Topological Ordering of Function Words in Hierarchical Phrase-based Translation

Hendra Setiawan*†(presenter), Min-Yen Kan*,
Haizhou Li ‡, Philip Resnik†

National University of Singapore *, Institute for Infocomm Research ‡, University of Maryland†

Outline

- Problem Statement
- Related Work
- Pairwise Dominance Model
- Experiments
- Discussion

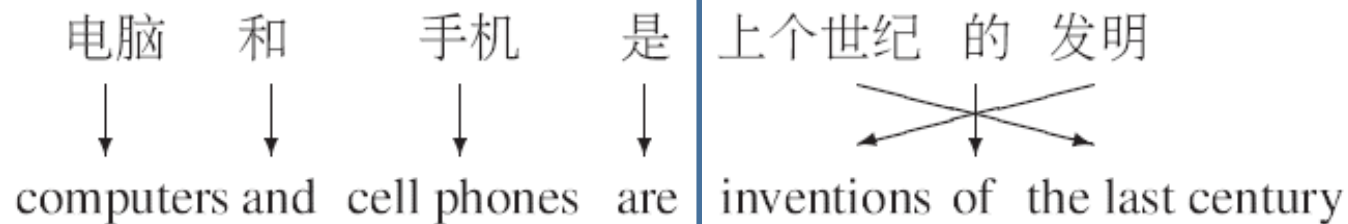# Problem Statement

Hierarchical Phrase-based Translation (Hiero) comes with **one type of non-terminal symbol**, e.g. $X \rightarrow \langle X\ 的\ 发明\ ,\ \text{inventions of } X \rangle$ thus permits *flexible* topological ordering

Problem Statement

We want to translate:

电脑　和　　手机　　是 | 上个世纪　的　发明

↓　　↓　　↓　　↓

computers and cell phones are | inventions of the last century

with：

r1: $X \rightarrow \langle$ 电脑 和 X_1, computers and $X_1 \rangle$

r2: $X \rightarrow \langle X_1$ 是 X_2, X_1 are $X_2 \rangle$

r3: $X \rightarrow \langle$ 手机 , cell phones $\rangle$

r4: $X \rightarrow \langle X_1$ 的 发明 , inventions of $X_1 \rangle$

r5: $X \rightarrow \langle$ 上个世纪 , the last century $\rangle$

Problem Statement

- There are two possible topological orderings:
  - r 1 ≺ r2 ≺ r3 ≺ (r4) ≺ r5   (correct)

    computers and cell phones are inventions of the last century
  - (r 4) ≺ r1 ≺ r2 ≺ r3 ≺ r5   (incorrect)

    inventions of computers and cell phones are the last century

- Standard treatment:

  n-gram (target) language model

Outline

- Problem Statement
- Related Work
- Pairwise Dominance Model
- Experiments
- Discussion

Related Work

Assume a parse tree: respect the boundary

[[[电脑] 和 [手机]] 是 [[上个世纪] 的 [发明]]]

- Constituent feature (Chiang 2005)

- Soft Syntactic Constraint (Marton and Resnik 2008)

- Maximum-Entropy based Soft Syntactic Constraint (Xiong et al. 2009)

Can we go about without parse tree?

Outline

- Problem Statement
- Related Work
- Pairwise Dominance Model
- Experiments
- Discussion

Pairwise Dominance Model

- In the paper:
  - r 1 ≺ r2 ≺ r3 ≺ r4 ≺ r5 ≈ 和/and ≺ 是/≺ 的/of ≺ ≺ ≺ ≺ ≺
  - r 4 r1 r2 r3 r5 ≈ 的 /of 和/and 是/are
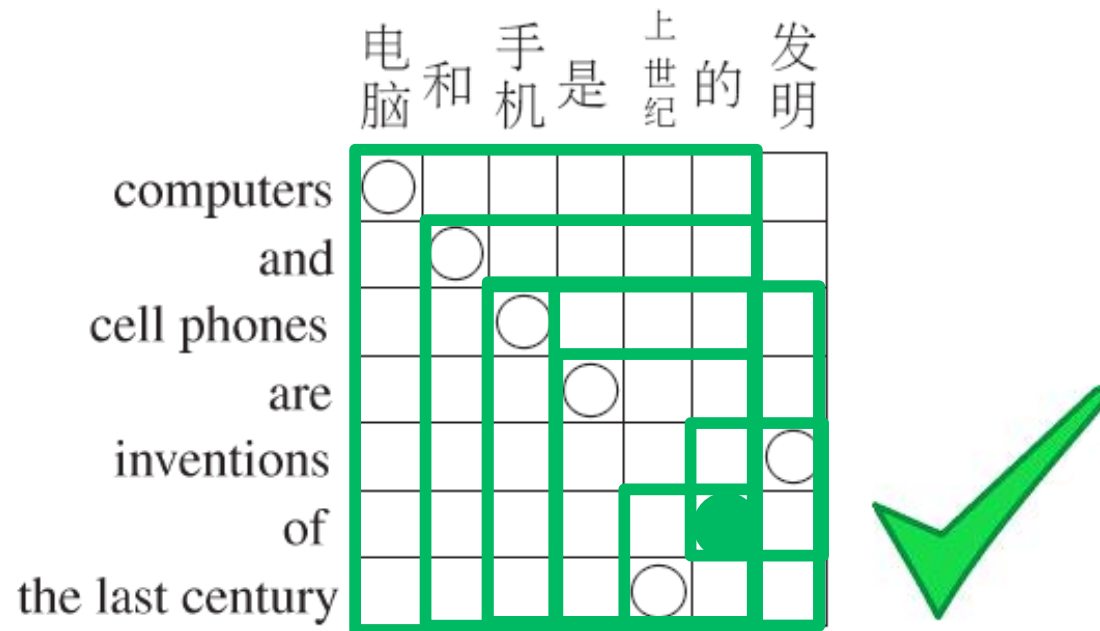
- In this presentation :

  Alignment information shall define

  [[[电脑] 和 [手机]] 是 [[上个世纪] 的 [发明]]]
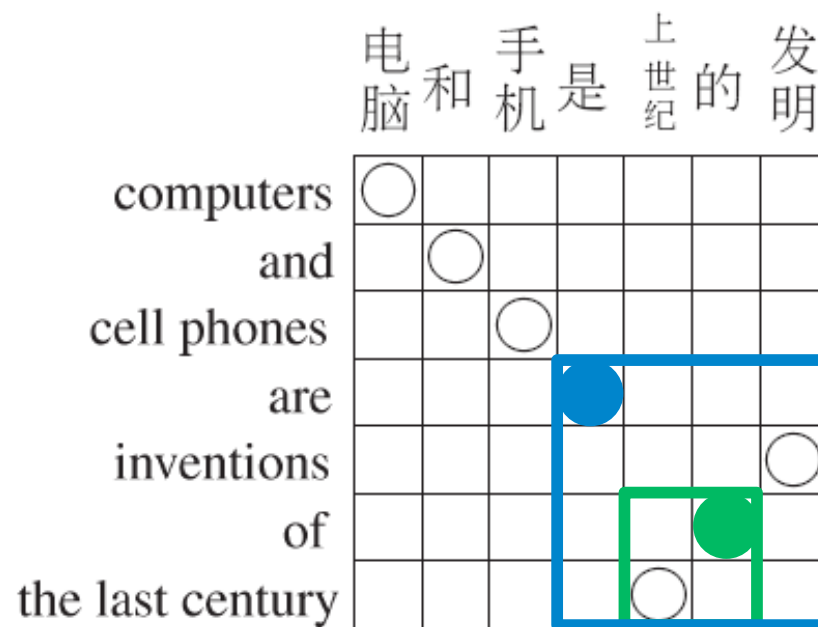
Estimating Span Boundary

Use *consistent alignment heuristic* (Och 2006)
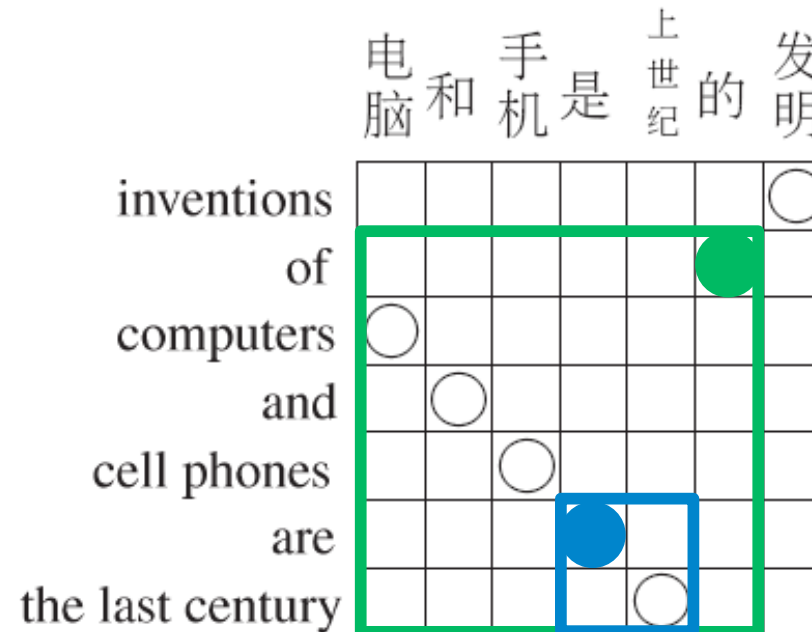
[[[电脑] 和 [手机]] 是 [[上个世纪] 的 [发明]]]

Dominance Relationship

- How to resolve topological ordering?
  Compare with the span of another anchor
  Larger span ≈ higher in topological ordering
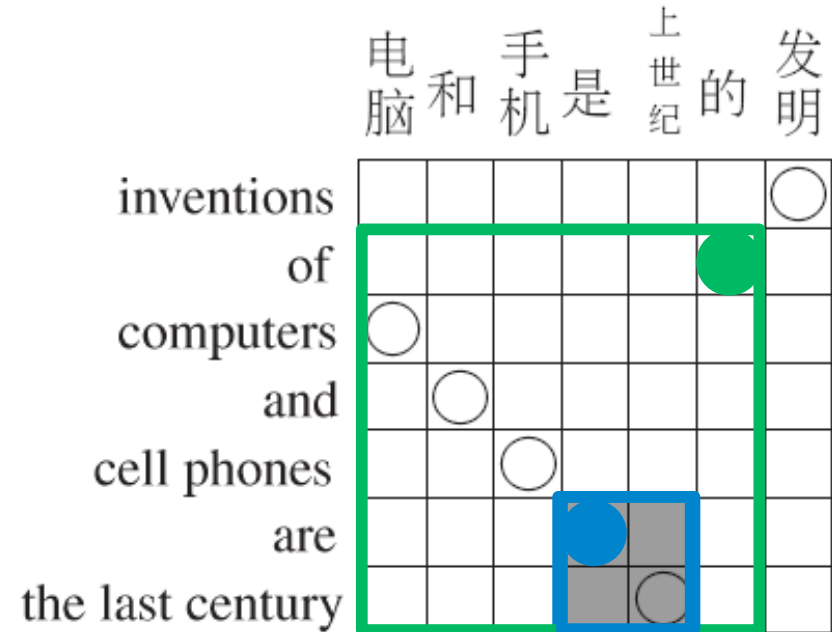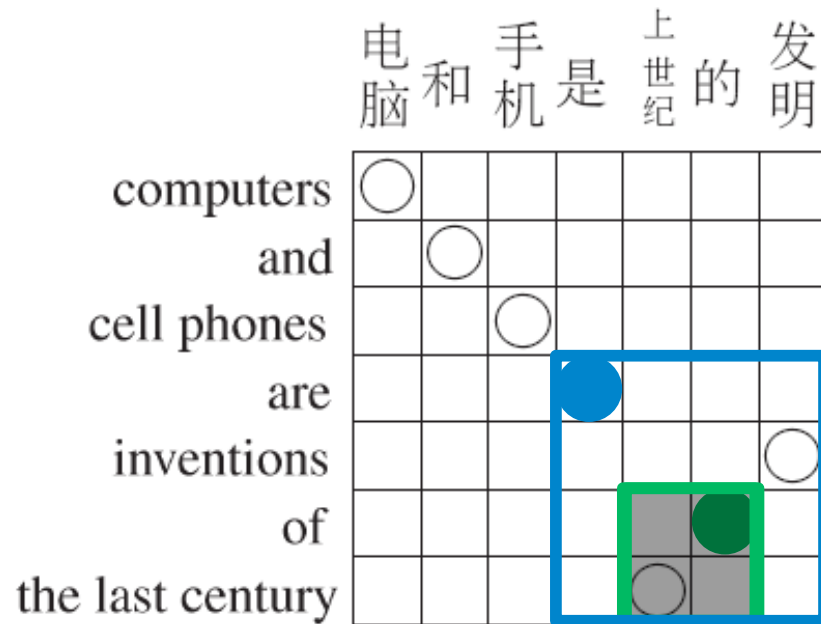
Dominance Relationship

- How about the incorrect topological ordering?

I can spot the difference !!!
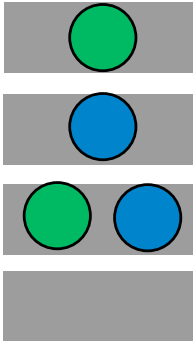
Dominance Relationship

- Correct versus Incorrect

The intersections are different!

Dominance Relationship

$$d(Y', Y'') = \begin{cases} \text{leftFirst,} & Y' \notin \mathrm{MCA}(Y'') \wedge Y'' \in \mathrm{MCA}(Y') \\ \text{rightFirst,} & Y' \in \mathrm{MCA}(Y'') \wedge Y'' \notin \mathrm{MCA}(Y') \\ \text{dontCare,} & Y' \in \mathrm{MCA}(Y'') \wedge Y'' \in \mathrm{MCA}(Y') \\ \text{neither,} & Y' \notin \mathrm{MCA}(Y'') \wedge Y'' \notin \mathrm{MCA}(Y') \end{cases}$$

Dominance Relationship

- Correct versus Incorrect

$$d(\textcolor{blue}{\bullet}, \textcolor{green}{\bullet}) = \text{leftFirst} \qquad d(\textcolor{blue}{\bullet}, \textcolor{green}{\bullet}) = \text{rightFirst}$$
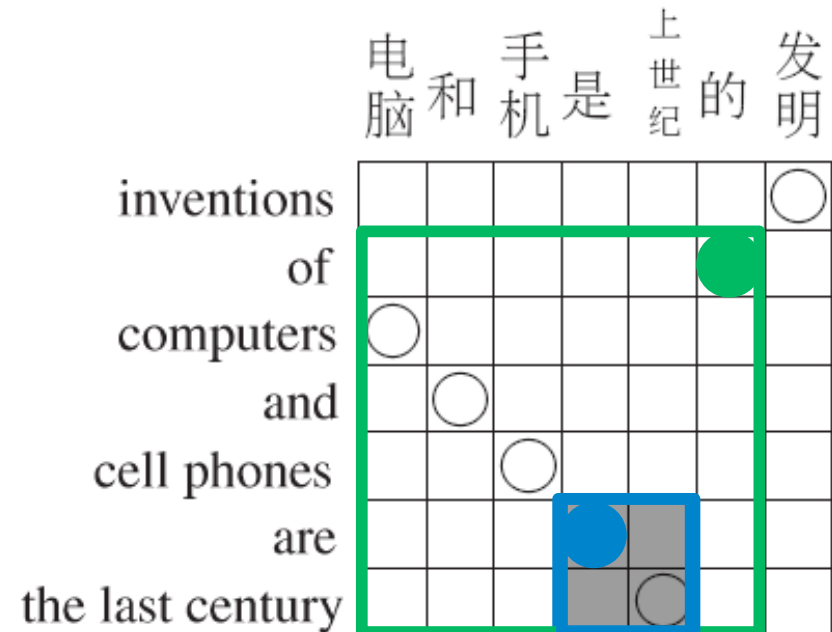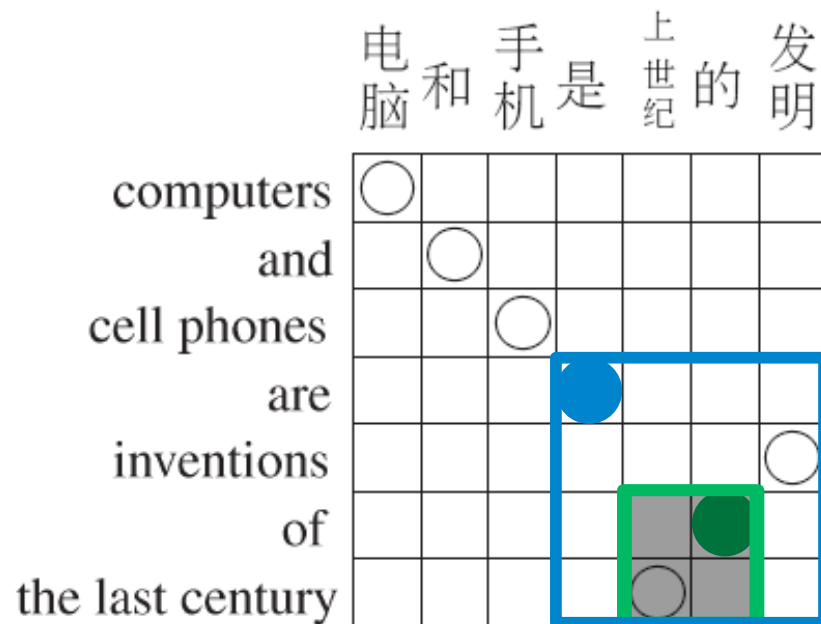
Dominance Relationship

$$d(Y', Y'') = \begin{cases} \text{leftFirst,} & Y' \notin \mathrm{MCA}(Y'') \wedge Y'' \in \mathrm{MCA}(Y') \\ \text{rightFirst,} & Y' \in \mathrm{MCA}(Y'') \wedge Y'' \notin \mathrm{MCA}(Y') \\ \text{dontCare,} & Y' \in \mathrm{MCA}(Y'') \wedge Y'' \in \mathrm{MCA}(Y') \\ \text{neither,} & Y' \notin \mathrm{MCA}(Y'') \wedge Y'' \notin \mathrm{MCA}(Y') \end{cases}$$

Pairwise dominance model: $\quad P(d(Y', Y'')|Y', Y'')$

What are Y' and Y''?

Neighboring function words

Outline

- Problem Statement
- Related Work
- Pairwise Dominance Model
- Experiments
- Discussion

Experiments

- NIST Chinese to English Translation Task

	MT06	MT08
baseline	30.58	24.08
$+dom(N = 32)$	30.43	23.91
$+dom(N = 64)$	30.96	24.45
$+dom(N = 128)$	**31.59**	**24.91**
$+dom(N = 256)$	**31.24**	24.26
$+dom(N = 512)$	**31.33**	24.39
$+dom(N = 1024)$	**31.22**	**24.79**
$+dom(N = 2048)$	30.75	23.92

Pairwise dominance model using 128 most frequent words gives **statistically significant improvement.**

Also in NIST Arabic to English MT Eval !!!

Outline

- Problem Statement
- Related Work
- Pairwise Dominance Model
- Experiments
- Discussion

Discussion

Pairwise Dominace model:

- improves Hiero via lexicalization idea,

- is a bilingual feature,

- is a non-local feature

What's Beyond?

- Better anchors

Thank you !

hendra@umiacs.umd.edu

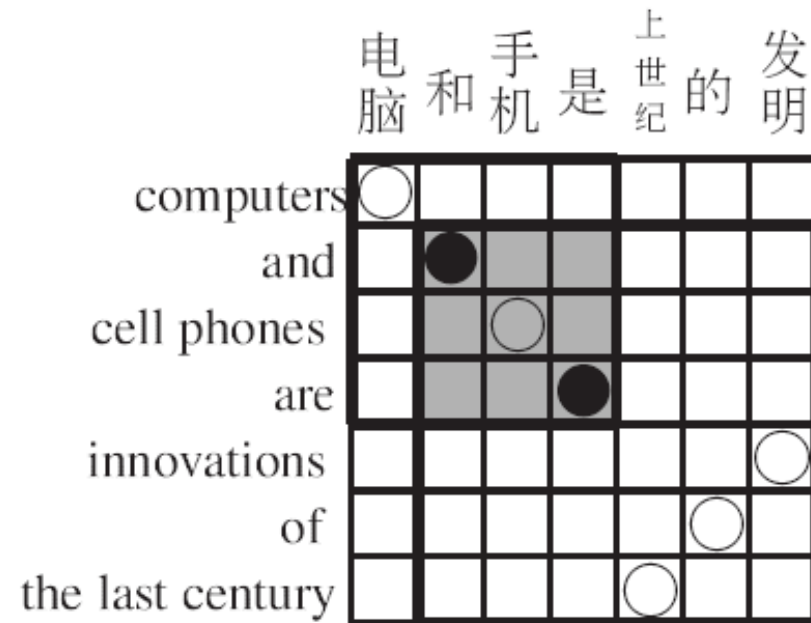# Learned Statistics

Y'	Y''	left-First	right-First	dont-Care	nei-ther
和 (and)	是 (are)	0.11	0.16	**0.68**	0.05
是 (are)	的 (of)	**0.57**	0.15	0.06	0.22

Illustration

- Don't care

Experiments

- NIST Arabic to English Translation

	MT06	MT08
baseline	41.56	40.06
$+dom(N = 32)$	41.66	40.26
$+dom(N = 64)$	**42.03**	**40.73**
$+dom(N = 128)$	**42.66**	**41.08**
$+dom(N = 256)$	**42.28**	40.69
$+dom(N = 512)$	41.97	**40.95**
$+dom(N = 1024)$	42.05	40.55
$+dom(N = 2048)$	**42.48**	**41.47**

Topological Ordering

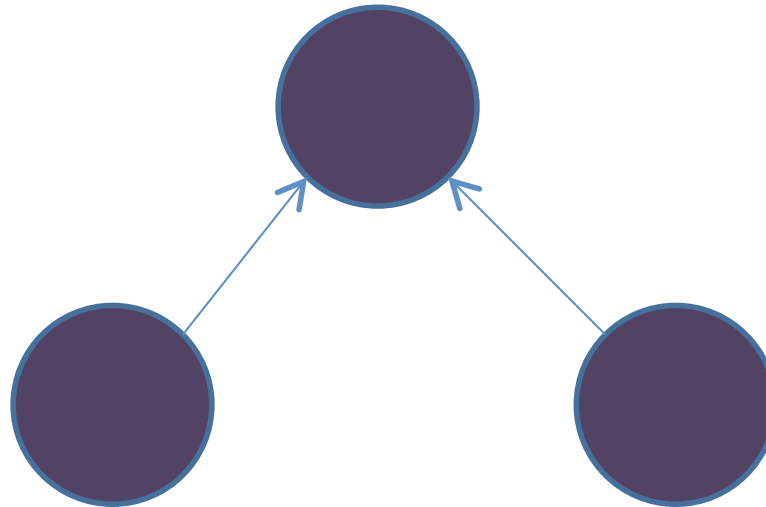In **non-Hiero** world, derivation tree:
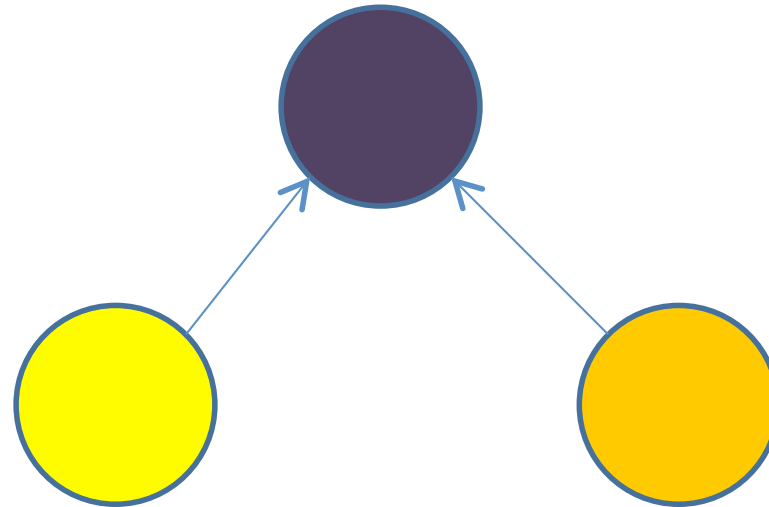
Topological Ordering: ⬤ ≺ ⬤ ≺ ⬤

Topological Ordering

In Hiero world, derivation tree:

Topological Ordering:

Pairwise Dominance Model

- How about **training**?

  Given f, e, a

$$P(d(Y', Y'') = \texttt{leftFirst} \mid Y', Y'') = \frac{Count(d(Y', Y'') = \texttt{leftFirst})}{\sum_{\forall val} Count(d(Y', Y'') = val)}$$

- What are Y' and Y''?

  Neighboring function words ≈ frequent words

- Why neighboring function words?

  Simplifies decoding (and training too!)

Pairwise Dominance Model

- How about **training**?

  Changes to the original Hiero's rules:

  Keep the alignment information!

$$X \rightarrow \langle X_1 \text{的 发明} , \text{inventions of } X_1 \rangle$$

$$X \rightarrow \langle X_1 \text{的}_2 \text{发明}_3 , \text{inventions}_3 \text{of}_2 X_1 \rangle$$

Pairwise Dominance Model

- How about **decoding**?

  It's like computing *n*-gram language model

(computers   …   century)
(computers … and)(cell   …   century)
                  (cell   …   are) (inventions   …   century)

电脑   和   手机   是   上个世
computers   and   cell phones   are   the last century   innovations of
纪   的   发明

Pairwise Dominance Model

- How about **decoding**?

It's like computing *n*-gram language model

(和/and   ...   的/of)
(和/and   )(是/are   ...   的/of)
              (是/are   ) (的/of   )

电脑   和   手机   是   上个世

纪 computers   的 and   cell 发 phones 明   are   the last century   of   innovations

Don't forget to keep alignment information!