The MITLL/AFRL IWSLT-2006 MT System

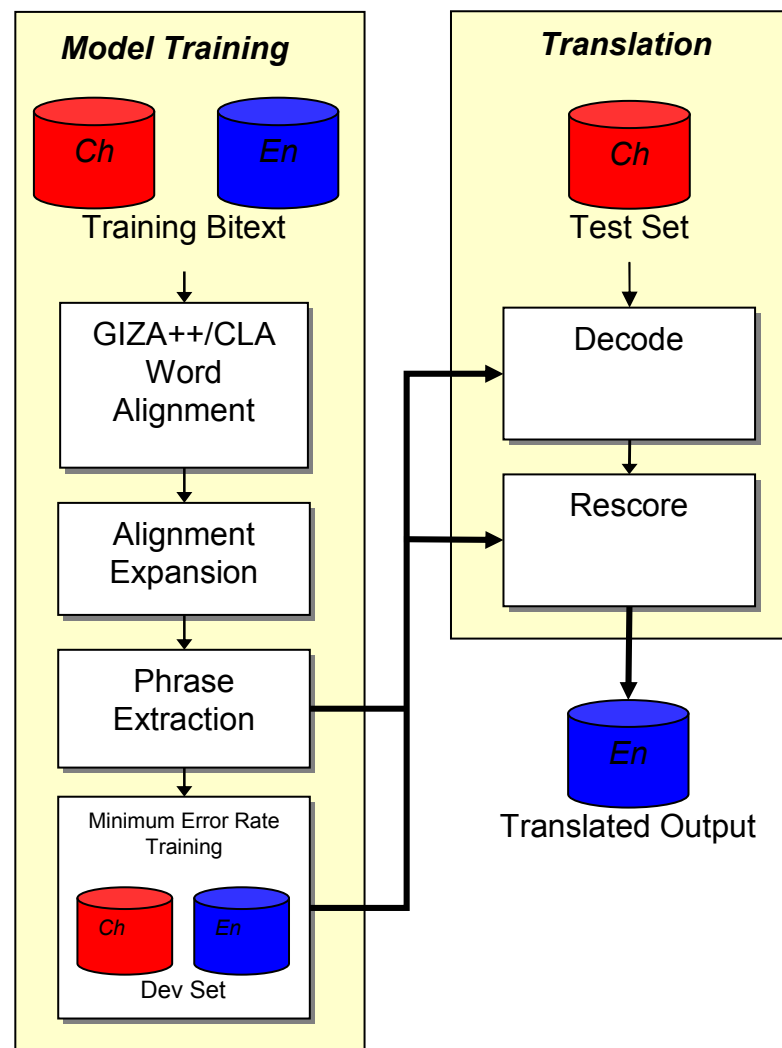**Wade Shen, Brian Delaney, and Tim Anderson**

27 November 2006

MIT Lincoln Laboratory

Statistical Translation System
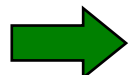## *Experimental Architecture*

- **Standard Statistical Architecture**

- **New this year**
  - **Improved Training**
  - **Better Decoders**
    - **internal decoder**
    - `moses`
  - **Rescoring Features**

- **Participated in**
  - **Chinese ⇨ English**
  - **Japanese ⇨ English**
  - **Italian ⇨ English**

- **Focus on Chinese ⇨ English**

Model Training

Ch   En

Training Bitext

GIZA++/CLA Word Alignment

Alignment Expansion

Phrase Extraction

Minimum Error Rate Training

Ch   En

Dev Set

Translation

Ch

Test Set

Decode

Rescore

En

Translated Output

The MITLL/AFRL MT System
Overview

- **Improved Training**
  - CLA and GIZA++ alignments
  - Factored Models: Hybrid Alignment templates

- **Better Decoding**
  - Reordering Constraints
  - Higher-order LMs

- **Development Experiments**
  - Segmentation and Alignment
  - Rescoring
  - Pre/Post-processing

- **Evaluation Results**
  - Manual Transcription
  - ASR Transcription

- **Next Steps**

Improved Training
Segmentation and Alignment

- **Added CLA Alignments**
  - **Assumes 1-1 alignment**
  - **Poor quality phrases without IBM models**
  - **Simple Greedy Algorithm [Melamed 2000]**
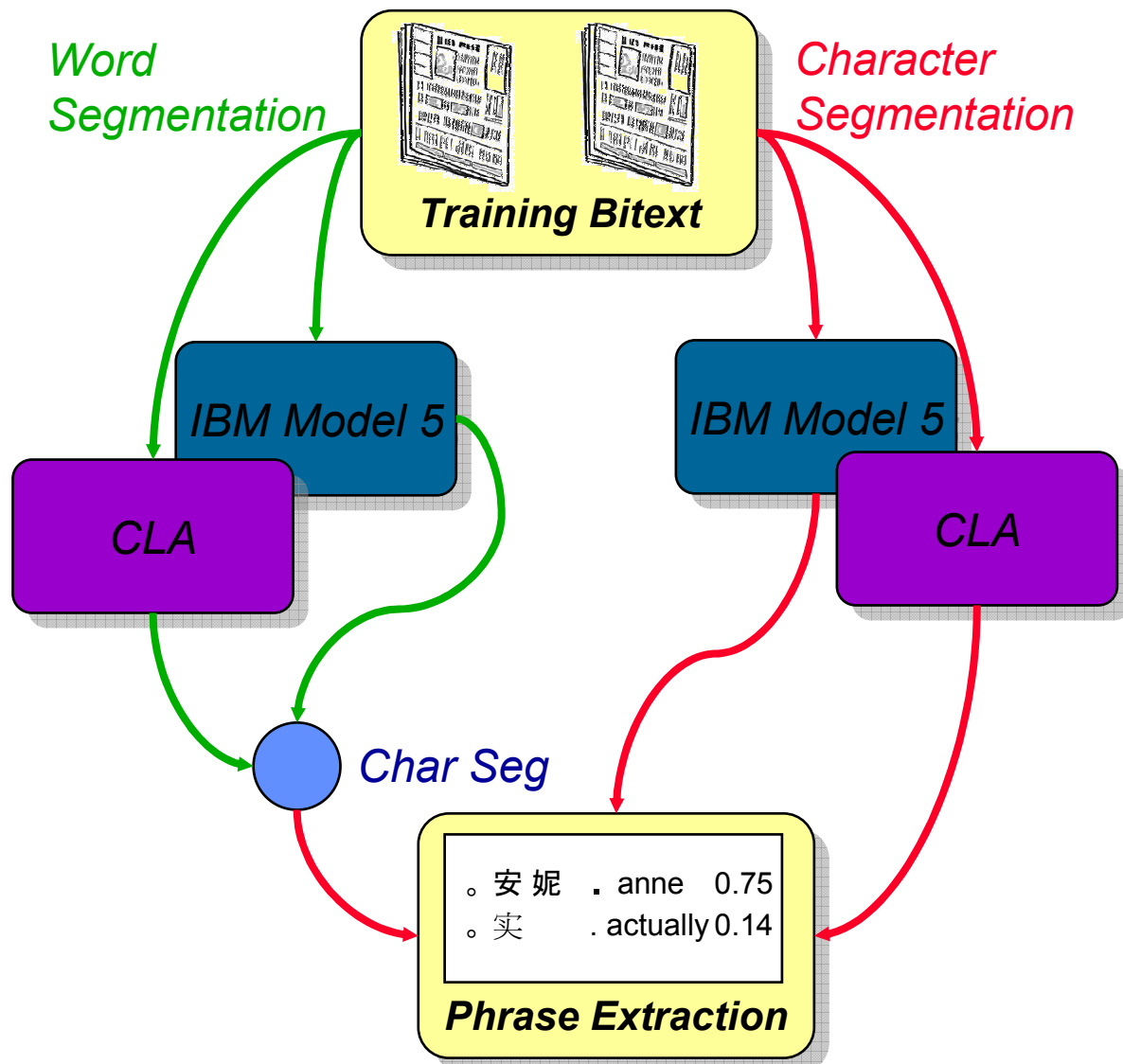 - **Simple symmetric Likelihood function**

$$L(e, f) = X_{ef}^2 = \sum_{s=i}^{N} \frac{(x_s - E_{ef})^2}{E_{ef}}$$

	Open	up	.
打	0.2	0.8	0.0
开	0.7	0.1	0.2
。	0.1	0.1	0.8

Improved Training
Alignment and Segmentation

Word Segmentation

Character Segmentation

Training Bitext

IBM Model 5

IBM Model 5

CLA

CLA

Char Seg

。安妮　． anne　0.75
。实　　． actually 0.14

Phrase Extraction

- **Phrase extraction**
  - **Word and Character Segmentation**
  - **CLA and IBM Model 5 alignments**
  - **Resegment Words (Chinese)**

- **All decoding done with char-seg (Chinese)**

Seg/Align	Word	Char
Model 5	**ALL**	**Chinese**
CLA	**ALL**	**Chinese**

MIT Lincoln Laboratory

Improved Training
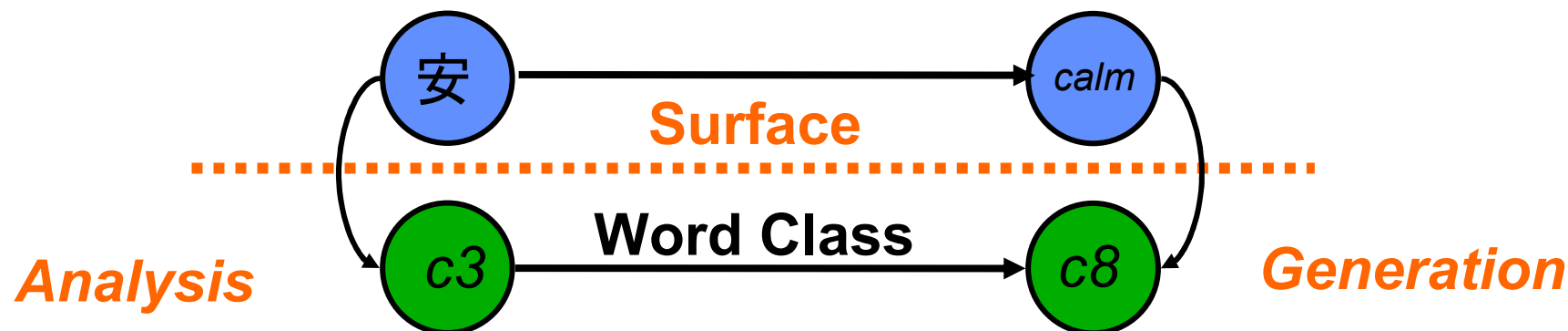## *Hybrid Alignment Template Models*

- **Apply Factored Models using Unsupervised Word Classes**
  - Words represented by different linguistic factors
  - Factors can be translated separately
  - LMs can be applied to factors independently

- **Can we use automatic word classes to learn general sequence constraints?**

- **First Experiment:** *Word Class LMs of varying orders*

Target Phrase

How much is it?

Source Phrase

总共 多少 钱 ?

c1 c22 c3 c55

word

class
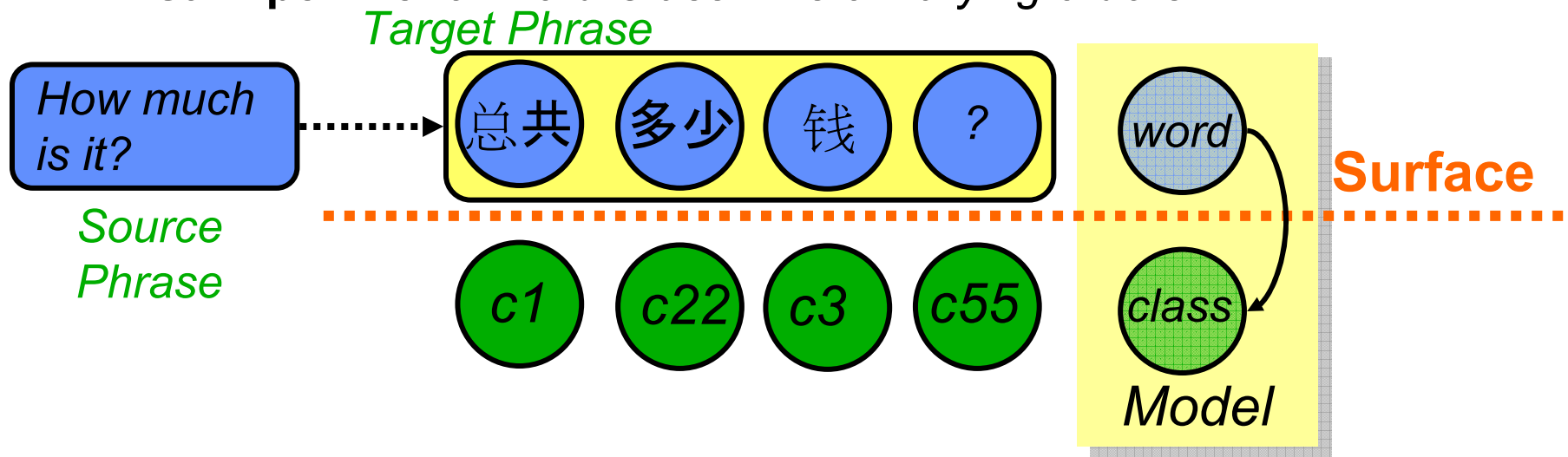
Surface

Model

IWSLT Chinese
Alignment Templates for Translation

安 → calm
Surface

Analysis c3 → c8 **Generation**
Word Class

- **Second Experiment:** *Extend Class-based LM to the translation Model*

- **Bigram word classes for source and target**

- **Translate alignment templates similar to [Och 98] + surface**

- **Apply LM to surface and Class**
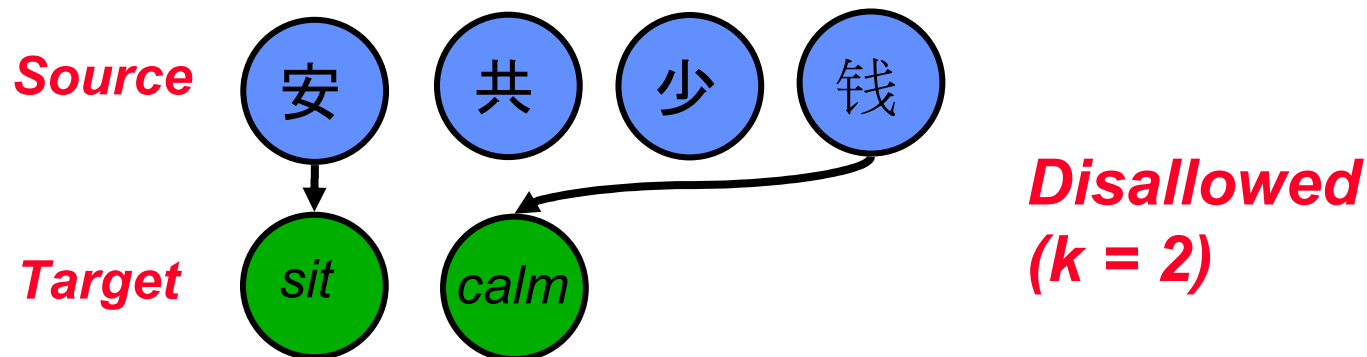
The MITLL/AFRL MT System
Overview

- **Improved Training**
    - CLA and GIZA++ alignments
    - Factored Models: Hybrid Alignment templates

- **Better Decoding**
    - Reordering Constraints
    - Higher-order LMs

- **Development Experiments**
    - Segmentation and Alignment
    - Rescoring
    - Pre/Post-processing

- **Evaluation Results**
    - Manual Transcription
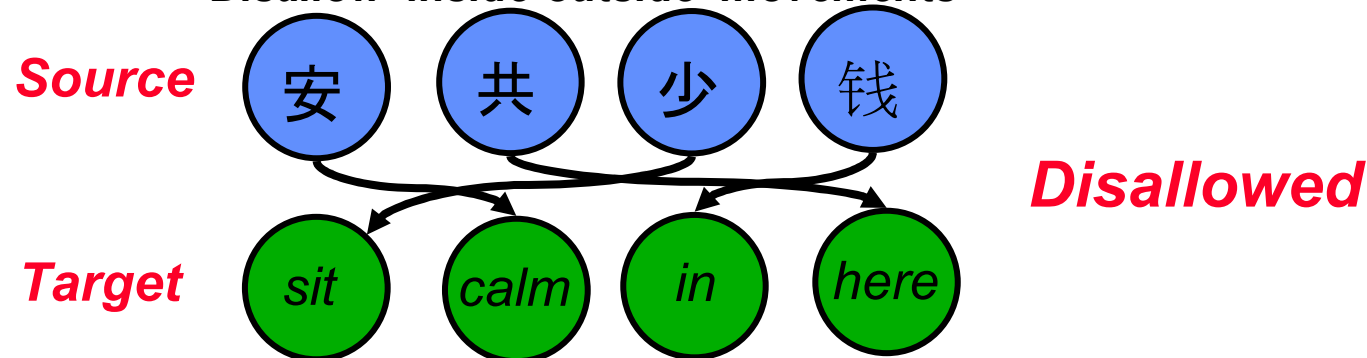   - ASR Transcription

- **Next Steps**

MIT Lincoln Laboratory

Better Decoding
Reordering Constraints

- **Revisit Reordering Constraints [Zens 03]**
  - **IBM Constraints:** *Must cover first K uncovered positions*

Source 安 共 少 钱

Target sit calm

Disallowed (k = 2)

  - **ITG Constraints [Wu 95, 97]**

Disallow 'inside-outside' movements

Source 安 共 少 钱

Target sit calm in here

Disallowed

- **Faster, some BLEU improvement in Japanese**
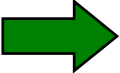
Better Decoding
Higher Order LMs

- **Used both `moses` and internal decoder for IWSLT-2006**

- **Both modified to support n-gram LMs > order 3**

- **Consistent and Significant Gain with 4-gram LM**

- **Allowed**
  - **high-order Class LM (with `moses`)**
  - **integrated TrueCasing with 4-gram LM**

The MITLL/AFRL MT System
Overview

- **Improved Training**
  - CLA and GIZA++ alignments
  - Factored Models: Hybrid Alignment templates

- **Better Decoding**
  - Reordering Constraints
  - Higher-order LMs

- **Development Experiments**
  - Segmentation and Alignment
  - Rescoring
  - Pre/Post-processing

- **Evaluation Results**
  - Manual Transcription
  - ASR Transcription

- **Next Steps**

Development Experiments
Segmentation and Alignment

- **All Results reported on Dev-4**

- **Official Eval Condition**
  - Mixed-case + Punctuation

- **Optimized using Dev-1**
  - With Rescoring Features
  - Development done w/Chinese

Japanese

Configuration	BLEU
Word Segmented	23.63
Word Seg + Char Backoff	23.82
Word Segmented + CLA	23.34
Word + Char Backoff + CLA	24.28

Chinese

Configuration	BLEU
Character Segmented	21.24
Word + Char Seg	21.21
Word Segmented	21.01
Word + Char Seg + CLA	22.18

Italian

Configuration	BLEU
Word Segmented	35.13
Word Segmented + CLA	37.40

Development Experiments
Rescoring

Chinese

Configuration	BLEU
Baseline 4-gram Decode	21.39
+ 5-gram rescore LM	21.55
+ 6-gram class-based LM	21.52
+ Model 1	21.86
+ Sentence Length Posterior*	22.10

Japanese

Configuration	BLEU
Baseline 4-gram Decode	21.92
All Rescoring Features	24.28

Italian

Configuration	BLEU
Baseline 4-gram Decode	36.92
All Rescoring Features	37.40

- **Eval Configuration**
  - 5-gram rescore
  - 6-gram class-based LM
  - Model 1

* Length posterior added post-eval

Development Experiments
Pre/Post-Processing

Chinese

Configuration	BLEU
Remove Source Punctuation	21.86
Re-punctuated Source	
TrueCase with 3gram	21.93
TrueCase with 4gram	22.10
TrueCase with 5gram	22.10

Japanese

Configuration	BLEU
Remove Source Punctuation	23.14
Re-punctuated Source	24.28

Italian

Configuration	BLEU
Remove Source Punctuation	36.64
Re-punctuated Source	37.40

- **Remove Source Punctuation**
  - **Truecased with 4gram LM**
- **Final Configuration**
  - **Repunctuate source**
  - **Truecase with 4gram**

Development Experiments
Hybrid Alignment Template

Chinese (dev-4)
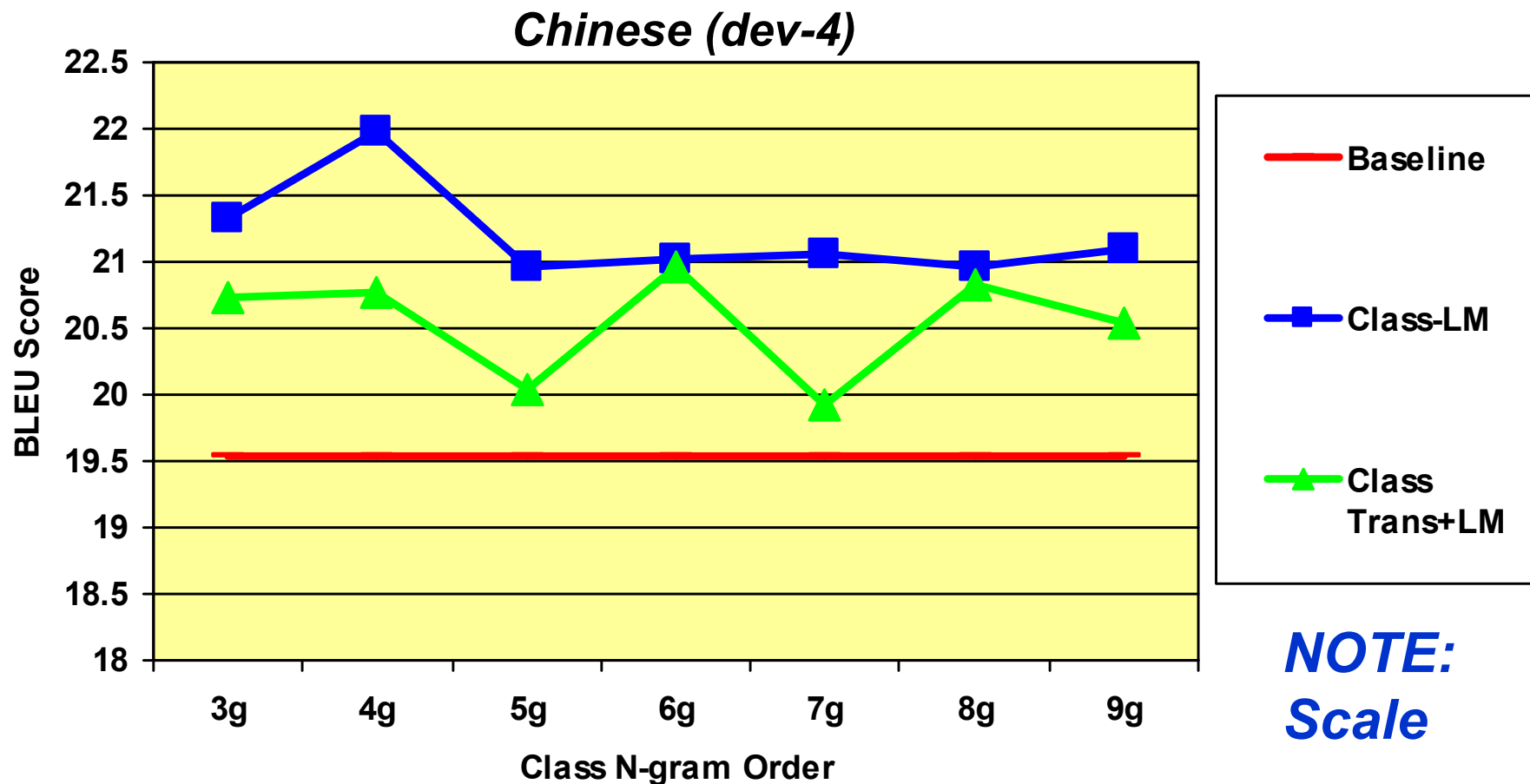
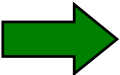- **Baseline: word segmented, 3-gram no rescore**
- **Class-LM significantly better (p=0.05, ~1.0 BLEU)**
- **Class-Trans may be limited by synchronous PT constraint**

NOTE: Scale

The MITLL/AFRL MT System
Overview

- **Improved Training**
  - CLA and GIZA++ alignments
  - Factored Models: Hybrid Alignment templates

- **Better Decoding**
  - Reordering Constraints
  - Higher-order LMs

- **Development Experiments**
  - Segmentation and Alignment
  - Rescoring
  - Pre/Post-processing

- **Evaluation Results**
  - Manual Transcription
  - ASR Transcription

- **Next Steps**

Evaluation Results
Manual Transcription

Chinese

Configuration	BLEU
Optimized (Dev Set 4)	21.57
Optimized (Dev Set 1)	20.66
No Rescore (Opt Dev4)	21.27

Japanese

Configuration	BLEU
Optimized (Dev Set 4)	20.99
Optimized (Dev Set 1)	20.24

- **Rank**
  - **Chinese: 2nd of 13**
  - **Italian: 2nd**
  - **Japanese: 4th**

Italian

Configuration	BLEU
Optimized (Dev Set 4)	35.74
Optimized (Dev Set 1)	34.40

Evaluation Results
ASR Transcription

Chinese (Read)

Configuration	BLEU
1-best ASR (Opt Dev4)	18.61
10-best ASR (Opt Dev4)	17.42
1-best ASR (Opt Dev1)	18.46

Japanese

Configuration	BLEU
1-best ASR (Opt Dev4)	18.91
1-best ASR (Opt Dev1)	18.43

Chinese (Spontaneous)

Configuration	BLEU
1-best ASR (Opt Dev4)	16.57
10-best ASR (Opt Dev4)	16.57*

Italian

Configuration	BLEU
1-best ASR (Opt Dev4)	27.98
10-best ASR (Opt Dev4)	28.81
1-best ASR (Opt Dev1)	27.64
10-best ASR (Opt Dev1)	27.81

- Rank
  - Chinese: 3rd of 13
  - Italian: 4th
    (secondary system: 2nd)
  - Japanese: 4th

The MITLL/AFRL MT System
Overview

- **Improved Training**
  - CLA and GIZA++ alignments
  - Factored Models: Hybrid Alignment templates

- **Better Decoding**
  - Reordering Constraints
  - Higher-order LMs

- **Development Experiments**
  - Segmentation and Alignment
  - Rescoring
  - Pre/Post-processing

- **Evaluation Results**
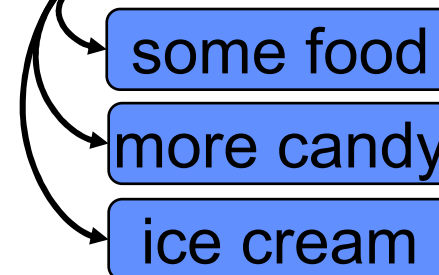 - Manual Transcription
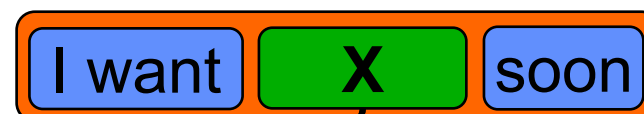 - ASR Transcription
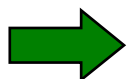
- **Next Steps**

Next Steps

- **Work with MIT/SLS on**
  - **Hybrid hierarchical translation for Chinese**

Current Model

| I want some | food soon |

New Model

| I want | **X** | soon |

some food

more candy

ice cream

- **Better sparse data models with Factors**
  - **Asynchronous Factor Decoding**
    - **Allow longer class templates**
  - **Integration of POS and morphology models**

安

him

Surface

Analysis

Lemma

he

CoGeneration

POS + case

pn

acc
pn