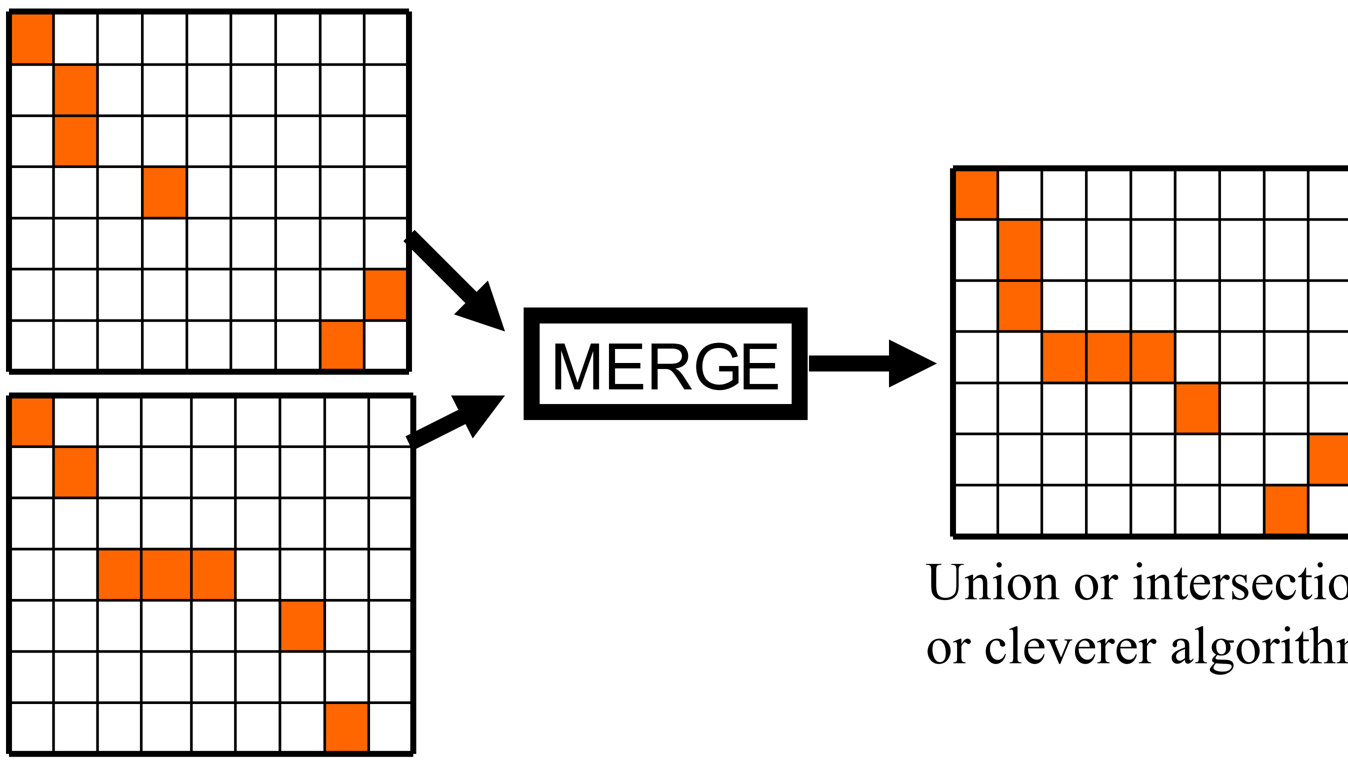
This is the best (or “Viterbi”) alignment.

IBM Models are 1-to-Many

- Run IBM-style aligner both directions, then merge:

E→F best alignment

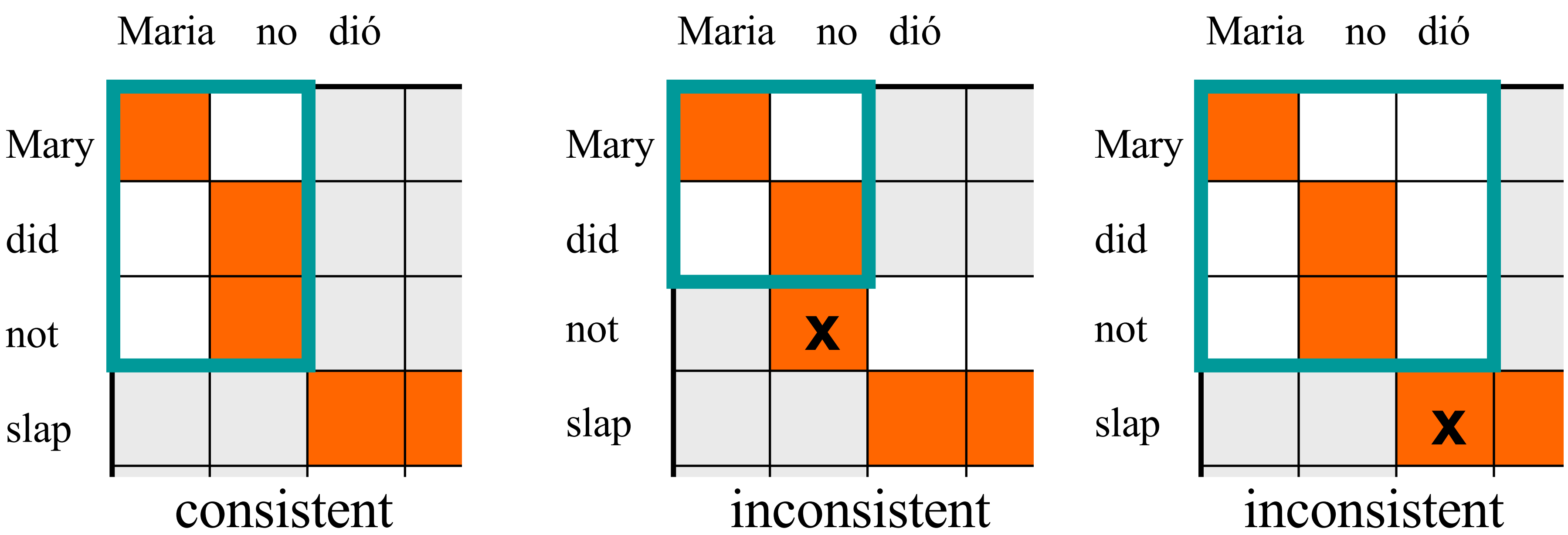
F→E best alignment



Union or intersection or cleverer algorithm

How to Learn the Phrase Translation Table?

- Collect all phrase pairs *that are consistent with the word alignment*



- Phrase alignment must contain all alignment points for all the words in both phrases!
- These phrase alignments are sometimes called *beads*

Phrase Pair Probabilities

- A certain phrase pair (f-f-f, e-e-e) may appear many times across the bilingual corpus.
- No EM training
- Just relative frequency:

$$P(\text{f-f-f} \mid \text{e-e-e}) = \frac{\text{count(f-f-f, e-e-e)}}{\text{count(e-e-e)}}$$

Phrase-Based Translation

这	7 人	中包括	来自	法国	和	俄罗斯	的	宇航	员	.
the	7 people	including	by some		**and**	the russian	**the**	the astronauts		,
it	7 people included		by france		and the	the russian		international astronautical	of rapporteur .	
this	7 out	including the	**from**	the french	and the russian		the fifth		.	
these	7 among	including from		the french and		of the russian	of	space	members	.
that	7 persons	including from the		of france	and to	russian	of the	aerospace	members .	
	7 include		from the	of france and		russian		**astronauts**		. the
	7 numbers include		**from france**		and russian		of astronauts who			. "
	7 populations include		those from france		and russian			astronauts .		
	7 deportees included		come from	**france**	**and russia**		in	astronautical	personnel	;
	7 philtrum	including those from		**france and**		**russia**	a space		**member**	
		including representatives from		france and the		**russia**		astronaut		
		include	came from	**france and russia**			by cosmonauts			
		include representatives from		french	**and russia**			cosmonauts		
		include	came from france		and russia 's			cosmonauts .		
		includes	coming from	french and		russia 's		cosmonaut		
				french and russian			's	astronavigation	member .	
				french	**and russia**		**astronauts**			
					and russia 's				special rapporteur	
					, and	**russia**			rapporteur	
					, and russia				rapporteur .	
					, and russia					
					or	russia 's				

Table 1: #11# the seven - member crew includes astronauts from france and russia .

Scoring: Try to use phrase pairs that have been frequently observed.
Try to output a sentence with frequent English word sequences.

Phrase-Based Translation

这	7 人	中包括	来自	法国	和	俄罗斯	的	宇航	员	.
the	7 people	including	by some		**and**	the russian	**the**	the astronauts		,
it	7 people included		by france		and the	the russian		international astronautical	of rapporteur .	
this	7 out	including the	**from**	the french	and the russian		the fifth		.	
these	7 among	including from		the french and		of the russian	of	space	members	.
that	7 persons	including from the		of france	and to	russian	of the	aerospace	members .	
	7 include		from the	of france and		russian		**astronauts**		. the
	7 numbers include		**from france**		and russian		of astronauts who			. "
	7 populations include		those from france		and russian			astronauts .		
	7 deportees included		come from	**france**	**and russia**		in	astronautical	personnel	;
	7 philtrum	including those from		**france and**		**russia**	a space		**member**	
		including representatives from		france and the		**russia**		astronaut		
		include	came from	**france and russia**			by cosmonauts			
		include representatives from		french	**and russia**			cosmonauts		
		include	came from france		and russia 's			cosmonauts .		
		includes	coming from	french and		russia 's		cosmonaut		
				french and russian			's	astronavigation	member .	
				french	**and russia**		**astronauts**			
					and russia 's				special rapporteur	
					, and	**russia**			rapporteur	
					, and russia				rapporteur .	
					, and russia					
					or	russia 's				

Table 1: #11# the seven - member crew includes astronauts from france and russia .

Scoring: Try to use phrase pairs that have been frequently observed.

Try to output a sentence with frequent English word sequences.

Phrase-Based Translation

这	7 人	中包括	来自	法国	和	俄罗斯	的	宇航	员	.
the	7 people	including	by some		**and**	the russian	**the**	the astronauts		,
it	7 people included		by france		and the	the russian		international astronautical	of rapporteur .	
this	7 out	including the	**from**	the french	and the russian		the fifth		.	
these	7 among	including from		the french and		of the russian	of	space	members	.
that	7 persons	including from the		of france	and to	russian	of the	aerospace	members	
	7 include		from the	of france and		russian		**astronauts**		. the
	7 numbers include		**from france**		and russian		of astronauts who			. "
	7 populations include		those from france		and russian			astronauts .		
	7 deportees included		come from	**france**	**and russia**		in	astronautical	personnel	;
	7 philtrum	including those from		**france and**		**russia**	a space		**member**	
		including representatives from		france and the		**russia**		astronaut		
		include	came from	**france and russia**			by cosmonauts			
		include representatives from		french	**and russia**			cosmonauts		
		include	came from france		and russia 's			cosmonauts .		
		includes	coming from	french and		russia 's		cosmonaut		
				french and russian			's	astronavigation	member .	
				french	**and russia**		**astronauts**			
					and russia 's				special rapporteur	
					, and	**russia**			rapporteur	
					, and russia				rapporteur .	
					, and russia					
					or	russia 's				

Table 1: #11# the seven - member crew includes astronauts from france and russia .

Scoring: Try to use phrase pairs that have been frequently observed.
Try to output a sentence with frequent English word sequences.

Phrase-Based Translation

这	7 人	中包括	来自	法国	和	俄罗斯	的	宇航	员	.
the	7 people	including	by some		and	the russian	the	the astronauts		,
it	7 people included		by france		and the	the russian		international astronautical	of rapporteur .	
this	7 out	including the	from	the french	and the russian		the fifth		.	
these	7 among	including from		the french and		of the russian	of	space	members	.
that	7 persons	including from the		of france	and to	russian	of the	aerospace	members	
	7 include		from the	of france and		russian		astronauts		. the
	7 numbers include		from france		and russian		of astronauts who			. "
	7 populations include		those from france		and russian			astronauts .		
	7 deportees included		come from	france	and russia		in	astronautical	personnel	;
	7 philtrum	including those from		france and		russia	a space		member	
		including representatives from		france and the		russia		astronaut		
		include	came from	france and russia			by cosmonauts			
		include representatives from		french	and russia			cosmonauts		
		include	came from france		and russia 's			cosmonauts .		
		includes	coming from	french and		russia 's		cosmonaut		
				french and russian			's	astronavigation	member .	
				french	and russia		astronauts			
					and russia 's				special rapporteur	
					, and	russia			rapporteur	
					, and russia				rapporteur .	
					, and russia					
					or	russia 's				

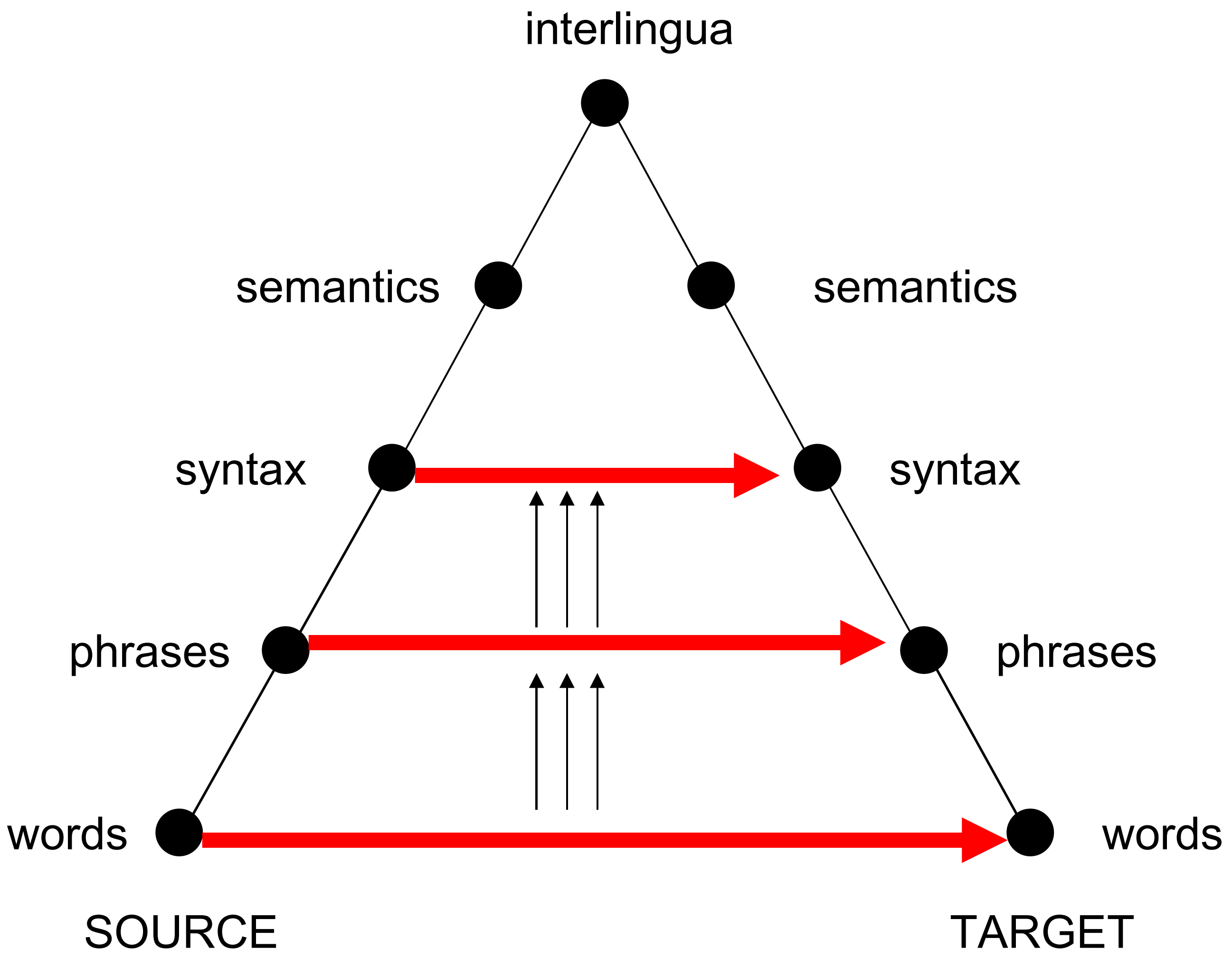
Table 1: #11# the seven - member crew includes astronauts from france and russia .

Scoring: Try to use phrase pairs that have been frequently observed.

Try to output a sentence with frequent English word sequences.

Syntax and Semantics in Statistical MT

Vauquois Triangle



Why Syntax?

- Need much more grammatical output
- Need accurate control over re-ordering
- Need accurate insertion of function words
- Word translations need to depend on grammatically-related words

Yamada and Knight (2001): The need for phrasal syntax

- He adores listening to music.

彼 は 音楽 を 聞く の が 大好き です

Kare ha ongaku wo kiku no ga daisuki desu

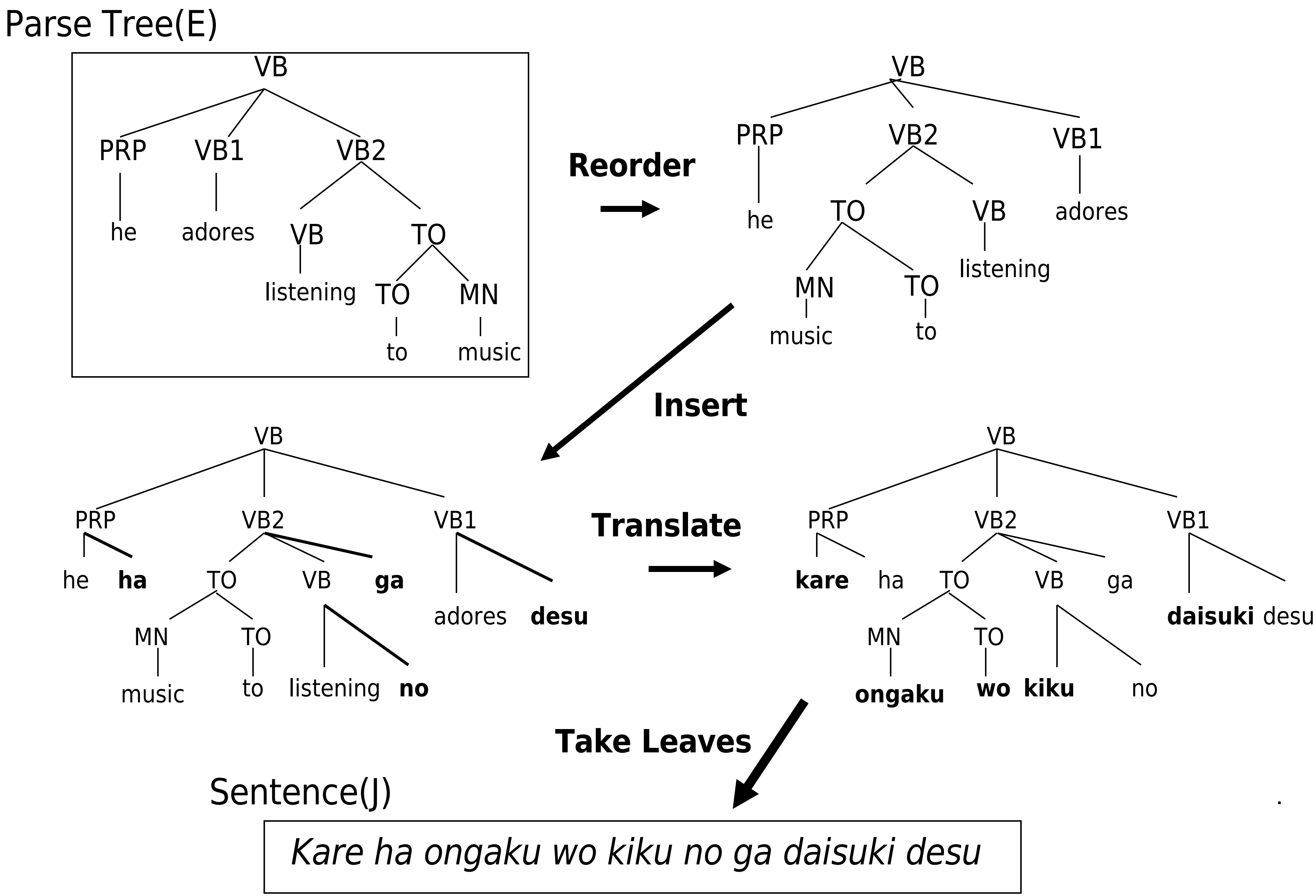
Syntax-based Model

- E→J Translation (Channel) Model

Parse Tree (English) ⇒ $\boxed{\text{Translation model}}$ ⇒ Sentence (Japanese)

- Preprocess English by a parser
- Probabilistic Operations on a parse-tree
  1. Reorder child nodes
  2. Insert extra nodes
  3. Translate leaf words

Parse Tree(E) → Sentence (J)



Experiment

- Training Corpus: J-E 2K sentence pairs
- J: Tokenized by Chasen [Matsumoto, et al., 1999]
- E: Parsed by Collins Parser [Collins, 1999]

  --- Trained: 40K Treebank, Accuracy: ~90%
- E: Flatten parse tree

  --- To Capture word-order difference (SVO->SOV)
- EM Training: 20 Iterations

  --- 50 min/iter (Sparc 200Mhz 1-CPU) or

  --- 30 sec/iter (Pentium3 700Mhz 30-CPU)

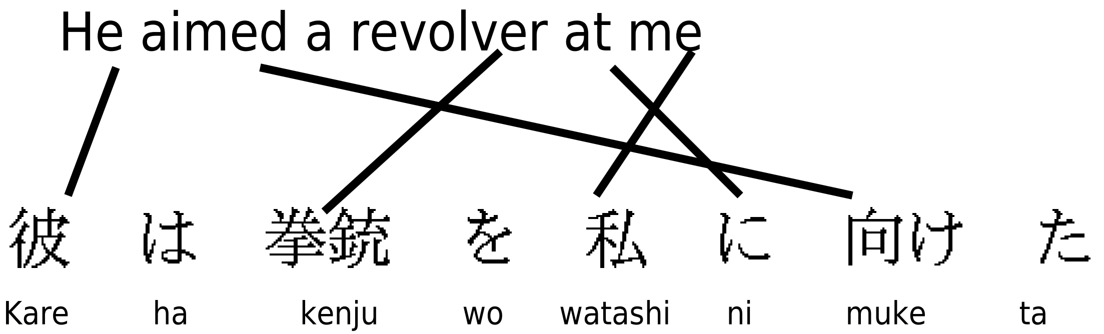
Result: Alignments

	Ave. Score	# perf sent
Y/K Model	0.582	10
IBM Model 5	0.431	0

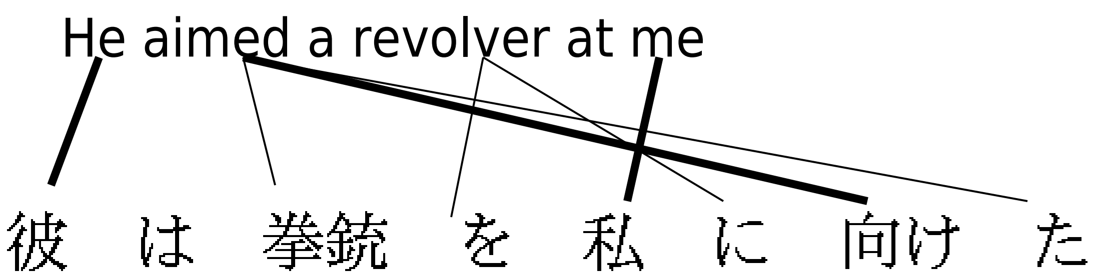
- Ave. by 3 humans for 50 sents
- okay(1.0), not sure(0.5), wrong(0.0)
- precision only

Result: Alignment 2

Syntax-based model

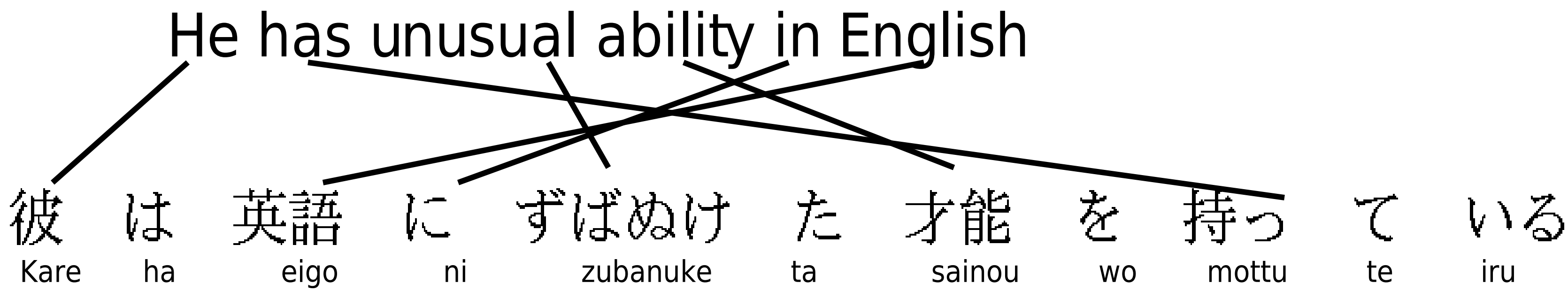


IBM Model 3

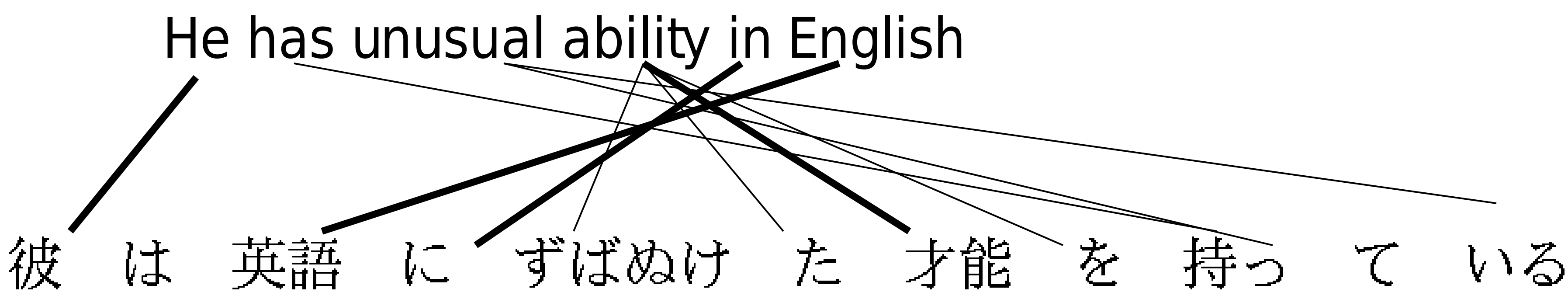


Result: Alignment 3

Syntax-based Model



IBM Model 3



MT Applications

Gerald Penn

CS 224N

2011

[Based on slides by Chris Manning]

MT: The early history (1950s)

- Earl
  less
- Fou
  lang
- First
- MT
  word
- Little
  sem
- Prob

MT Applications: 1. Traditional

- Traditional scenario:
  - Documents had to be translated for your company/organization. Document production for organization
  - Generally, the quality/accuracy demands are high
  - High cost
    - Though most of it is now done as outsourced piecework
- MT tends to be ineffective: The cost of post-translation error correction is too high
- Main technology in the game: translation memory/translation workbench/terminology management
  - E.g., TRADOS.
    - Very slowly, MT technology is starting to be incorporated, but most of the action is in terminology lexicon management

The person shown here is a basic user of this site. He or she may be contacted directly for language-related services.

Working languages:
English to Japanese
Portuguese to Japanese
Japanese to Portuguese

More

Local time
23:50 JST (GMT+9)

T. I. E. Japan Co., Ltd - T. I. E. Japan Co., Ltd

Accuracy. Reliability. Reasonable price.

Contact:

Minato-ku, Tokyo, Japan / Native in: Japanese

Profile | Company profile | Contact

	Freelancer and outsourcer
Services	Translation, Website localization, Software localization, Desktop publishing
Expertise	**Specializes in:** Business/Commerce (general) Mechanics / Mech Engineering IT (Information Technology) Electronics / Elect Eng Computers (general) Mathematics & Statistics Journalism Internet, e-Commerce Telecom(munications) Computers: Systems, Networks
Rates	English to Japanese - Rates: 0.10 - 0.06 USD per word / 40 - 20 USD per hour Portuguese to Japanese - Rates: 0.13 - 0.08 USD per word / 45 - 25 USD per hour Japanese to Portuguese - Rates: 0.13 - 0.08 USD per word / 45 - 25 USD per hour Spanish to Japanese - Rates: 0.13 - 0.08 USD per word / 45 - 25 USD per hour
KudoZ activity	Questions answered: 2, Questions asked: 0 Easy / 0 PRO
Translation education	Univ. de Sao Paulo
Experience	Years of translation experience: 14. Registered at ProZ.com: Oct 2006.
ProZ.com Certified PRO certificate(s)	N/A
Credentials	N/A
Memberships	N/A
Software	Adobe Acrobat, Adobe Photoshop, Microsoft Excel, Microsoft Word, Pagemaker, Powerpoint
Website	http://www.logos-net.com
CV/Resume	CV/Resume

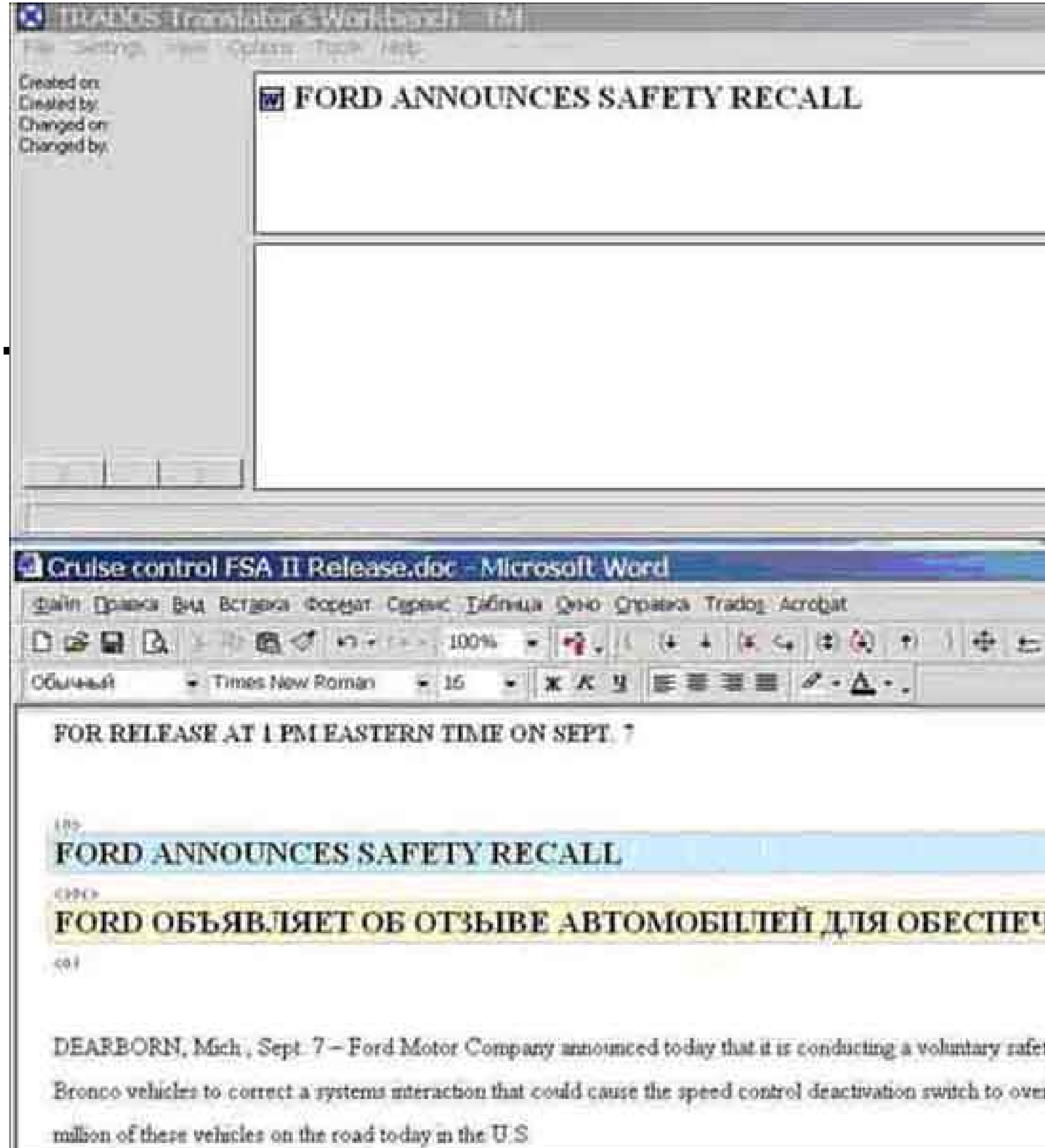
About me

Languages - Japanese / Brazilian Portuguese / Spanish / English

English-Japanese
Brazilian Portuguese-Japanese, Japanese-Brazilian Portuguese
English-Brazilian Portuguese
Spanish-Japanese

Mother tongue: Japanese

Bad TRADOS Screenshot…



Trados is relatively pricey (high hundreds for PC versions, thousands for server version); seen as necessary productivity tool (Photoshop for translators)

MT Applications: 2. Web

- Web applications:
  - Dominant scenario: User-initiated translation
    - Crucial difference: The quality doesn't have to be great. The user is usually okay with just understanding the gist of what is going on
  - Second scenario
    - Somehow on the web people will accept medium quality results. Accessible information is better than no information
- MT is saved!!! "It's the web, stupid."
    - (But is there money in it?)

AltaVista BabelFish

1997:

Free, automatic translation for the masses.

Revolutionary.

But, what was the underlying technology?

SYSTRAN.

MacOS Dashboard?
SYSTRAN

Google until 2006?

SYSTRAN

AltaVista
A DIGITAL Internet Service

Palo Alto, CA - USA

About AltaVista

Our Strengths Our Technology Our Story Our Team Our Press Room

DIGITAL ANNOUNCES FIRST INTERNET TRANSLATION CAPABILITY

... AltaVista Search Services Delivers Free Web Translation Service in Spanish, French, German, Portuguese and Italian

MAYNARD, Mass., December 9, 1997

Digital Equipment Corporation today announced it has broken the Internet language barrier, with the first European language translation service for Web content. For the first time, non-English speaking Web surfers can translate information on the predominantly English speaking Web in real time. The new free service, provided on Digital's AltaVista Search service, also gives English-only Web surfers the ability to understand information in five of the world's most popular languages.

"We are launching this technology as a global experiment and we invite our 12 million monthly users to test drive this breakthrough technology for the Web," said Louis Monier, director of technology for Digital's AltaVista Search service. "For the first time, users will be able to understand, in real-time, Web sites written in other languages. With 70 percent of the Web pages written in English, our technology provides access to a world of information to a greater population."

A test version of the technology that provides real-time language translation capabilities is now available on the AltaVista Search site (http://altavista.digital.com). The new technology allows users to easily search, retrieve and translate Web search results from:

- Spanish to English, English to Spanish
- French to English, English to French
- German to English, English to German
- Portuguese to English, English to Portuguese
- Italian to English, English to Italian

Instant Access to Business-Critical Information

Through a technology partnership with SYSTRAN Corporation, a pioneer of automatic language translation technology, AltaVista has access to SYSTRAN's tested multilingual dictionaries, containing more than 2.5 million terms and a wide range of vocabulary and professional specialties. The AltaVista Web Translation Service provides nearly complete translation results, equipping the end-user with a greater comprehension of foreign information. With the new technology, users from around the world will have access to:

SYSTRAN

From Wikipedia, the free encyclopedia

SYSTRAN, founded by Dr. Peter Toma in 1968[1], is one of the oldest machine translation companies. SYSTRAN has done extensive work for the United States Department of Defense and the European Commission.

SYSTRAN provides the technology for Yahoo! and AltaVista's (Babel Fish) among others, but use of it was ended (circa 2007) for all of the language combinations offered by Google's language tools[2].

SYSTRAN is used by the Dashboard Translation widget in Mac OS X.

Commercial versions of SYSTRAN operate with operating systems Microsoft Windows (including Windows Mobile), Linux and Solaris.

Contents [hide]

History [edit]

With its origin in the Georgetown machine translation effort, SYSTRAN was one of the few machine translation systems to survive the major decrease of funding after the ALPAC Report of the mid-1960s. The company was established in La Jolla, California to work on translation of Russian to English text for the United States Air Force during the Cold War. Large numbers of Russian scientific and technical documents were translated using SYSTRAN under the auspices of the USAF Foreign Technology Division (later the National Air and Space Intelligence Center) at Wright-Patterson Air Force Base, Ohio. The quality of the translations, although only approximate, was usually adequate for understanding content.

The company was sold in 1986 to the Gachot family, based in Paris, France, and is now traded publicly by the French stock exchange. It has a main office at the Grande Arche in La Defense and maintains a secondary office in La Jolla, San Diego, California.

During the dot com boom, language industry started a new era, and Systran entered into agreements with a number of translation integrators, the most successful of these being WorldLingo, today many millions of people use Systran's Translation Engines through WorldLingo.

WIKIPEDIA
The Free Encyclopedia
navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

search
Go Search

interaction
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link
- Cite this page

languages
- Español
- Français

English

SYSTRAN

The leading supplier of language translation software

PRODUCTS | SOLUTIONS | SUPPORT | ABOUT SYSTRAN | CONTACT | PURCHASE

Translate with **SYSTRAN**

Text | Web Page

English ⇆ French

SYSTRAN Translation Software

SYSTRAN's wide range of translation software products and solutions will help you create multilingual documents and understand foreign language content in real-time.

Home & Home Office
- Web Translator
- Home Translator
- Office Translator
- Mobile Translator
- All products

Small Business
- Business Translator
- Premium Translator
- Enterprise Server 6
- SYSTRANLinks

Enterprise Solutions
- Premium Translator
- Business Translator
- Enterprise Server 6
- SYSTRANLinks
- Professional Services

Buy Desktop Product Online Choose your product!

New: Financials 2008 [en] (PDF - 55 Kb)
February 13, 2009

Download **The Gilbane Group** Report "Multilingual Communication as a Business Imperative"

An International Benchmark in **Translation Software**

SYSTRAN is the market-leading provider of language translation software products and solutions for desktop, enterprise and Internet that facilitate communication in 52 language combinations and in 20 vertical domains. With over four decades of expertise in Machine Translation dedicated to multilingual translation, SYSTRAN's software is the choice of leading global corporations, Internet portals and public agencies including the US Intelligence Community and the European Commission.

What is automatic translation?

Automatic translation is a translation produced by state-of-the-art technology, without the intervention of human translators. Automatic translation is also often referred to as "Machine Translation".

Automatic translation

LW ENTERPRISE TRANSLATION SERVER

- Multiple input and output formats provide greater flexibility
- Fluent, accurate, and understandable translations provided by the translation software
- Fast, simple integration and deployment

SYSTEM REQUIREMENTS

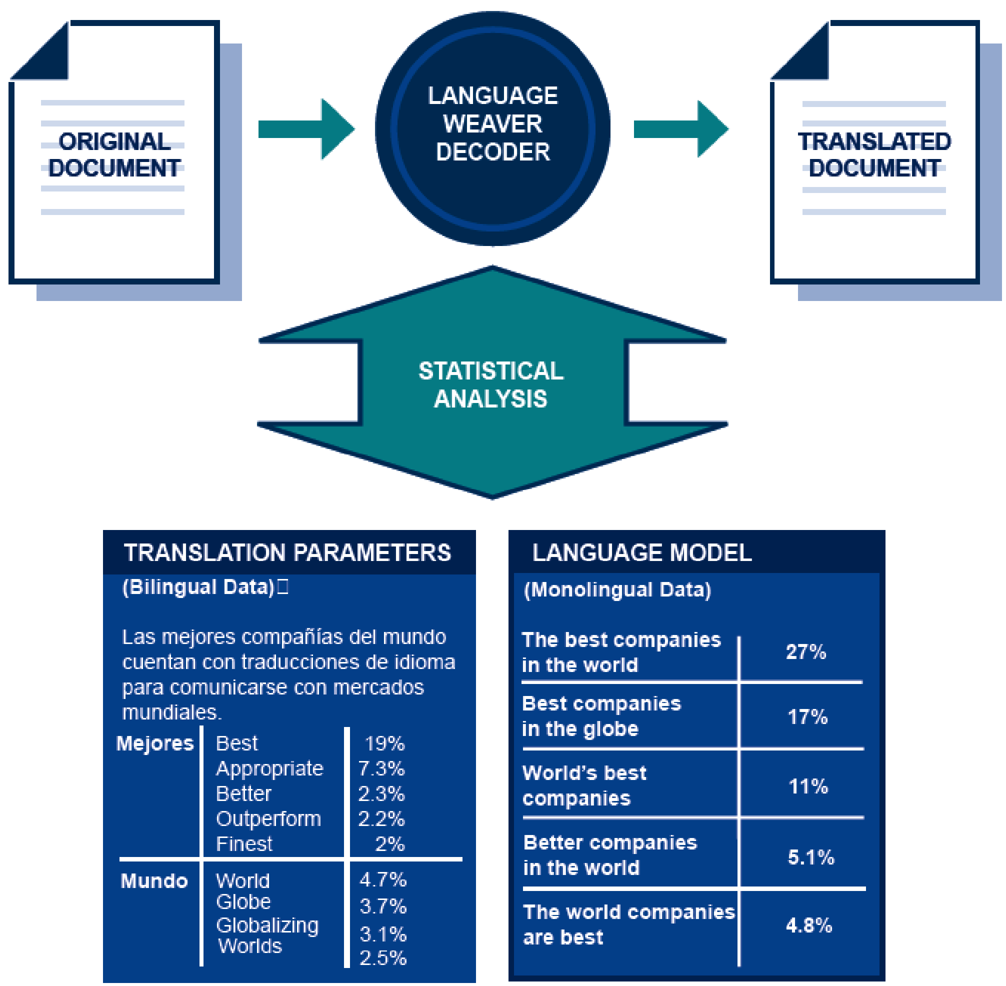
Client Server 32-Bit Platform

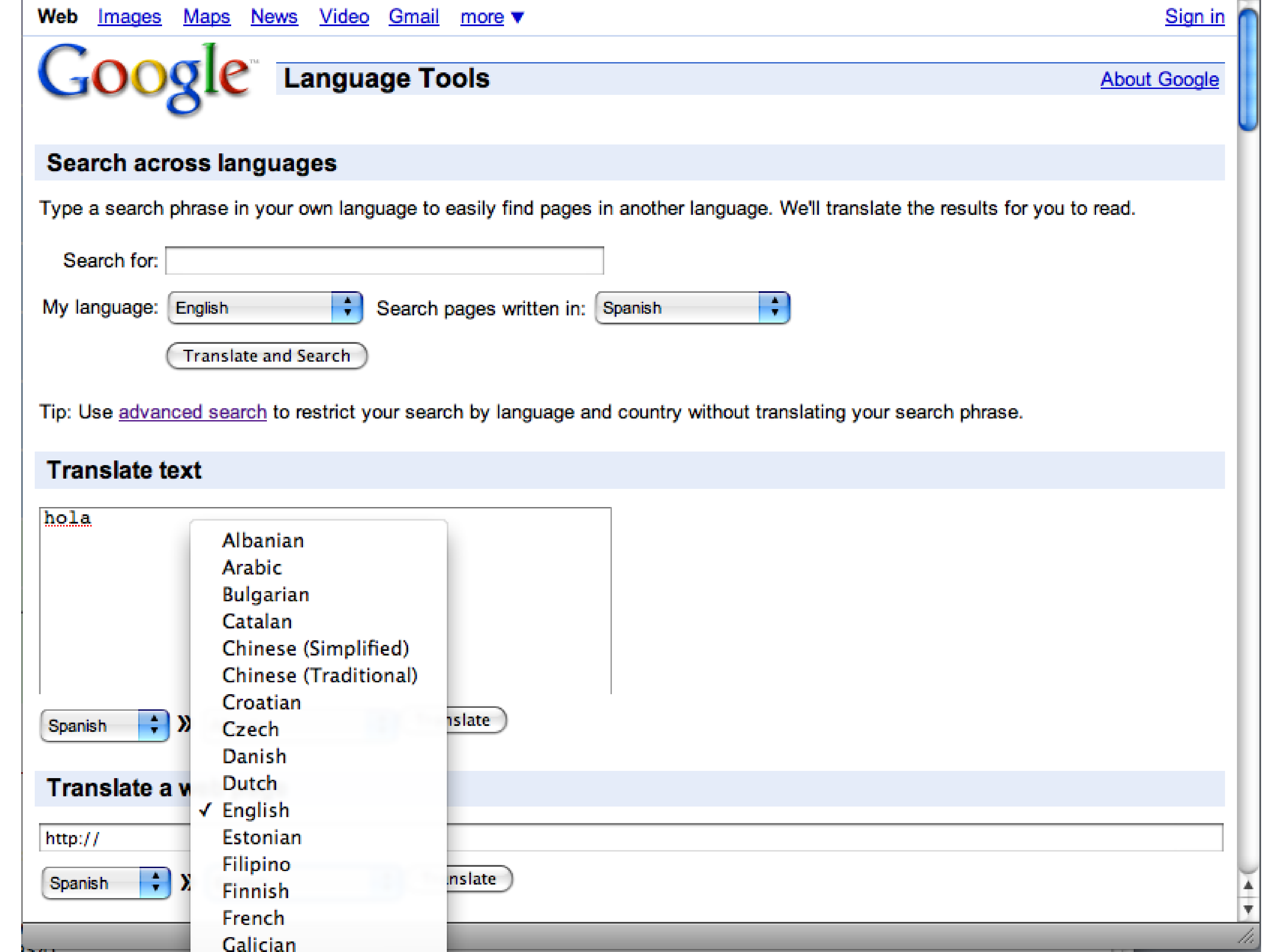
- **OS:** Windows 2003 Server, Windows 2000 Advanced Server (SP3+)
- **Processor:** 2.8 GHz Pentium 4
- **RAM:** 4 GB

Client Server 64-Bit Platform

- **OS:** Windows 2003 Server, Windows 2000 Advanced Server (SP3+) or Linux Fedora 8 or similar (contact us for details)
- **Processor:** 2.8 GHz Intel Xeon
- **RAM:** 8 GB

RUN TIME TRANSLATION PROCESS




Web Images Maps News Video Gmail more
Sign in
Google
Language Tools
About Google
Search across languages
Type a search phrase in your own language to easily find pages in another language. We'll translate the results for you to read.
Search for:
My language: English
Search pages written in: Spanish
Translate and Search
Tip: Use advanced search to restrict your search by language and country without translating your search phrase.
Translate text
hola
Albanian
Arabic
Bulgarian
Catalan
Chinese (Simplified)
Chinese (Traditional)
Croatian
Czech
Danish
Dutch
✓ English
Estonian
Filipino
Finnish
French
Galician
Spanish
Translate a w
http://
Spanish

Microsoft .NET Framework 2.0

http://support.microsoft.com/ph/8291

Barkly Ta... | cs224n-s... | YouTube ... | cs224n st... | java jedit... | Make JEdi... | JavaTech... | JFileChoo... | Microsoft...

- CORREGIR: es que recibas un mensaje de error cuando intenta cerrar una aplicación basada en formularios de Windows cuando la aplicación se ejecuta en .NET Framework 2.0
  Corrige un problema que se puede producir cuando intenta cerrar una aplicación basada en formularios de Windows que utiliza la información sobre herramientas. Es posible que reciba un " System.NullReferenceException: objetos de referencia no se...

Más Descargas...

Solución de problemas - .NET Framework 2.0

- El control de usuario de ActiveX no está visible en el diseñador de formularios cuando se vuelve a abrir un proyecto existente
  Explica que el control de usuario de ActiveX no es visible en el diseñador de Windows Forms cuando se vuelve a abrir un proyecto existente. Proporciona un método para solucionar este problema.
- No puede establecer el foco a los controles secundarios de un control de usuario mediante la tecla TAB
  Explica que no establece el foco a los controles secundarios de un control de usuario mediante la tecla TAB. Proporciona una solución para solucionar este problema.
- ERROR:" Puede de iniciar la depuración en el servidor Web" mensaje de error cuando ejecuta unas aplicaciones Web ASP.NET
  Describe un problema que se produce cuando utiliza localhost como el servidor Web para crear una aplicación ASP.NET y utiliza la dirección IP para la Identificación del sitio Web. Para resolver el problema, modifique el archivo de proyecto .webinfo y...

Más Solución de problemas...

Última revisión : miércoles, 05 de noviembre de 2008

¿Necesita más ayuda?

Contactar con un profesional de soporte técnico por correo electrónico, online o por teléfono

El control de usuario de ActiveX no está visible en el diseñador de formularios cuando se vuelve a abrir un proyecto existente

http://support.microsoft.com/kb/818220/es?spid=8291&sid=global

Barkly Ta... | cs224n-s... | YouTube ... | cs224n s... | java jedit... | Make JEd... | JavaTech... | JFileChoo... | El contro...

Id. de artículo: 818220 - Última revisión: lunes, 03 de diciembre de 2007 - Versión: 3.11

El control de usuario de ActiveX no está visible en el diseñador de formularios cuando se vuelve a abrir un proyecto existente

Advertencia: Artículo de Traducción Automática, vea la exención de responsabilidad.

¿Necesita má ayuda?

Contactar con profesional de soporte técnic correo electró online o por

Solución

Para solucionar este problema, quitar las existentes las referencias de control de usuario de ActiveX y, a continuación, agregue las referencias actualizadas de nuevo. Para ello, siga uno de los métodos siguientes.

Volver al principio

1 (Método)

1. En el Explorador de soluciones, bajo **referencias** , haga clic con el botón secundario del mouse en **AxProject1** y, a continuación, haga clic en **Quitar** .
2. En **referencias** , haga clic con el botón secundario del mouse en **Project1** y, a continuación, haga clic en **Quitar** .
3. En el cuadro de herramientas, haga doble clic en **Project1.TestControl** para agregar el control de usuario de ActiveX modificado en el formulario.

Article ID: 818220 - Last Review: Monday, 03 December 2007 - Version: 3.11

The ActiveX user control is not visible in the form designer when you reopen an existing project

Warning: Article on Machine Translation, see disclaimer.

View products that this article applies

Solution

To solve this problem, remove the existing references to user control of ActiveX, and then add references updated again. To do this, follow one of the following methods.

Back to top

1 (Method)

1. In Solution Explorer, under **References,** right-click the secondary mouse **AxProject1,** and then click **Remove.**
2. **References,** click with the secondary mouse button in **Project1,** and then click **Remove.**
3. In the toolbox, double-click **Project1.TestControl** to add the user control of ActiveX in the amended form.

   Note that the ActiveX user control in the Windows Forms designer and reference to **Project1** and **AxProject1** listed under **references** to **UserControlDemo.**
4. In the **menu,** click **Build** Solution.

Article ID: 818220 - Last Review: December 3, 2007 - Revision: 3.11

The ActiveX user control is not visible in the Form Designer when you reopen an existing project

View products that this article applies to.

On This Page

Expand all | Collapse all

WORKAROUND

To work around this problem, remove the existing references of the ActiveX user control, and then add the updated references again. To do this, follow one of the following methods.

Back to the top

Method 1

1. In Solution Explorer, under **References**, right-click **AxProject1**, and then click **Remove**.
2. Under **References**, right-click **Project1**, and then click **Remove**.
3. In the toolbox, double-click **Project1.TestControl** to add the modified ActiveX user control to the form.

   Notice that the ActiveX user control in the Windows Form Designer and the references to **AxProject1** and **Project1** appear under **References** for **UserControlDemo**.
4. On the **Build** menu, click **Build Solution**.

   You do not receive any errors.

Machine Translation Summary

- Usable Technologies
  - "Translation memories" to aid translator
  - Low quality screening/web translators
- Technologies
  - Traditional: Systran (Altavista Babelfish, what you got till mid-2006 on Google) is now seen as a limited success
  - Statistical MT over huge training sets is successful (ISI/LanguageWeaver, Microsoft, Google)
- Key ideas of the present/future
  - Statistical phrase based models
  - Syntax based models
  - Better language models (e.g., bigger, using grammar)
  - Better decoding models (e.g., by restricting model?)