Multi-Engine Machine Translation (MT Combination)

Weiyun Ma

2012/02/17

Why MT combination?

- A wide range of MT approaches have emerged
  - We want to leverage strengths and avoid weakness of individual systems through MT combination

Scenario 1

Source:我 想要 蘋果
(I would like apples)

Sys1: I prefer fruit
Sys2: I would like apples
Sys3: I am fond of apples

Is it possible to select sys2:
“I would like apples”?

Sentence-based Combination

Scenario 2

Source:我 想要 蘋果
(I would like apples)

Sys1: I would like fruit
Sys2: I prefer apples
Sys3: I am fond of apples

Is it possible to create:
“I would like apples”?

Word-based Combination
Or
Phrase-based Combination

Outline

- Sentence-based Combination (4 papers)
- Word-based Combination (11 papers)
- Phrase-based Combination (10 papers)
- Comparative Analysis (3 papers)
- Conclusion

Abbreviations

- Evaluation Metrics
  - Bilingual Evaluation Understudy (BLEU)
    - N-gram agreement of target and reference
  - Translation Error Rate (TER)
    - The number of edits (word insertion, deletion and substation, and block shift) from target to reference
- Performance compared to the best MT system
  - BLEU:+1.2, TER:-0.8

Outline

- Sentence-based Combination
- Word-based Combination
- Phrase-based Combination
- Comparative Analysis
- Conclusion

Sentence-based Combination

Source:我 想要 蘋果
(I would like apples)
Sys1: I prefer fruit
Sys2: I would like apples
Sys3: I am fond of apples

Sentence-based Combination
(Selection)

sys2 – “I would like apples”

1. What are the features
for distinguishing translation quality?
2. How to model those features?

MT combination paper

MT paper

Features

*** Language model**
*** Translation model**
(* Agreement model)

Nomoto 2003

Hildebr and Vogel. 2008

***Syntactic model**

Zwarts and Dras. 2008

***Agreement model**

Kumar and Byrne. 2004

MT combination paper
MT paper

Features

*** Language model**
*** Translation model**
(* Agreement model)

Nomoto 2003

Hildebr and Vogel. 2008

***Syntactic model**

Zwarts and Dras. 2008

***Agreement model**

Kumar and Byrne. 2004

Sentence-based Combination

Nomoto 2003

- Fluency-based model (FLM): 4-gram LM
- Alignment-based model (ALM): lexical translation model - IBM model
- Regression toward sentence-based BLEU for FLM, ALM or FLM+ALM
- Evaluation: Regression for FLM is the best (Bleu:+1)

Hildebrand and Vogel. 2008

- Six Chinese-English MT systems (topN-prov, b-box)
- 4-gram and 5-gram LM, and lexical translation models (Lex)
- Two agreement models:
  - Position-dependent word agreement model (WordAgr)
  - Position-dependent N-gram agreement model (NgrAgr)
- Evaluation:
  - All features: Bleu:+2.3, TER:-0.4
  - Importance: LM>NgrAgr>WordAgr>Lex

Nomoto 2003 Predictive Models of Performance in Multi-Engine Machine Translation

Hildebrand and Vogel. 2008 Combination of machine translation systems via hypothesis selection from combined n-best lists

Sentence-based Combination

Nomoto 2003

- Four English-Japanese MT systems (top1-prov, b-box)
- Fluency-based model (FLM): 4-gram LM
- Alignment-based model (ALM): lexical translation model - IBM model
- Regression toward sentence-based BLEU for FLM, ALM or FLM+ALM
- Evaluation: Regression for FLM is the best (Bleu:+1)

Hildebrand and Vogel. 2008

- 4-gram and 5-gram LM, and lexical translation models (Lex)
- Difference with Nomoto 2003
  - Add two agreement models:
    - Position-dependent word agreement model (WordAgr)
    - Position-independent N-gram agreement model (NgrAgr)
  - Log linear model
- Evaluation:
  - Importance: LM>NgrAgr>WordAgr>Lex

Nomoto 2003 Predictive Models of Performance in Multi-Engine Machine Translation

Hildebrand and Vogel. 2008 Combination of machine translation systems via hypothesis selection from combined n-best lists

MT combination paper
MT paper

Features

*** Language model**
*** Translation model**
(* Agreement model)

Nomoto 2003

Hildebr and Vogel. 2008

***Syntactic model**

Zwarts and Dras. 2008

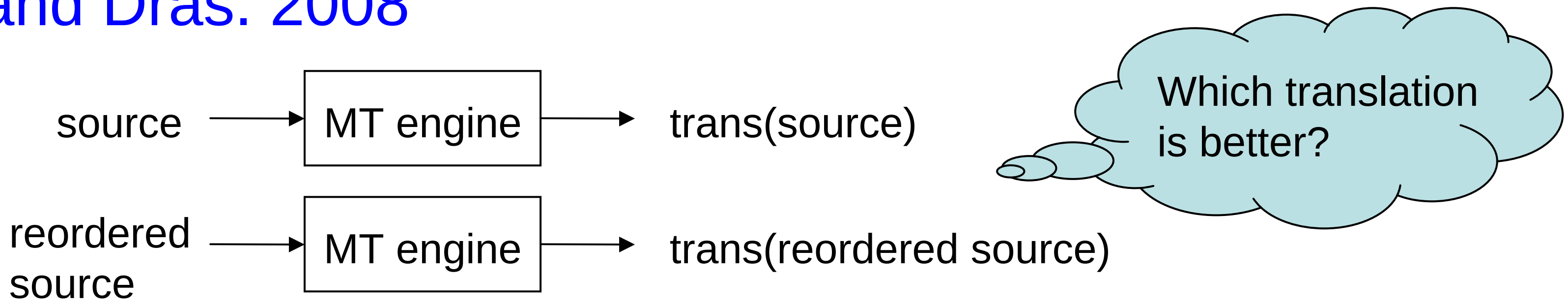
***Agreement model**

Kumar and Byrne. 2004

Sentence-based Combination

Zwarts and Dras. 2008

- Goal



- Syntactic features
  - Parsing scores of (non)reordered sources and their translations
- Binary SVM Classifier
- Evaluation
  - Parsing score of Target is more useful than Source
  - Decision accuracy is related to classifier’s prediction scores

---

Zwarts and Dras. 2008 Choosing the Right Translation: A Syntactically Informed classification Approach

MT combination paper

MT paper

Features

*** Language model**
*** Translation model**
(* Agreement model)

Nomoto 2003

Hildebr and Vogel. 2008

***Syntactic model**

Zwarts and Dras. 2008

***Agreement model**

Kumar and Byrne. 2004

Sentence-based Combination

Kumar and Byrne. 2004

- Minimum Bayes-Risk (MBR) Decoding for SMT
  - Could apply to N-best reranking

$$\hat{i} = \underset{i \in \{1,2,\ldots,N\}}{\operatorname{argmin}} \sum_{j=1}^{N} L((E_j, A_j), (E_i, A_i)) P(E_j, A_j | F)$$

- The loss function can be 1-BLEU, WER, PER, TER, Target-parse-tree-based function or Bilingual parse-tree-based function

---

Kumar and Byrne. 2004 Minimum Bayes-Risk Decoding for Statistical Machine Translation

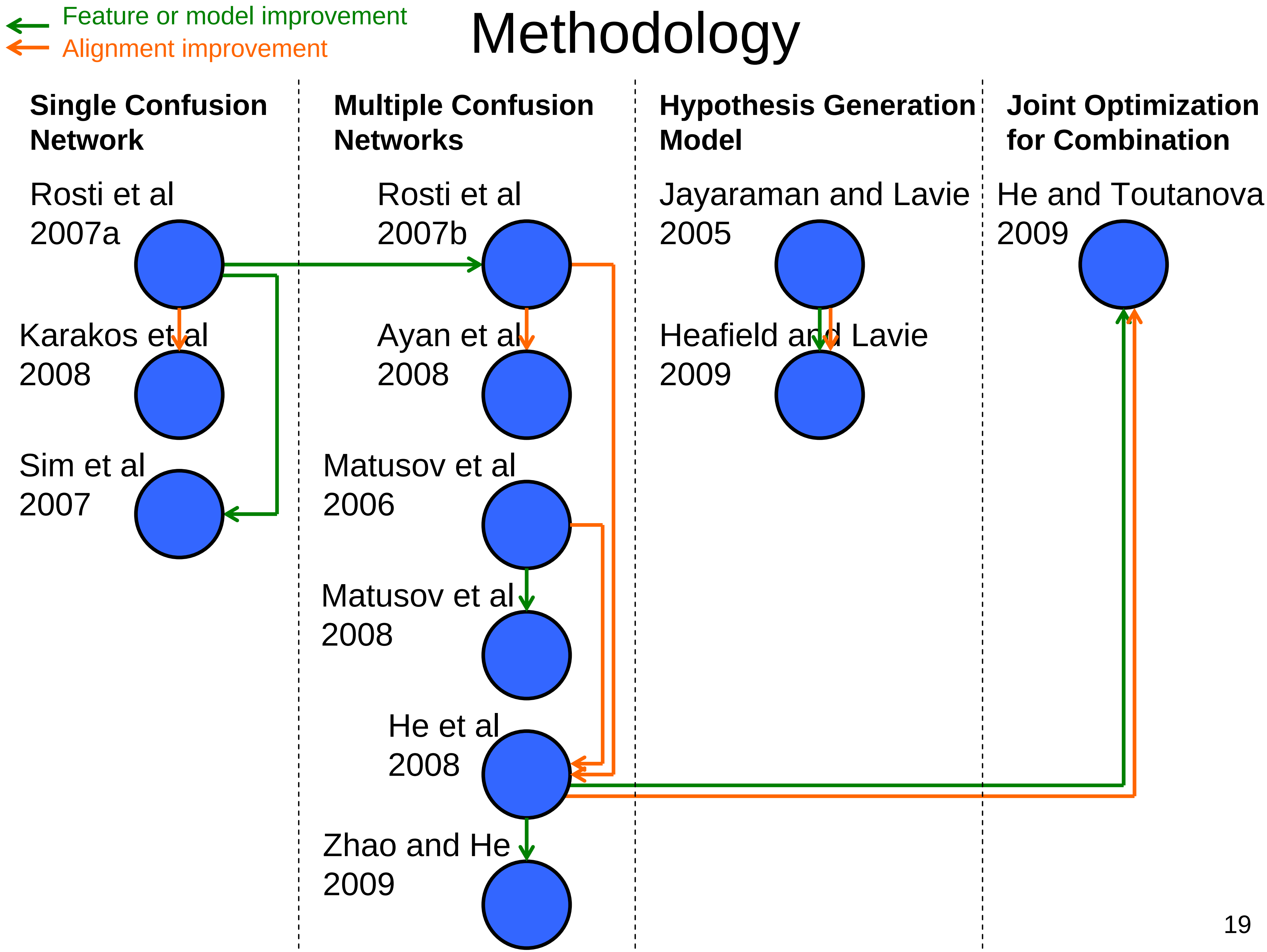
Synthesis: Sentence Based Combination

- My comments
  - Deep syntactic or even semantic relation could help
    - For example, semantic roles (who, what, where, why, how) in source are supposed to remain in target

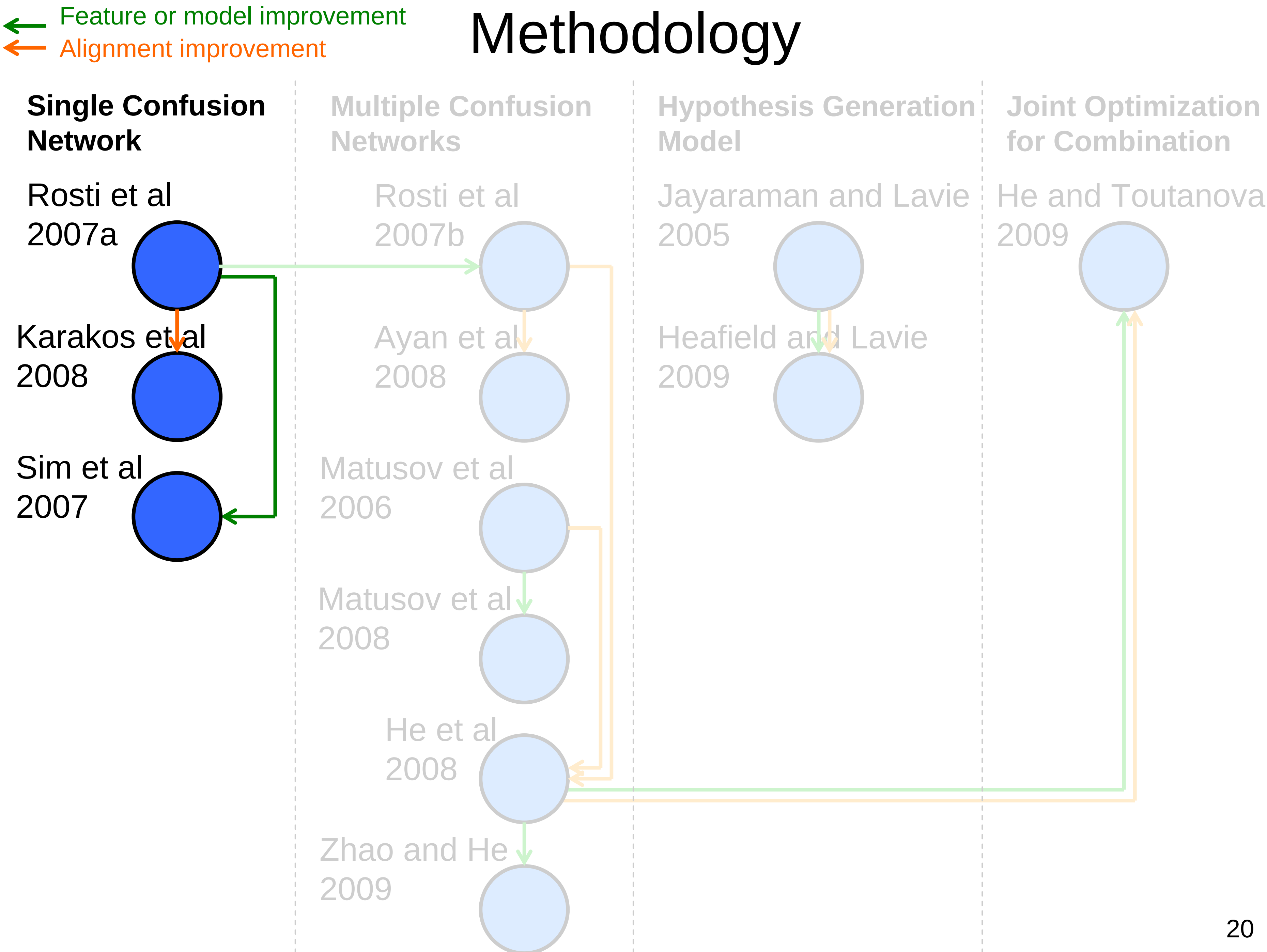
Outline

- Sentence-based Combination
- Word-based Combination
- Phrase-based Combination
- Comparative Analysis
- Conclusion

Methodology

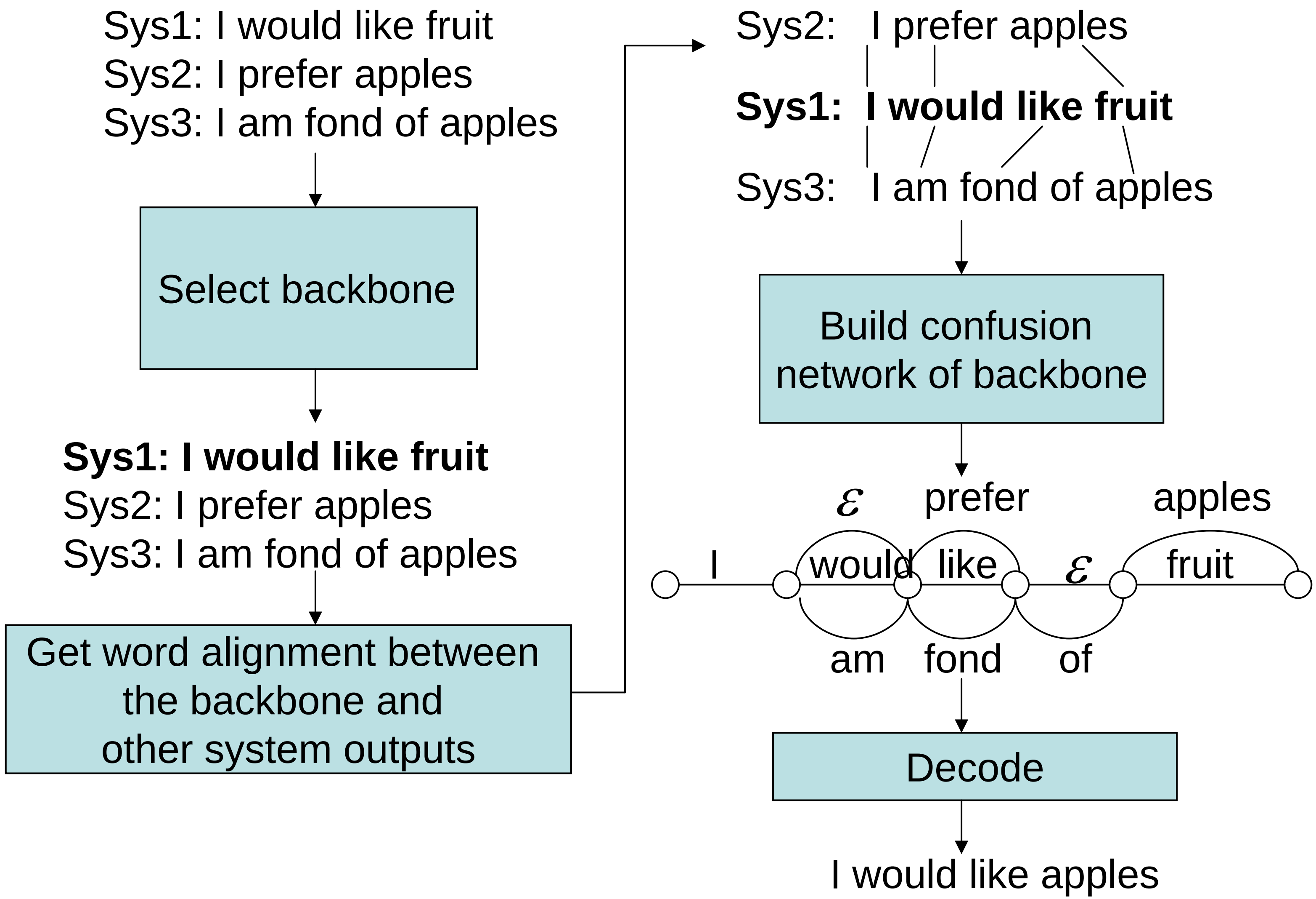


Methodology

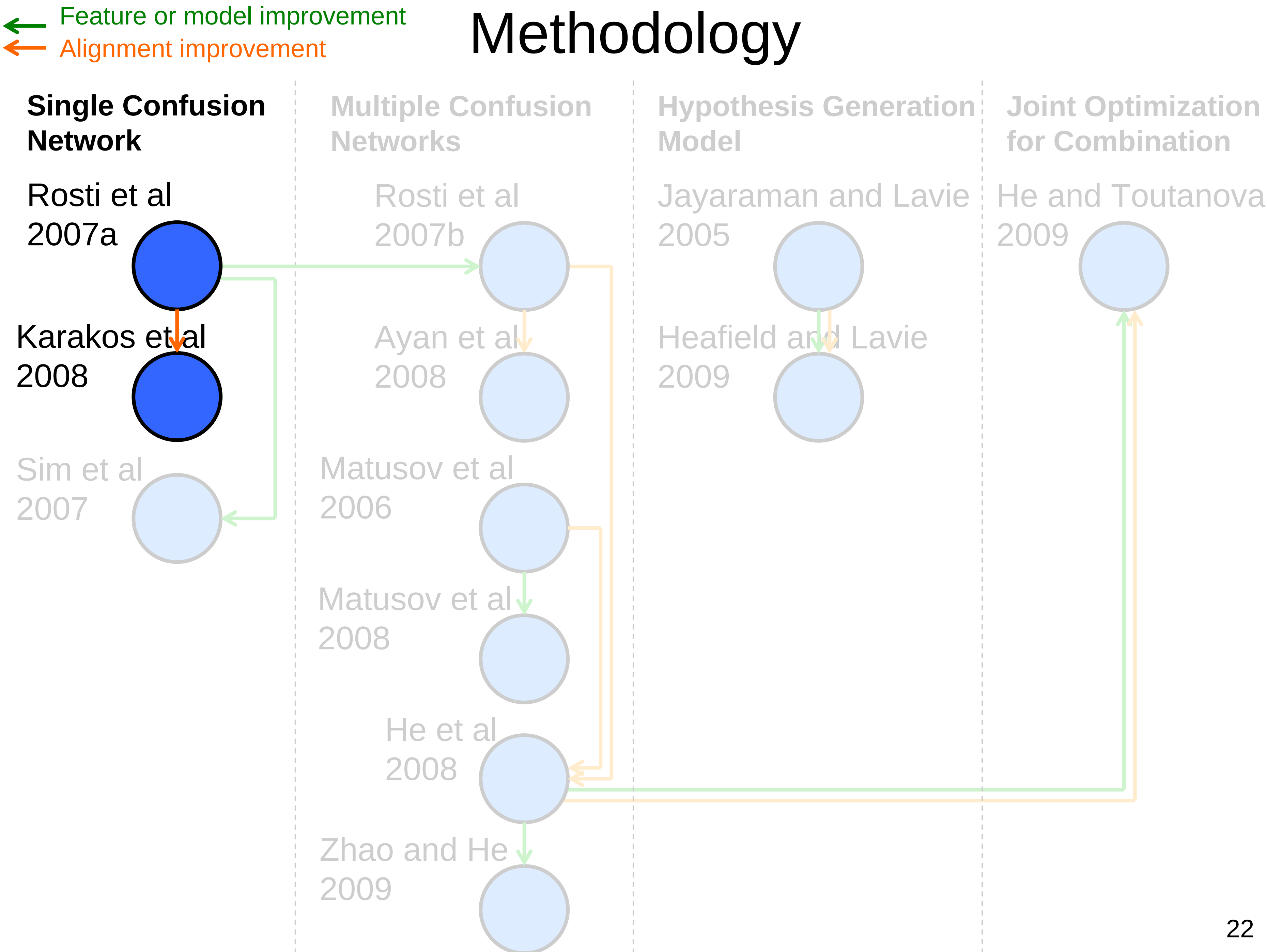


Word-based Combination

Single Confusion Networks



Methodology



Word-based Combination
Single Confusion Network

Rosti et al 2007a

- Each system provides TopN hypotheses
- Select Backbone and get alignment: TER (tool: tercom)
- Confidence score for each work (arc): 1/(1+N)
- Decoding:

$$E_r = \arg\min_i \sum_{j=1}^{N_s} \mathrm{TER}(E_j, E_i)$$

$$c(E_{j,n}|F_j) = \sum_{i=1}^{N_j-1} \sum_{l=1}^{N_s} \lambda_l c_{wli} + \mu N_{nulls}(E_{j,n})$$

- Evaluation
  - Arabic-English(News): BLEU:+2.3 TER:-1.34,
  - Chinese-English(News): BLEU:+1.1 TER:-1.96

Karakos et al 2008

- Nine Chinese-English MT systems (top1-prov, b-box)
- tercom is only an approximation of TER movements
- ITG-based alignment:
  edits allowed by the ITG grammar
  (nested block movements)

Ex : "thomas jefferson says eat your vegetables"
"eat your cereal thomas edison says"
tercom: 5 edits(wrong)
ITG-based alignment: 3 edits (correct)

- Combination evaluation shows ITG-based alignment outperforms tercom by BLEU of 0.6 and TER of 1.3, but it is much slower.

Rosti et al 2007a Combining outputs from multiple machine translation systems

Karakos et al 2008 Machine Translation System Combination using ITG-based Alignments

Word-based Combination

Single Confusion Network

Rosti et al 2007a

- Six Arabic-English and six Chinese-English MT systems (topN-prov, g-box)
- Select Backbone and get alignment: TER (tool: tercom) $E_b = \arg\min_{i} \sum_{j=1}^{N_s} \mathrm{TER}(E_j, E_i)$
- Confidence score for each work (arc): 1/(1+rank)
- Decoding: $\ell(E_{l}, F) = \sum_{l=1}^{L} \sum_{i=1}^{N_s} \lambda_i s_{il} + \nu N_{nul}(E_{l})$
- Evaluation
  - Arabic-English(News): BLEU:+2.3 TER:-1.34,
  - Chinese-English(News): BLEU:+1.1 TER:-1.96

Karakos et al 2008

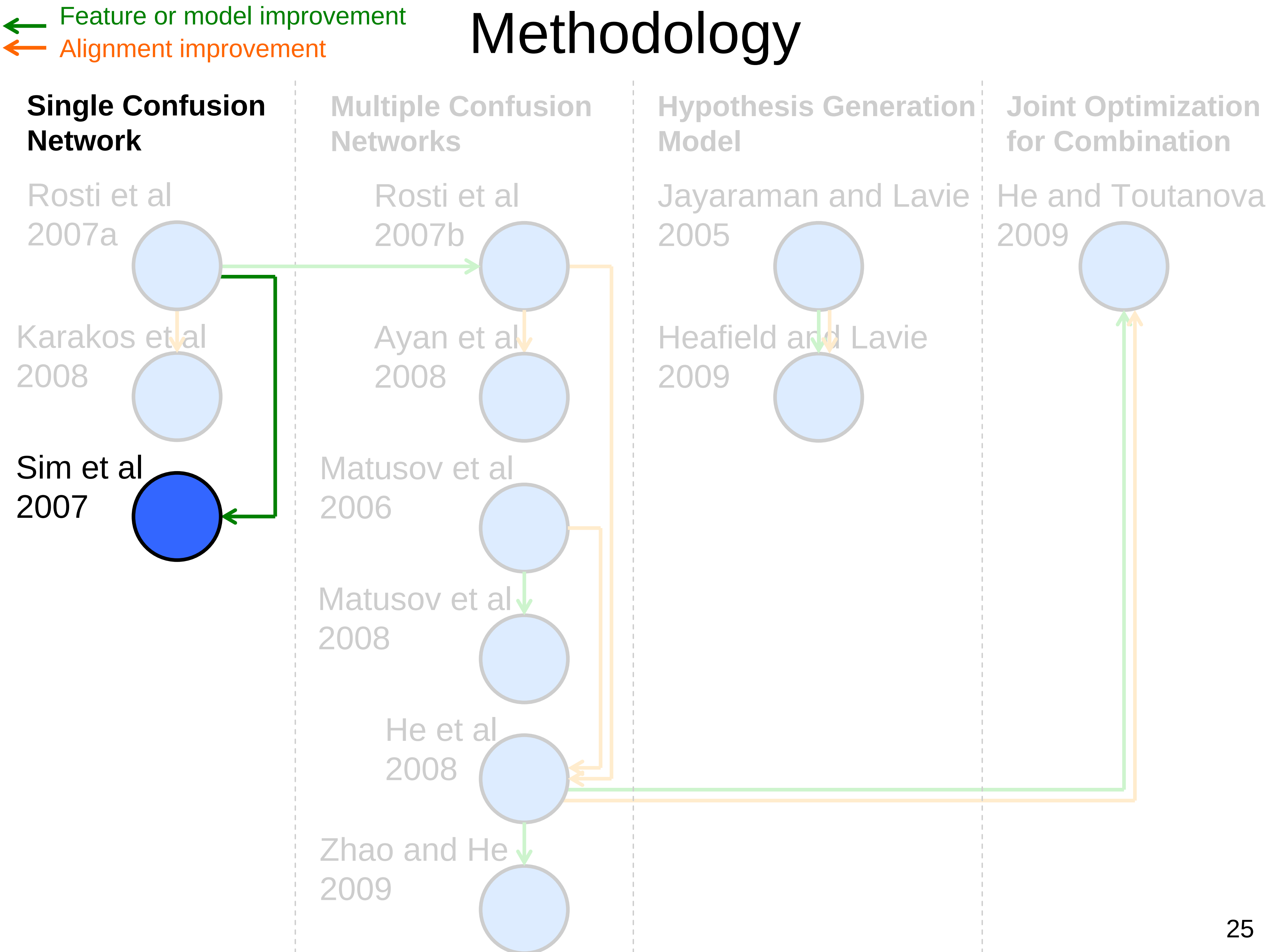
- tercom is only an approximation of TER movements
- Improvement on Rosti et al 2007a
  - ITG-based alignment:
    edits allowed by the ITG grammar
    (nested block movements)

Ex : "thomas jefferson says eat your vegetables"
"eat your cereal thomas edison says"
tercom: 5 edits(wrong)
ITG-based alignment: 3 edits (correct)

- Evaluation
  - ITG-based alignment outperforms tercom by BLEU of 0.6 and TER of 1.3, but it is much slower.

Rosti et al 2007a Combining outputs from multiple machine translation systems
Karakos et al 2008 Machine Translation System Combination using ITG-based Alignments

Methodology



Word-based Combination

Single Confusion Network

Sim et al 2007

- Six Arabic-English MT systems (top1-prov, b-box)
- Improvement on Rosti et al 2007a
  - Consensus Network MBR (ConMBR)
    - Goal: Retain the coherent phrases in the original translations
    - Procedure:
      - Step1: get decoded hypothesis (E_{con}) from confusion network
      - Step2: Select the original translation which is most similar with E_{con}

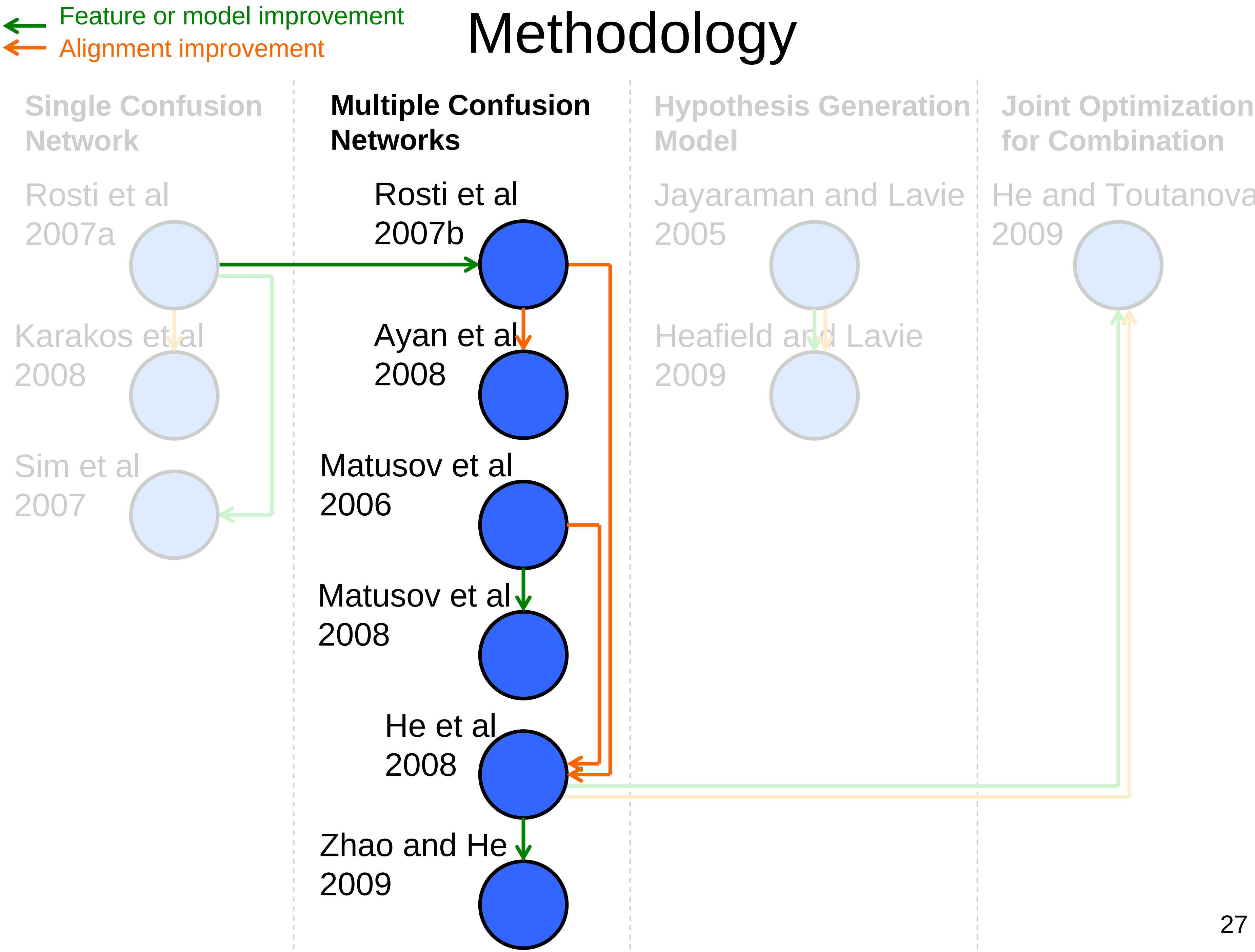
$$E_{\mathtt{ConMBR}} = \arg\min_{E'} L(E', E_{\mathtt{con}})$$

- Evaluation

System/ Combination	2003		2004	
	TER	BLEU	TER	BLEU
BBN Phrase	41.56	53.32	41.71	45.40
BBN Hiero	42.36	52.03	44.26	42.67
Edinburgh	42.05	52.5	44.20	**47.76**
ISI Hiero	**40.53**	**54.54**	**42.21**	46.49
ISI Phrase	41.94	52.35	43.09	45.21
ISI Syntax	42.96	52.36	45.00	44.11
MBR-BLEU	39.71	56.16	41.29	48.37
Confusion	39.37	55.67	41.21	46.45
ConMBR-BLEU	**39.02**	**56.64**	**40.23**	**48.93**

Sim et al 2007 Consensus network decoding for statistical machine translation system combination

Methodology



Word-based Combination
Multiple Confusion Networks

Sys1: I would like fruit
Sys2: I prefer apples
Sys3: I am fond of apples

top1-prov:
no backbone selection
topN-prov:
For each system, select
a backbone from its N-best

Sys1: I would like fruit
Sys2: I prefer apples
Sys3: I am fond of apples

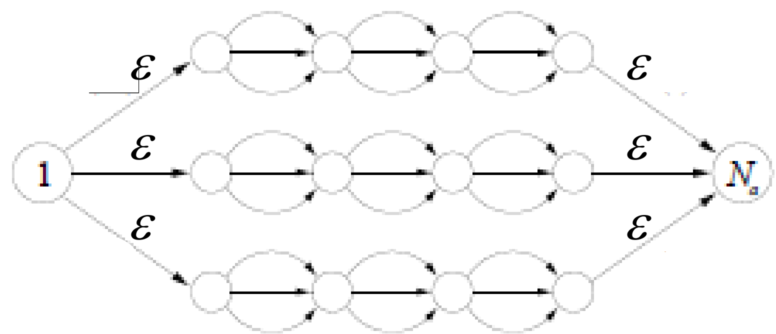
Get word alignment between
each backbone and
all other system outputs

Sys2: I prefer apples
Sys1: I would like fruit
Sys3: I am fond of apples

Sys1: I would like fruit
Sys2: I prefer apples
Sys3: I am fond of apples

Sys1: I would like fruit
Sys3: I am fond of apples
Sys2: I prefer apples

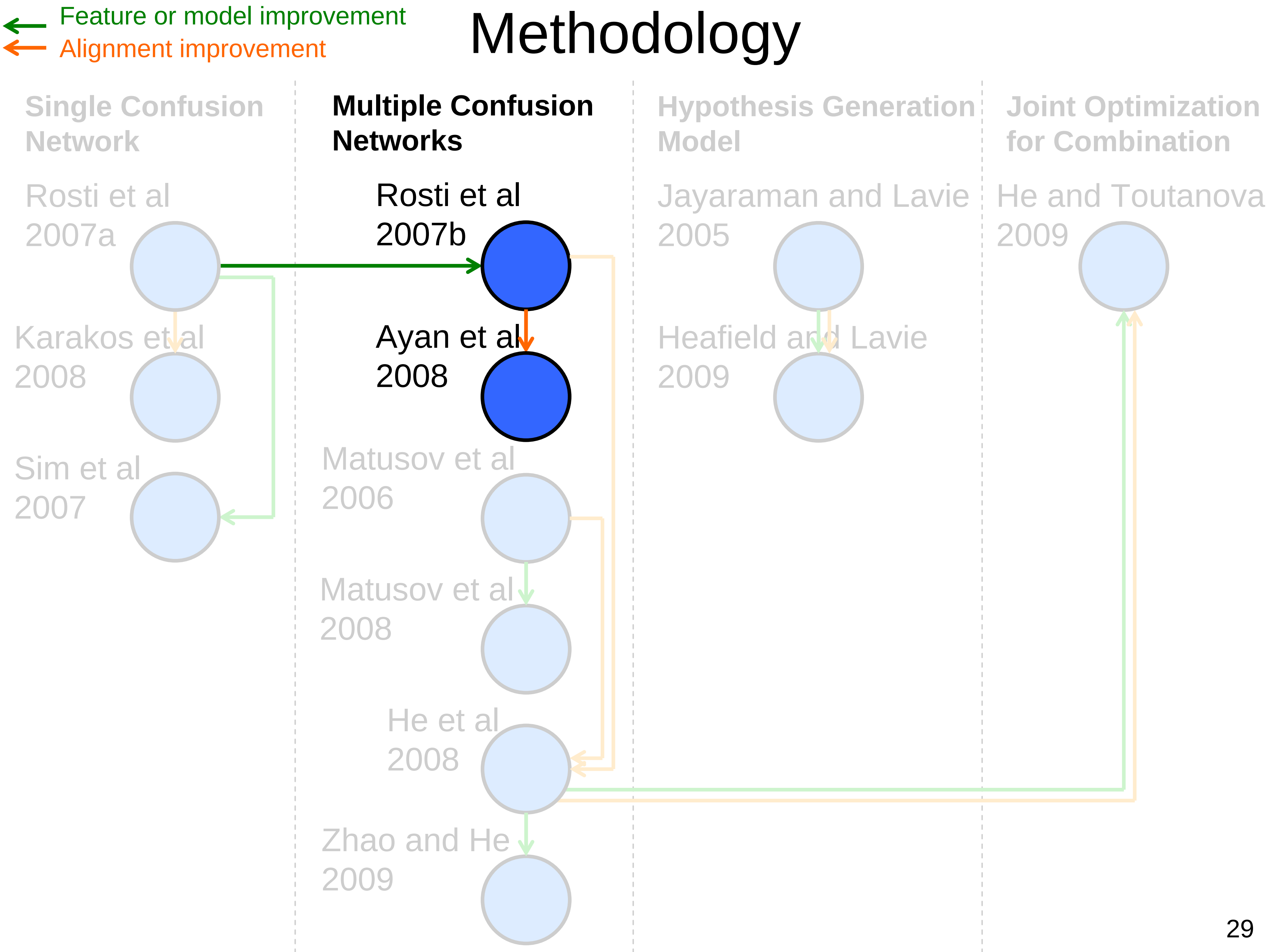
Build confusion networks for
each backbones



decode

I would like apples

Methodology



Word-based Combination

Multiple Confusion Networks

Rosti et al 2007b

- Improvement on Rosti et al 2007a
  - Structure: multiple Confusion Networks
  - Scoring: arbitrary features, such as LM and word number

$$\log p(E_{j,n}|F_j) = \sum_{i=1}^{N_j-1} \log\Big(\sum_{l=1}^{N_s} \lambda_l p(w|l,i)\Big) + \nu L(E_{j,n}) + \mu N_{nulls}(E_{j,n}) + \xi N_{words}(E_{j,n})$$

- Evaluation
  - Arabic-English: BLEU:+3.2, TER:-1.7 (baseline:BLEU:+2.4, TER:-1.5)
  - Chinese-English: BLEU:+0.5, TER:-3.4 (baseline:BLEU:+1.1, TER:-2)

Ayan et al 2008

- Three Arabic-English and three Chinese-English MT systems (topN-prov, g-box)
  - Only one engine but use different training data
- Difference with Rosti et al 2007b
  - word confidence score: add system-provided translation score
  - Extend TER script (tercom) with synonym matching operation using WordNet
  - Two-pass alignment strategy to improve the alignment performance
    - Step1: align backbone with all other hypotheses to produce confusion network
    - Step2: get decoded hypothesis (E_{con}) form confusion network
    - Step3: align E_{con} with all other hypotheses to get the new alignment
- Evaluation
  - No synon+No Two-pass: BLEU:+1.6 synon+No Two-pass: BLEU:+1.9
  - No synon+Two-pass: BLEU:+2.6 synon+Two-pass: BLEU:+2.9

Rosti et al 2007b Improved Word-Level System Combination for Machine Translation
Ayan et al 2008 Improving alignments for better confusion networks for combining machine translation systems

Word-based Combination

Multiple Confusion Networks

Rosti et al 2007b

- Six Arabic-English and six Chinese-English MT systems (topN-prov, b-box)
- Difference with Rosti et al 2007a
  - Structure: multiple Confusion Networks
  - Scoring: arbitrary features, such as LM

$$\log p(E_{j,a}|F_j) = \sum_{i=1}^{N_j} \log\Big(\sum_{k=1}^{N_s} \lambda_k p(w|k,i)\Big) + \nu L(E_{j,a}) + \mu N_{nulls}(E_{j,a}) + \xi N_{words}(E_{j,a})$$

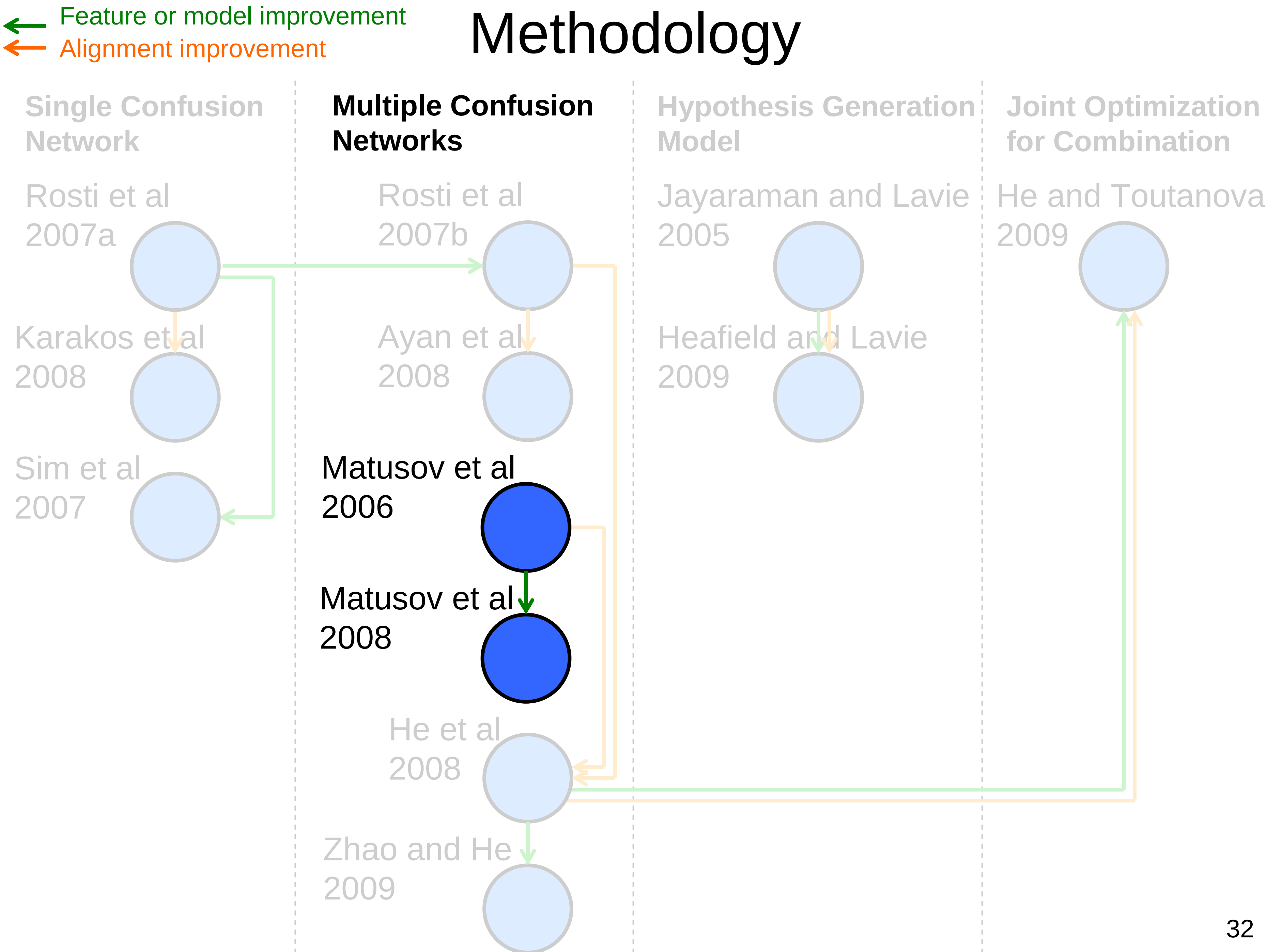
- Evaluation
  - Arabic-English: BLEU:+3.2, TER:-1.7 (baseline:BLEU:+2.4, TER:-1.5)
  - Chinese-English: BLEU:+0.5, TER:-3.4 (baseline:BLEU:+1.1, TER:-2)

Ayan et al 2008

- Only one MT engine but use different training data
- Improvement on Rosti et al 2007b
  - word confidence score: add system-provided translation score
  - Extend TER script (tercom) with synonym matching operation using WordNet
  - Two-pass alignment strategy to improve the alignment performance
    - Step1: align backbone with all other hypotheses to produce confusion network
    - Step2: get decoded hypothesis (E_{con}) form confusion network
    - Step3: align E_{con} with all other hypotheses to get the new alignment
- Evaluation
  - No synon+No Two-pass: BLEU:+1.6 synon+No Two-pass: BLEU:+1.9
  - No synon+Two-pass: BLEU:+2.6 synon+Two-pass: BLEU:+2.9

Rosti et al 2007b Improved Word-Level System Combination for Machine Translation
Ayan et al 2008 Improving alignments for better confusion networks for combining machine translation systems

Methodology



Word-based Combination

Multiple Confusion Networks

Matusov et al 2006

- Alignment approach: HMM model bootstrapped from IBM model1

$$p(e_1'^J \mid e_1^I) = \sum_{a_1^J} \prod_{j=1}^{J} \left[p(a_j \mid a_{j-1}, I)\, p(e_j' \mid e_{a_j}) \right] \qquad p(a_j = i \mid a_{j-1} = i', I) = \frac{c(i-i')}{\sum_{l=1}^{I} c(l-i')}$$

- Rescoring for confusion network outputs by general LM

Matusov et al 2008

- Six English-Spanish and six Spanish-English MT systems (top1-prov, b-box)
- Difference with Matusov et al 2006
  - Integrate general LM and adapted LM (online LM) into confusion network decoding
    - adapted LM (online LM): N-gram based on system outputs
  - Handling long sentences by splitting them
- Evaluation
  - English-Spanish: BLEU:+2.1 Spanish-English: BLEU:+1.2
  - adapted LM is more useful than general LM in either confusion network decoding or rescoring

---

Matusov et al 2006 Computing consensus translation from multiple machine translation systems using enhanced hypotheses alignment

Matusov et al 2008 System combination for machine translation of spoken and written language

Word-based Combination
Multiple Confusion Networks

Matusov et al 2006

- Five Chinese-English and four Spanish-English MT systems (top1-prov, b-box)
- Alignment approach: HMM model bootstrapped from IBM model1

$$p(e_1'^{J} \mid e_1^{I}) = \sum_{a_1^J} \prod_{j=1}^{J} \left[p(a_j \mid a_{j-1}, I)\, p(e_j' \mid e_{a_j}) \right] \qquad p(a_j = i \mid a_{j-1} = i', I) = \frac{c(i-i')}{\sum_{l=1}^{I} c(l-i')}$$

- Rescoring for confusion network outputs by general LM
- Evaluation
  - Chinese-English: BLEU:+5.9 Spanish-English: BLEU:+1.6

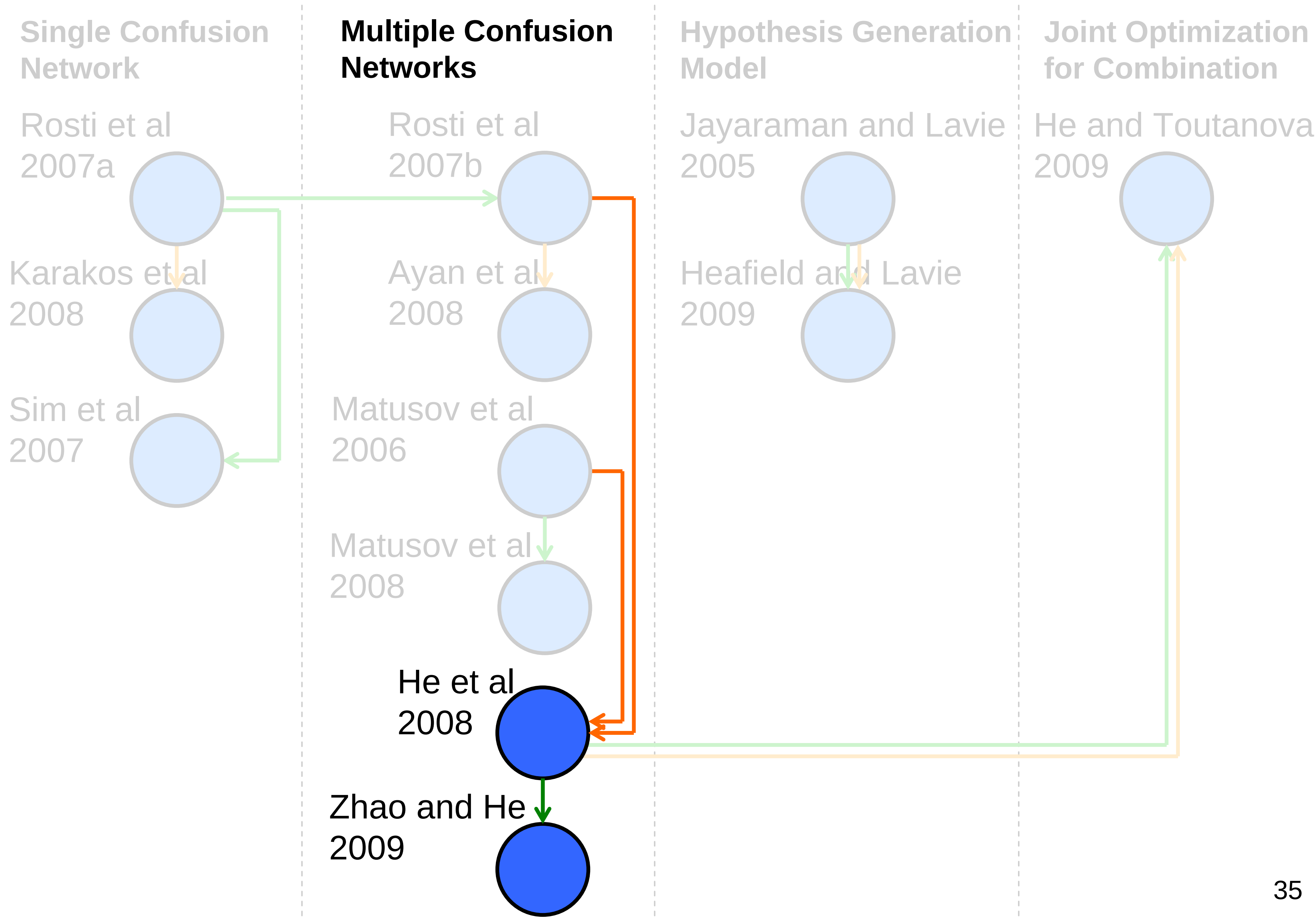
Matusov et al 2008

- Improvement on Matusov et al 2006
  - Integrate general LM and adapted LM (online LM) into confusion network decoding
    - adapted LM (online LM): N-gram based on system outputs
  - Handling long sentences by splitting them
- Evaluation
  - adapted LM is more useful than general LM in either confusion network decoding or rescoring

Matusov et al 2006 Computing consensus translation from multiple machine translation systems using enhanced hypotheses alignment
Matusov et al 2008 System combination for machine translation of spoken and written language

Methodology

Feature or model improvement
Alignment improvement



Word-based Combination

Multiple Confusion Networks

He et al 2008

- Alignment approach: Indirect HMM (IHMM)

HMM

$$p(e_1'^J \mid e_1^I) = \sum_{a_1^J} \prod_{j=1}^{J} \left[p(a_j \mid a_{j-1}, I) \boxed{p(e_j' \mid e_{a_j})} \right]$$

$$p(a_j = i \mid a_{j-1} = i', I) = \frac{\boxed{c(i-i')}}{\sum_{l=1}^{I} c(l-i')}$$

IHMM

$$p(e_j' \mid e_i) = \alpha \cdot p_{sem}(e_j' \mid e_i) + (1-\alpha) \cdot p_{sur}(e_j' \mid e_i)$$

Grouping c(i-I') with 11 buckets: c(<=-4), c(-3) ... c(0), ..., c(5), C(>=6) and use the following to give the value

$$c(d) = \left(1 + |d-1|\right)^{-\kappa}, \; d = -4, \ldots, 6$$

- Evaluation
  - Baseline (alignment: TER): BLEU:+3.7 This paper (alignment: IHMM): BLEU:+4.7

Zhao and He 2009

- Some Chinese-English MT systems (topN-prov, b-box)
- Difference with He et al 2008
  - Add agreement model: two online N-gram LM models
- Evaluation
  - Baseline (He et al 2008): BLEU:+4.3 This paper: BLEU:+5.11

He et al 2008 Indirect-hmm-based hypothesis alignment for computing outputs from machine translation systems

Zhao and He 2009 Using n-gram based features for machine translation system combination

Word-based Combination
Multiple Confusion Networks

He et al 2008

- Eight Chinese-English MT systems (topN-prov, b-box)
- Alignment approach: Indirect HMM (IHMM)

HMM

$$p(e_1'^{J} \mid e_1^{I}) = \sum_{a_1^J} \prod_{j=1}^{J} \left[p(a_j \mid a_{j-1}, I)\, p(e_j' \mid e_{a_j}) \right]$$

$$p(a_j = i \mid a_{j-1} = i', I) = \frac{c(i-i')}{\sum_{l=1}^{I} c(l-i')}$$

IHMM

$$p(e_j' \mid e_i) = \alpha \cdot p_{sem}(e_j' \mid e_i) + (1-\alpha) \cdot p_{sur}(e_j' \mid e_i)$$

Grouping c(i-I') with 11 buckets: c(<=-4), c(-3) ... c(0), ..., c(5), C(>=6) and use the following to give the value

$$c(d) = \left(1 + |d-1|\right)^{-K}, \; d = -4, \ldots, 6$$

- Evaluation
  – Baseline (alignment: TER): BLEU:+3.7 This paper (alignment: IHMM): BLEU:+4.7

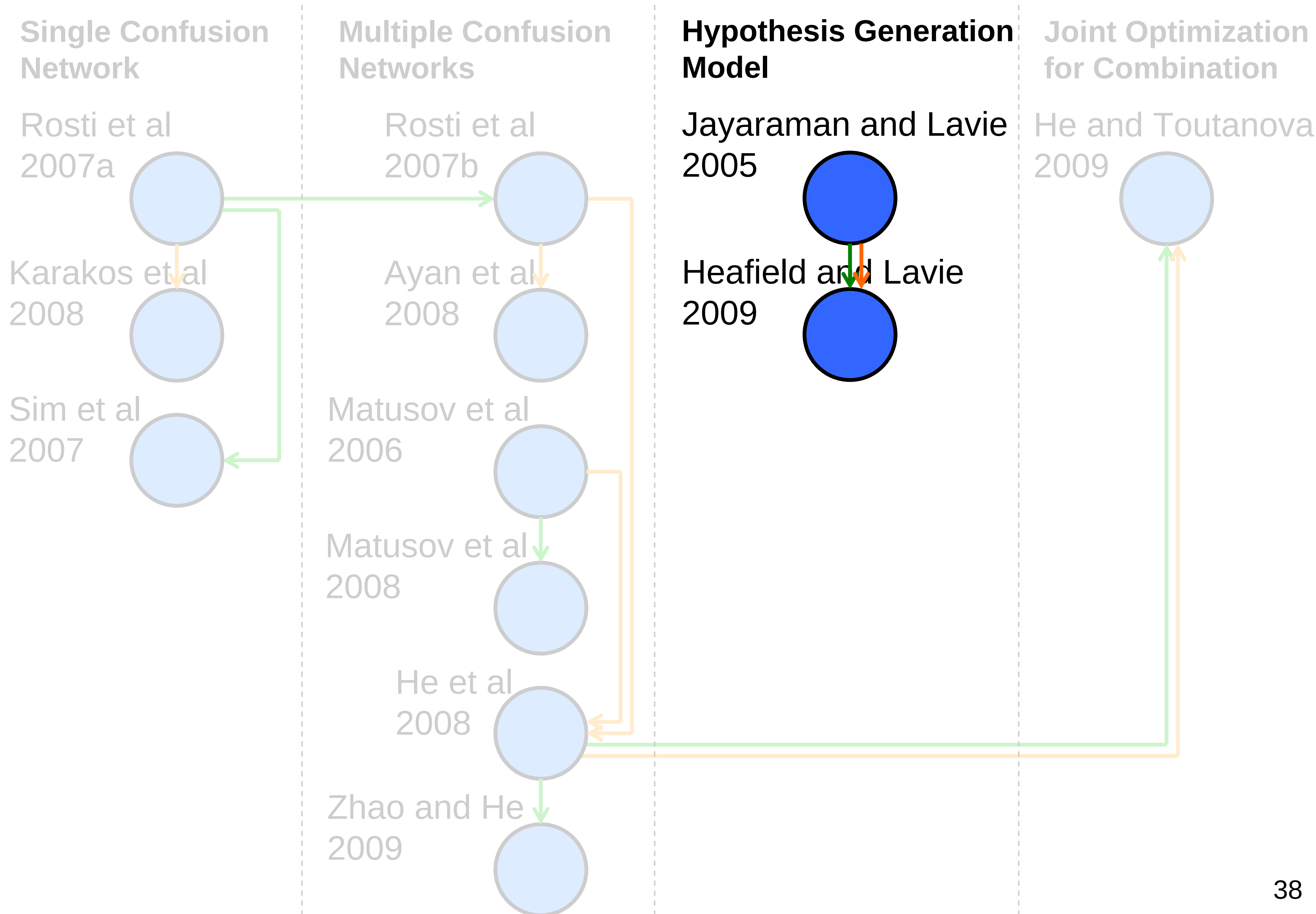
Zhao and He 2009

- Improvement on He et al 2008
  – Add agreement model: two online N-gram LM models
- Evaluation
  – Baseline (He et al 2008): BLEU:+4.3 This paper: BLEU:+5.11

He et al 2008 Indirect-hmm-based hypothesis alignment for computing outputs from machine translation systems
Zhao and He 2009 Using n-gram based features for machine translation system combination

Feature or model improvement
Alignment improvement

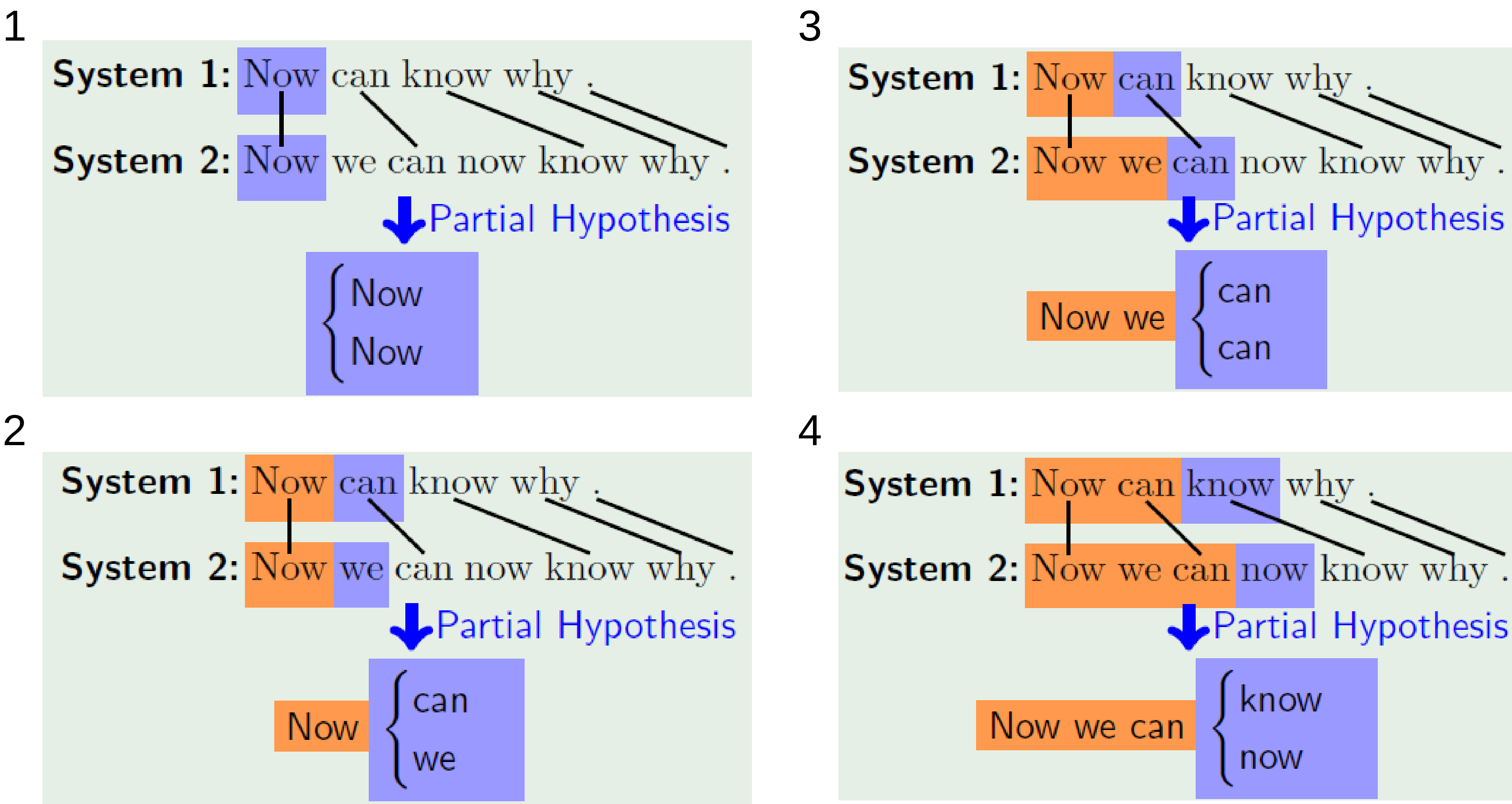
Methodology



Word-based Combination

Hypothesis Generation Model

Algorithm: Repeatedly extend hypothesis by appending a word from a system



Word-based Combination

Multiple Confusion Networks

Jayaraman and Lavie 2005

- Heuristic word alignment approach
- Feature: LM+N-gram agreement model

Heafield and Lavie 2009

- Three German-English and three French-English MT systems (top1-prov, b-box)
- Difference with Jayaraman and Lavie 2005
  - Word alignment tool: METEOR
  - Switching between systems is not permitted within a phrase
    - Phrase Definition is based on word aligned situations
  - Synchronize extensions of hypotheses
- Evaluation
  - German-English: BLEU:+0.16 TER:-2.3
  - French-English: BLEU:-0.1 TER:-0.2

Jayaraman and Lavie 2005 Multi-Engine Machine Translation Guided by Explicit Word Matching

Heafield and Lavie 2009 Machine Translation System Combination with Flexible Word Ordering

Word-based Combination

Multiple Confusion Networks

Jayaraman and Lavie 2005

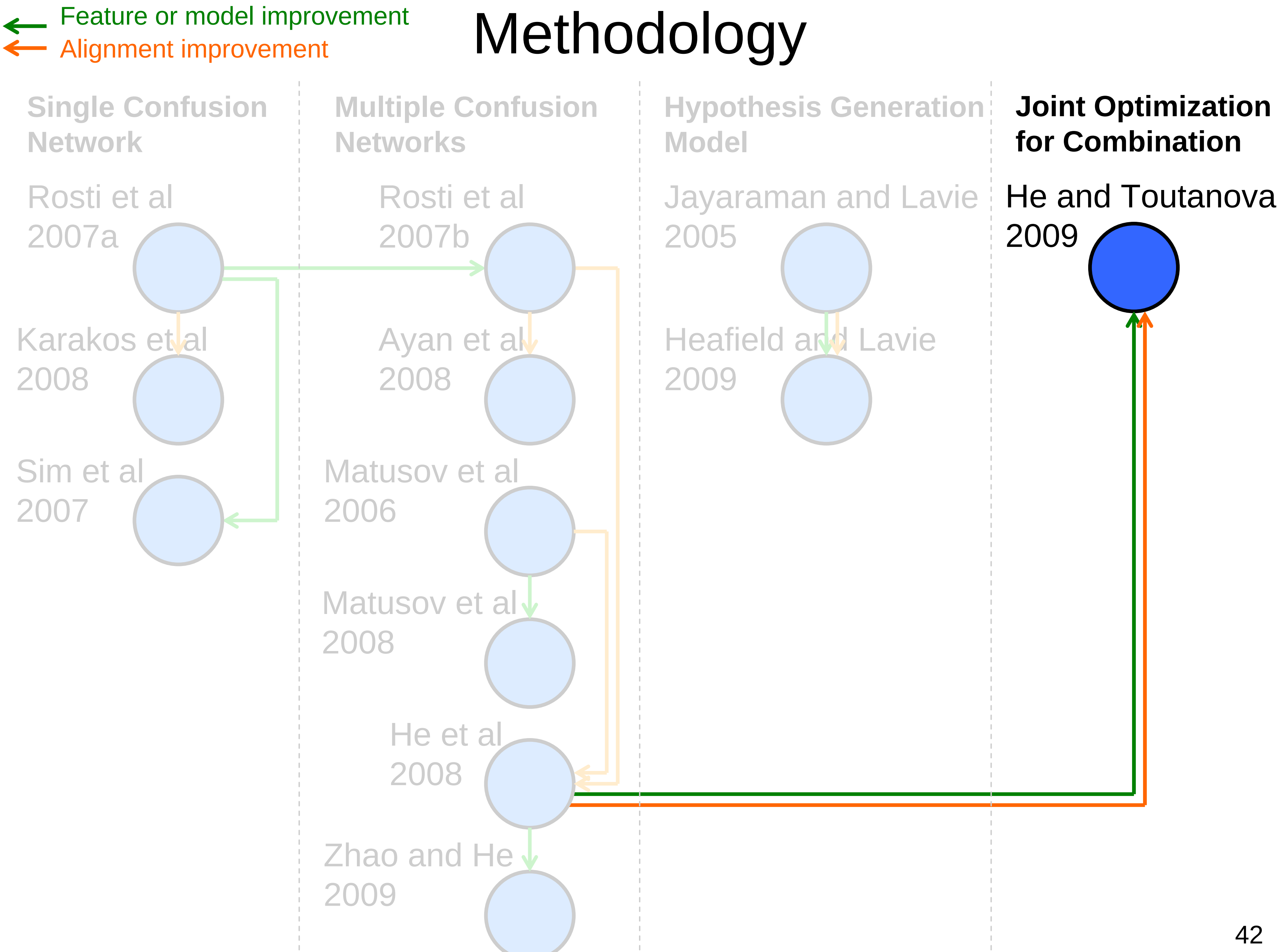
- Three Arabic-English MT systems (top1-prov, b-box)
- Heuristic word alignment approach
- Feature: LM+N-gram agreement model
- Evaluation
  - BLEU:+7.78

Heafield and Lavie 2009

- Improvement on Jayaraman and Lavie 2005
  - Word alignment tool: METEOR
  - Switching between systems is not permitted within a phrase
    - Phrase Definition is based on word aligned situations
  - Synchronize extensions of hypotheses

Jayaraman and Lavie 2005 Multi-Engine Machine Translation Guided by Explicit Word Matching
Heafield and Lavie 2009 Machine Translation System Combination with Flexible Word Ordering

Methodology



Word-based Combination

Joint Optimization for Combination

He and Toutanova 2009

- Motivation: poor alignment
- Joint log-linear model integrating the following features
  - Word posterior model (agreement model)
  - Bi-gram voting model (agreement model)
  - Distortion model
  - Alignment model
  - Entropy model
- Decoding: A beam search algorithm
  - Pruning: prune down alignment space
  - Estimate the future cost of an unfinished path
- Evaluation
  - Baseline (IHMM in He et al 2008): BLEU:+3.82 This paper: BLEU+5.17

He and Toutanova 2009 Joint optimization for machine translation system combination

Outline

- Sentence-based Combination
- Word-based Combination
- Phrase-based Combination
- Comparative Analysis
- Conclusion

Methodology

MT combination paper

MT paper

← Feature or model improvement

Related work from MT

Koehn et al 2003

Callison-Burch et al 2006

Utilizing MT Engine

Rosti et al 2007a

Chen et al 2009

Huang and Papineni 2007

Mellebeek et al 2006

Without utilizing MT Engine

Frederking and Nirenburg 1994

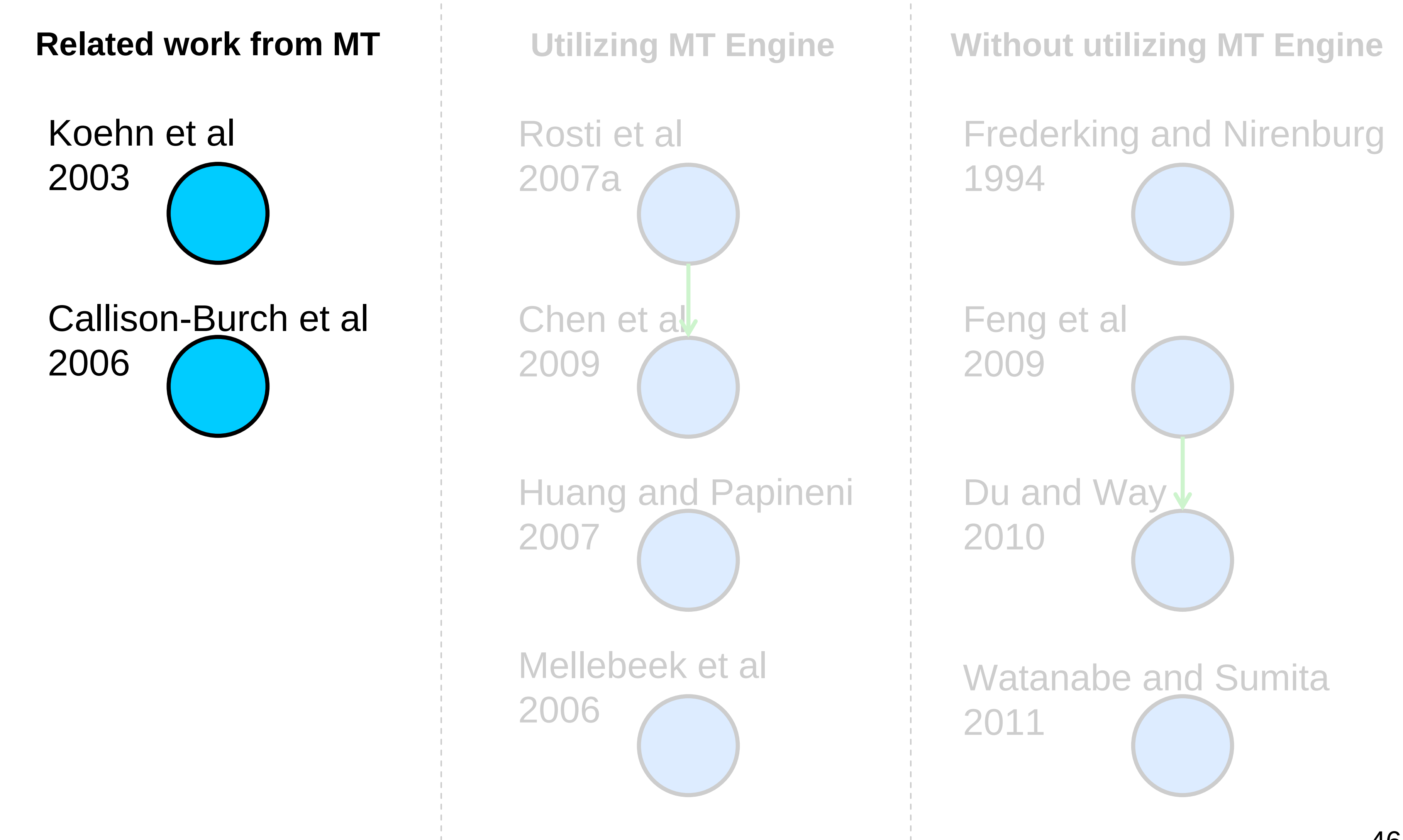
Feng et al 2009

Du and Way 2010

Watanabe and Sumita 2011

Methodology

MT combination paper
MT paper
Feature or model improvement



Phrase-based Combination
Related work from MT

Koehn et al 2003

- A set of experiments tells us:
  - Phrase-based translations is better than word-based translation
  - Heuristic learning of phrase translations form word-based alignment works
  - Lexical weighting of phrase translations helps
  - Phrases longer than three words do not help
  - Syntactically motivated phrases degrade the performance
- My comment
  - Are they also true for MT combination?

Callison-Burch et al 2006

- The paper tells us that augmenting a state-of-the-art SMT system with paraphrases helps.
- Acquiring paraphrases through bilingual parallel corpora
  - Paraphrase probabilities

$$p(e_2|e_1) = \sum_f p(f|e_1)p(e_2|f)$$

- My comment
  - Do paraphrase probabilities helps for MT combination?

Koehn et al 2003 Statistical phrase-based translation

Callison-Burch et al 2006 Improved Statistical Machine Translation Using Paraphrases

Phrase-based Combination
Related work from MT

Koehn et al 2003

- A set of experiments tells us:
  - Phrase-based translations is better than word-based translation Probably, but…
  - Heuristic learning of phrase translations form word-based alignment works Probably, but…
  - Lexical weighting of phrase translations helps not sure so far
  - Phrases longer than three words do not help not sure so far
  - Syntactically motivated phrases degrade the performance not sure so far
- My comment
  - Are they also true for MT combination?

Callison-Burch et al 2006

- The paper tells us that augmenting a state-of-the-art SMT system with paraphrases helps.
- Acquiring paraphrases through bilingual parallel corpora
  - Paraphrase probabilities

$$p(e_2|e_1) = \sum_f p(f|e_1)p(e_2|f)$$

- My comment
  - Do paraphrase probabilities helps for MT combination?

Koehn et al 2003 Statistical phrase-based translation

Callison-Burch et al 2006 Improved Statistical Machine Translation Using Paraphrases

Phrase-based Combination
Related work from MT

Koehn et al 2003

- A set of experiments tells us:
  - Phrase-based translations is better than word-based translation
  - Heuristic leaning of phrase translations form word-based alignment works
  - Lexical weighting of phrase translations helps
  - Phrases longer than three words do not help
  - Syntactically motivated phrases degrade the performance
- My comment
  - Are they also true for MT combination?

Callison-Burch et al 2006

- The paper tells us that augmenting a state-of-the-art SMT system with paraphrases helps.
- Acquiring paraphrases through bilingual parallel corpora
  - Paraphrase probabilities

$$p(e_2|e_1) = \sum_f p(f|e_1)p(e_2|f)$$

- My comment
  - Do paraphrase probabilities helps for phrase-based combination?

Koehn et al 2003 Statistical phrase-based translation

Callison-Burch et al 2006 Improved Statistical Machine Translation Using Paraphrases

Phrase-based Combination

Related work from MT

Koehn et al 2003

- A set of experiments tells us:
  - Phrase-based translations is better than word-based translation
  - Heuristic leaning of phrase translations form word-based alignment works
  - Lexical weighting of phrase translations helps
  - Phrases longer than three words do not help
  - Syntactically motivated phrases degrade the performance
- My comment
  - Are they also true for MT combination?

Callison-Burch et al 2006

- The paper tells us that augmenting a state-of-the-art SMT system with paraphrases helps.
- Acquiring paraphrases through bilingual parallel corpora
  - Paraphrase probabilities

$$p(e_2|e_1) = \sum_f p(f|e_1)p(e_2|f)$$

- My comment
  - Do paraphrase probabilities helps for phrase-based combination? not sure so far

---

Koehn et al 2003 Statistical phrase-based translation

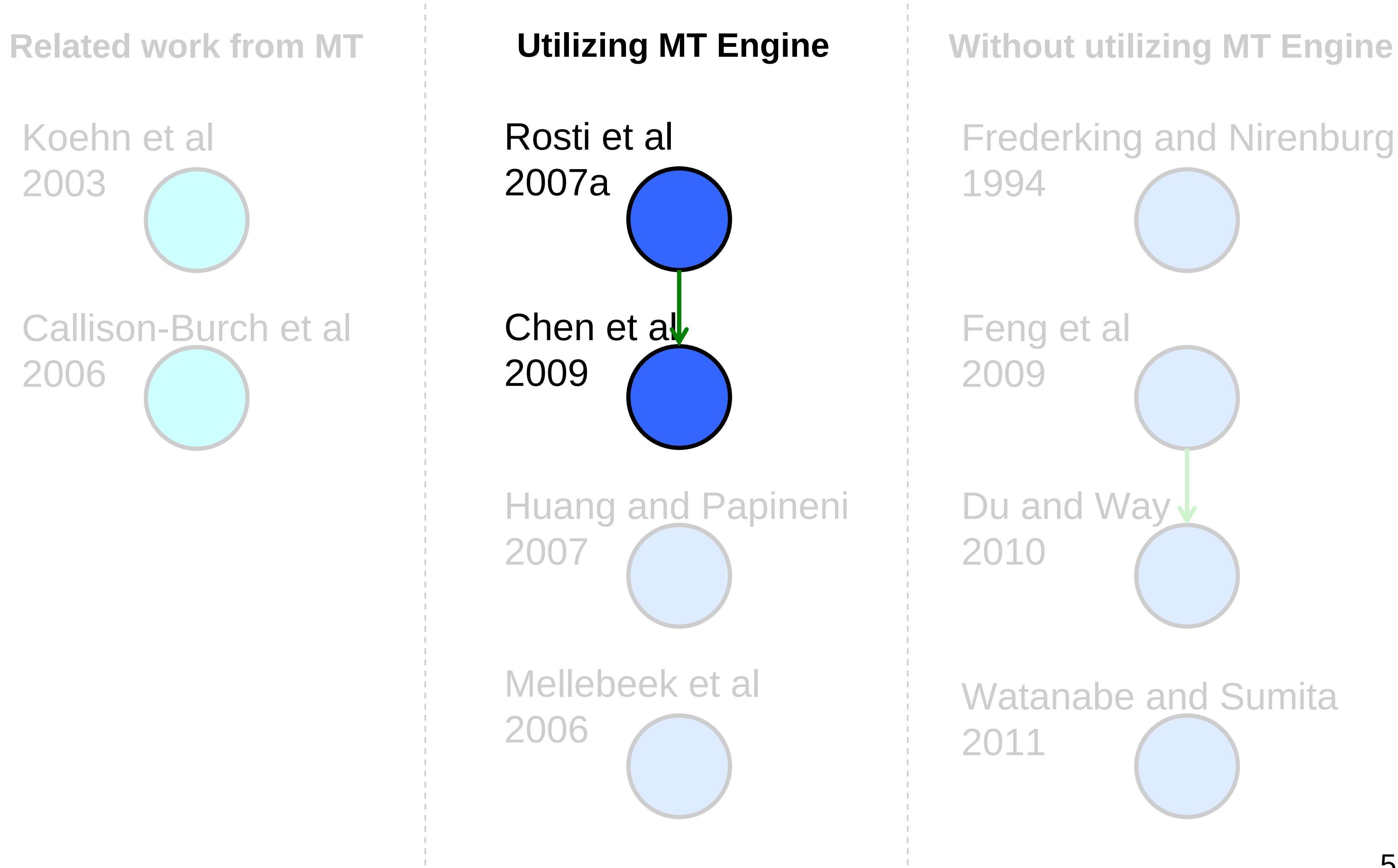
Callison-Burch et al 2006 Improved Statistical Machine Translation Using Paraphrases

Methodology

MT combination paper

MT paper

Feature or model improvement



Phrase-based Combination
Utilizing MT Engine

Rosti et al 2007a

- Algorithm
  - Extracting a new phrase table from provided phrase alignment
  - Re-decoding source based on the new phrase table
- Phrase confidence score
  - Agreement model on four levels of similarity
  - Integrating weights of systems and levels of similarity
- Re-decoding: a standard beam search – Pharaoh
- Evaluation
  - Performance Comparison
    - Arabic-English: word-based comb. > phrase-based comb. > sentence-based comb.
    - Chinese-English: word-based comb. > sentence-based comb. > phrase-based comb.

Chen et al 2009

- Three German-English and three French-English MT systems (top1-prov, b-box)
- Two Re-decoding approach using Moses
  - A. Use the new phrase table
  - B. Use the new phrase table + existing phrase table
- Evaluation
  - German-English: Performance of A is almost the same as B
  - French-English: Performance of A is worse than B

Rosti et al 2007a Combining outputs from multiple machine translation systems

Chen et al 2009 Combining Multi-Engine Translations with Moses

Phrase-based Combination

Utilizing MT Engine

Rosti et al 2007a

- Six Arabic-English and six Chinese-English MT systems (topN-prov, g-box)
- Algorithm
  - Extracting a new phrase table from provided phrase alignment
  - Re-docoding source based on the new phrase table
- Phrase confidence score
  - Agreement model on four levels of similarity
  - Integrating weights of systems and levels of similarity
- Re-docoding: a standard beam search – Pharaoh
- Evaluation
  - Arabic-English: BLEU:+1.61 TER:-1.42 Chinese-English:BLEU:+0.03 TER:+0.20
  - Performance Comparison
    - Arabic-English: word-based comb. > phrase-based comb. > sentence-based comb.
    - Chinese-English: word-based comb. > sentence-based comb. > phrase-based comb.

Chen et al 2009

- Improvement on Rosti et al 2007a
  - Two Re-decoding approach using Moses
    - A. Use the new phrase table
    - B. Use the new phrase table + existing phrase table
- Evaluation
  - German-English: Performance of A is almost the same as B
  - French-English: Performance of A is worse than B

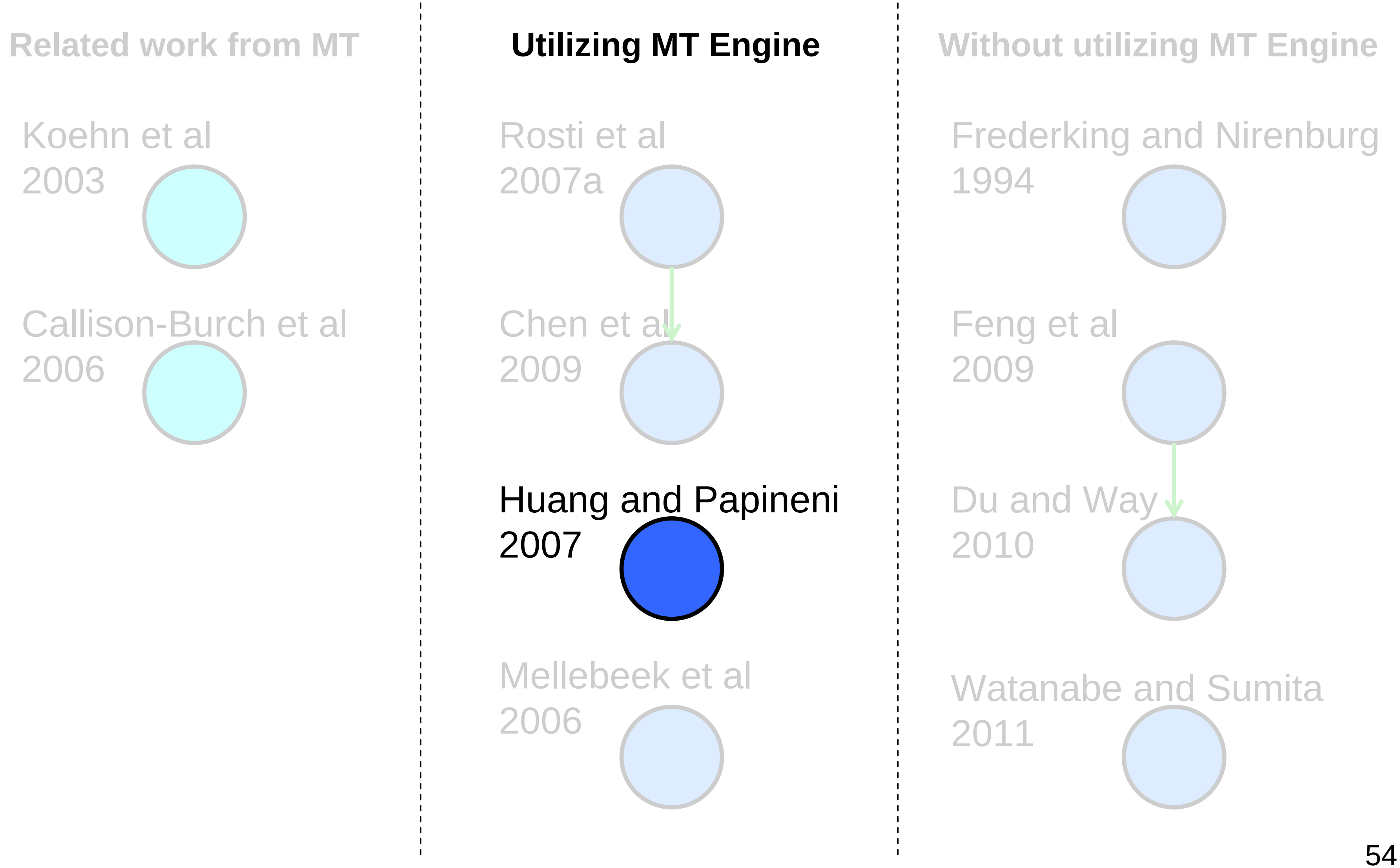
Rosti et al 2007a Combining outputs from multiple machine translation systems
Chen et al 2009 Combining Multi-Engine Translations with Moses

Methodology

MT combination paper

MT paper

Feature or model improvement



Phrase-based Combination

Utilizing MT Engine

Huang and Papineni 2007

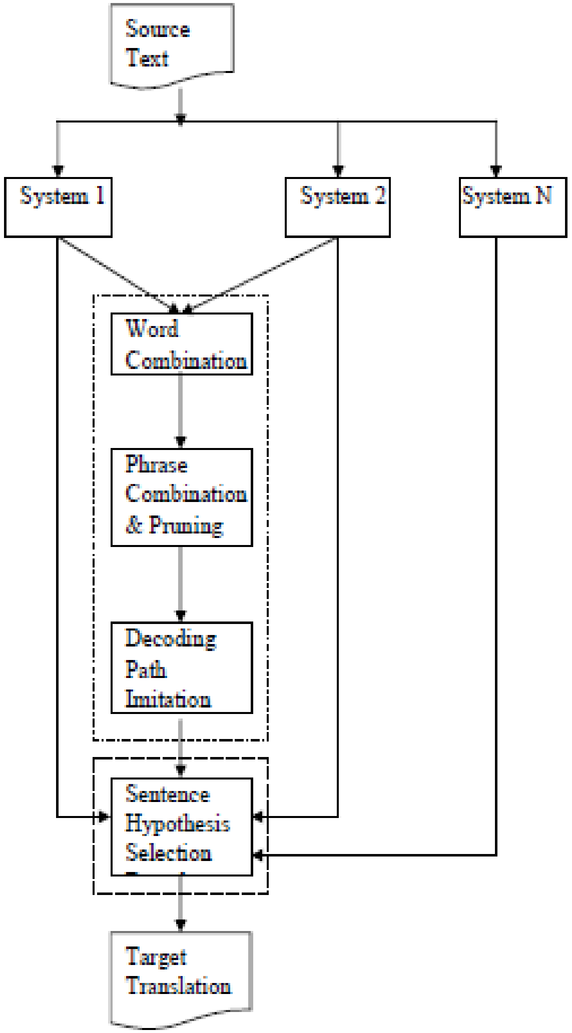
- Word-based Combination

$$t''(e|f) = \gamma t'(e|f) + (1-\gamma)t(e|f);$$

- Phrase-based Combination

$$P'(e|f) = \frac{C_b(f,e) + \sum \alpha_m C_m(f,e)}{C_b(f) + \sum \alpha_m C_m(f)},$$

  - Decoding path imitation of word order of system outputs
- Sentence-based Combination
  - Word LM and POS LM
- Evaluation
  - Decoding path imitation helps



Huang and Papineni 2007 Hierarchical system combination for machine translation

Methodology

MT combination paper

MT paper

Feature or model improvement

Related work from MT

Koehn et al
2003

Callison-Burch et al
2006

Utilizing MT Engine

Rosti et al
2007a

Chen et al
2009

Huang and Papineni
2007

Mellebeek et al
2006

Without utilizing MT Engine

Frederking and Nirenburg
1994

Feng et al
2009

Du and Way
2010

Watanabe and Sumita
2011

Phrase-based Combination

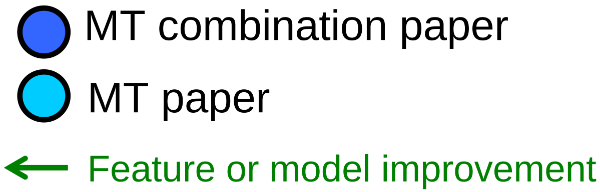
Utilizing MT Engine

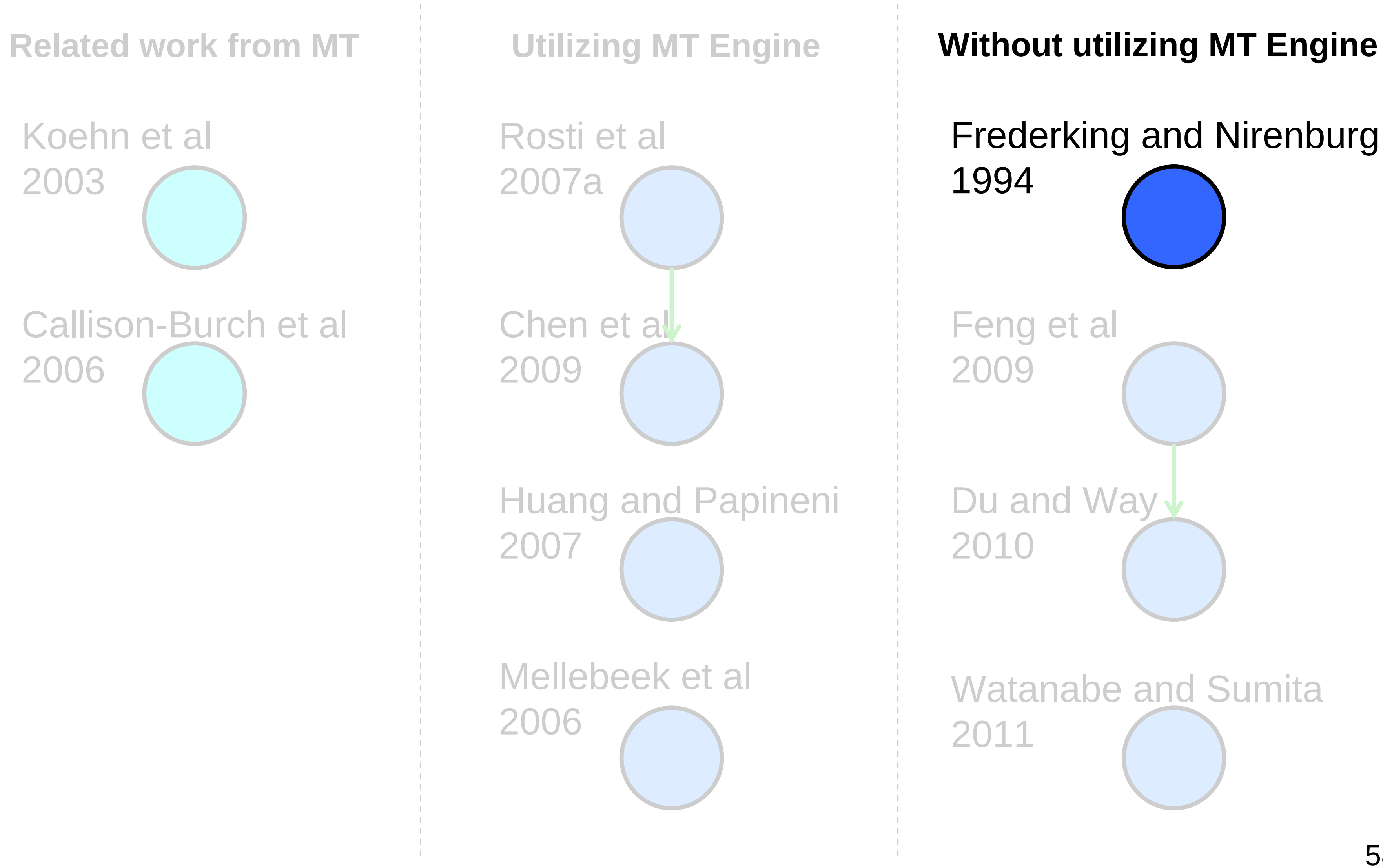
Mellebeek et al 2006

- Recursively do the following
  - decomposing source
  - translate each chunk by using different MT engines
  - select the best chunk translations through agreement, LM and confidence score.

Mellebeek et al 2006 Multi-Engine Machine Translation by Recursive Sentence Decomposition

Methodology





Phrase-based Combination
Without utilizing MT Engine

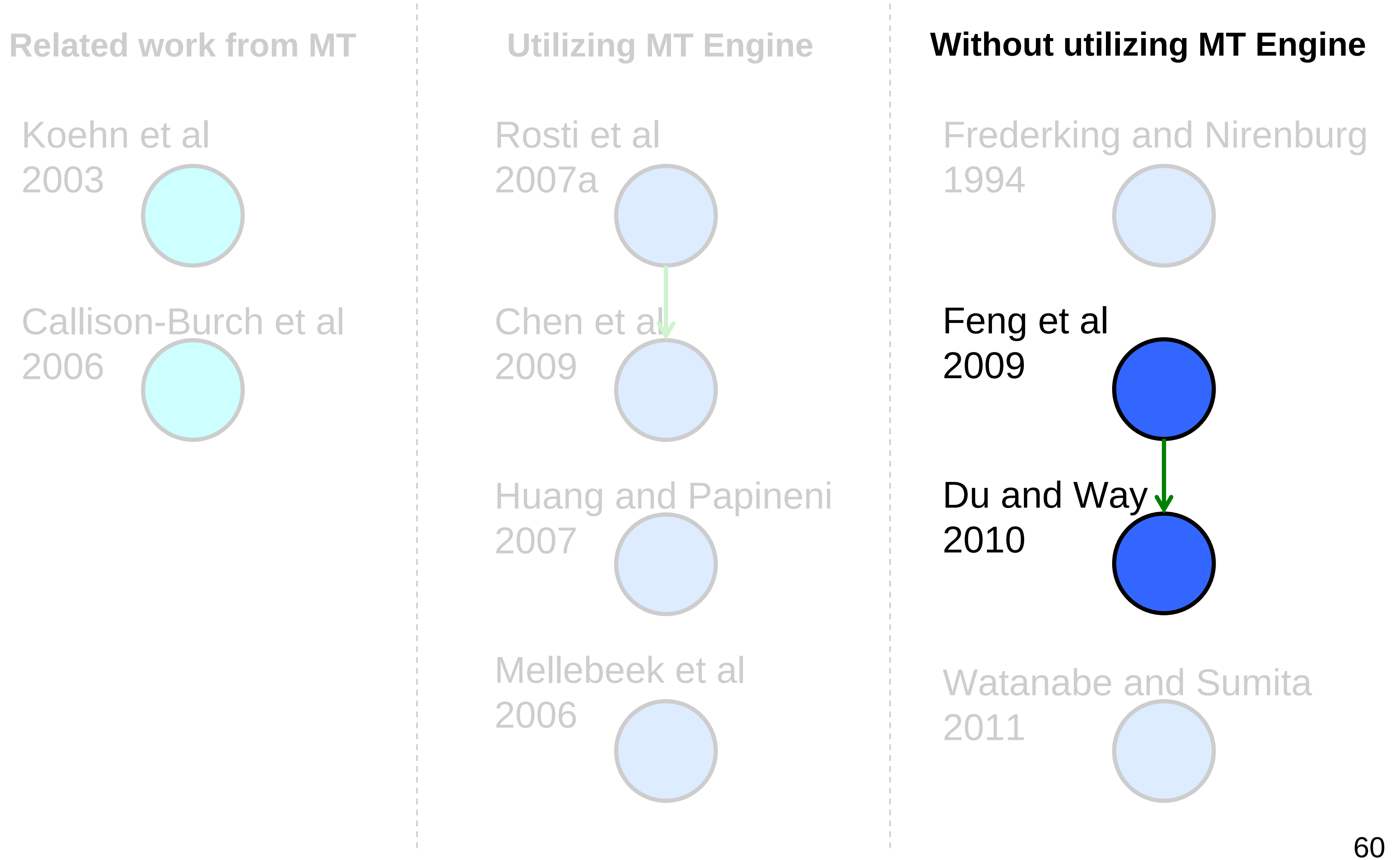
Frederking and Nirenburg 1994

- First MT combination paper
- Algorithm
  - Record target words, phrases and their source positions in a chart
  - Normalize the provided translation scores
  - Select the highest-score sequence of the chart that covers the source using a divide-and-conquer algorithm

Frederking and Nirenburg 1994 Three Heads are Better than One

Methodology

MT combination paper
MT paper
Feature or model improvement

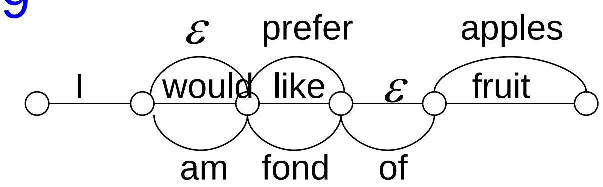


Phrase-based Combination

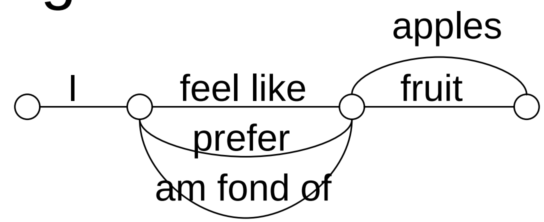
Without utilizing MT Engine

Feng et al 2009

- Motivation



VS



- Convert IHMM word alignments into phrase alignments by heuristic rules
- Construct Lattice based on phrase alignments by heuristic rules
- Evaluation
  - Baseline (IHMM word-based combination):+2.50 This paper: BLEU:+3.73

Du and Way 2010

- Difference with Feng et al 2009
  - Alignment tool: TERp (extending TER by using morphology, synonymy and paraphrases)
- Improvement on Feng et al 2009
  - Two-pass decoding algorithm
    - Combine synonym arcs or paraphrase arcs

- Evaluation: BLEU:+2.4

Feng et al 2009 Lattice-based system combination for statistical machine translation
Du and Way 2010 Using TERp to Augment the System Combination for SMT

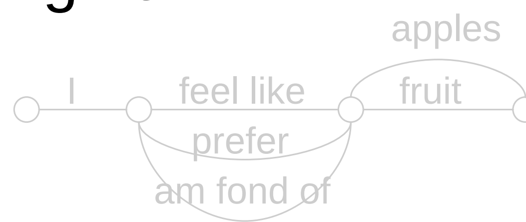
Phrase-based Combination

Without utilizing MT Engine

Feng et al 2009

- Motivation

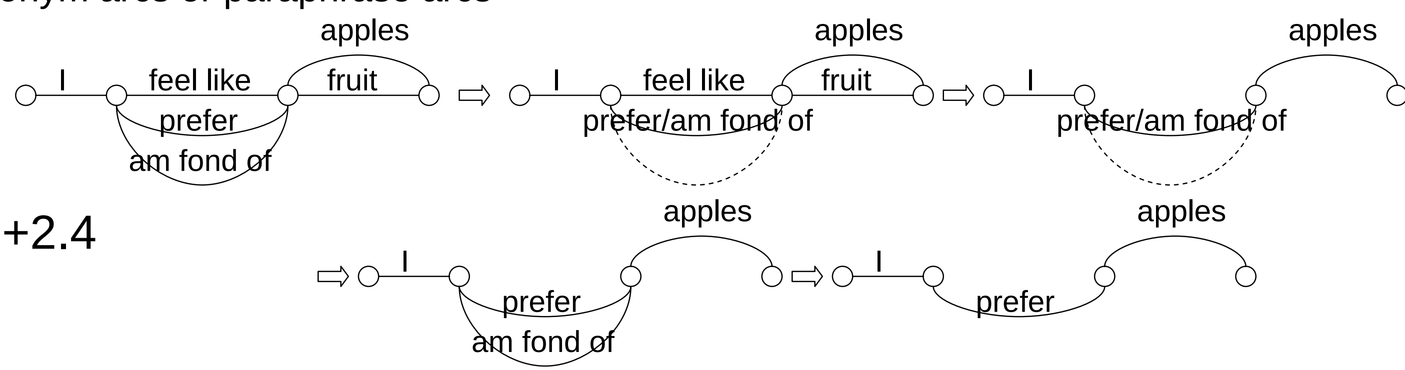
VS



- Convert IHMM word alignments into phrase alignments by heuristic rules
- Construct Lattice based on phrase alignments by heuristic rules
- Evaluation
  - Baseline (IHMM word-based combination):+2.50 This paper: BLEU:+3.73

Du and Way 2010

- Difference with Feng et al 2009
  - Alignment tool: TERp (extending TER by using morphology, synonymy and paraphrases)
- Improvement on Feng et al 2009
  - Two-pass decoding algorithm
    - Combine synonym arcs or paraphrase arcs



- Evaluation: BLEU:+2.4

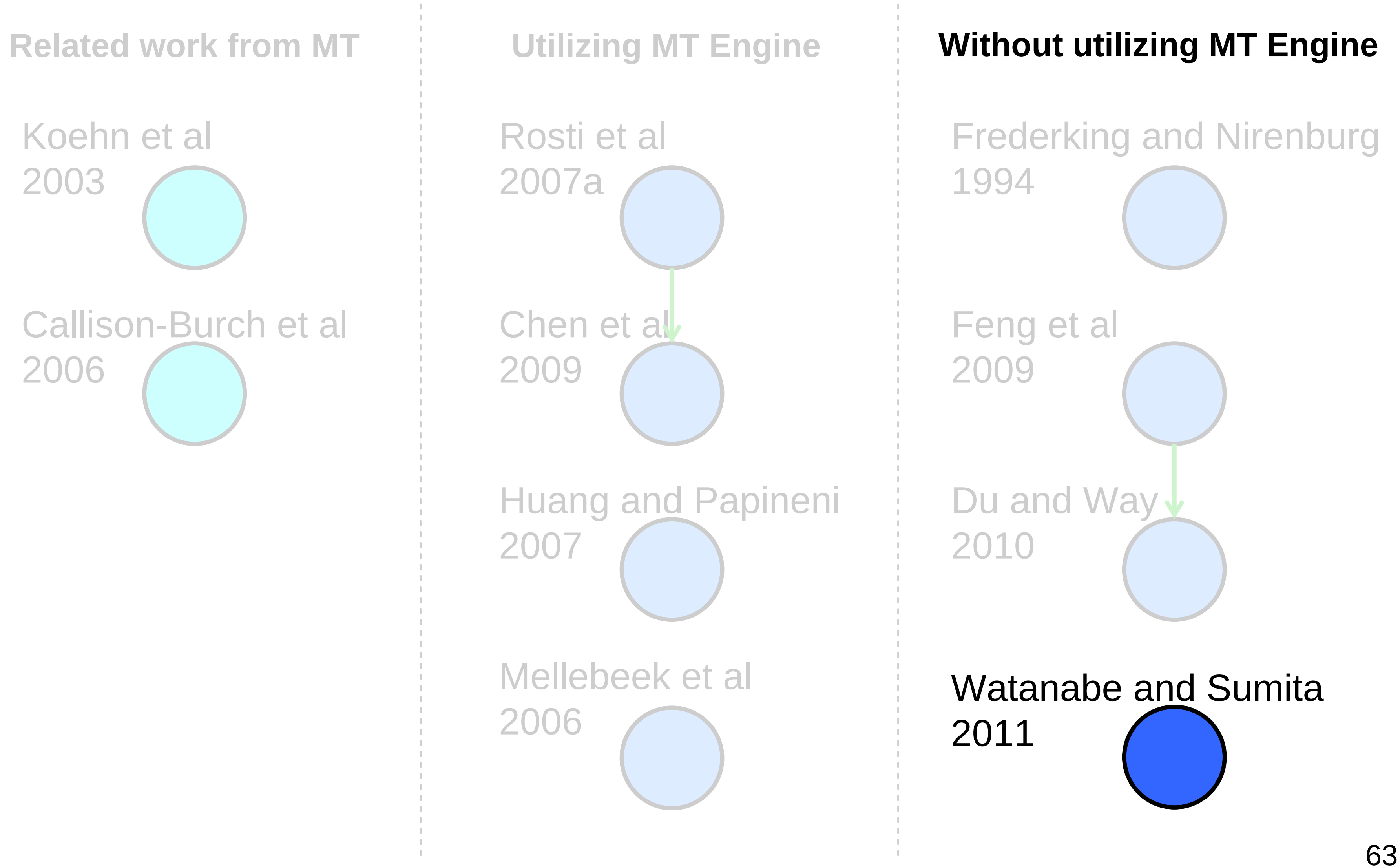
Feng et al 2009 Lattice-based system combination for statistical machine translation
Du and Way 2010 Using TERp to Augment the System Combination for SMT

Methodology

MT combination paper

MT paper

← Feature or model improvement



Phrase-based Combination

Without utilizing MT Engine

Watanabe and Sumita 2011

- Goal
  - Exploiting the syntactic similarity of system outputs
- Syntactic Consensus Combination
  - Step 1: parse MT outputs
  - Step 2: extract CFG rules
  - Step 3: generate forest by merging CFG rules
  - Step 4: searching the best derivation in the forest
- Evaluation
  - German-English:+0.48 French-English:+0.40

Watanabe and Sumita 2011 Machine Translation System Combination by Confusion Forest

Outline

- Sentence-based Combination
- Word-based Combination
- Phrase-based Combination
- Comparative Analysis
- Conclusion

Comparative Analysis

MT system analysis

Macherey and Och
2007

Alignment analysis

Chen et al
2009

Contest report

Callison-Burch et al
2011

Phrase-based Combination

Related work from MT

Macherey and Och 2007

- A set of experiments about system selection tells us:
  - The systems to be combined should be of similar quality and need to be almost uncorrelated
  - More systems are better

Chen et al 2009

- A set of experiments about word alignment used in single confusion network tells us:
  - For IWSLT corpus: IHMM(BLEU:31.74)>HMM(BLEU:31.40)>TER(31.36)
  - For NIST corpus: IHMM(BLEU:25.37)>HMM(BLEU:25.11)>TER(24.88)

Callison-Burch et al 2011

- The contest of MY combination tells us that what are the best MT combination systems in the world
- Three winners
  - BBN(Rosti et al 2007b)
  - CMU(Heafield and Lavie 2009)
  - RWTH(Matusov et al 2008)

Macherey and Och 2007 An Empirical Study on Computing Consensus Translations from Multiple Machine Translation Systems

Chen et al 2009 A Comparative Study of Hypothesis Alignment and its Improvement for Machine Translation System Combination

Callison-Burch et al 2011 Findings of the 2011 Workshop on Statistical Machine Translation

Outline

- Sentence-based Combination
- Word-based Combination
- Phrase-based Combination
- Comparative Analysis
- Conclusion

Conclusion

- Three Kinds of Combination Units
  – Sentence-based Combination
  – Word-based Combination
  – Phrase-based Combination
    - Retranslation from Source to Target
    - Target Phrase-based Combination
- Components
  – Alignments
    - HMM, TER, TERp, METEOR, IHMM
  – Scoring
    - LM, agreement model, confidence score

backup

Nomoto 2003

$$\begin{aligned} ALM(e, j_{(e)}) &= \log P(j_{(e)} \mid e) \\ &\approx \log P(j_{(e)}) P(e \mid j_{(e)}) \end{aligned}$$

Assume in addition that:

$$P(e \mid j_{(e)}) = \sum_{\mathbf{a}} P(e, \mathbf{a} \mid j_{(e)})$$

Regressive FLM (rFLM)

$$h(FLM(e, j)) = w \cdot FLM(e, j) + b$$

Regressive ALM (rALM)

$$h(ALM(e, j)) = w \cdot ALM(e, j) + b$$

A variant of rALM is also possible, where the fluency and alignment estimates are assigned to separate parameters, and takes the following form.

Regressive ALM$^+$ (rALM$^+$)

$$h(\vec{x}) = \vec{w} \cdot \vec{x} + b,$$

where $\vec{x} = (\log P(j), \log P(e \mid j))$.

Sentence-based Combination

Nomoto 2003

- Four English-Japanese MT systems (top1-prov, b-box)
- Fluency-based model (FLM): 4-gram LM
- Alignment-based model (ALM): lexical translation model - IBM model
- Regression toward sentence-based BLEU for
  - FLM $h(FLM(e,j)) = w \cdot FLM(e,j) + b$
  - ALM $h(ALM(e,j)) = w \cdot ALM(e,j) + b$
  - FLM+ALM $h(\vec{x}) = \vec{w} \cdot \vec{x} + b,$ where $\vec{x} = (\log P(j), \log P(e \mid j)).$
- Evaluation
  - Regression for FLM is the best (Bleu:+1)
- My comments
  - Unique MT combination paper using regression
  - Only sentence-based BLEU for regression is not enough, could try other metrics, such as TER

Nomoto 2003 Predictive Models of Performance in Multi-Engine Machine Translation

Sentence-based Combination

Hildebrand and Vogel. 2008

- Six Chinese-English MT systems (N-best-prov, b-box)
- 4-gram LM and 5-gram LM
- Six lexical translation models (Lex)
- Two agreement models:
  - Sum of position dependent N-best list word agreement score (WordAgr)

    Sys1: I prefer apples
    Sys2: I would like apples
    Freq(apples,3)=1, Freq(apples,4)=1

  - Sum of position independent N-best list N-gram agreement score (NgrAgr)

    Freq(prefer apples)=1, Freq(like apples)=1, Freq(apples)=2
- Evaluation
  - All features: Bleu:+2.3, TER:-0.4
  - Importance: LM>NgrAgr>WordAgr>Lex
- My comments
  - Valuable feature performance comparison
  - No system weight

Hildebrand and Vogel. 2008 Combination of machine translation systems via hypothesis selection from combined n-best lists

Sentence-based Combination

Zwarts and Dras. 2008

- The same Dutch-English MT engine but two systems (top1-prov, b-box)
  - $Source_{nonord}$ -> $Trans(Source_{nonord})$
  - $Source_{ord}$ -> $Trans(Source_{ord})$
- Syntactical features
  - Score of Parse($Source_{nonord}$), Score of Parse($Source_{ord}$), Score-of-Parse(Trans($Source_{nonord}$)), Score-of-Parse(Trans($Source_{ord}$))…etc
- Binary SVM Classifier to decide which one is better

  Trans($Source_{nonord}$) or Trans($Source_{ord}$)
- Evaluation
  - Score of Parsing Target is more useful than Score of Parsing Source
  - The SVM classifier's prediction score helps.
- My comments
  - Could add LM and translation model (also in the paper's future work)

Zwarts and Dras. 2008 Choosing the Right Translation: A Syntactically Informed Approach

MBR

$$\delta(F) = \underset{E',A'}{\operatorname{argmin}} \sum_{E,A} L((E,A),(E',A');F)P(E,A|F).$$

$$\hat{i} = \underset{i \in \{1,2,\ldots,N\}}{\operatorname{argmin}} \sum_{j=1}^{N} L((E_j,A_j),(E_i,A_i))P(E_j,A_j|F)$$

Loss Function	Functional Form
Lexical	$L(E,E')$
Target Language Parse-Tree	$L(T_E,T_{E'})$
Bilingual Parse-Tree	$L((T_E,A),(T_{E'},A');T_F)$

Word-based Combination

Single Confusion Network

Rosti et al 2007a

- Six Arabic-English and six Chinese-English MT systems (top10-prov, g-box)

Top10 Sys1 hyps
Top10 Sys2 hyps
Top10 Sys3 hyps

Backbone selection: MBR
(Loss function: TER)

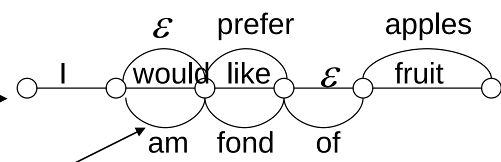
$$E_r = \arg\min_i \sum_{j=1}^{N_s} \mathrm{TER}(E_j, E_i)$$

⇨ **Sys1(3th): I would like fruit**

Alignment approach: TER
(tool: tercom)

Sys1(3th): I would like fruit
Sys2(2th): I prefer apples

Sys1(3th): I would like fruit
Sys3(5th): I am fond of apples



Score of this arc: SysWeight$_3$*1/(1+5)

Confidence score
for each word: 1/(1+rank)

- Evaluation
  - Arabic-English(News): BLEU:+2.3 TER:-1.34,
  - Chinese-English(News): BLEU:+1.1 TER:-1.96

Karakos et al 2008

- Nine Chinese-English MT systems (top1-prov, b-box)
- The well-known TER tool (tercom) is only an approximation of TER movements
- ITG-based alignment: minimum number of edits allowed by the ITG (nested block movements)

Ex : “thomas jefferson says eat your vegetables”
“eat your cereal thomas edison says”
tercom: 5 edits, ITG-based alignment: 3 edits

- Evaluation shows the combination using ITG-based alignment outperforms the combination using tercom by BLEU of 0.6 and TER of 1.3, but it is much slower.

Rosti et al 2007a Combining outputs from multiple machine translation systems
Karakos et al 2008 Machine Translation System Combination using ITG-based Alignments

Word-based Combination

Multiple Confusion Networks

Rosti et al 2007b

- Six Arabic-English and six Chinese-English MT systems (topN-prov, b-box)
- Difference with Rosti et al 2007a
  - Structure: From Single Confusion Network to Multiple Confusion Networks
  - Scoring: From only confidence scores to arbitrary features, such as LM

$$\log p(E_{j,n}|F_j) = \sum_{i=1}^{N_j-1} \log\left(\sum_{l=1}^{N_s} \lambda_l p(w|l,i)\right) + \nu L(E_{j,n}) + \mu N_{nulls}(E_{j,n}) + \xi N_{words}(E_{j,n})$$

- Evaluation
  - Arabic-English: BLEU:+3.2, TER:-1.7 (baseline:BLEU:+2.4, TER:-1.5)
  - Chinese-English: BLEU:+0.5, TER:-3.4 (baseline:BLEU:+1.1, TER:-2)

Ayan et al 2008

- Three Arabic-English and three Chinese-English MT systems (topN-prov, g-box)
  - Only one engine but use different training data
- Difference with Rosti et al 2007b
  - Extend TER script (tercom) with synonym matching operation using WordNet
  - Two-pass alignment strategy
  - Use translation score

Sys1: I like big blue balloons
Sys2: I like balloons
Sys3: I like blue kites

⇨

Intermediate ref. sent.:
I like blue balloons

⇨

I like blue balloons
Sys1: I like big blue balloons

I like blue balloons
Sys2: I like balloons

I like blue balloons
Sys3: I like blue kites

- Evaluation
  - No synon+No Two-pass: BLEU:+1.6 synon+No Two-pass: BLEU:+1.9
  - No synon+Two-pass: BLEU:+2.6 synon+Two-pass: BLEU:+2.9

Rosti et al 2007b Improved Word-Level System Combination for Machine Translation
Ayan et al 2008 Improving alignments for better confusion networks for combining machine translation systems

Word-based Combination

Multiple Confusion Networks

Matusov et al 2006

- Five Chinese-English and four Spanish-English MT systems (top1-prov, b-box)
- Alignment approach: HMM model bootstrapped from IBM model1

$$p(e_1'^J \mid e_1^I) = \sum_{a_1^J} \prod_{j=1}^{J} \left[p(a_j \mid a_{j-1}, I) p(e_j' \mid e_{a_j}) \right] \qquad p(a_j = i \mid a_{j-1} = i', I) = \frac{c(i - i')}{\sum_{l=1}^{I} c(l - i')}$$

- Confidence score for each word: system-weighted vo
- Rescoring for confusion network outputs by general LM
- Evaluation
  - Chinese-English: BLEU:+5.9   Spanish-English: BLEU:+1.6
- My comments
  - Efficiency for online system could be a problem

Matusov et al 2008

- Six English-Spanish and six Spanish-English MT systems (top1-prov, b-box)
- Difference with Matusov et al 2006
  - Integrate general LM and adapted LM into confusion network decoding
    - adapted LM: N-gram based on system outputs
  - Handling long sentences by splitting them
- Evaluation
  - English-Spanish: BLEU:+2.1   Spanish-English: BLEU:+1.2
  - adapted LM is more useful than general LM in either confusion network decoding or rescoring

---

Matusov et al 2006 Computing consensus translation from multiple machine translation systems using enhanced hypotheses alignment
Matusov et al 2008 System combination for machine translation of spoken and written language

Word-based Combination

Multiple Confusion Networks

He et al 2008

- Eight Chinese-English (topN-prov, b-box)
- Alignment approach: Indirect HMM (IHMM) $p(e_1'^J \mid e_1^I) = \sum_{a_1^J} \prod_{j=1}^{J} \left[p(a_j \mid a_{j-1}, I) p(e_j' \mid e_{a_j}) \right]$

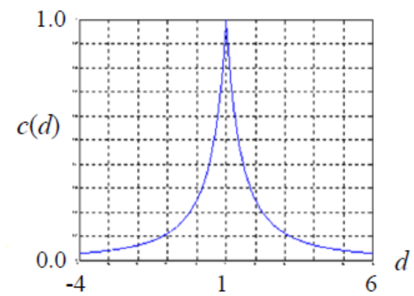
$$p(e_j' \mid e_i) = \alpha \cdot p_{sem}(e_j' \mid e_i) + (1-\alpha) \cdot p_{sur}(e_j' \mid e_i)$$

$$p_{sem}(e_j' \mid e_i) = \sum_{k=0}^{K} p(f_k \mid e_i) p(e_j' \mid f_k, e_i) \approx \sum_{k=0}^{K} p(f_k \mid e_i) p(e_j' \mid f_k)$$

$$p(a_j = i \mid a_{j-1} = i', I) = \frac{c(i-i')}{\sum_{l=1}^{I} c(l-i')}$$

define 11 buckets: c(<=-4), c(-3), ... c(0), ..., c(5), C(>=6)

$$c(d) = \left(1 + |d-1|\right)^{-\kappa}, d = -4, \dots,$$



- Evaluation
  - Baseline (alignment: TER): BLEU:+3.7
  - This paper (alignment: IHMM): BLEU:+4.7

Zhao and He 2009

- Some Chinese-English MT systems (topN-prov, b-box)
- Difference with He et al 2008
  - Add agreement model: online N-gram LM and N-gram voting feature
- Evaluation
  - Baseline (He et al 2008): BLEU:+4.3 This paper: BLEU:+5.11

He et al 2008 Indirect-hmm-based hypothesis alignment for computing outputs from machine translation systems

Zhao and He 2009 Using n-gram based features for machine translation system combination

IHMM

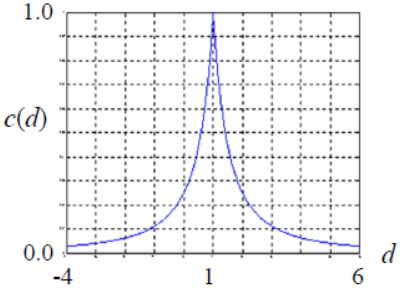
$$p(e_1'^J \mid e_1^I) = \sum_{a_1^J} \prod_{j=1}^{J} \left[p(a_j \mid a_{j-1}, I) p(e_j' \mid e_{a_j}) \right]$$

$$p(e_j' \mid e_i) = \alpha \cdot p_{sem}(e_j' \mid e_i) + (1-\alpha) \cdot p_{sur}(e_j' \mid e_i)$$

$$p(a_j = i \mid a_{j-1} = i', I) = \frac{c(i-i')}{\sum_{l=1}^{I} c(l-i')}$$

define 11 buckets: c(<=-4), c(-3), ... c(0), ..., c(5), C(>=6)

$$c(d) = \left(1 + \left|d-1\right|\right)^{-\kappa},\ d = -4, \ldots,$$



$$p_{sem}(e_j' \mid e_i)$$
$$= \sum_{k=0}^{K} p(f_k \mid e_i) p(e_j' \mid f_k, e_i)$$
$$\approx \sum_{k=0}^{K} p(f_k \mid e_i) p(e_j' \mid f_k)$$

$$p_{sur}(e_j' \mid e_i) = \exp\left\{\rho \cdot \left[s(e_j', e_i) - 1 \right]\right\}$$

$$p(e_j' \mid f_k) = p_{s2t}(e_j' \mid f_k)$$

where $s(e_j', e_i)$ is computed as

$$s(e_j', e_i) = \frac{M(e_j', e_i)}{\max(|e_j'|, |e_i|)}$$

$$p(f_k \mid e_i) = \frac{p_{t2s}(f_k \mid e_i)}{\sum_{k=0}^{K} p_{t2s}(f_k \mid e_i)}$$

Joint Optimization

$$w^* = \underset{w \in \boldsymbol{W}, O \in \boldsymbol{O}, C \in \boldsymbol{C}}{\operatorname{argmax}} \; exp\left\{\sum_{i=1}^{F} \alpha_i \cdot f_i(w, O, C, \boldsymbol{H})\right\}$$

$$f_{wp}(w, O, C, \boldsymbol{H}) = \sum_{m=1}^{M} log\big(P(w_m | CS_m)\big)$$

$$P\left(w_{i,l_i} | CS\right) = P\left(w_{i,l_i} \middle| CS(l_1, \ldots, l_N)\right) = \sum_{k=1}^{N} W(k)\, \delta(w_{k,l_k} = w_{i,l_i})$$

$$d(CS_m, CS_{m+1}) = \sum_{k=1}^{N} W(k) \cdot |l_{m,k} - l_{m+1,k}|$$

$$f_{dis}(w, O, C, \boldsymbol{H}) = -\sum_{m=1}^{M-1} d(CS_m, CS_{m+1})$$

$$P(\langle w_i, w_{i+1}\rangle | \boldsymbol{H}) = \sum_{k=1}^{N} W(k)\, \delta(\langle w_i, w_{i+1}\rangle \in h_i)$$

And the global bi-gram voting feature is defined as:

$$f_{bgv}(w, O, C, \boldsymbol{H}) = \sum_{i=1}^{|w|-1} log\big(P(\langle w_i, w_{i+1}\rangle | \boldsymbol{H})\big)$$

$$p\left(w_{j,l_j}, w_{k,l_k}\right) = \frac{1}{2}\Big(p(a_{l_j} = l_k | h_j, h_k) + p(a_{l_k} = l_j | h_k, h_j)\Big)$$

$$p(j|CS) = \prod_{\substack{k=1 \\ k \neq j}}^{N} p\left(w_{j,l_j}, w_{k,l_k}\right)$$

$$f_{aln}(w, O, C, \boldsymbol{H}) = \sum_{m=1}^{M} S_{aln}(CS_m)$$

$$Ent(CS) = Ent(CS(l_1, \ldots, l_N)) = \sum_{i=1}^{N'} P(w_{i,l_i} | CS) log P(w_{i,l_i} | CS)$$

$$f_{ent}(w, O, C, \boldsymbol{H}) = \sum_{m=1}^{M} Ent(CS_m)$$

Synchronize extensions of hypotheses

Phrases

Detect phrases using maximal consecutive alignments

Tie punctuation to the preceding word

Constrain decoding to complete phrases if possible

However , it is not yet won .

However , it is still not won .

Synchronization Example

1. The decoder can pick the first unused word from either system.

   Most people always takes over a cell phone .

   The majority of the people is always a mobile .

2. Suppose the decoder picks "Most", marking it used.

   Most people always takes over a cell phone .

3. Looking at alignments, system 2 is behind by 4 words.

   Most people always takes over a cell phone .

   The majority of the people is always a mobile .

   4 3 2 1 0

4. Words are marked used to synchronize within tolerance.

   The majority of the people is always a mobile .

Watanabe and Sumita 2011

*	I	saw	the		forest		
	I	walked	the	blue	forest		
	I	saw	the		green	trees	
			the		forest	was	found