The UKA/CMU translation system for IWSLT 2006

Matthias Eck, Ian Lane, Nguyen Bach, Sanjika Hewavitharana, Muntsin Kolss, Bing Zhao, Almut Silja Hildebrand, Stephan Vogel, and Alex Waibel

InterACT Research Laboratories:
University of Karlsruhe, Karlsruhe, Germany
Carnegie Mellon University, Pittsburgh, USA

interACT research

Carnegie Mellon

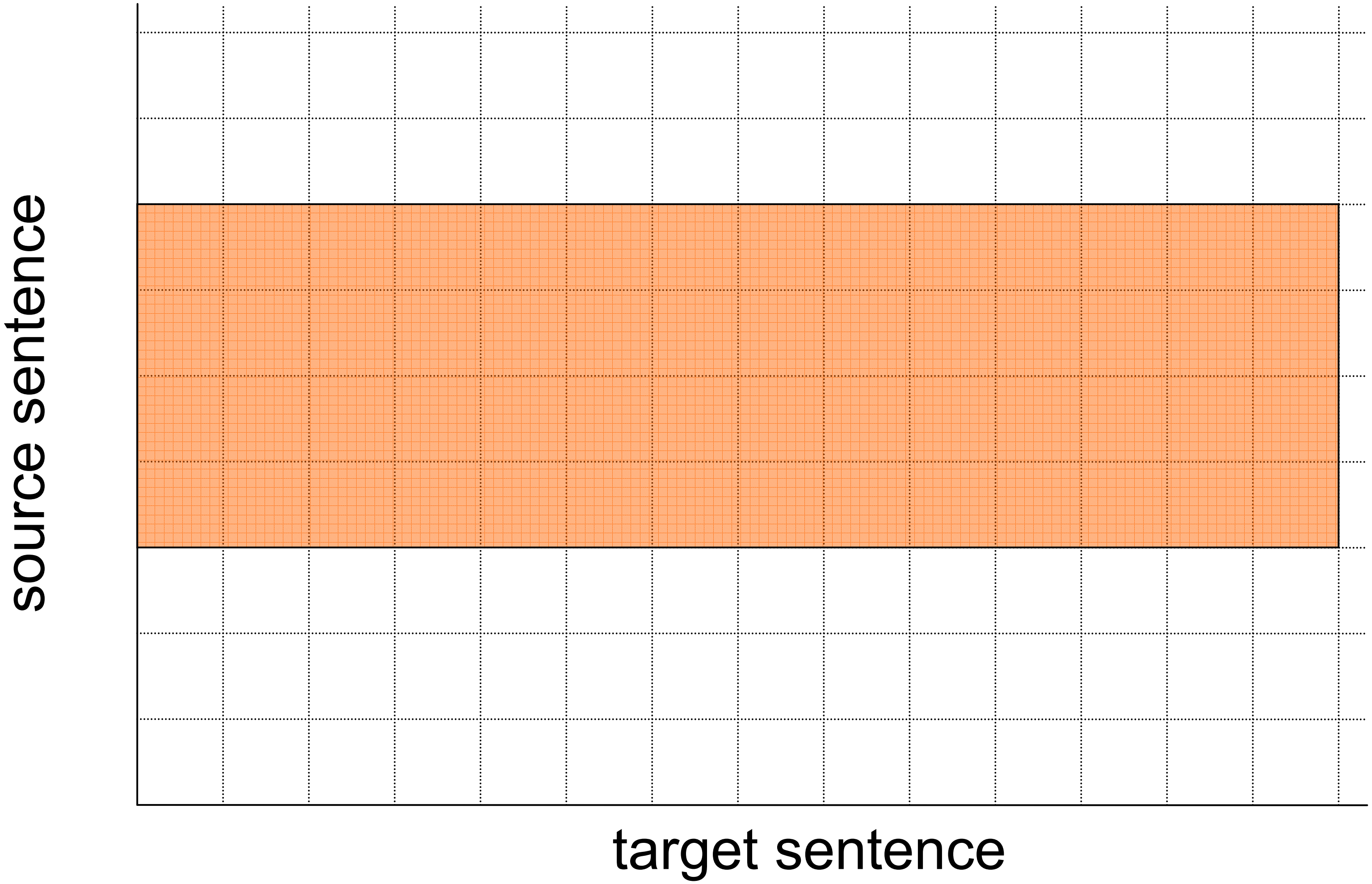
Overview

- SMT System components
  - Phrase Alignment Models
    - PESA
    - Log-Linear Phrase Alignment (LogLin)
  - Language Model
  - Decoder

- Experimental Results
- Analysis
- Conclusions

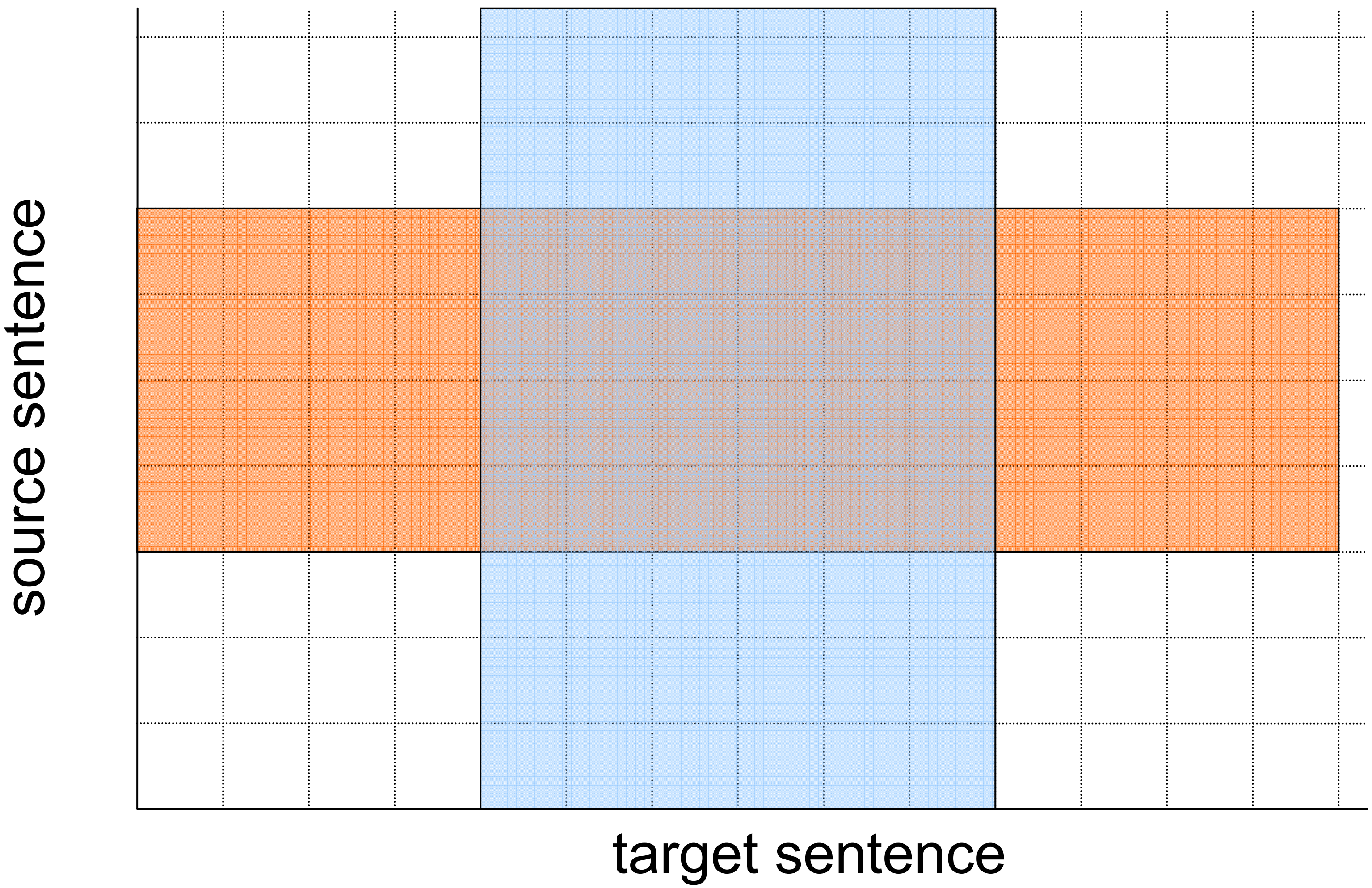
PESA Alignment

- Given source phrase



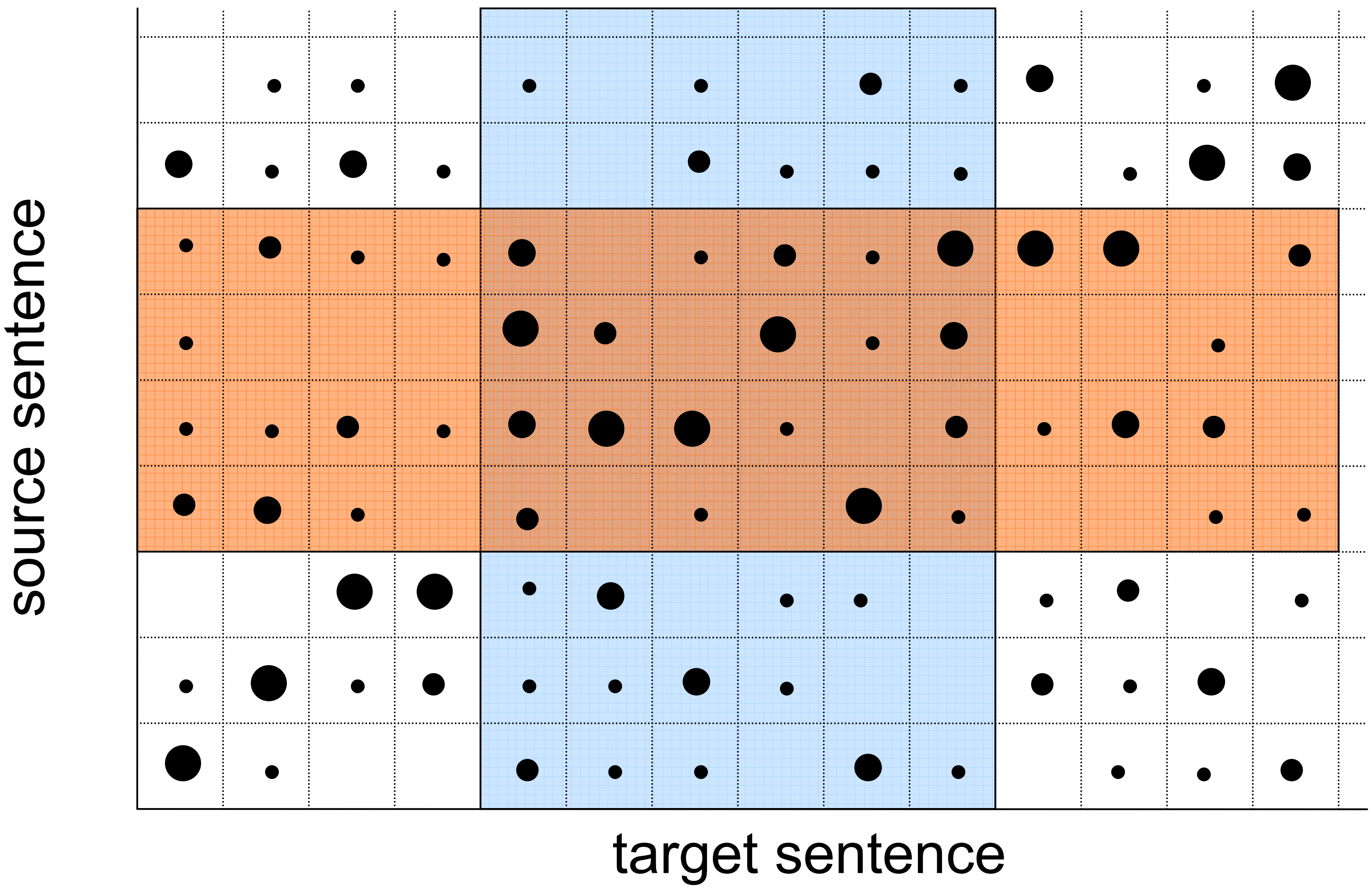
PESA Alignment

- What is the translation of the source phrase?



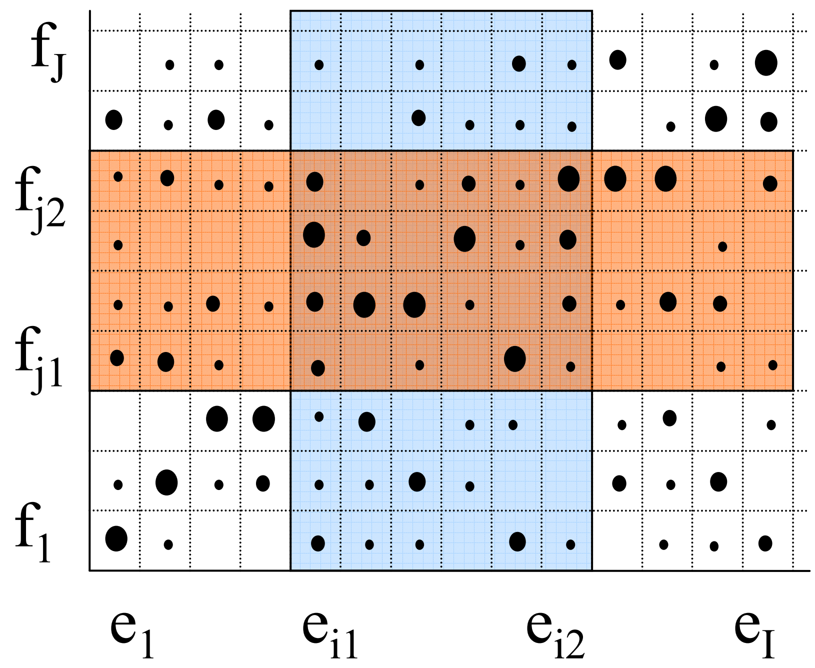
PESA Alignment

- Back to IBM-1 probabilities…



PESA Alignment

- Probability for this split:

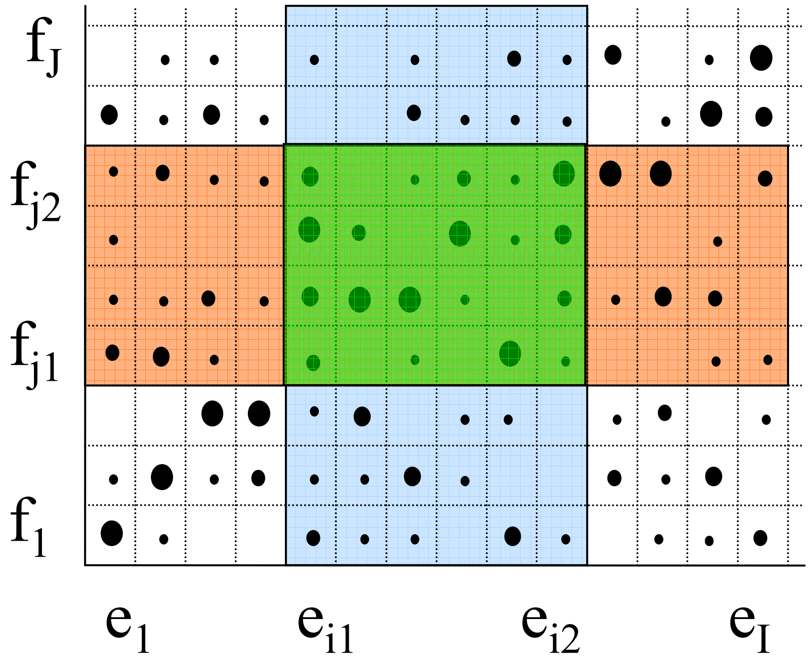


PESA Alignment

- Probability for this split:

$$\prod_{j=j_1}^{j_2}\left(\sum_{i=i_1}^{i_2} p(f_j \mid e_i)\right)$$

„Inside Alignment Probability"



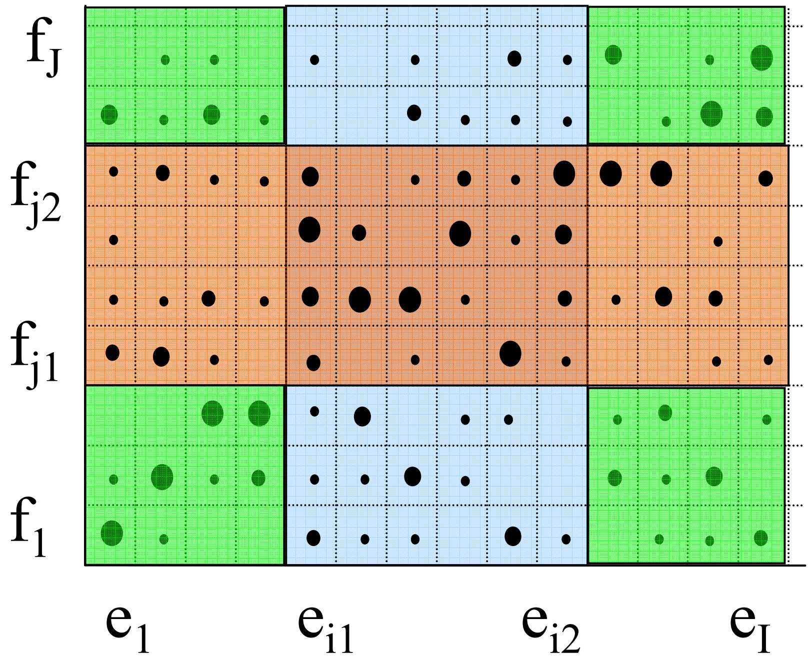
Word Alignment Matrix

- Probability for this split:

$$\prod_{j=j_1}^{j_2}\left(\sum_{i=i_1}^{i_2} p(f_j \mid e_i)\right)$$

$$*\prod_{j=1}^{j_1-1}\left(\sum_{i\notin(i_1\ldots i_2)} p(f_j \mid e_i)\right)$$

$$*\prod_{j=j_2+1}^{J}\left(\sum_{i\notin(i_1\ldots i_2)} p(f_j \mid e_i)\right)$$



„Outside Alignment Probability“

PESA Alignment

- Optimize over target boundaries to find optimal split
- Look from both directions

$$p(f_j \mid e_i) \quad p(e_i \mid f_j)$$

- Online phrase extraction
  - Phrases are extracted as needed during decoding process
  - No restriction on phrase length

LogLin Alignment

General idea:

LogLin extends idea of PESA by adding multiple features

e.g.

- Word alignment
- Fertility (Phrase length)
- Relative position in sentence pair
- Lexical features (IBM-1)

- Some feature functions might overlap

⇨ Framework of Log Linear Model is applied

LogLin Alignment

- Log-linear model to combine more feature functions

$$Pr(X|\mathbf{e},\mathbf{f}) = \frac{exp(\sum_{m=1}^{M} \lambda_m \phi_m(X,\mathbf{e},\mathbf{f}))}{\sum_{\{X'\}} exp(\sum_{m=1}^{M} \lambda_m \phi_m(X',\mathbf{e},\mathbf{f}))}$$

Feature functions

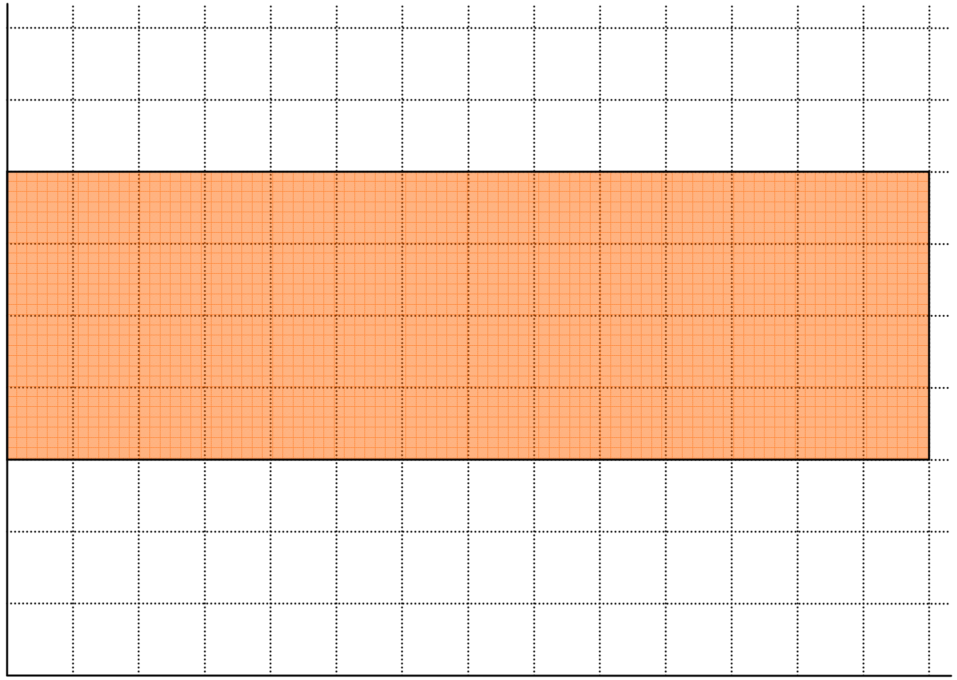
Model parameters: weights of the feature functions

- (**e**,**f**) is a sentence pair
- *X* is a phrase pair extracted from (e,f)

LogLin Alignment

2 Step approach:

1. Find candidates using simple heuristics
2. Score candidates using feature functions

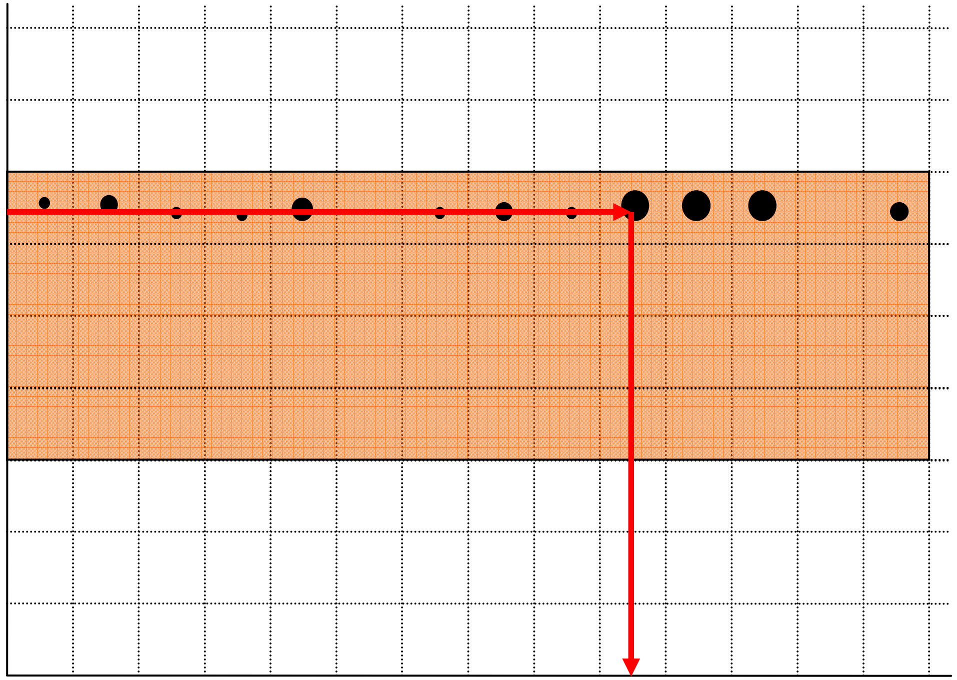


LogLin Alignment

Find projected center of target phrase

- For each source word:
  Find „center of gravity" of IBM1 probabilities

⇨ Projected center for this source word

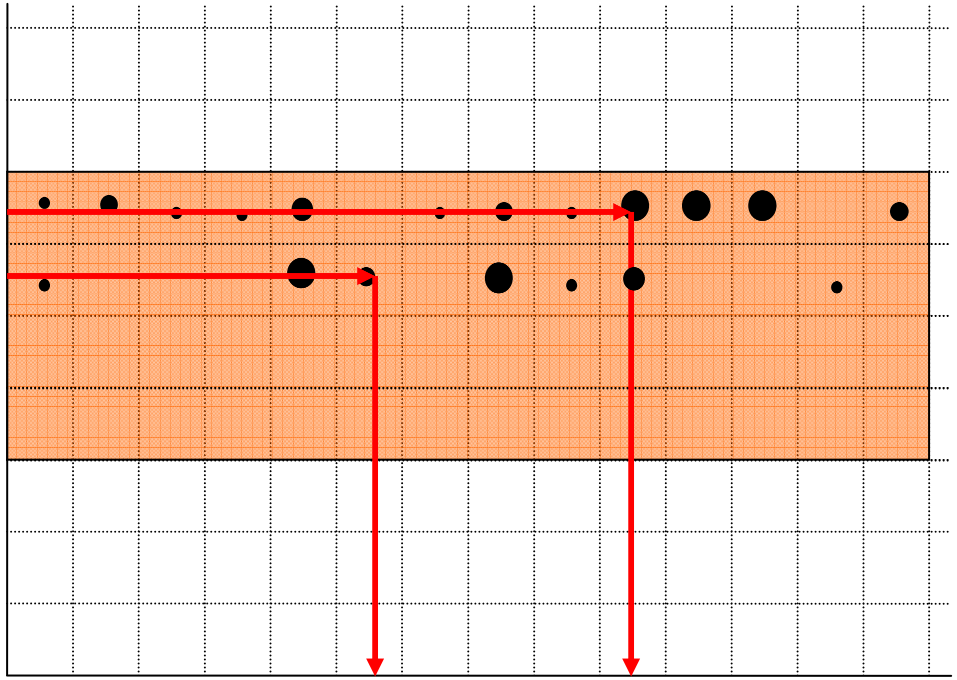


LogLin Alignment

Find projected center of target phrase

- For each source word:
  Find „center of gravity“ of IBM1 probabilities

⇨ Projected center for this source word

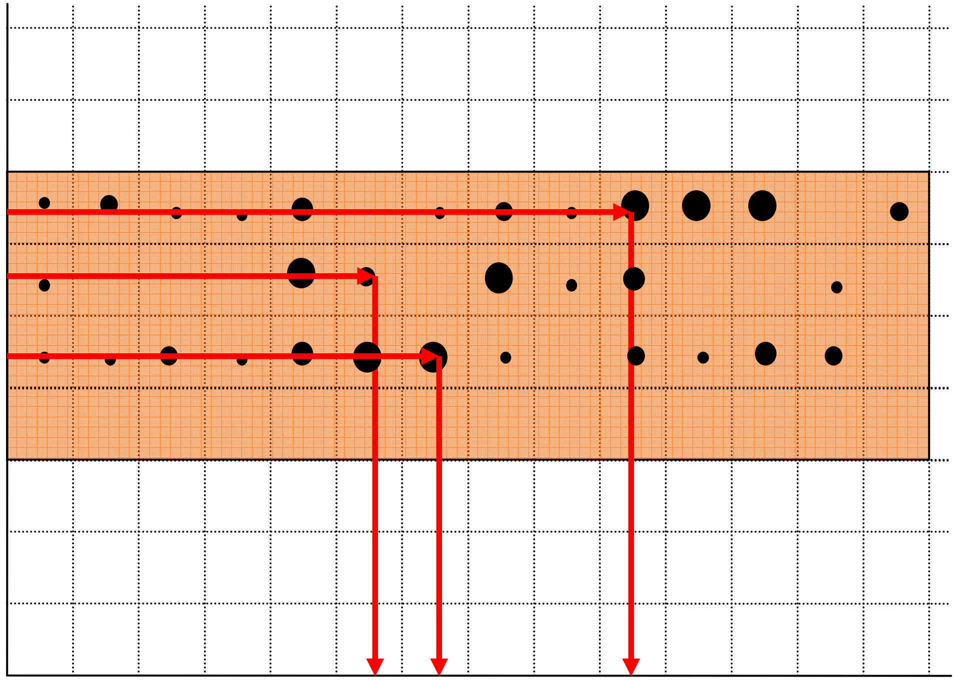


LogLin Alignment

Find projected center of target phrase

- For each source word:
  Find „center of gravity“ of IBM1 probabilities

⇨ Projected center for this source word

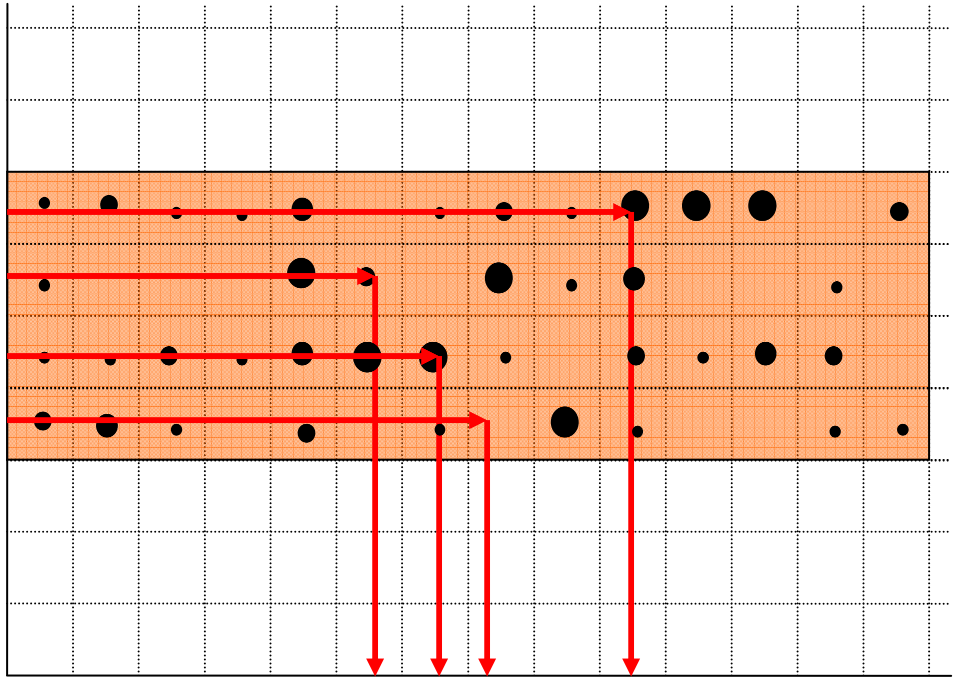


LogLin Alignment

Find projected center of target phrase

- For each source word:
  Find „center of gravity" of IBM1 probabilities

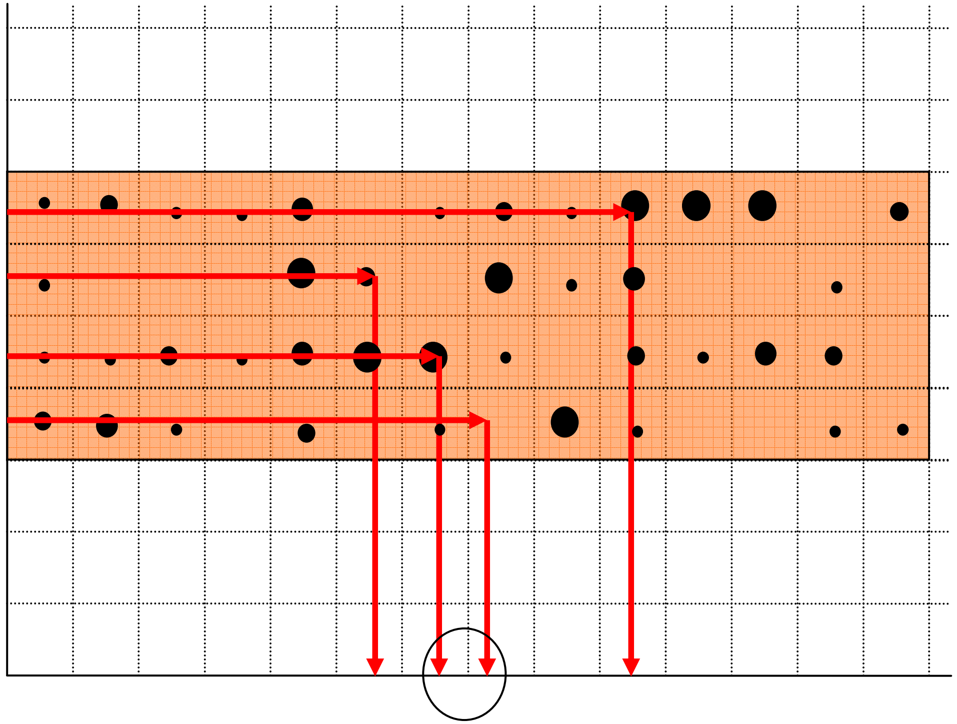
⇨ Projected center for this source word



LogLin Alignment

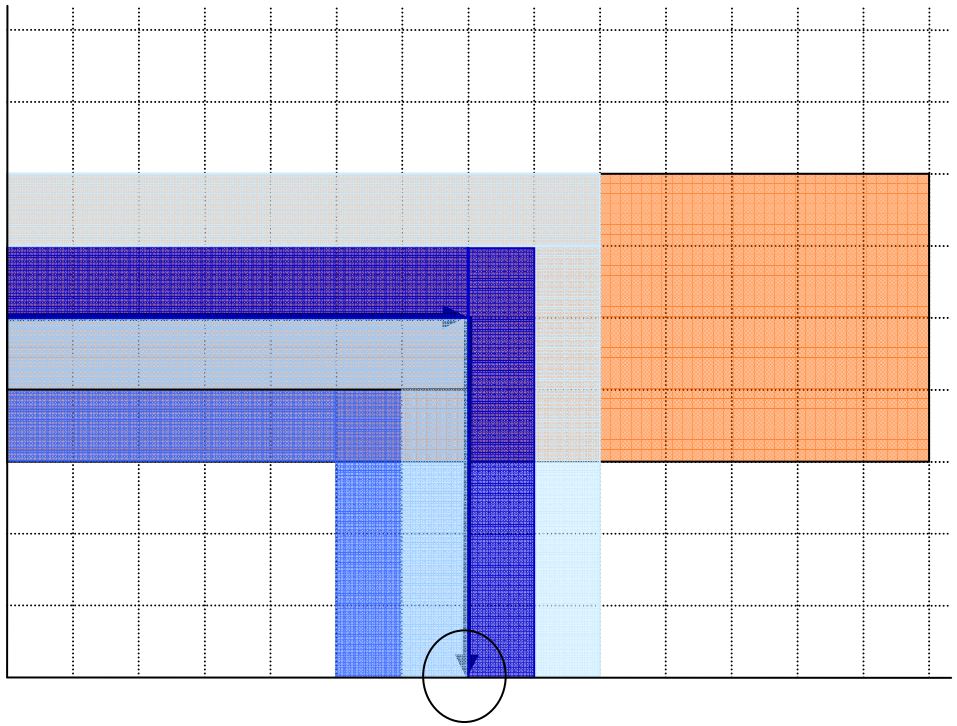
Find projected center of target phrase

- Average of centers to get projected target center for source phrase



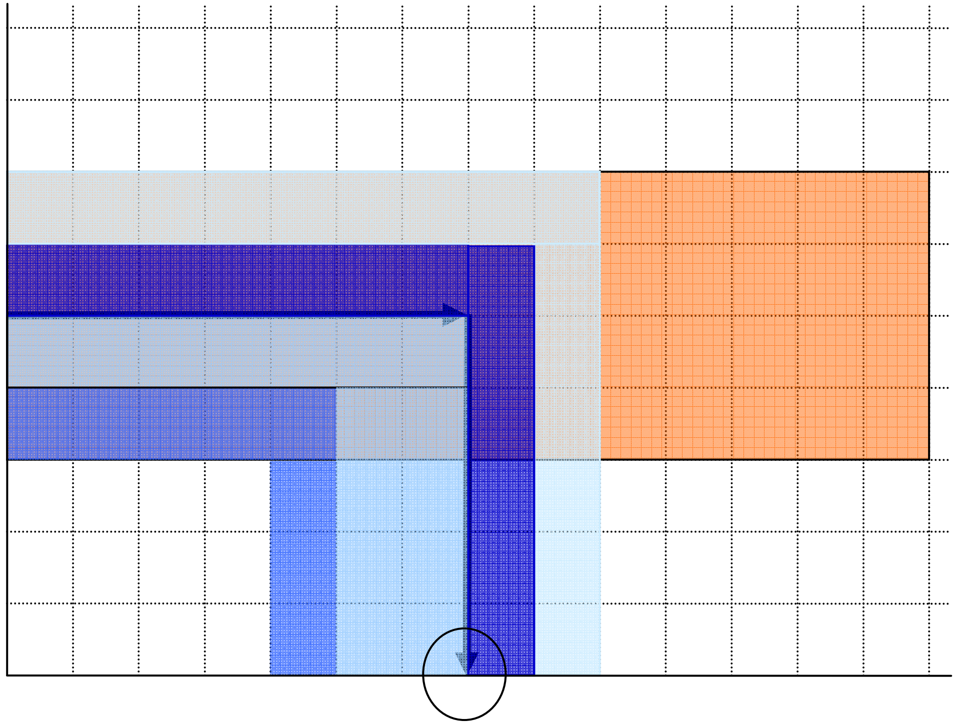
LogLin Alignment

- Predict target length using IBM-4 fertilities



LogLin Alignment

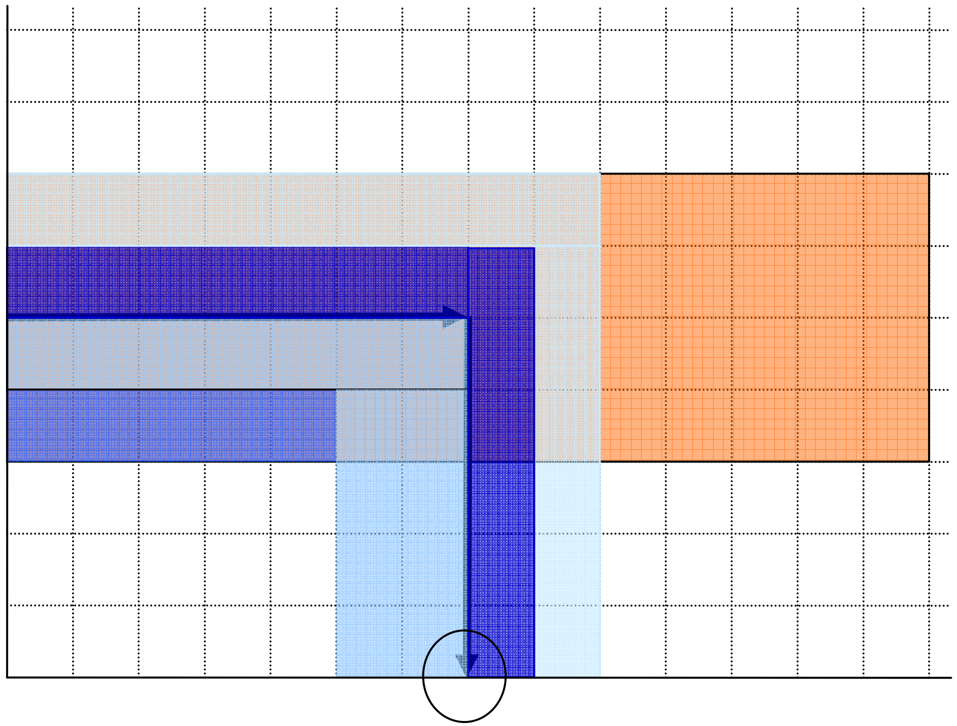
- Predict target length using IBM-4 fertilities



LogLin Alignment

- Predict target length using IBM-4 fertilities
- Generate candidates using the predictions for center and target length
- Target phrase does not have to have the projected center in the middle but it has to contain it

⇨ First step generates a (relatively small) number of phrase translation candidates

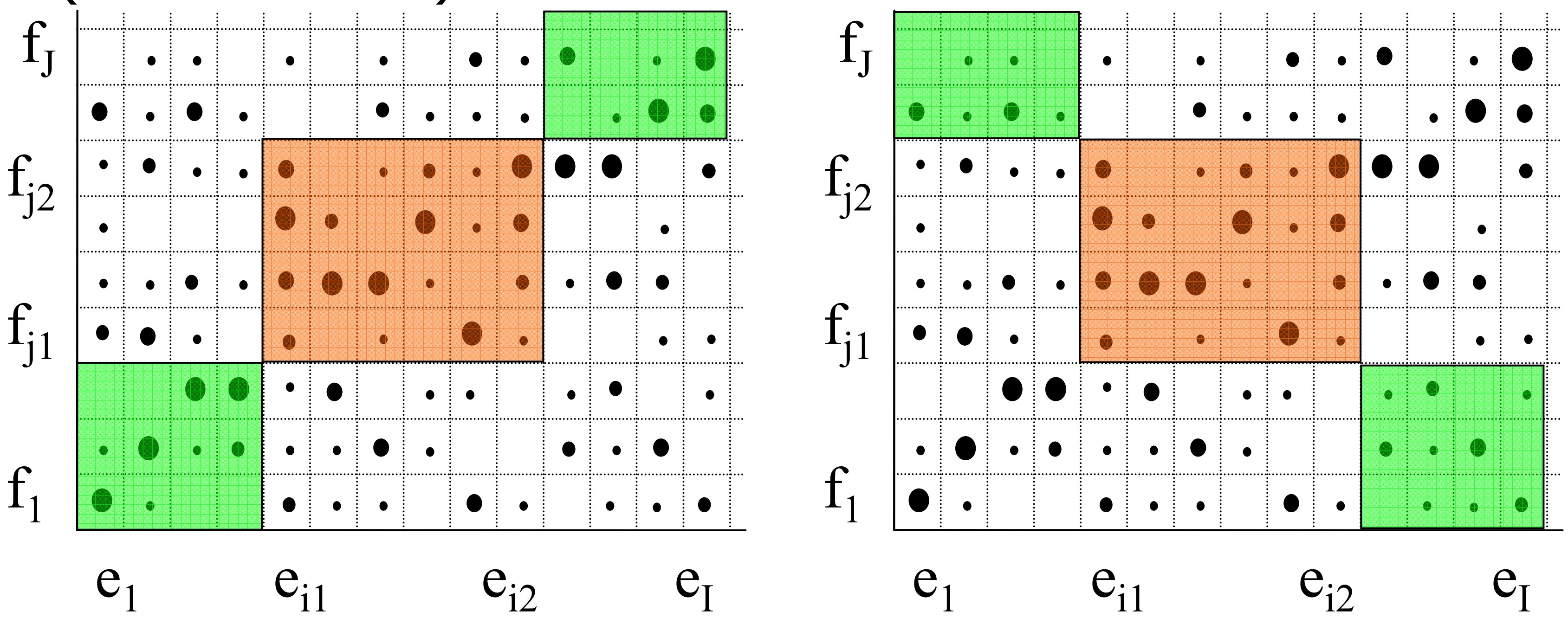


13 Features for candidate scoring

- 4: Phrase-level length relevance
  - Source phrase generates target phrase of this length
  - „Rest of sentence“ generates „Rest of sentence“ of this length
  - + reverse direction
- 4: IBM Model-1 scores
  - similar to PESA
  - Source phrase generates target phrase
  - „Rest of sentence“ generates „Rest of sentence“
  - + reverse direction

13 Features for candidate scoring

- 4: Bracket the sentence pair diagonal and inverse diagonal (both directions)



- 1: average alignment links per source word
  - Every block should contain at least one word alignment from the Viterbi path

Feature weights

- Weights for each feature function are learned using human aligned *„gold standard phrase pairs"*
- Weights are adjusted to optimize accuracy on these phrases

Problems:

- For BTEC data no human word alignment available to extract gold-standard phrase pairs
- Used previously trained weights (Chinese – English newswire data)

⇨ Should work reasonably well on Chinese BTEC

⇨ Questionable on other language pairs

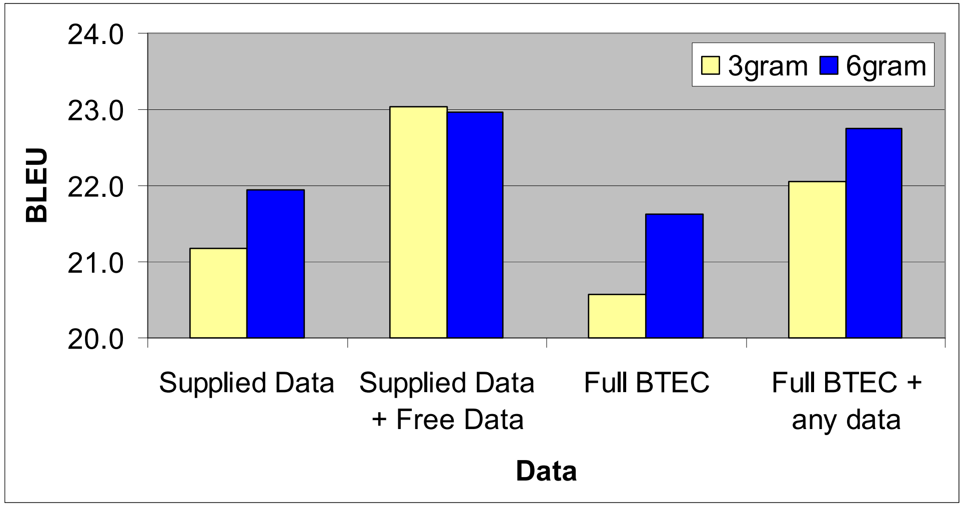
- Overfitting possible due to overlapping features

Language Model

2 Options:

- 3-gram SRI language model (Kneser-Ney discounting)
- 6-gram Suffix Array language model (Good-Turing discounting)

- 6-gram consistently gave better results
- Only used 6-gram LM



Decoding

2 stage decoding process

- Build translation lattice using the extracted phrase pairs
- Search for best path through lattice
- Word reordering possible within reordering window (best results at ~4-5)

- ASR output translation:
  Only translated 1best

Italian – English results

Open Track

- 20k lines supplied data

C-STAR Track

- 55k lines „Full BTEC"
- 3k lines web data (travel phrases)

	Open Track		**C-STAR Track**	
	BLEU	NIST	BLEU	NIST
PESA	**0.2388**	**6.20**	**0.2630**	**6.66**
LogLin	0.2719	6.61	0.2912	7.08

Arabic – English results

Open Track

- 20k lines supplied data

C-STAR Track

- 20k lines supplied data
- 20k lines additional translated BTEC
- 31k lines typed travel books (English)

	Open Track		**C-STAR Track**	
	BLEU	NIST	BLEU	NIST
PESA	0.1908	5.38	0.1989	5.62
LogLin	**0.1995**	**5.34**	**0.2123**	**5.87**

Chinese – English results

Open Track

- 40k lines supplied data

C-STAR Track

- 163k lines Full BTEC
- 106k lines newswire data (gathered with IR technique)
- 31k lines typed travel books (English)

		Open Track		**C-STAR Track**	
		BLEU	NIST	BLEU	NIST
PESA	read	0.1501	4.87	**0.1622**	**5.19**
	spont	0.1654	5.08	**0.1645**	**5.24**
LogLin	**read**	**0.1630**	**4.97**	**-**	**-**
	spont	**0.1710**	**5.08**	**-**	**-**

Japanese – English results

Open Track

- 40k lines supplied data

C-STAR Track

- 163k lines Full BTEC
- 4k medical dialogs

	Open Track		**C-STAR Track**	
	BLEU	NIST	BLEU	NIST
PESA	**0.1868**	**5.63**	**0.1841**	**5.40**
LogLin	0.1830	5.93	-	-

Chinese – English

Influence of additional data

- tested with PESA alignment:

	Supplied Data	**Supplied Data + IR data**	
spont.	0.1393	0.1501	+7.8%
read	0.1539	0.1654	+7.5%

	Full BTEC	**Full BTEC + IR data + travel books**	
spont.	0.1388	0.1622	+16.9%
read	0.1439	0.1645	+14.9%

Analysis

Chinese and Japanese:

- No improvements
  Open Data Track ⇨ C-STAR Data track

Alignment problem with Full BTEC data for Chinese - English

Word segmentation problems:

- Provided segmentation could not be used for the C-STAR Data track ⇨ Re-segmentation was necessary
- Worse word segmentation quality especially on ASR output

Word segmentation - Japanese

Provided-segmentation

- **ASR:** 御　荷物 は　に　持つ　引き取り　と　に　ございます **(3-errors)**
- **REF:** 御　荷物　は　荷物　引き取り　所　に　ございます
- **3-ASR errors ⇨ 3 segmentation errors**

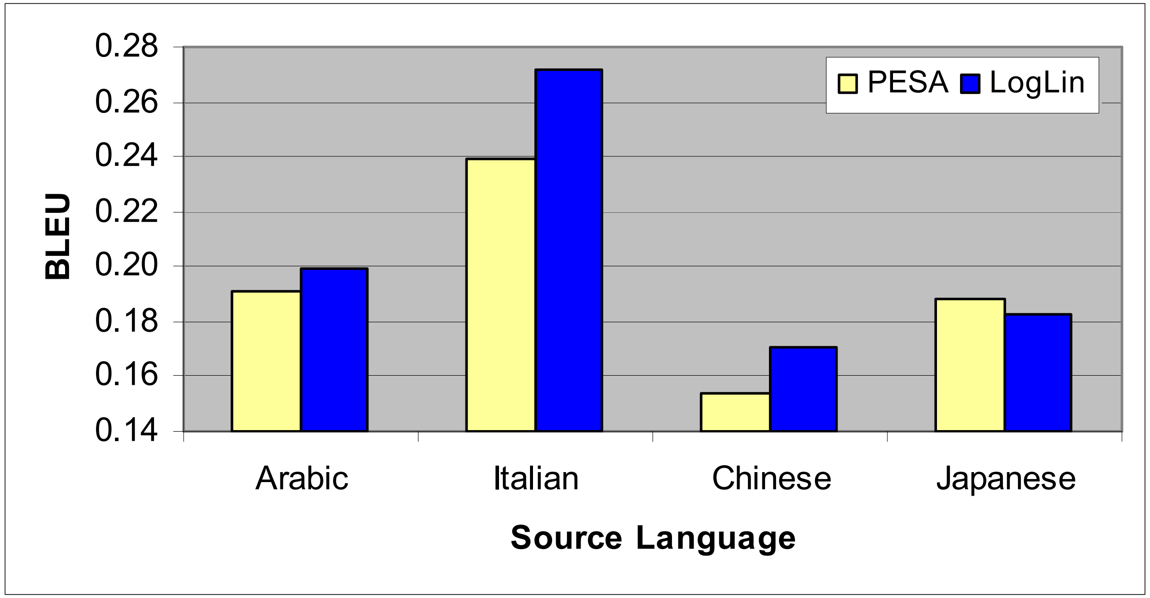
MeCab-segmentation (used on C-STAR track)

- **ASR:**御　荷物　は　に　持つ　引き　取り　と　に　ござい　ます (**5-errors)**
- **REF:**御　荷物　は　荷物　引き取り　所　に　ござい　ます
- **3-ASR errors ⇨ 5 segmentation errors**

	BLEU (% degradation)	
Word Segmentation	**Provided**	**MeCab**
Transcriptions	24.3	23.5
ASR Output	21.1 (13%)	19.6 (17%)

Analysis: Phrase alignments

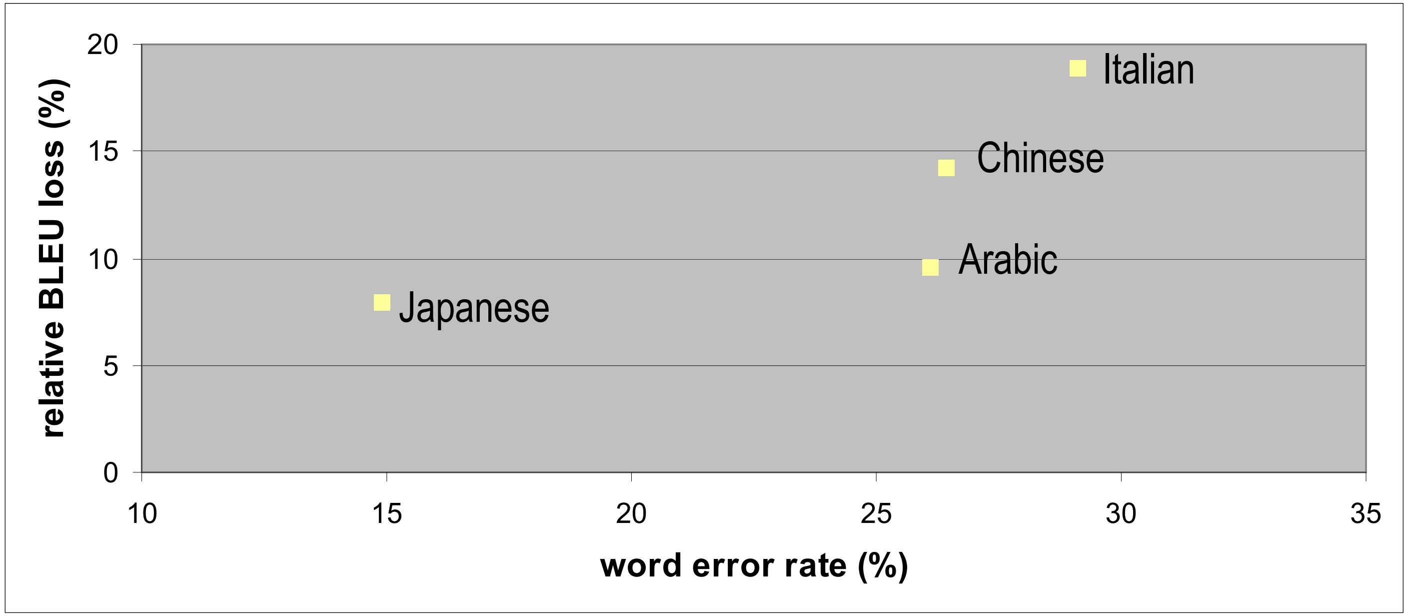
- LogLin outperforms PESA on Chinese, Arabic
- Best improvements on Italian (+0.03 BLEU)
- Slight drop on Japanese



Analysis BLEU - WER

- Correlation BLEU degradation
  CRR ⇨ ASR
  with WER of ASR output

	CRR	ASR (read)		WER
Japanese	0.2030	0.1868	**-8.0%**	14.9%
Arabic	0.2208	0.1995	**-9.6%**	26.1%
Chinese	0.1996	0.1710	**-14.3%**	26.4%
Italian	0.3353	0.2719	**-18.9%**	29.1%



Future Work

- Use lattice/nbest information for translation of ASR output
- Provide LogLin with better hand-aligned data (in-domain) in different languages
- Limit influence of overfitting