



Context-Dependent Alignment Models

Jamie Brunning, Adria de Gispert, Bill Byrne
Cambridge University

Context Dependent Translation in Alignment



Target Words	e_1	e_2	...	e_j
Source Words	f_1	f_2	...	f_j
Source Context	c_1	c_2	...	c_j

$$P(e_1^j, a_1^j | f_1^j, c_1^j) \approx \prod_{i=1}^j T(e_i | f_{a_i}, c_{a_i})$$

- **Basic models rely on context-independent word-to-word translation -- Models 1, 2, ..., HMMs and variants**
- **Not easy to introduce translation dependency on context**
 - need to control model size and computational complexity
- **New: Context clustering via decision trees**
 - Each source word gets its own tree (many, many trees...)
 - Questions are asked about the source word contexts
 - Clustering is performed during iterative EM training
 - EM auxiliary function is used as the clustering purity function
 - identical to HTK acoustic clustering

Context-dependent alignment models for statistical machine translation.

Context-dependent alignment models: Clustering by Part of Speech tags



- POS taggers were run on both sides of the parallel texts
- Created question sets for POS contexts
 - questions covered 5-word contexts
 - Grew four sets of trees: two for Ar/Zh->En and two for En->Ar/Zh

Most frequent root node questions for Ar-En

English Question	Count	Arabic Questions	Count
Is_Next_Preposition	1523	Is_Prev_Preposition	1110
Is_Prev_Determiner	1444	Is_Next_Preposition	993
Is_Prev_Preposition	1209	Is_Prev_Noun	981
Is_Prev_Adjective	864	Is_Prev_Coordinating_Conj	627

- 26% and 40% of Arabic and English words have CD translations
- Question combinations get fairly complex further down the tree

CD models improve both alignment and translation quality



- CD Model 1: Large reductions in Alignment Error Rate
- CD HMMs: Improvements in alignment and in BLEU

BLEU scores on development training and test sets

	Ar - En		Zh - En	
Alignment	mt02-05test	MT08-nw	test-nw	MT08-nw
CI-HMM	49.4	46.3	28.5	26.9
CD-HMM	49.7	46.9	29.0	27.7

- Using decision trees for context dependent translation models scales well to larger data sets:
 - Parallelized implementation using MTTK alignment tools
 - Good gains on the full AGILE training sets (0.5 -- 1.0 BLEU)



Hierarchical Phrase-based Translation with Weighted Finite State Transducers

Adria de Gispert, Bill Byrne

Cambridge University

Gonzalo Iglesias, Eduardo Banga

University of Vigo

HiFST: WFST implementation of Hiero



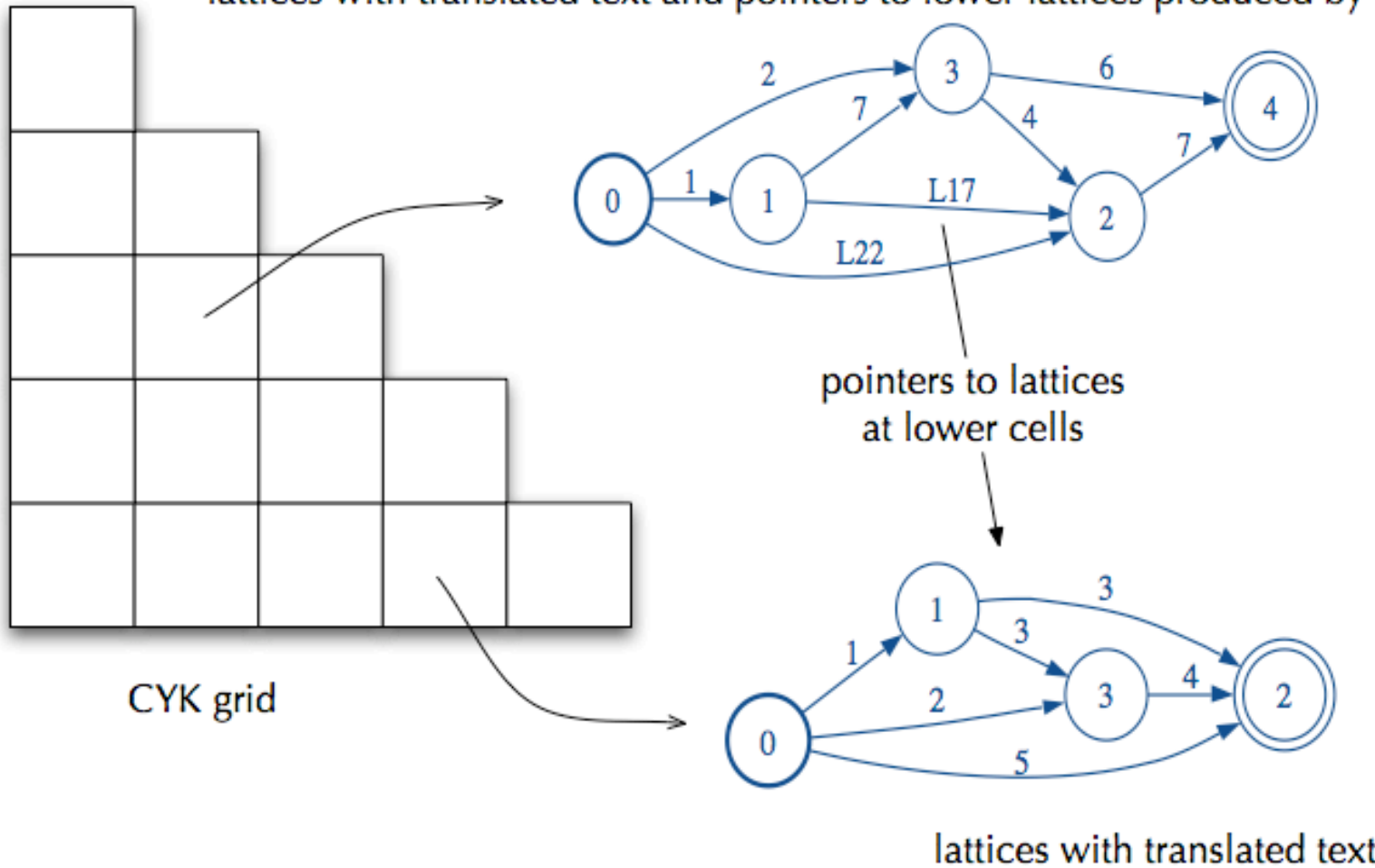
- **WFST formulation of Hiero-style translation**
 - Alternative implementation to Cube Pruning (CPH)
 - Unioning/concatenating/pruning/etc. is done with standard WFST ops
 - Based on Google OpenFST toolkit
- **Key idea: work with lattices containing many hypotheses and avoid working with individual translation hypotheses**
 - vastly larger hypothesis spaces
- **Source language sentence is parsed using source-side of the Hiero translation rules**
 - CYK grid is built up and applicable rules are kept in each cell
- **A translation lattice is built for every cell, based on its rules**
 - lattice arcs have translated text or pointers to other lattices
 - these intermediate lattices can be minimized, determinized, etc.
 - lattice pointers are only expanded when translation is complete or if intermediate pruning is required
 - fast, and controls memory usage

Hierarchical phrase-based translation with weighted finite state transducers.
G. Iglesias, A. de Gispert, E. R. Banga, and W. Byrne. NAACL-HLT, 2009.

Delayed Translation



lattices with translated text and pointers to lower lattices produced by hierarchical rules



Translation improves with denser search spaces



- **More efficient search:**
 - 48% reduction in search errors for ZH->EN translation
 - AR->EN translation runs without pruning
- **Direct generation of translation lattices**
- **Richer search space improves subsequent rescoring**
 - HiFST is comparable to CPH in first pass [(a) vs (c)]
 - HiFST produces richer/better hypotheses for rescoring
 - increased gains from 5Gram LM + Minimum Bayes Risk rescoring [(b) vs (d)]

	Decoder	ZH->EN		AR->EN	
		test-nw	NIST MT08	mt02-05-test	NIST MT08
		BLEU / TER		BLEU / TER	
(a)	CPH	32.2 / 59.9	27.2 / 60.2	51.5 / 42.2	42.5 / 48.6
(b)	CPH +5G+MBR	32.7 / 59.4	28.1 / 59.3	52.6 / 41.4	43.4 / 48.1
(c)	HiFST	32.2 / 60.0	27.1 / 60.5	51.6 / 42.1	42.4 / 48.7
(d)	HiFST+5G+MBR	33.4 / 58.5	28.9 / 58.9	53.3 / 40.9	44.0 / 48.0