“It’s reasonably easy to determine what language the data is in…”

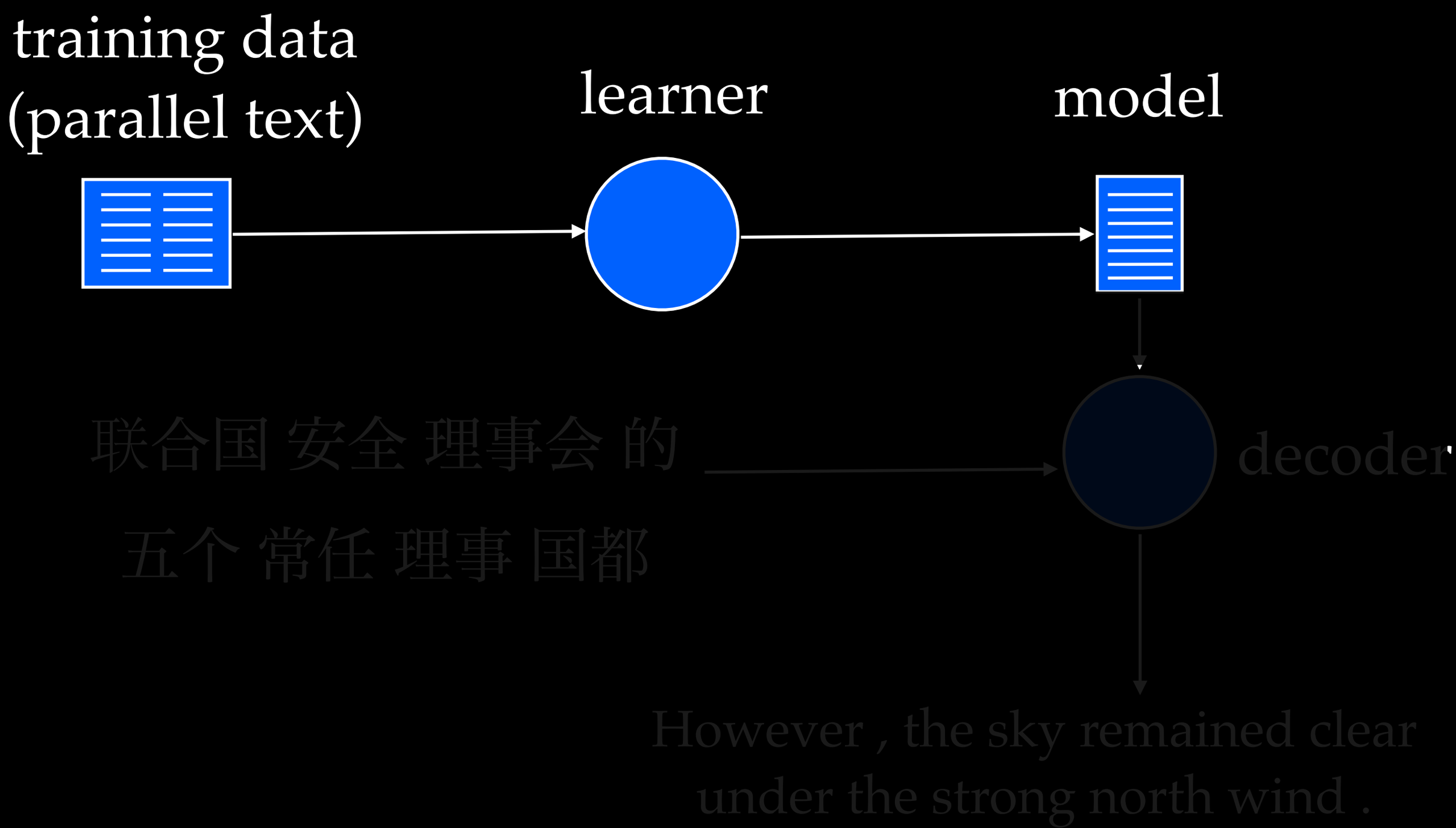
"It's reasonably easy to determine what language the data is in..."

"It's reasonably easy to determine what language the data is in..."

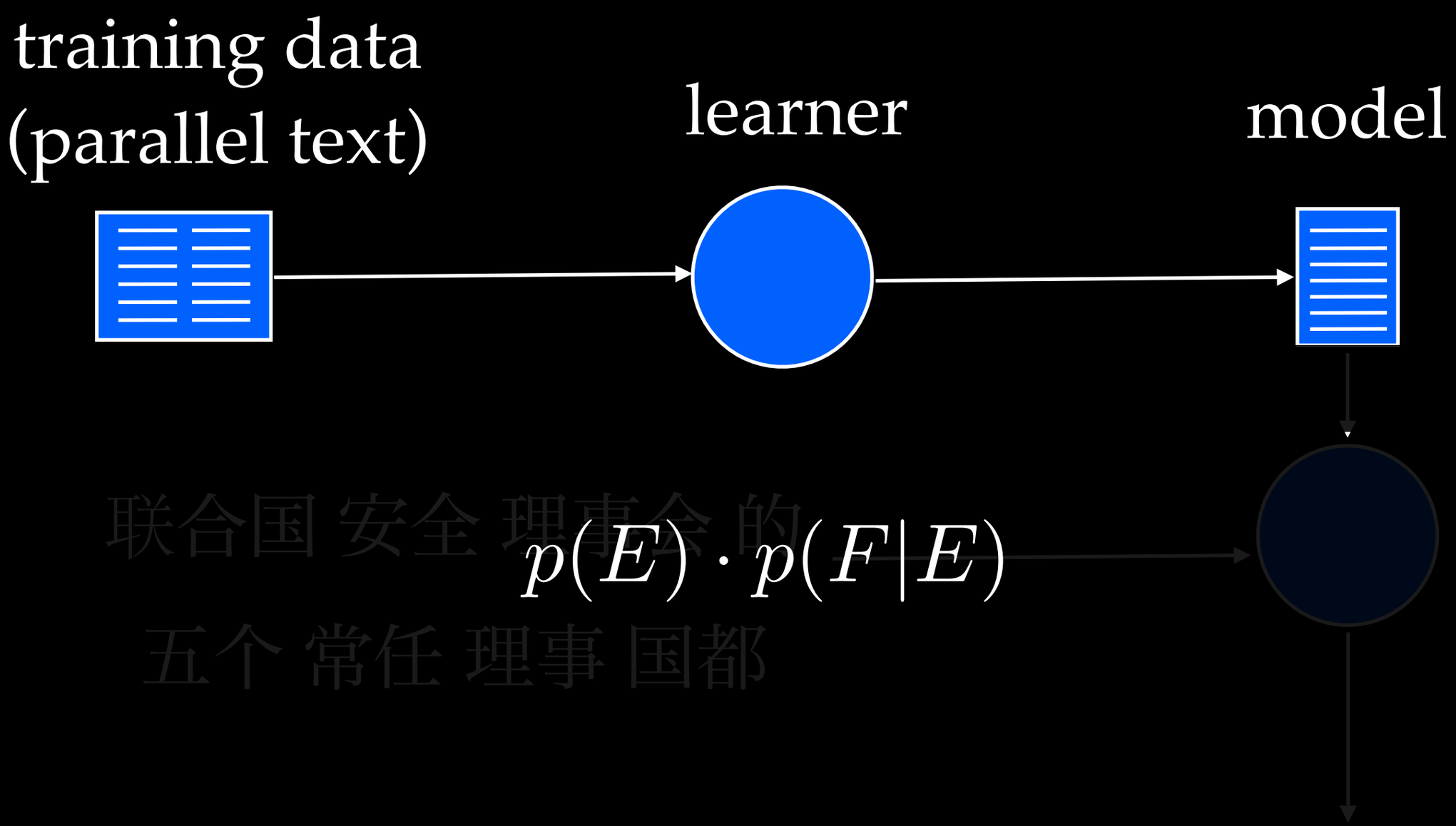
linguist deduces that French = FORTRAN

Learning Better Translation Models

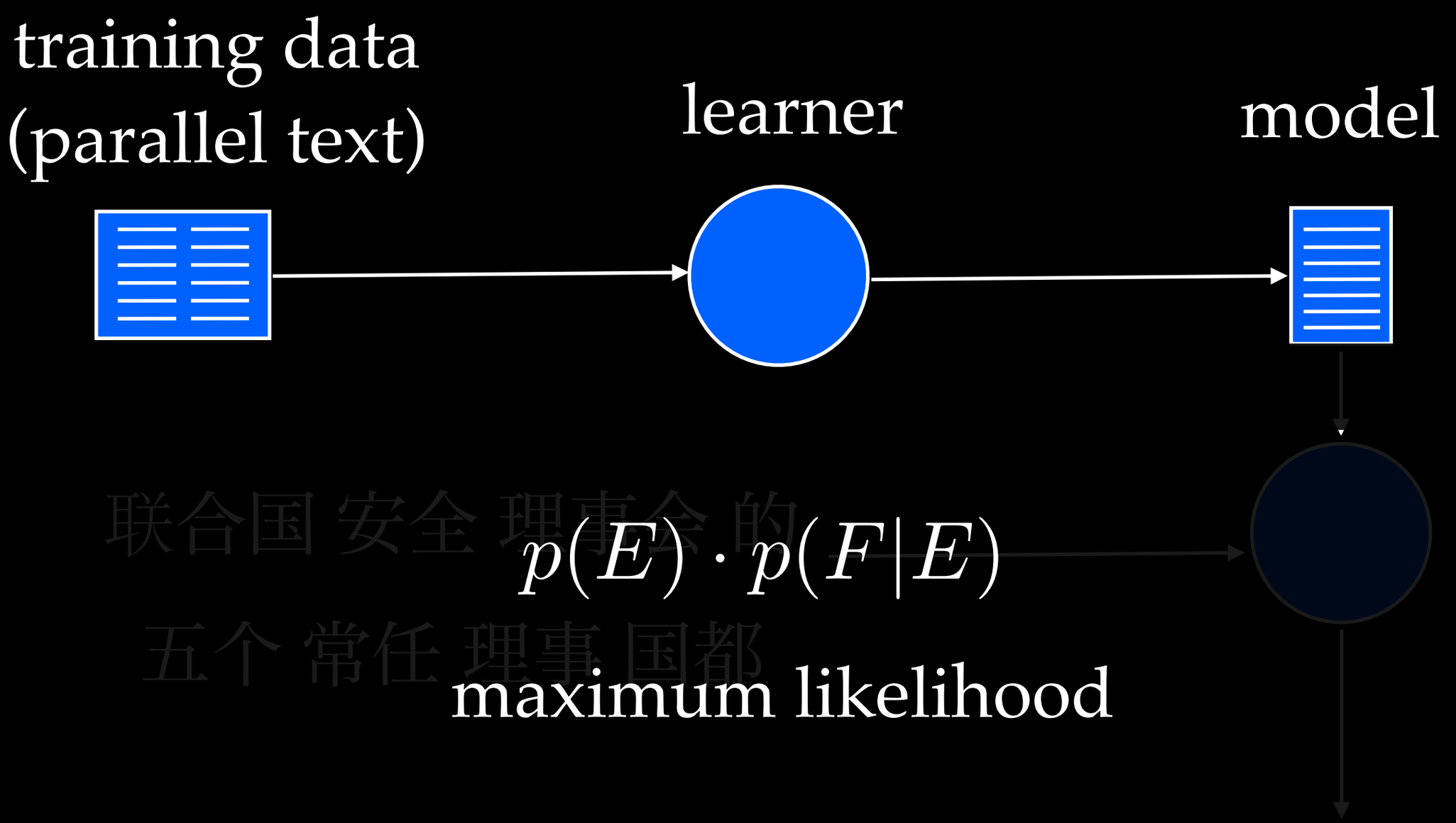
The Story So Far...



The Story So Far...



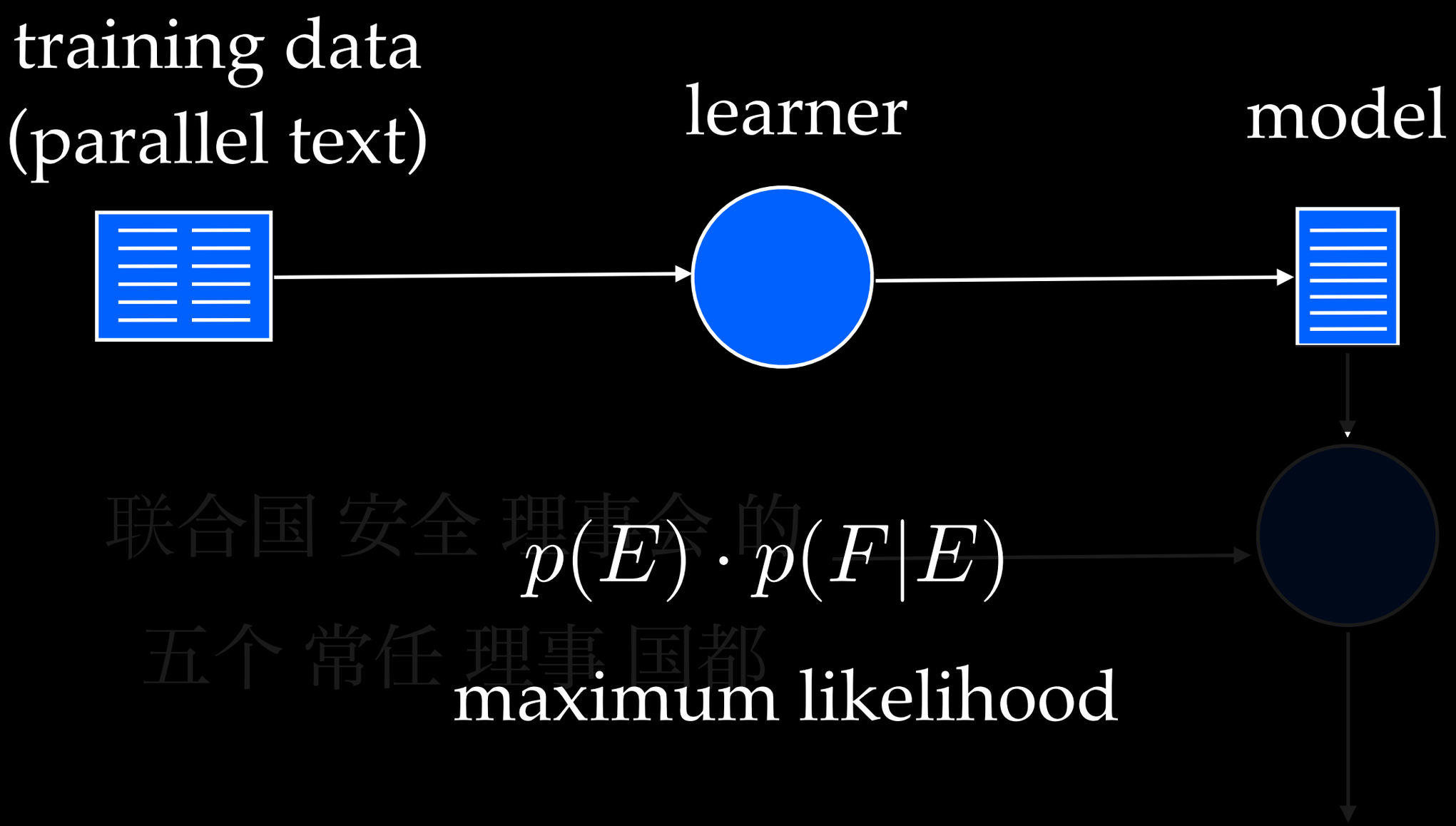
The Story So Far...



$$p(E) \cdot p(F|E)$$

maximum likelihood

The Story So Far...

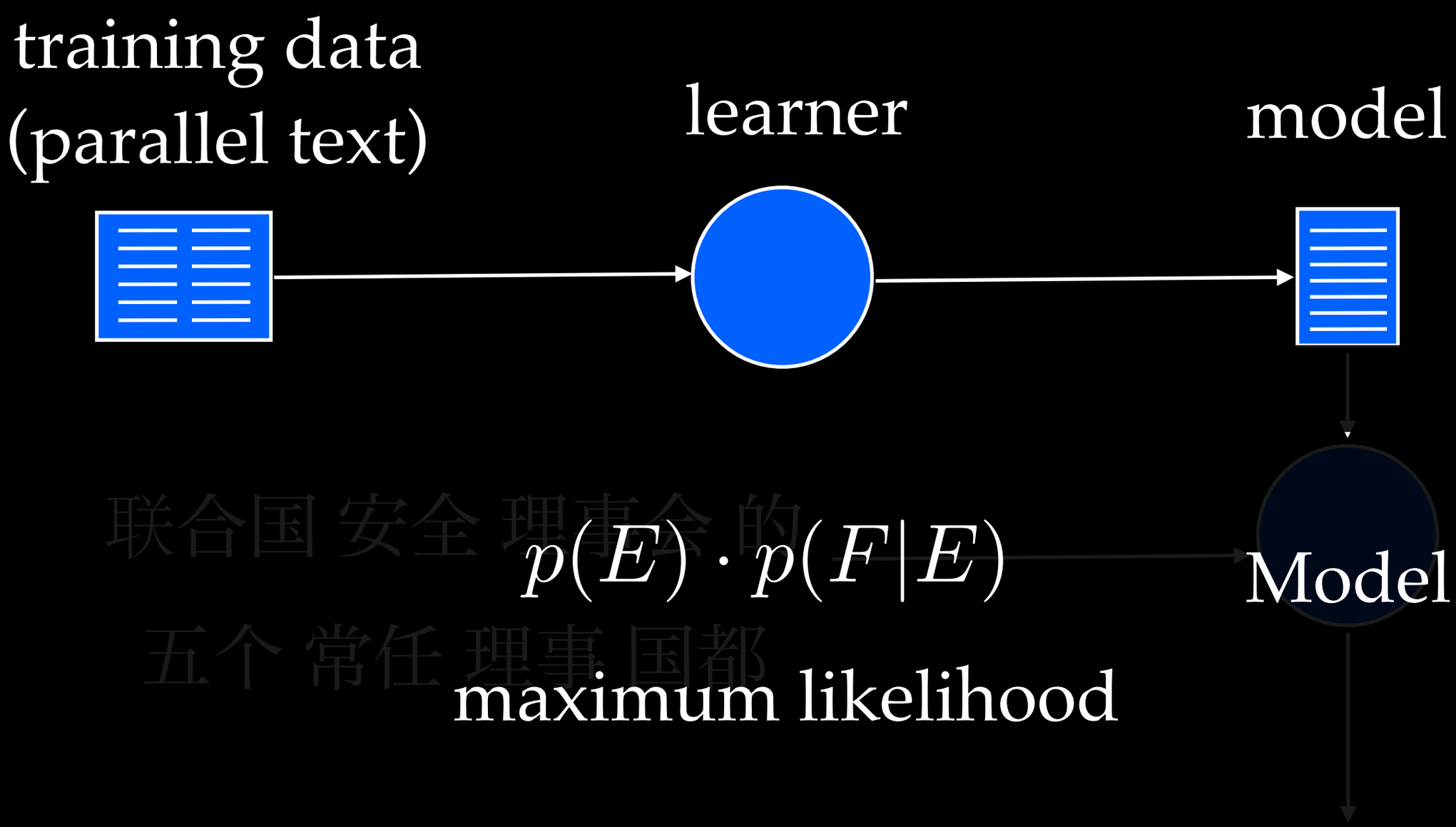


$$p(E) \cdot p(F|E)$$

maximum likelihood

expectation maximization

The Story So Far...



Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

Model 1

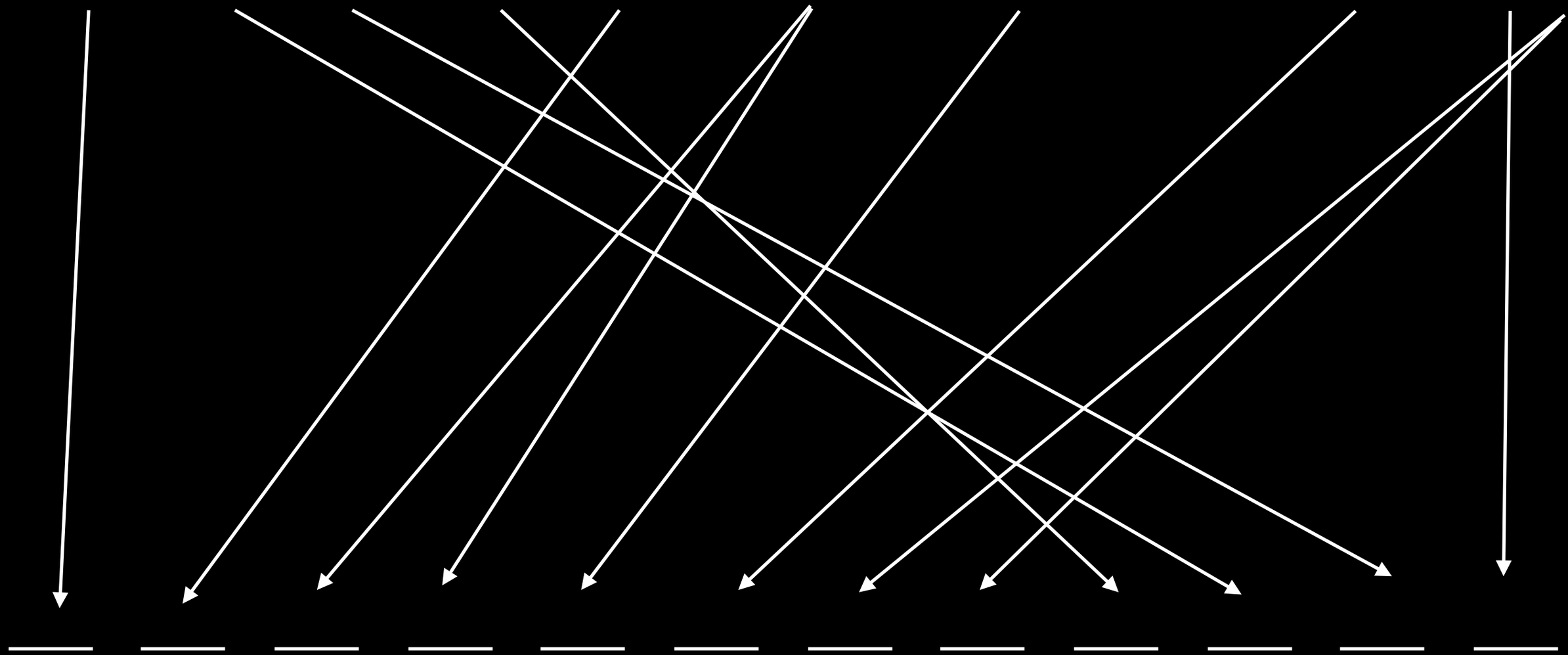
Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

Model 1

Although north wind howls , but sky still very clear .

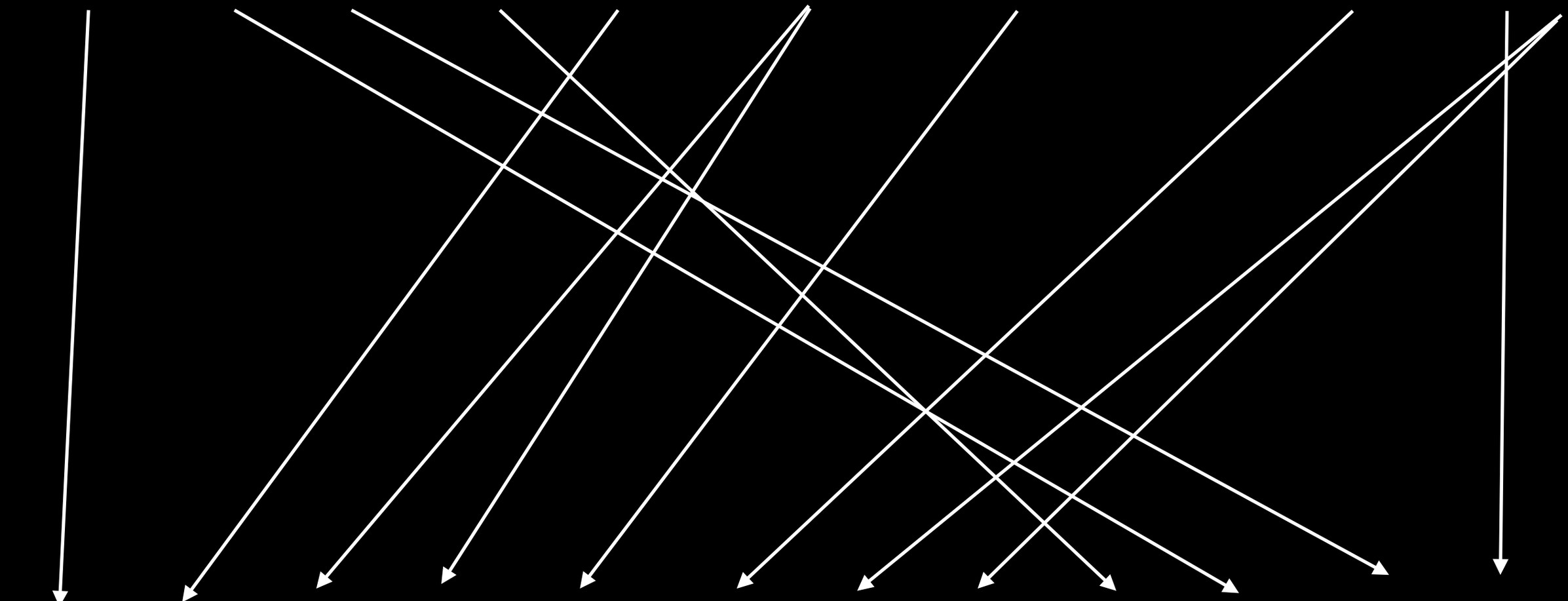
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε



Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

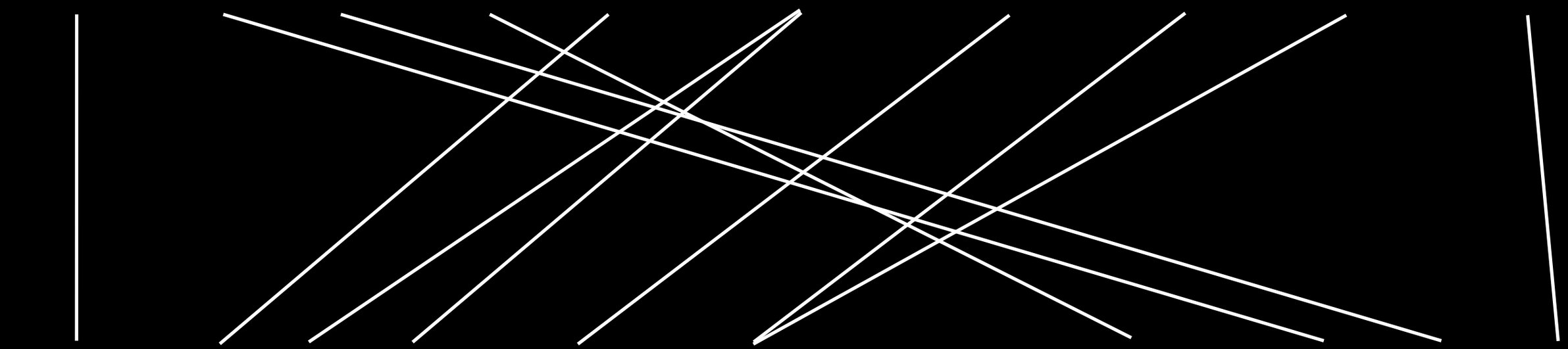


However , the sky remained clear under the strong north wind .

Model 1

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



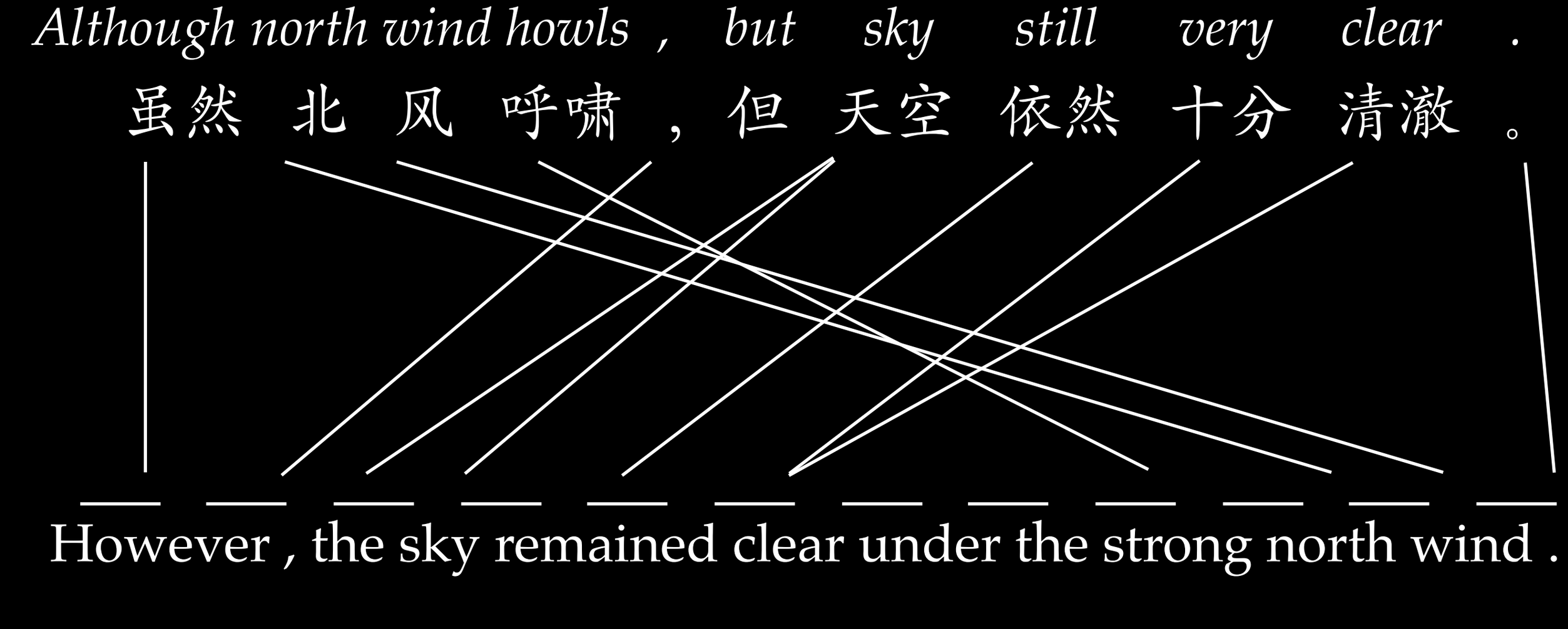
However , the sky remained clear under the strong north wind .

Model 1

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j) p(f_i|e_j)$$

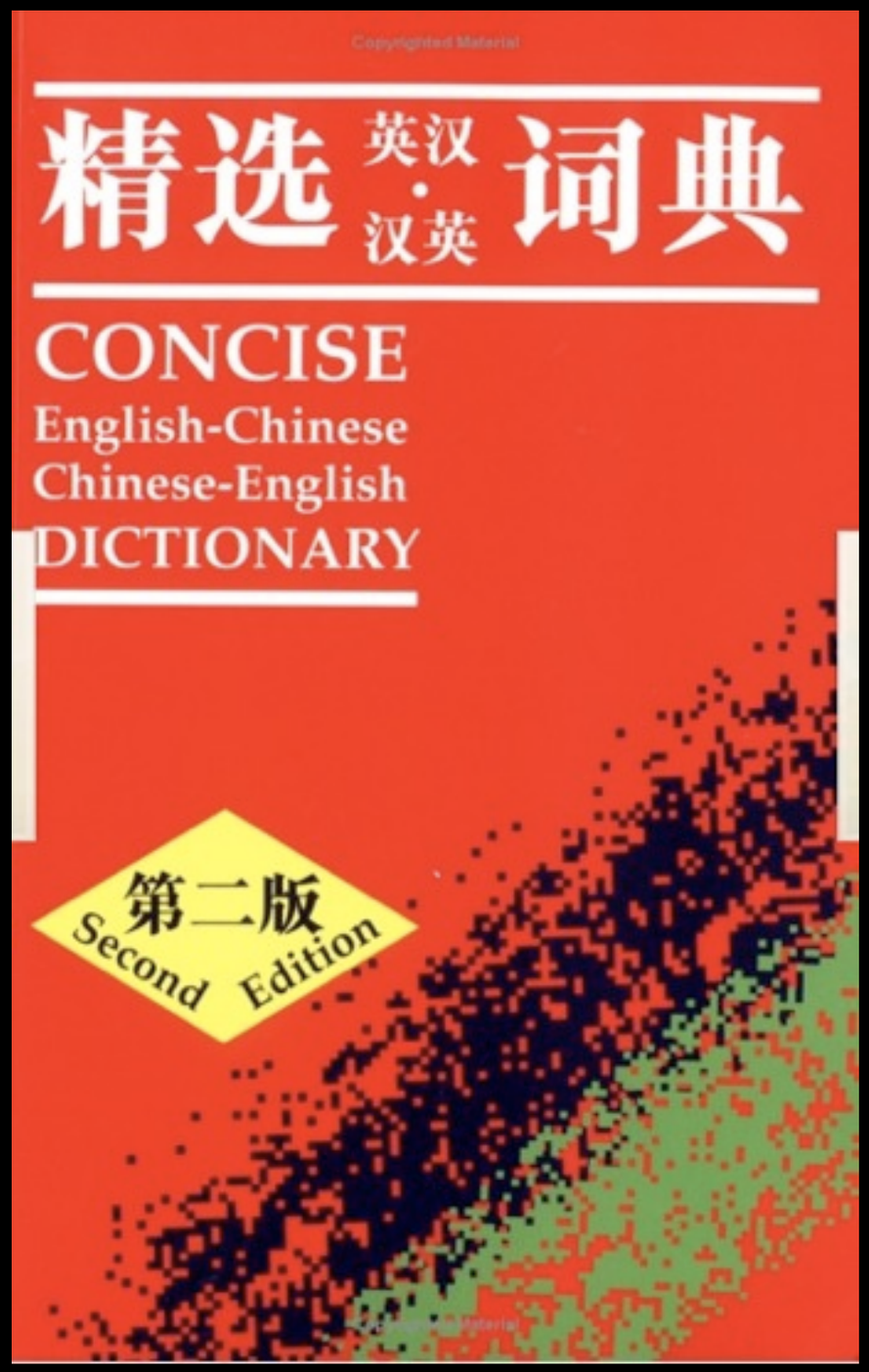
Model 1

$$p(F, A|E) = \overbrace{p(I|J)}^{\text{length}} \prod_{a_i} \overbrace{p(a_i = j)}^{\text{alignment}} \overbrace{p(f_i|e_j)}^{\text{translation}}$$



IBM Model 1: Pros

- Easy to understand.
- Model of lexical translation: seems somewhat natural.
- EM objective is convex.
- Expectations can be computed efficiently.



IBM Model 1: Cons

IBM Model 1: Cons

- No account of word order

IBM Model 1: Cons

- No account of word order
- No control over multi-word alignments

IBM Model 1: Cons

- No account of word order
- No control over multi-word alignments
- "Garbage collection"

IBM Model 1: Cons

- No account of word order
- No control over multi-word alignments
- “Garbage collection”
- Asymmetry

IBM Model 1: Cons

- No account of word order
- No control over multi-word alignments
- “Garbage collection”
- Asymmetry
- No awareness of morphology

IBM Model 1: Cons

- No account of word order
- No control over multi-word alignments
- “Garbage collection”
- Asymmetry
- No awareness of morphology
- No syntactic generalization

IBM Model 1: Cons

- No account of word order
- No control over multi-word alignments
- "Garbage collection"
- Asymmetry
- No awareness of morphology
- No syntactic generalization
- Can we do better?

IBM Model 1: Cons

- *No account of word order*
- No control over multi-word alignments
- "Garbage collection"
- Asymmetry
- No awareness of morphology
- No syntactic generalization
- Can we do better?

Model 1 Again ...

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

Model 1 Again ...

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

Model 1 Again ...

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

Model 1 Again ...

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

虽然 ，天空 天空 依然清澈 ε ε 呼啸 北 风 。

— — — — — — — — — — — — — —

Model 1 Again ...

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

虽然 ，天空 天空 依然清澈 ε ε 呼啸 北 风 。

However , the sky remained clear under the strong north wind .

Model 1 Again ...

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

hidden state sequence

虽然 ，天空 天空 依然清澈 ε ε 呼啸 北 风 。

However , the sky remained clear under the strong north wind .

Model 1 Again ...

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

hidden state sequence

虽然 ，天空 天空 依然清澈 ε ε 呼啸 北 风 。

However , the sky remained clear under the strong north wind .

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j) p(f_i|e_j)$$

... to Hidden Markov Model

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

hidden state sequence

However , the sky remained clear under the strong north wind .

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j|a_{i-1}) p(f_i|e_j)$$

... to Hidden Markov Model

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

hidden state sequence

However , the sky remained clear under the strong north wind .

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j|a_{i-1}) p(f_i|e_j)$$

... to Hidden Markov Model

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

states

hidden state sequence

However , the sky remained clear under the strong north wind .

$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j|a_{i-1} - j) p(f_i|e_j)$$

... to Hidden Markov Model

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

distortion=4

states

hidden state sequence

虽然 ，天空 天空 依然清澈 ε ε 呼啸 北 风 。

However , the sky remained clear under the strong north wind .

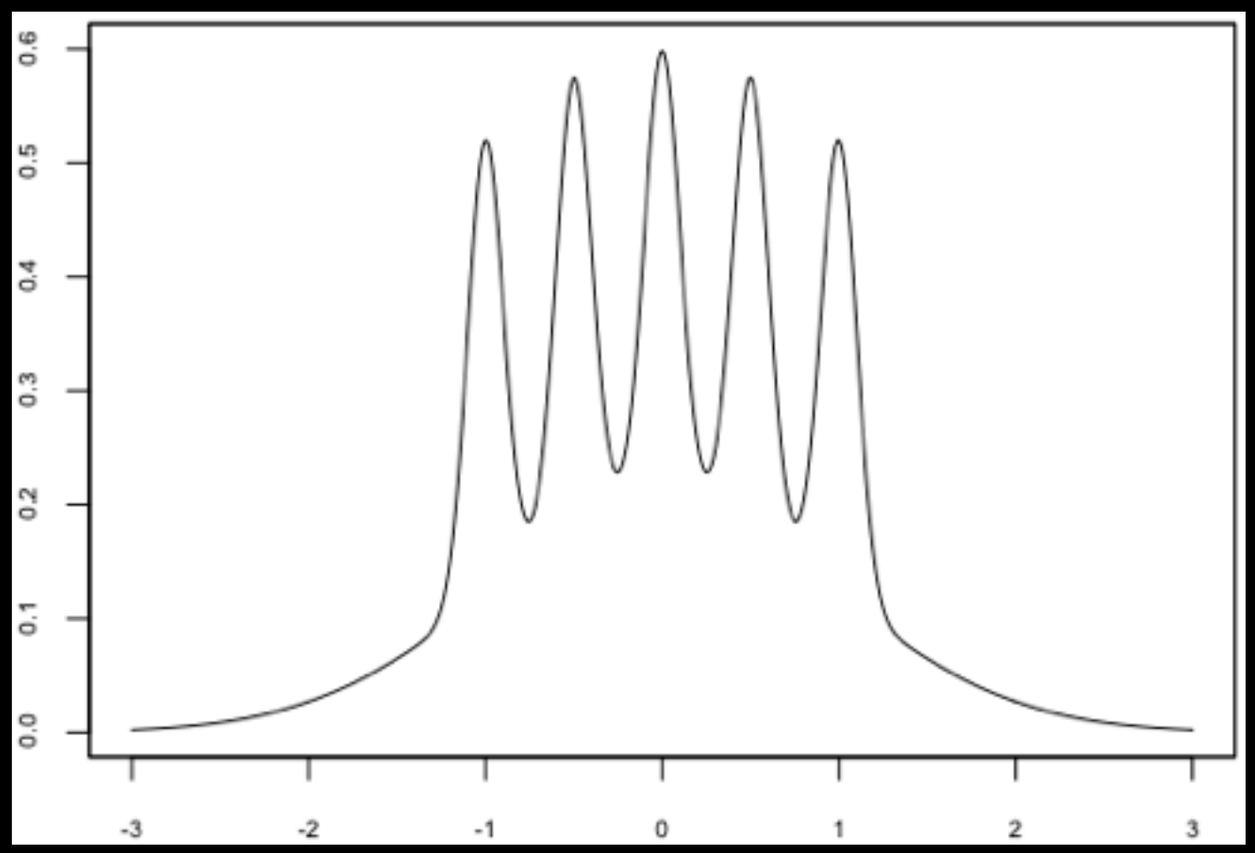
$$p(F, A|E) = p(I|J) \prod_{a_i} p(a_i = j|a_{i-1} - j) p(f_i|e_j)$$

EM for HMM

- Forward-backward algorithm: computation of marginals.
- Good: still exact, polynomial-time.
- Bad: EM objective is no longer convex.

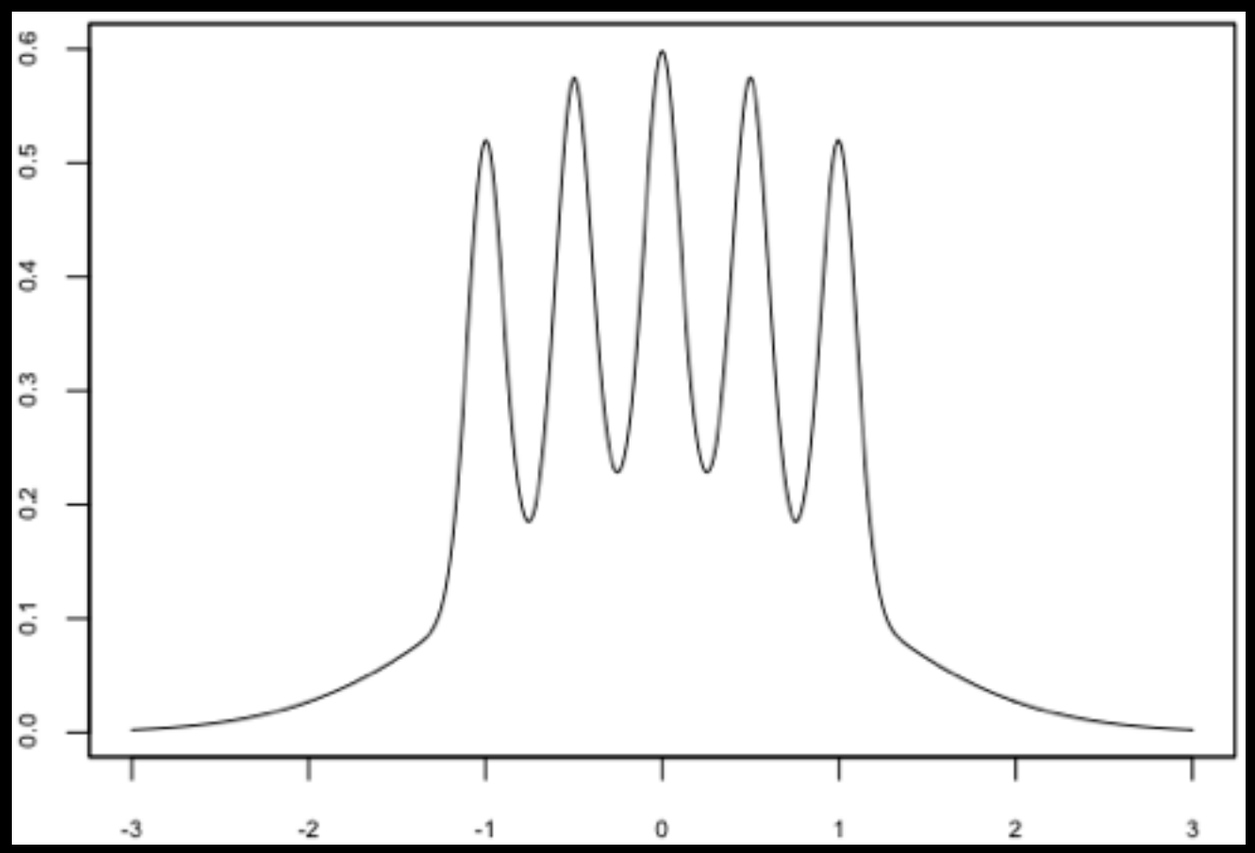
EM for HMM

- Forward-backward algorithm: computation of marginals.
- Good: still exact, polynomial-time.
- Bad: EM objective is no longer convex.



EM for HMM

- Forward-backward algorithm: computation of marginals.
- Good: still exact, polynomial-time.
- Bad: EM objective is no longer convex.



Use Model 1 to initialize translation parameters!

Hidden Markov Model: Cons

- ~~*No account of word order*~~
- No control over multi-word alignments
- "Garbage collection"
- Asymmetry
- No awareness of morphology
- No syntactic generalization
- Can we do better?

Hidden Markov Model: Cons

- ~~*No account of word order*~~
- *No control over multi-word alignments*
- "Garbage collection"
- Asymmetry
- No awareness of morphology
- No syntactic generalization
- Can we do better?

IBM Model 4

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

IBM Model 4

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

↓

虽然

IBM Model 4

Although north wind howls , but sky still very clear .

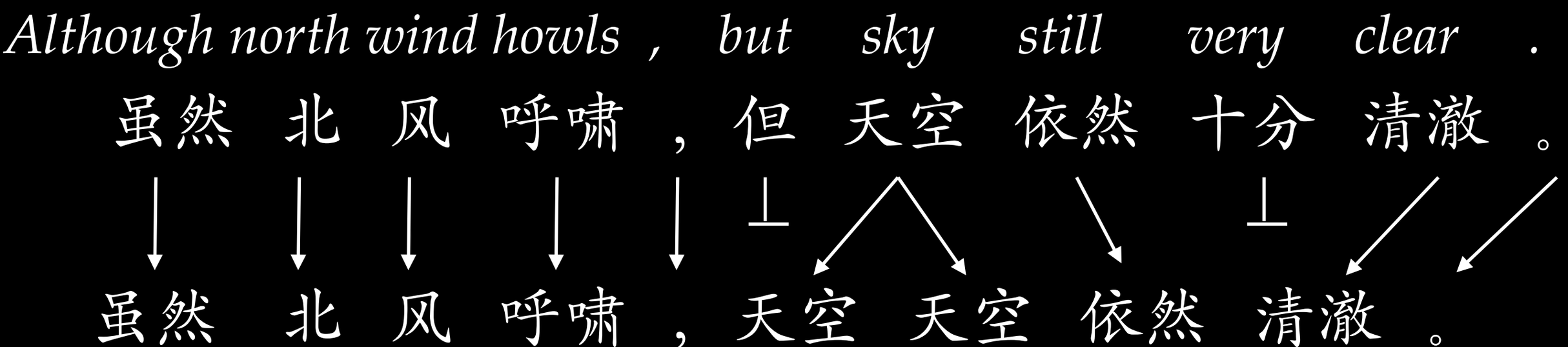
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

↓

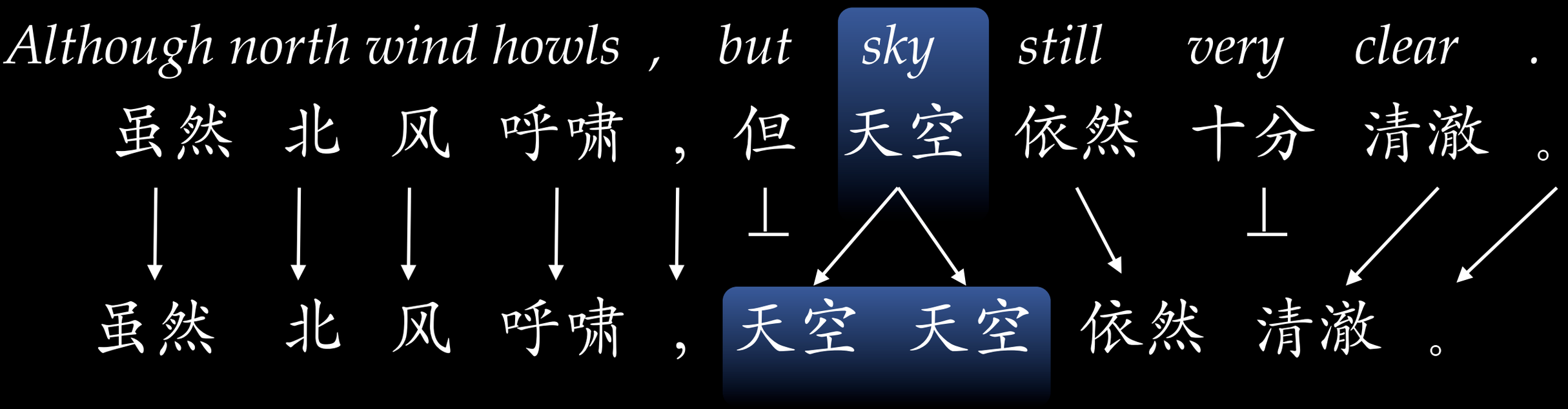
虽然

$$p_f(1|\text{虽然})$$

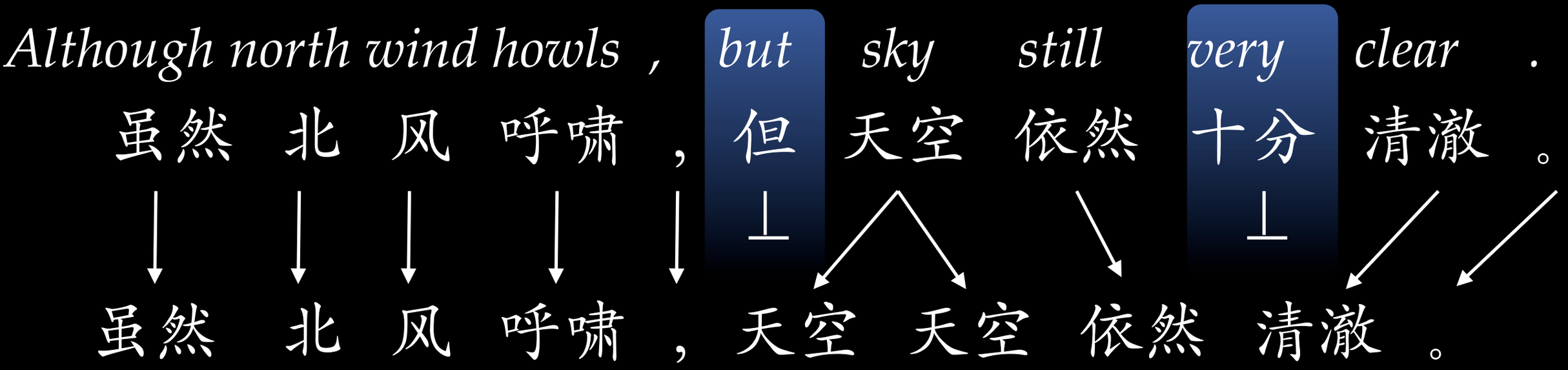
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。

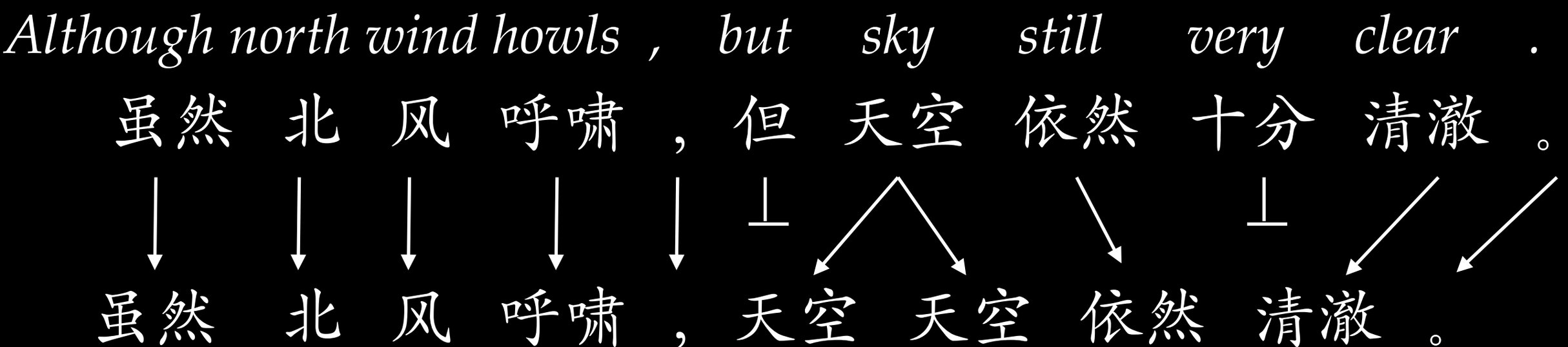
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。

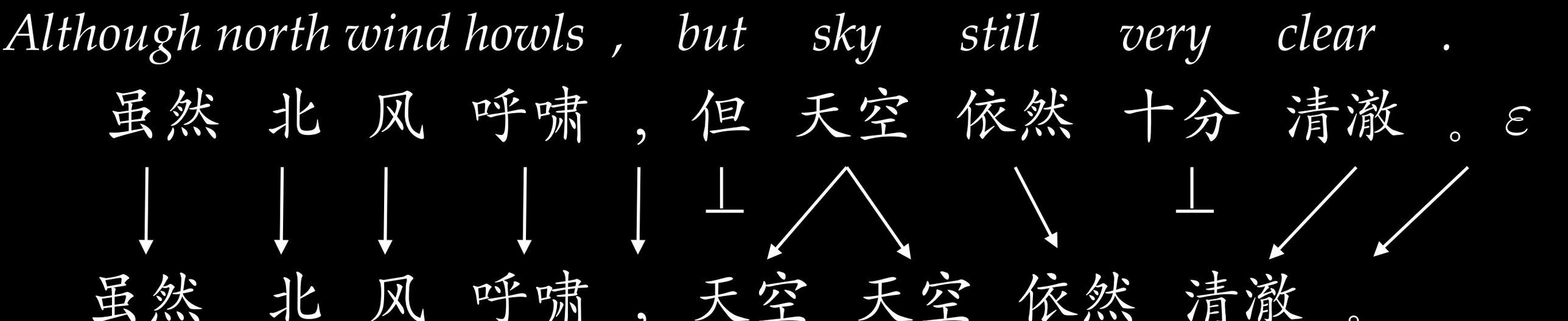
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。

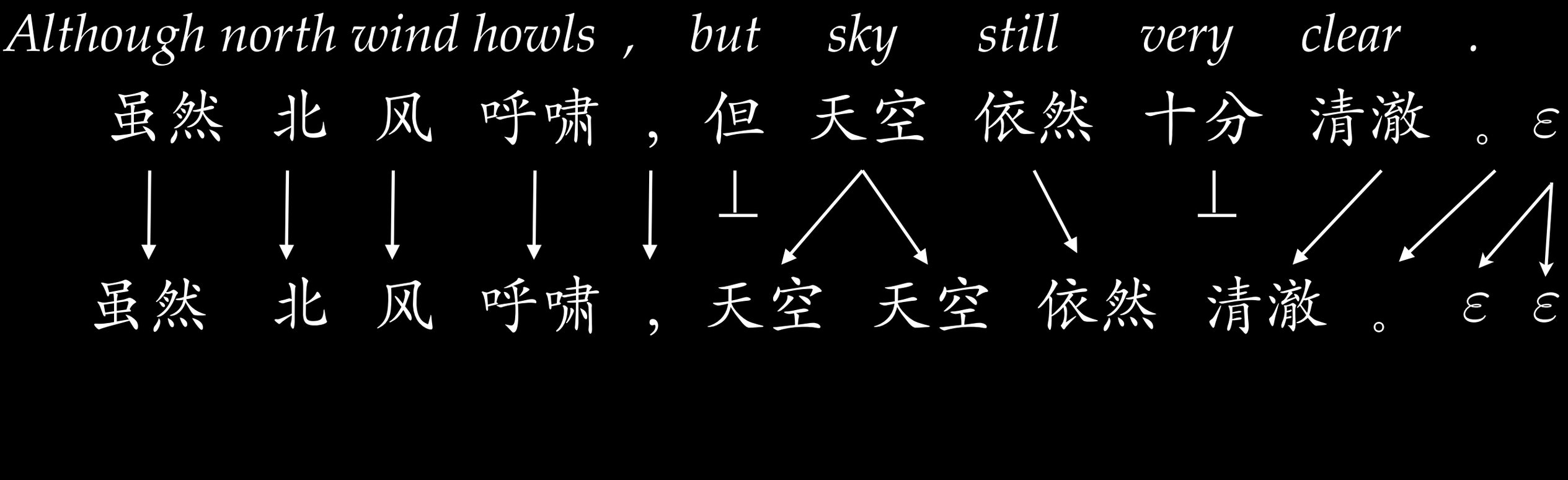
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。
⊥
⊥
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。

IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε
⊥
⊥
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。

IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。 ε
⊥ ⊥
虽然 北 风 呼啸 ， 天空 天空 依然 清澈 。 ε ε

IBM Model 4

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

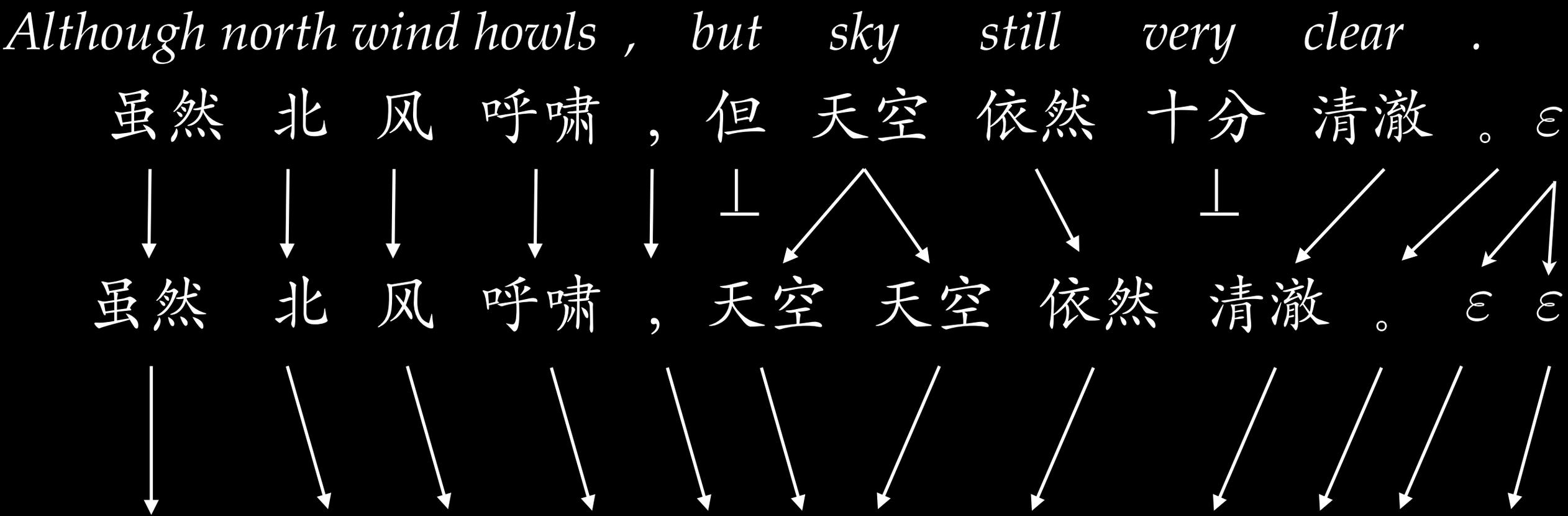
虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

However

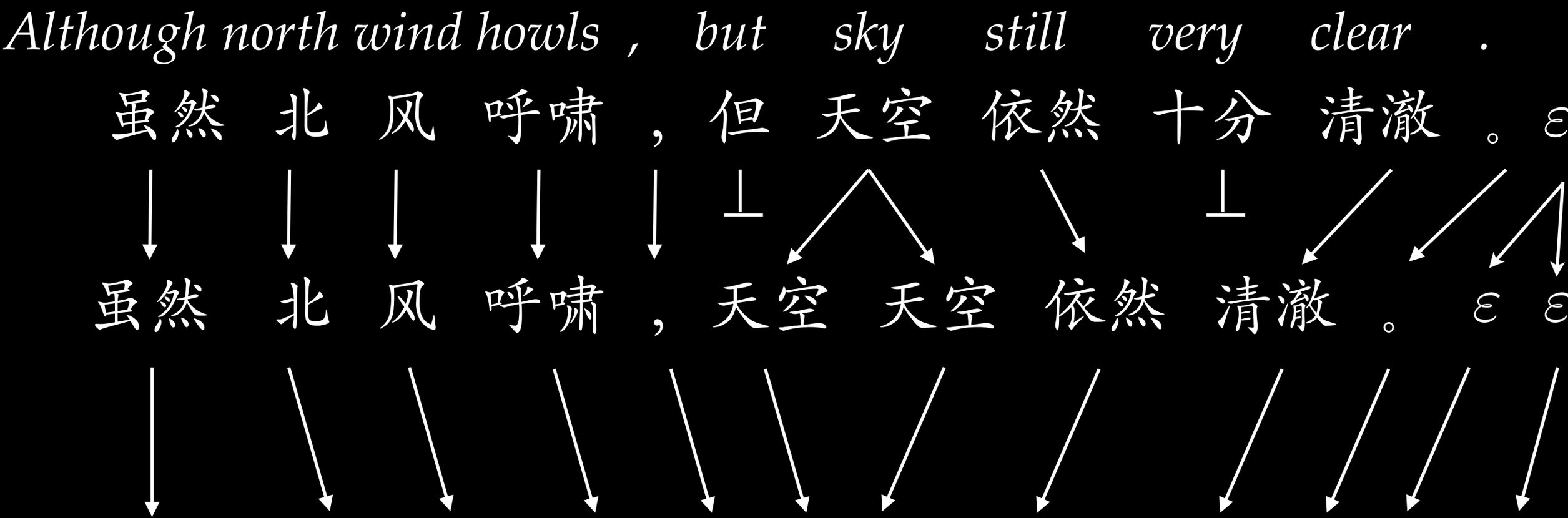
IBM Model 4

$$p_t(However|\text{虽然})$$

IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。 ε
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

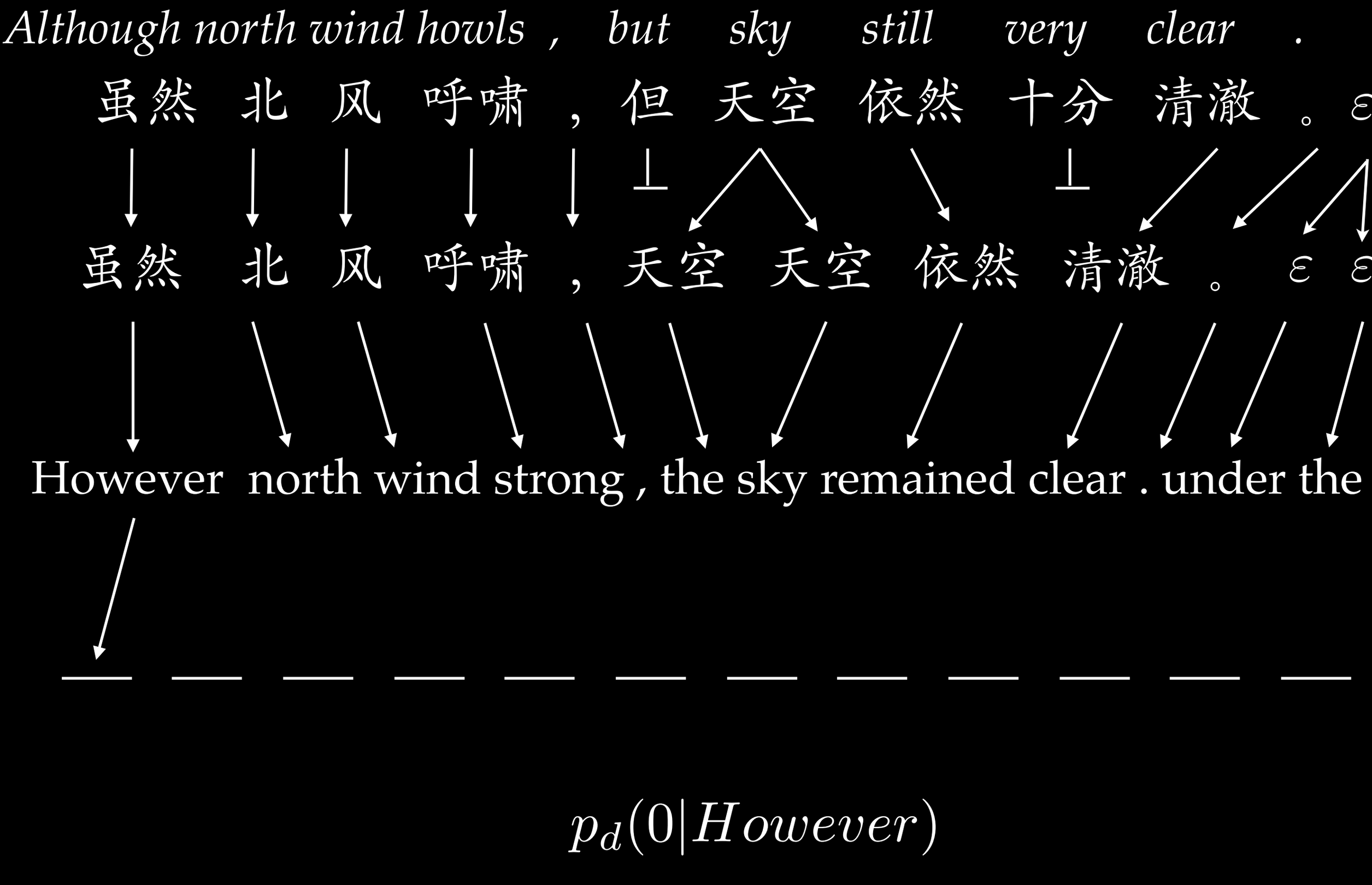
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。 ε
⊥ ⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

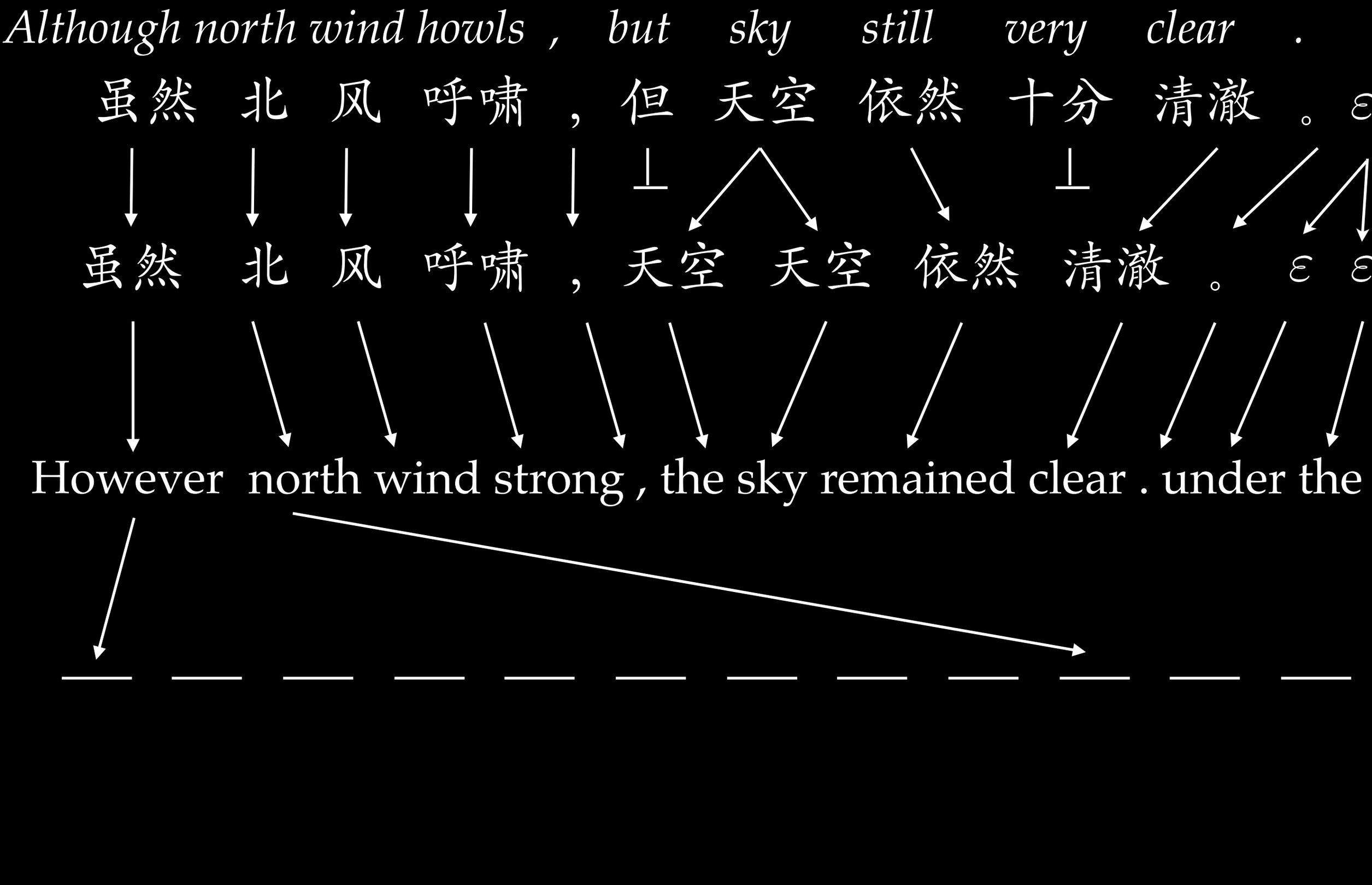
IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。 ε
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。 ε
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the
$p_d(0|However)$

IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。 ε
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the

IBM Model 4

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

However north wind strong , the sky remained clear . under the

$p_d(8|north)$

IBM Model 4

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

However north wind strong , the sky remained clear . under the

IBM Model 4


Although north wind howls , but sky still very clear .
虽然 北 风 呼啸 , 但 天空 依然 十分 清澈 。 ε
⊥
⊥
虽然 北 风 呼啸 , 天空 天空 依然 清澈 。 ε ε
However north wind strong , the sky remained clear . under the
However , the sky remained clear under the strong north wind .

IBM Model 4

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。ε

虽然 北 风 呼啸 ，天空 天空 依然 清澈 。 ε ε

However north wind strong , the sky remained clear . under the

However , the sky remained clear under the strong north wind .

$$p(English, alignment|Chinese) = \prod_{p_f} \prod_{p_t} \prod_{p_d}$$

IBM Model 4

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

$$p(English, alignment|Chinese) = \prod_{p_f} \prod_{p_t} \prod_{p_d}$$

IBM Model 4

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。

However , the sky remained clear under the strong north wind .

$$p(English|Chinese) = \sum_{alignments} \prod_{p_f} \prod_{p_t} \prod_{p_d}$$

IBM Model 4

$$\Pr(\tau, \pi|\mathbf{e}) = \prod_{i=1}^{l} \Pr(\phi_i|\phi_1^{i-1}, \mathbf{e}) \Pr(\phi_0|\phi_1^{l}, \mathbf{e}) \times$$

$$\prod_{i=0}^{l} \prod_{k=1}^{\phi_i} \Pr(\tau_{ik}|\tau_{i1}^{k-1}, \tau_0^{i-1}, \phi_0^{l}, \mathbf{e}) \times$$

$$\prod_{i=1}^{l} \prod_{k=1}^{\phi_i} \Pr(\pi_{ik}|\pi_{i1}^{k-1}, \pi_1^{i-1}, \tau_0^{l}, \phi_0^{l}, \mathbf{e}) \times$$

$$\prod_{k=1}^{\phi_0} \Pr(\pi_{0k}|\pi_{01}^{k-1}, \pi_1^{l}, \tau_0^{l}, \phi_0^{l}, \mathbf{e})$$

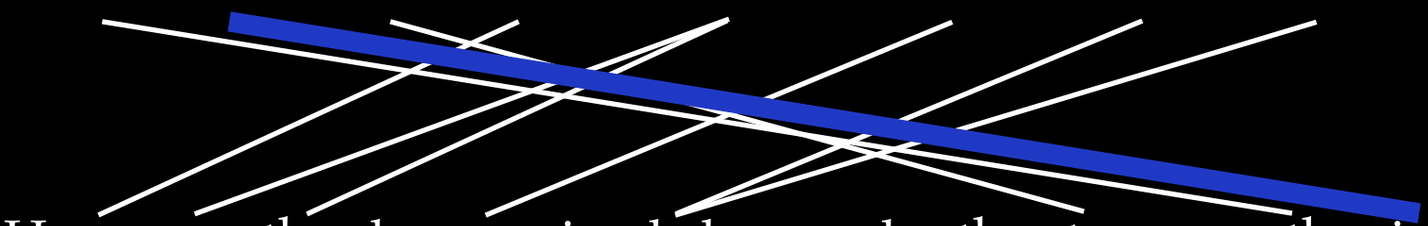
EM for IBM Model 4

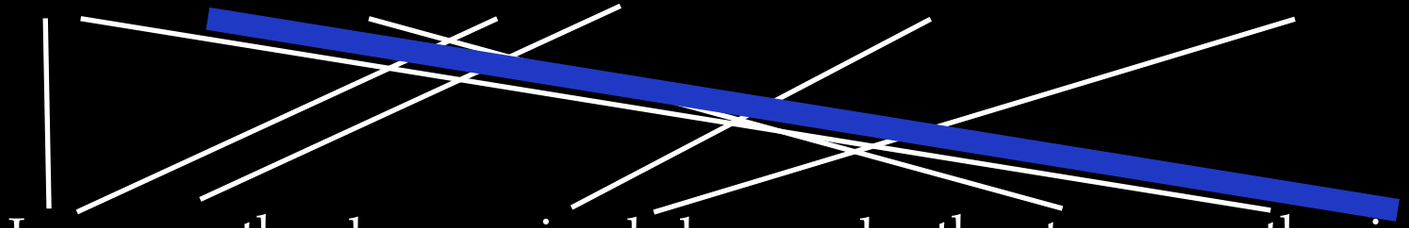
EM for IBM Model 4

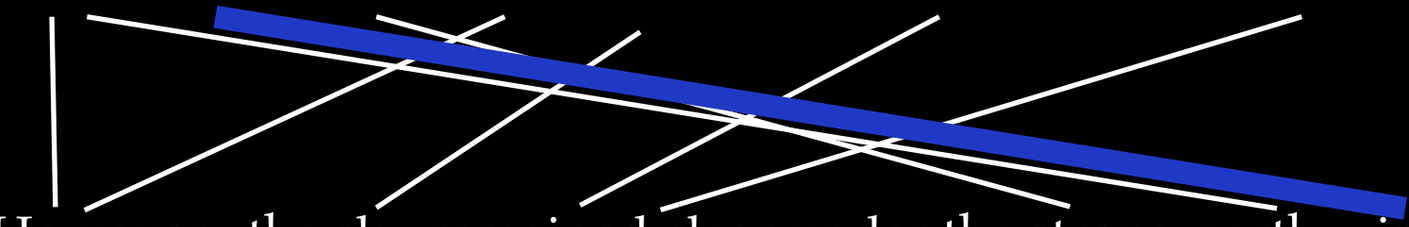
EM for IBM Model 4

Word-for-word independence assumptions do not hold!

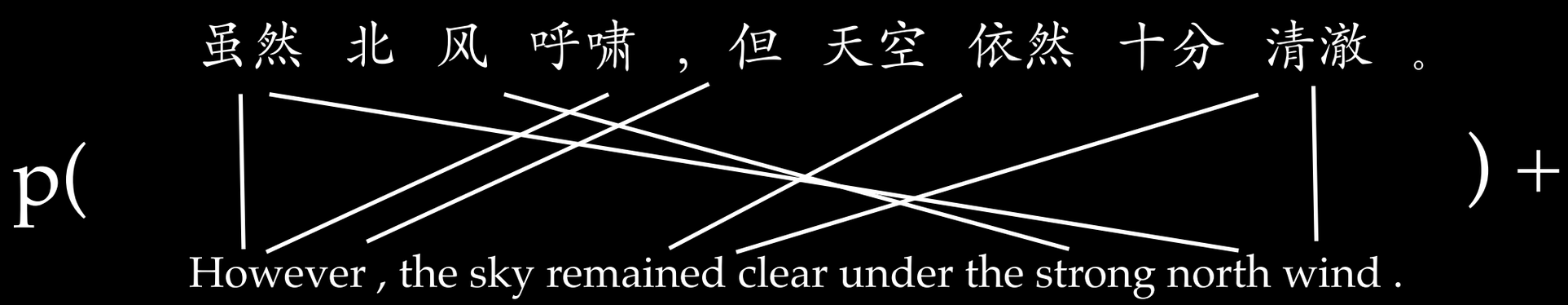
Marginalize: sum all alignments containing the link

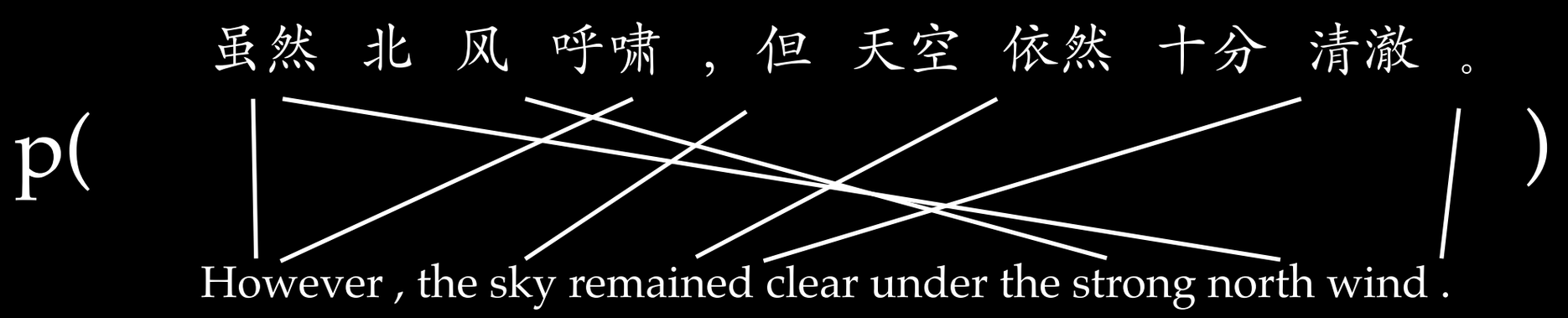
p(
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .) +

p(
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .) +

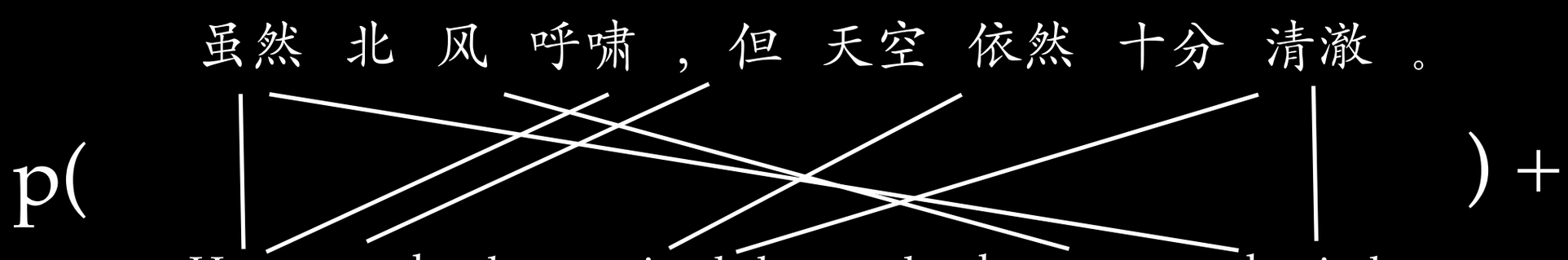
p(
虽然 北 风 呼啸 ， 但 天空 依然 十分 清澈 。
However , the sky remained clear under the strong north wind .)

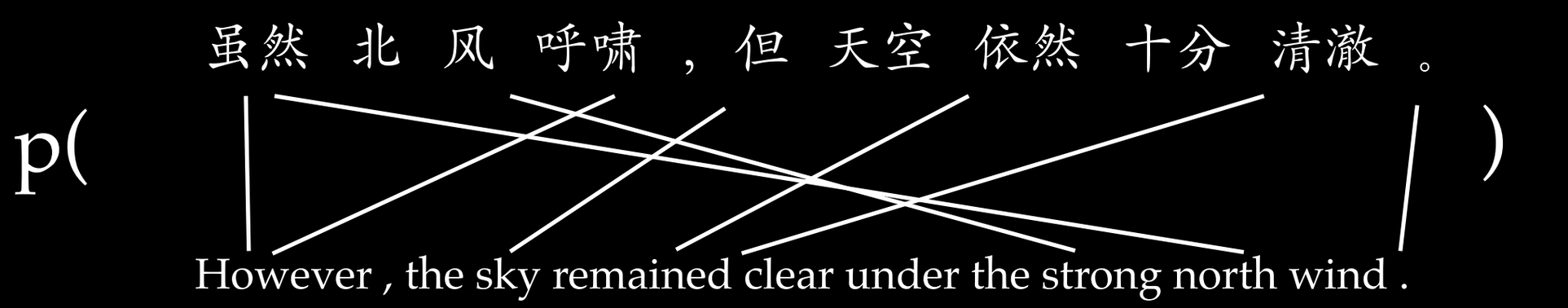
Divide by sum of all *possible* alignments





Divide by sum of all *possible* alignments





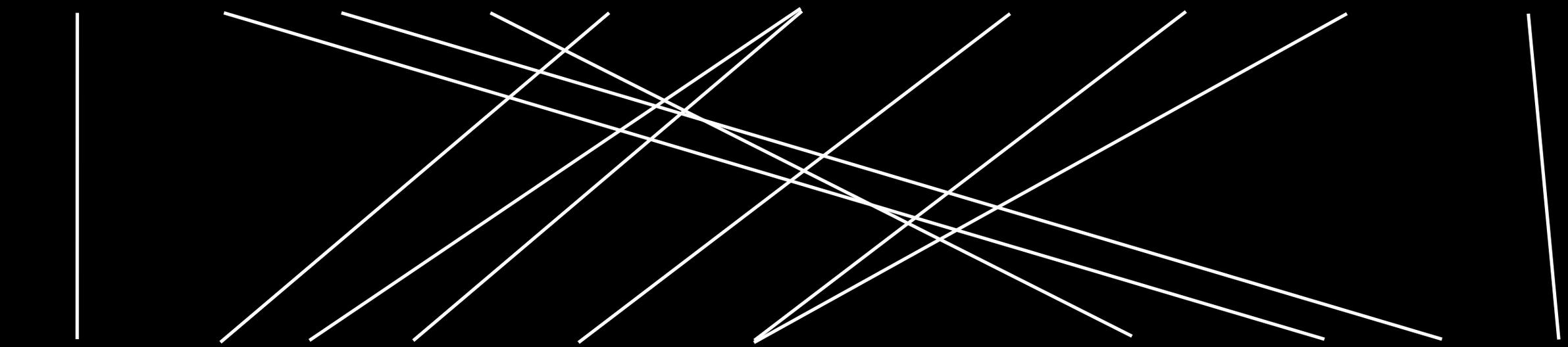
We have to sum over exponentially many alignments!

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



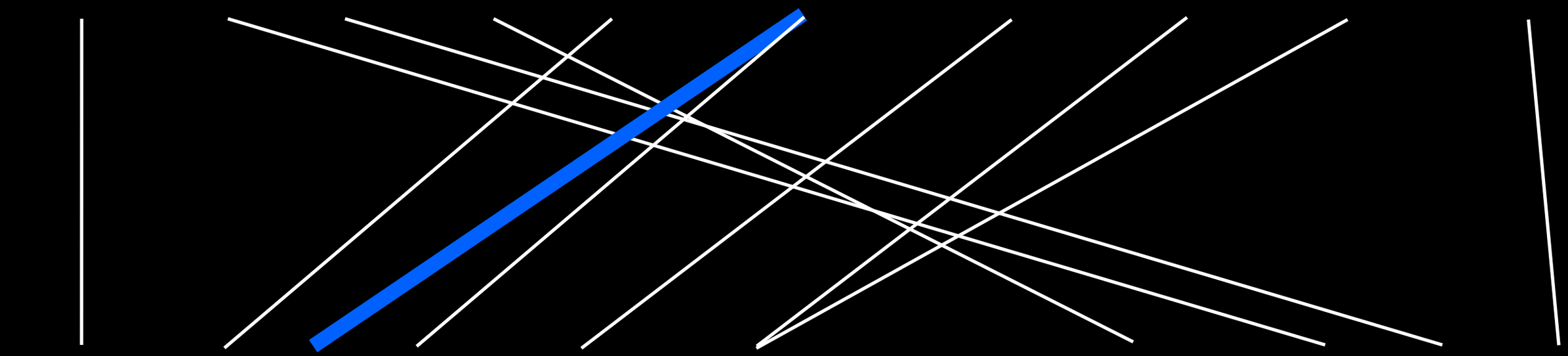
However , the sky remained clear under the strong north wind .

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



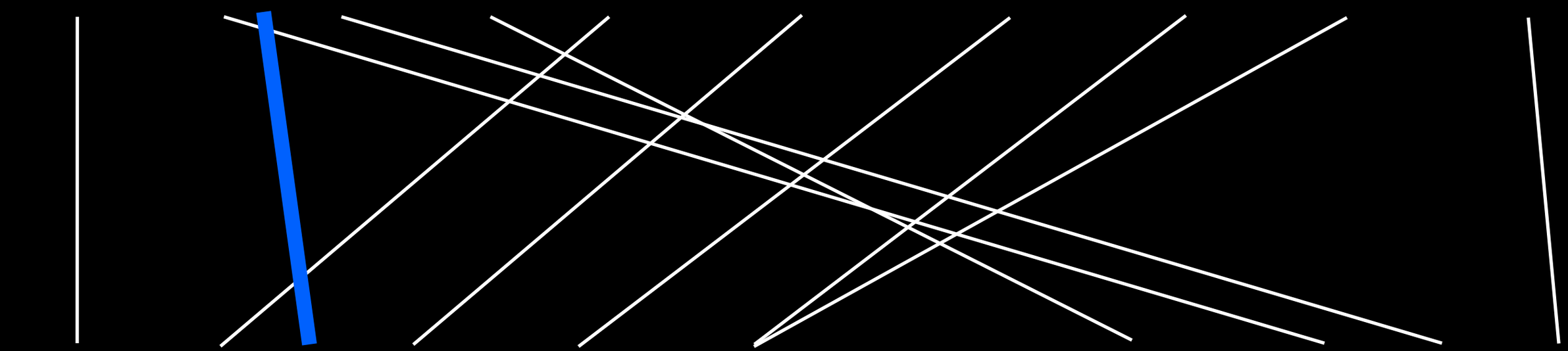
However , the sky remained clear under the strong north wind .

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



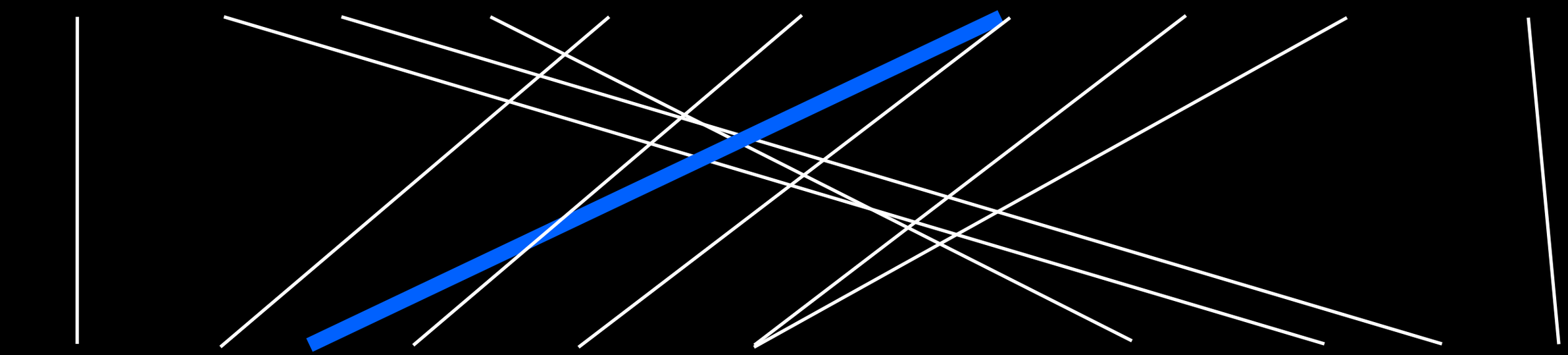
However , the sky remained clear under the strong north wind .

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



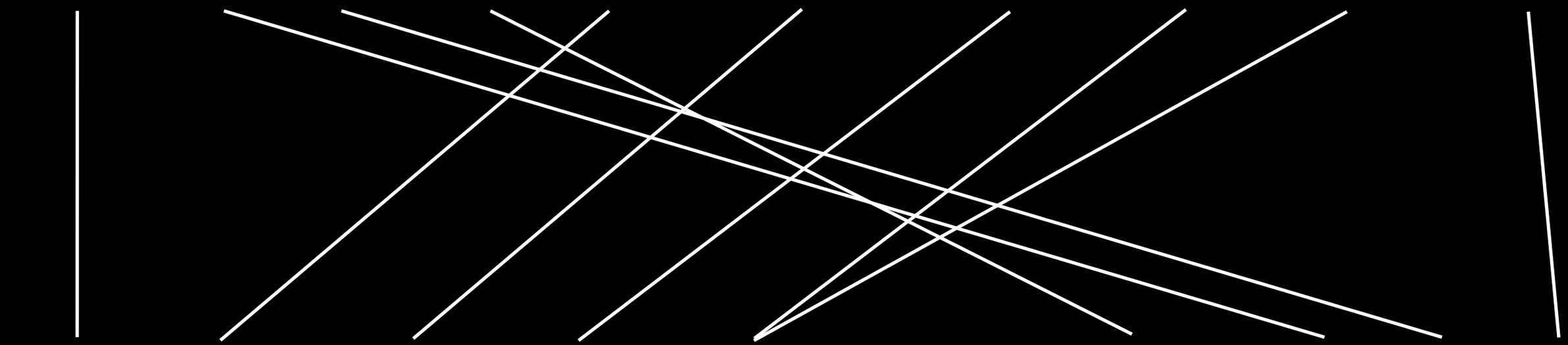
However , the sky remained clear under the strong north wind .

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



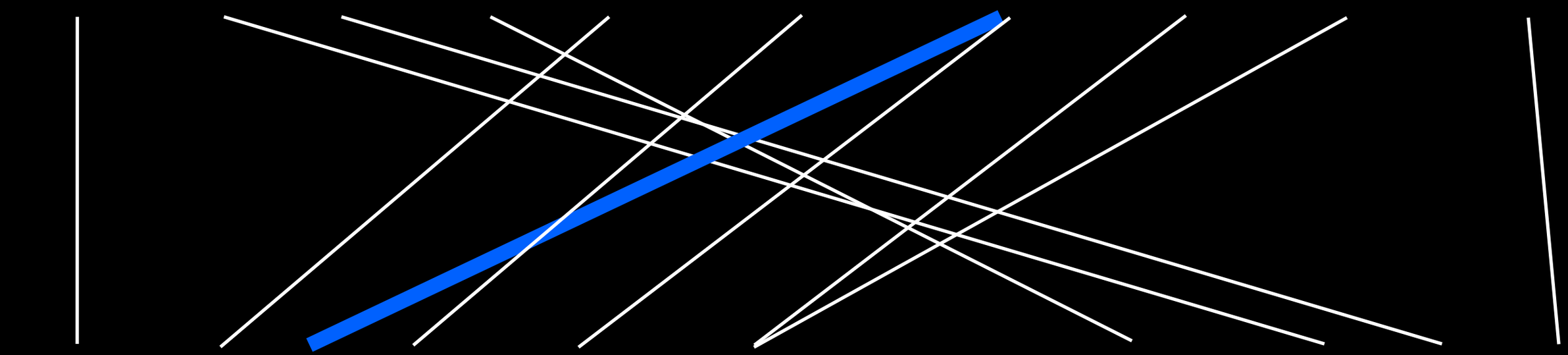
However , the sky remained clear under the strong north wind .

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



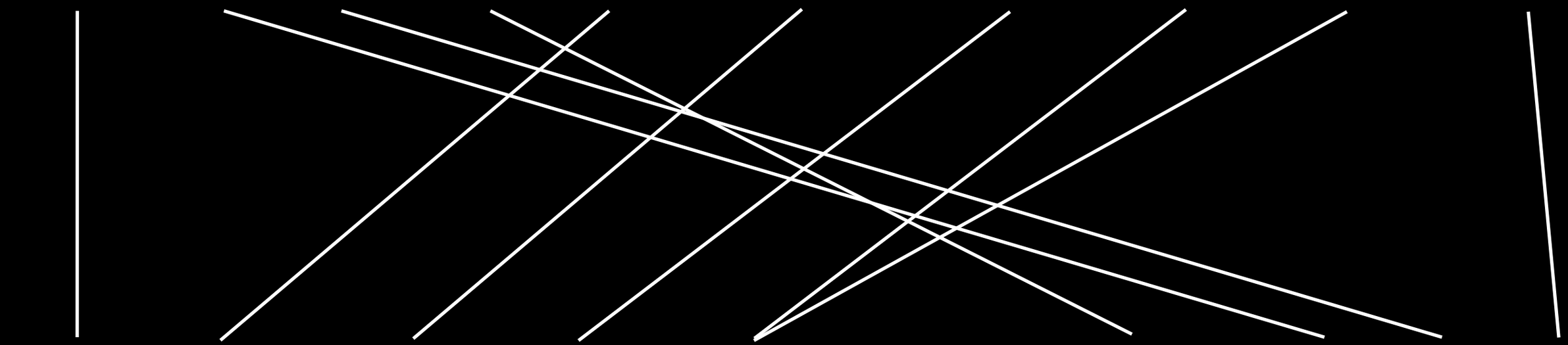
However , the sky remained clear under the strong north wind .

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



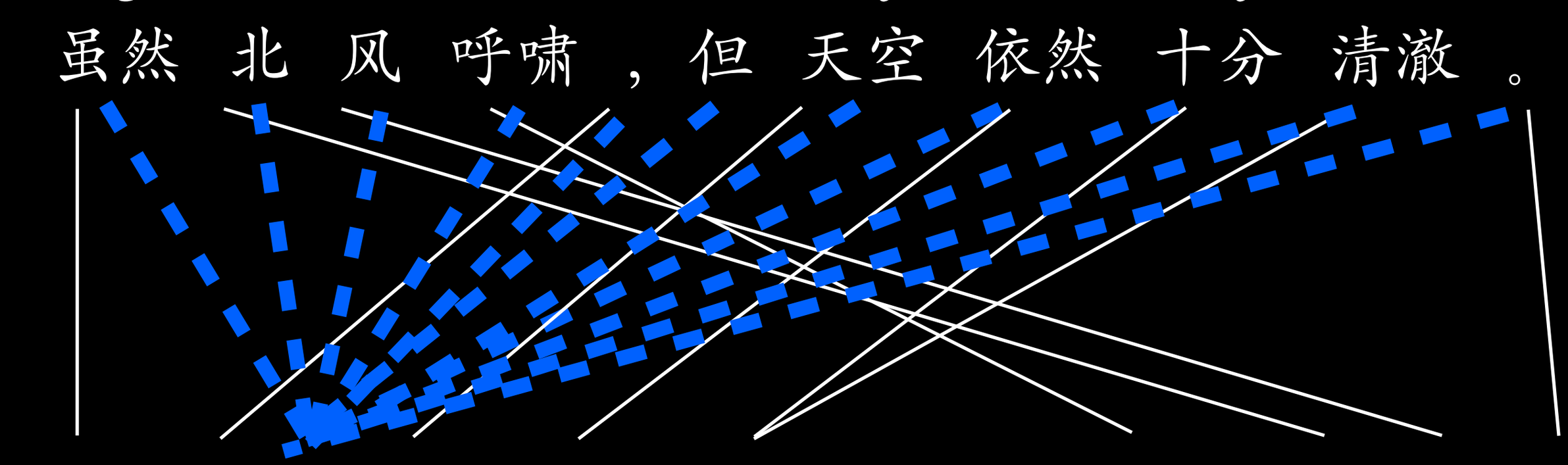
However , the sky remained clear under the strong north wind .

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

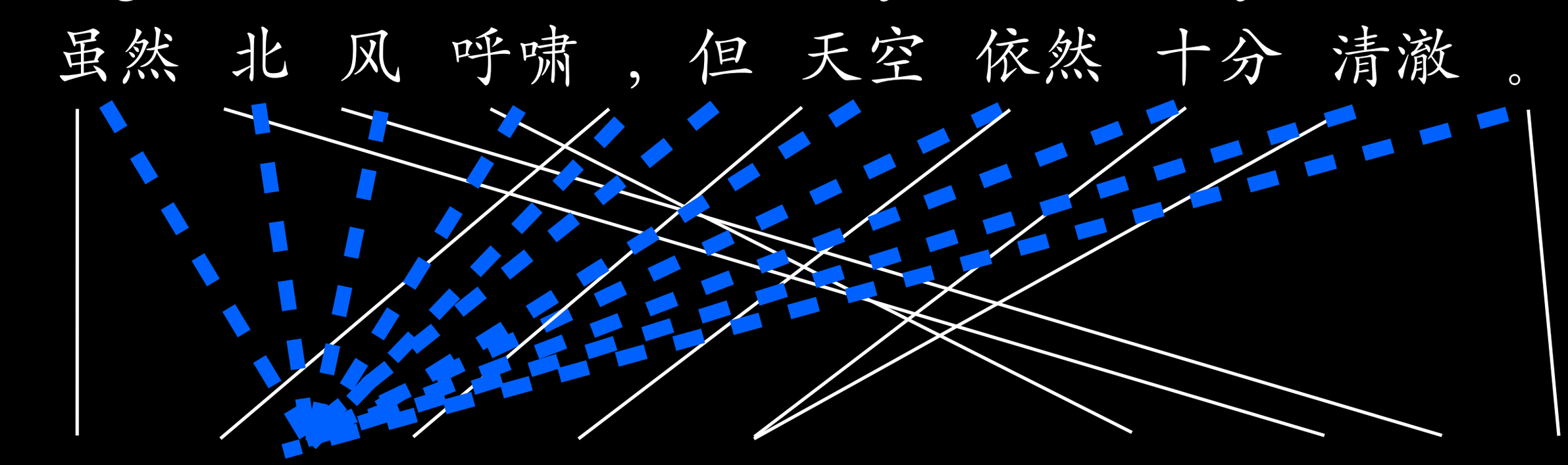
choose probabilistically among all possible alignments.

Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

IBM (1993): choose best among all possible alignments.

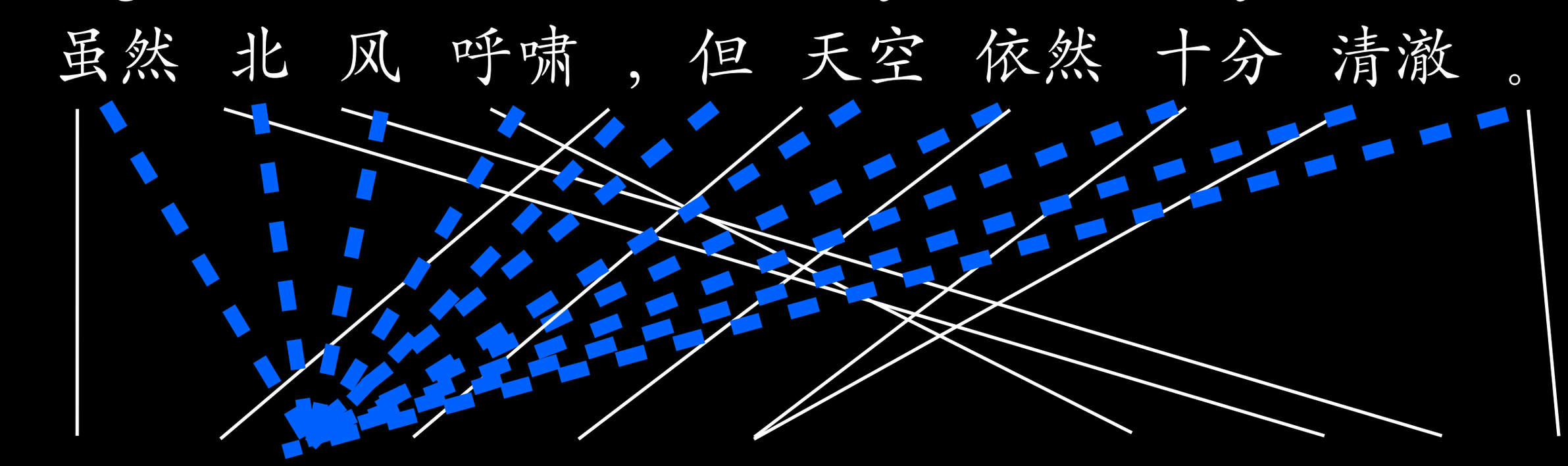
Monte Carlo EM for IBM Model 4

Idea: approximate sums with a representative sample

(similar to strategies for modern Bayesian inference)

Although north wind howls , but sky still very clear .

虽然 北 风 呼啸 ，但 天空 依然 十分 清澈 。



However , the sky remained clear under the strong north wind .

choose probabilistically among all possible alignments.

Tradeoffs: Modeling v. Learning

Lexical Translation

Local ordering dependency

Fertility

Convex

Tractable Exact Inference

Tradeoffs: Modeling v. Learning

Lexical Translation	Local ordering dependency	Fertility	Convex	Tractable Exact Inference
IBM Model 1				
HMM				
IBM Model 4				

Tradeoffs: Modeling v. Learning

	Lexical Translation	Local ordering dependency	Fertility	Convex	Tractable Exact Inference
IBM Model 1	✔				
HMM	✔	✔			
IBM Model 4	✔		✔		

Tradeoffs: Modeling v. Learning

	Lexical Translation	Local ordering dependency	Fertility	Convex	Tractable Exact Inference
IBM Model 1	✔	✘			
HMM	✔	✔			
IBM Model 4	✔	✔			

Tradeoffs: Modeling v. Learning

	Lexical Translation	Local ordering dependency	Fertility	Convex	Tractable Exact Inference
IBM Model 1	✔	✘	✘		
HMM	✔	✔	✘		
IBM Model 4	✔	✔	✔		

Tradeoffs: Modeling v. Learning

	Lexical Translation	Local ordering dependency	Fertility	Convex	Tractable Exact Inference
IBM Model 1	✔	✘	✘	✔	
HMM	✔	✔	✘	✘	
IBM Model 4	✔	✔	✔	✘	

Tradeoffs: Modeling v. Learning

	Lexical Translation	Local ordering dependency	Fertility	Convex	Tractable Exact Inference
IBM Model 1	✔	✘	✘	✔	✔
HMM	✔	✔	✘	✘	✔
IBM Model 4	✔	✔	✔	✘	✘

Non-Learning Uses of Alignment

- Lexicography
- Cross-lingual information retrieval
- Computer-aided translation
- Comparative linguistics
- General natural language processing
  - Parsers, taggers, etc. can be projected across alignments

Alignment for Alignment's Sake

- We might compute best alignment (or posteriors) and treat as observed for future application.
- Suggests a natural (and common) strategy for breaking asymmetry.
  - Learn English → French model.
  - Learn French → English model.
  - Combine their predictions in some way.
- Further along these lines: alignment by agreement.

IBM Model 4

What are some things this model doesn't account for?

What are some things this model doesn't account for?

Other IBM Models?

- Model 2: chooses alignment based on absolute word position.
- Model 3: fertility, but no Markov depedency.
- Model 5: non-deficient estimation.
- Original purpose: initialize Model N parameters from Model N-1 parameters.
- See also: Och & Ney, 2003, *A Systematic Comparison of Various Statistical Alignment Models*

Historical Note

- Why ***IBM Models***?

Historical Note

- Why ***IBM Models***?

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center

We describe a series of five statistical models of the translation process and give algorithms for estimating the parameters of these models given a set of pairs of sentences that are translations of one another. We define a concept of word-by-word alignment between such pairs of sentences. For any given pair of such sentences each of our models assigns a probability to each of the possible word-by-word alignments. We give an algorithm for seeking the most probable of these alignments. Although the algorithm is suboptimal, the alignment thus obtained accounts well for the word-by-word relationships in the pair of sentences. We have a great deal of data in French and English from the proceedings of the Canadian Parliament. Accordingly, we have restricted our work to these two languages; but we feel that because our algorithms have minimal linguistic content they would work well on other pairs of languages. We also feel, again because of the minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

Historical Note

- Why *IBM Models*?

Fred Jelinek
(1932-2010)

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center

We describe a series of five statistical models of the translation process and give algorithms for estimating the parameters of these models given a set of pairs of sentences that are translations of one another. We define a concept of word-by-word alignment between such pairs of sentences. For any given pair of such sentences each of our models assigns a probability to each of the possible word-by-word alignments. We give an algorithm for seeking the most probable of these alignments. Although the algorithm is suboptimal, the alignment thus obtained accounts well for the word-by-word relationships in the pair of sentences. We have a great deal of data in French and English from the proceedings of the Canadian Parliament. Accordingly, we have restricted our work to these two languages; but we feel that because our algorithms have minimal linguistic content they would work well on other pairs of languages. We also feel, again because of the minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

Historical Note

- Why ***IBM Models***?

Fred Jelinek
(1932-2010)

Some of us started to wonder in the mid 1980s whether our [speech recognition] methods could be successfully applied to new fields. Bob Mercer and I spent many of our after-lunch "periphery" walks discussing possible candidates. We soon came up with two: machine translation and stock market modeling

minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

Historical Note

- Why *IBM Models*?

Fred Jelinek
(1932-2010)

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center

We describe a series of five statistical models of the translation process and give algorithms for estimating the parameters of these models given a set of pairs of sentences that are translations of one another. We define a concept of word-by-word alignment between such pairs of sentences. For any given pair of such sentences each of our models assigns a probability to each of the possible word-by-word alignments. We give an algorithm for seeking the most probable of these alignments. Although the algorithm is suboptimal, the alignment thus obtained accounts well for the word-by-word relationships in the pair of sentences. We have a great deal of data in French and English from the proceedings of the Canadian Parliament. Accordingly, we have restricted our work to these two languages; but we feel that because our algorithms have minimal linguistic content they would work well on other pairs of languages. We also feel, again because of the minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

Historical Note

- Why ***IBM Models***?

Fred Jelinek
(1932-2010)

"The validity of a statistical (information theoretic) approach to MT has indeed been recognized, as the authors mention, by Weaver as early as 1949. And was universally recognized as mistaken by 1950 (cf. Hutchins, MT – Past, Present, Future, Ellis Horwood, 1986, p. 30ff and references therein). The crude force of computers is not science. The paper is simply beyond the scope of COLING."

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

Historical Note

- Why *IBM Models*?

Fred Jelinek
(1932-2010)

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center

We describe a series of five statistical models of the translation process and give algorithms for estimating the parameters of these models given a set of pairs of sentences that are translations of one another. We define a concept of word-by-word alignment between such pairs of sentences. For any given pair of such sentences each of our models assigns a probability to each of the possible word-by-word alignments. We give an algorithm for seeking the most probable of these alignments. Although the algorithm is suboptimal, the alignment thus obtained accounts well for the word-by-word relationships in the pair of sentences. We have a great deal of data in French and English from the proceedings of the Canadian Parliament. Accordingly, we have restricted our work to these two languages; but we feel that because our algorithms have minimal linguistic content they would work well on other pairs of languages. We also feel, again because of the minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

Historical Note

- Why *IBM Models*?

Renaissance

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center

We describe a series of five statistical models of the translation process and give algorithms for estimating the parameters of these models given a set of pairs of sentences that are translations of one another. We define a concept of word-by-word alignment between such pairs of sentences. For any given pair of such sentences each of our models assigns a probability to each of the possible word-by-word alignments. We give an algorithm for seeking the most probable of these alignments. Although the algorithm is suboptimal, the alignment thus obtained accounts well for the word-by-word relationships in the pair of sentences. We have a great deal of data in French and English from the proceedings of the Canadian Parliament. Accordingly, we have restricted our work to these two languages; but we feel that because our algorithms have minimal linguistic content they would work well on other pairs of languages. We also feel, again because of the minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

Fred Jelinek
(1932-2010)

Historical Note

- Why *IBM Models*?

Renaissance

The Mathematics of Statistical Machine Translation: Parameter Estimation

Peter F. Brown*
IBM T.J. Watson Research Center

Stephen A. Della Pietra*
IBM T.J. Watson Research Center

Vincent J. Della Pietra*
IBM T.J. Watson Research Center

Robert L. Mercer*
IBM T.J. Watson Research Center

We describe a series of five statistical models of the translation process and give algorithms for estimating the parameters of these models given a set of pairs of sentences that are translations of one another. We define a concept of word-by-word alignment between such pairs of sentences. For any given pair of such sentences each of our models assigns a probability to each of the possible word-by-word alignments. We give an algorithm for seeking the most probable of these alignments. Although the algorithm is suboptimal, the alignment thus obtained accounts well for the word-by-word relationships in the pair of sentences. We have a great deal of data in French and English from the proceedings of the Canadian Parliament. Accordingly, we have restricted our work to these two languages; but we feel that because our algorithms have minimal linguistic content they would work well on other pairs of languages. We also feel, again because of the minimal linguistic content of our algorithms, that it is reasonable to argue that word-by-word alignments are inherent in any sufficiently large bilingual corpus.

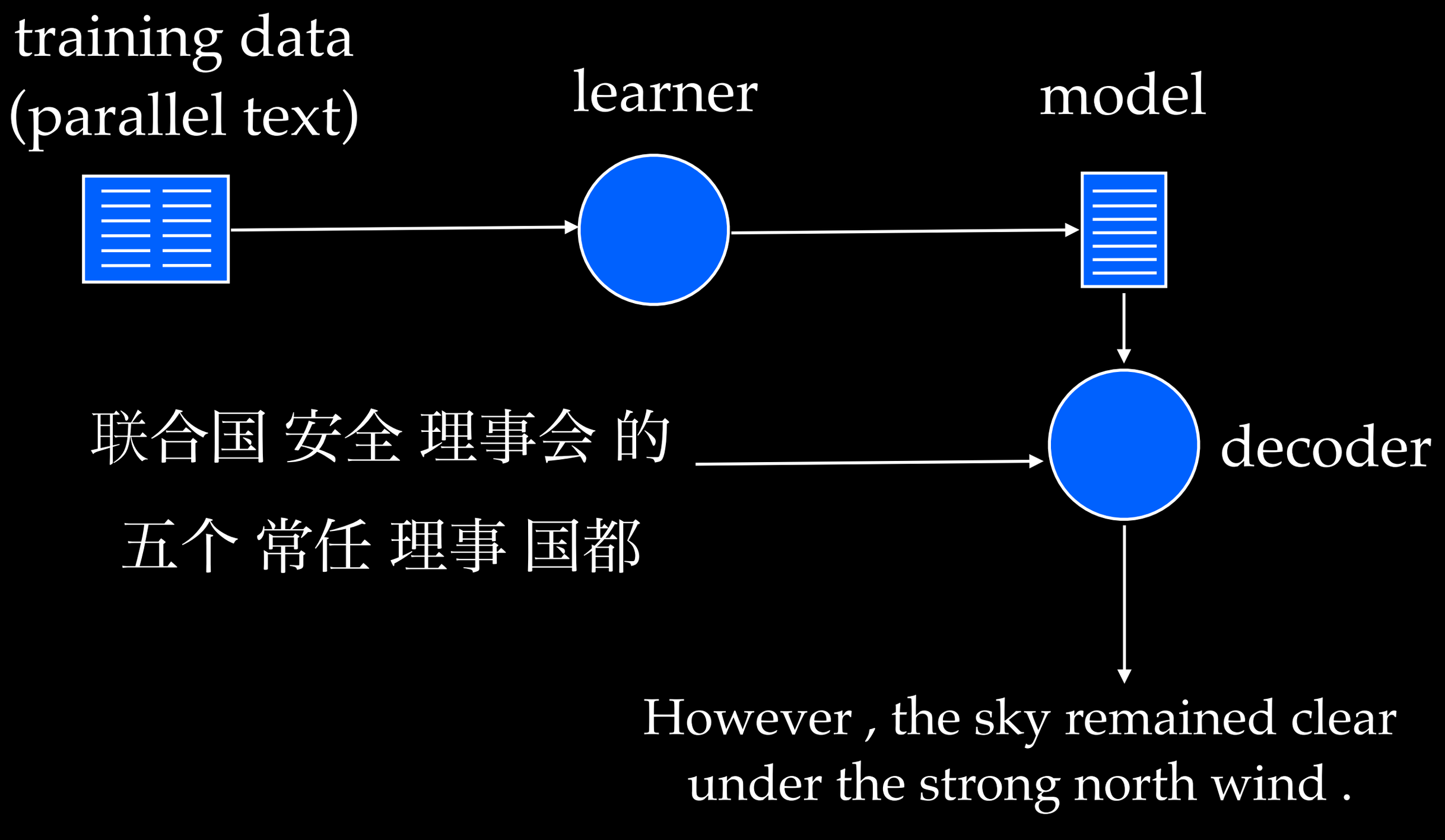
1. Introduction

The growing availability of bilingual, machine-readable texts has stimulated interest

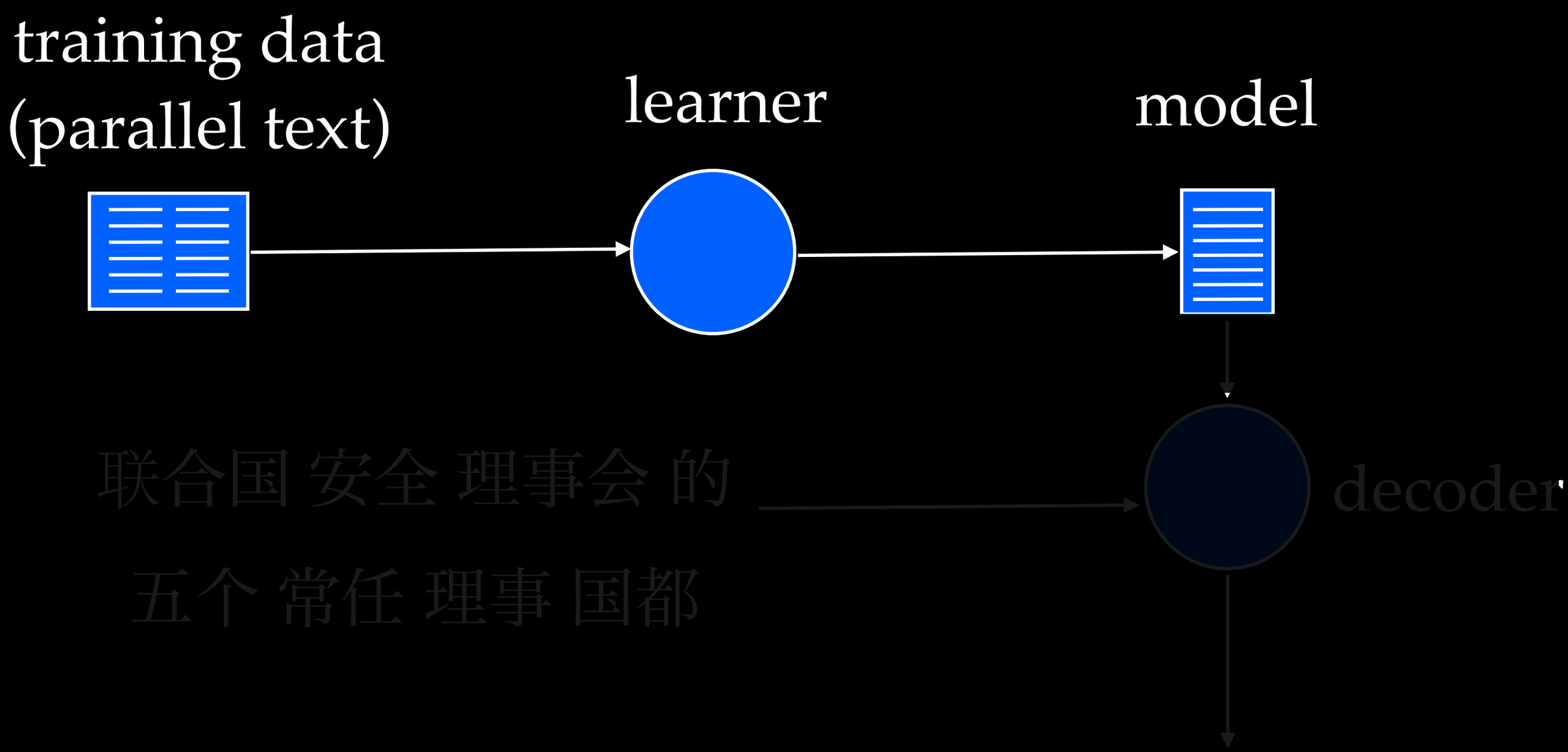
Fred Jelinek
(1932-2010)

The Center For Language and Speech Processing
at the Johns Hopkins University

Where to Next?



Where to Next?



Where to Next?