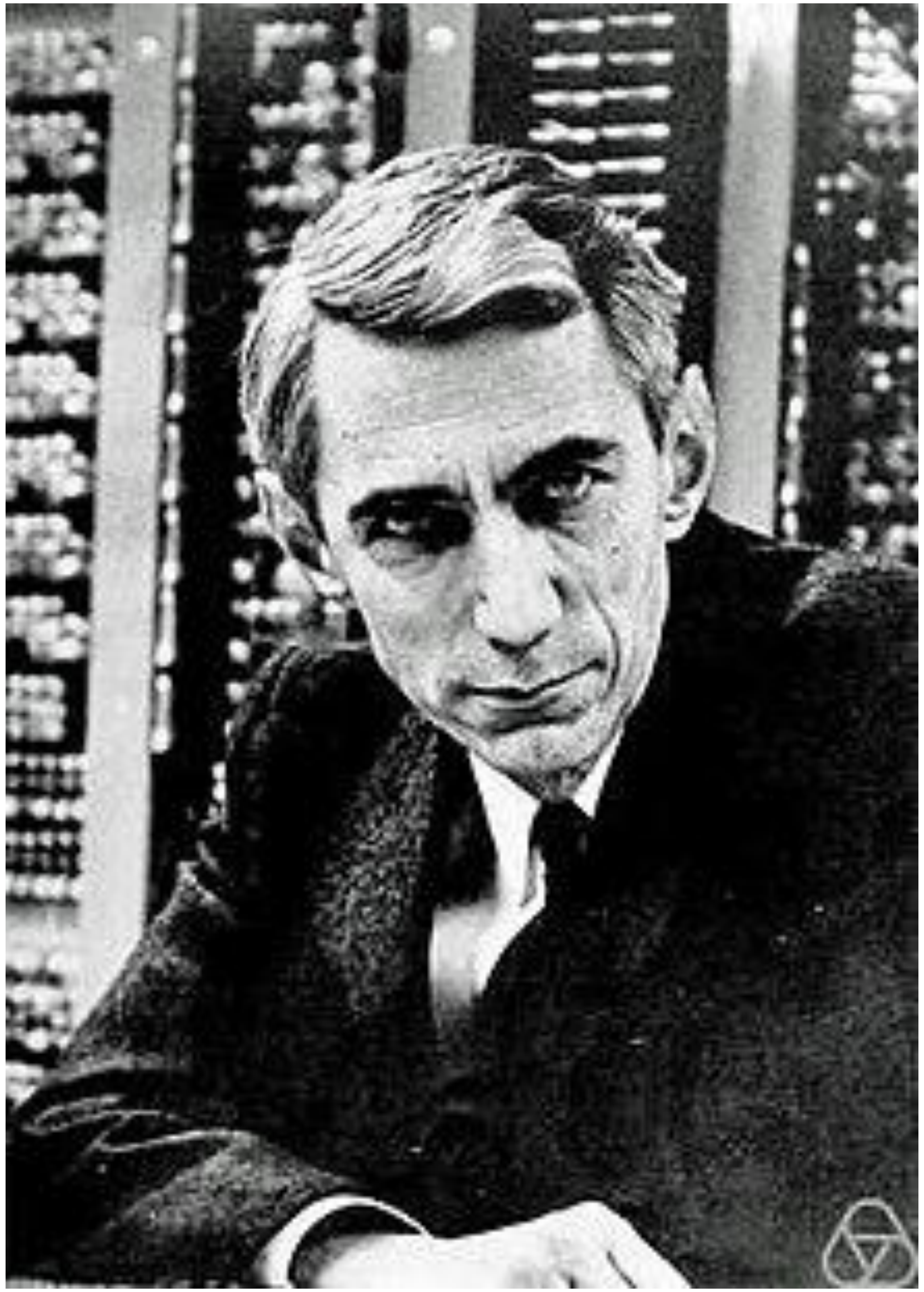
Shannon Revisited: N-grams and autoregressive



Prediction and Entropy of Printed English

By C. E. SHANNON

(*Manuscript Received Sept. 15, 1950*)

A new method of estimating the entropy and redundancy of a language is described. This method exploits the knowledge of the language statistics possessed by those who speak the language, and depends on experimental results in prediction of the next letter when the preceding text is known. Results of experiments in prediction are given, and some properties of an ideal predictor are developed.

Markov Chain

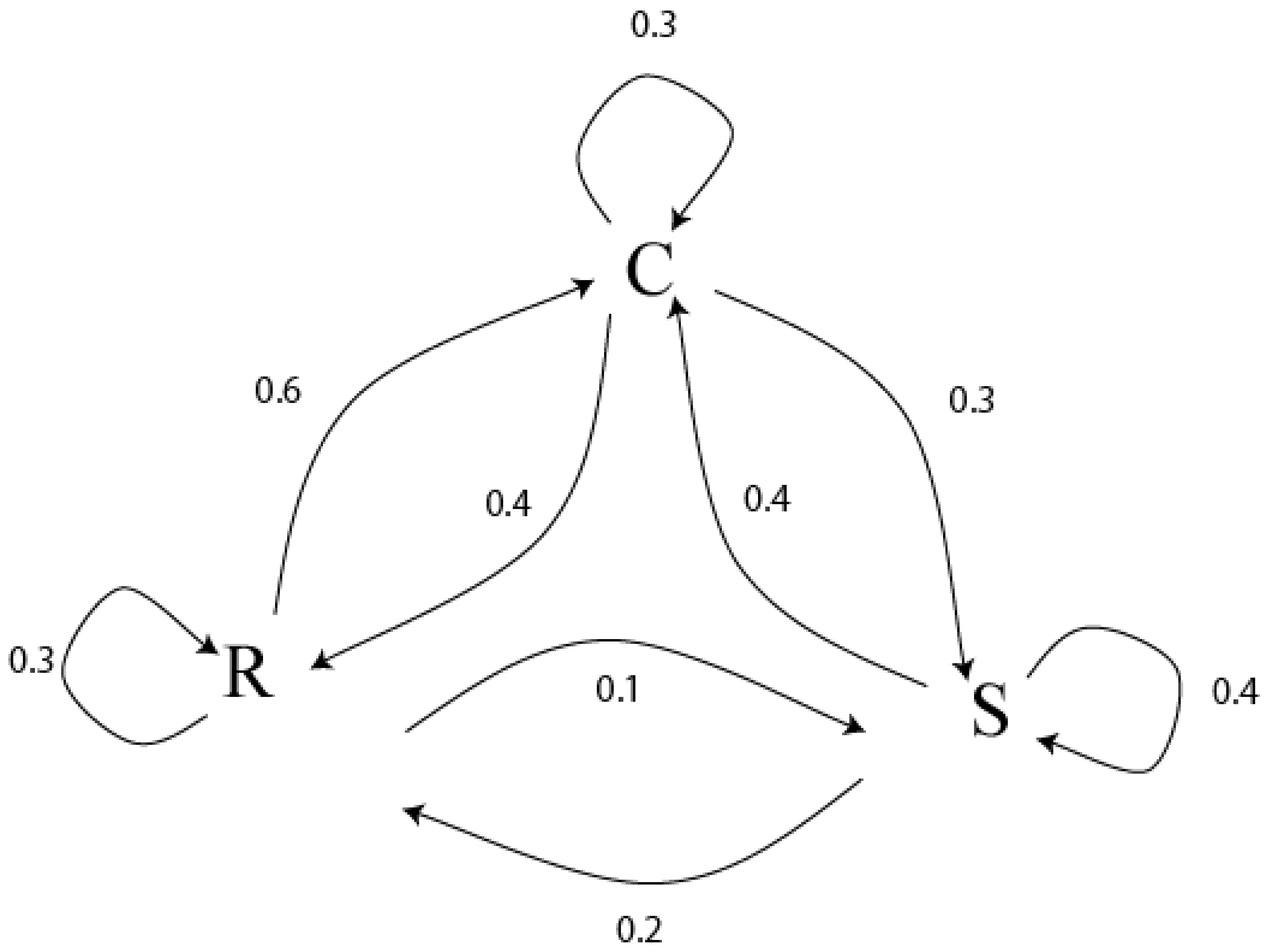
Let’s predict weather:

- Given today’s weather only, we want to know tomorrow’s
- Suppose weather can only be {Sunny, Cloudy, Raining}

The “Weather Channel” algorithm:

- Over a long period of time, record:
  - How often S followed by R
  - How often S followed by S
  - Etc.
- Compute percentages for each state:
  - P(R|S), P(S|S), etc.
- Predict the state with highest probability!
- It’s a Markov Chain

Markov Chain



$$\begin{pmatrix} 0.3 & 0.6 & 0.1 \\ 0.4 & 0.3 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}$$

What if we know today and yesterday's weather?

Text Synthesis

[Shannon,'48] proposed a way to generate English-looking text using N-grams:

- Assume a generalized Markov model
- Use a large text to compute prob. distributions of each letter given N-1 previous letters
- Starting from a seed repeatedly sample this Markov chain to generate new letters
- Also works for whole words

WE NEED TO EAT CAKE

Mark V. Shaney (Bell Labs)

Results (using `alt.singles` corpus):

- *"As I've commented before, really relating to someone involves standing next to impossible."*
- *"One morning I shot an elephant in my arms and kissed him."*
- *"I spent an interesting evening recently with a grain of salt"*

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Bob Dylan, *Tangled up in Blue*

slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

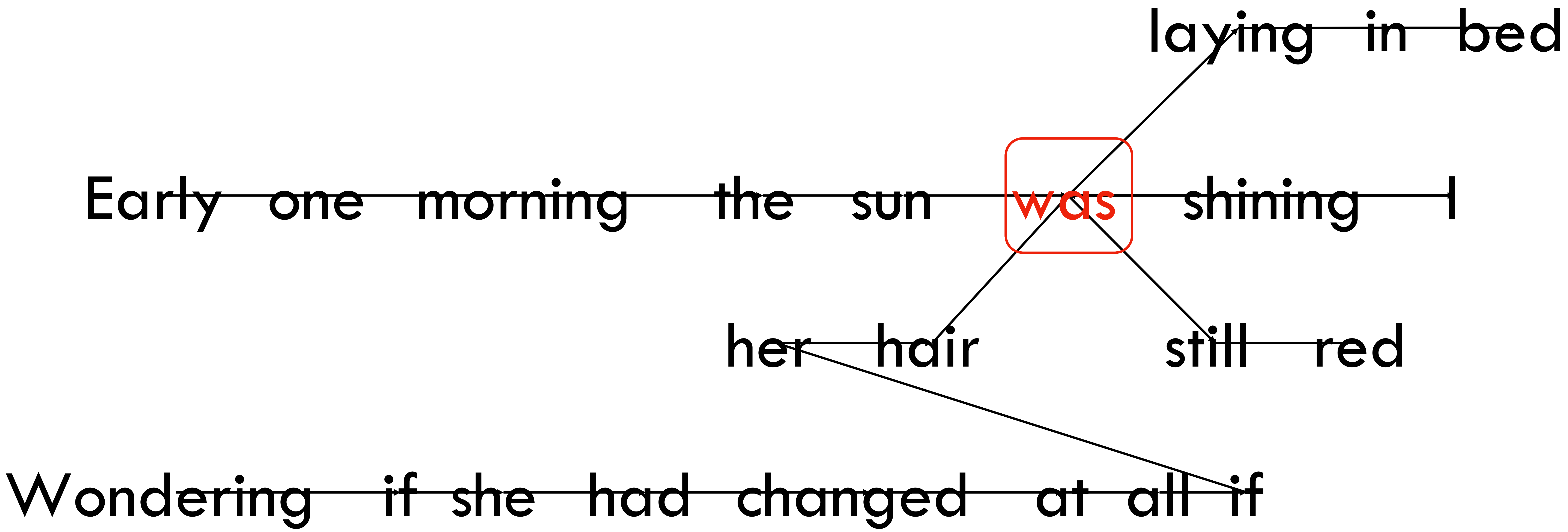
Wondering if she had changed at all if her hair was still red

slide from Steve Seitz's [video]

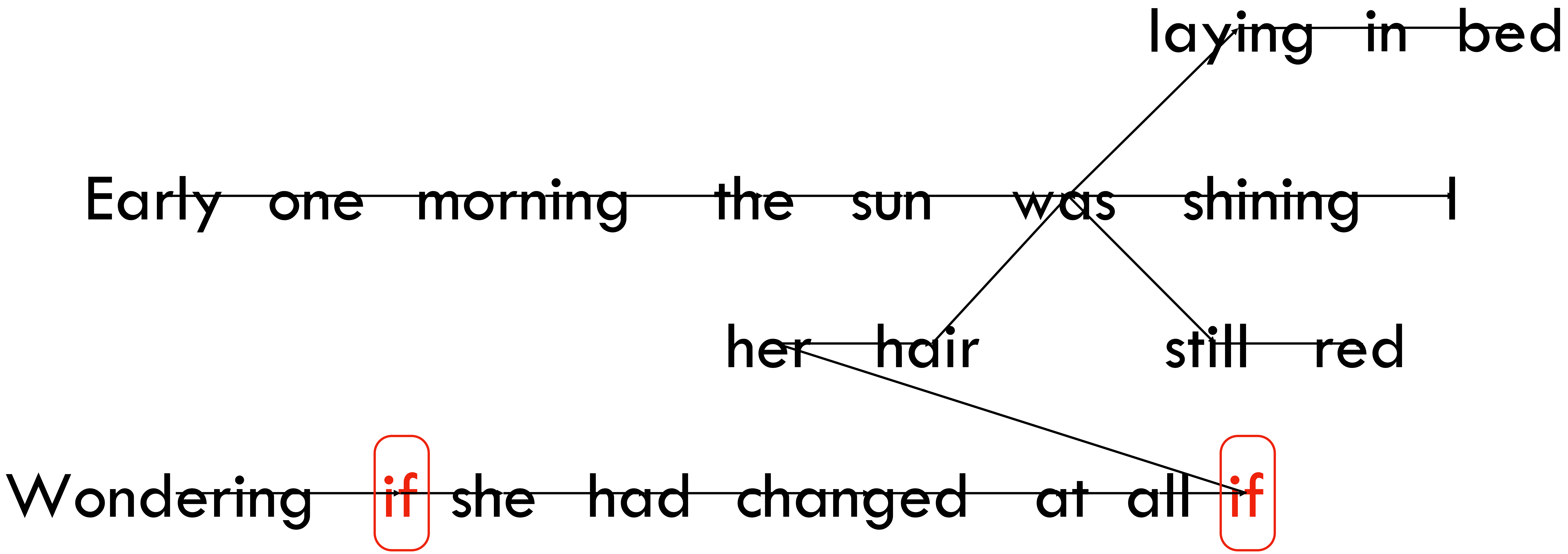
Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

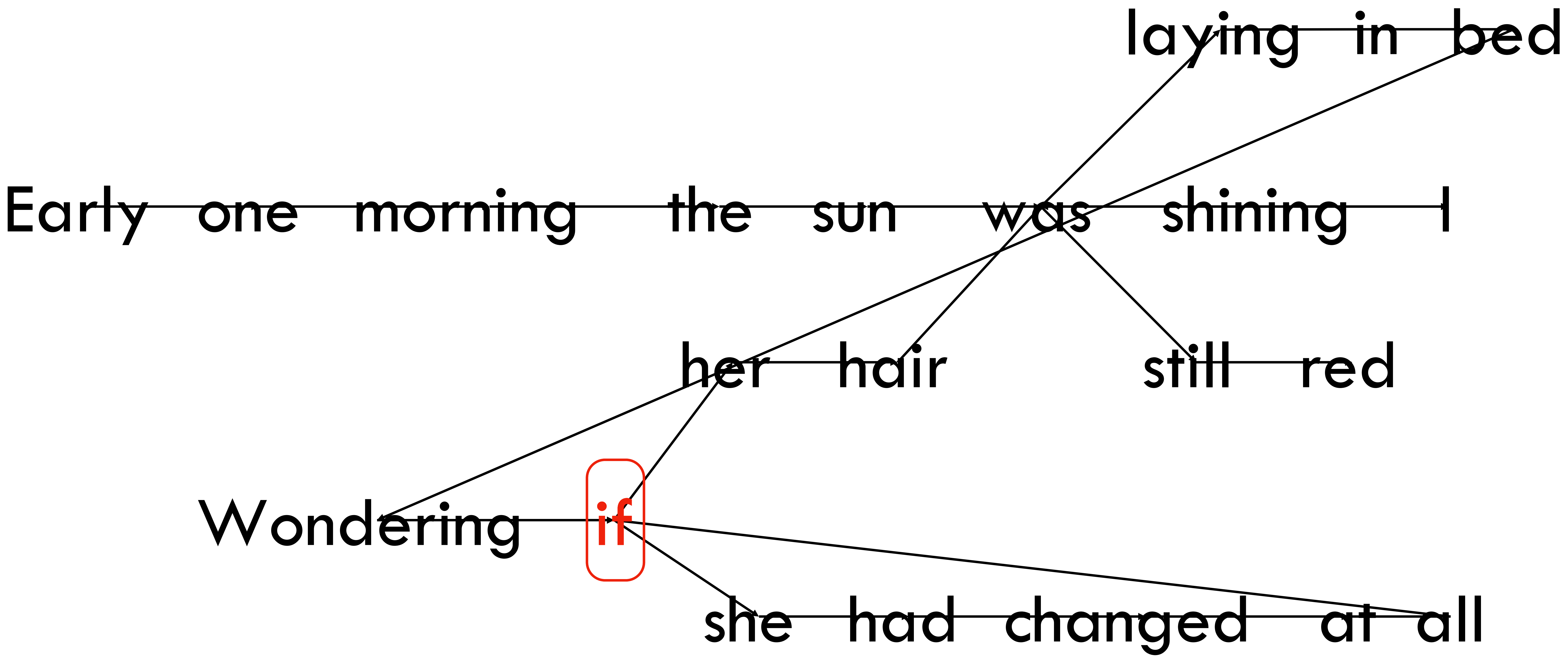
slide from Steve Seitz's video



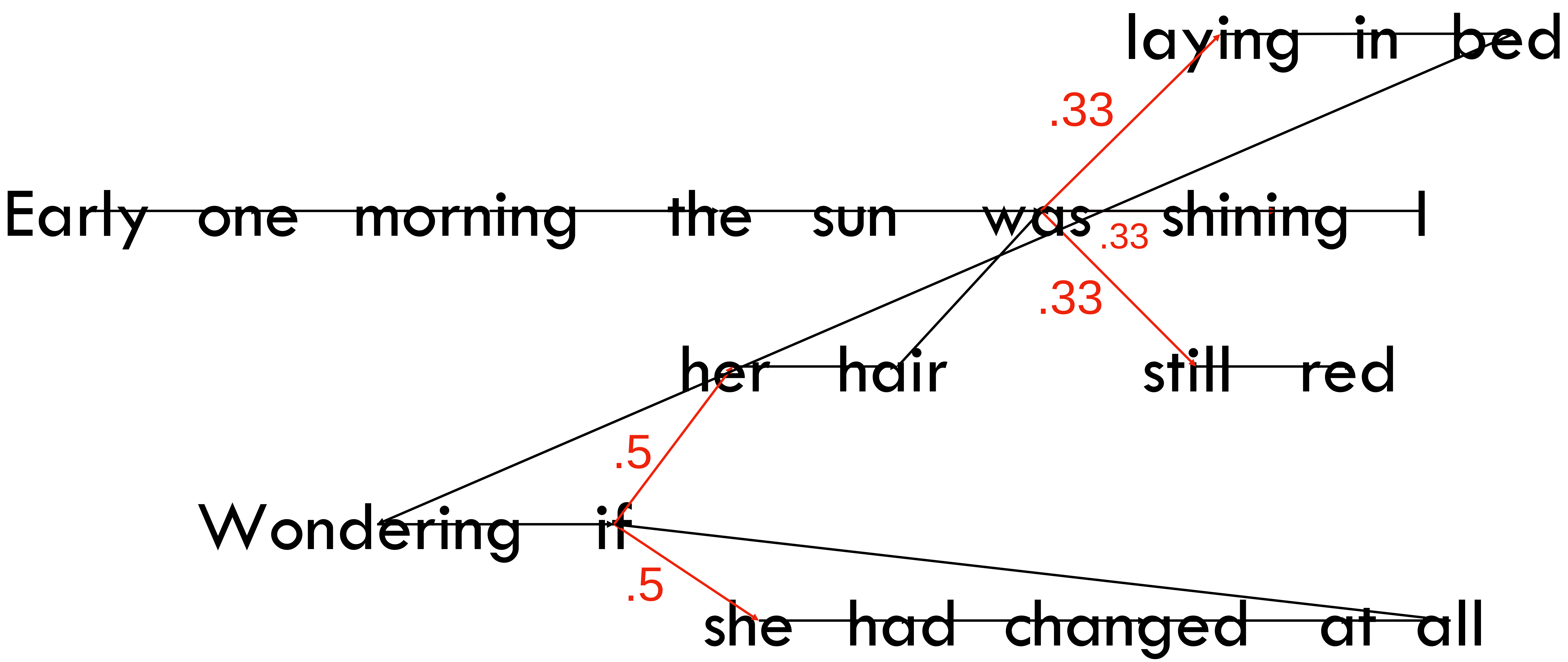
slide from Steve Seitz's video



slide from Steve Seitz's [video]

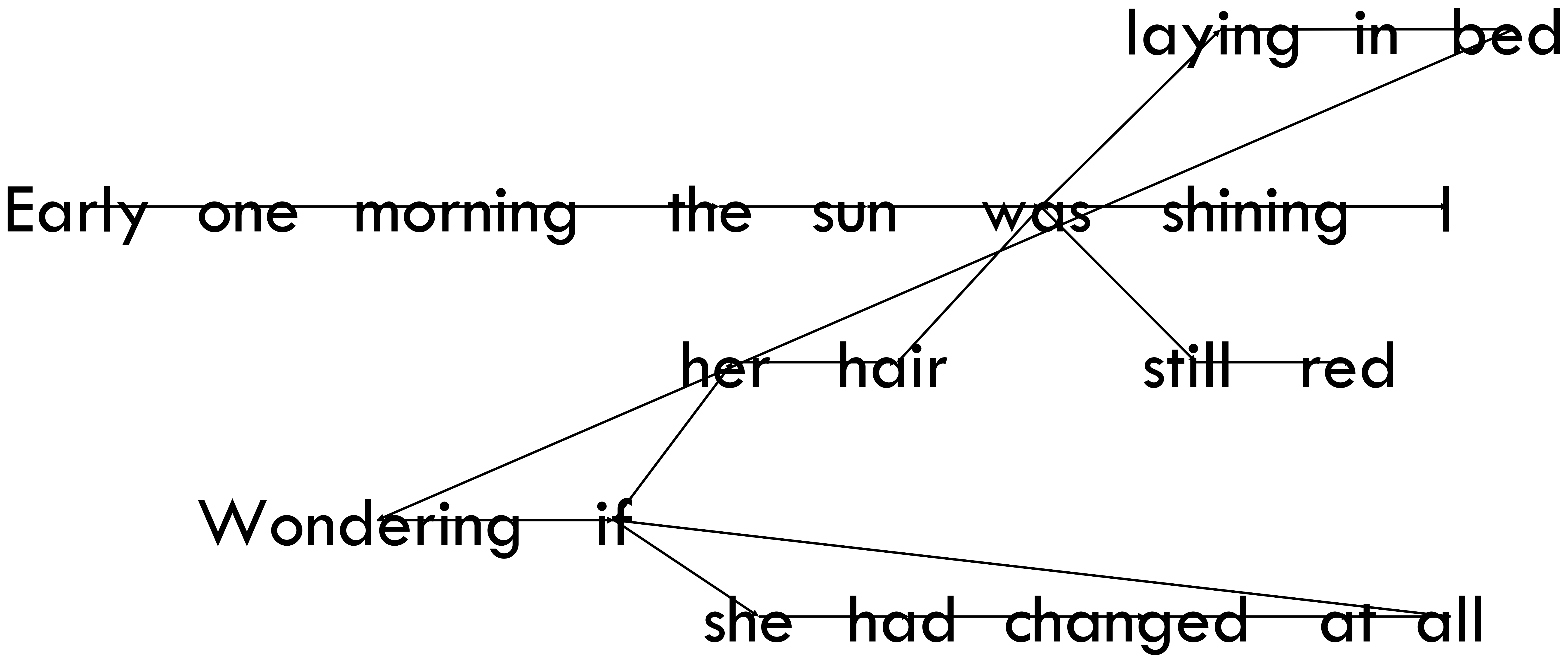


slide from Steve Seitz's video



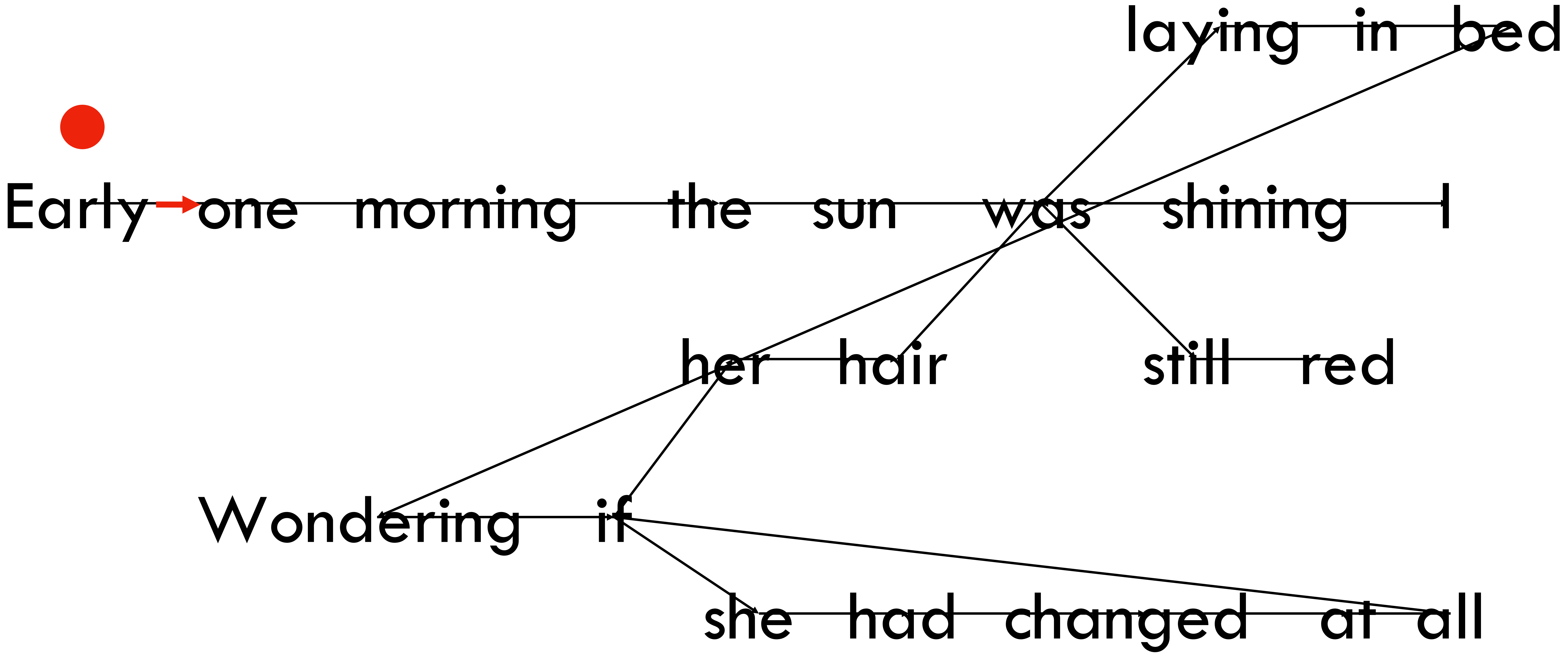
Language Model

slide from Steve Seitz's [video]

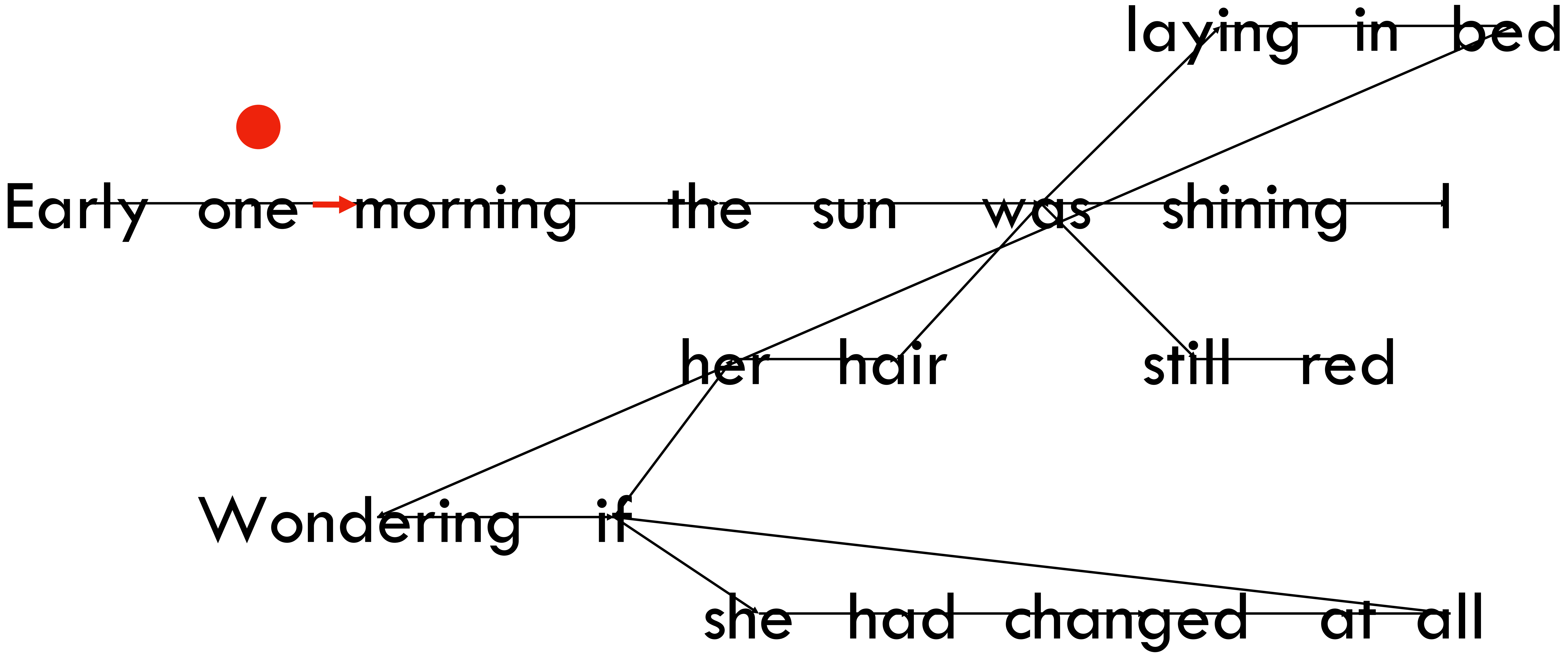


slide from Steve Seitz’s [video]

Early

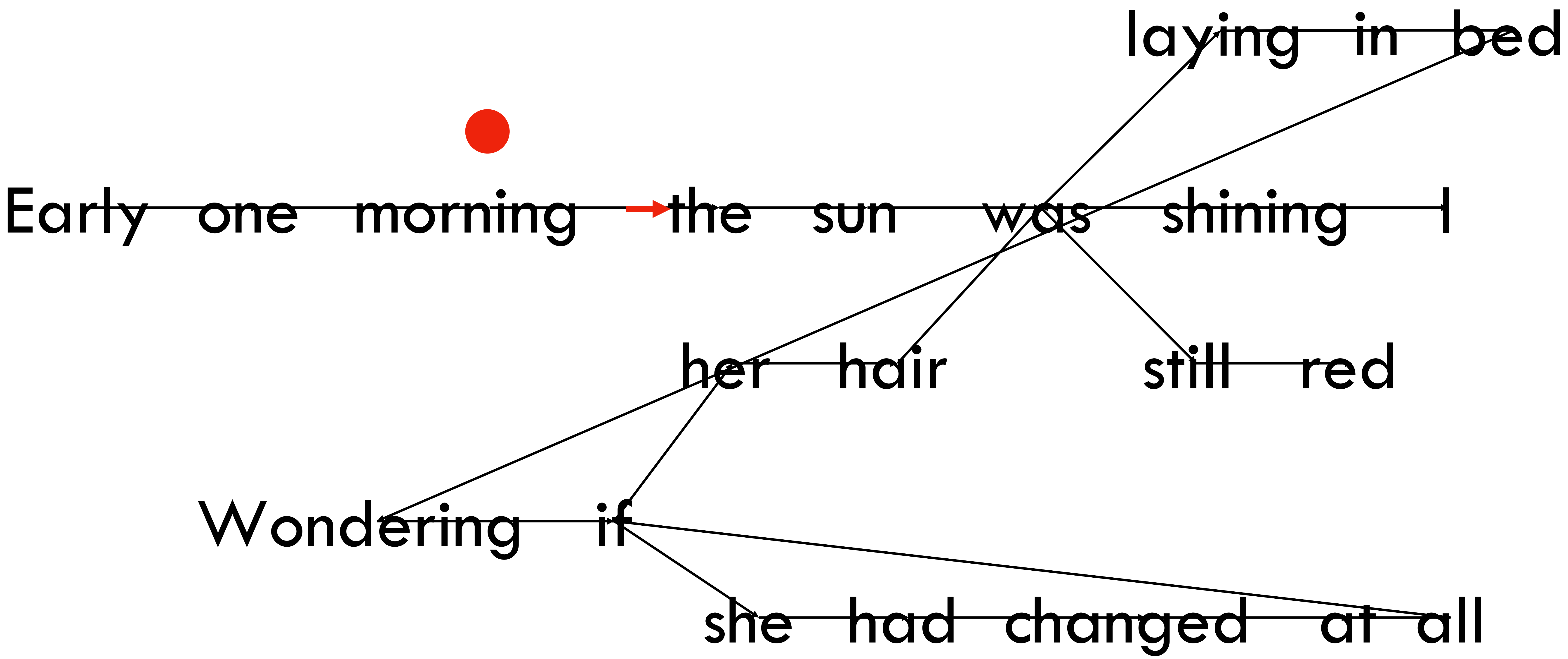

laying in bed
Early one morning the sun was shining I
her hair still red
Wondering if
she had changed at all

Early one



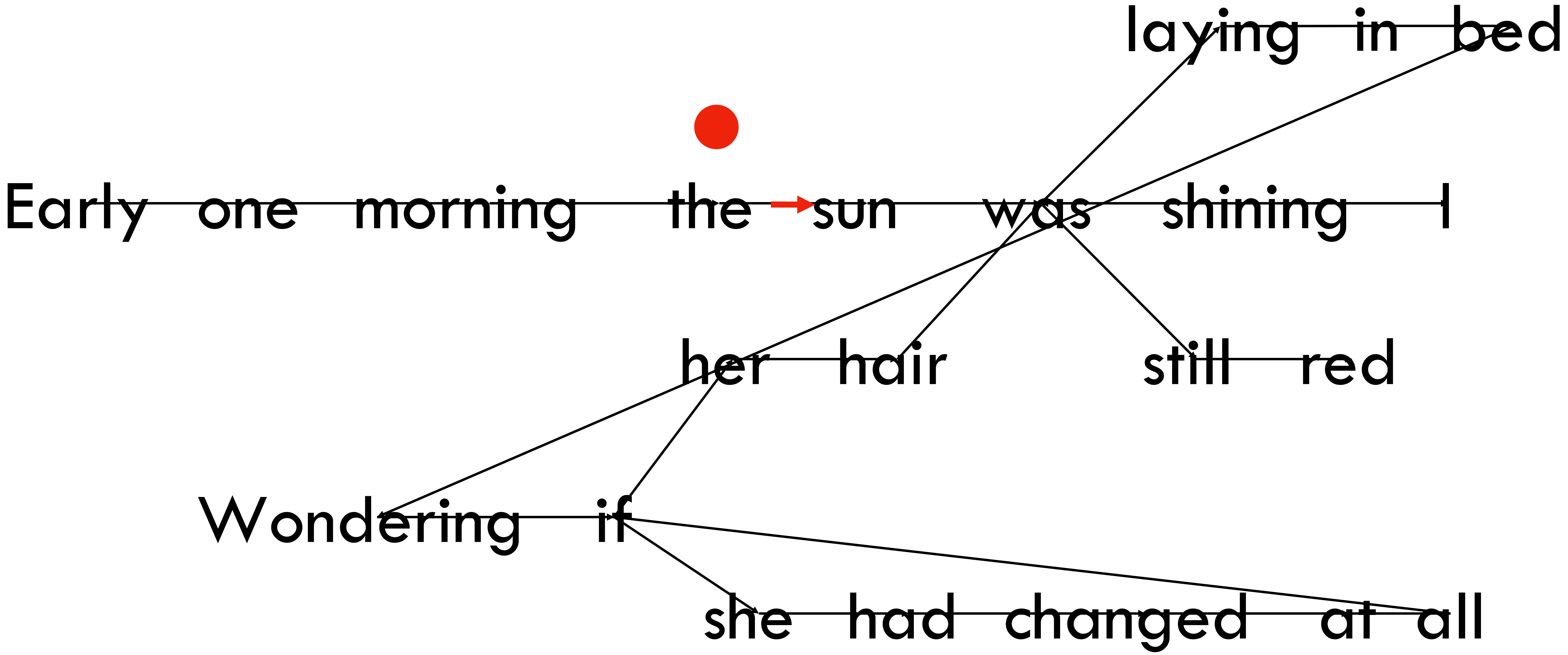
slide from Steve Seitz's [video]

Early one morning



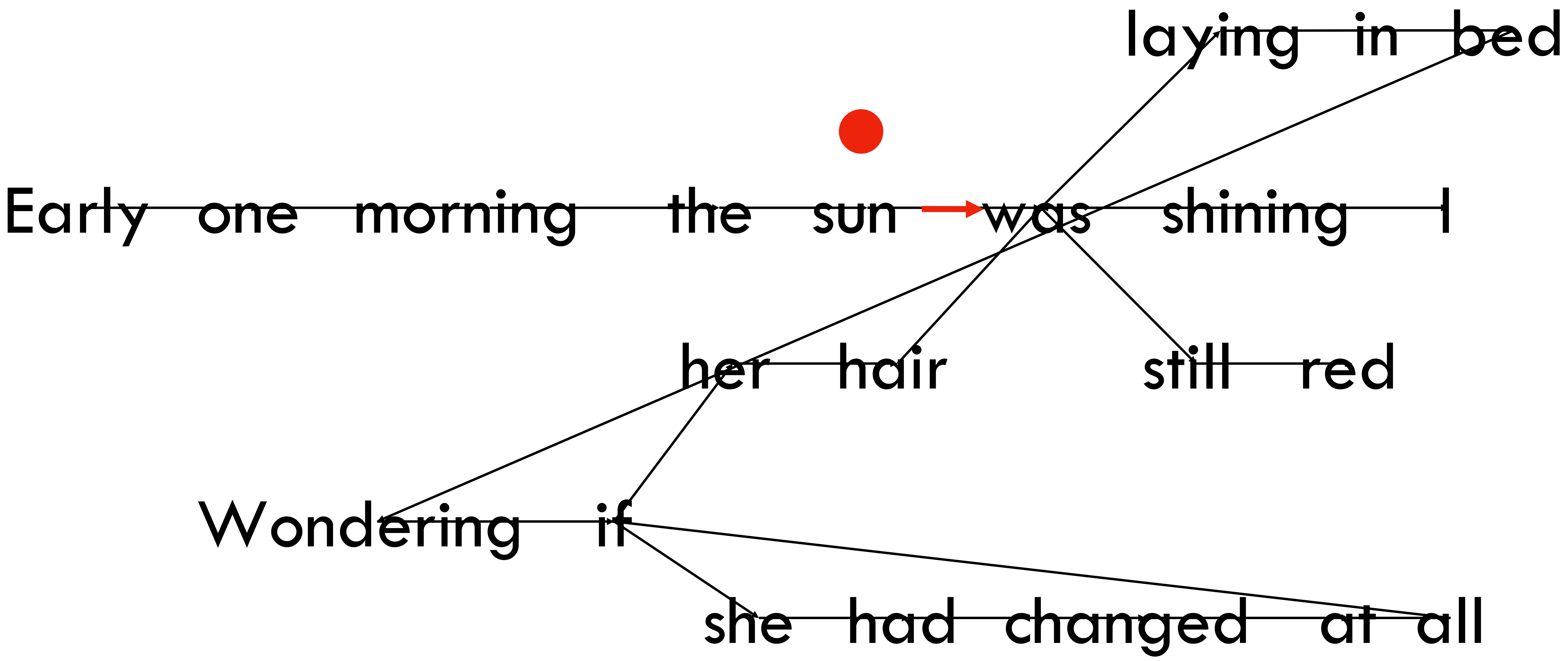
slide from Steve Seitz's [video]

Early one morning the



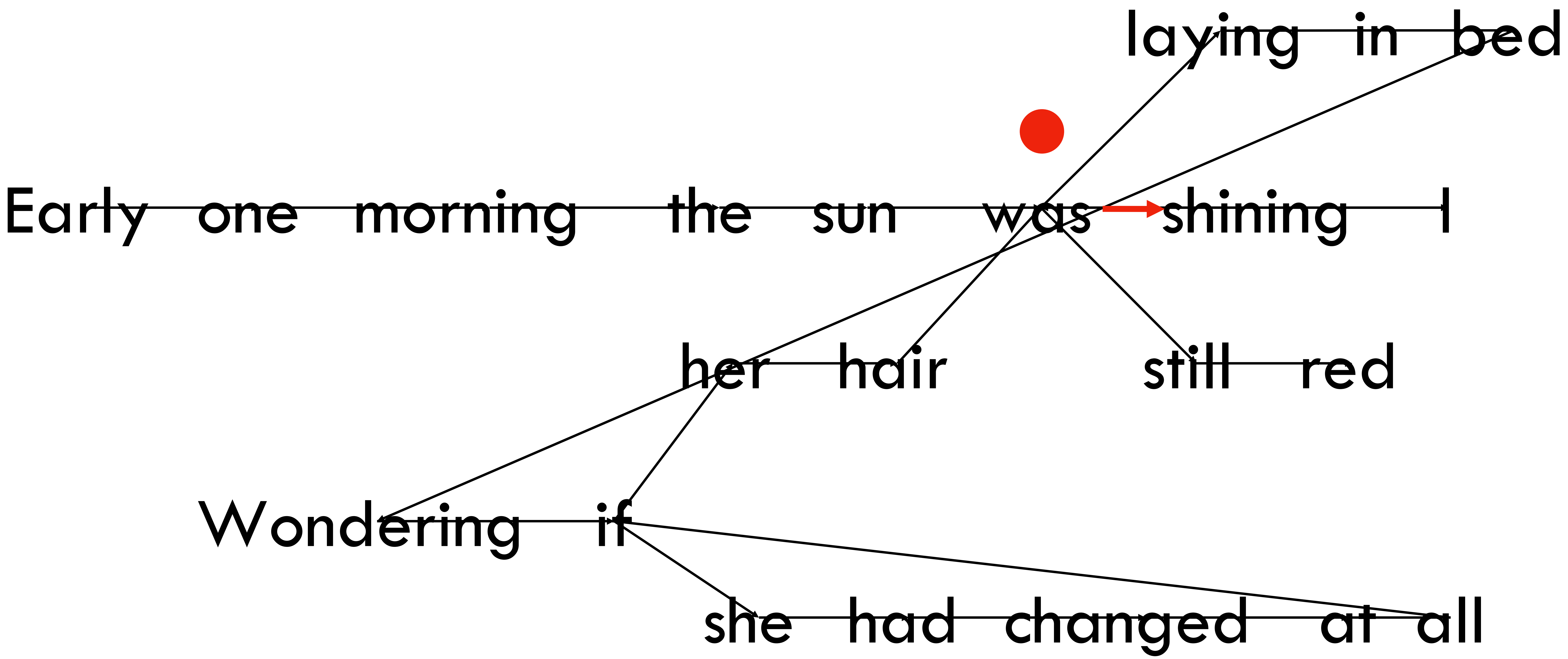
slide from Steve Seitz's [video]

Early one morning the sun



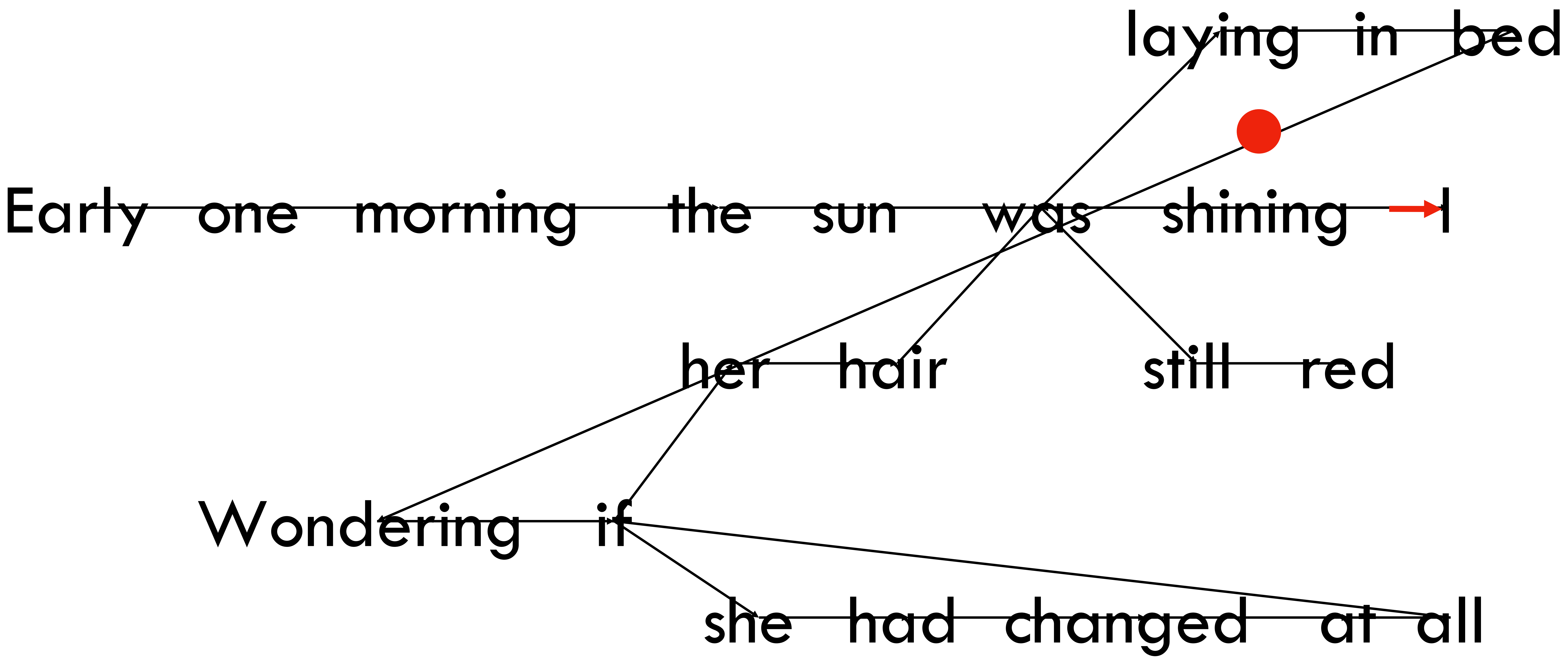
slide from Steve Seitz's video

Early one morning the sun was



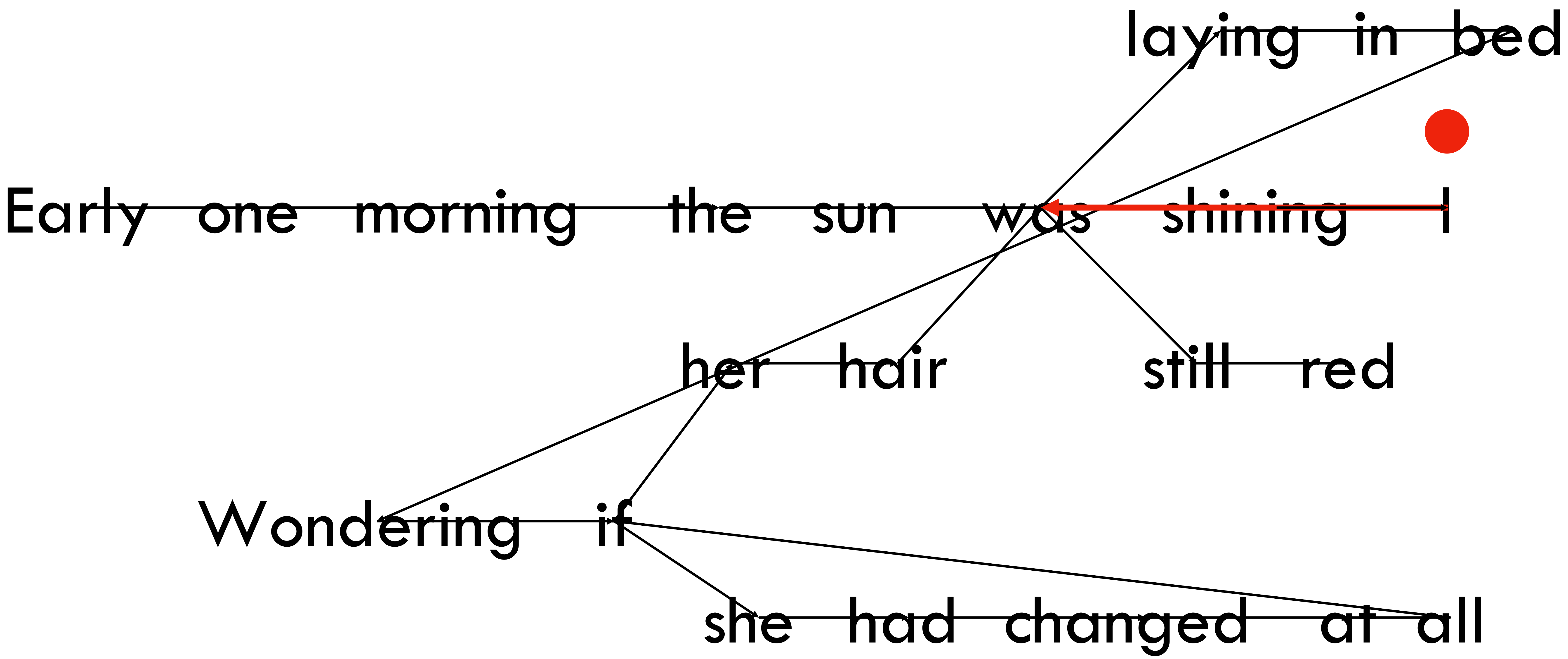
slide from Steve Seitz's [video]()

Early one morning the sun was shining



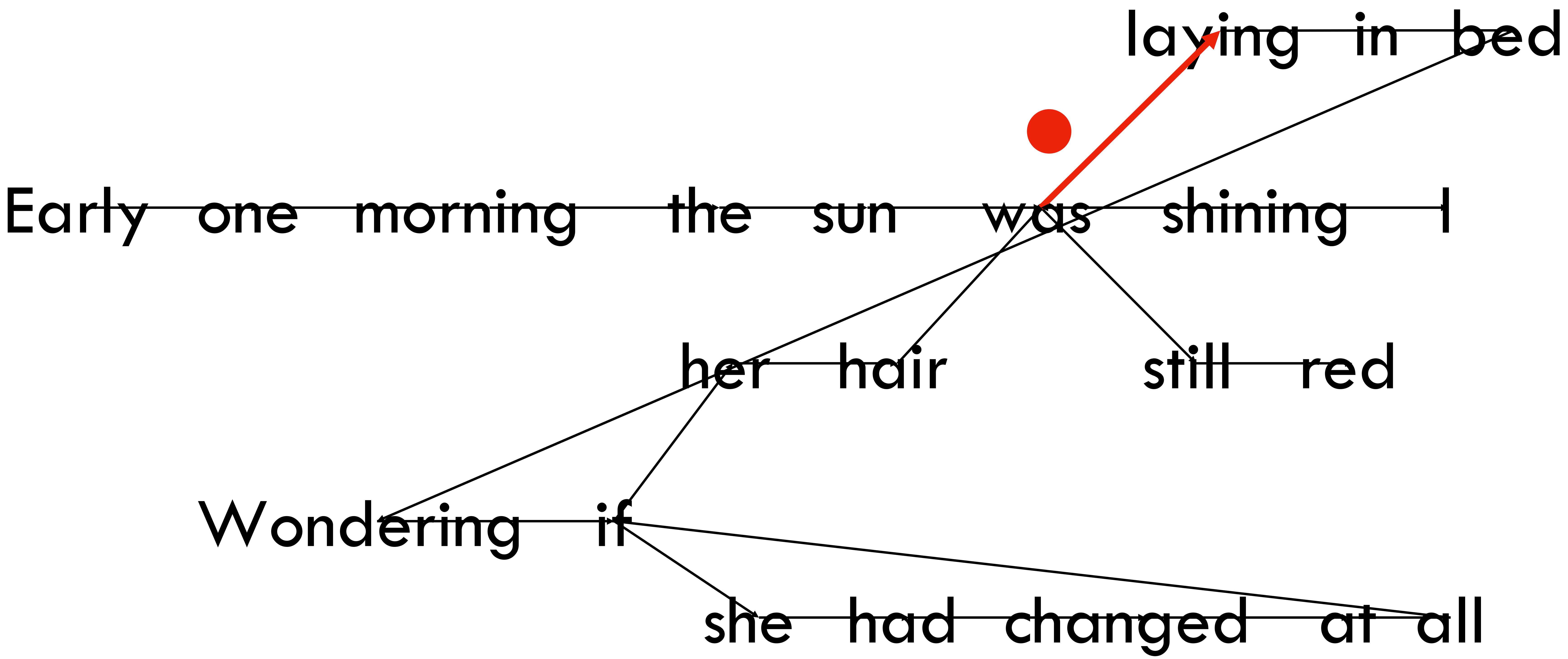
slide from Steve Seitz’s [video]

Early one morning the sun was shining I



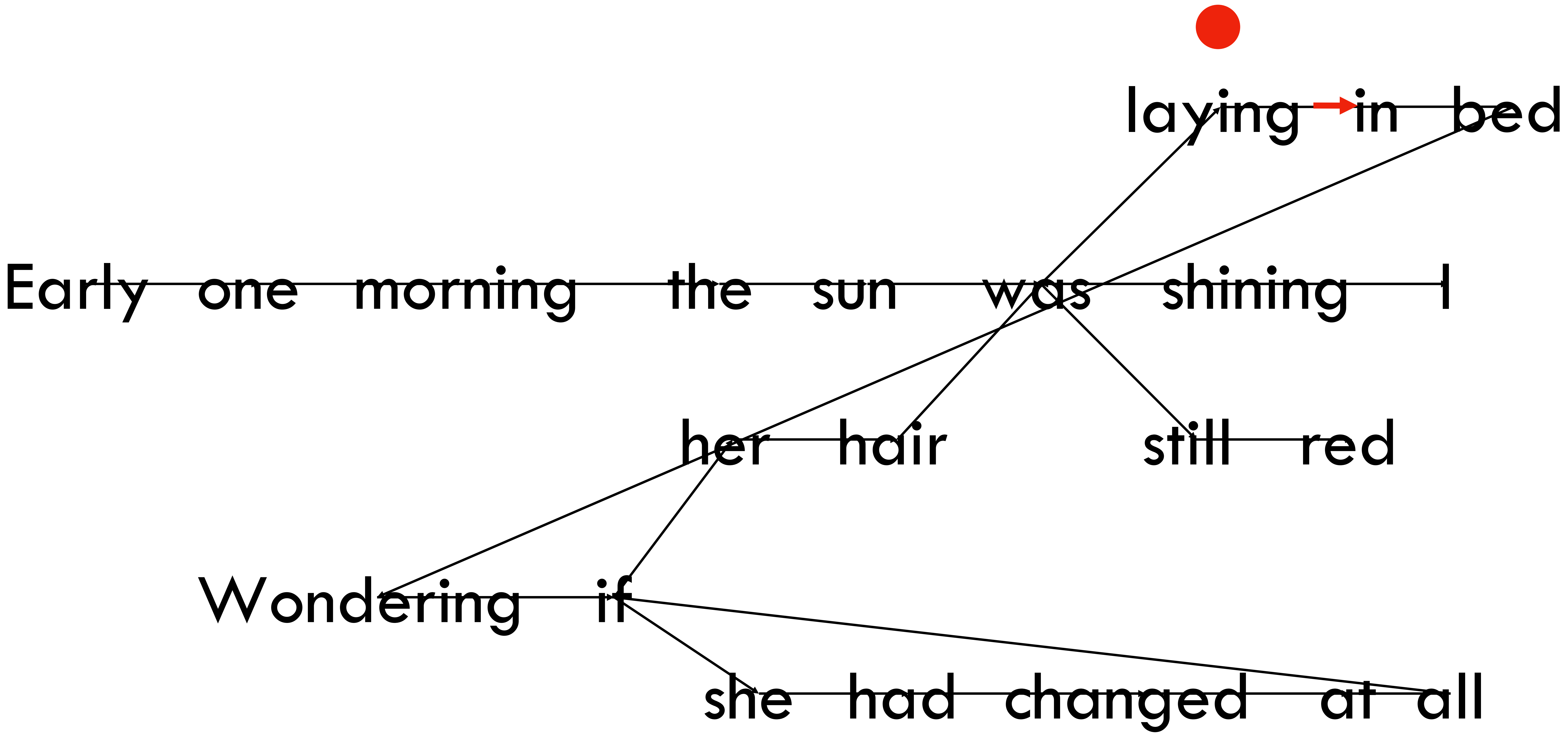
slide from Steve Seitz's video

Early one morning the sun was shining I was



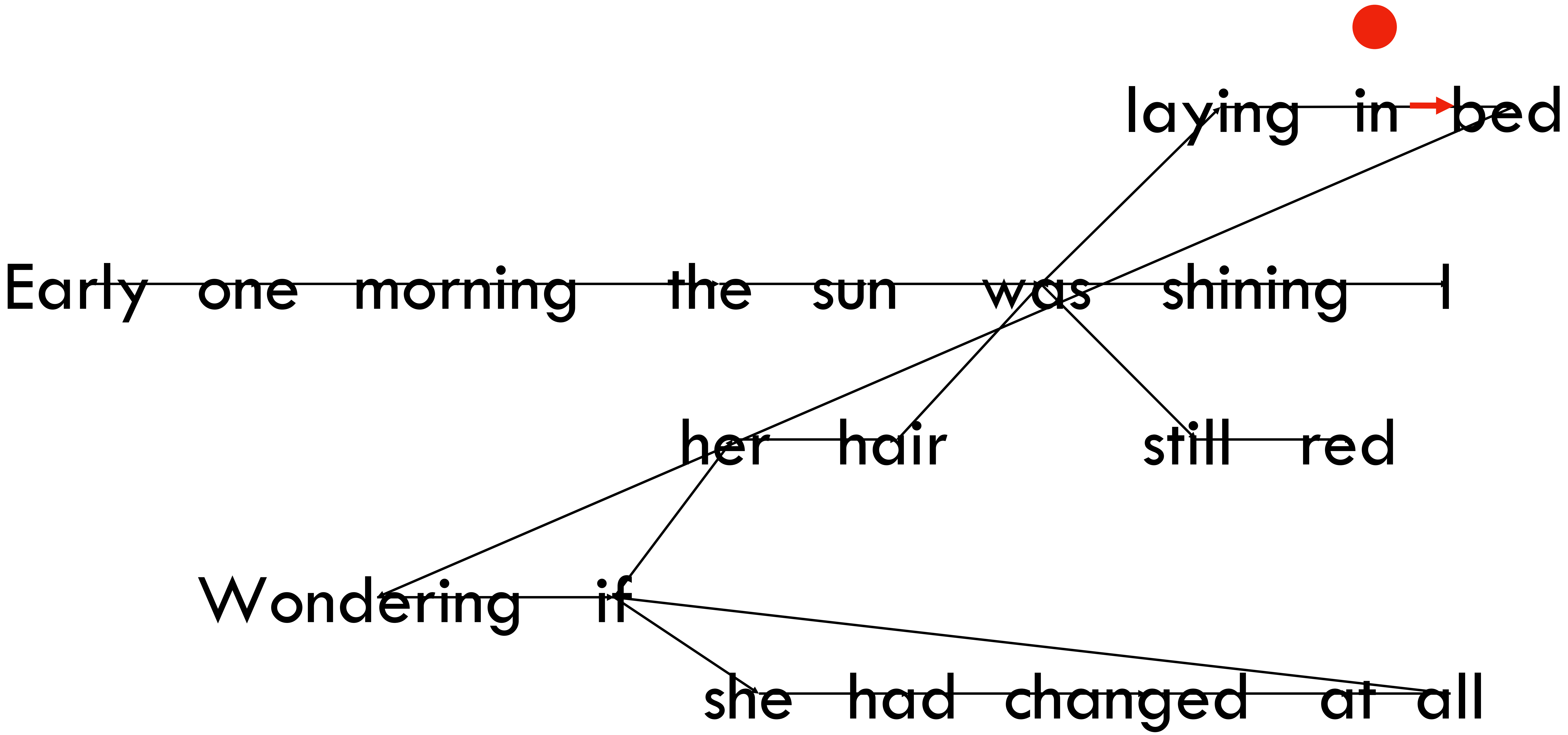
slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying



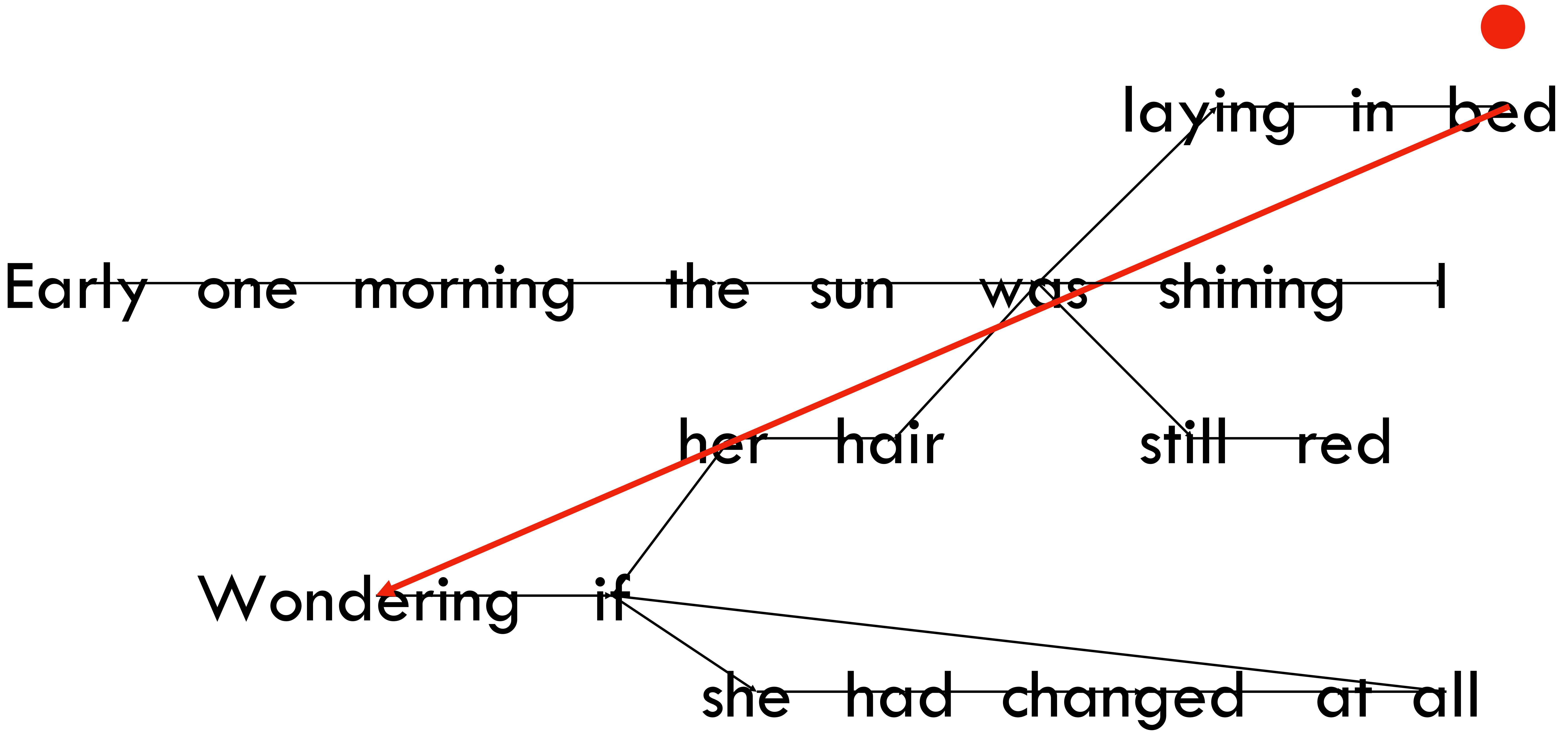
slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in



slide from Steve Seitz's video

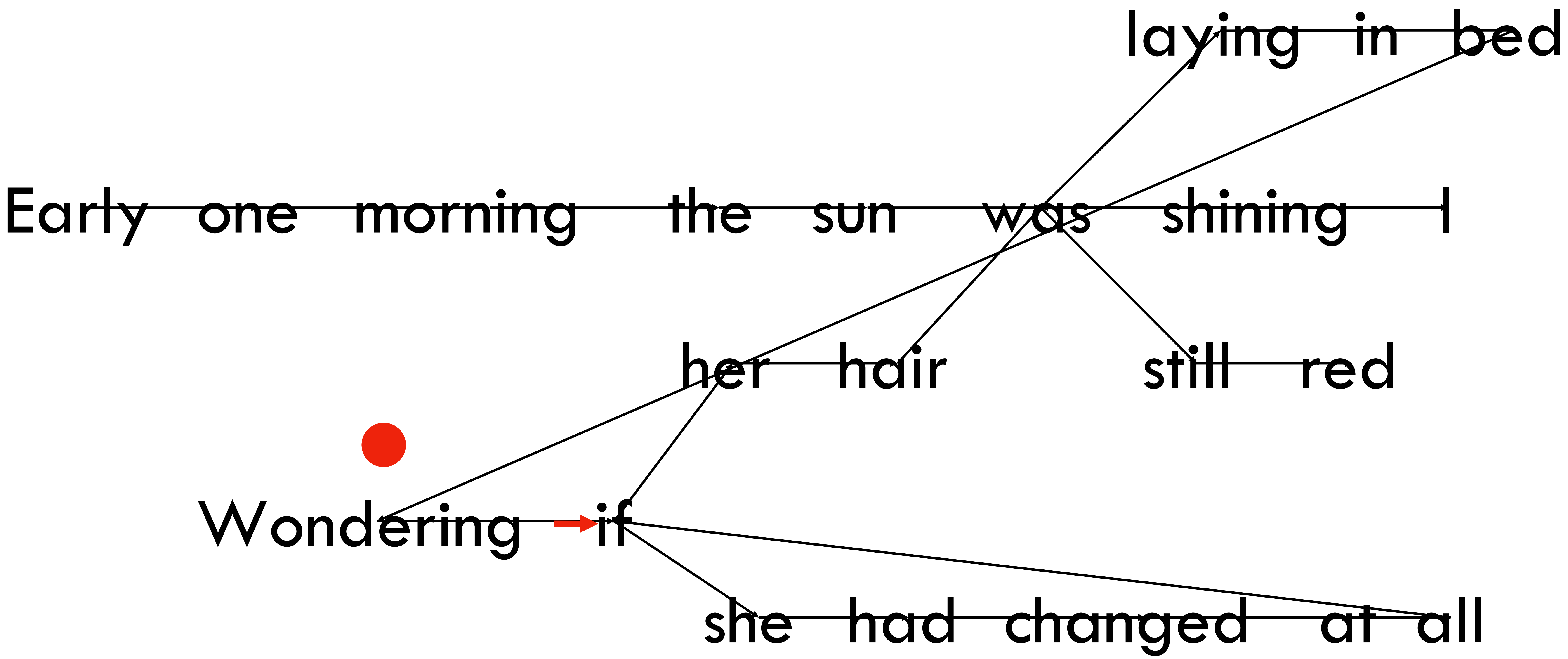
Early one morning the sun was shining I was laying in bed



slide from Steve Seitz's video

Early one morning the sun was shining I was laying in bed

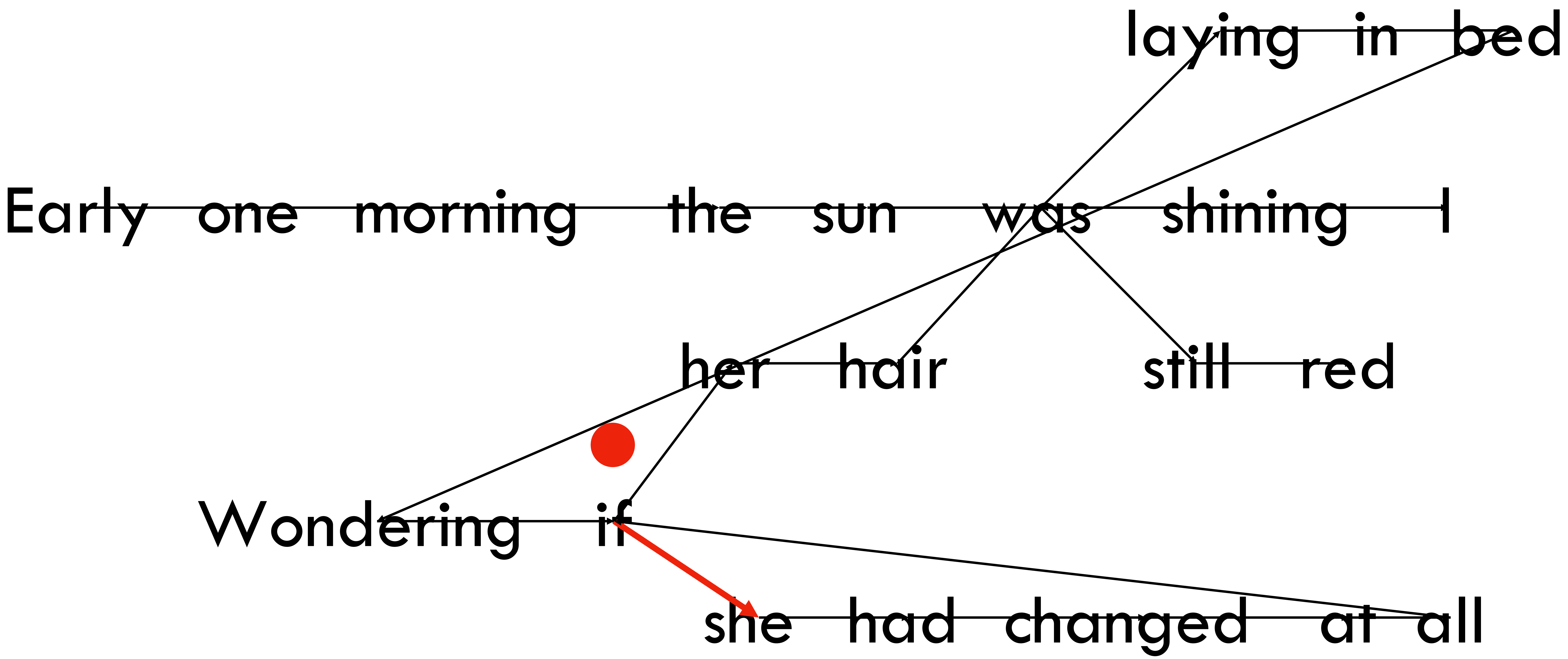
Wondering



slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

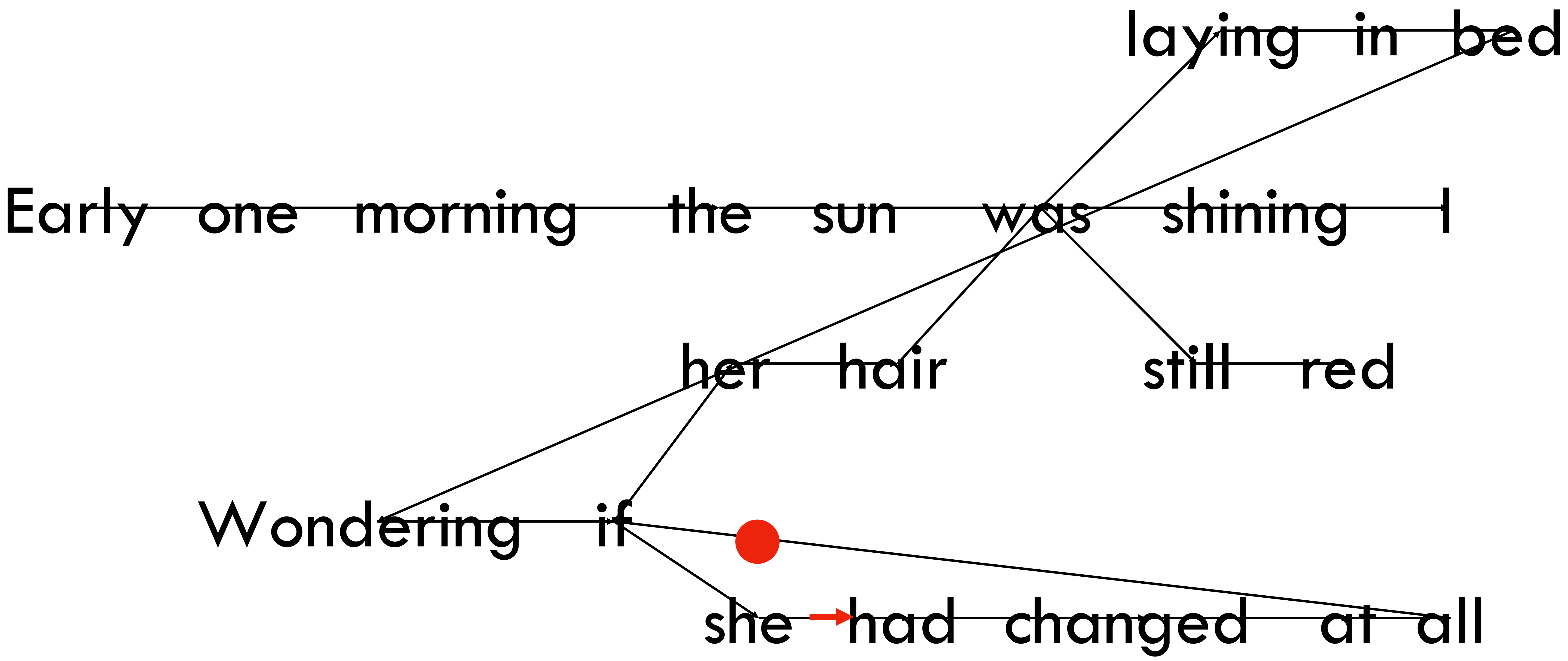
Wondering if



slide from Steve Seitz's video

Early one morning the sun was shining I was laying in bed

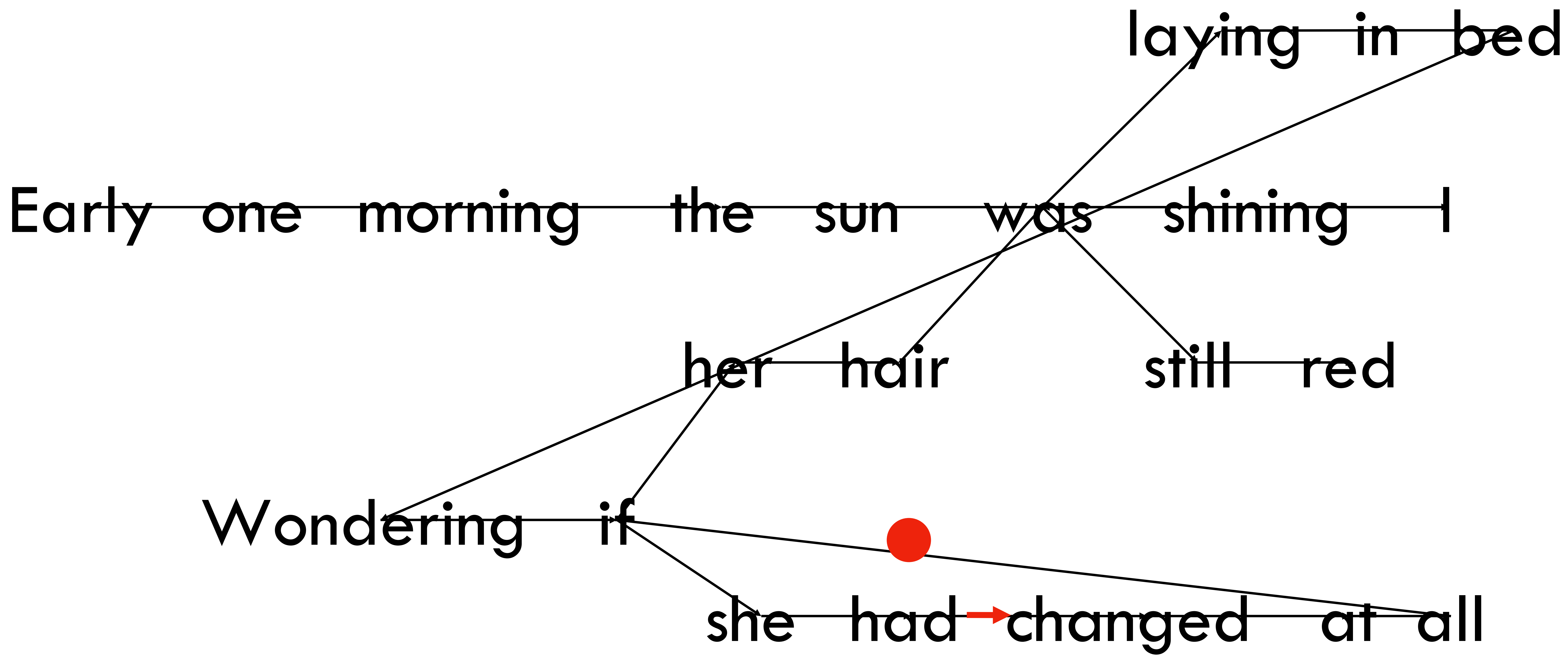
Wondering if she



slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

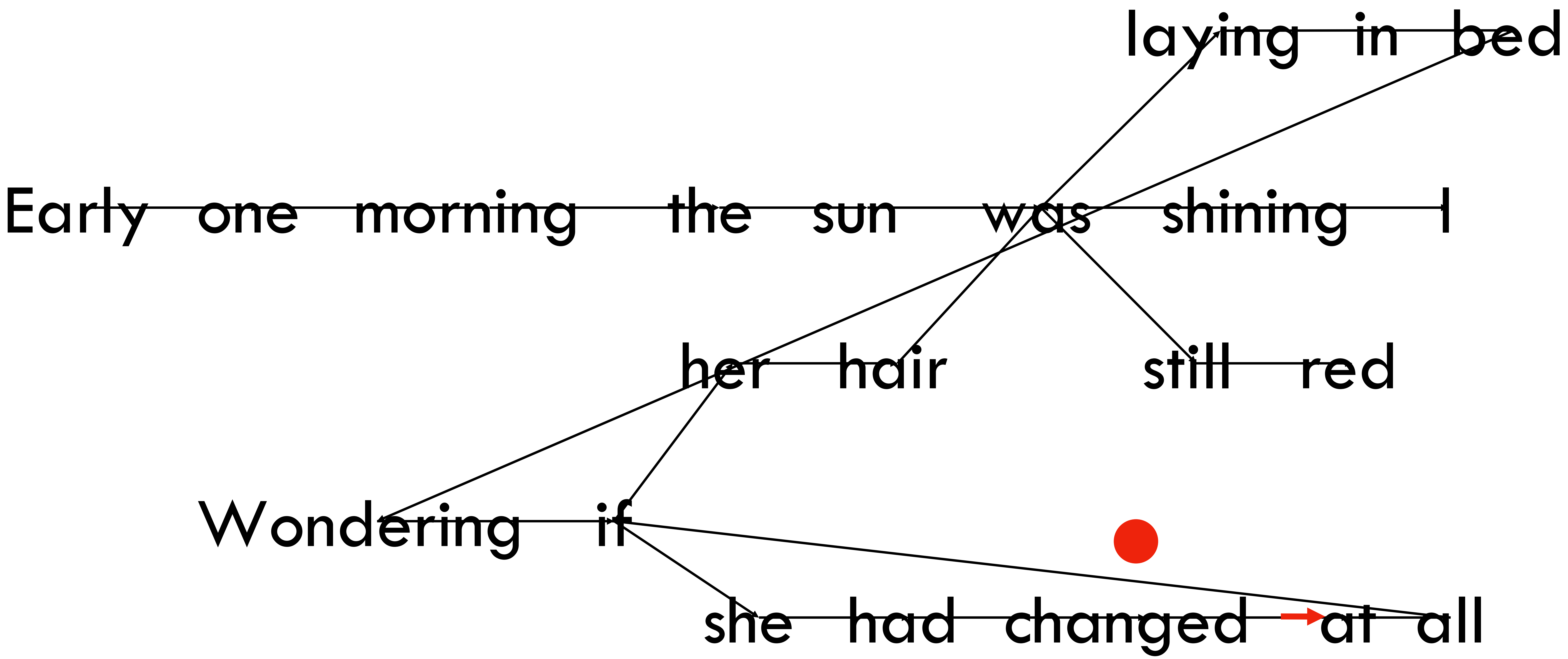
Wondering if she had



slide from Steve Seitz's video

Early one morning the sun was shining I was laying in bed

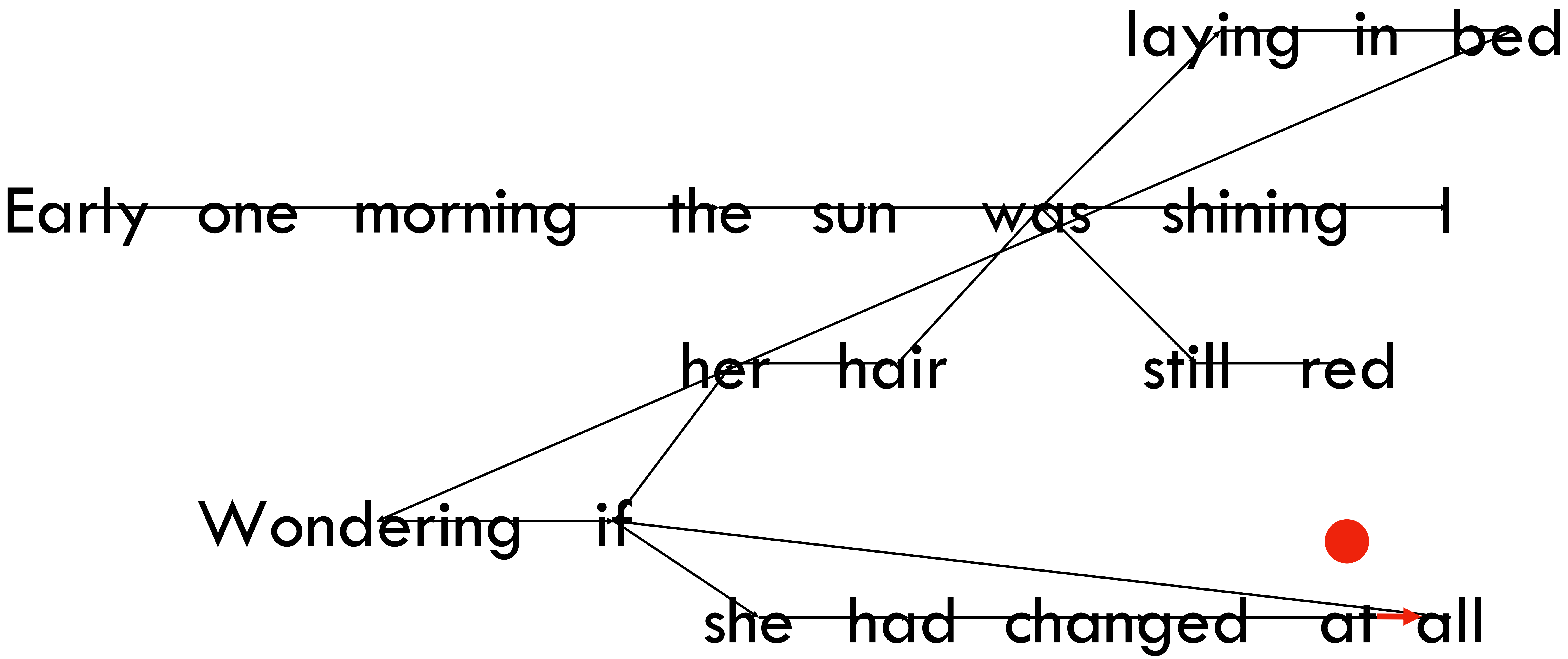
Wondering if she had changed



slide from Steve Seitz's video

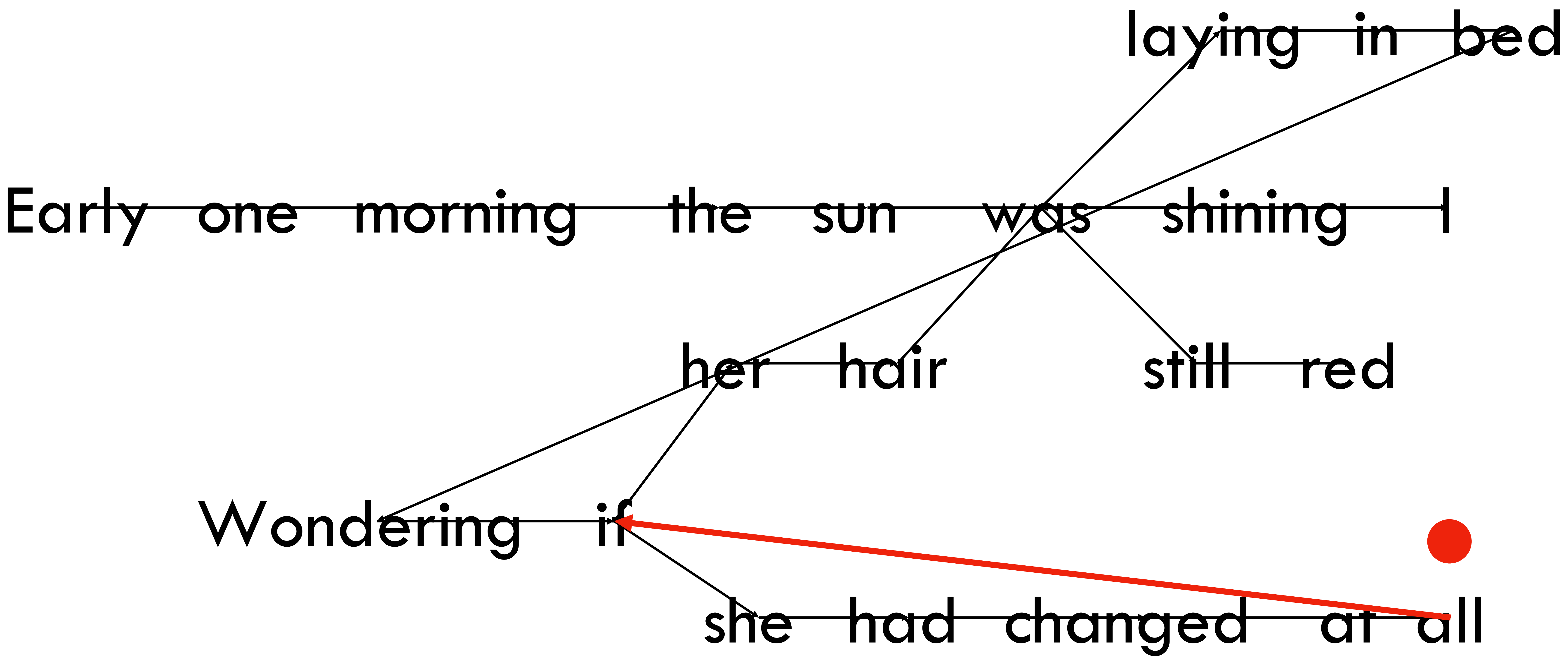
Early one morning the sun was shining I was laying in bed

Wondering if she had changed at



slide from Steve Seitz's video

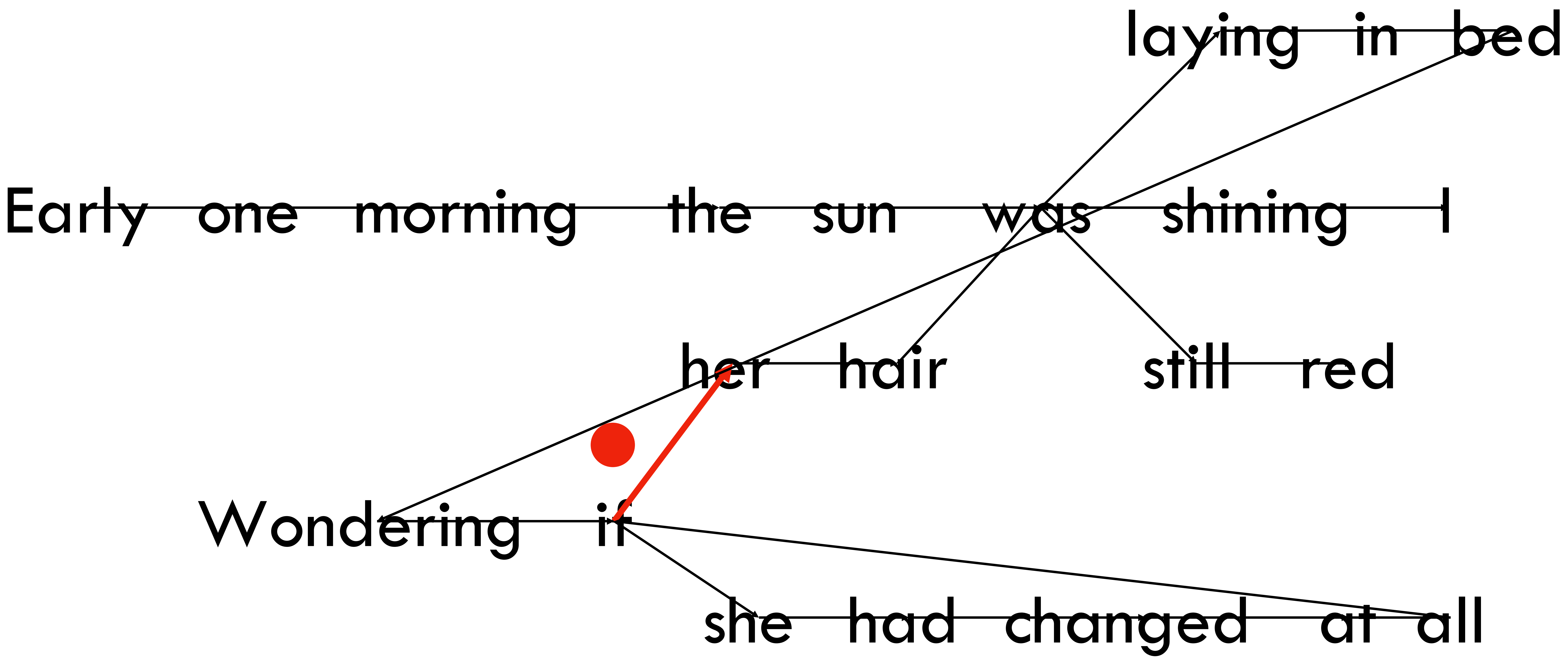
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all



slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

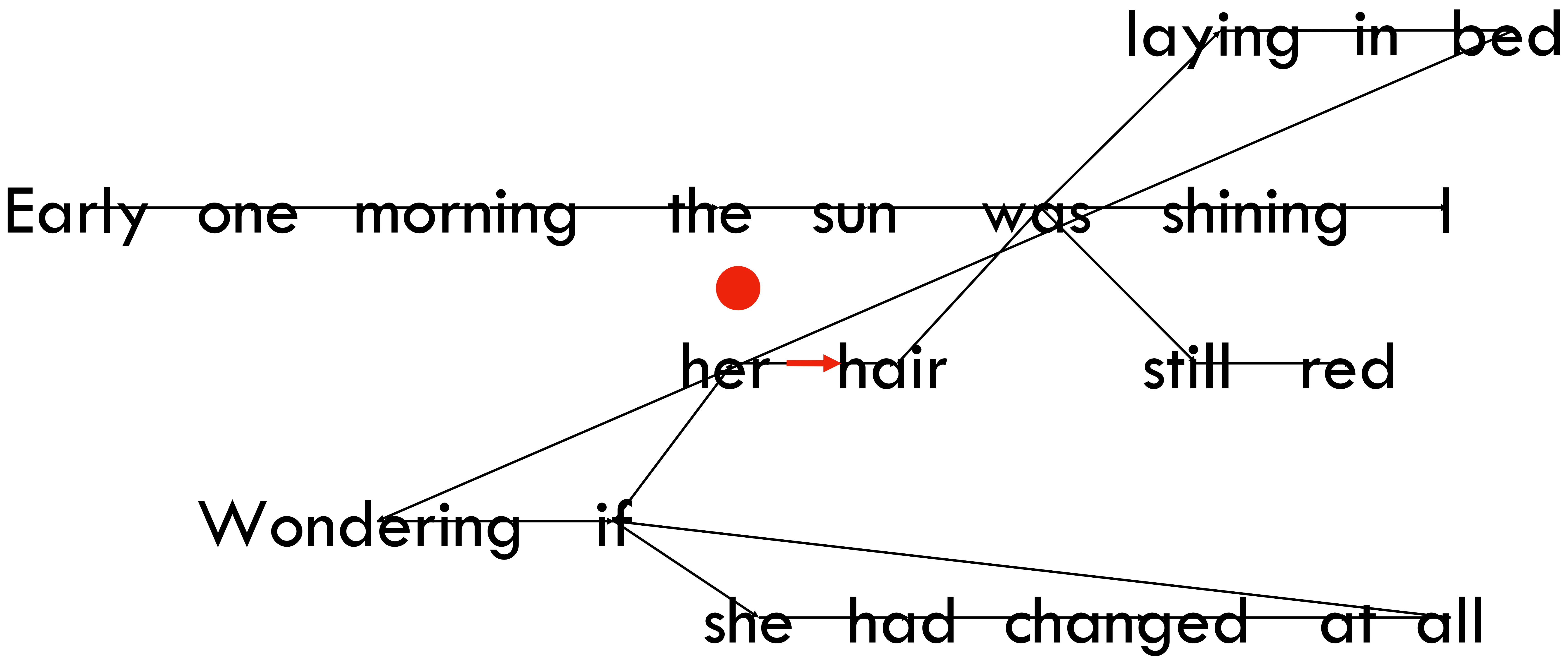
Wondering if she had changed at all if



slide from Steve Seitz's video

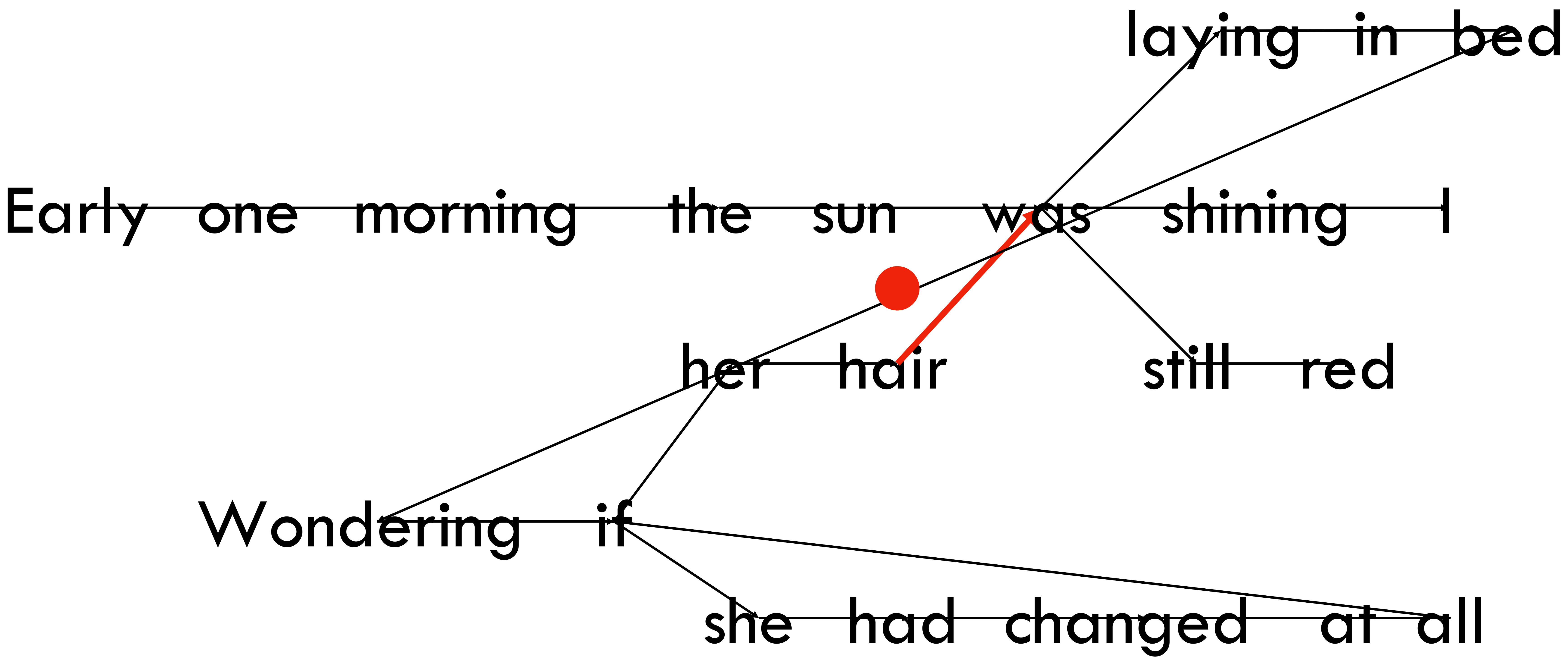
Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her



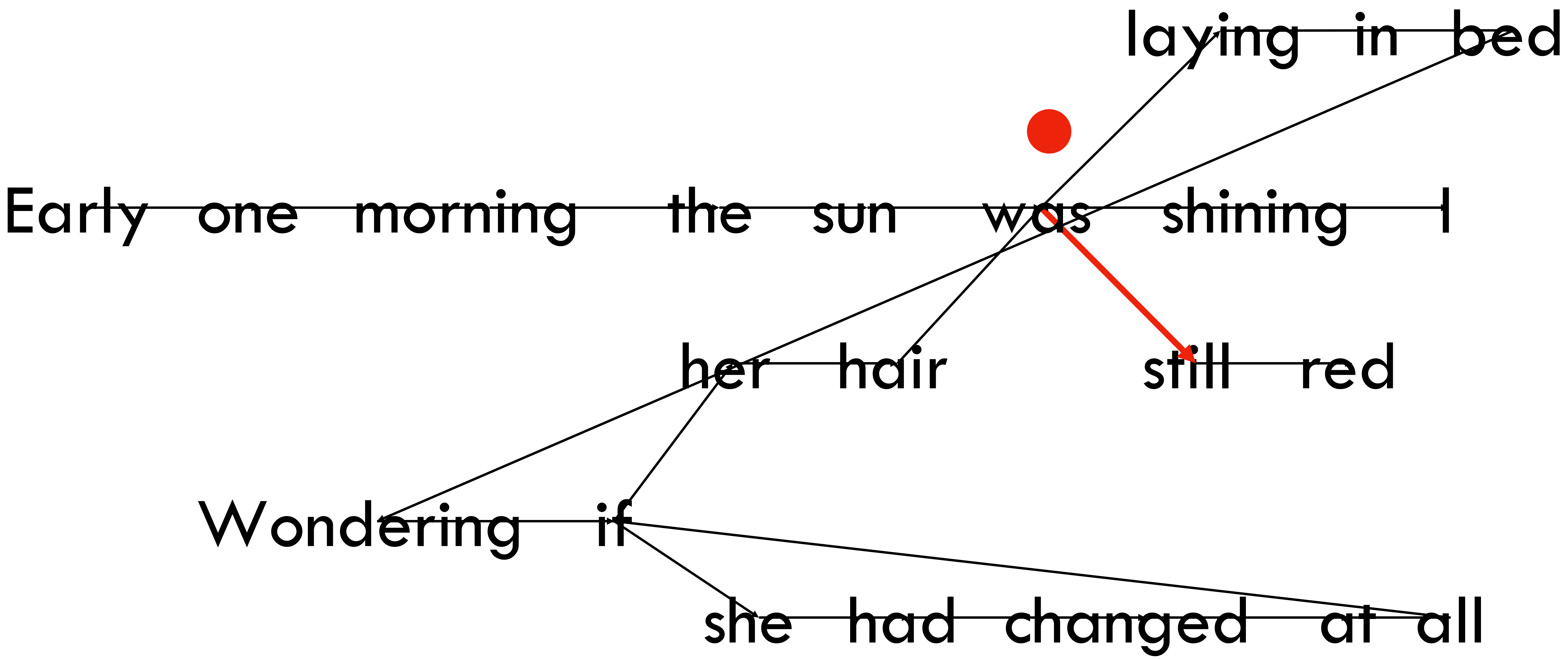
slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair



slide from Steve Seitz's [video]

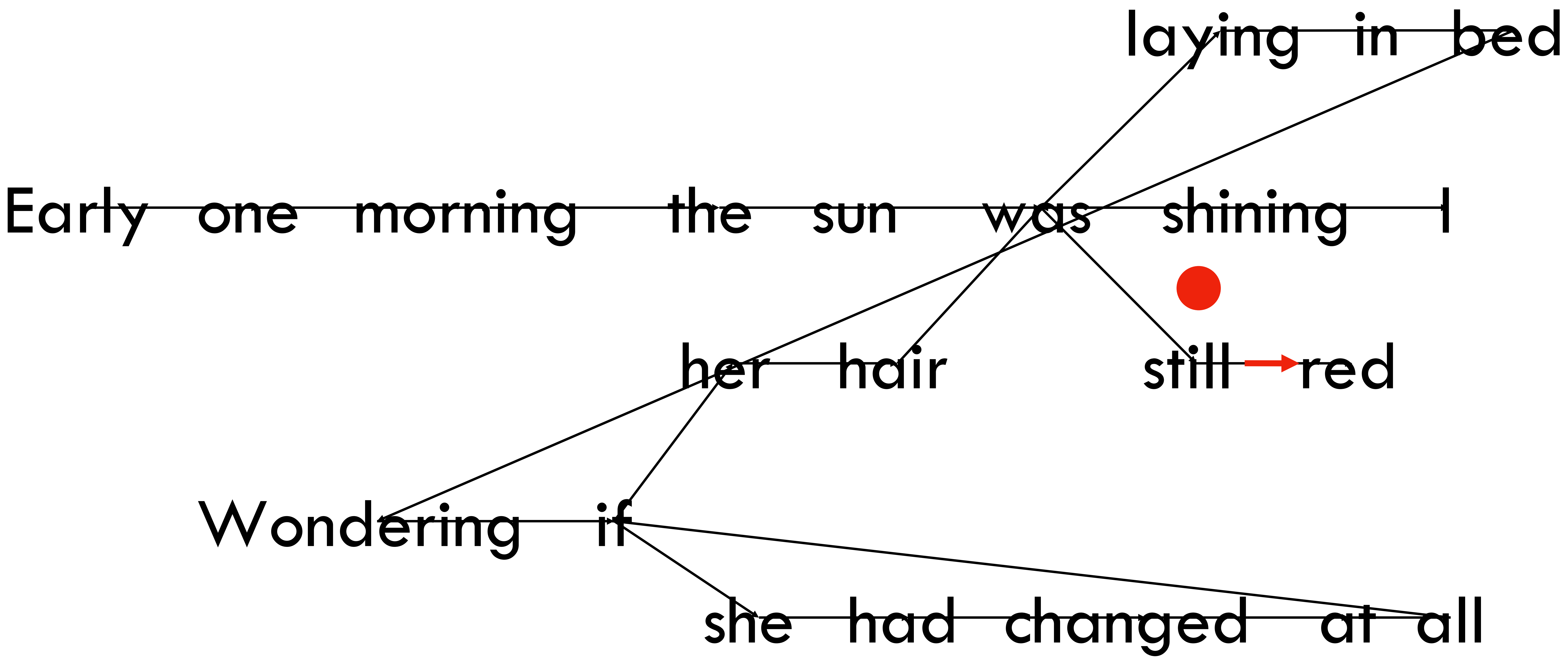
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was



slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

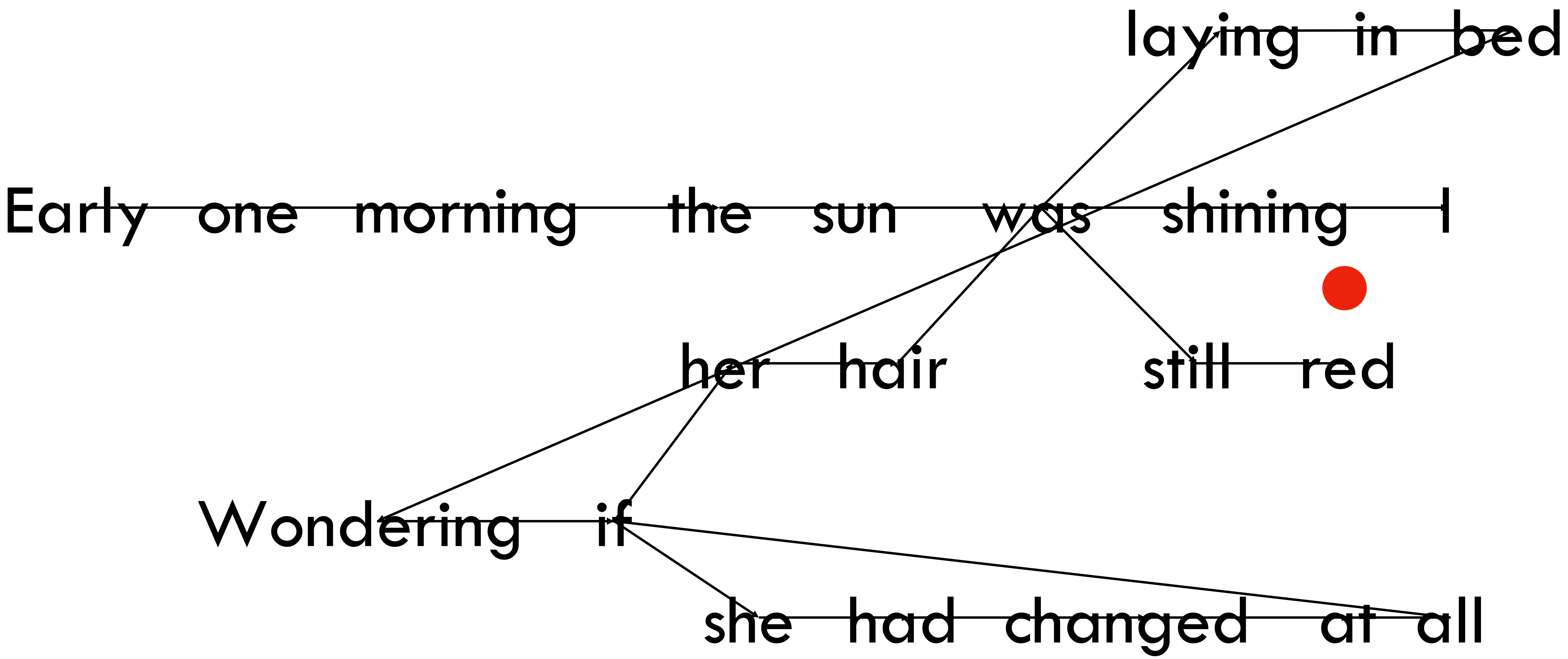
Wondering if she had changed at all if her hair was still



slide from Steve Seitz's video

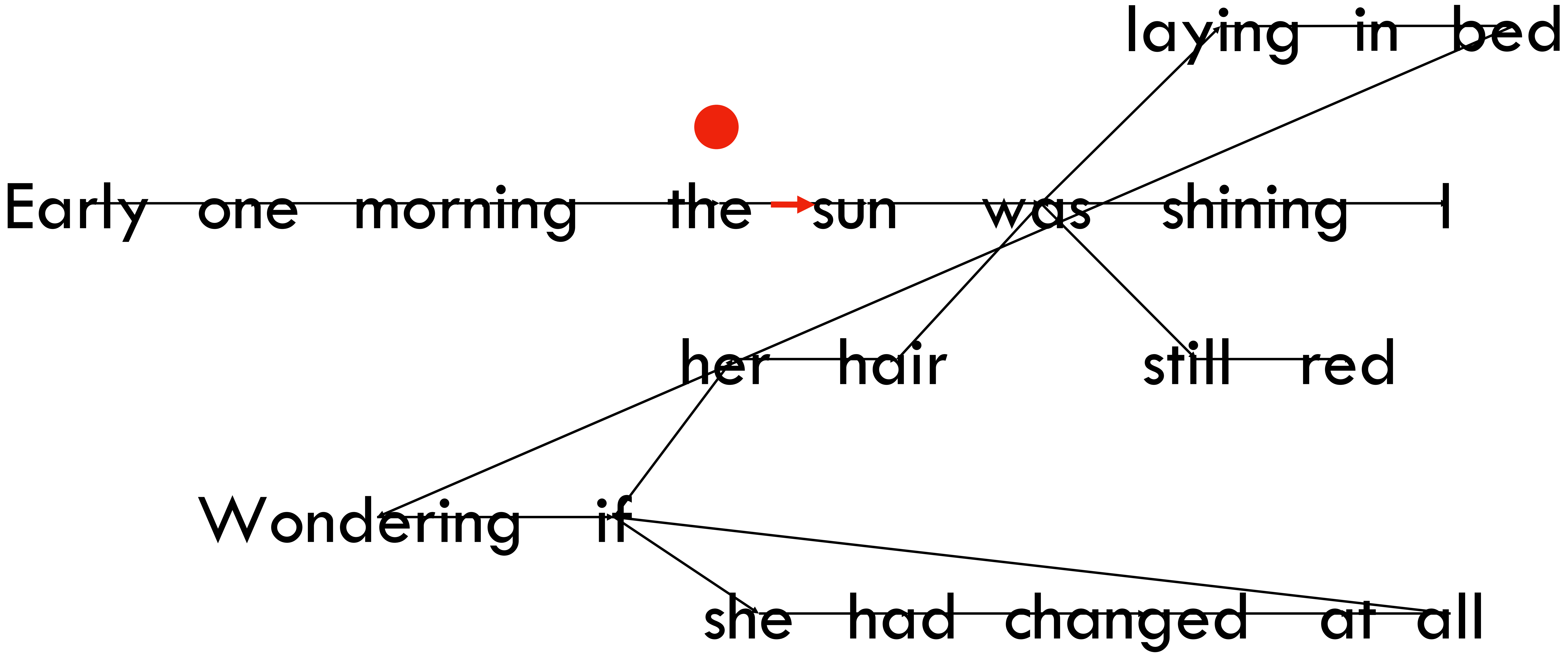
Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red



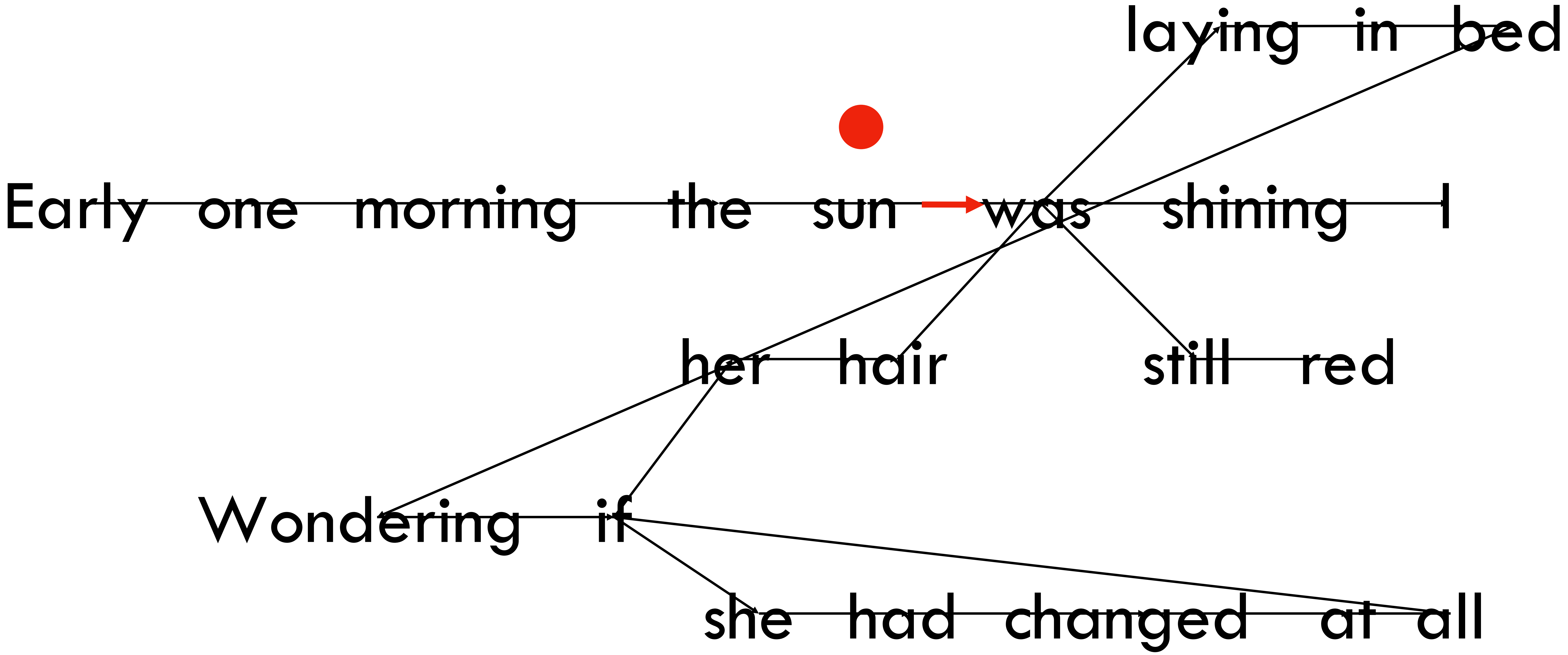
slide from Steve Seitz's [video]

the



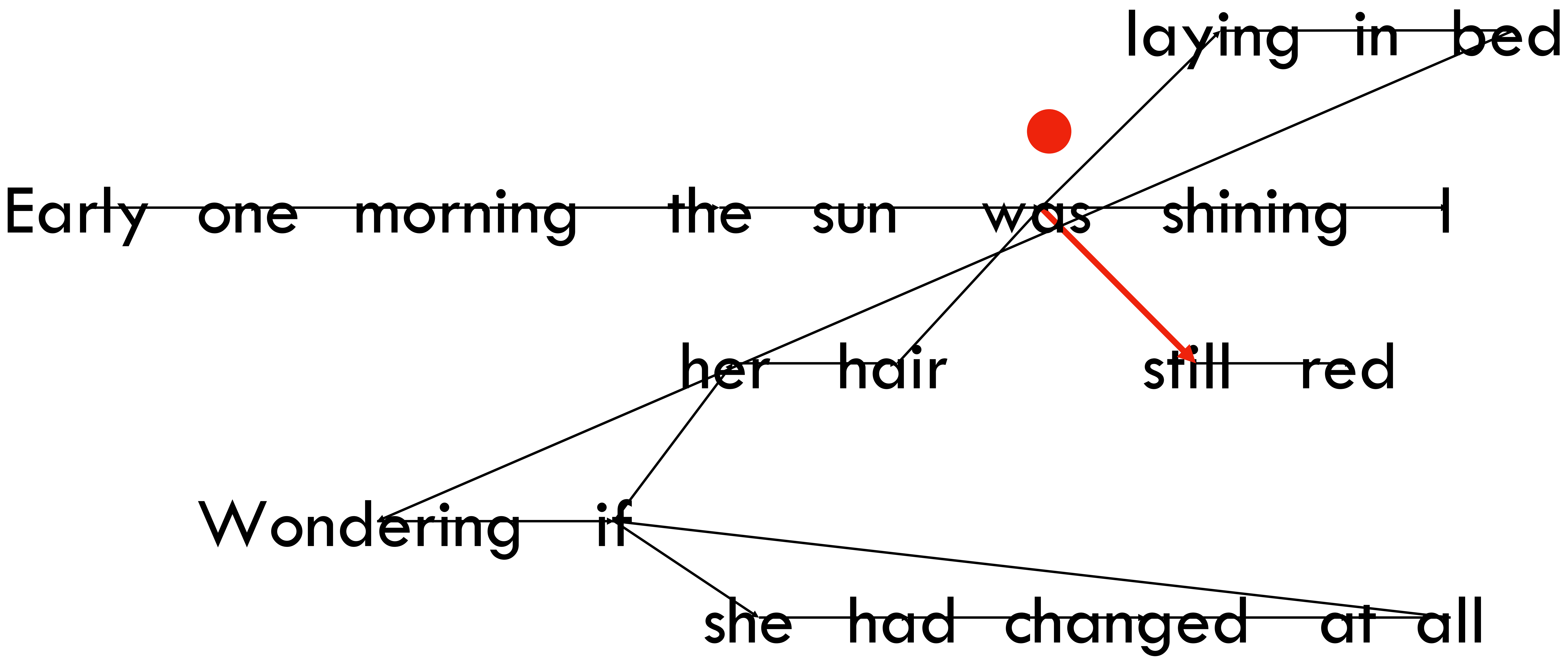
slide from Steve Seitz's [video]

the sun



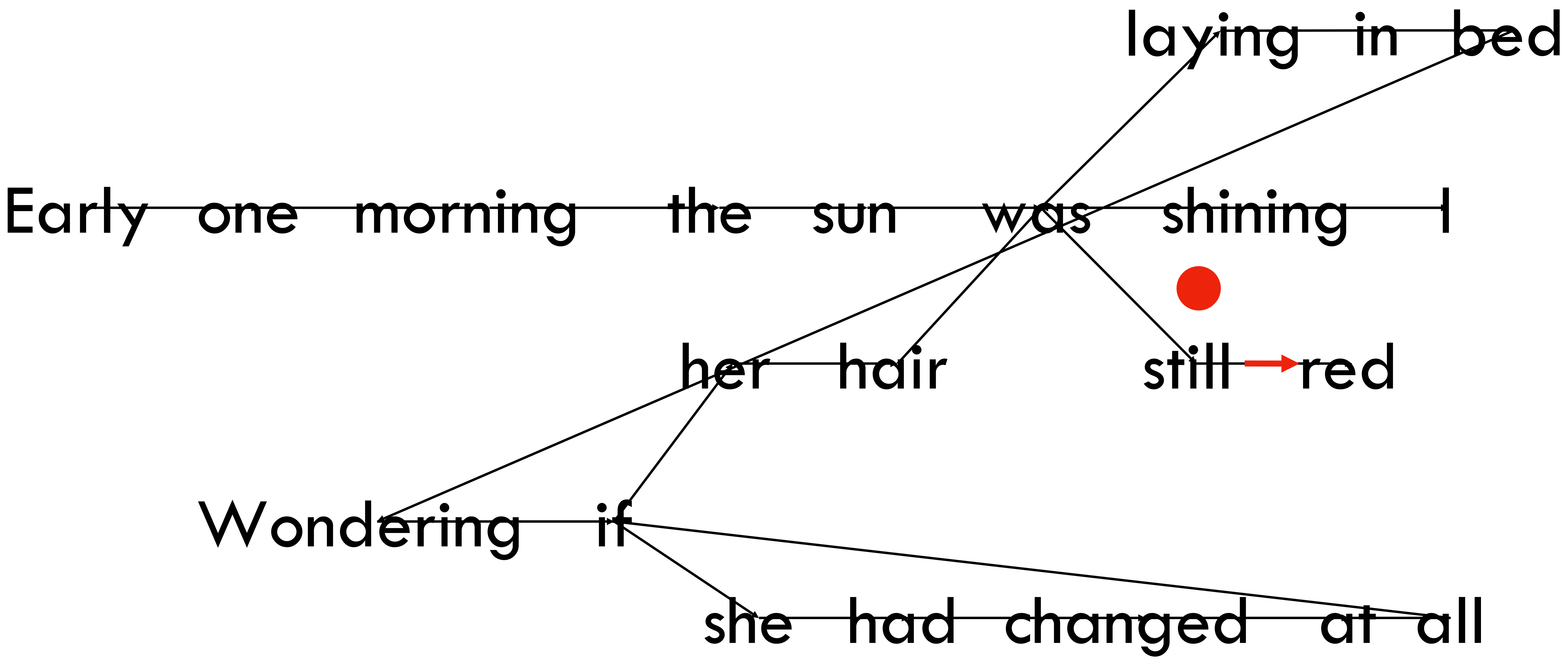
slide from Steve Seitz's [video]

the sun was



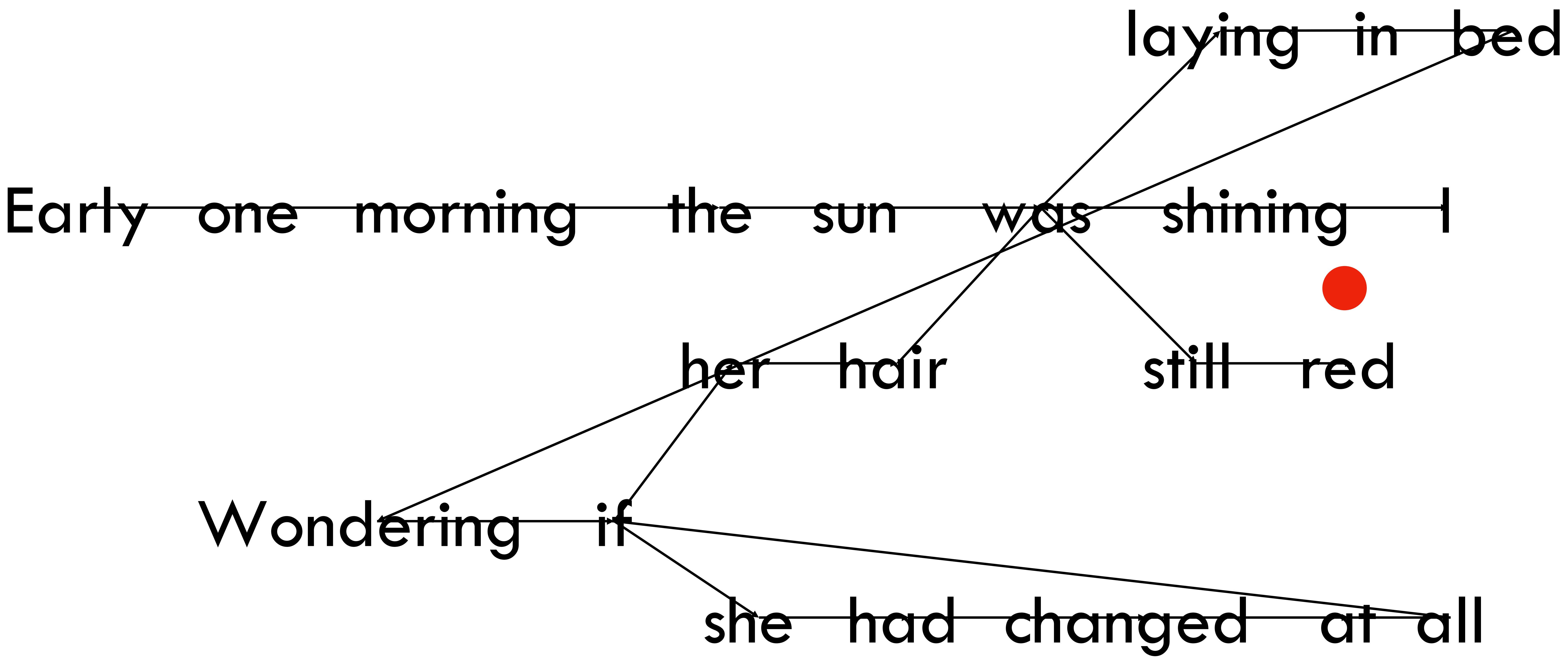
slide from Steve Seitz's [video]

the sun was still



slide from Steve Seitz's [video]

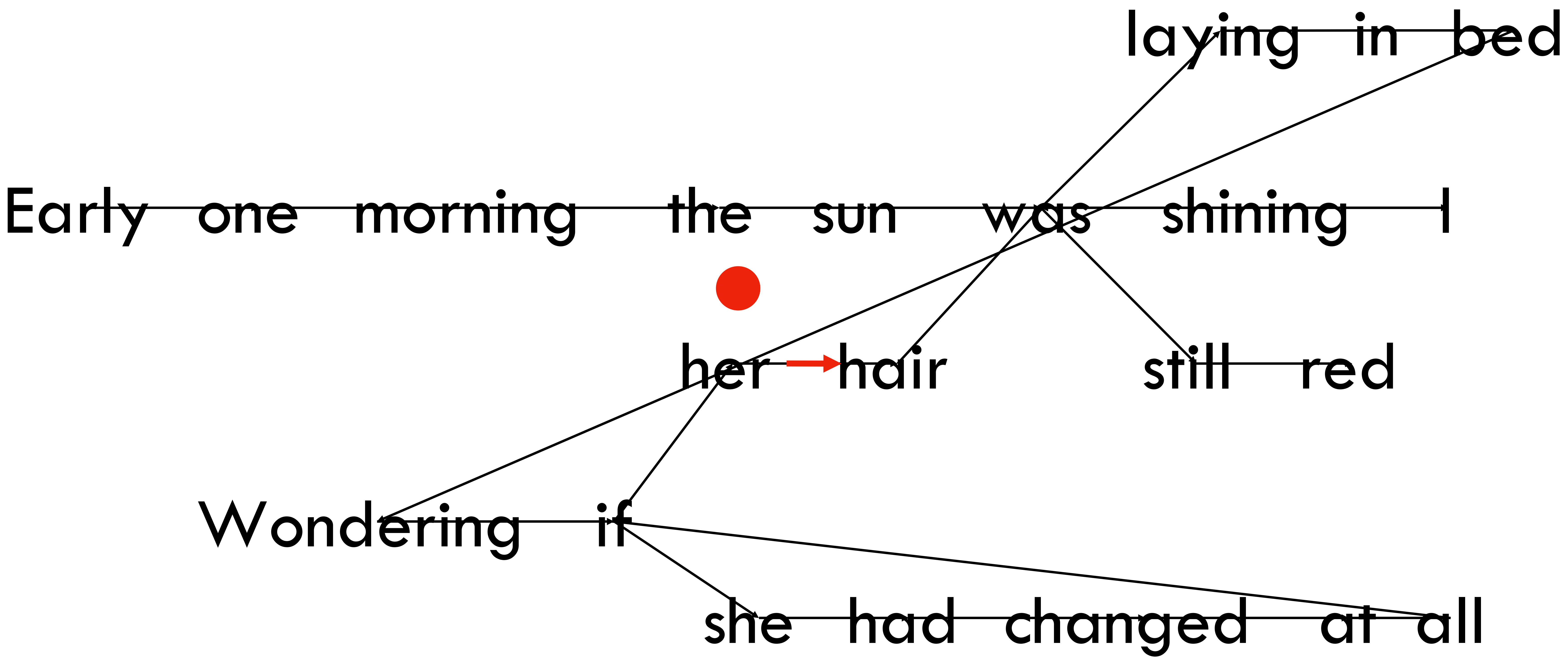
the sun was still red



slide from Steve Seitz's video

the sun was still red

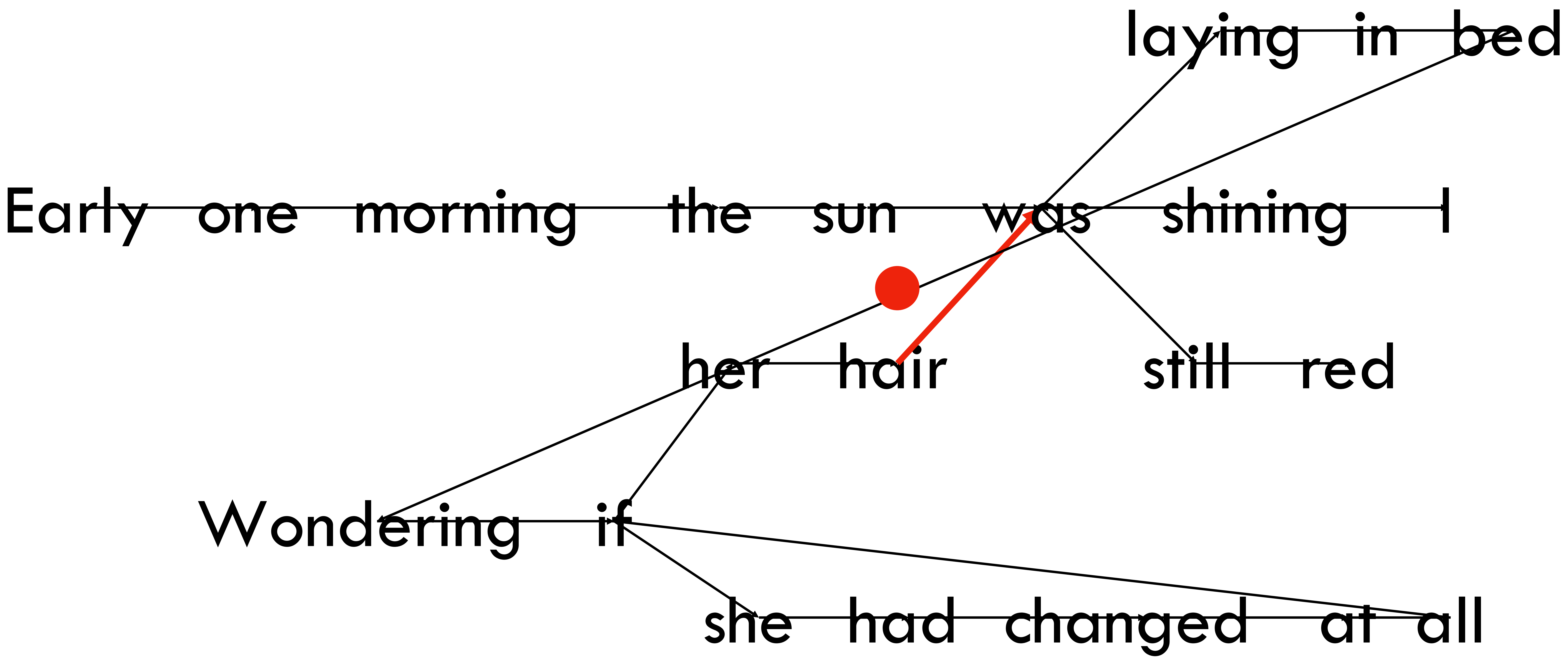
her



slide from Steve Seitz's video

the sun was still red

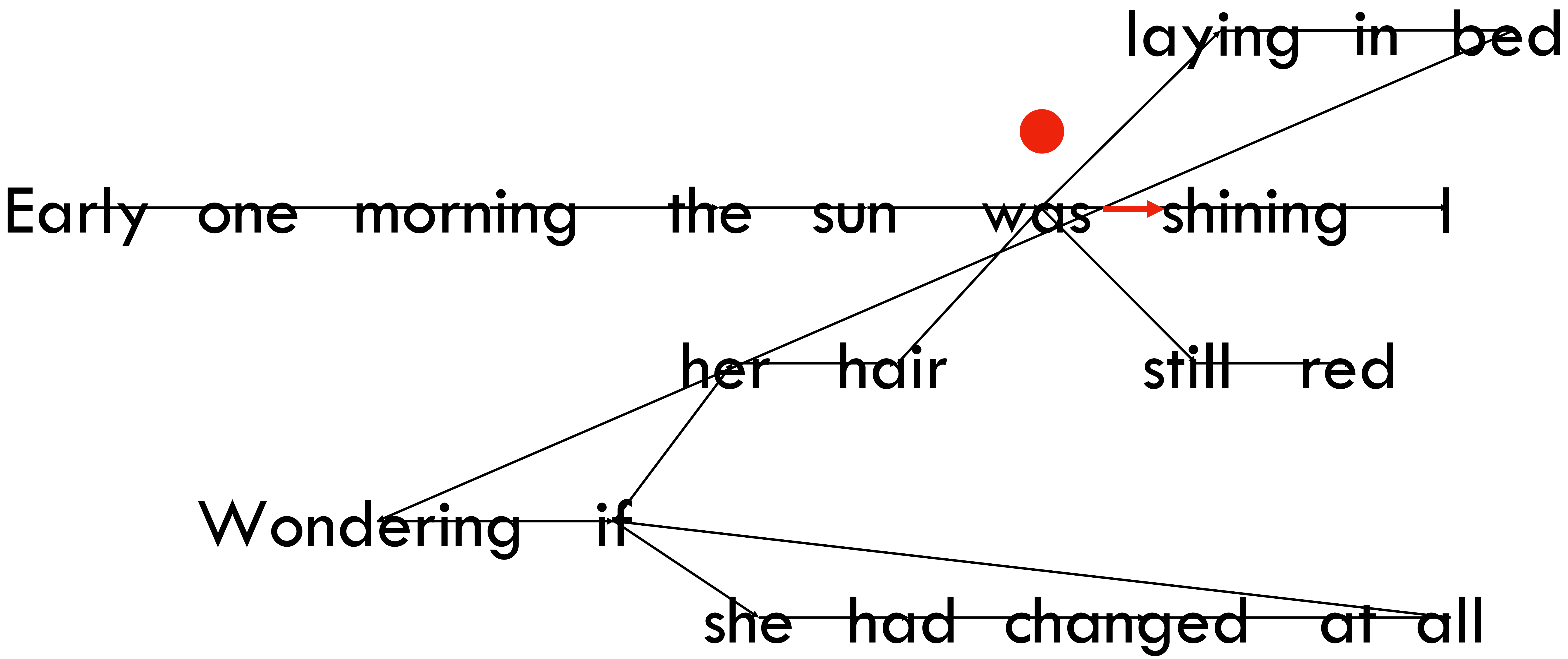
her hair



slide from Steve Seitz’s [video]

the sun was still red

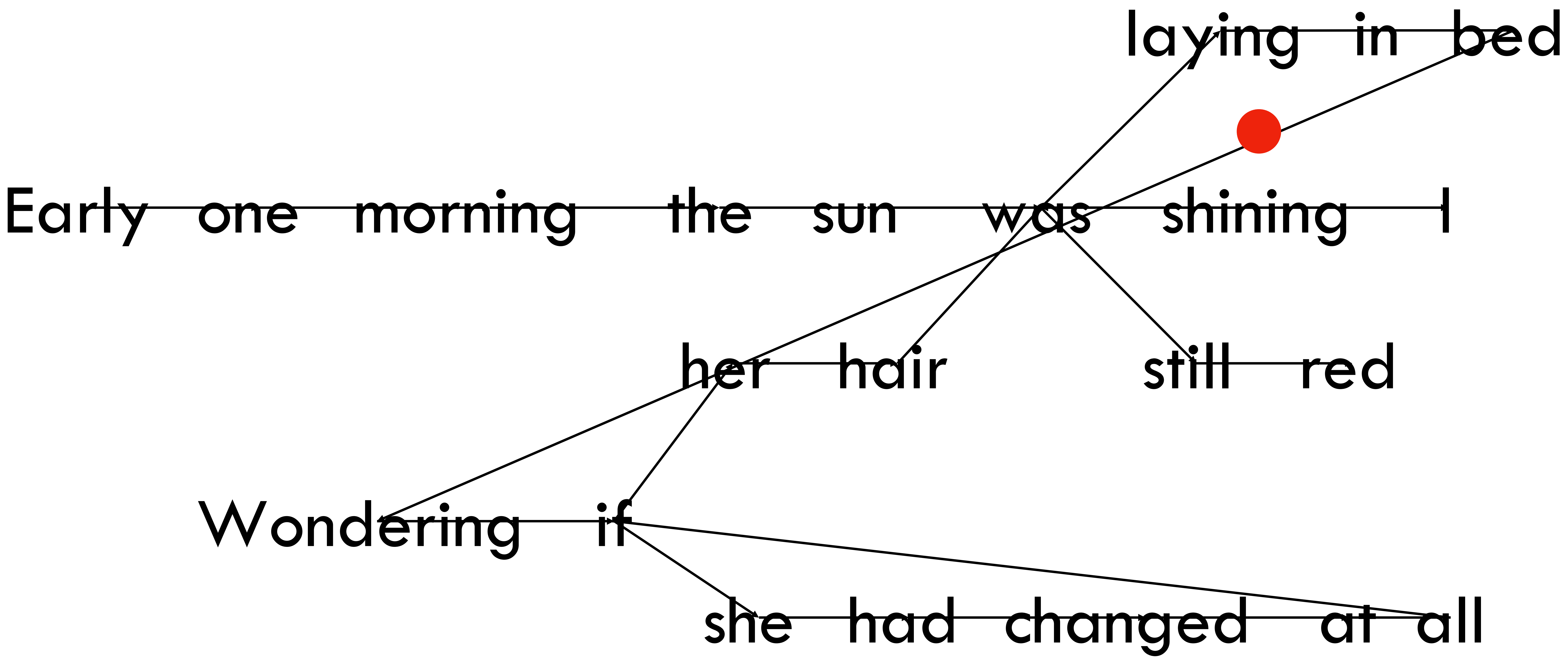
her hair was



slide from Steve Seitz's [video]

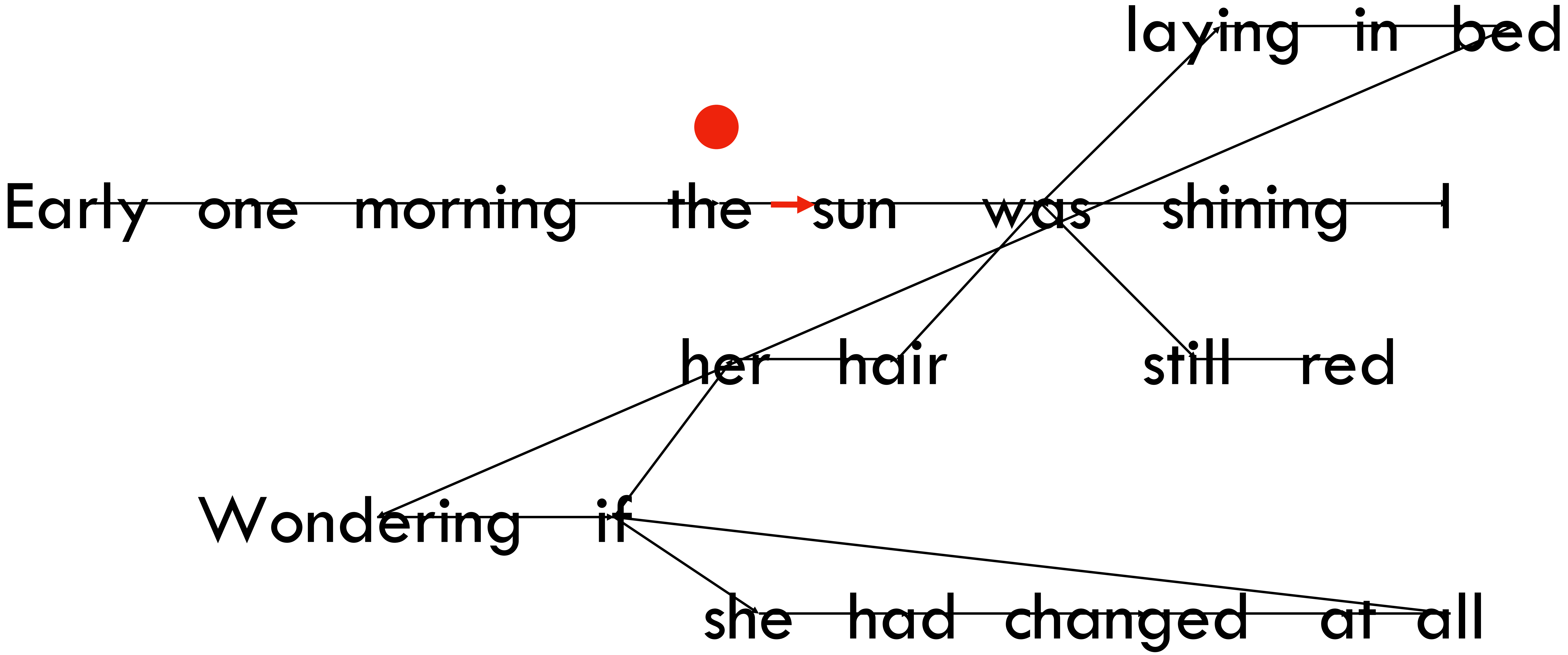
the sun was still red

her hair was shining



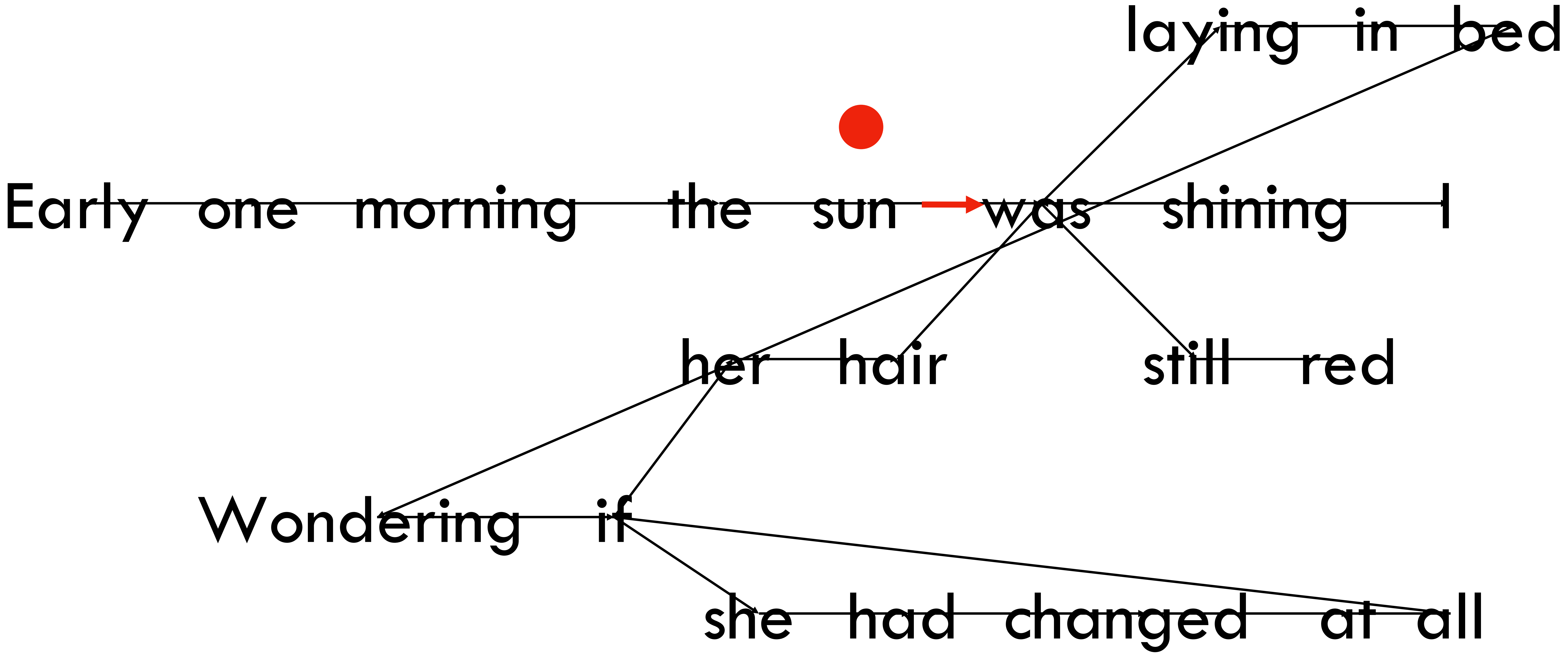
slide from Steve Seitz's [video]

the



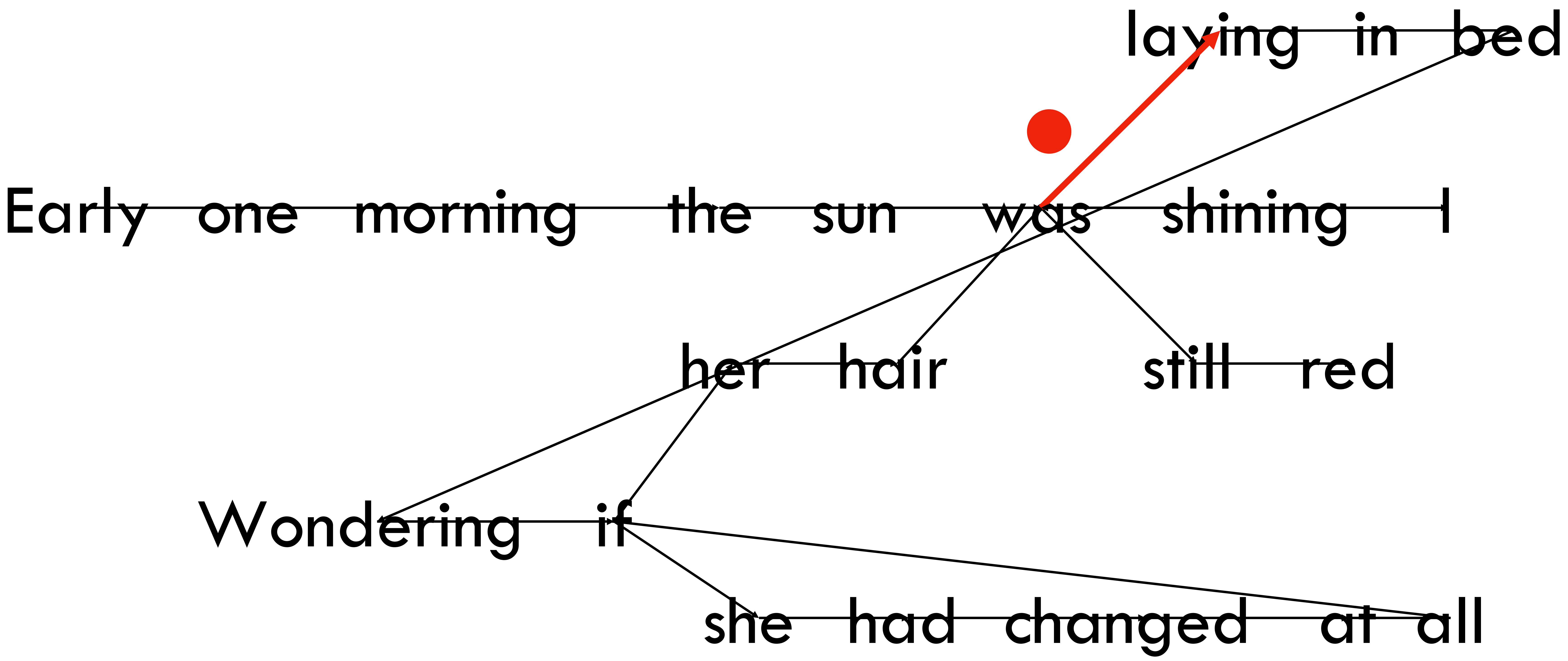
slide from Steve Seitz's video

the sun



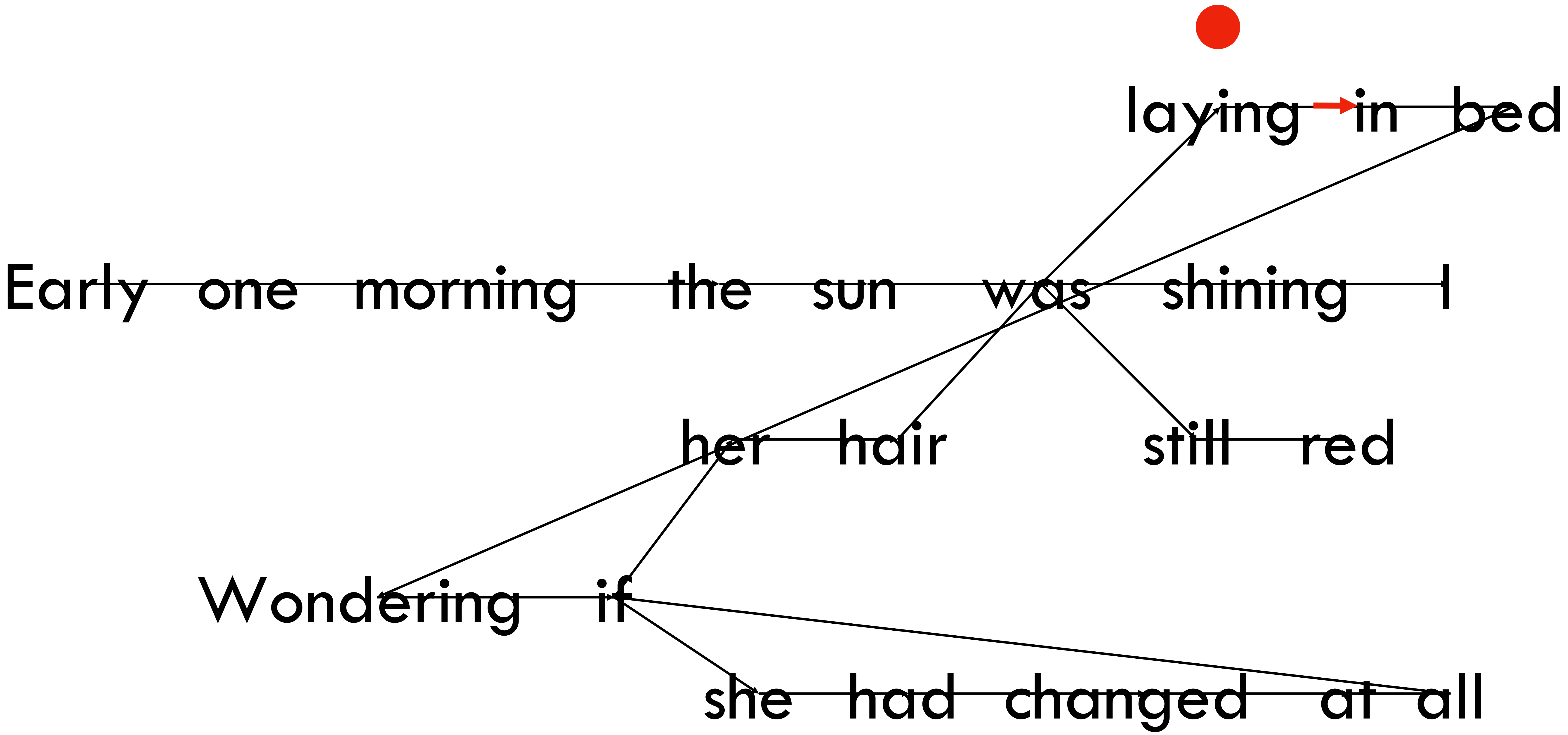
slide from Steve Seitz's [video]

the sun was



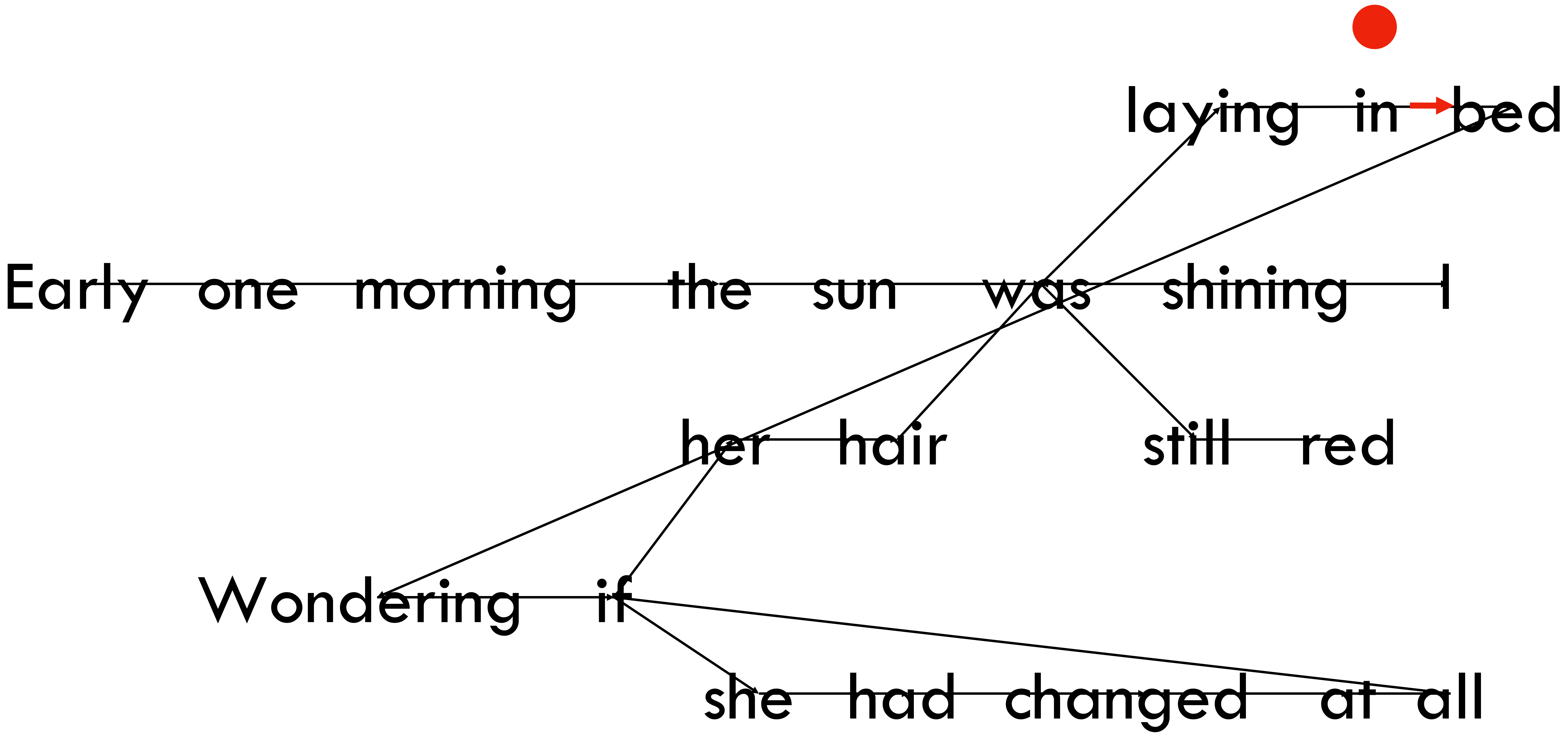
slide from Steve Seitz's [video]

the sun was laying



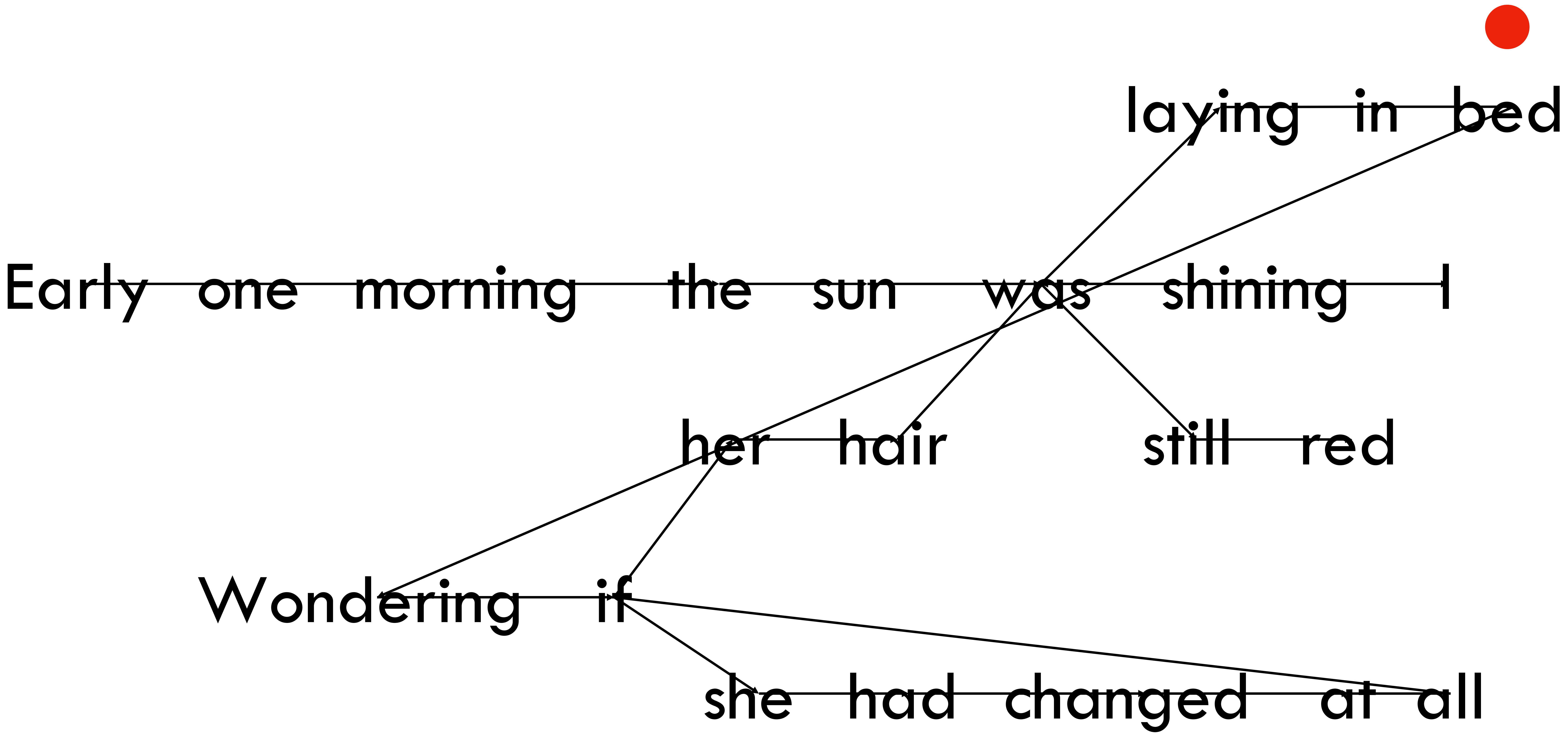
slide from Steve Seitz's video

the sun was laying in



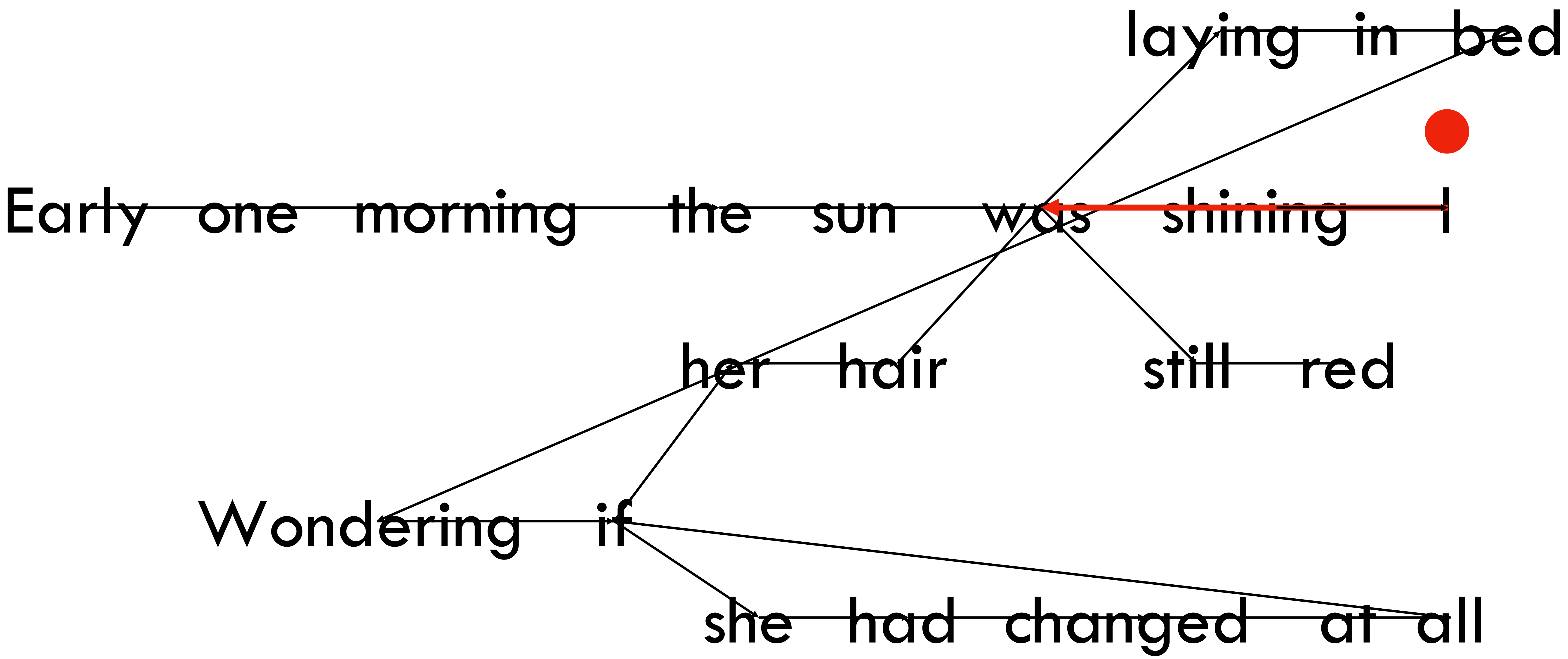
slide from Steve Seitz's video

the sun was laying in bed



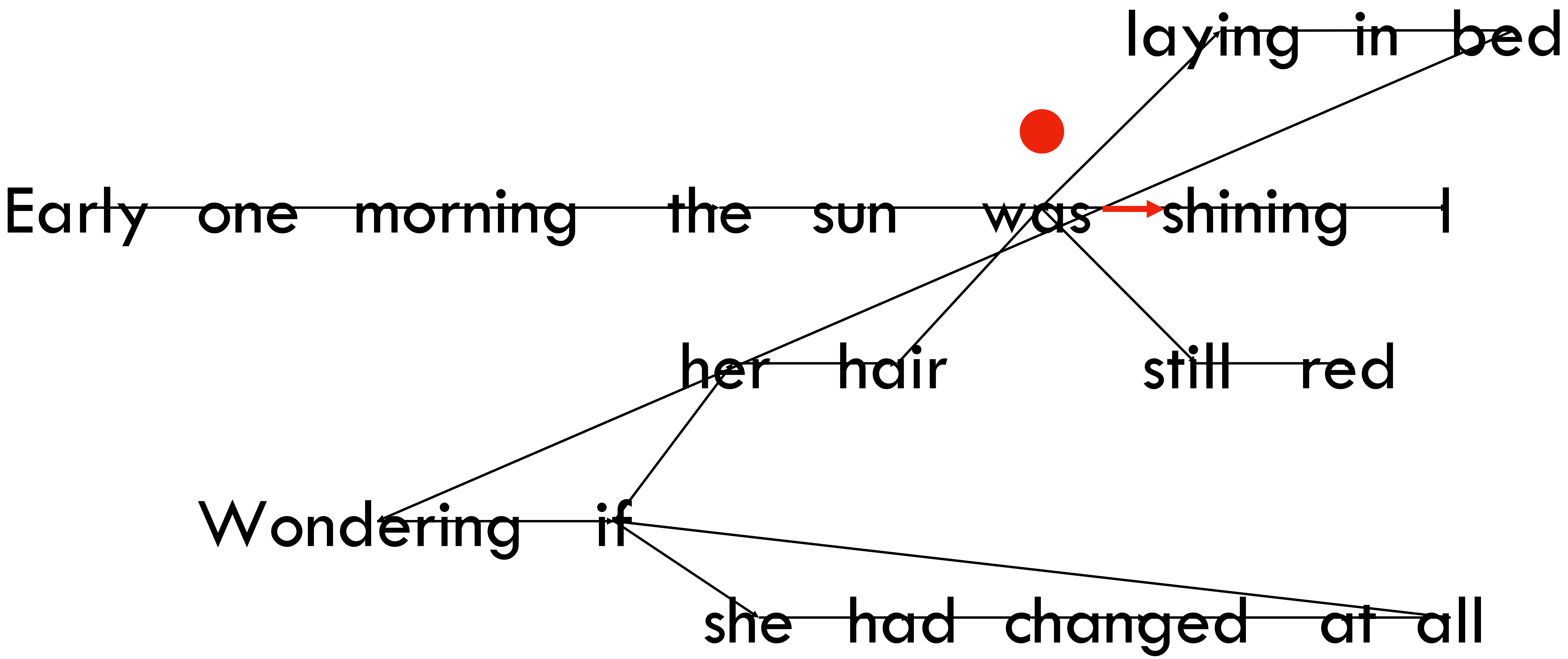
slide from Steve Seitz's video

I



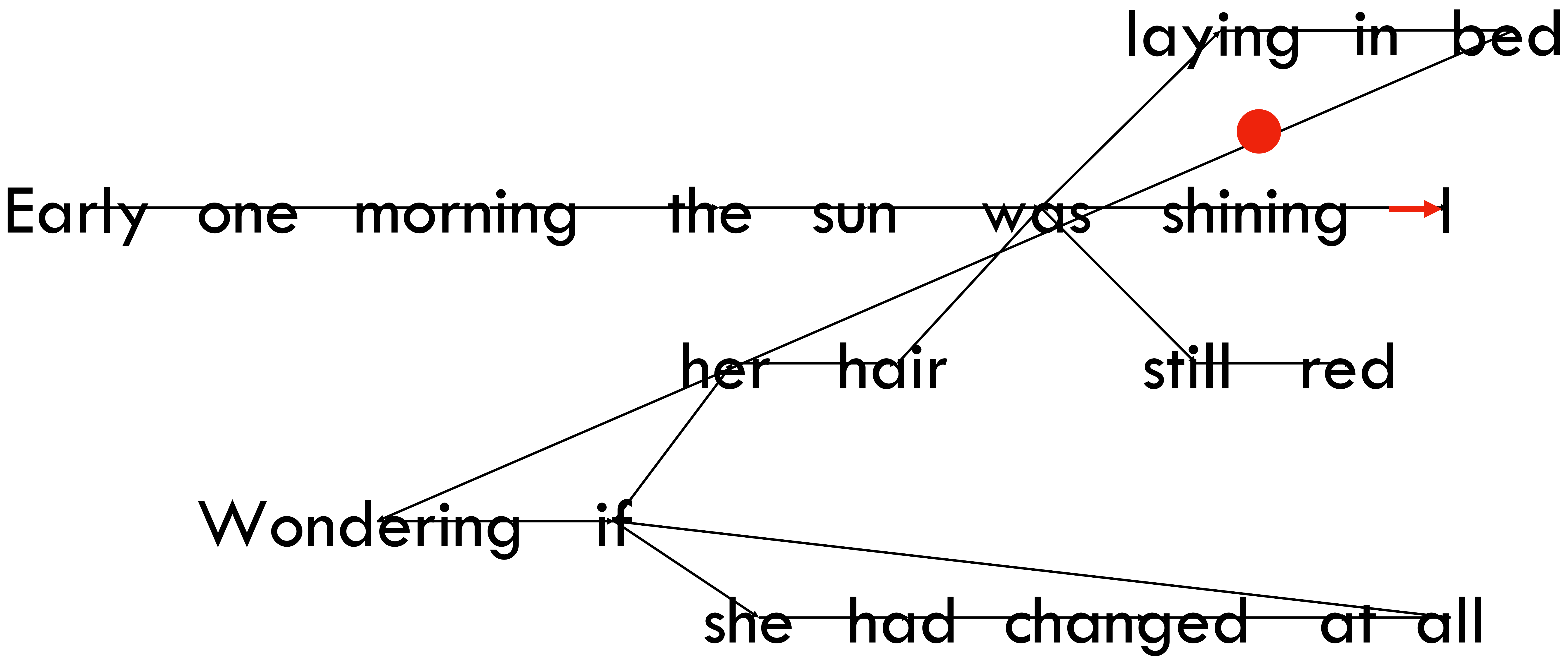
slide from Steve Seitz's [video]

I was



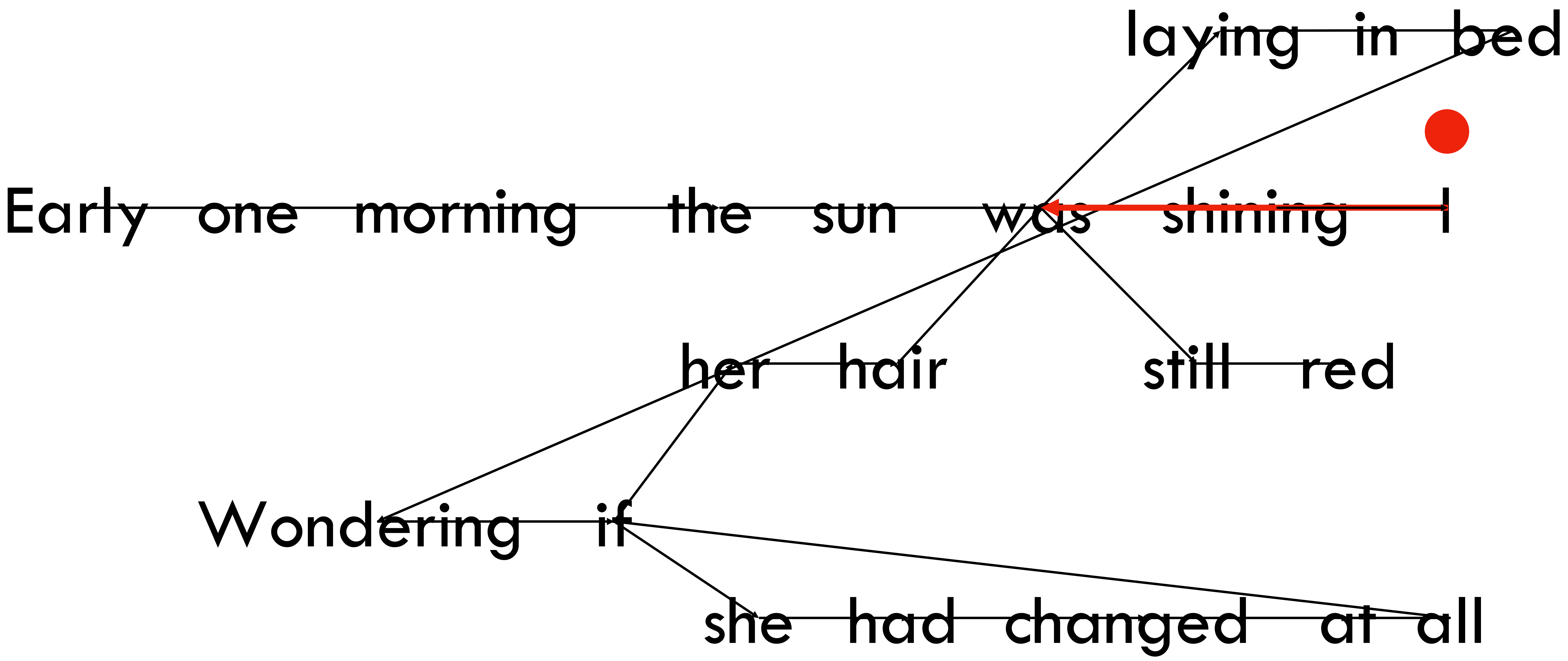
slide from Steve Seitz's [video]

I was shining



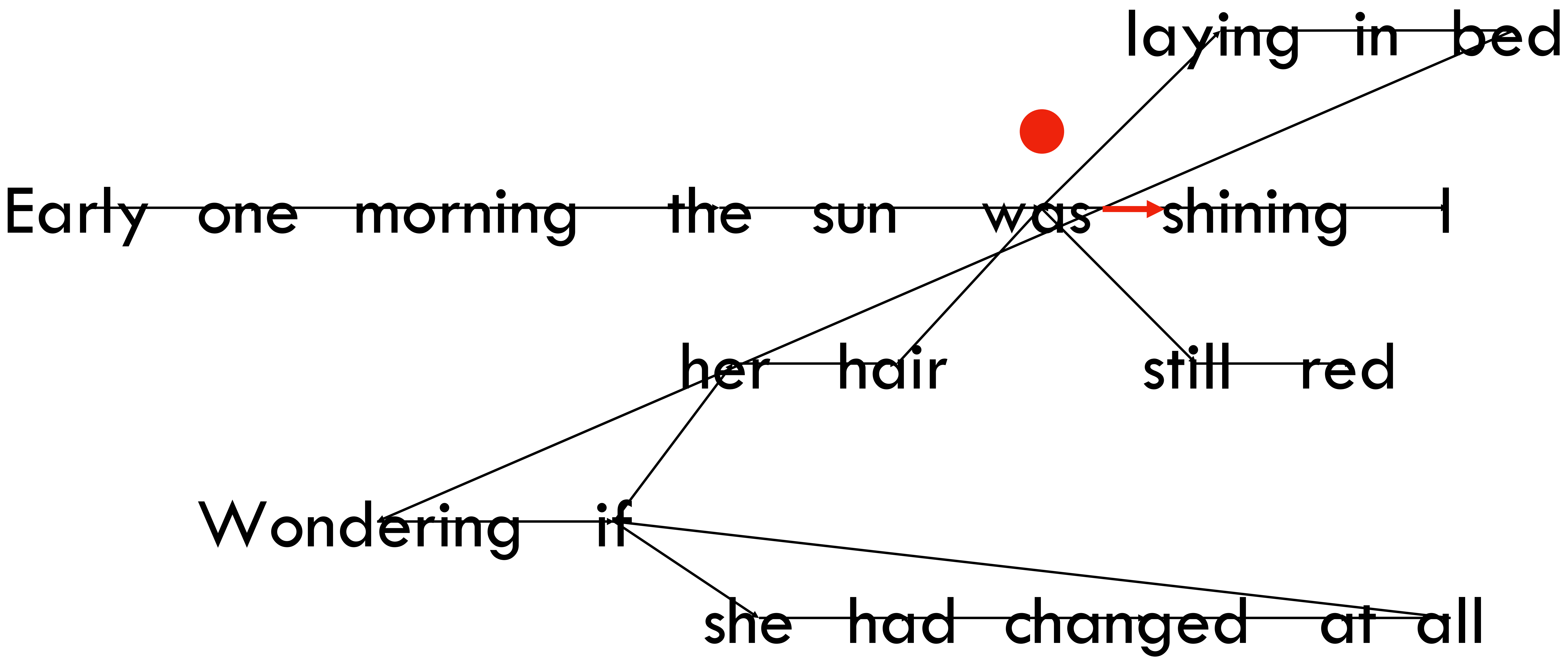
slide from Steve Seitz's [video]

I was shining I



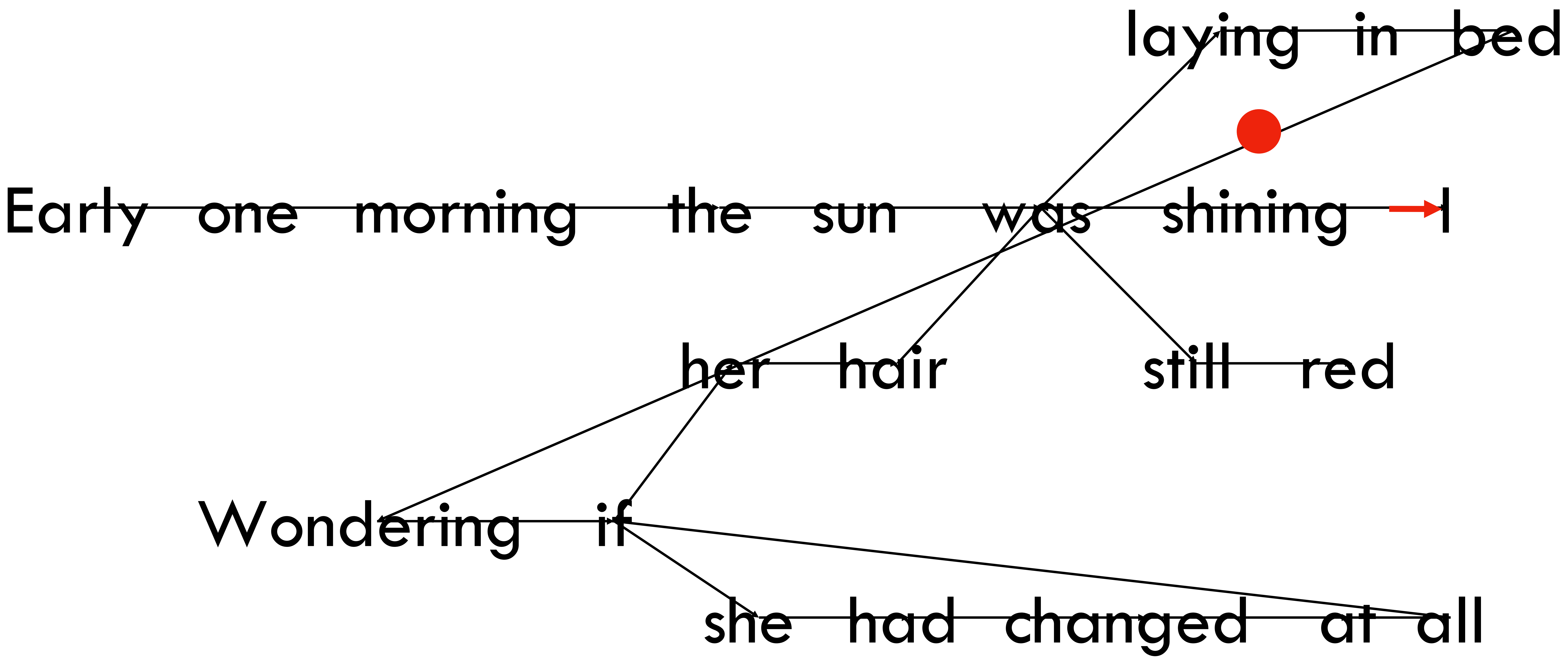
slide from Steve Seitz's [video]

I was shining I was



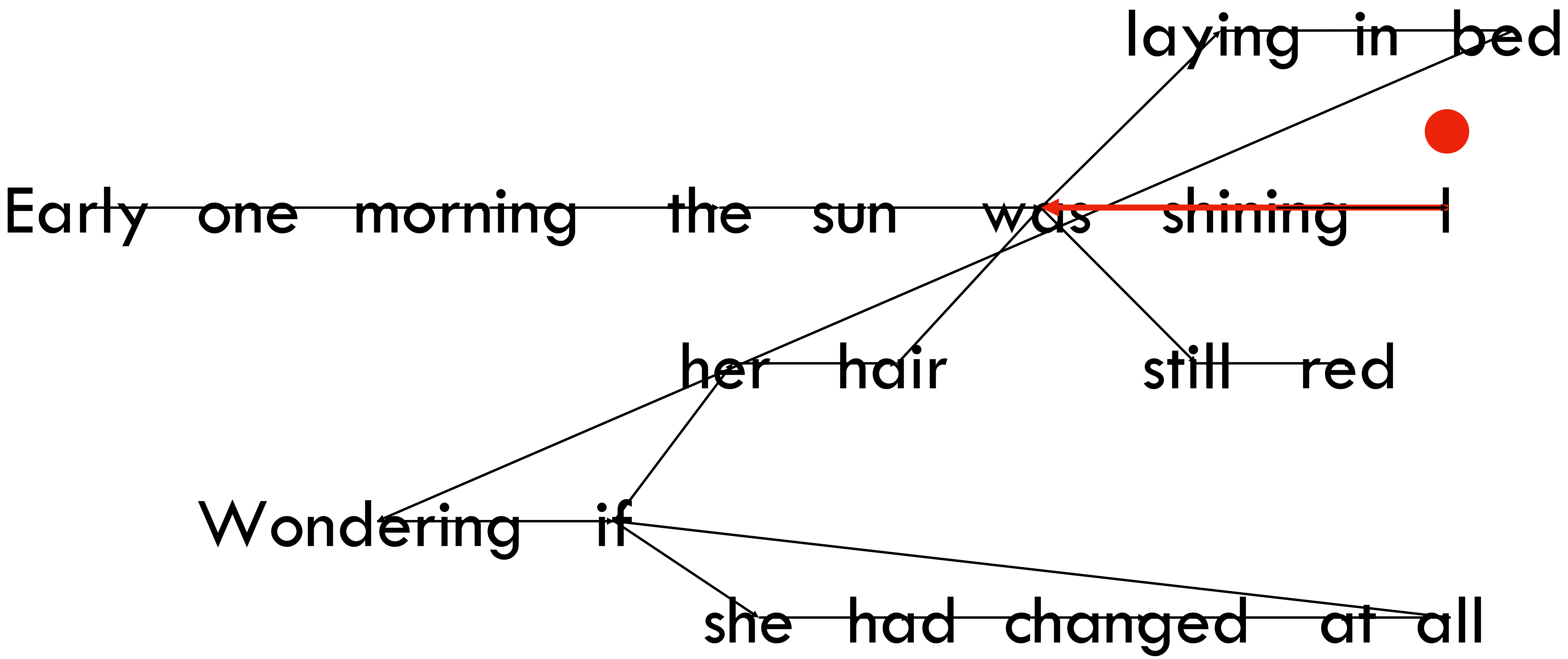
slide from Steve Seitz's [video]

I was shining I was shining



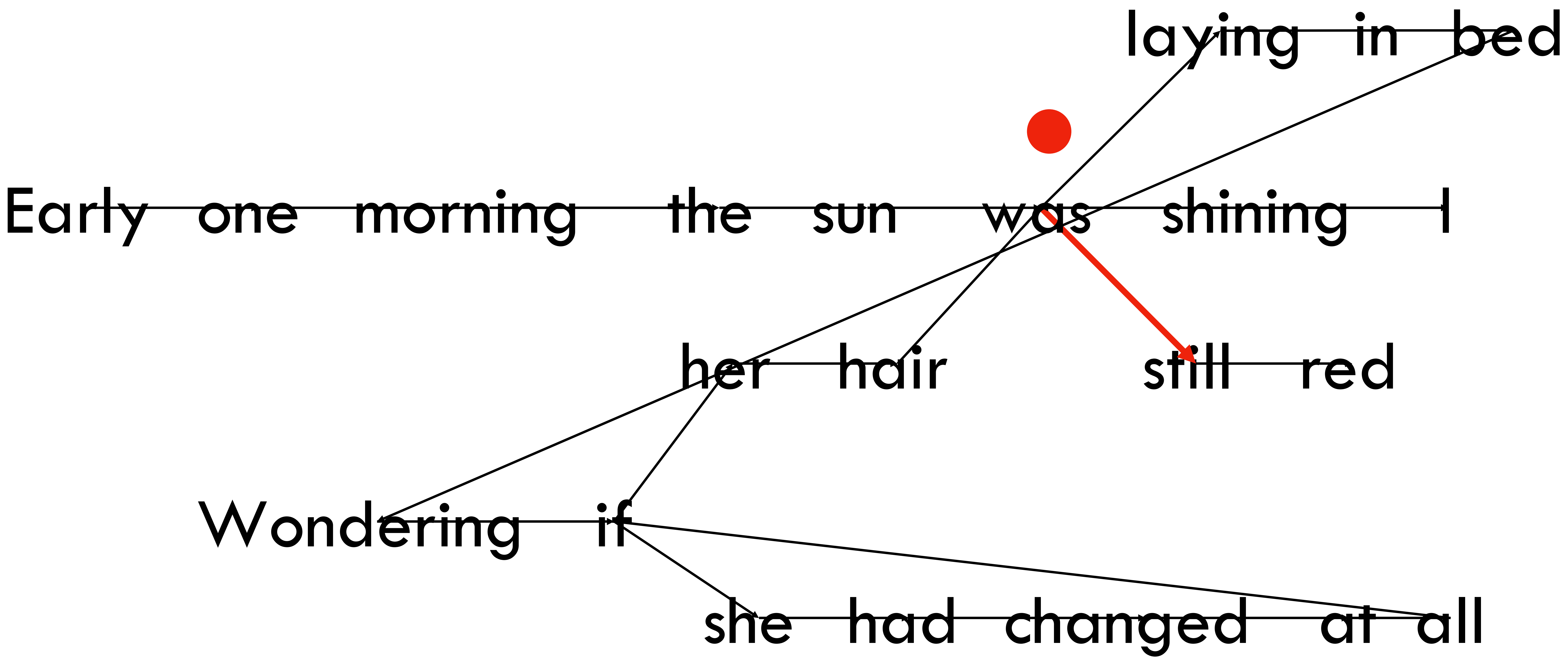
slide from Steve Seitz's [video]

I was shining I was shining I



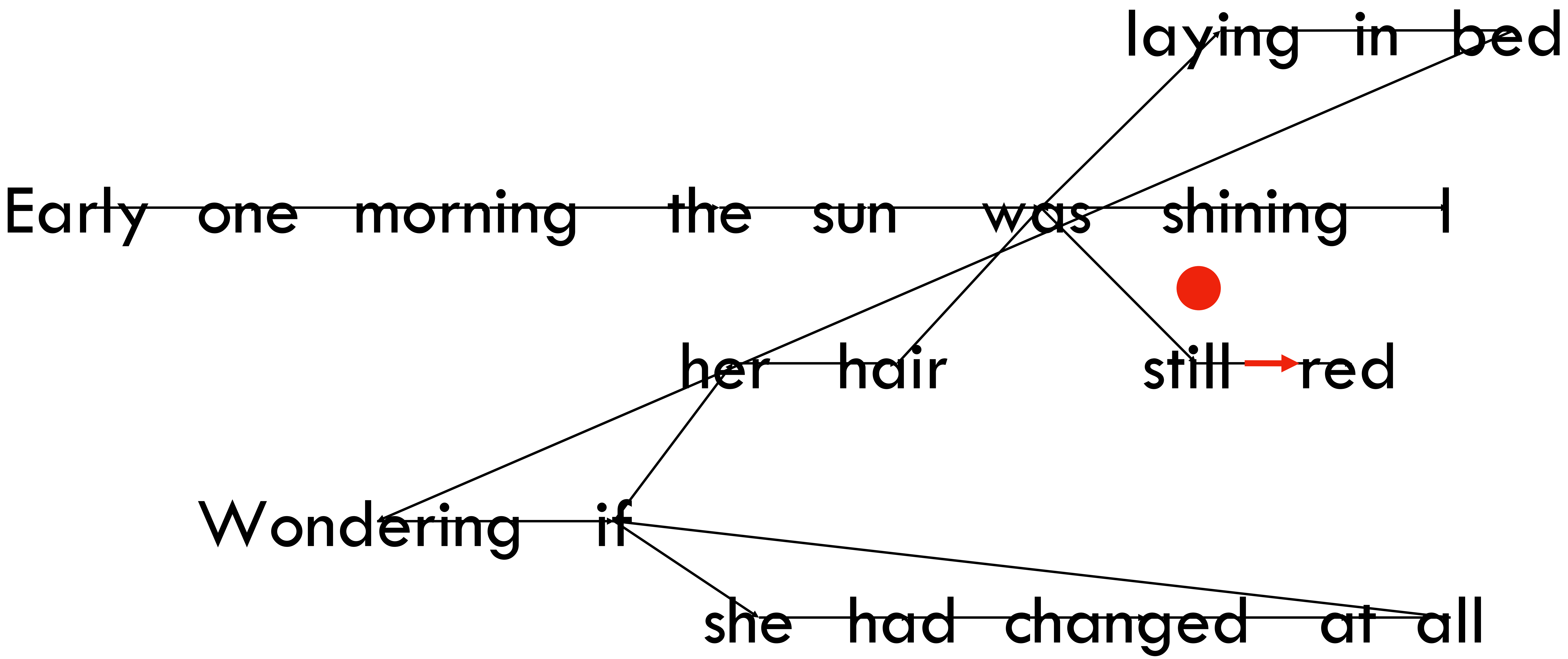
slide from Steve Seitz's video

I was shining I was shining I was



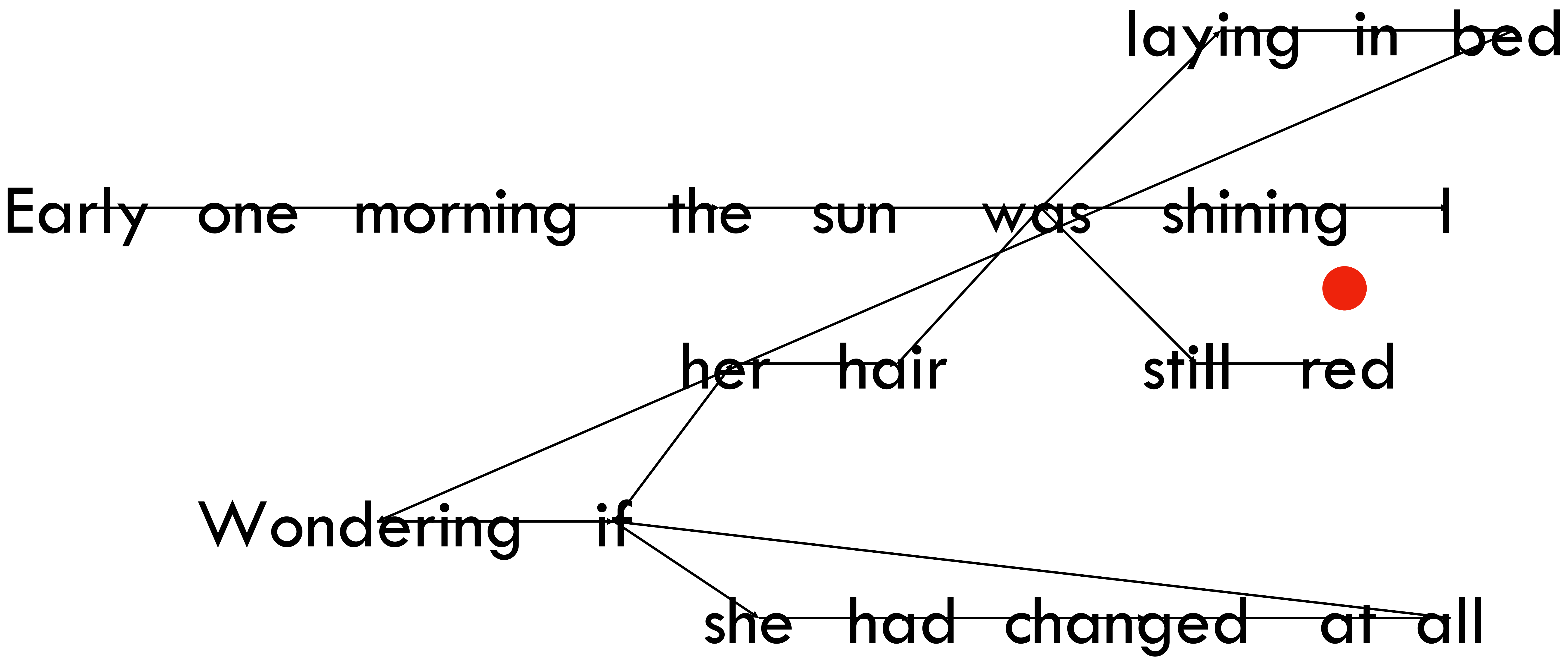
slide from Steve Seitz's video

I was shining I was shining I was still

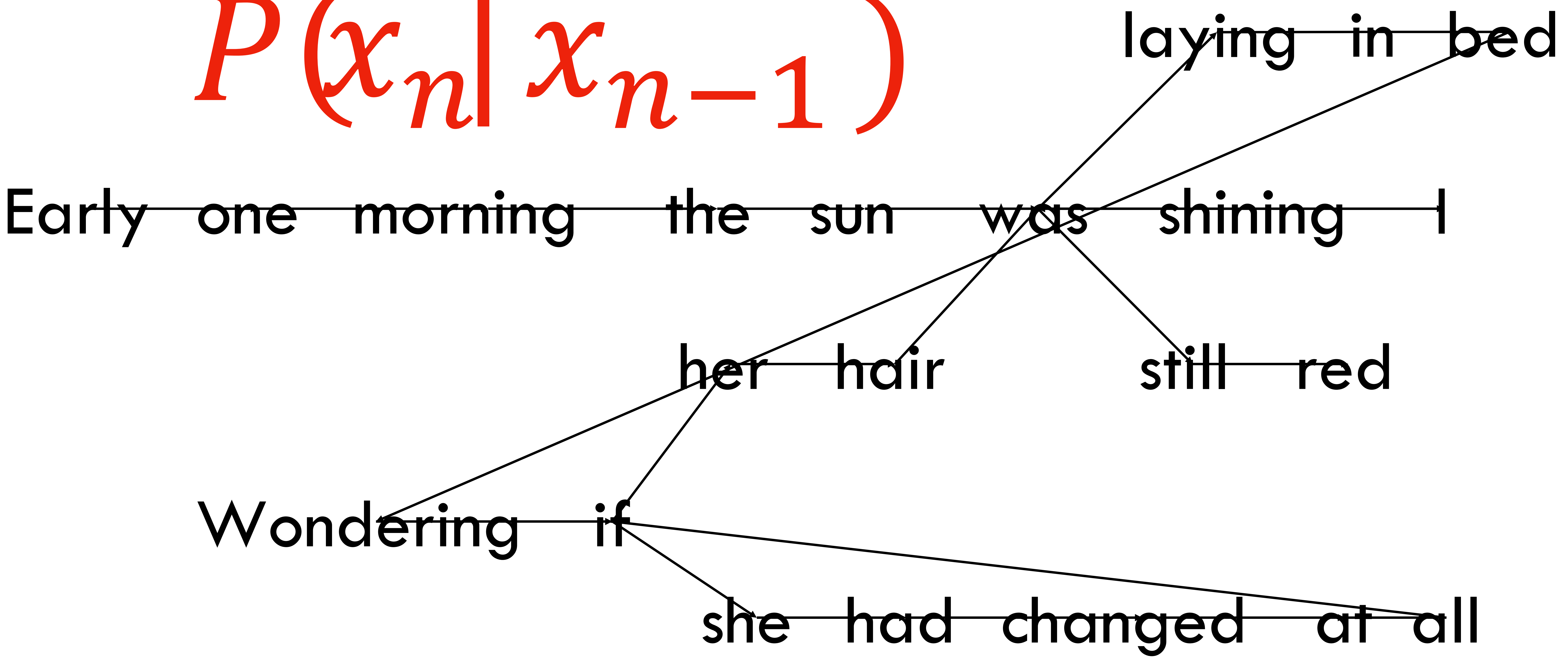


slide from Steve Seitz's [video]

I was shining I was shining I was still red



$$P(x_n \mid x_{n-1})$$


laying in bed
Early one morning the sun was shining I
her hair still red
Wondering if
she had changed at all

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning

trigrams

slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning

one morning the

trigrams

slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun

trigrams

slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun
the sun was

trigrams

slide from Steve Seitz's video

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun
the sun was

trigrams

slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

Early one morning
one morning the
morning the sun
the sun was
sun was shining
was shining I
shining I was
I was laying
...

trigrams

slide from Steve Seitz's video

$$P(x_n \mid x_{n-1}, x_{n-2})$$

Early one	→ morning
one morning	→ the
morning the	→ sun
the sun	→ was
sun was	→ shining
was shining	→ I
shining I	→ was
I was	→ laying

...

Video Textures

Arno Schödl

Richard Szeliski

David Salesin

Irfan Essa

Microsoft Research, Georgia Tech

SIGGRAPH 2000

Still photos




Video clips



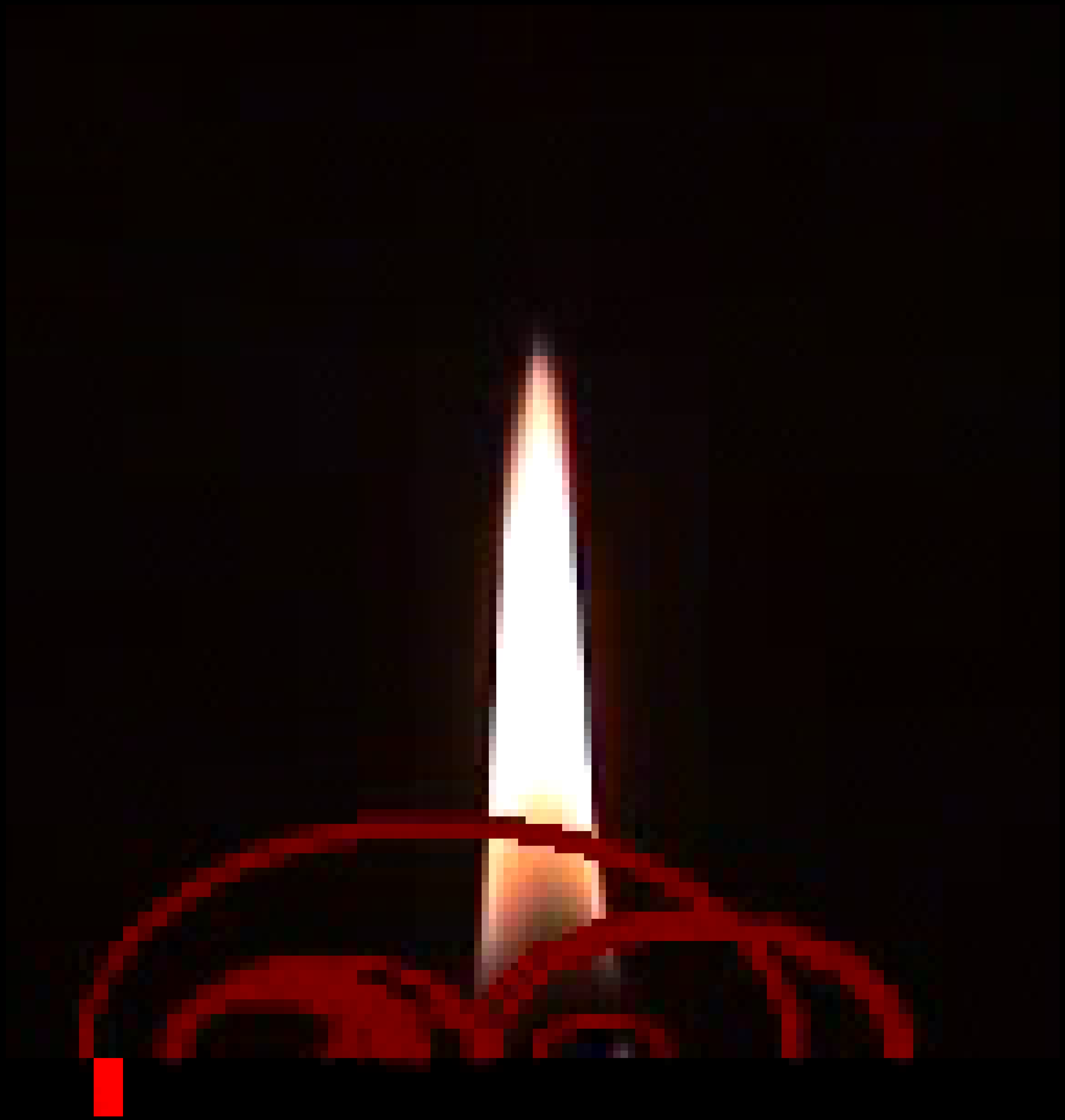

Video textures

Problem statement

video clip

video texture

Our approach



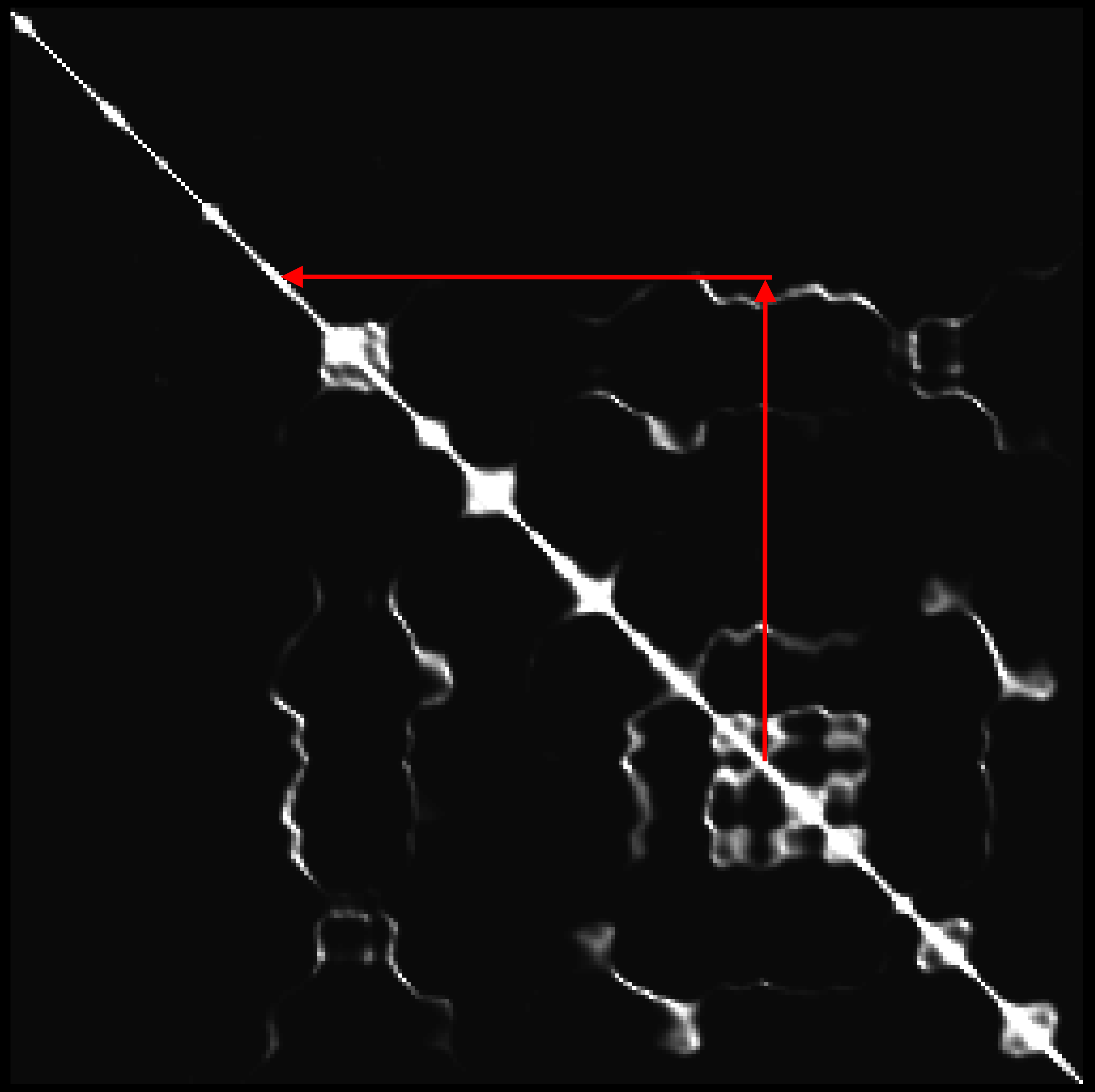
- How do we find good transitions?

Finding good transitions

- Compute L_2 distance $D_{i,\,j}$ between all frames

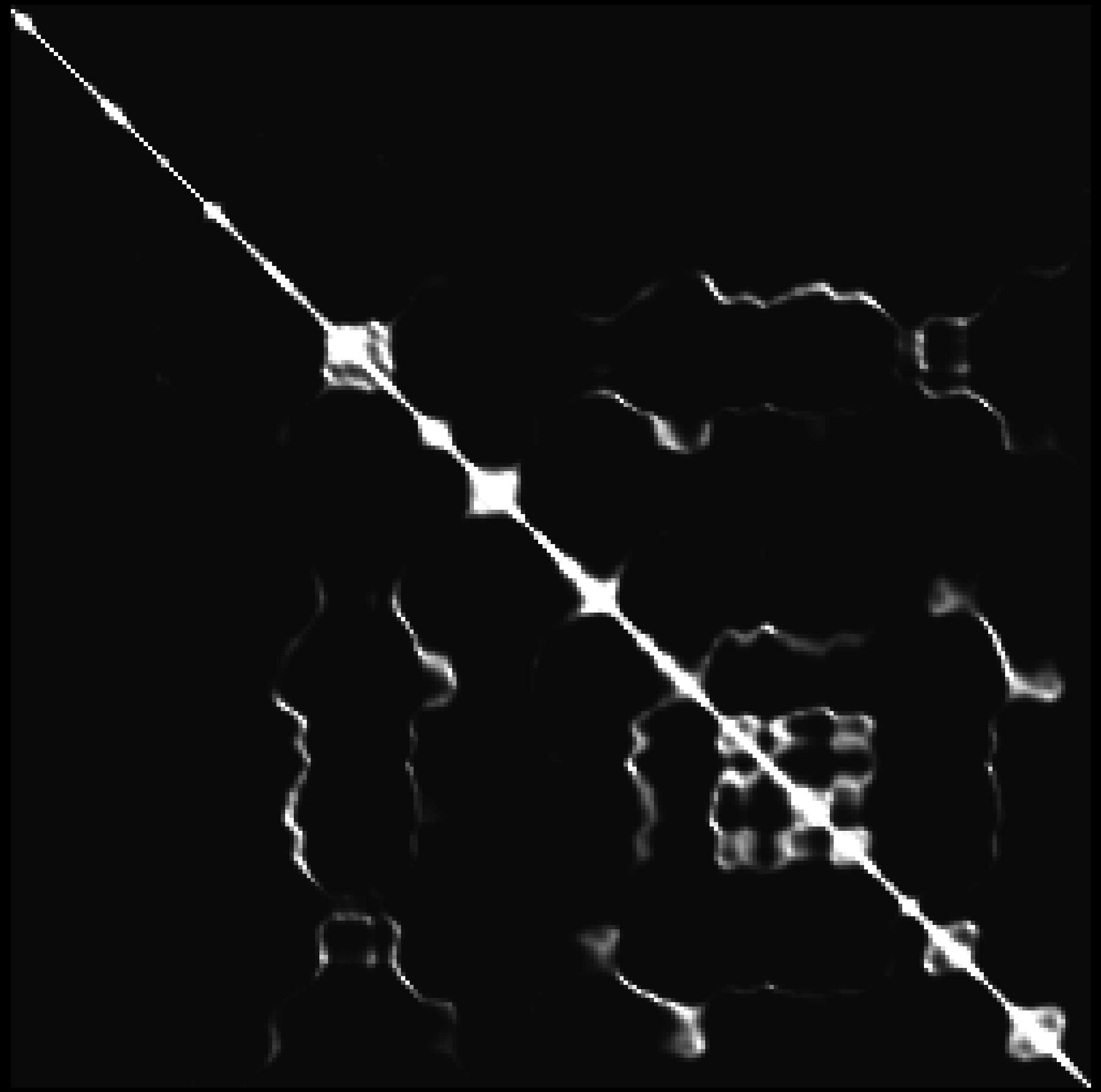
vs. → frame *i*

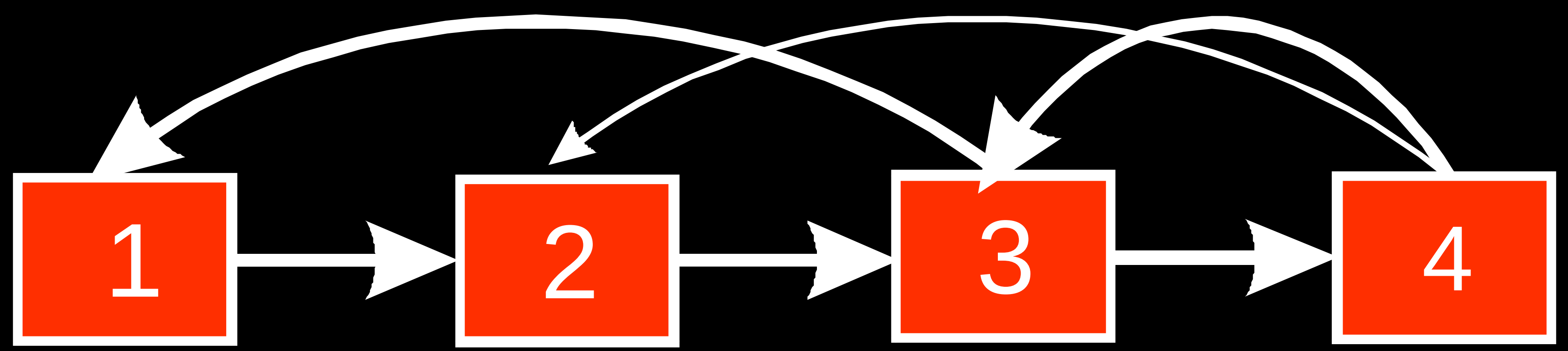
↓ frame *j*



Similar frames make good transitions

Markov chain representation

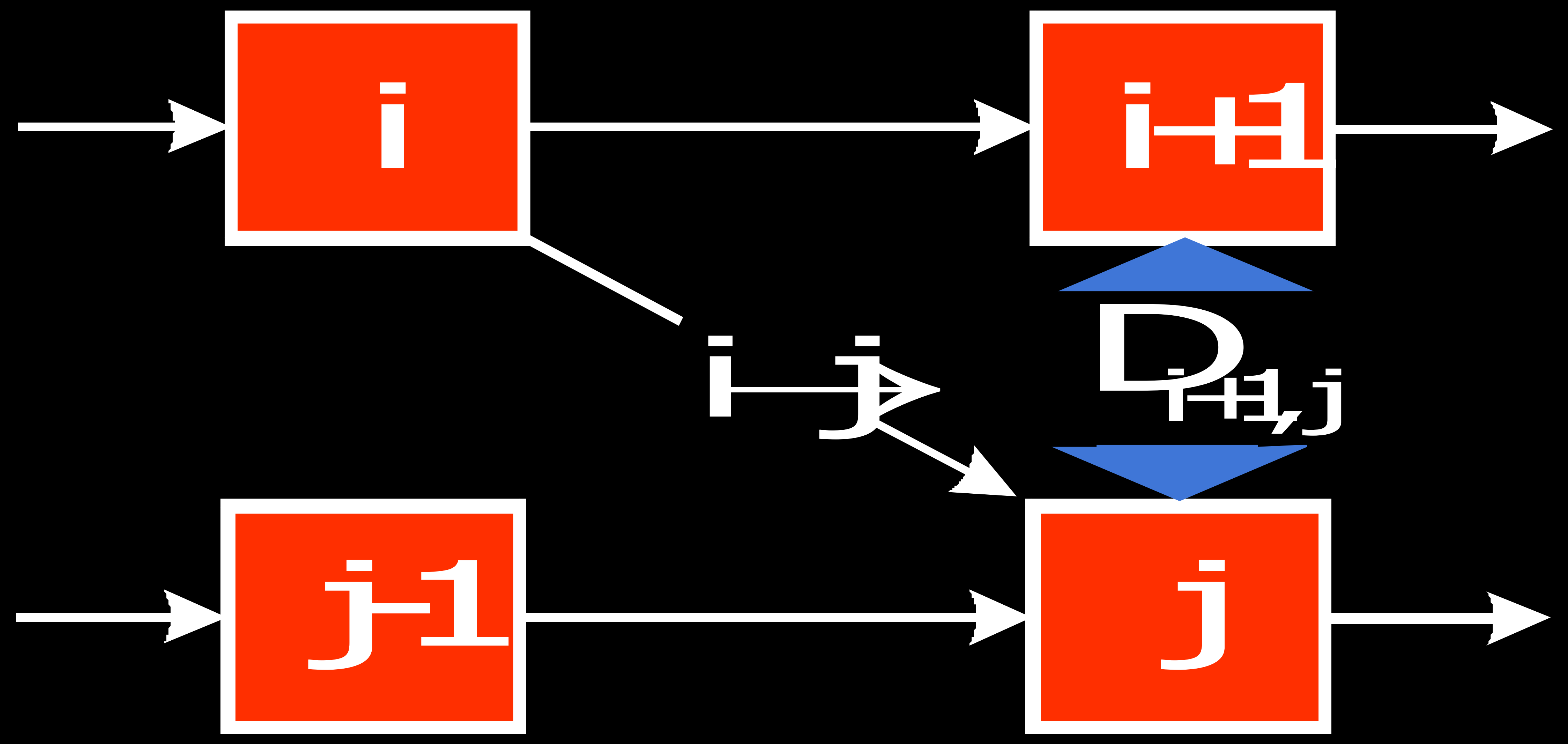




Similar frames make good transitions

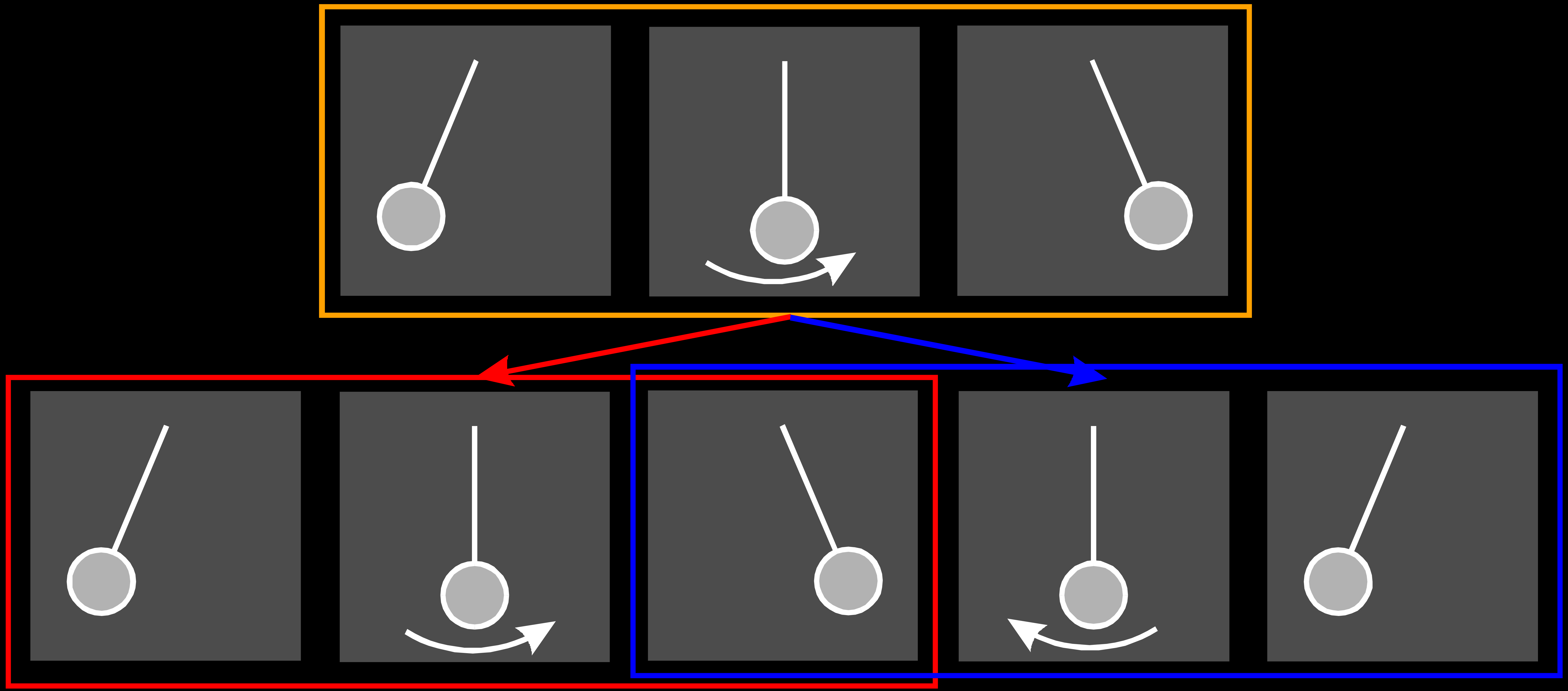
Transition costs

- Transition from i to j if successor of i is similar to j
  - Cost function: $C_{i\to j} = D_{i+1,\,j}$
-



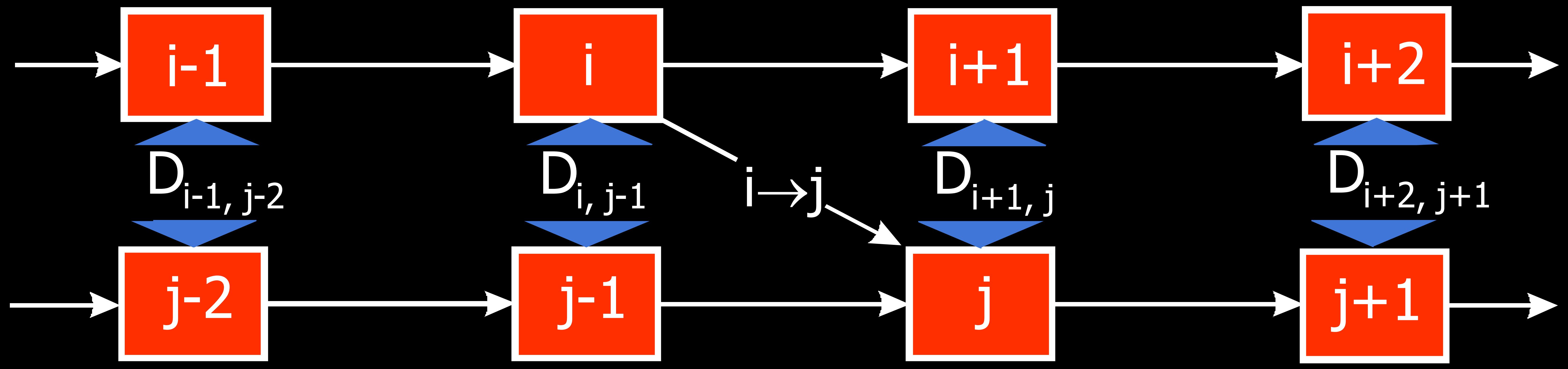
Preserving dynamics

Preserving dynamics



Preserving dynamics

- Cost for transition $i \to j$
  - $$C_{i \to j} = \sum_{k=-N}^{N-1} w_k D_{i+k+1,\, j+k}$$



Preserving dynamics – effect

- Cost for transition $i \to j$
  - $C_{i \to j} = \sum_{k=-N}^{N-1} w_k D_{i+k+1,\, j+k}$

Video portrait

- c.f. Harry Potter

Region-based analysis

- Divide video up into regions

- Generate a video texture for each region

User-controlled video textures

User selects target frame range

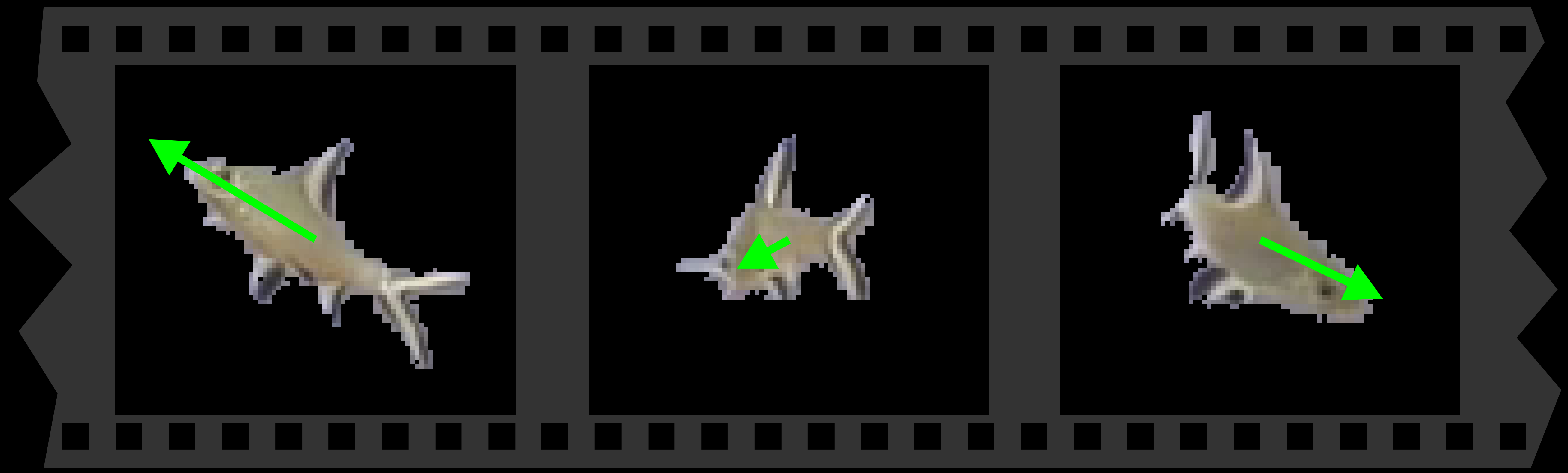
Video-based animation

- Like sprites computer games
- Extract sprites from real video
- Interactively control desired motion

©1985 Nintendo of America Inc.

Video sprite extraction

blue screen matting
and velocity estimation



Video sprite control

- Augmented transition cost:

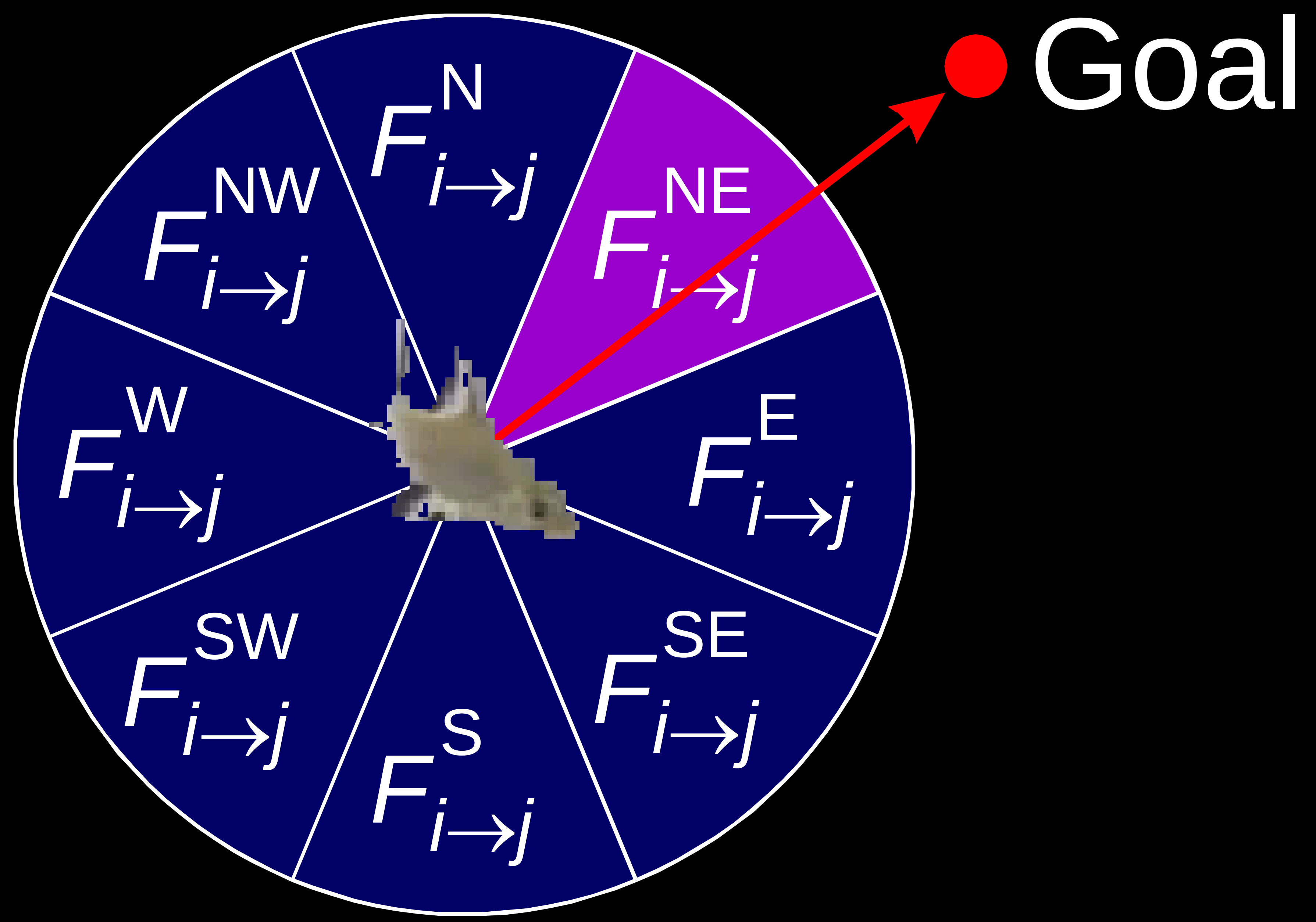
$$C^{\text{Animation}}_{i\to j} = \alpha \underbrace{C_{i\to j}}_{\text{Similarity term}} + \beta \underbrace{\text{angle}}_{\text{Control term}}$$

vector to
mouse pointer

velocity vector

Video sprite control

- Need future cost computation
- Precompute future costs for a few angles.
- Switch between precomputed angles according to user input
- [GIT-GVU-00-11]



Interactive fish

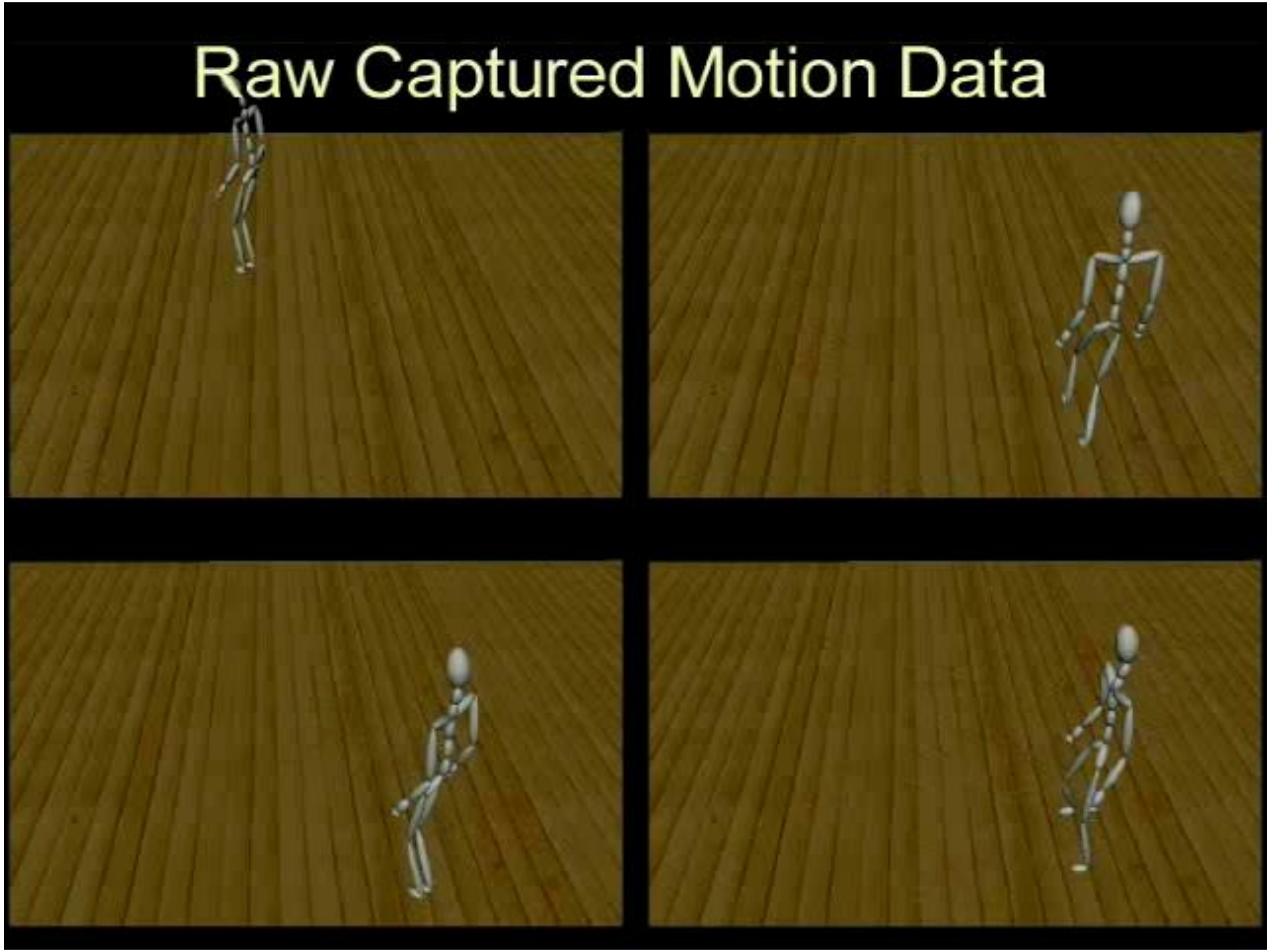
Summary / Discussion

- Some things are relatively easy

Discussion

- Some are hard

Motion Caption Database

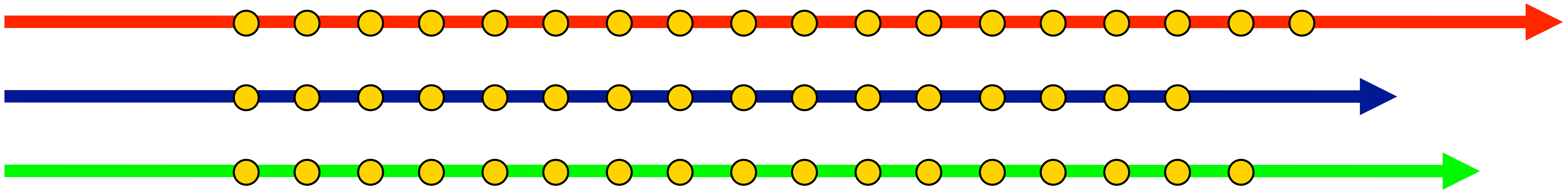


Jehee Lee, Jinxiang Chai, Paul Reitsma, Nancy Pollard, Jessica Hodgins SIGGRAPH 2002

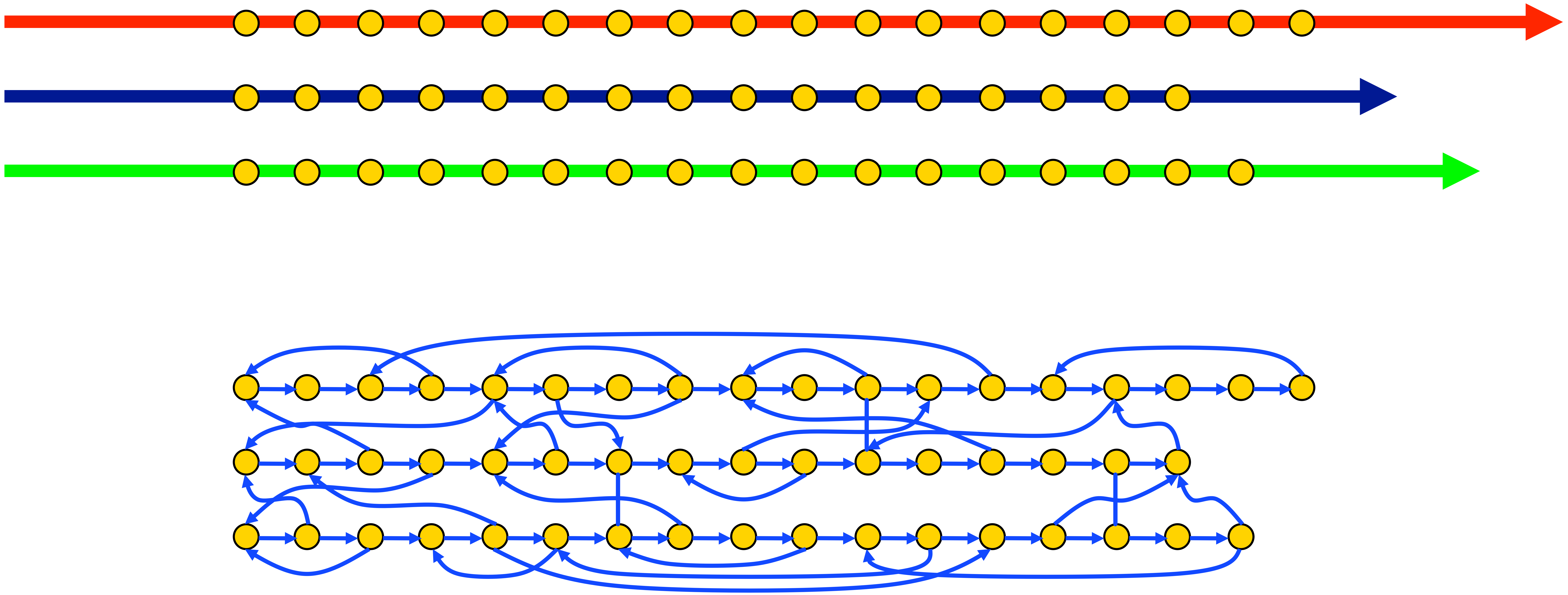
Sketch Interface

Jehee Lee, Jinxiang Chai, Paul Reitsma, Nancy Pollard, Jessica Hodgins SIGGRAPH 2002

Unstructured Input Data

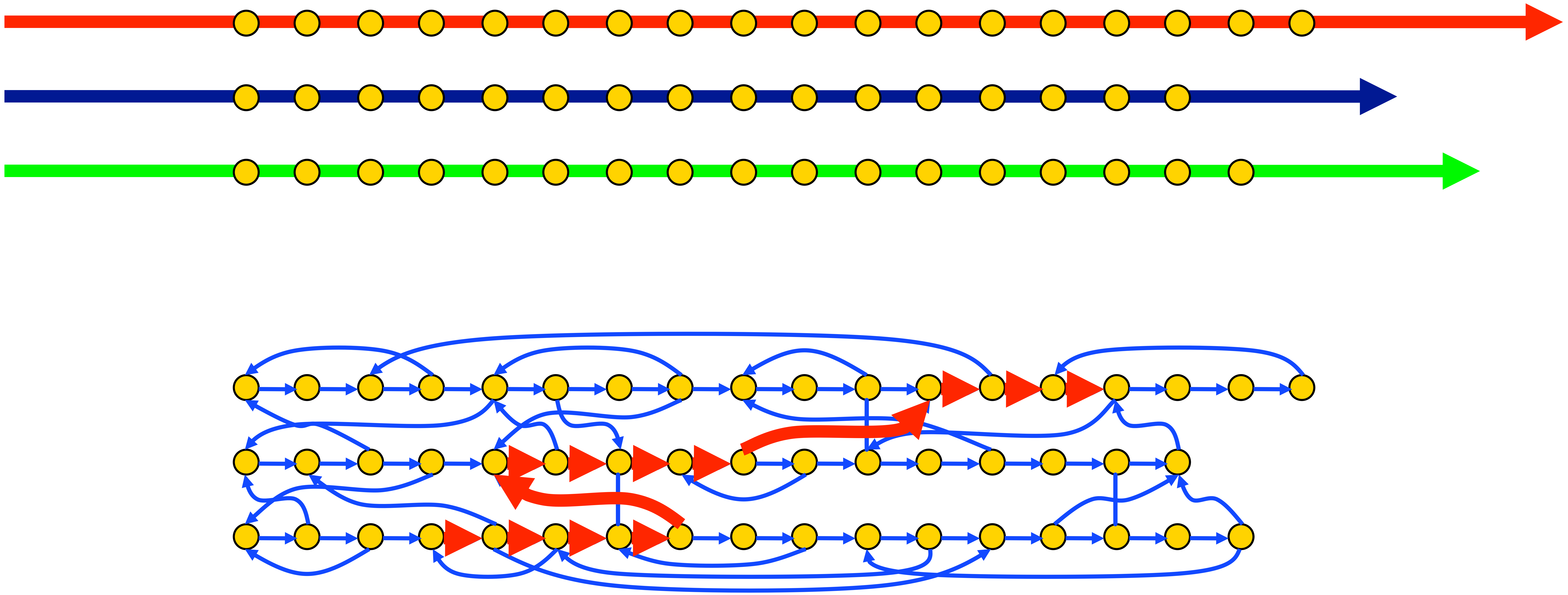


Connecting Transitions



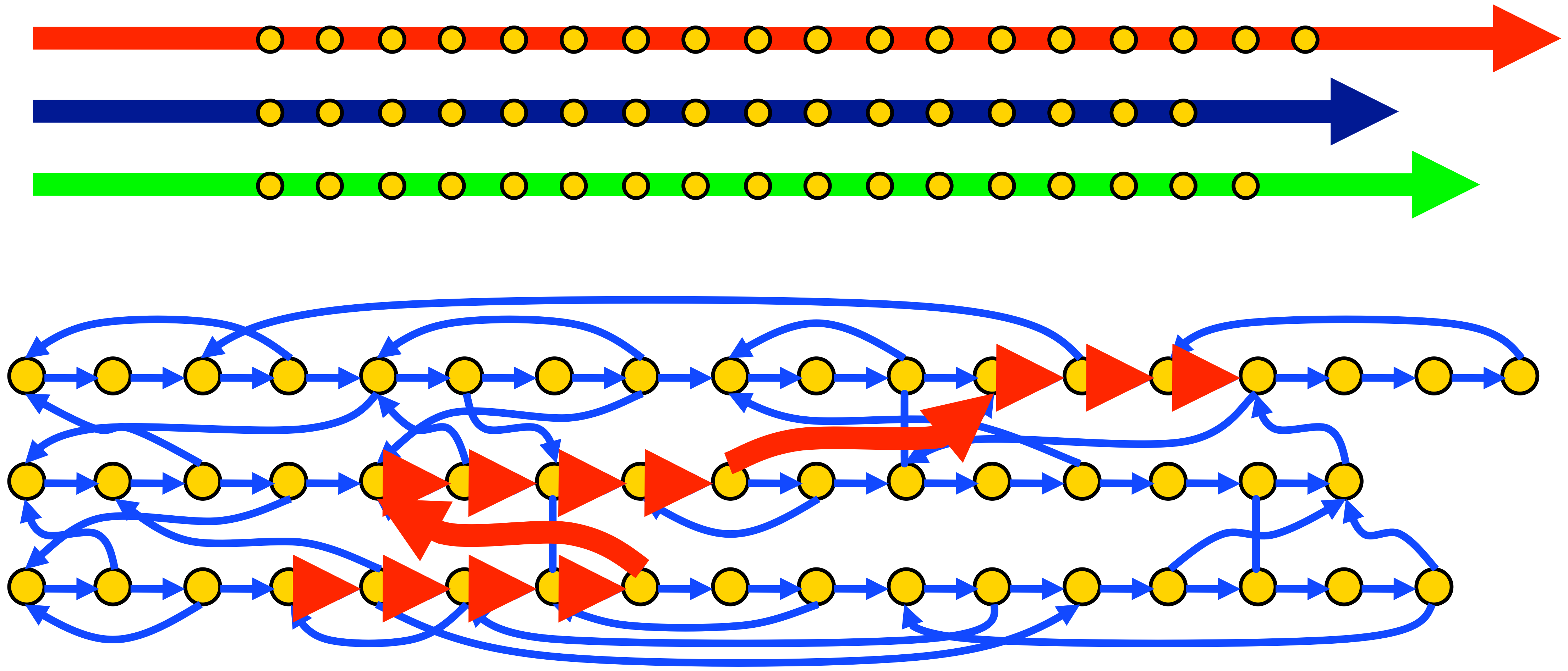
Jehee Lee, Jinxiang Chai, Paul Reitsma, Nancy Pollard, Jessica Hodgins SIGGRAPH 2002

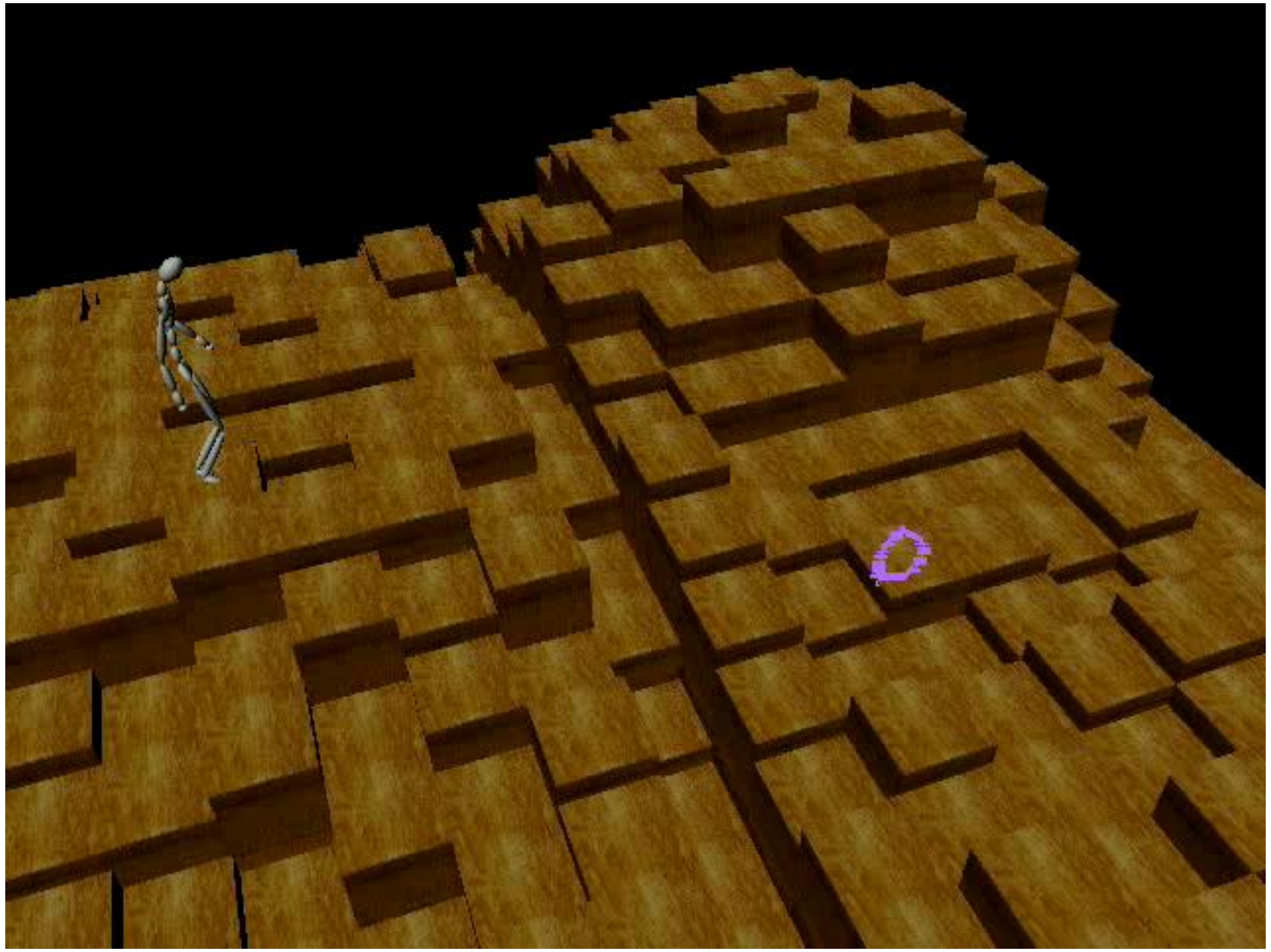
Search to Find Path



Jehee Lee, Jinxiang Chai, Paul Reitsma, Nancy Pollard, Jessica Hodgins SIGGRAPH 2002

Motion Data for Rough Terrain





Jehee Lee, Jinxiang Chai, Paul Reitsma, Nancy Pollard, Jessica Hodgins SIGGRAPH 2002

What did they learn?

Databases do not need to be nearly as big as we would have guessed

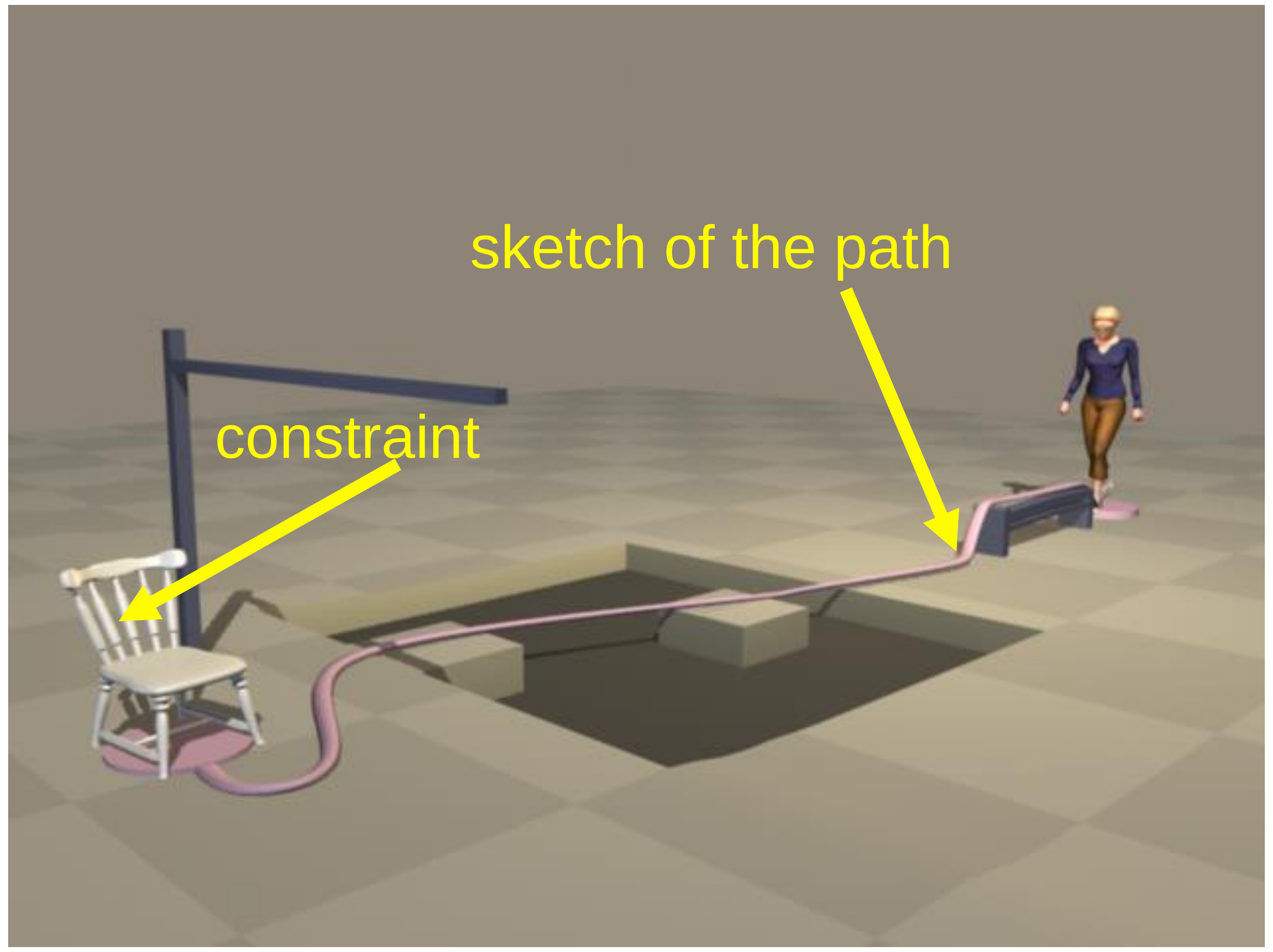
~5 minutes is sufficient for walking

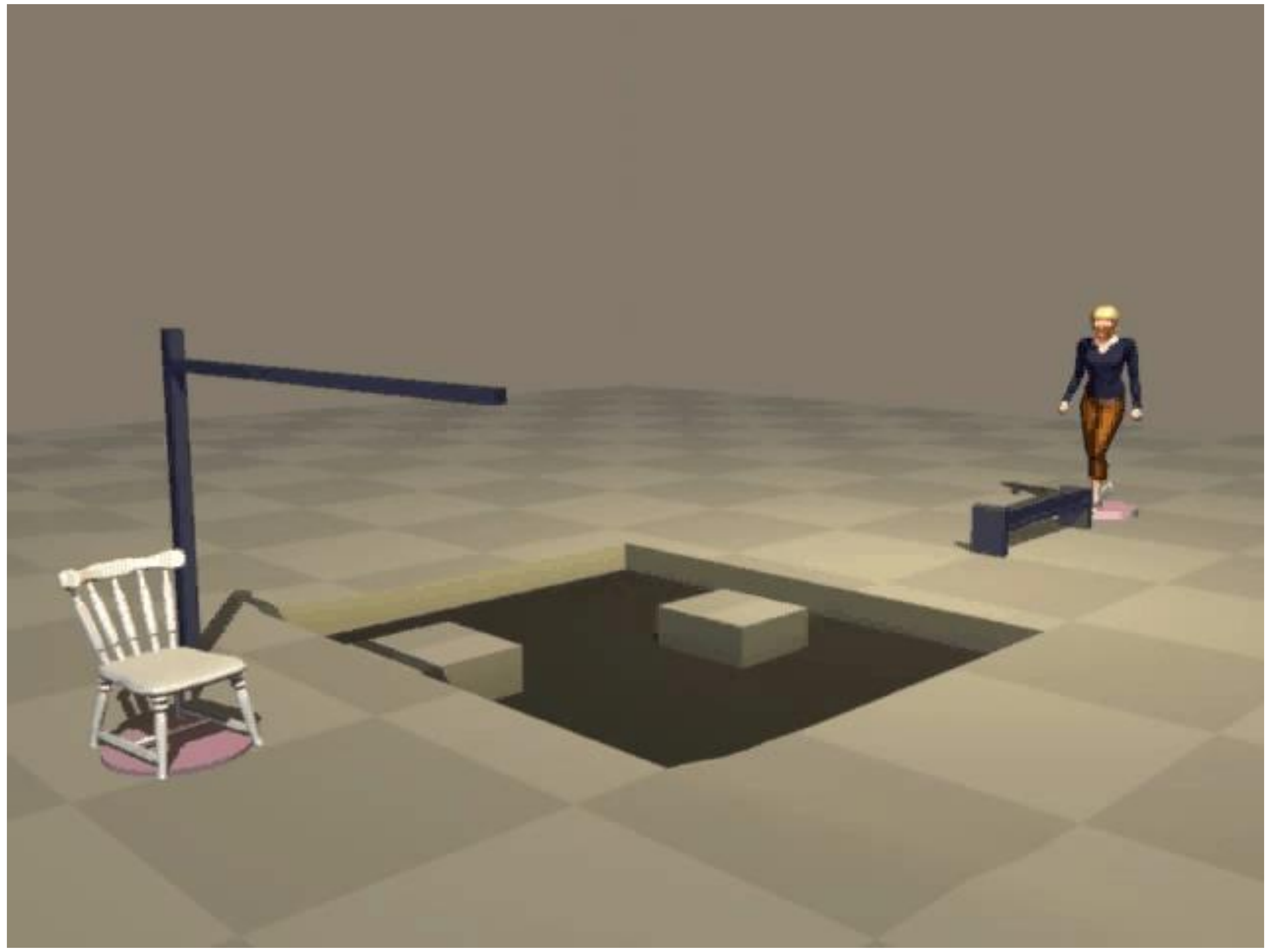
Planning in the space of possible behaviors (as represented by the data) is quite efficient

Interpolated Motion Graphs

Synthesized motion

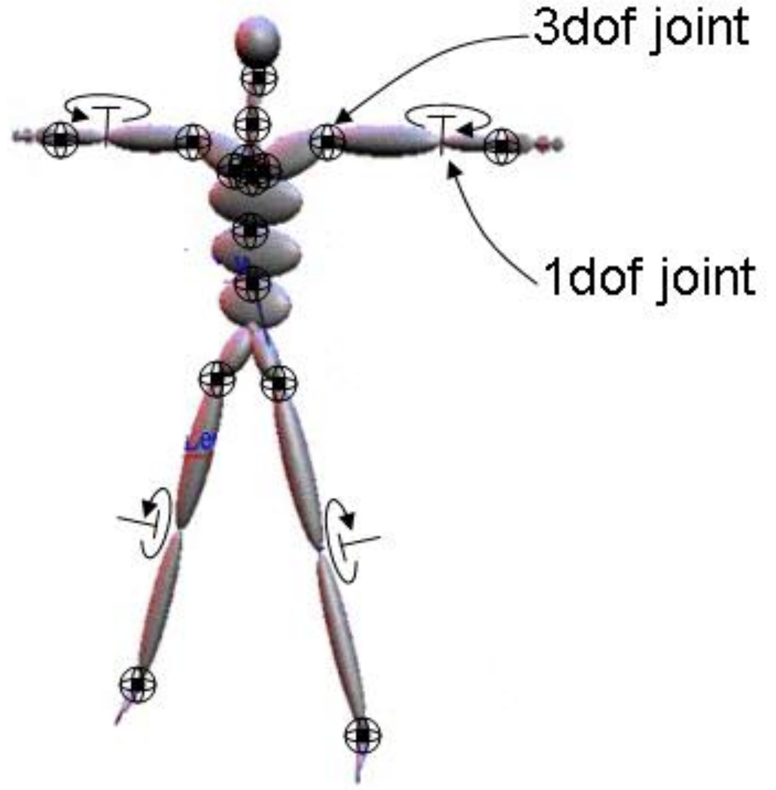
Sketch of the path





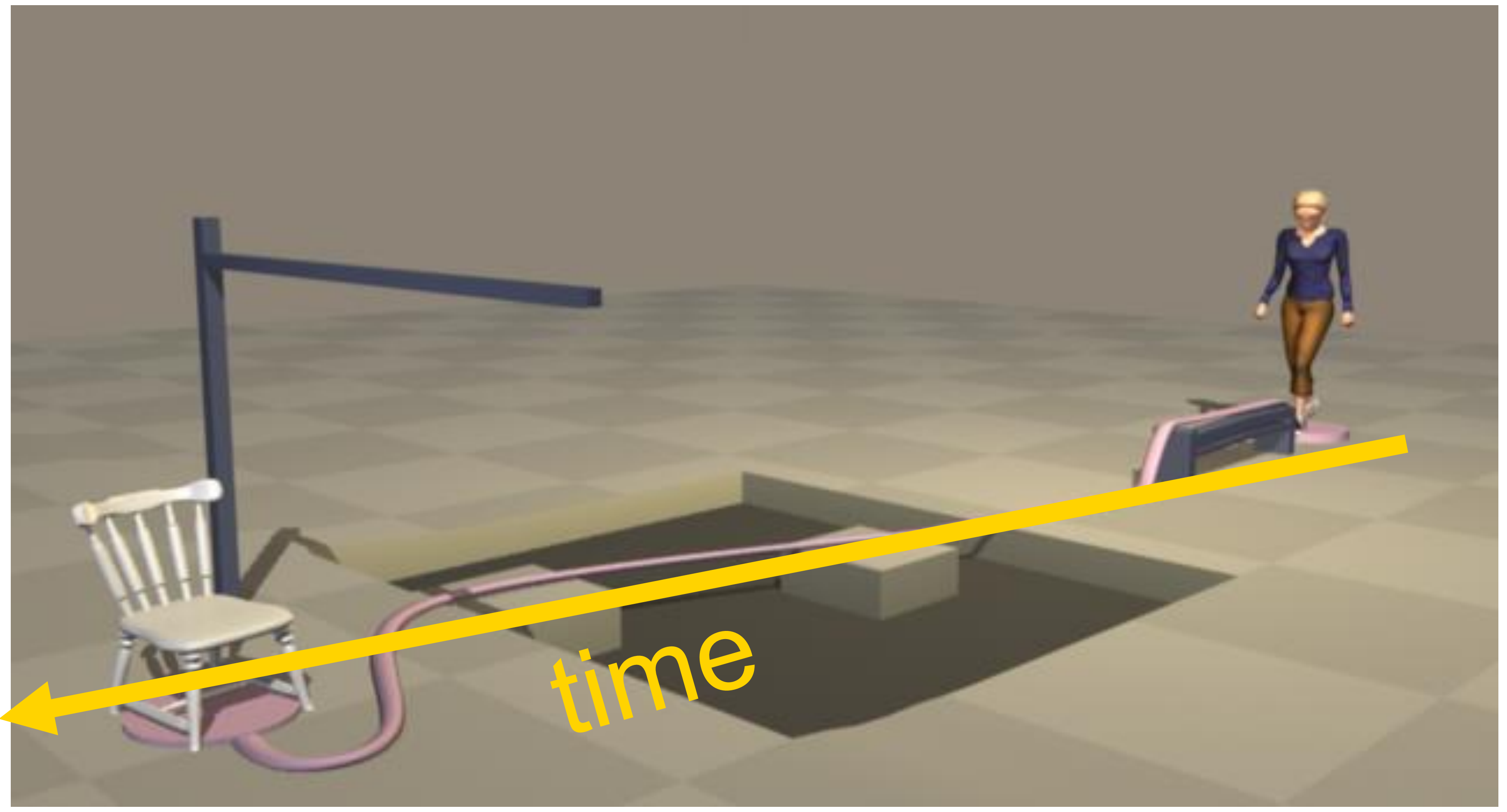
Alla Safonova & Jessica Hodgins, SIGGRAPH 2007

Optimization Problem



50 degrees of freedom

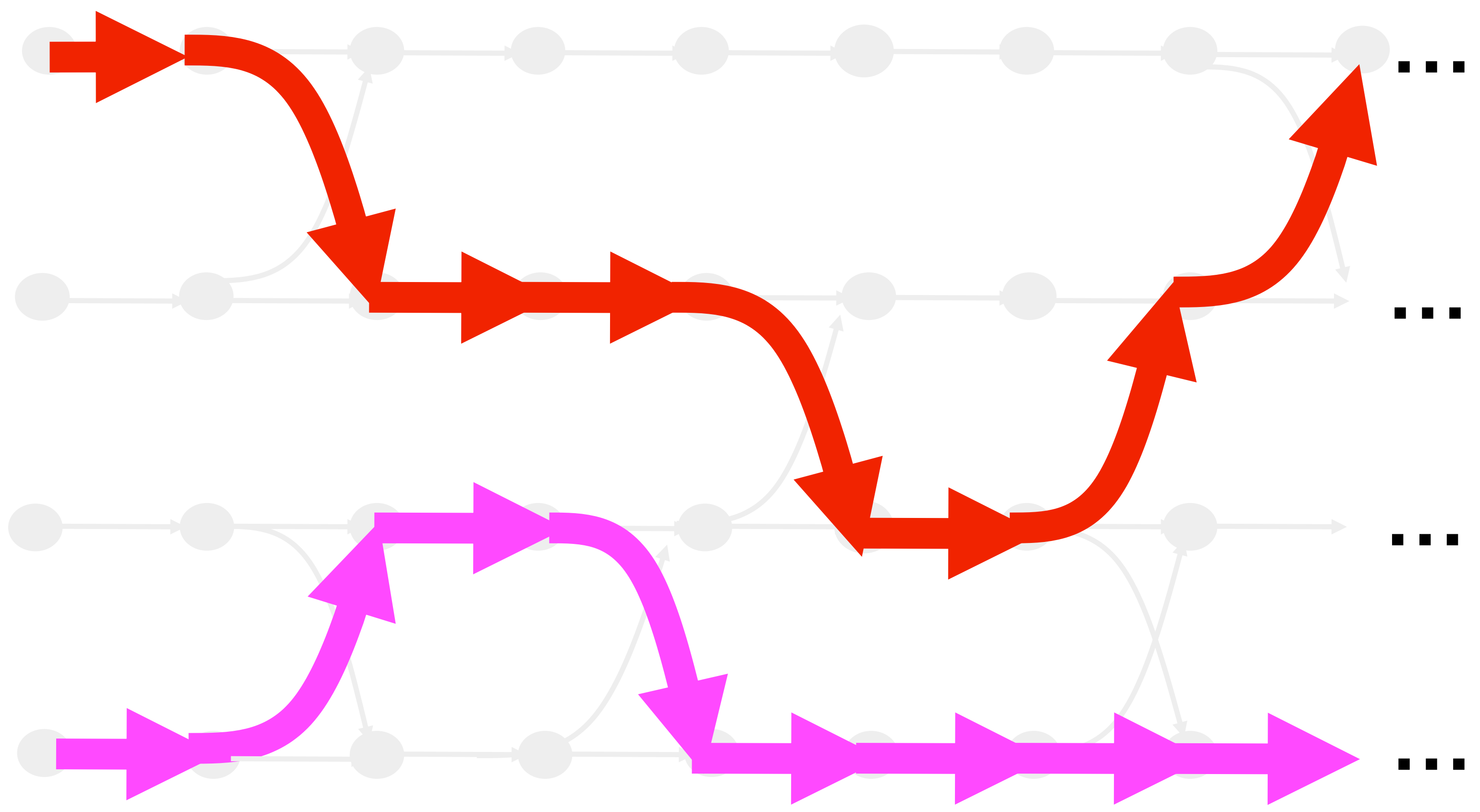
Unknowns: poses of the character over time



Minimize: sum of squared torques
Constraints: user, environmental and physics

Naïve Search space is very large: 50 x T unknowns

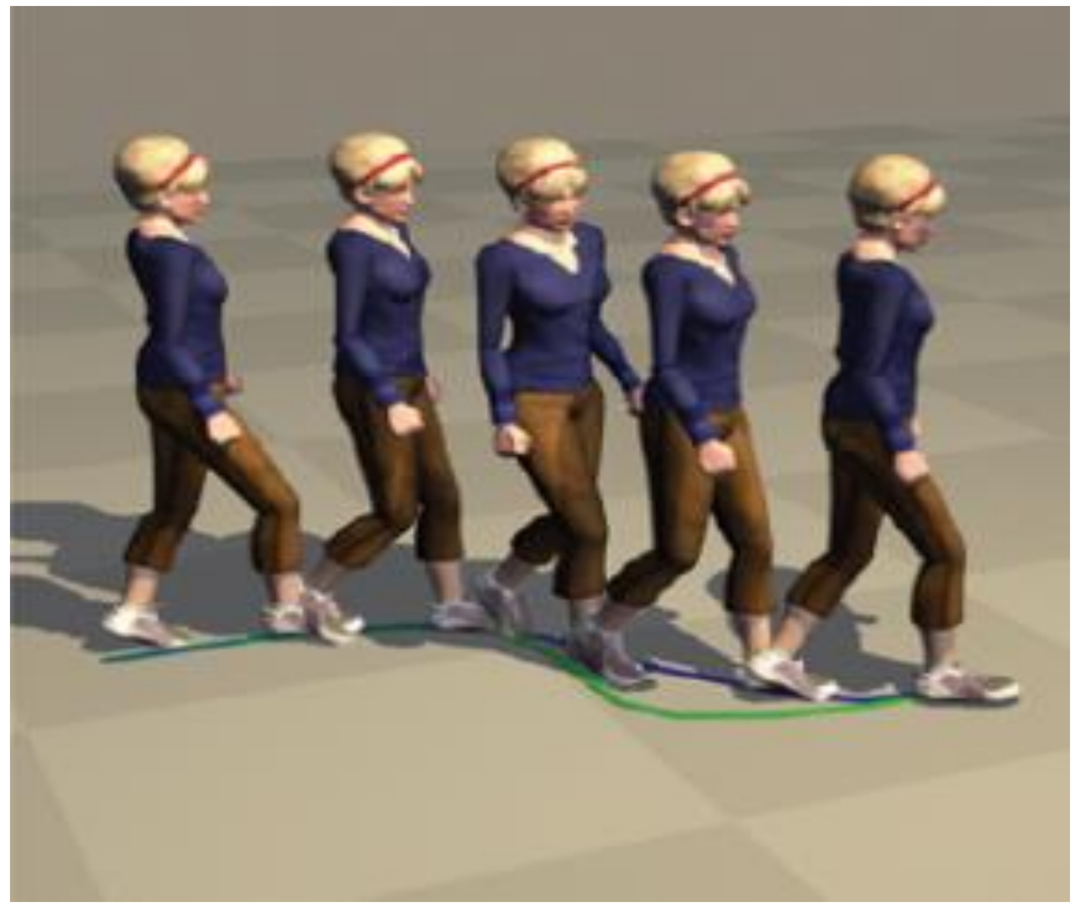
Key Idea 1: Interpolated Motion Graphs



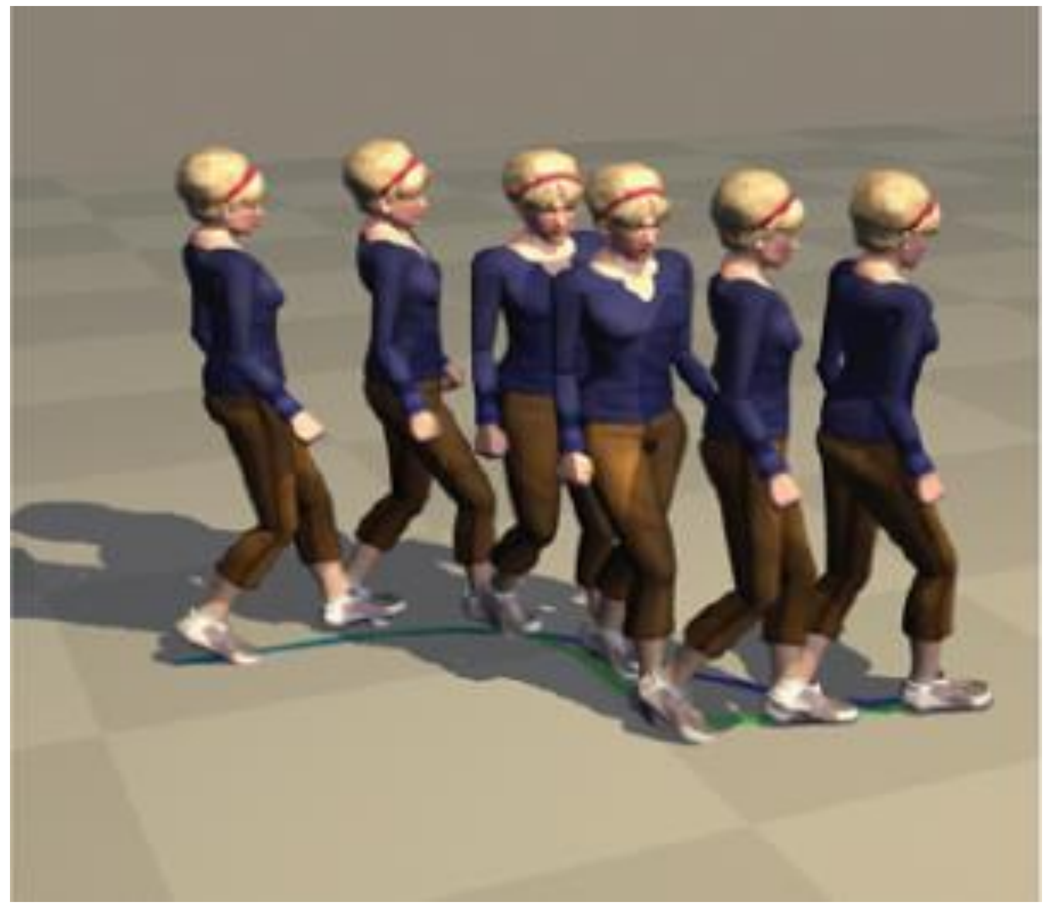
Compact representation of motion

- Long, multi-behavior motions
- Novel motions

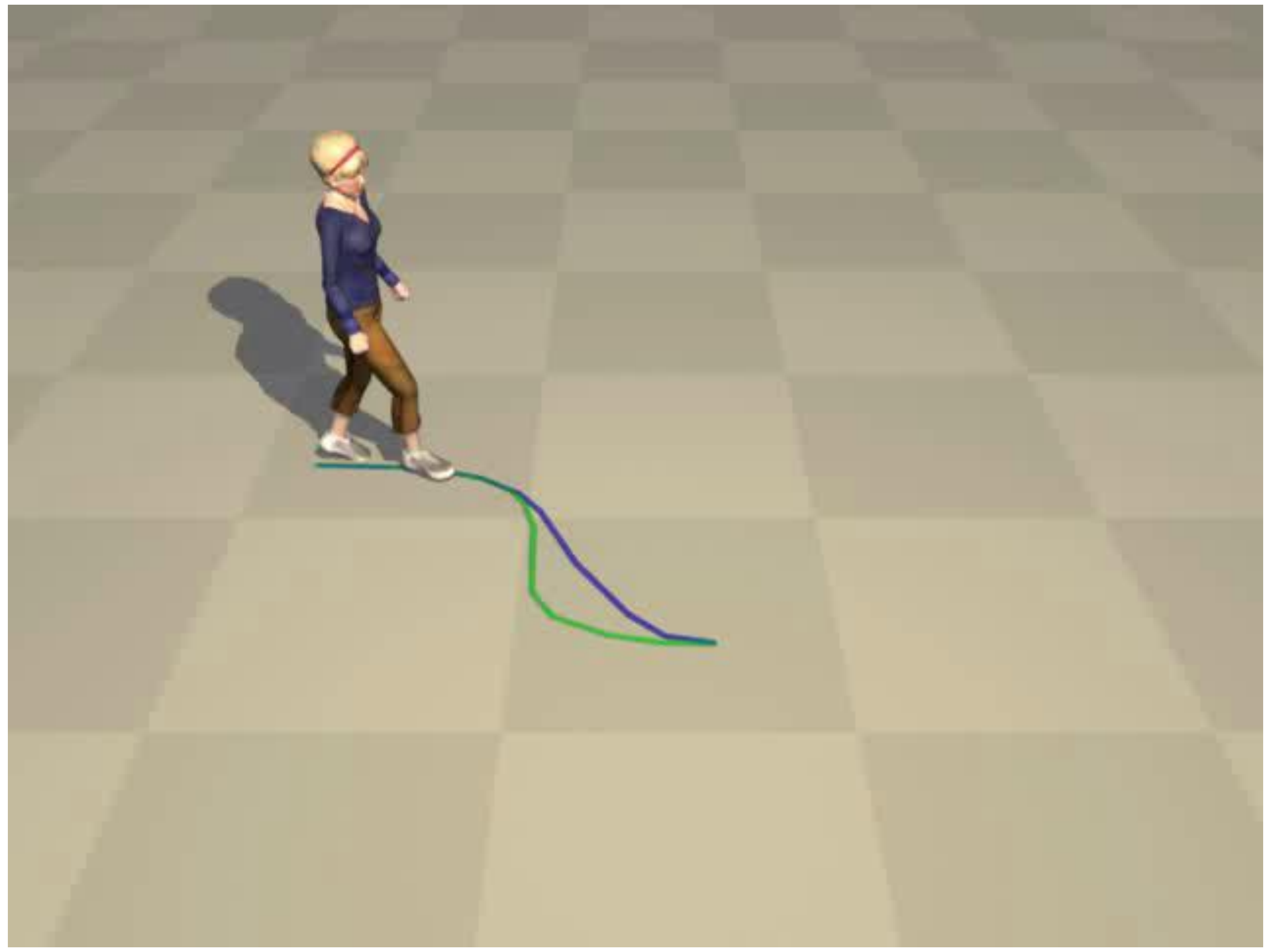
Interpolation



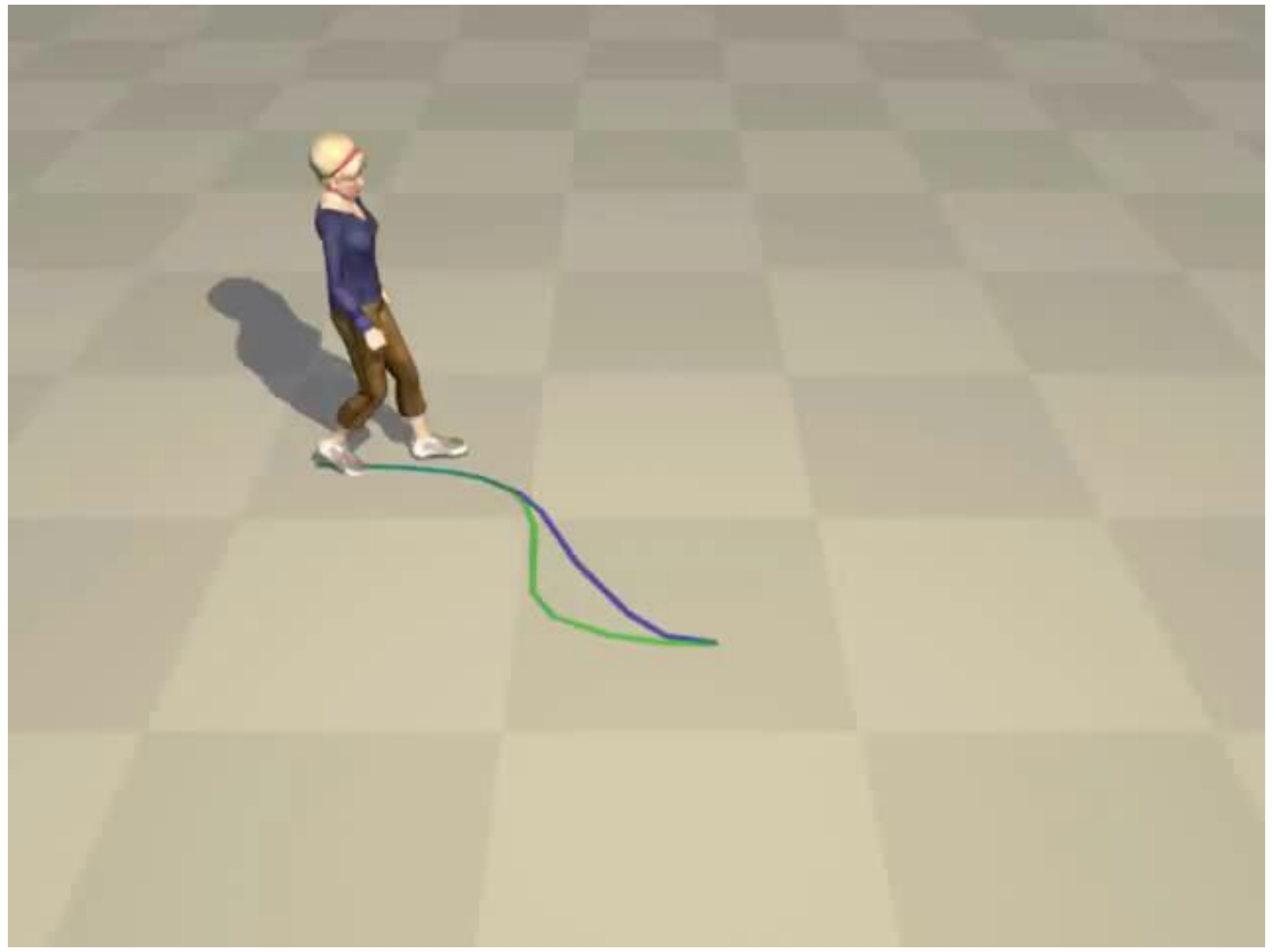
no interpolation



with interpolation



no interpolation



with interpolation

Optimality

sub-optimal solutions

optimal solution

Alla Safonova & Jessica Hodgins, SIGGRAPH 2007

Texture

- Texture depicts spatially repeating patterns
- Many natural phenomena are textures

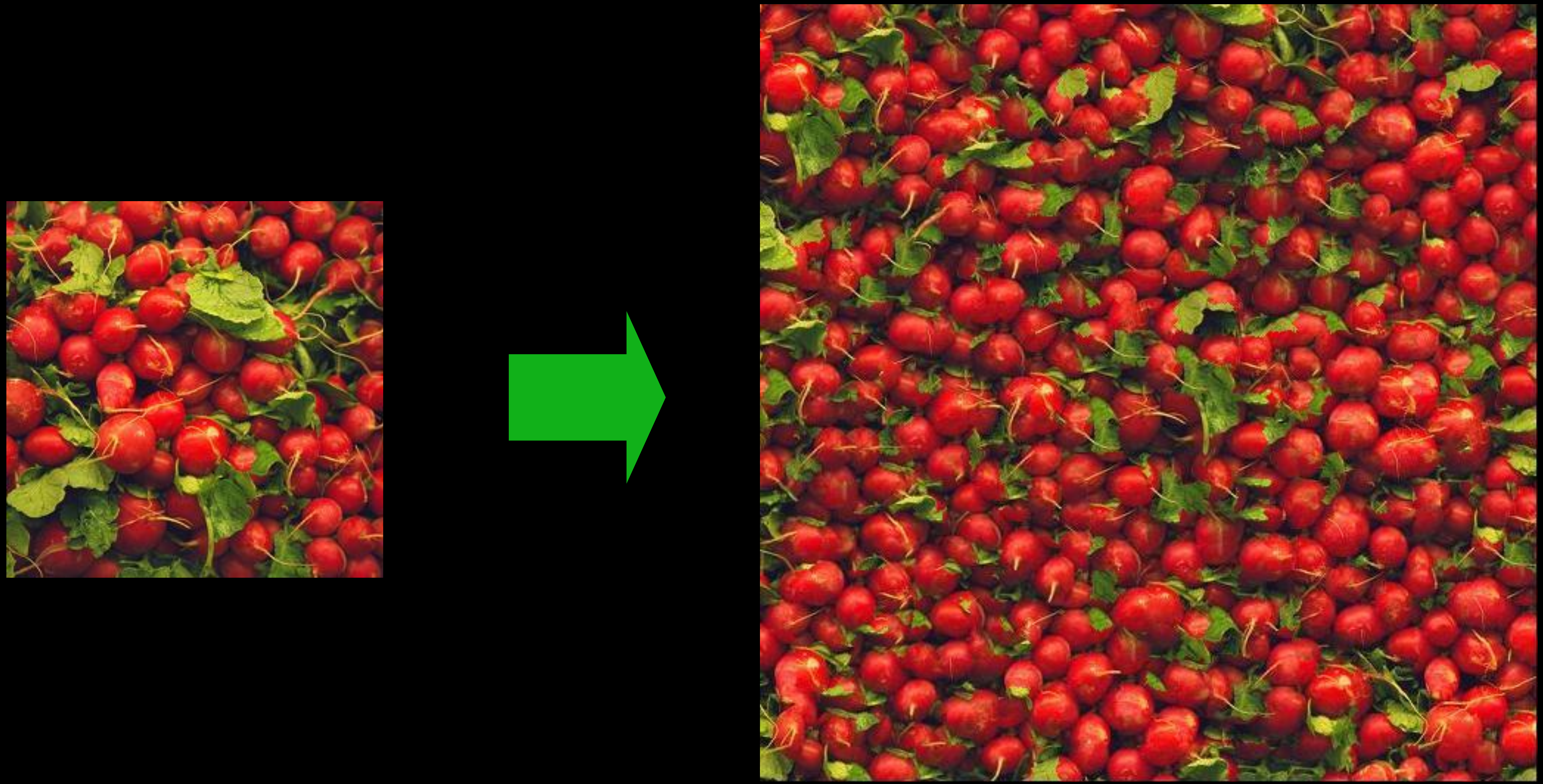
radishes

rocks

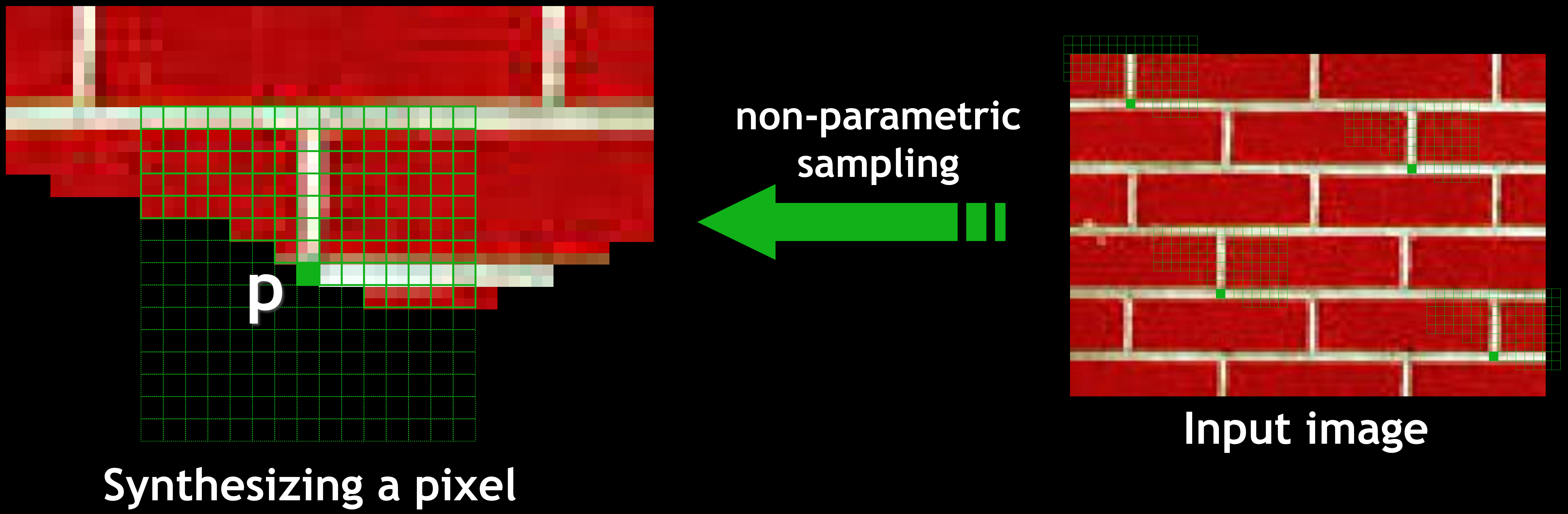
yogurt

Texture Synthesis

- Goal of Texture Synthesis: create new samples of a given texture
- Many applications: virtual environments, hole-filling, texturing surfaces



Efros & Leung Algorithm (ICCV 1999)

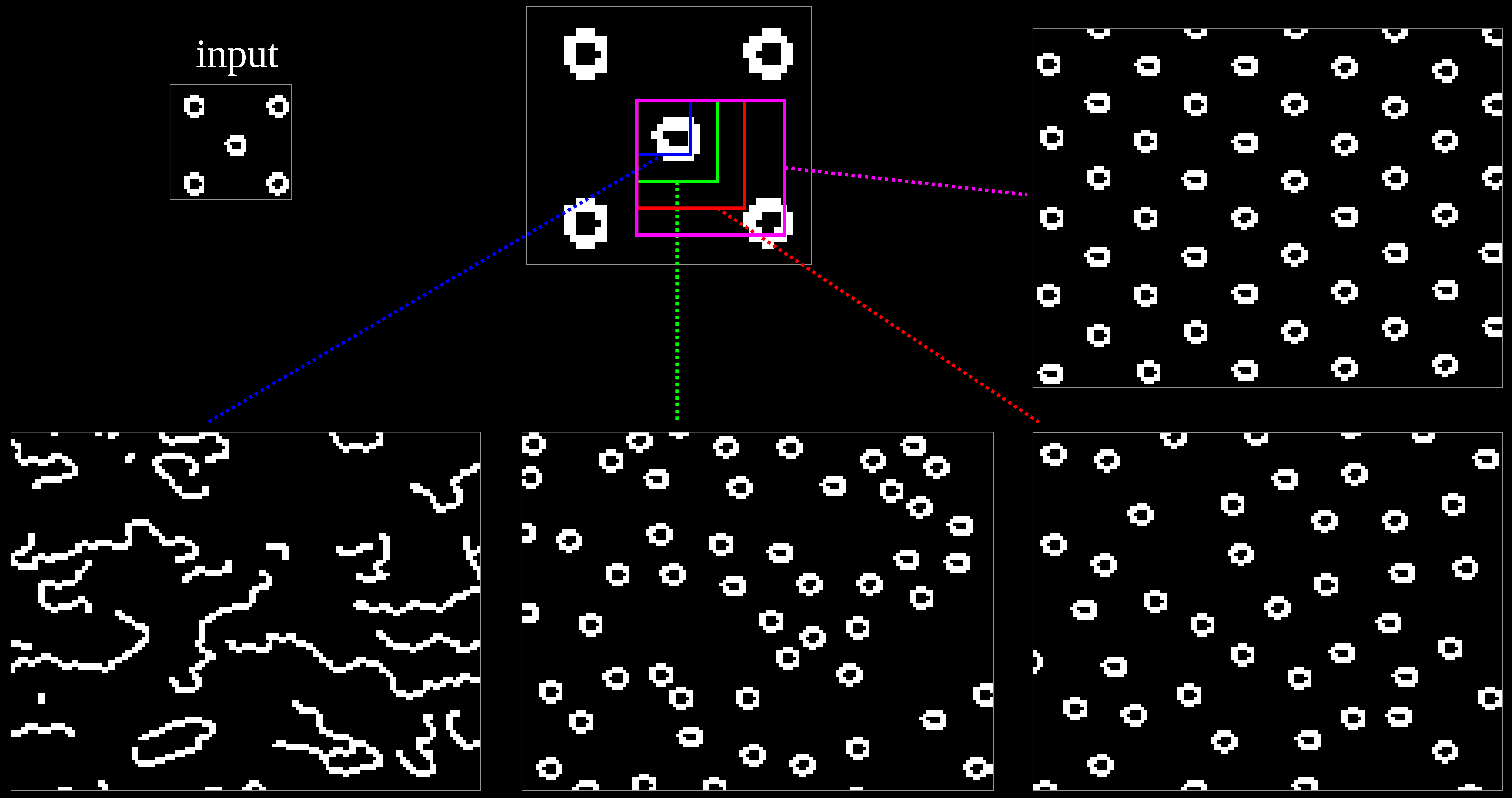


- Assuming Markov property, compute P(**p**|N(**p**))
  - Building explicit probability tables infeasible
  - Instead, we *search the input image* for all similar neighborhoods — that's our pdf for **p**
  - To sample from this pdf, just pick one match at random

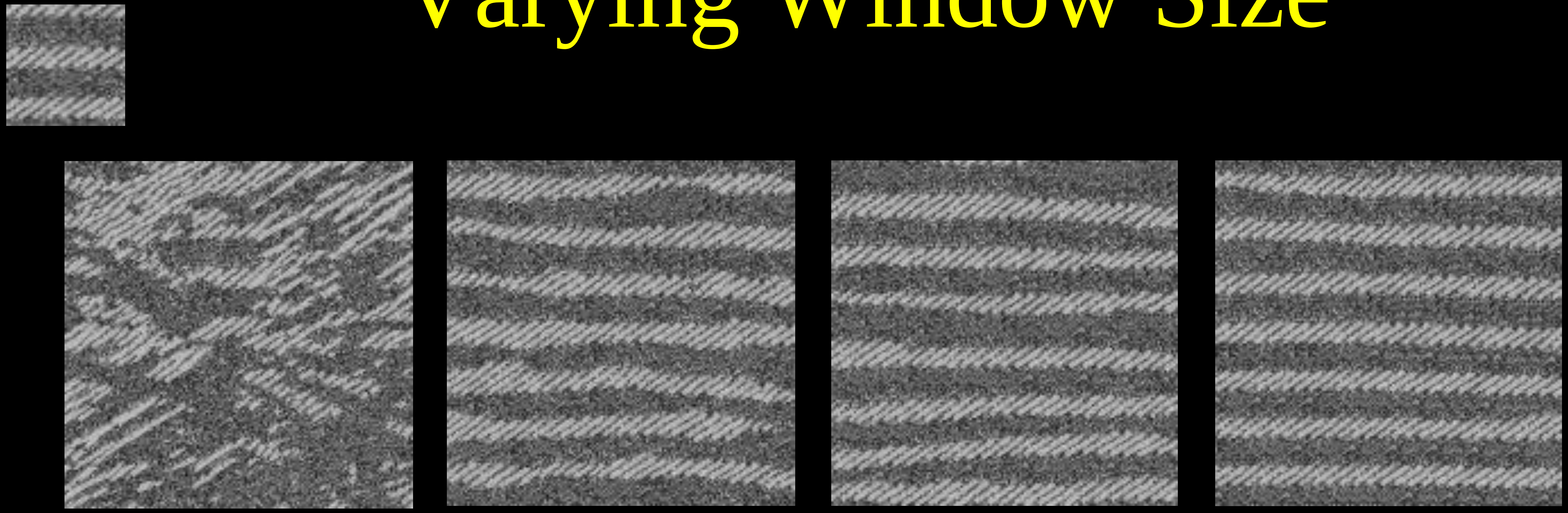
Some Details

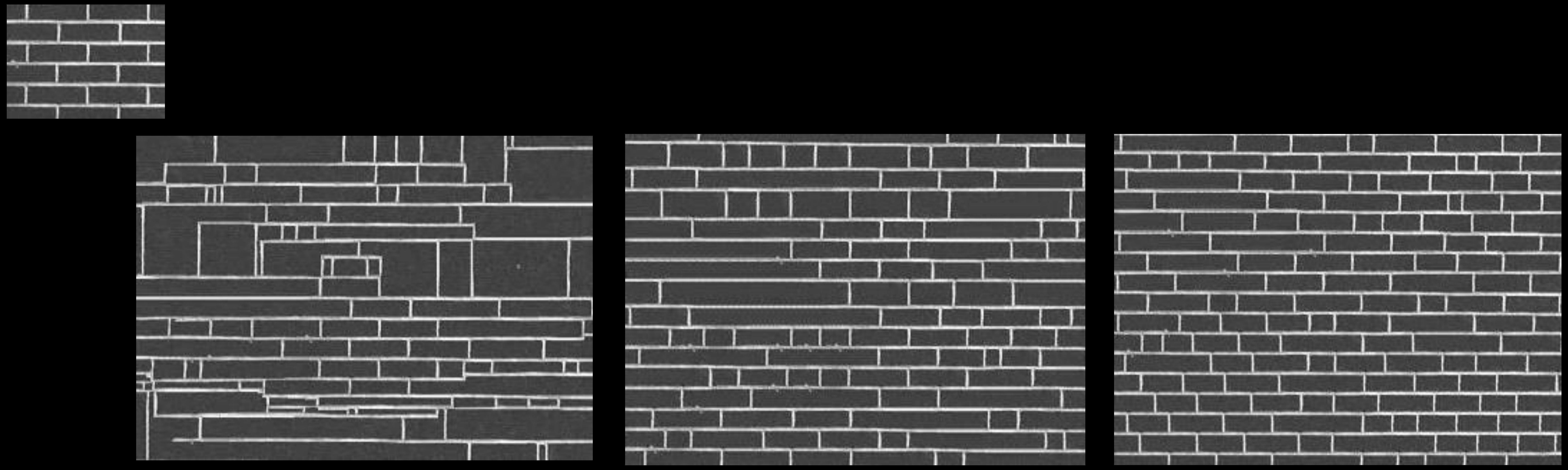
- Growing is in “onion skin” order
  - Within each “layer”, pixels with most neighbors are synthesized first
  - If no close match can be found, the pixel is not synthesized until the end
- Using *Gaussian-weighted* SSD is very important
  - to make sure the new pixel agrees with its closest neighbors
  - Approximates reduction to a smaller neighborhood window if data is too sparse

Neighborhood Window



Varying Window Size

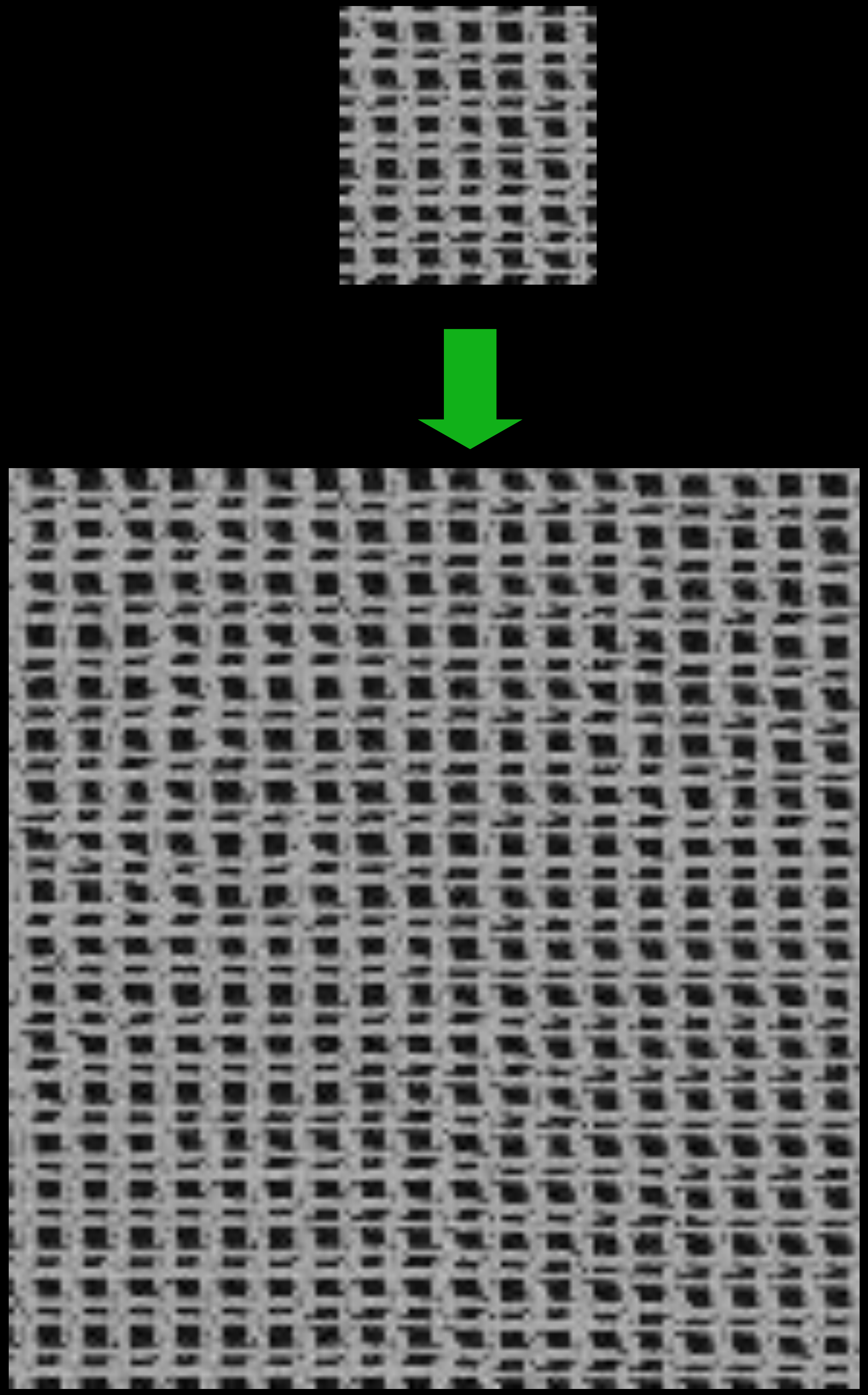




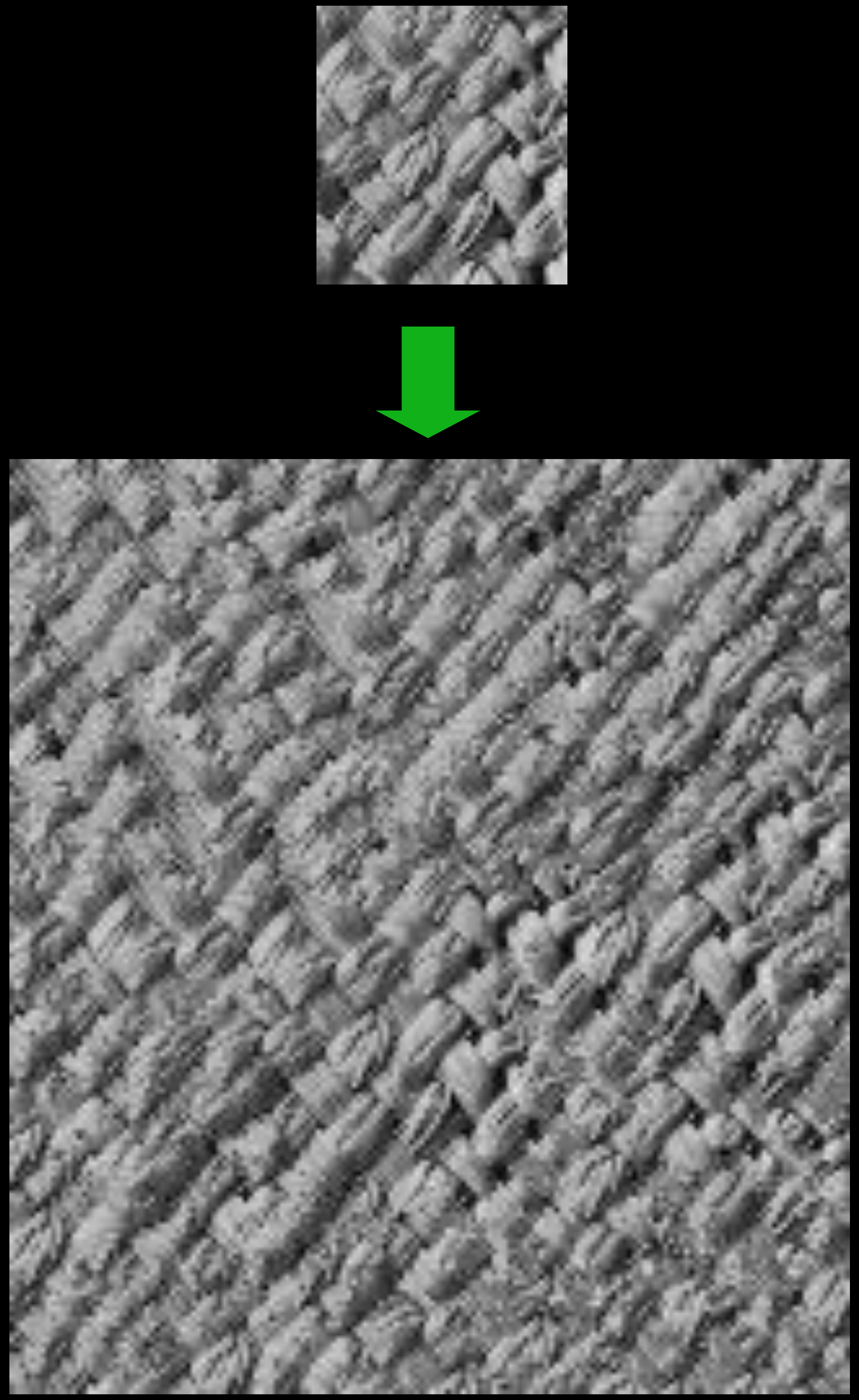
Increasing window size

Synthesis Results

french canvas

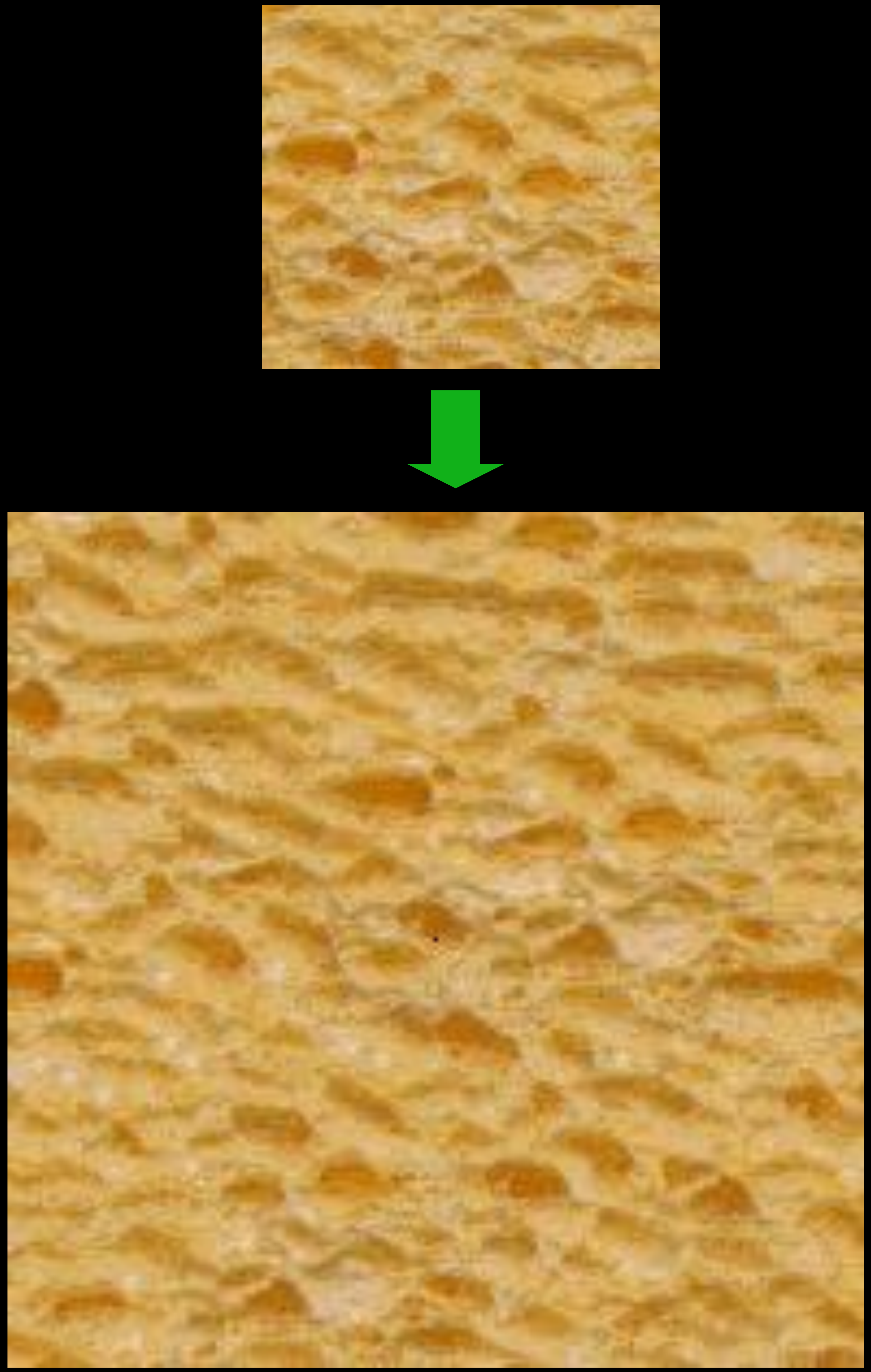


rafia weave

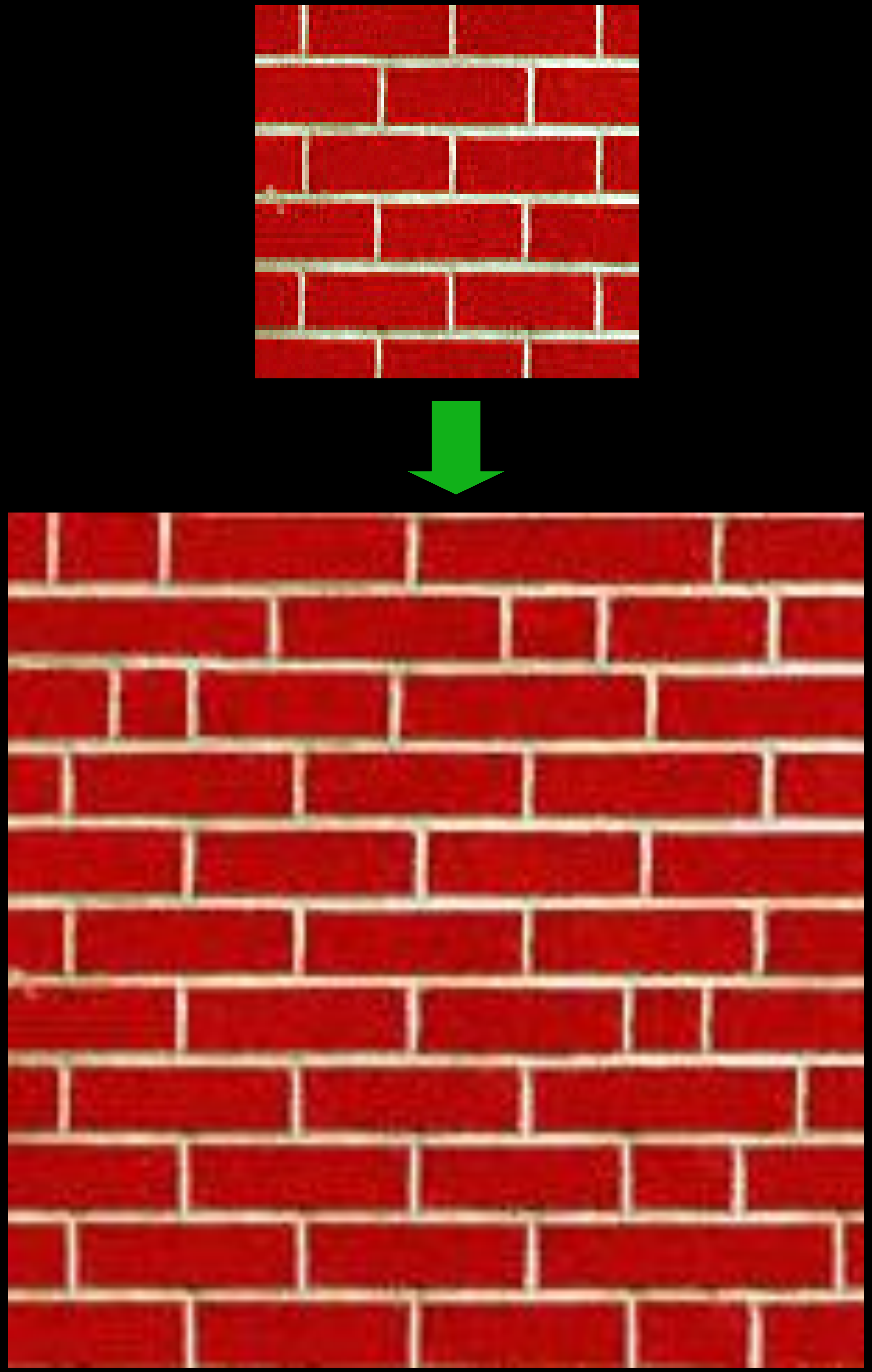


More Results

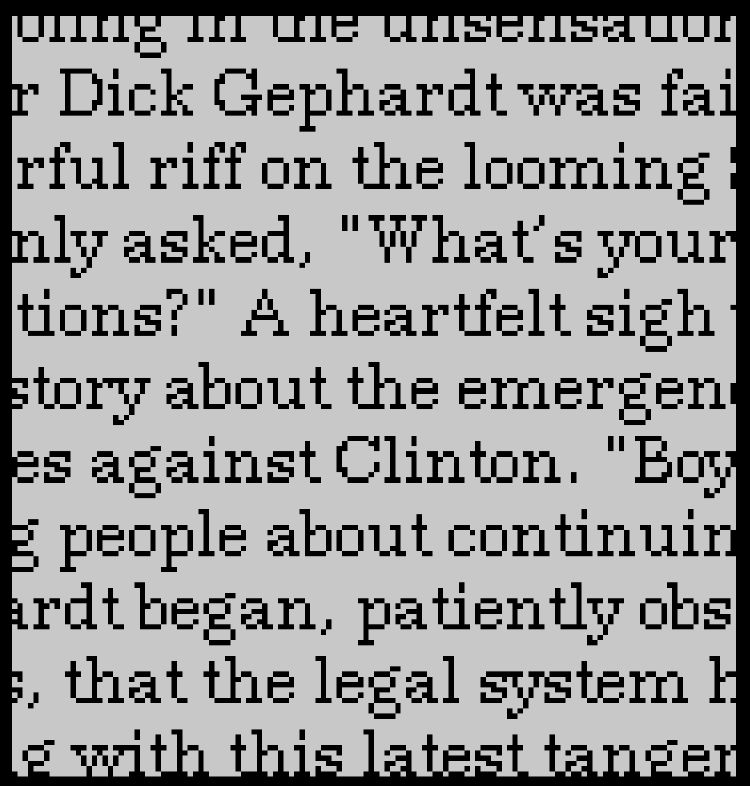
white bread

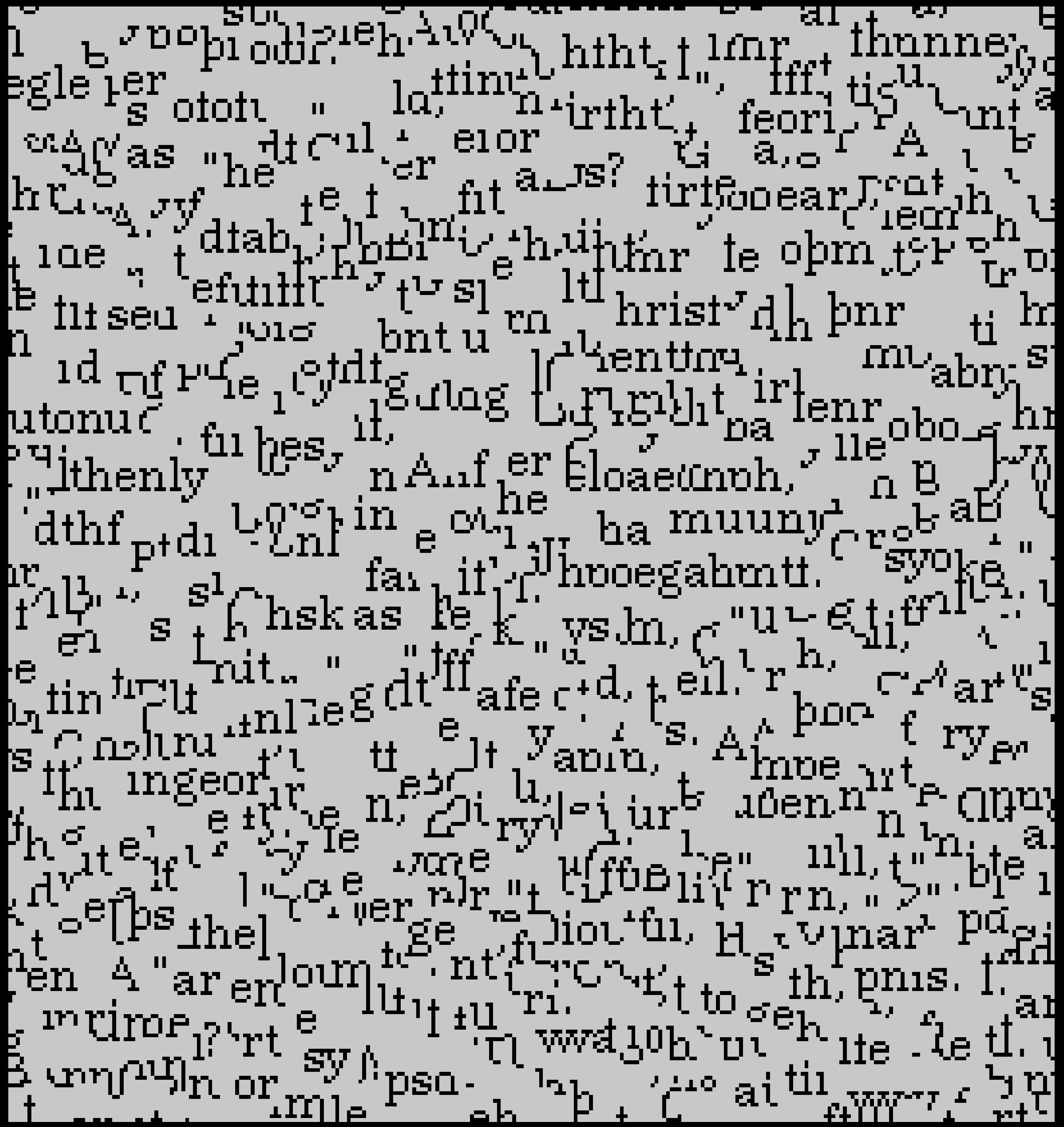


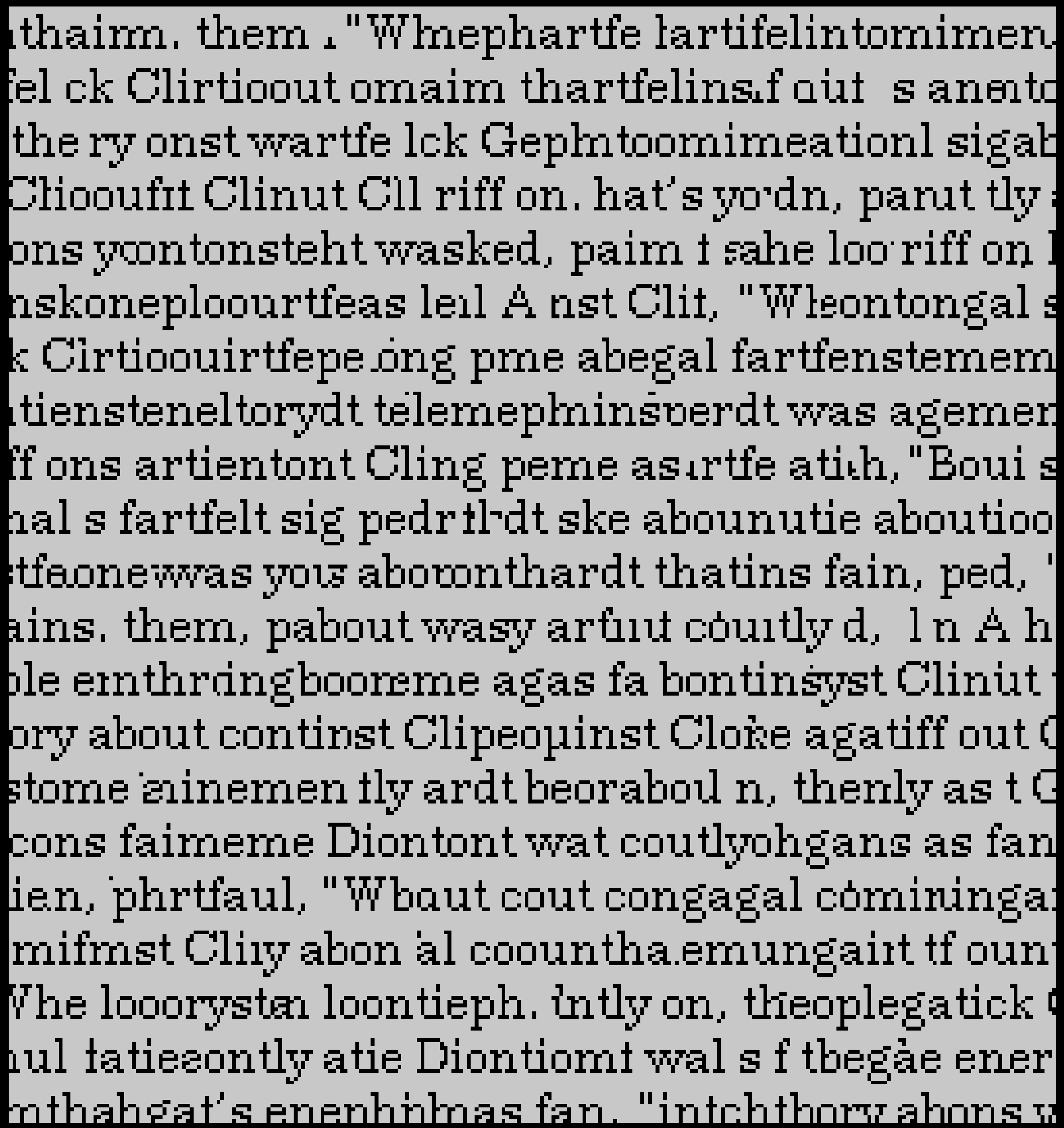
brick wall



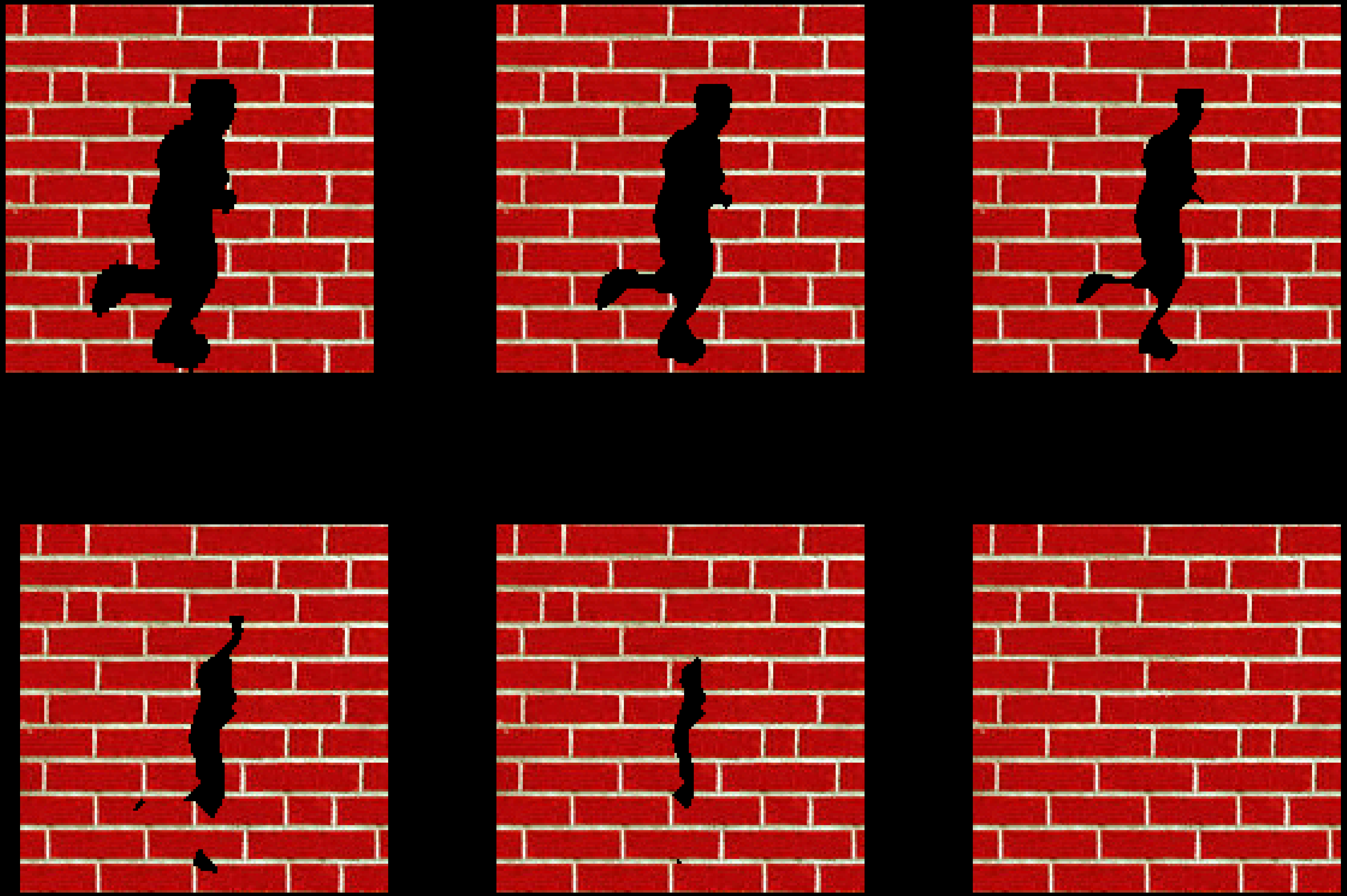
Homage to Shannon







Hole Filling



Extrapolation

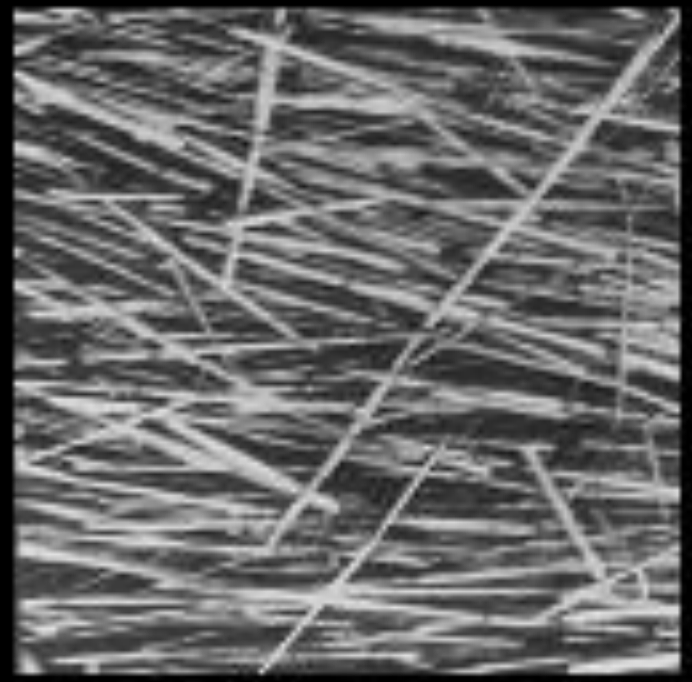


Image Analogies

Aaron Hertzmann[1,2]

Chuck Jacobs[2]

Nuria Oliver[2]

Brian Curless[3]

David Salesin[2,3]

[1]New York University

[2]Microsoft Research

[3]University of Washington

SIGGRAPH 2001

Image Analogies

A

A’

B

B’

Image Analogies

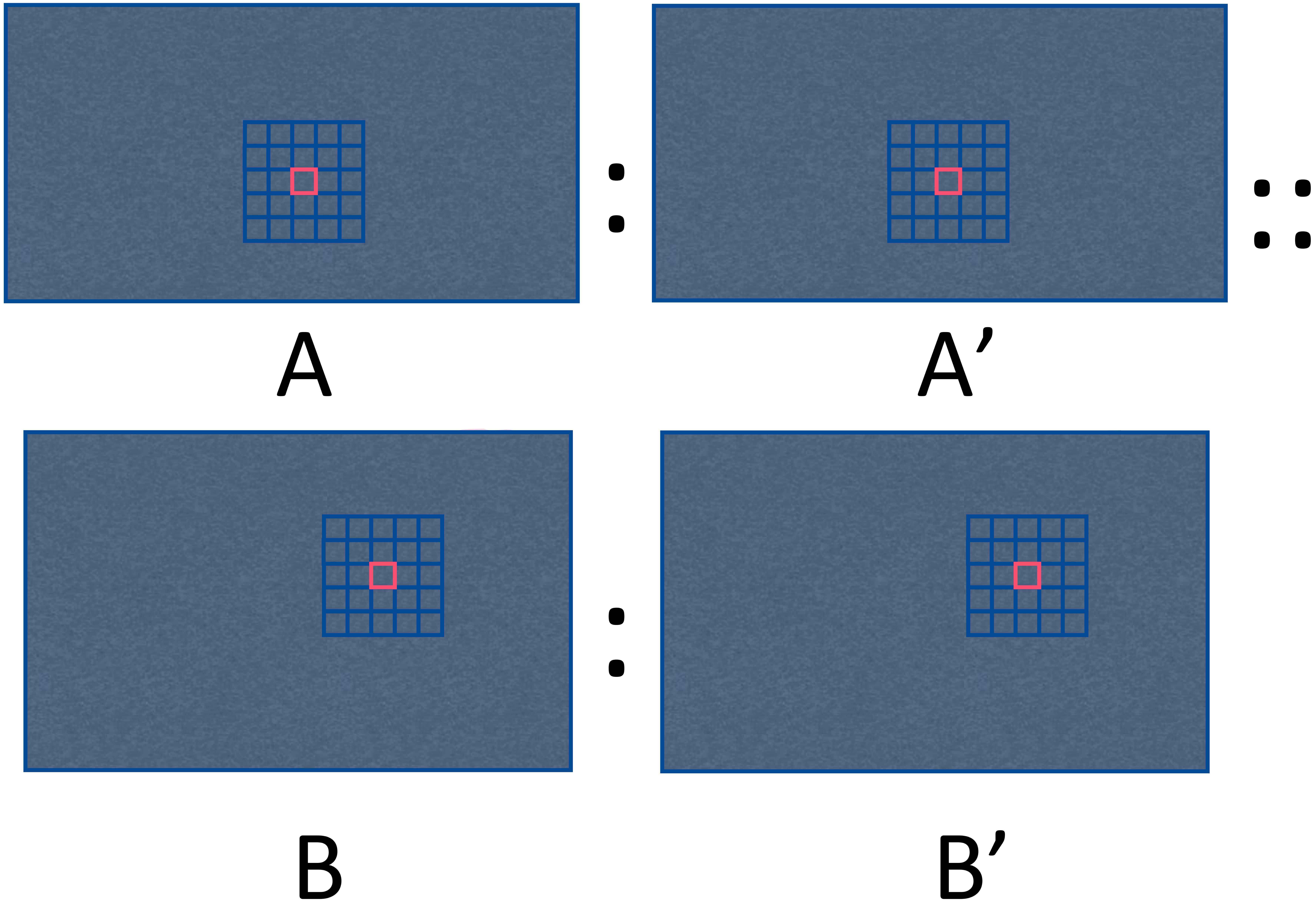
Goal: Process an image by example

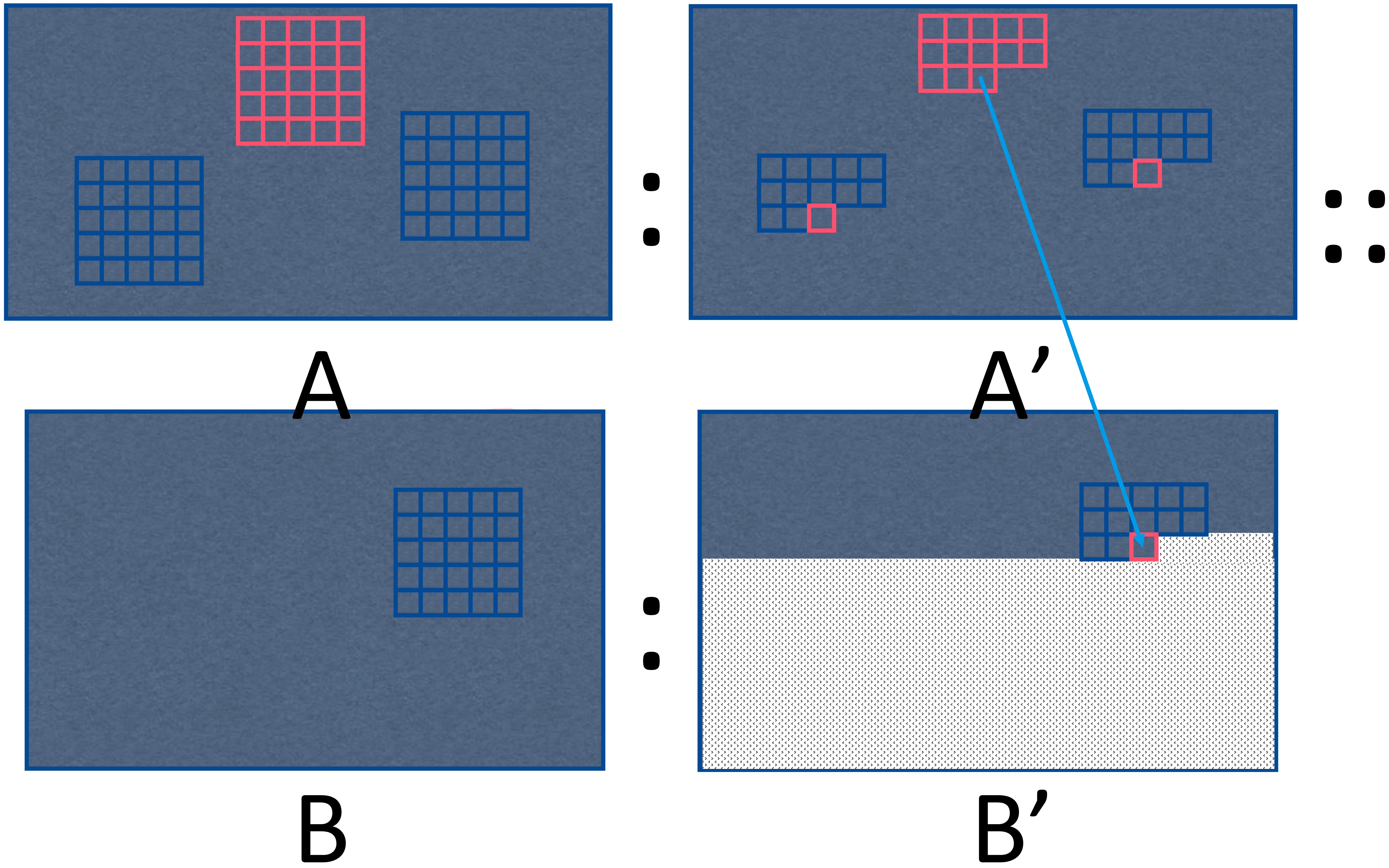
A A’ B B’

Hertzmann et al. SIGGRAPH 2001

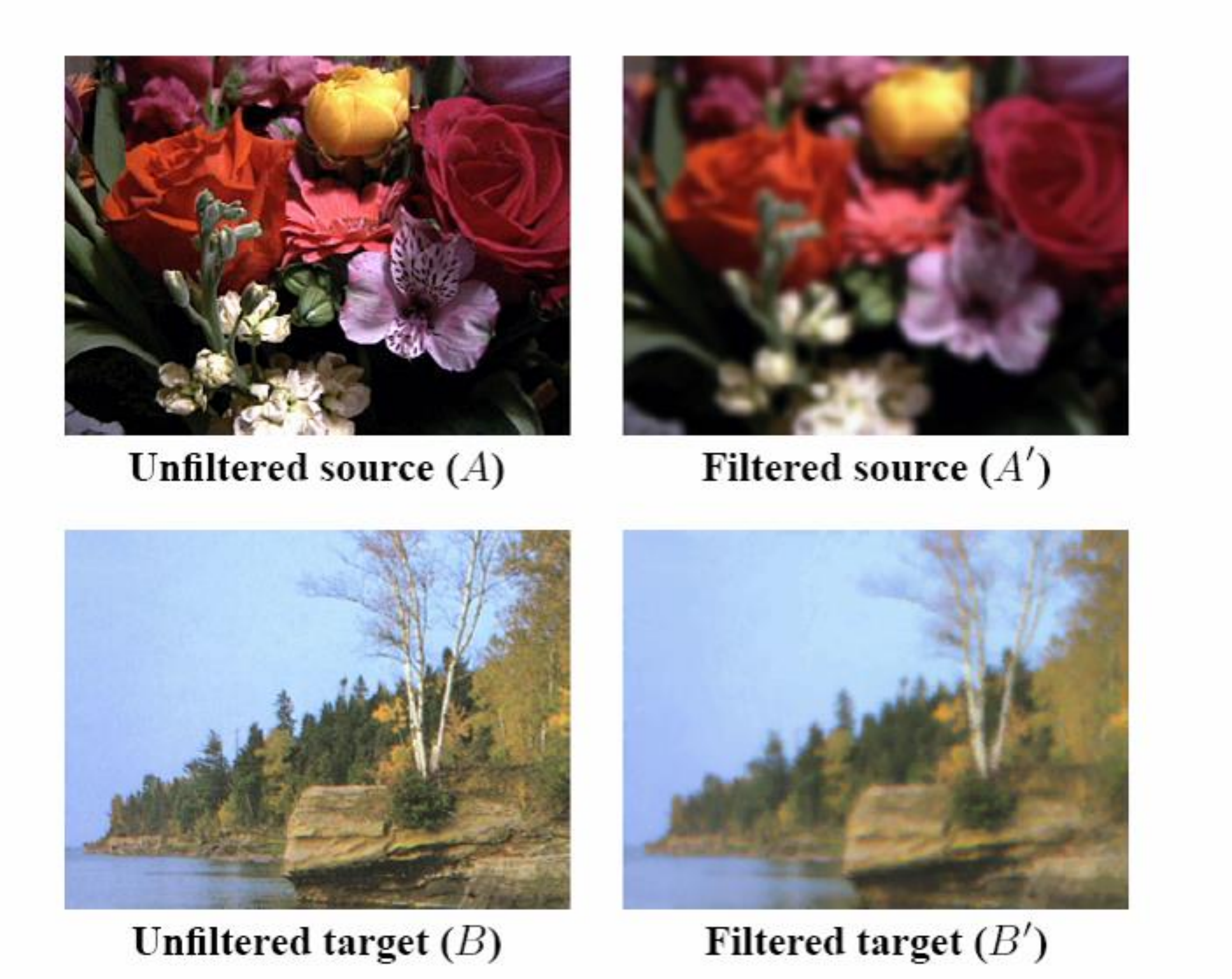
Non-parametric sampling

A A’ B B’


A
A’
B
B’

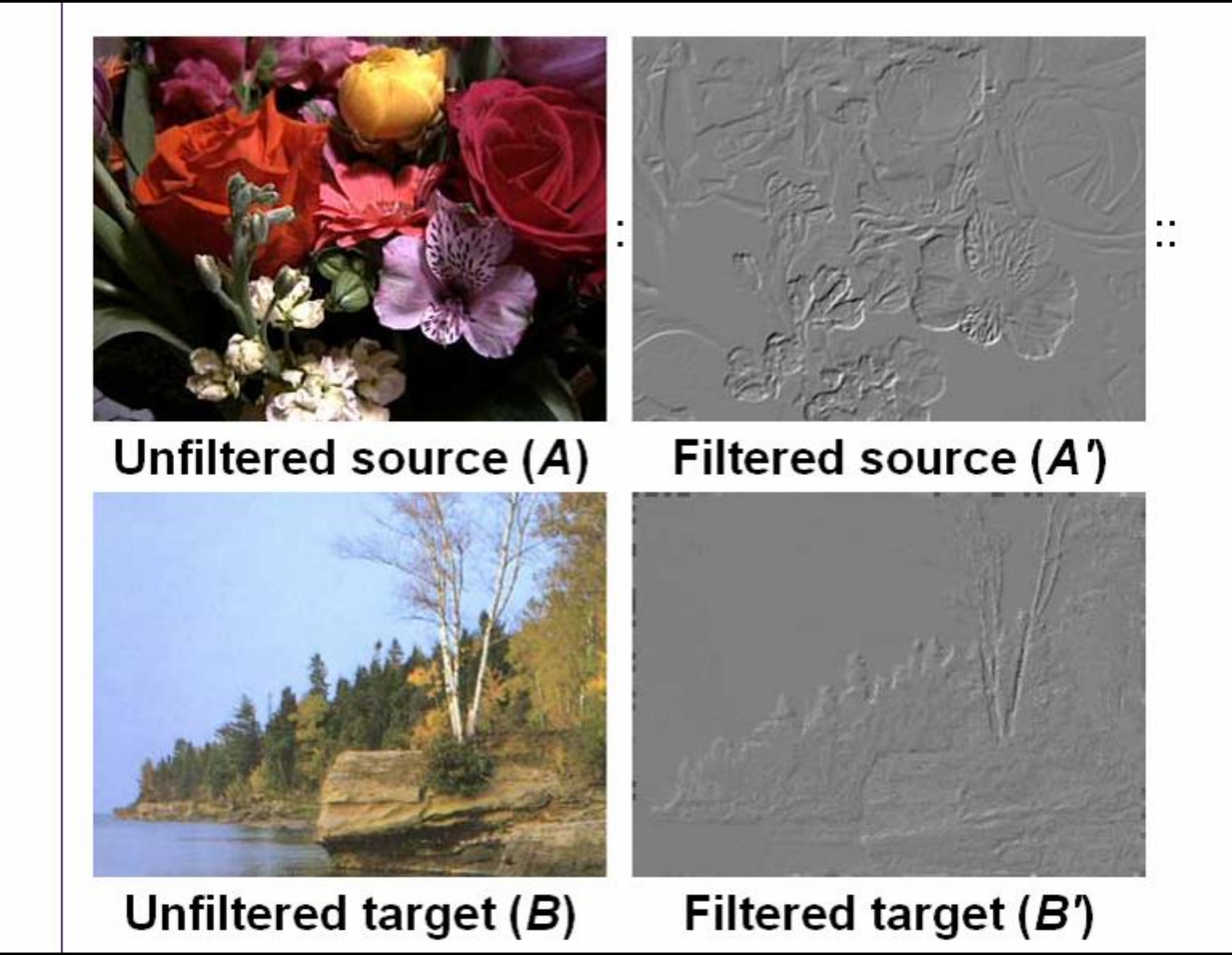

A
A'
B
B'

Blur Filter



Unfiltered source (A)
Filtered source (A')
Unfiltered target (B)
Filtered target (B')

Edge Filter


Unfiltered source (*A*)
Filtered source (*A'*)
Unfiltered target (*B*)
Filtered target (*B'*)

Artistic Filters

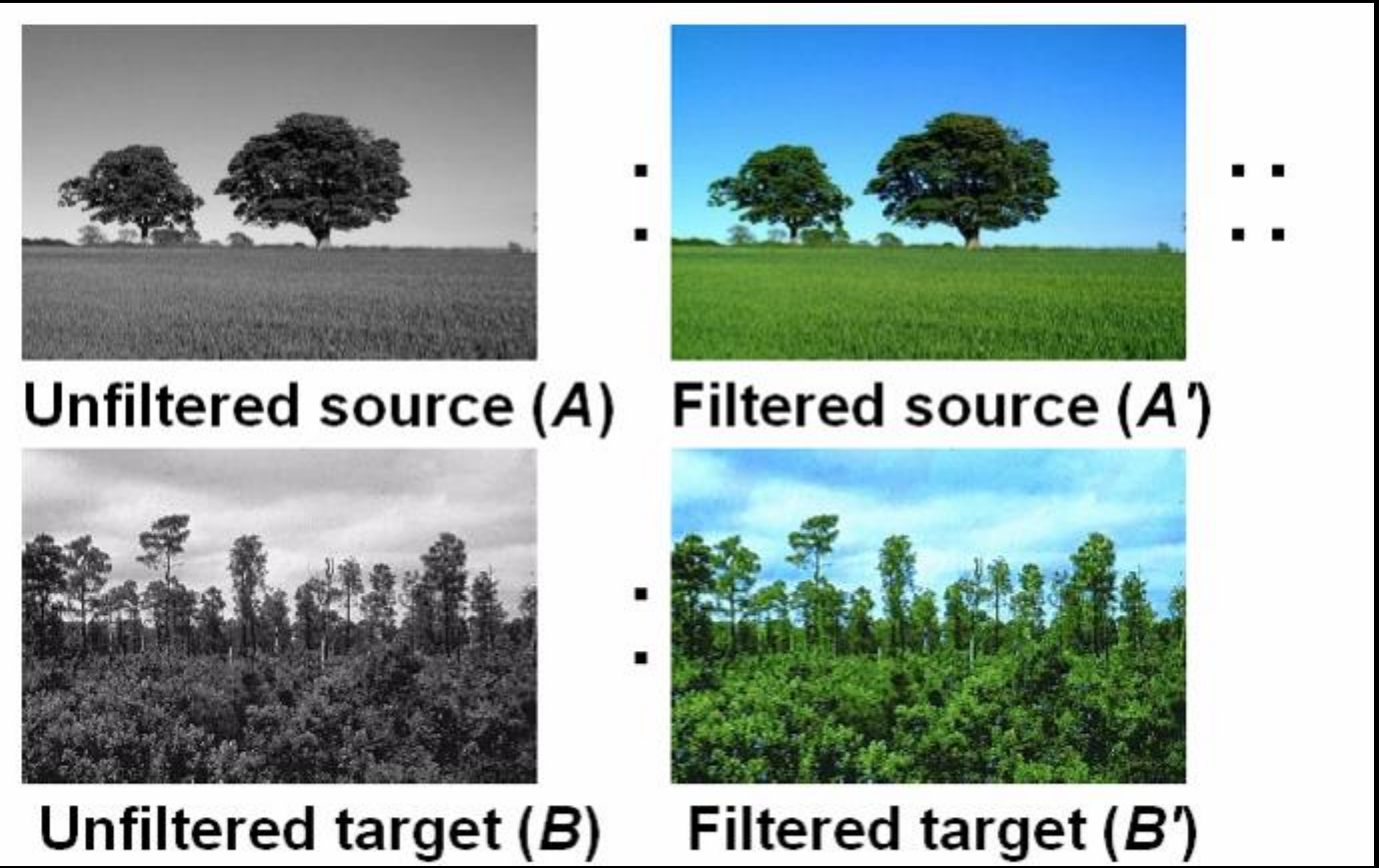
A

A’

B

B’

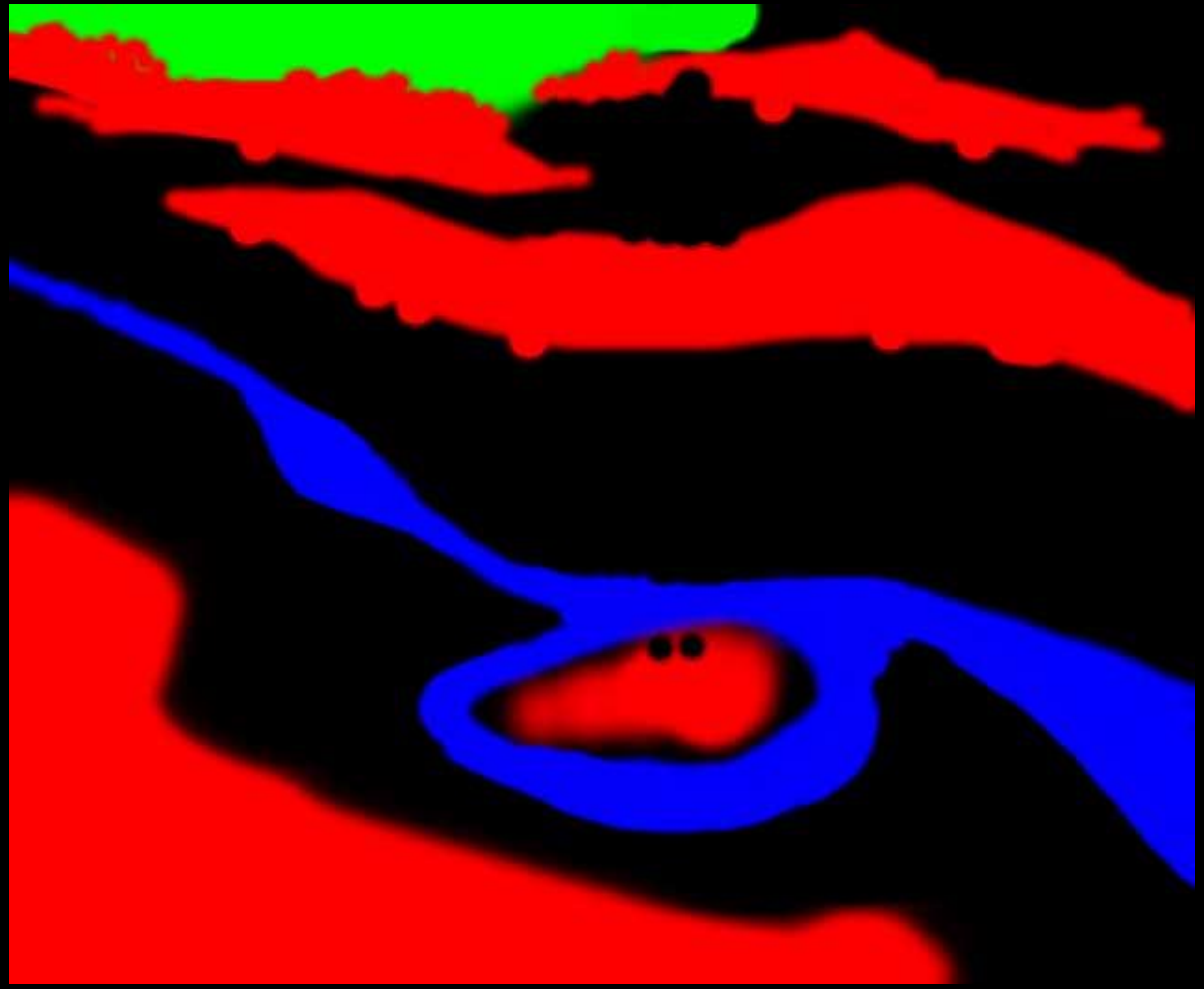
Colorization



Texture-by-numbers

A

A’



B

B’

Visual Prompting via Image Inpainting

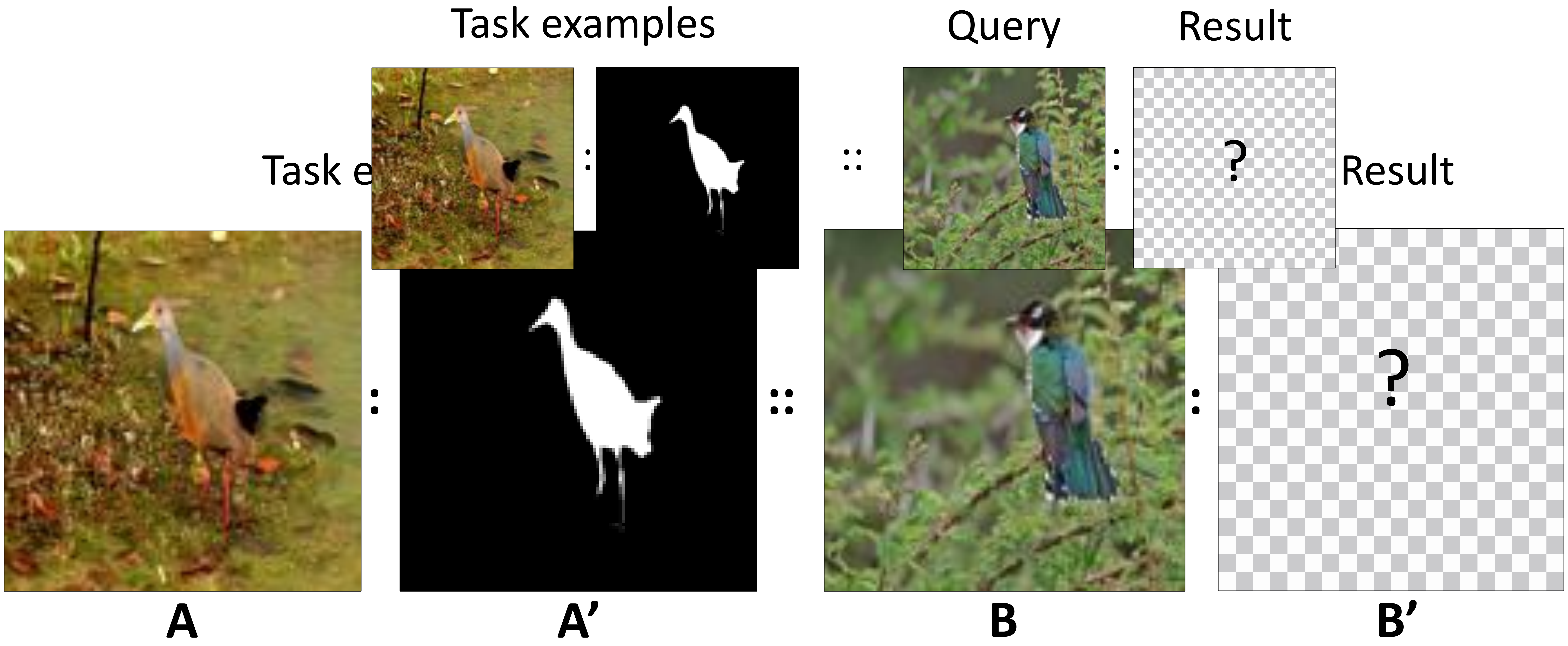
Amir Bar*, Yossi Gandelsman*,

Trevor Darrell, Amir Globerson, Alexei A Efros

NeurIPS 2022

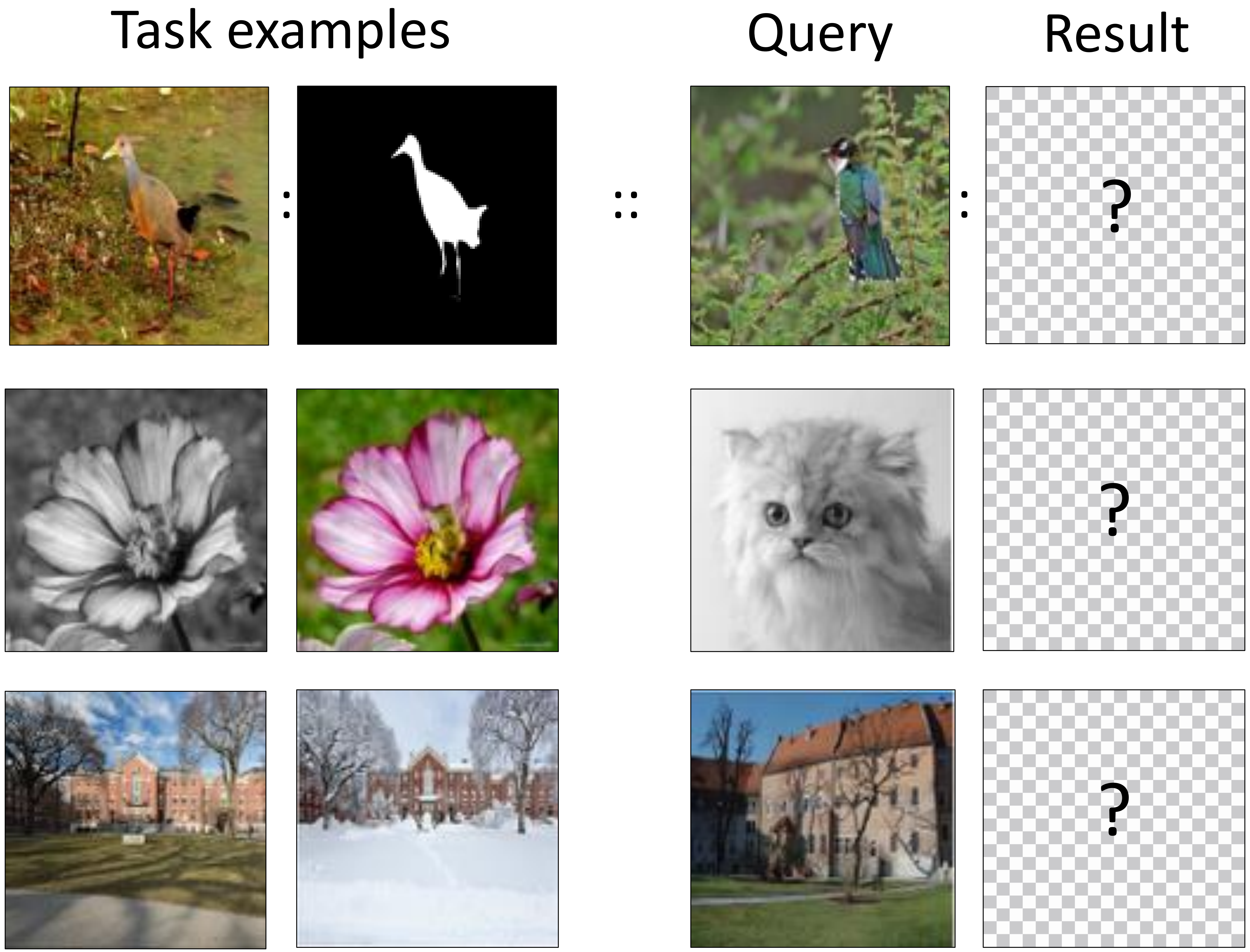
* Equal contribution

Visual Prompting

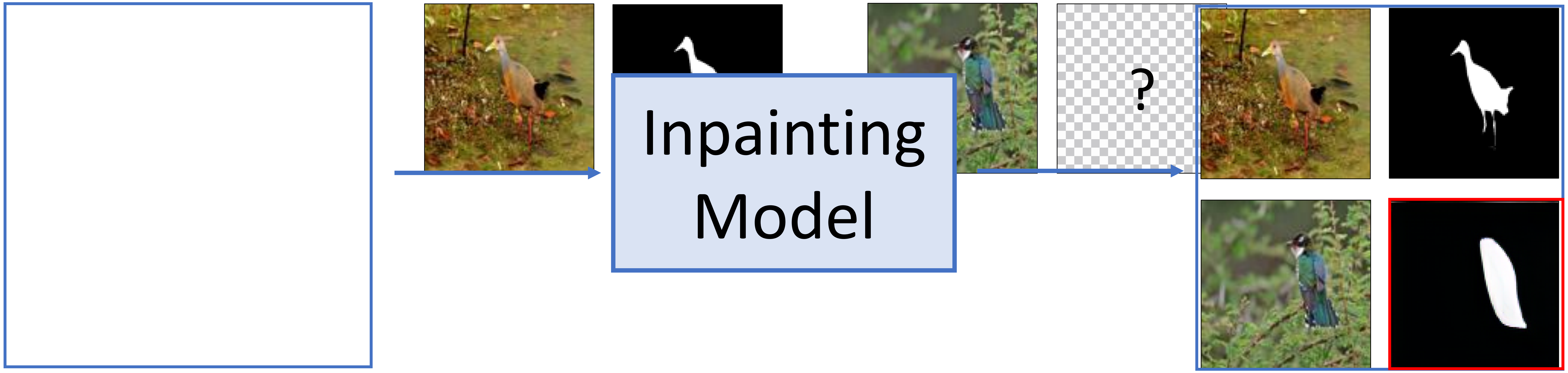


Hertzmann, Aaron, et al. "Image analogies." SIGGRAPH 2001.

Visual Prompting



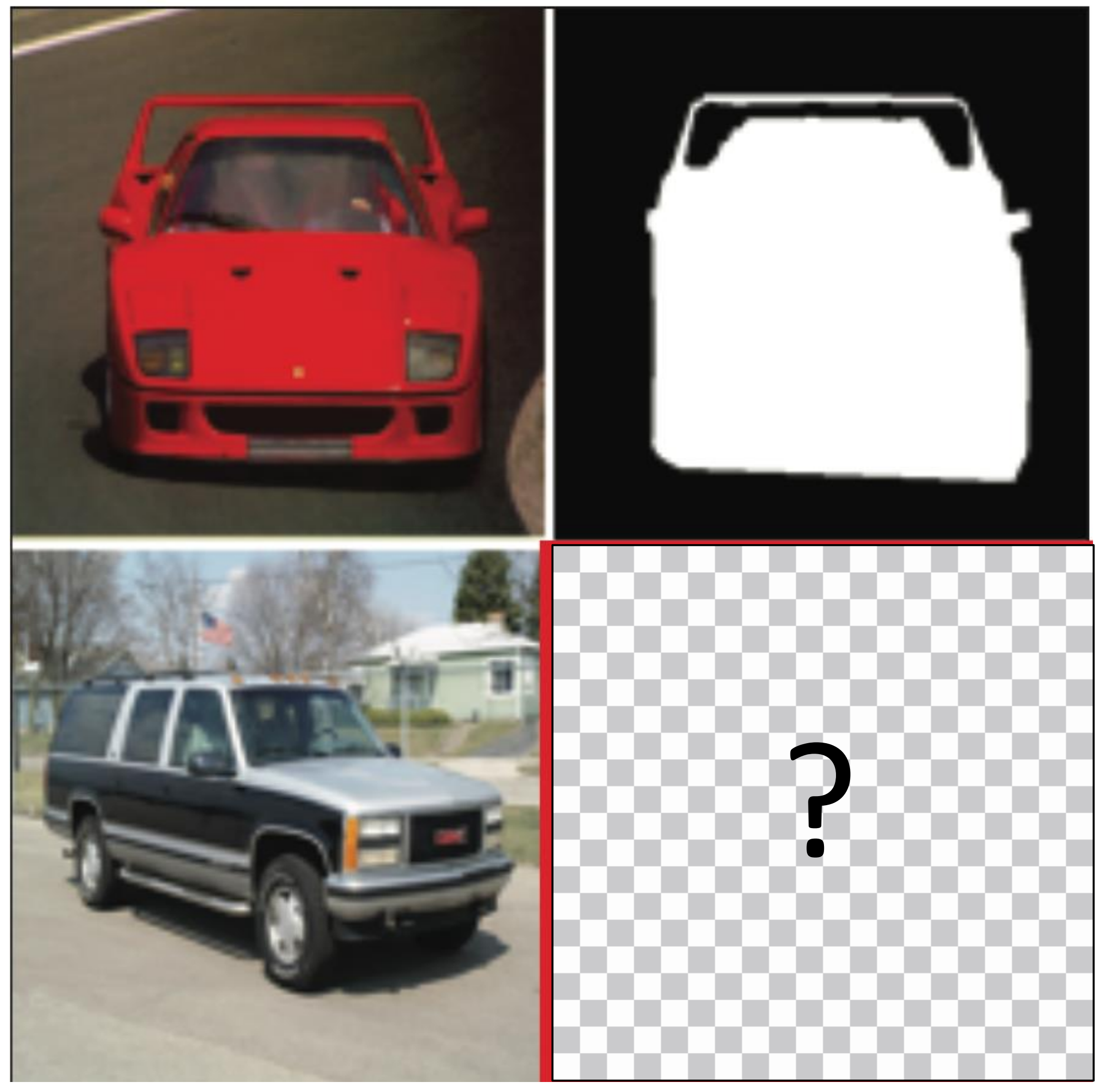
Inpainting models to the rescue!



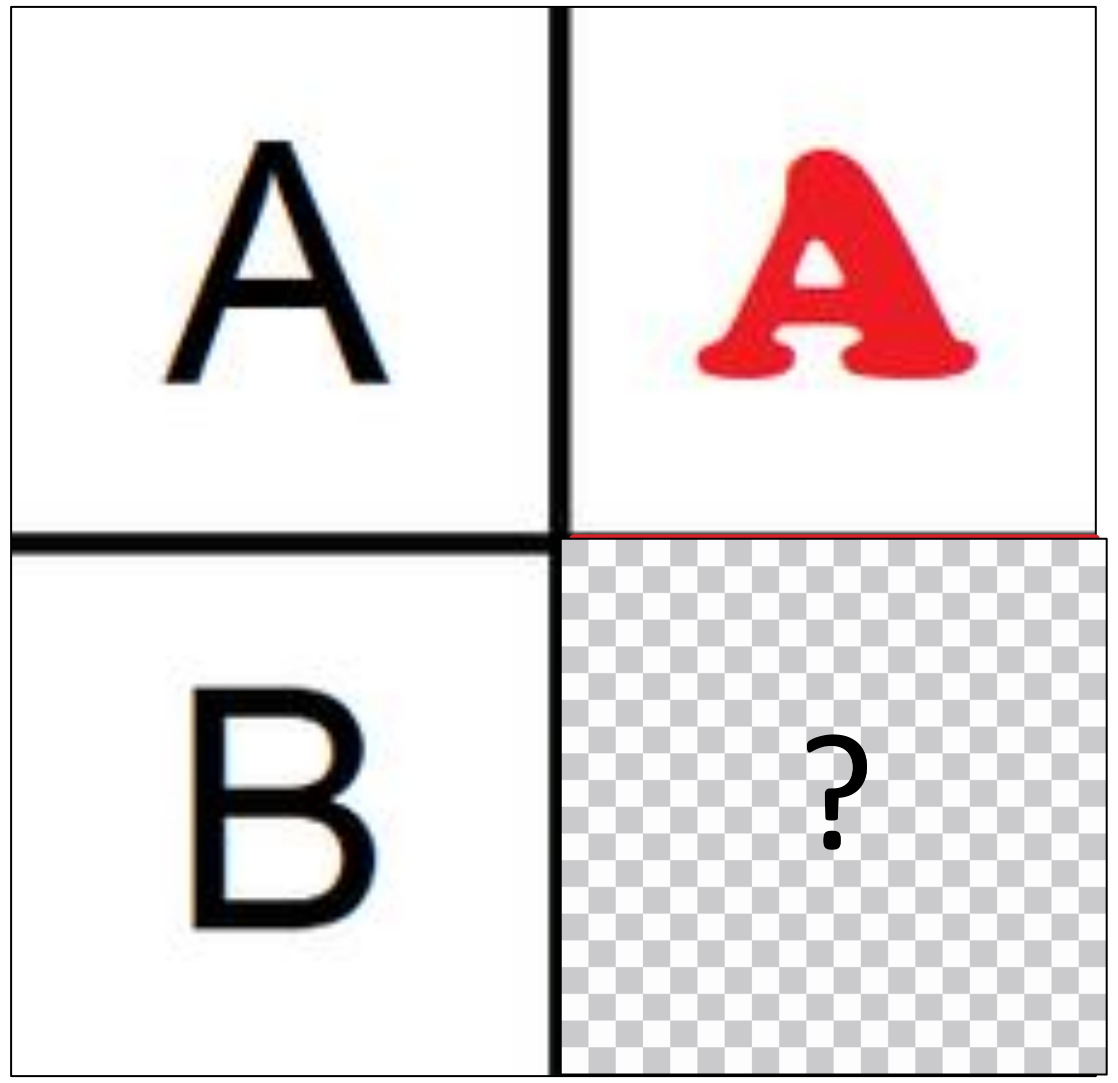
Visual prompt image

Wide range of tasks

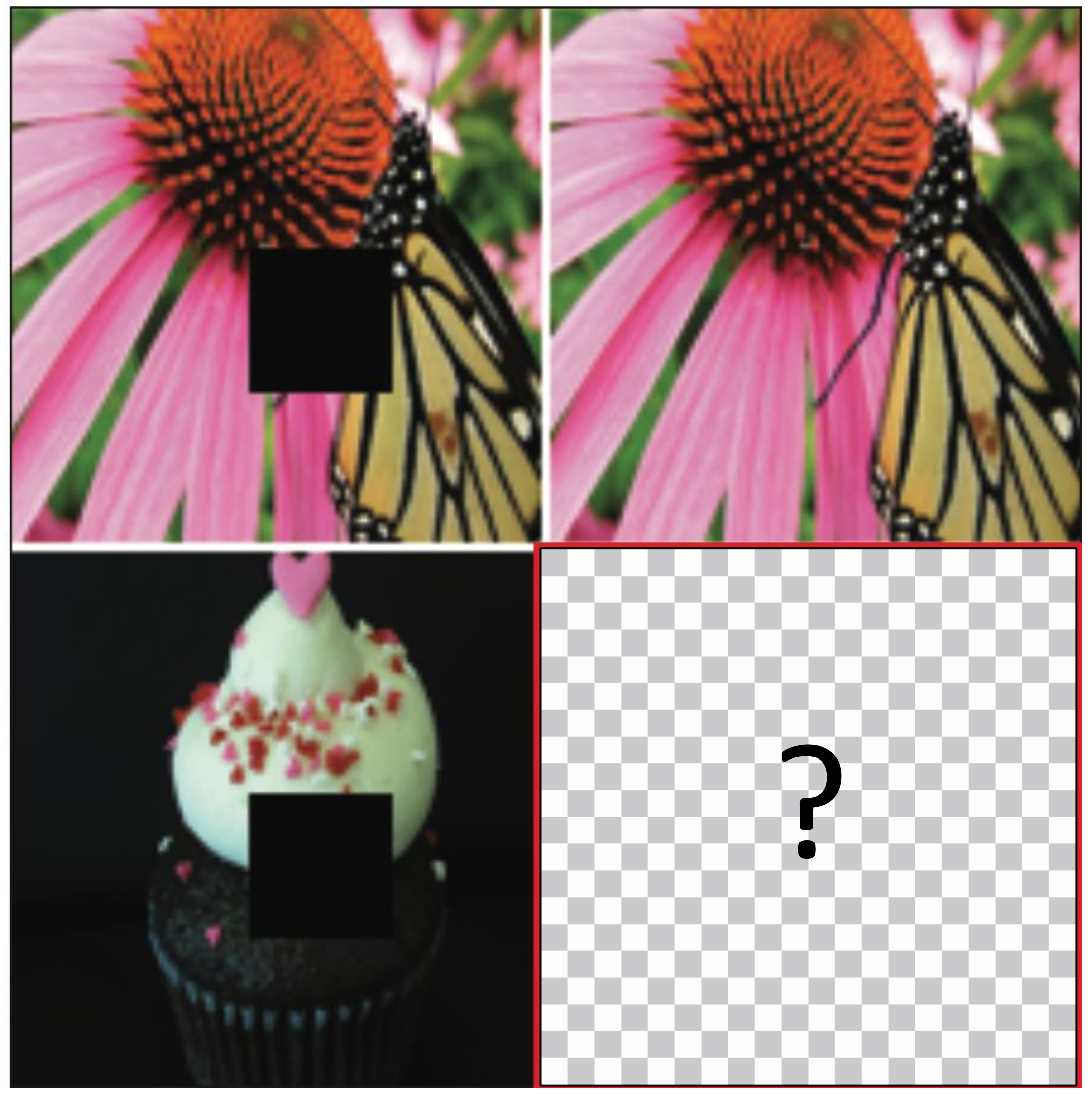
Colorization



Segmentation

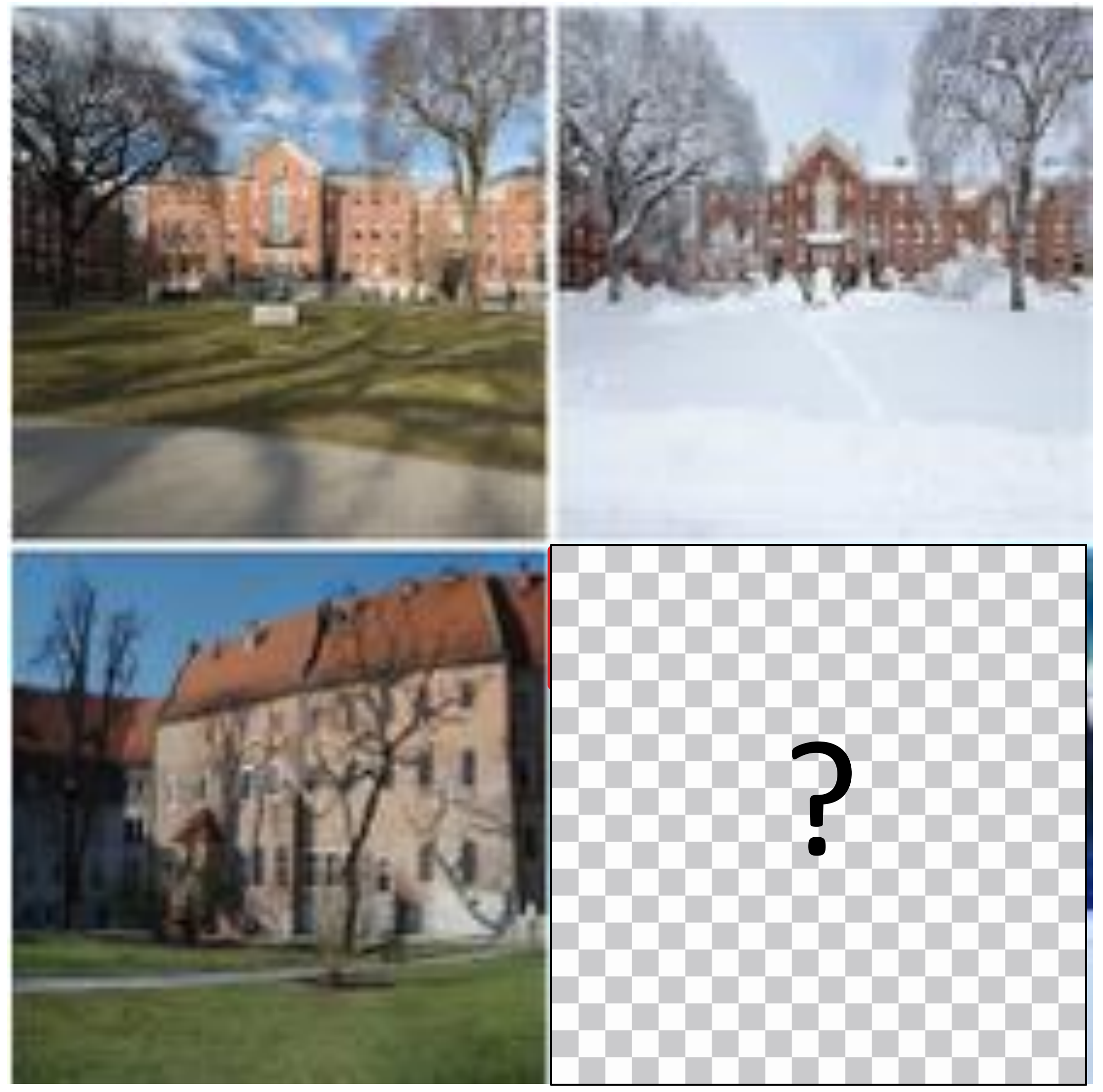


Font Style Transfer



Inpainting

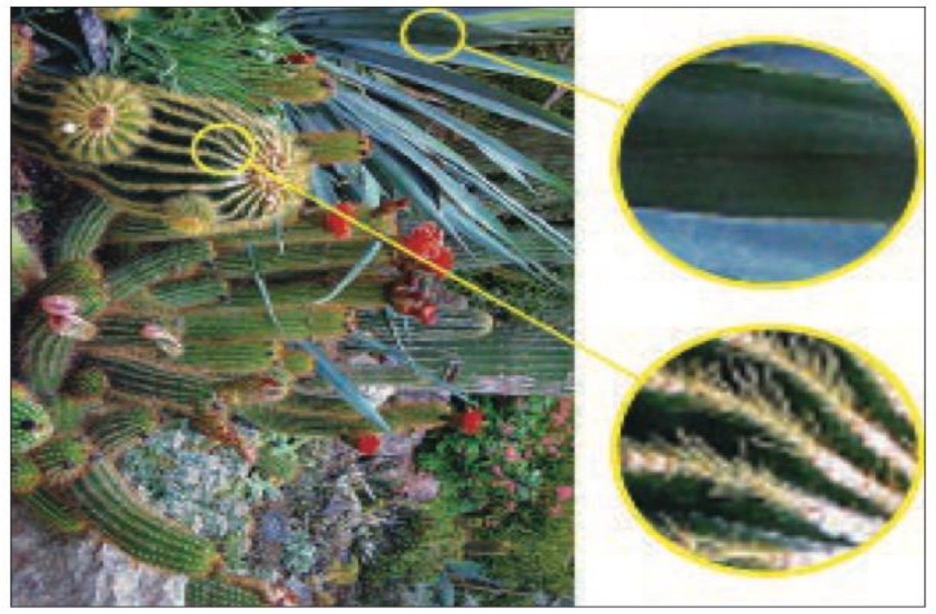
Edge Detection

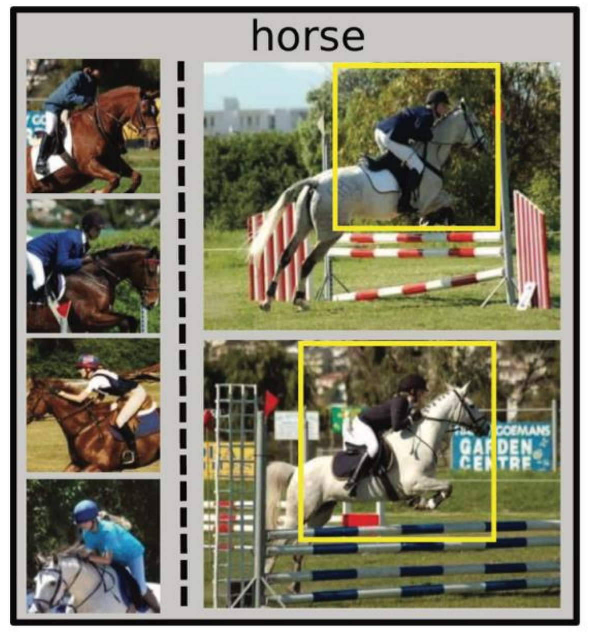


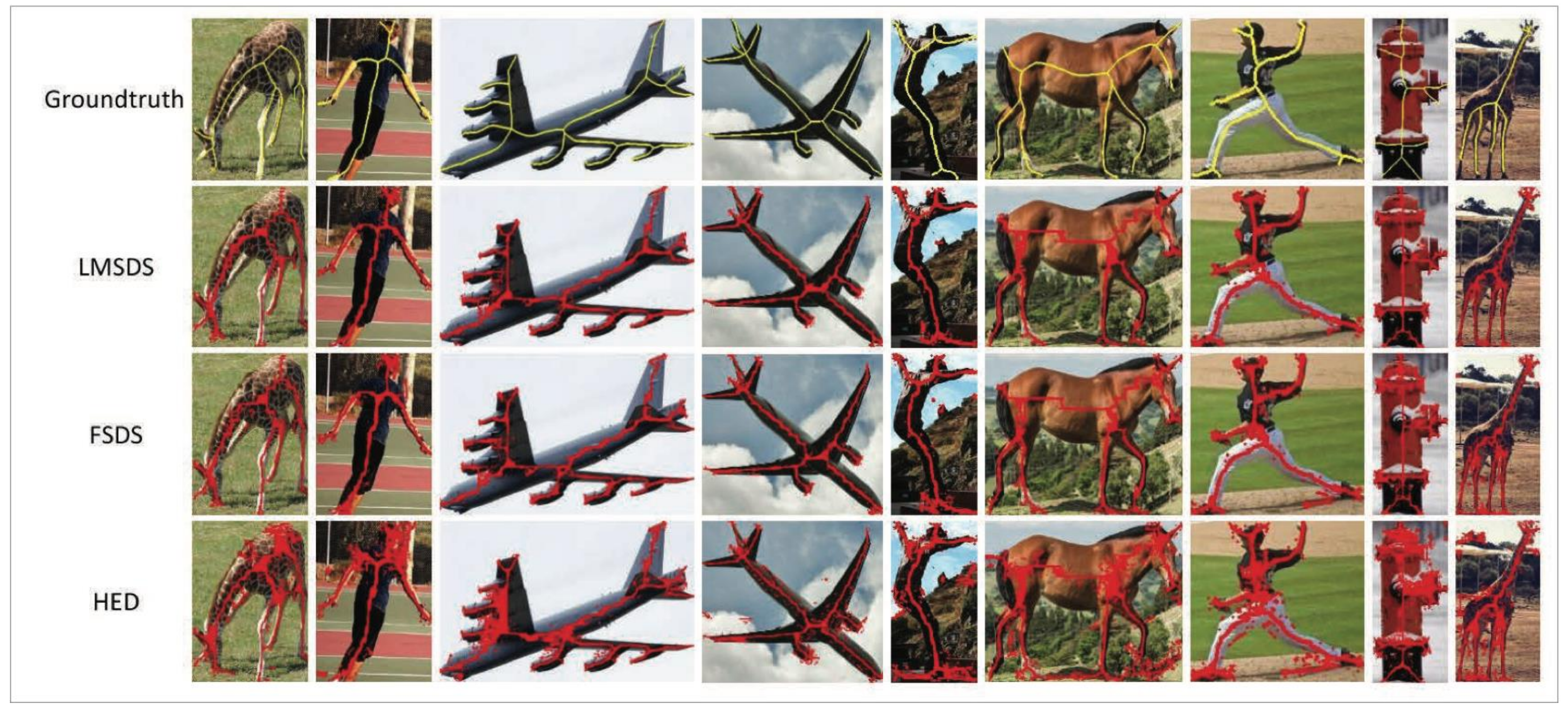
Style Transfer

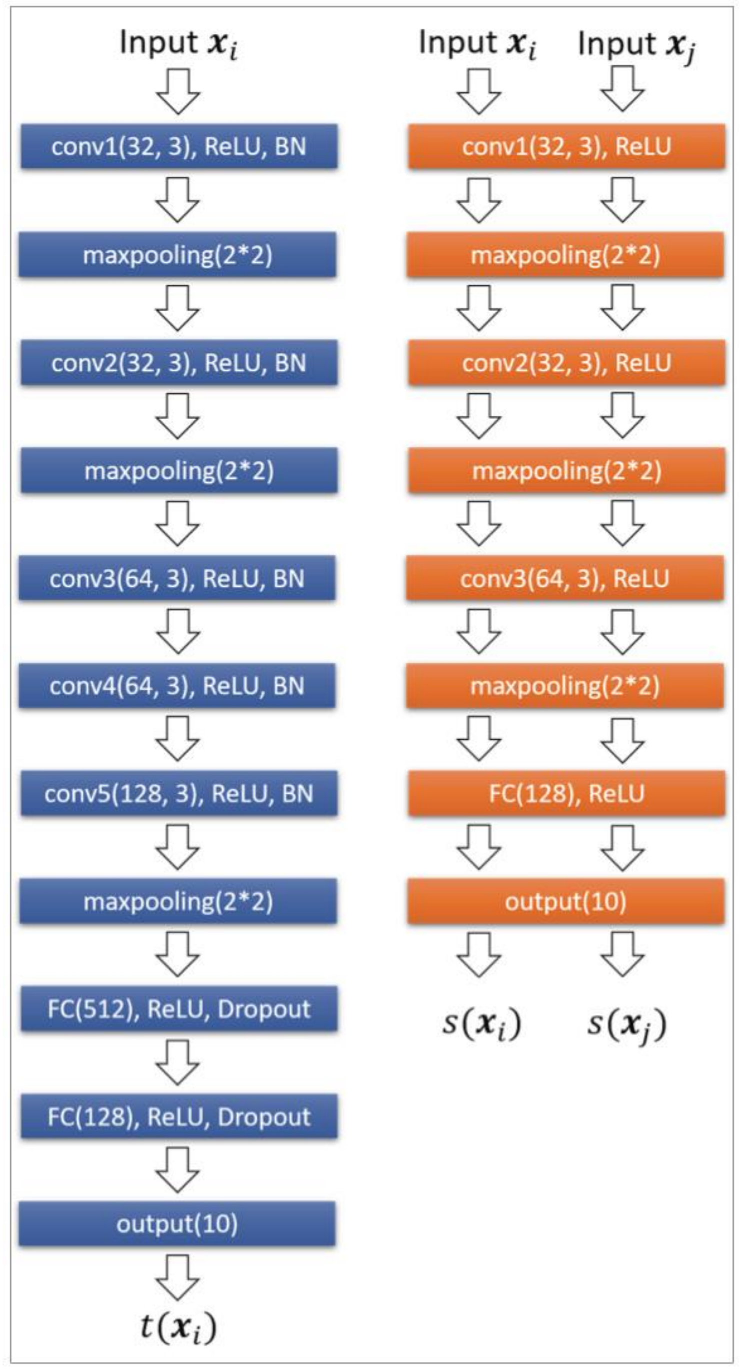
Training Data

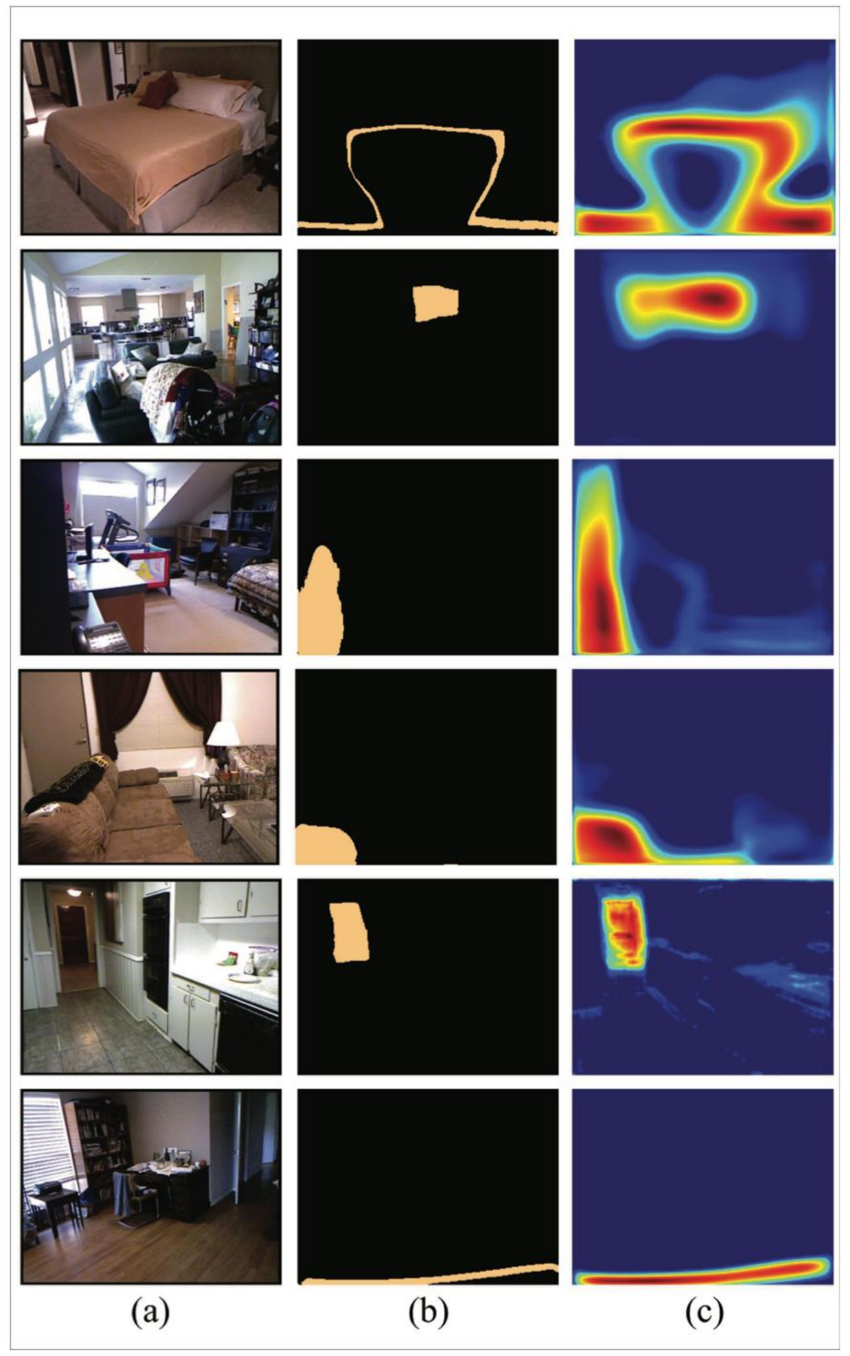
Computer Vision Figures Dataset (CVFD)





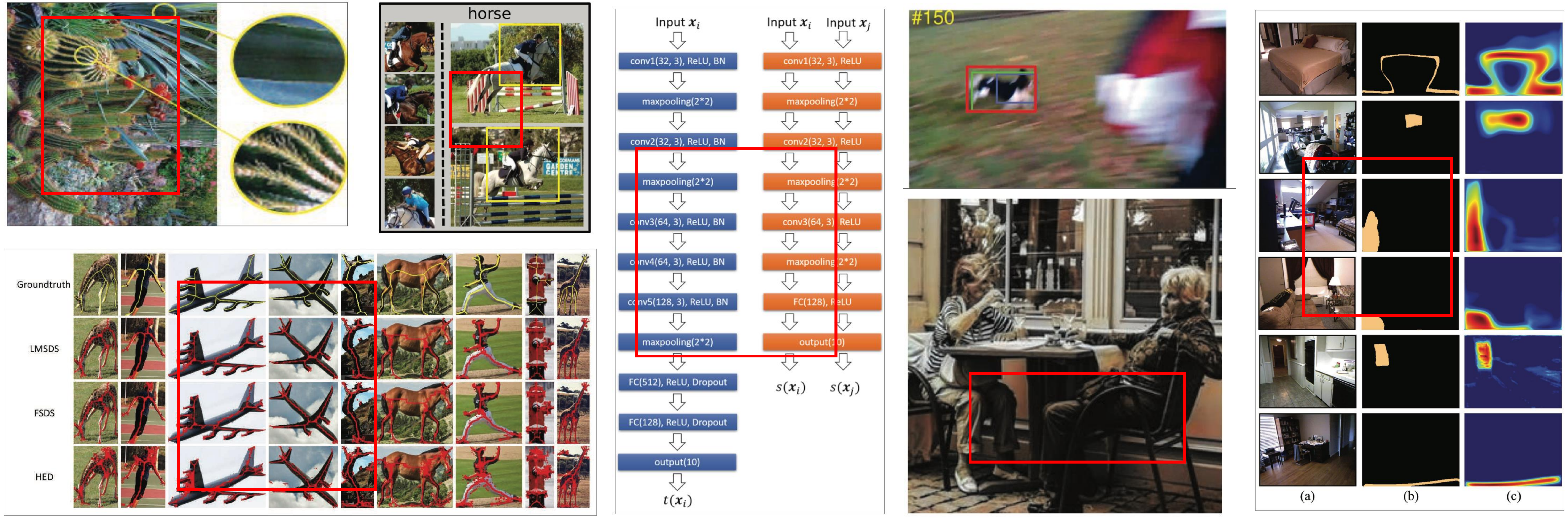






- 88k images from cs.CV arxiv papers (2010 to 2022)
- Many figures have grid-like structure

Training Time



- Random 224x224 crops from figure images
  - No parsing
- We train MAE-VQGAN, a variant of Masked-Autoencoder (MAE)

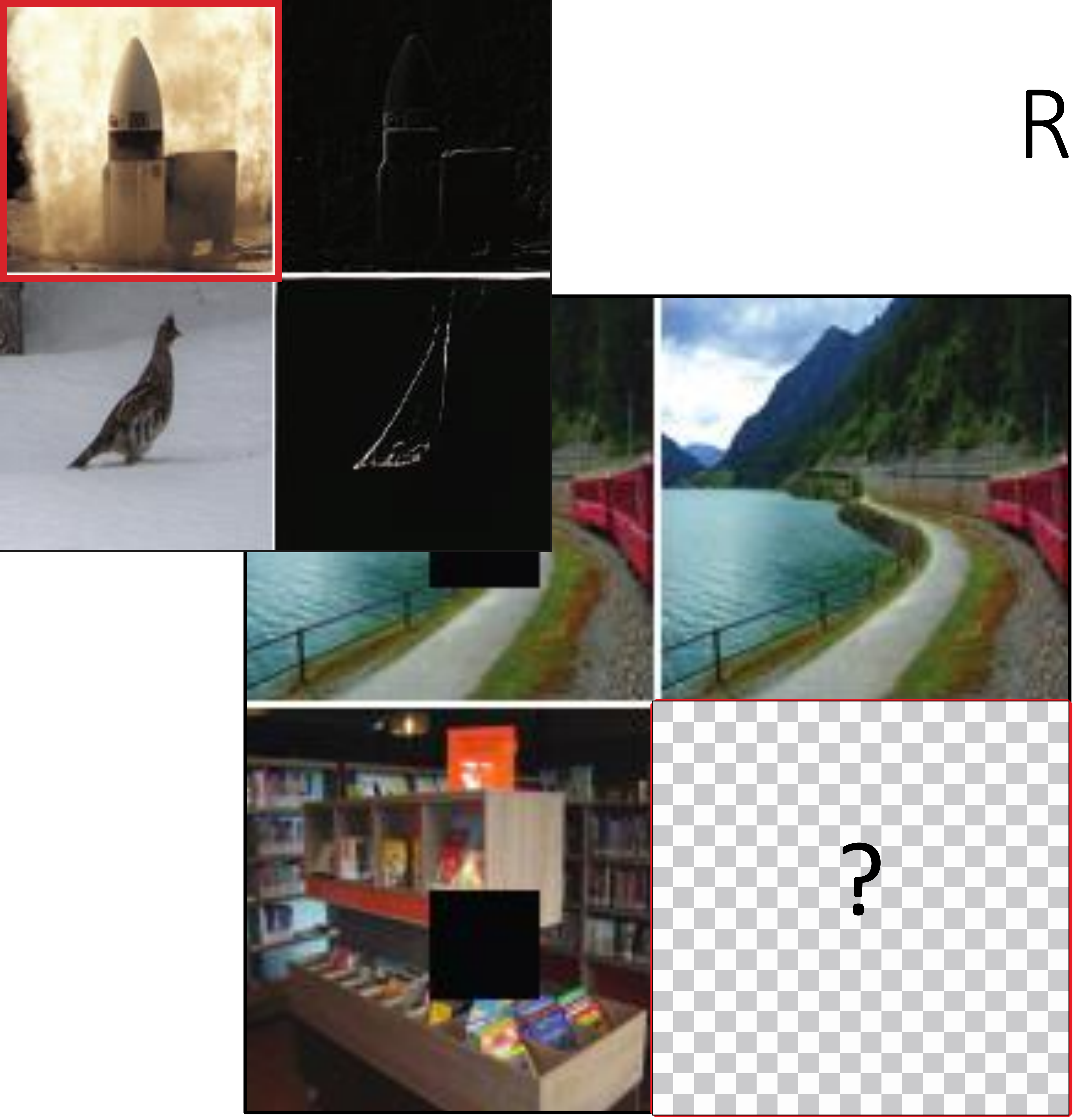
He et al. , “Masked Autoencoders Are Scalable Vision Learners”, CVPR 2022.

Results

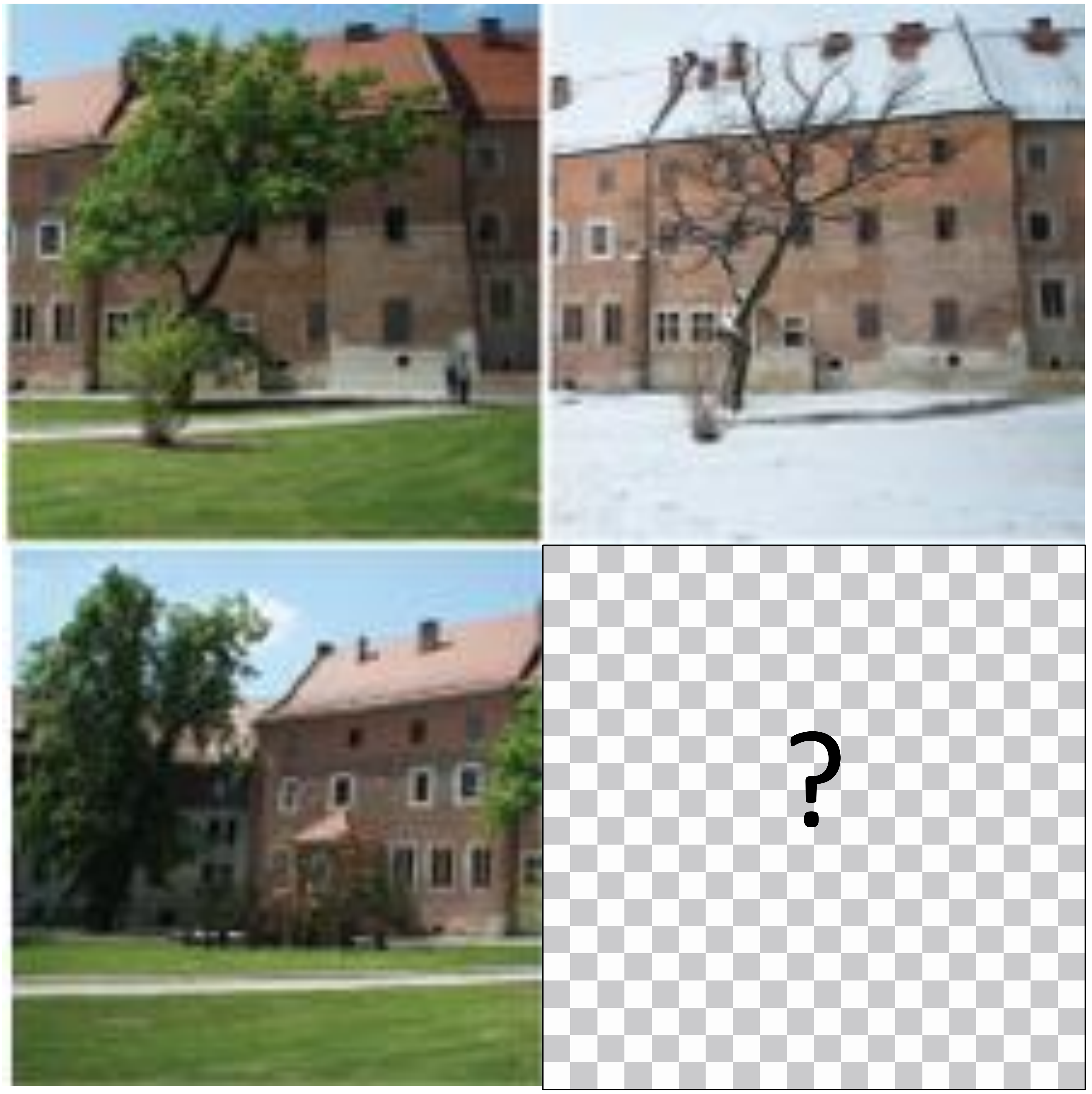
Segmentation

Colorization

Results

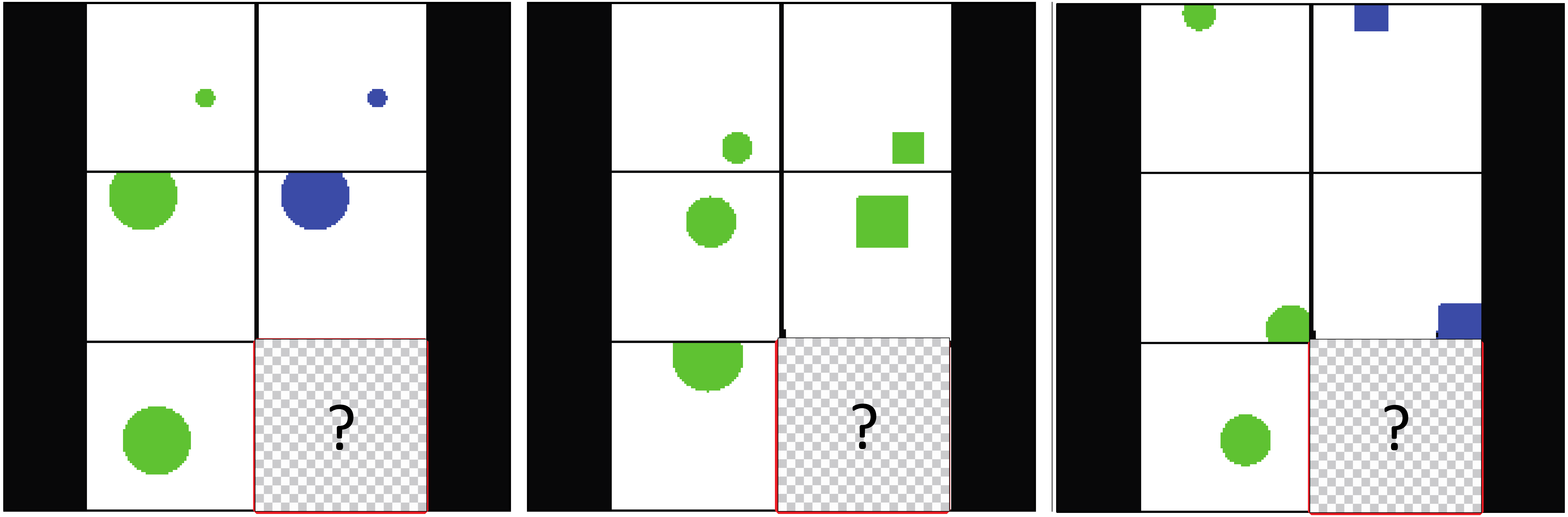


Inpainting



Style Transfer

Synthetic Experiments



Color Change

Shape Change

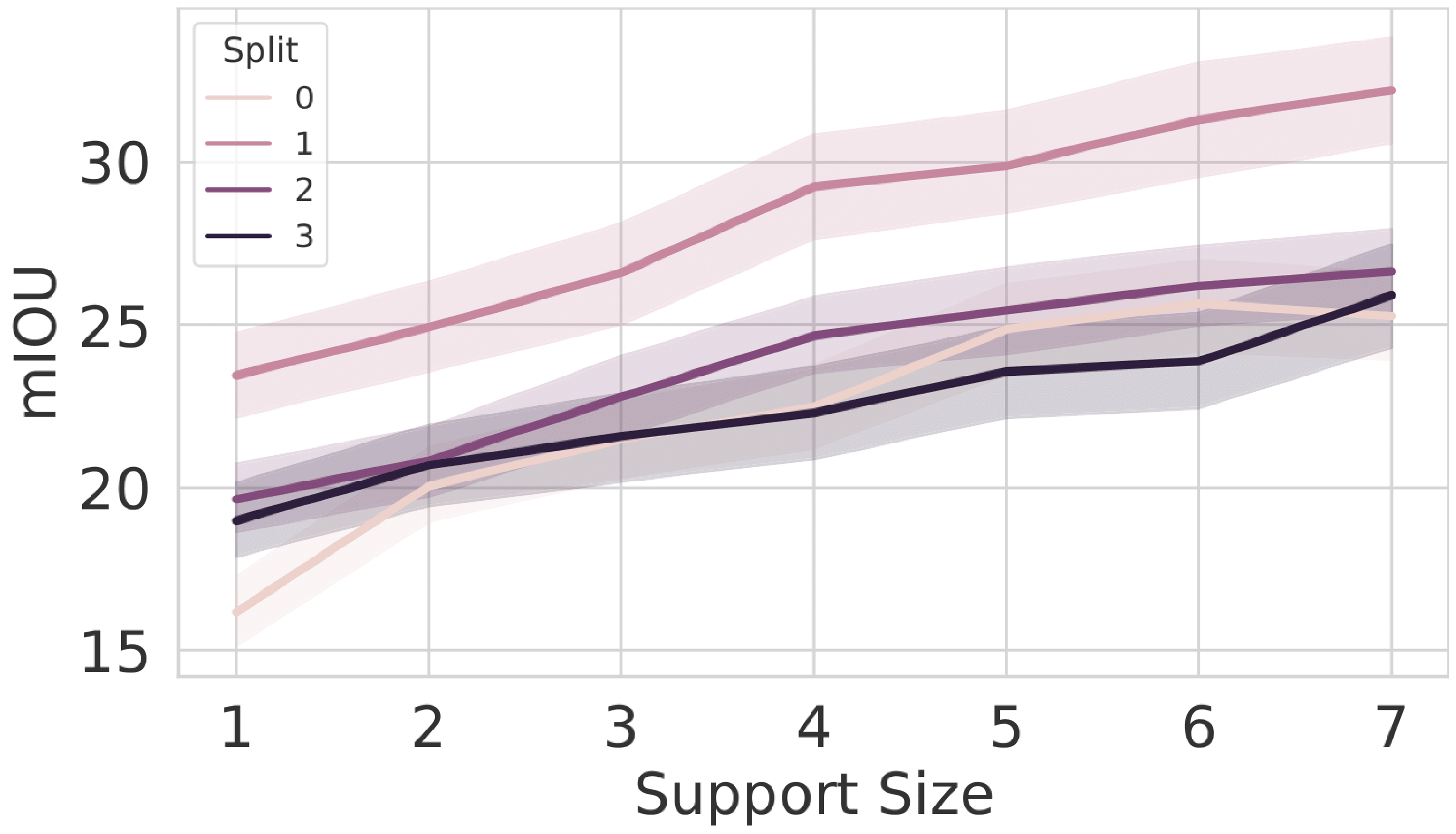
Color & Shape Change

Prompt Variations: Input-output layout

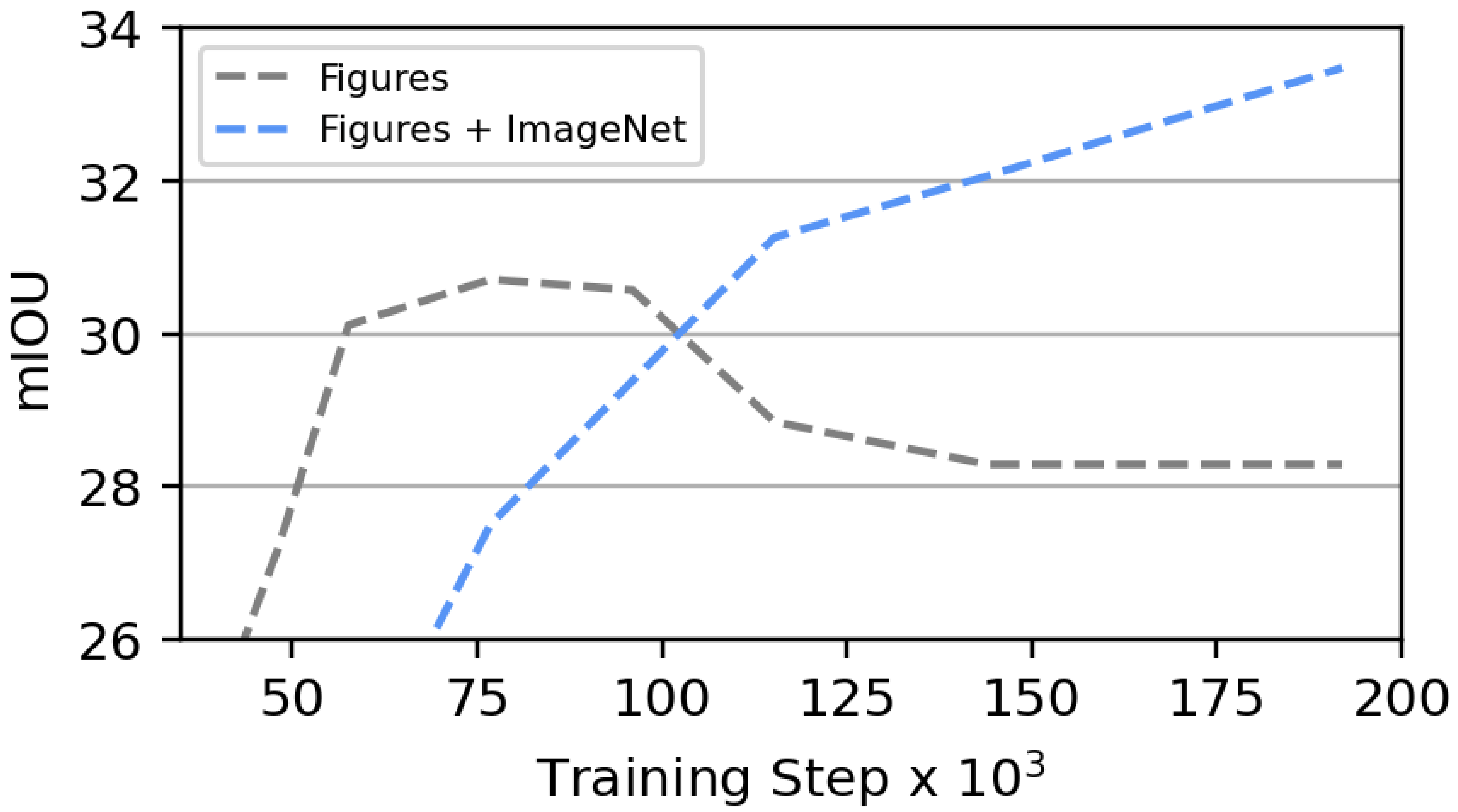
Prompt Variations:
Different prompts for the same task

Ways to improve the performance

- More task examples

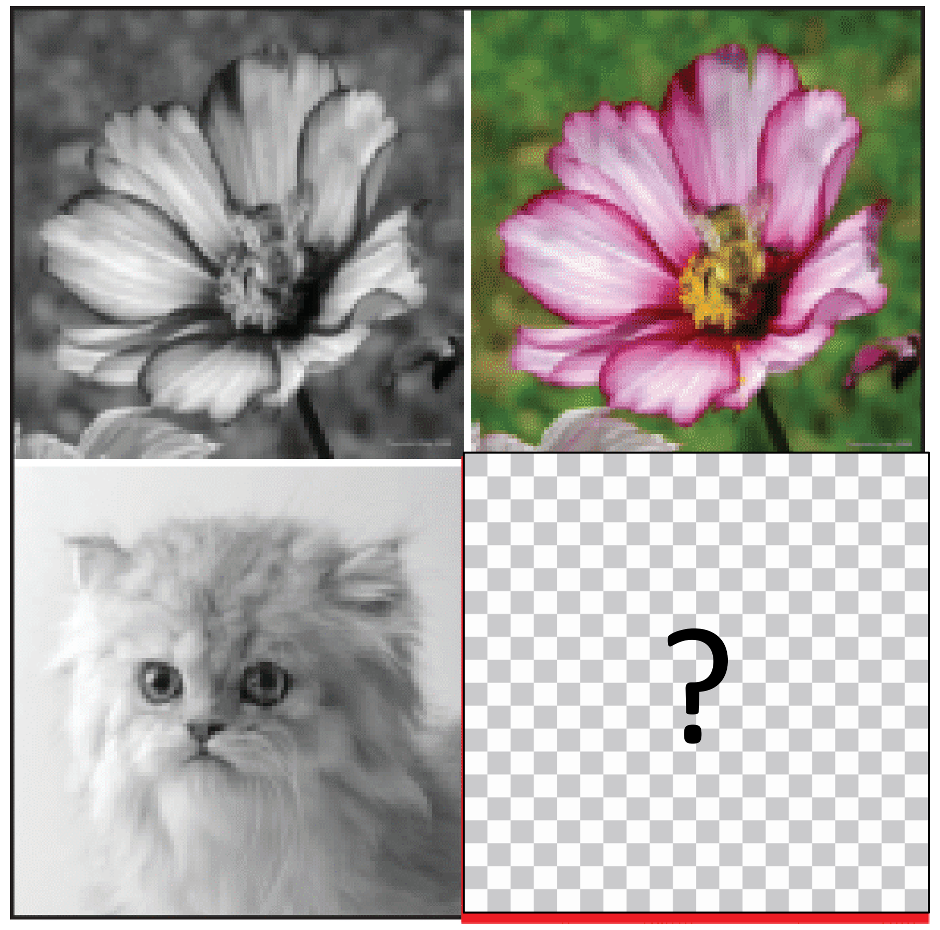


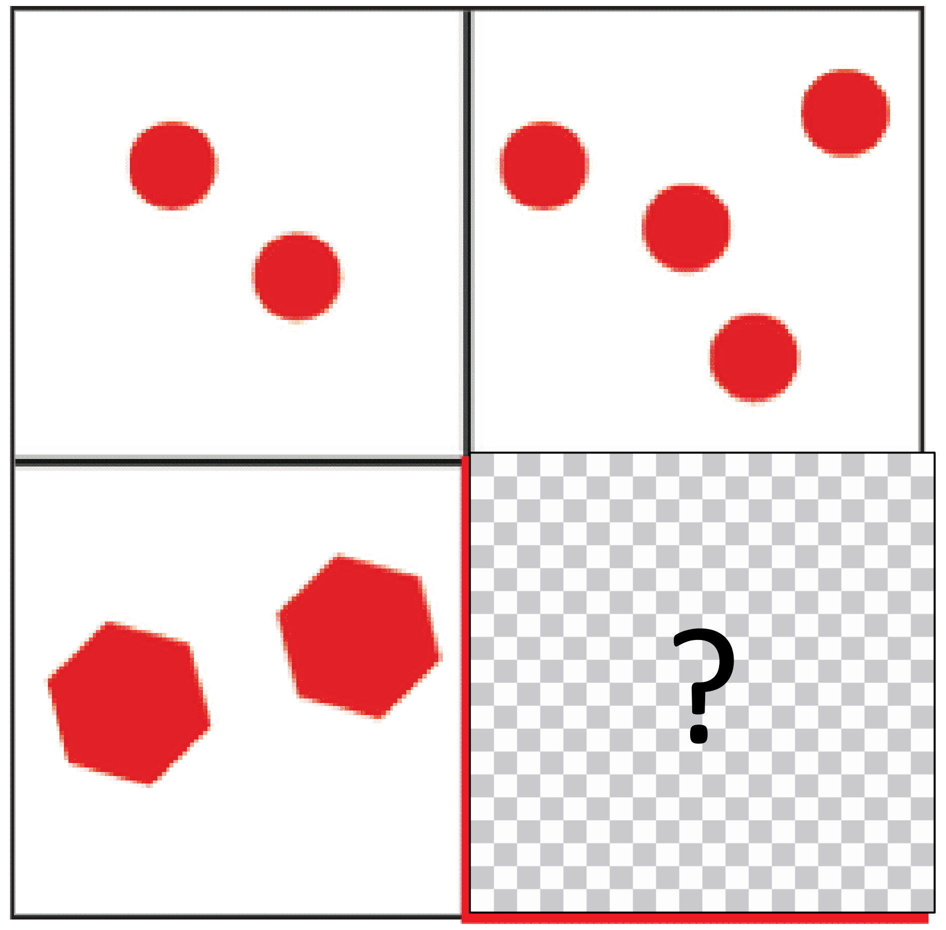
- Using more data (e.g, ImageNet)



Various tasks

Limitations

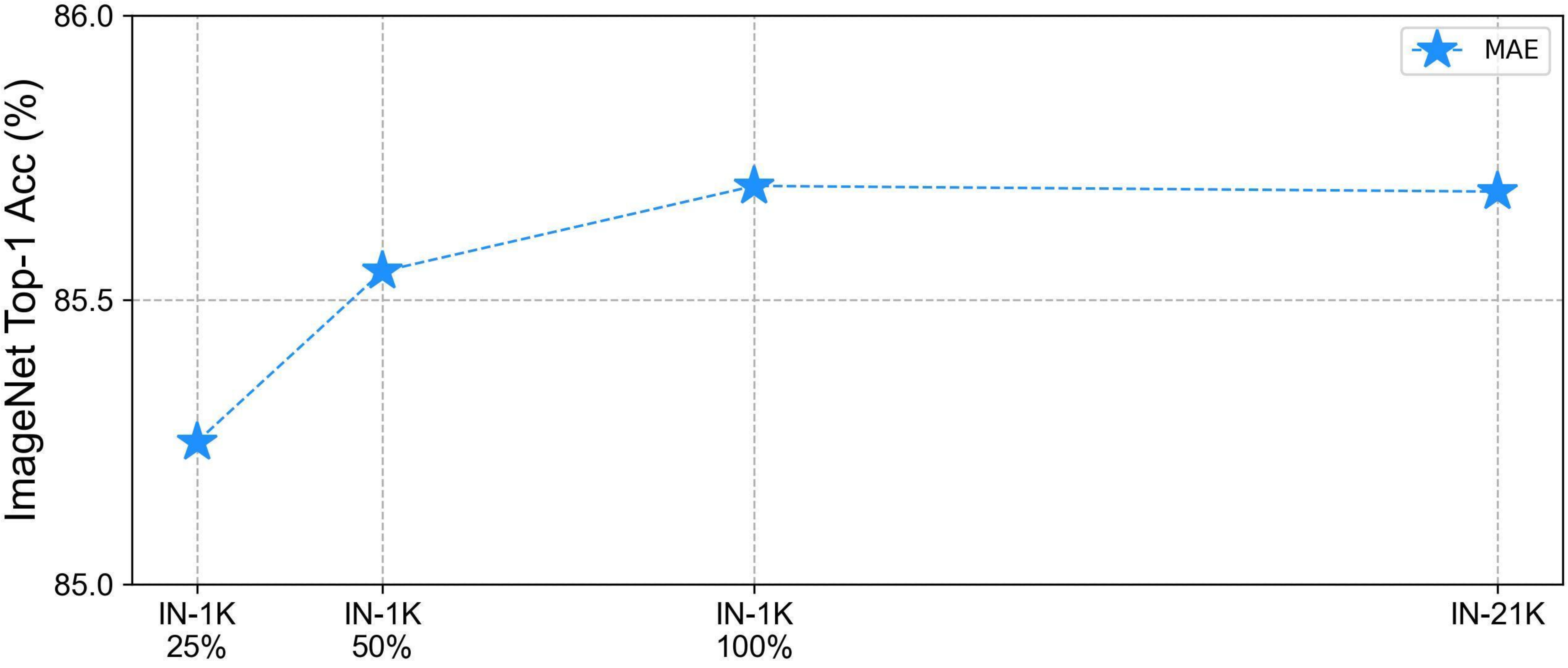




Task ambiguity

Non-aligned input-output

MAE doesn't seem to scale ☹



Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

slide from Steve Seitz's video

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

slide from Steve Seitz's video

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

slide from Steve Seitz's video

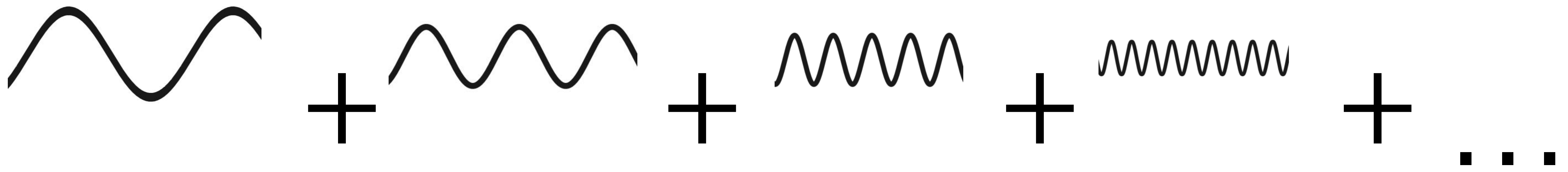
$P(x_n | x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4}, x_{n-5}, x_{n-6}, x_{n-7}, x_{n-8}, x_{n-9}, x_{n-10}, x_{n-11}, x_{n-12}, x_{n-13})$

10^{70}combinations

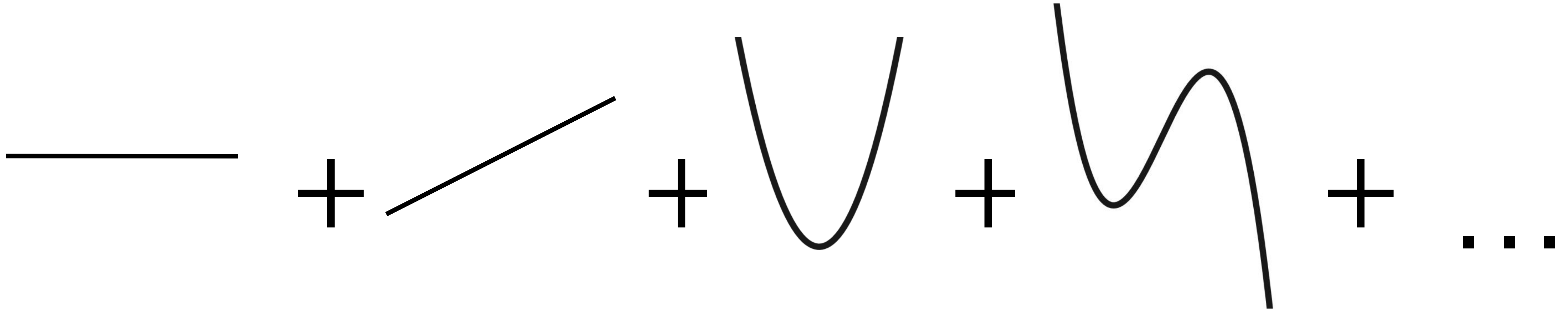
slide from Steve Seitz's [video]

Function Approximation

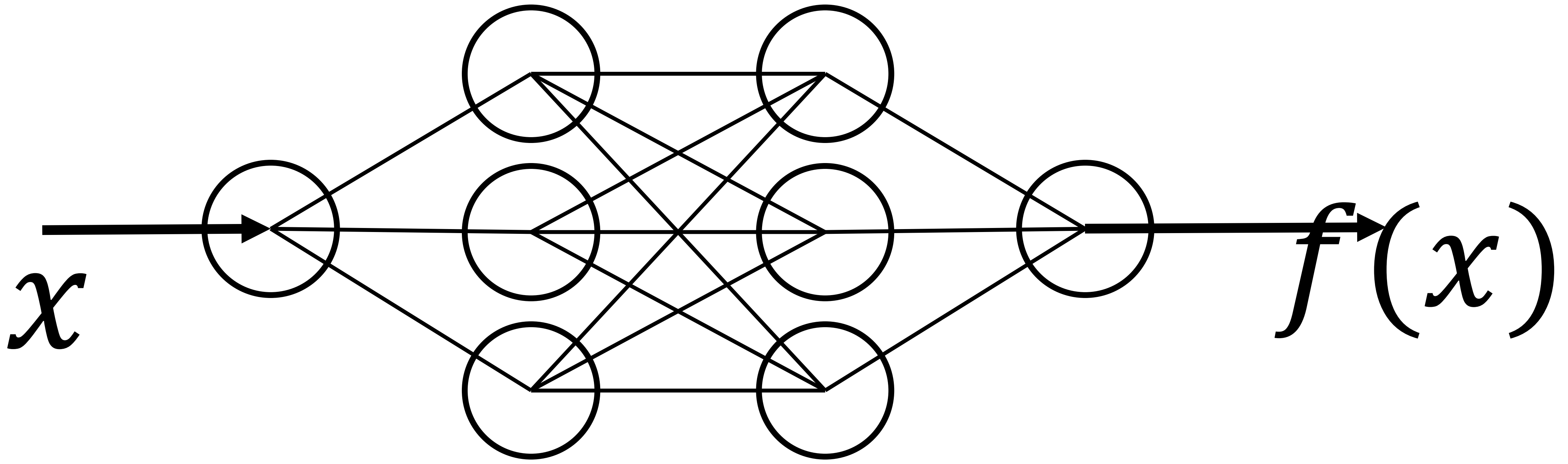
Fourier Series: $f(x) =$



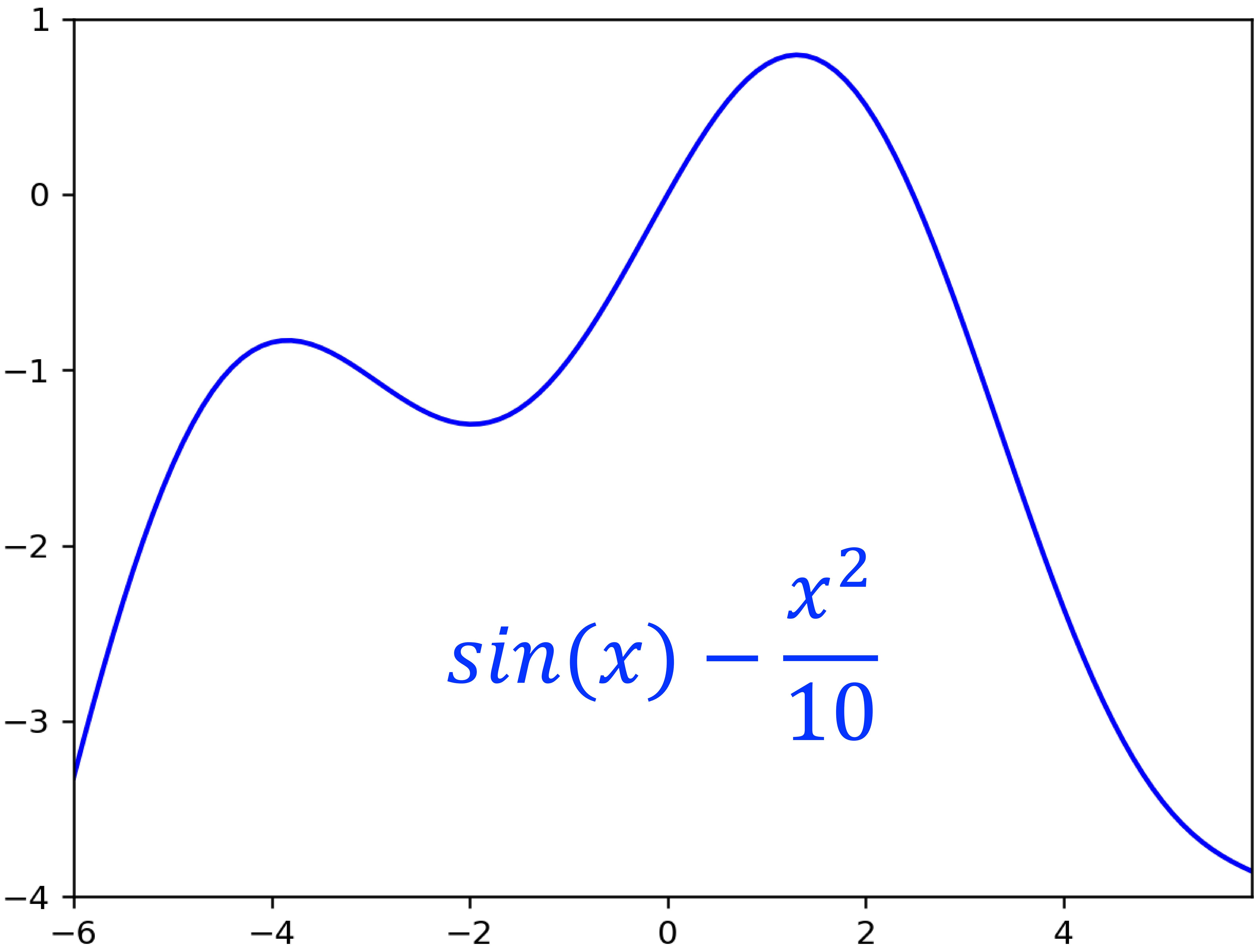
Taylor Series: $f(x) =$



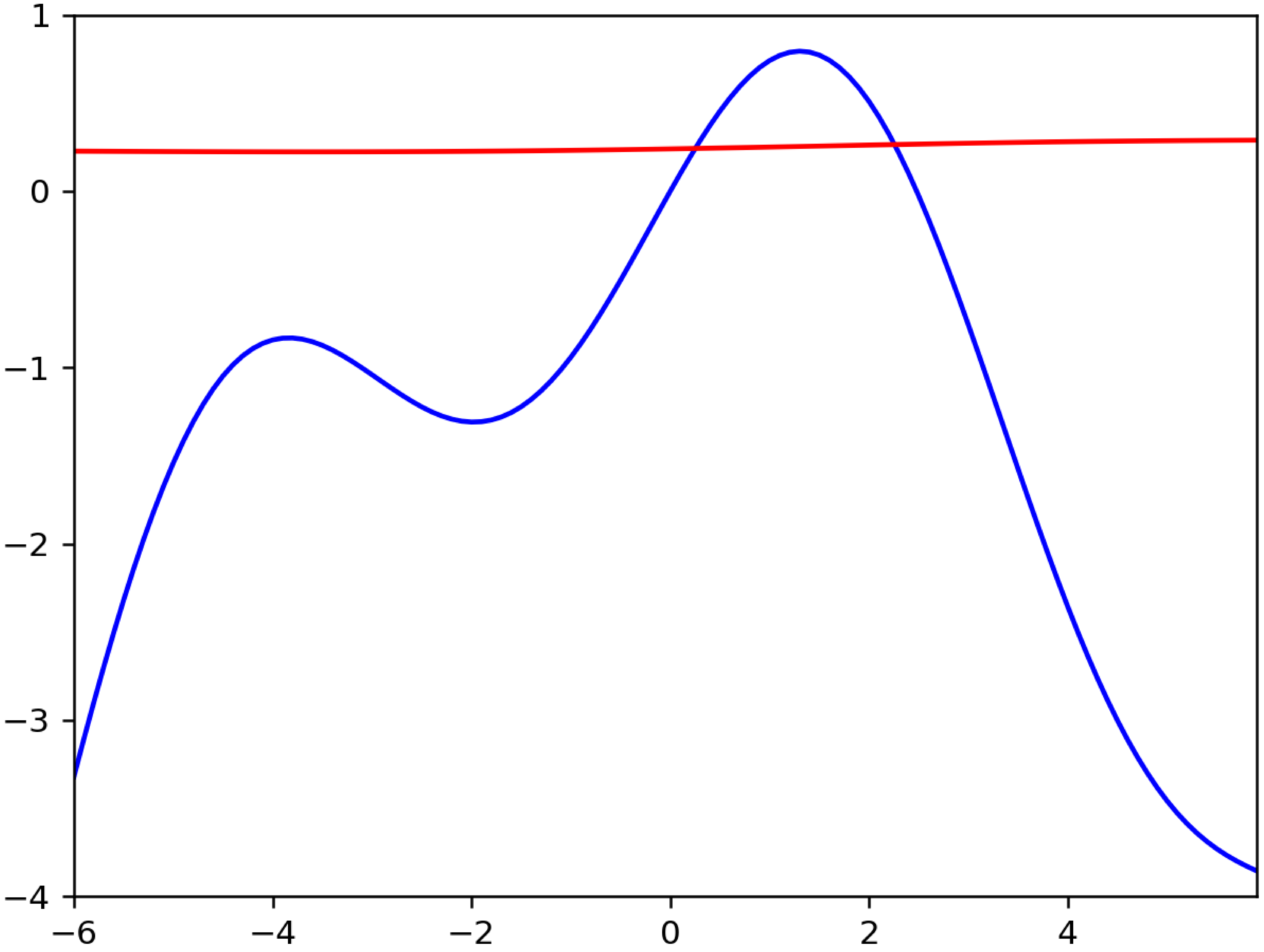
Neural Network:



slide from Steve Seitz's [video]



slide from Steve Seitz's video



slide from Steve Seitz's [video]

$$P(x_n|x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4}, x_{n-5}, x_{n-6}, x_{n-7}, \ldots)$$

slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still red

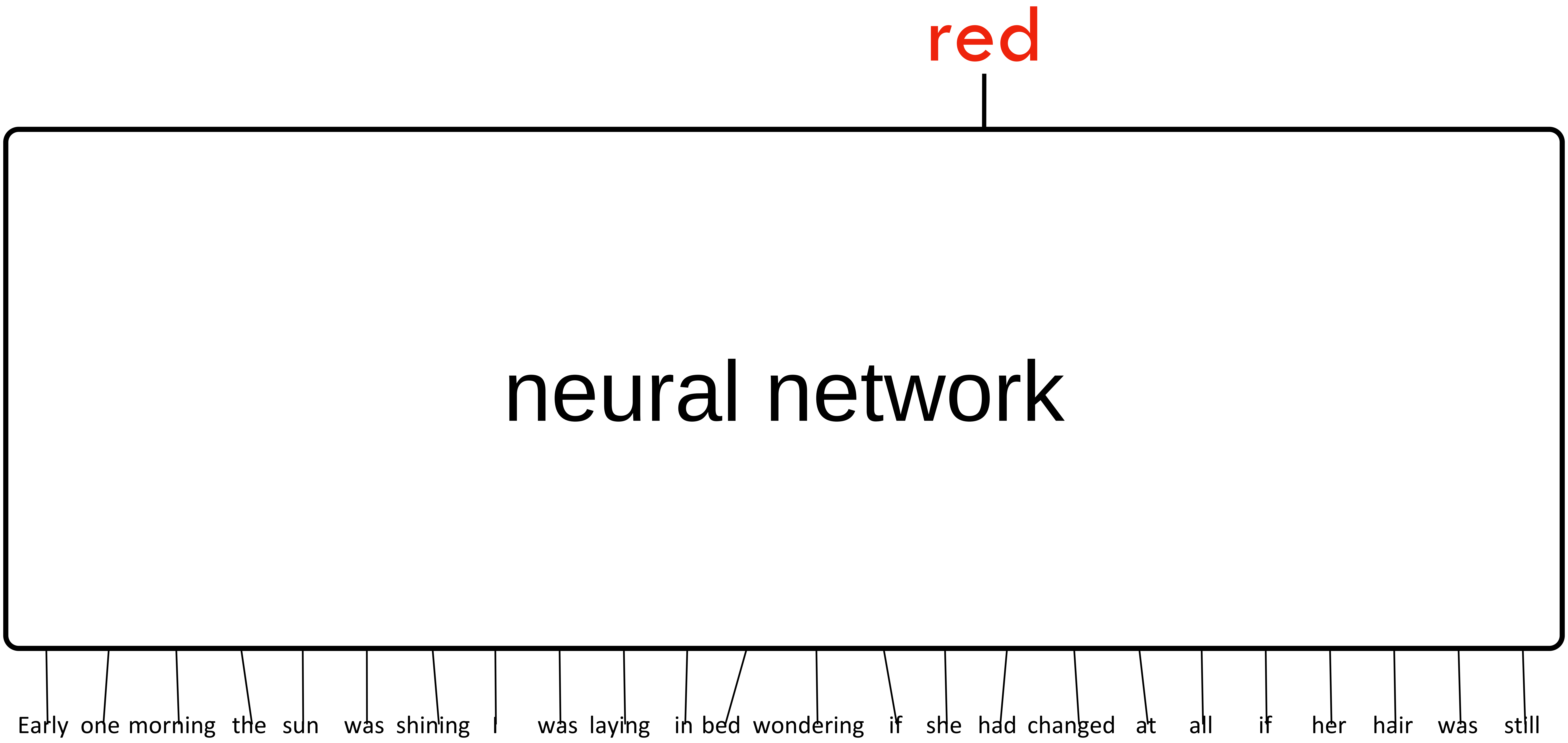
slide from Steve Seitz's video

red

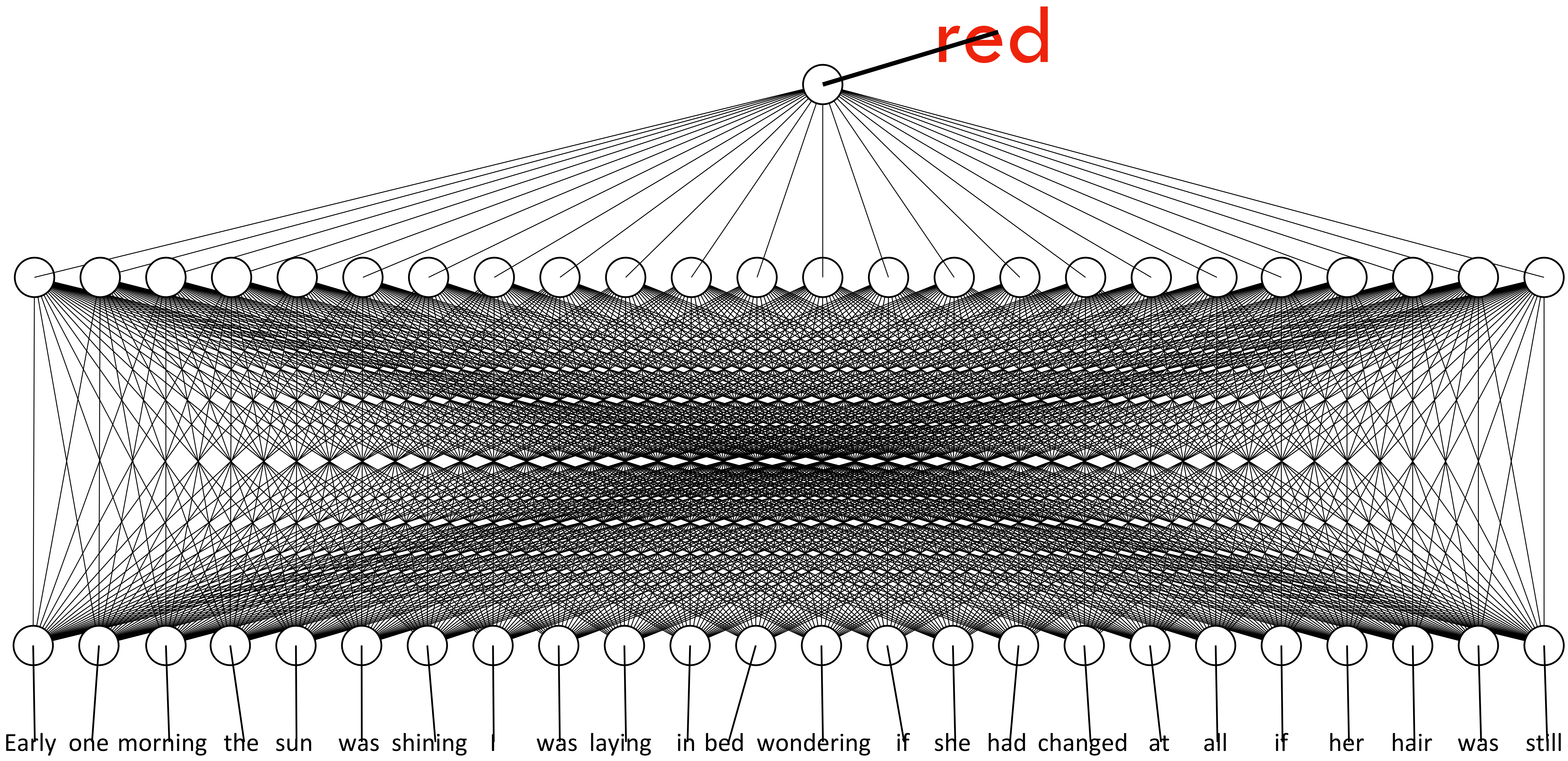
Early one morning the sun was shining I was laying in bed

Wondering if she had changed at all if her hair was still

slide from Steve Seitz's video



slide from Steve Seitz's video



slide from Steve Seitz's [video]

Early one morning the sun was shining I was laying in bed

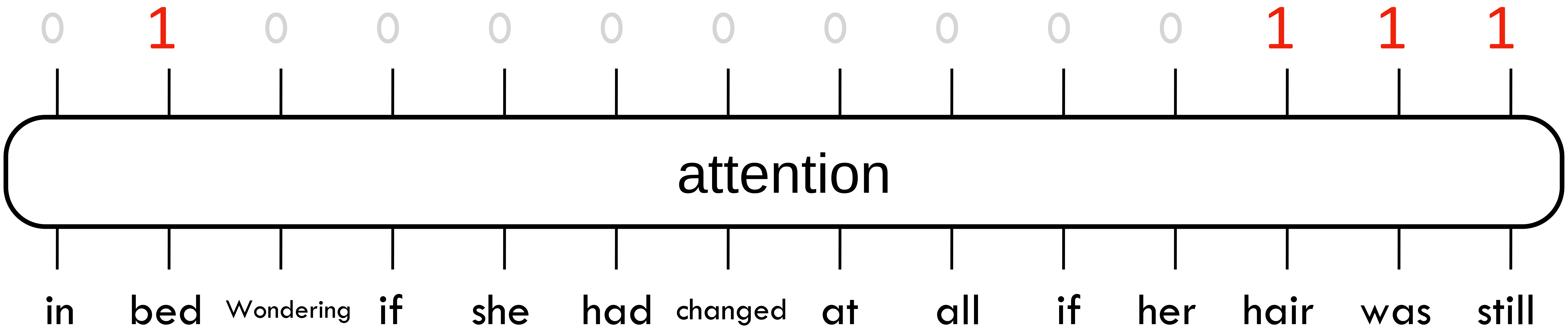
Wondering if she had changed at all if her hair was still ___?___

slide from Steve Seitz's [video]

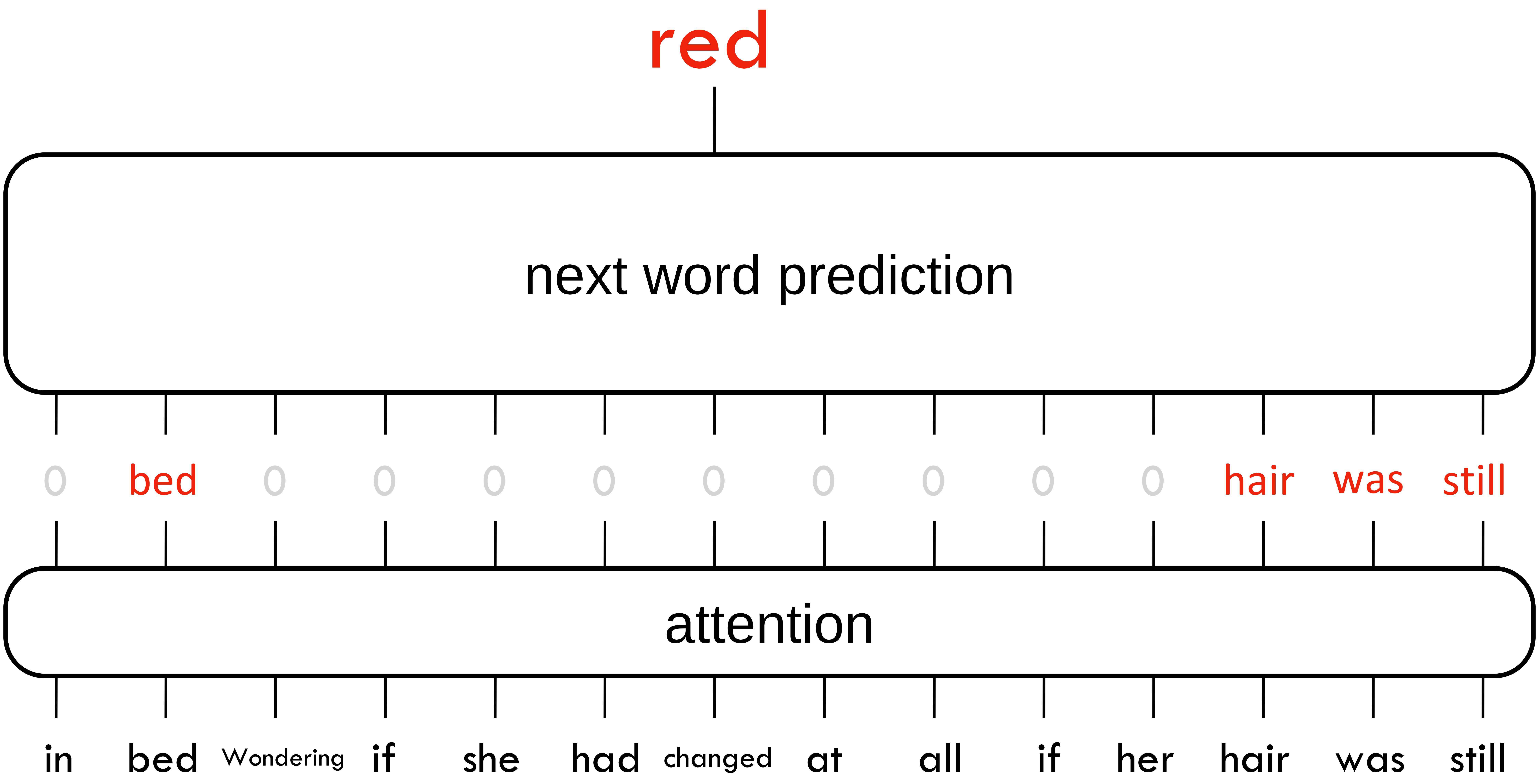
bed

hair was still red

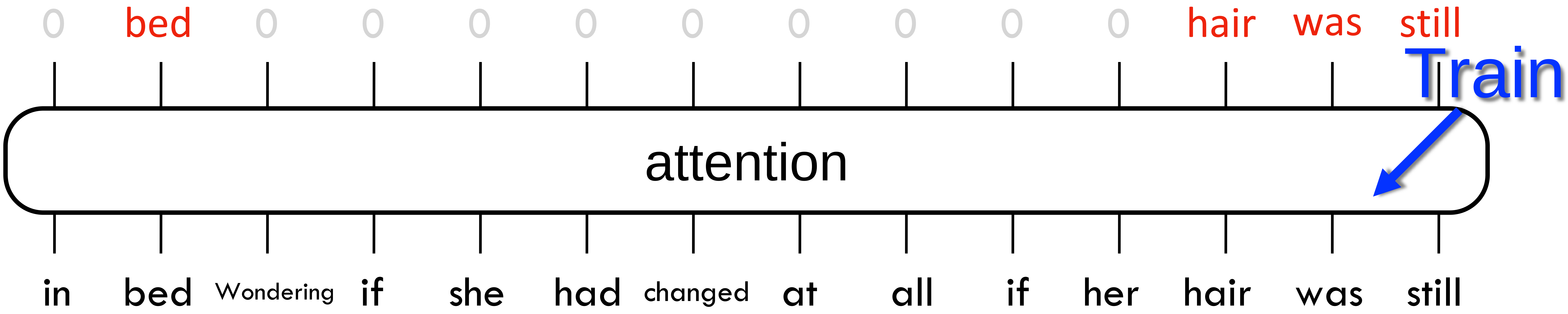
slide from Steve Seitz's video



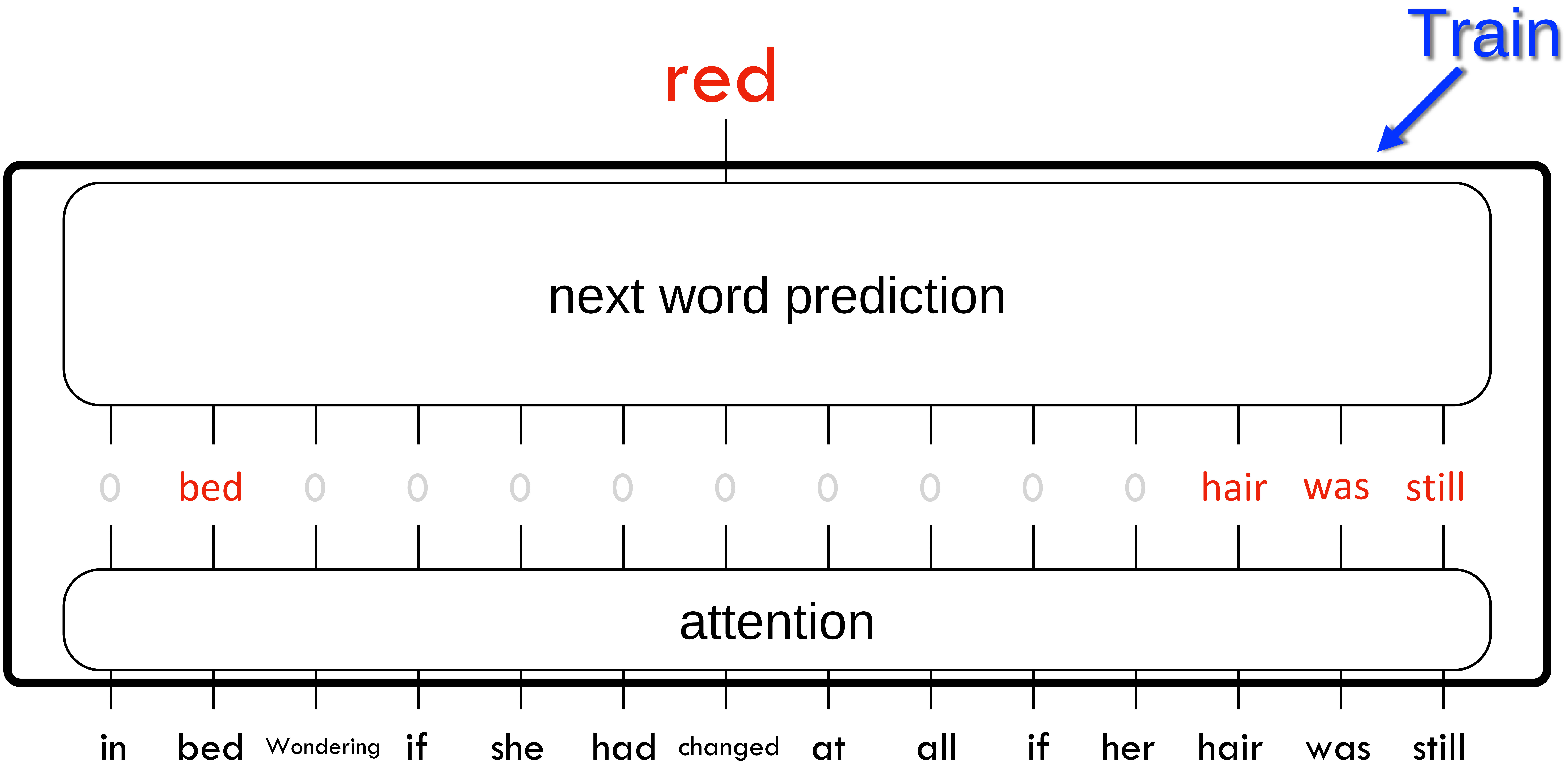
slide from Steve Seitz's [video]



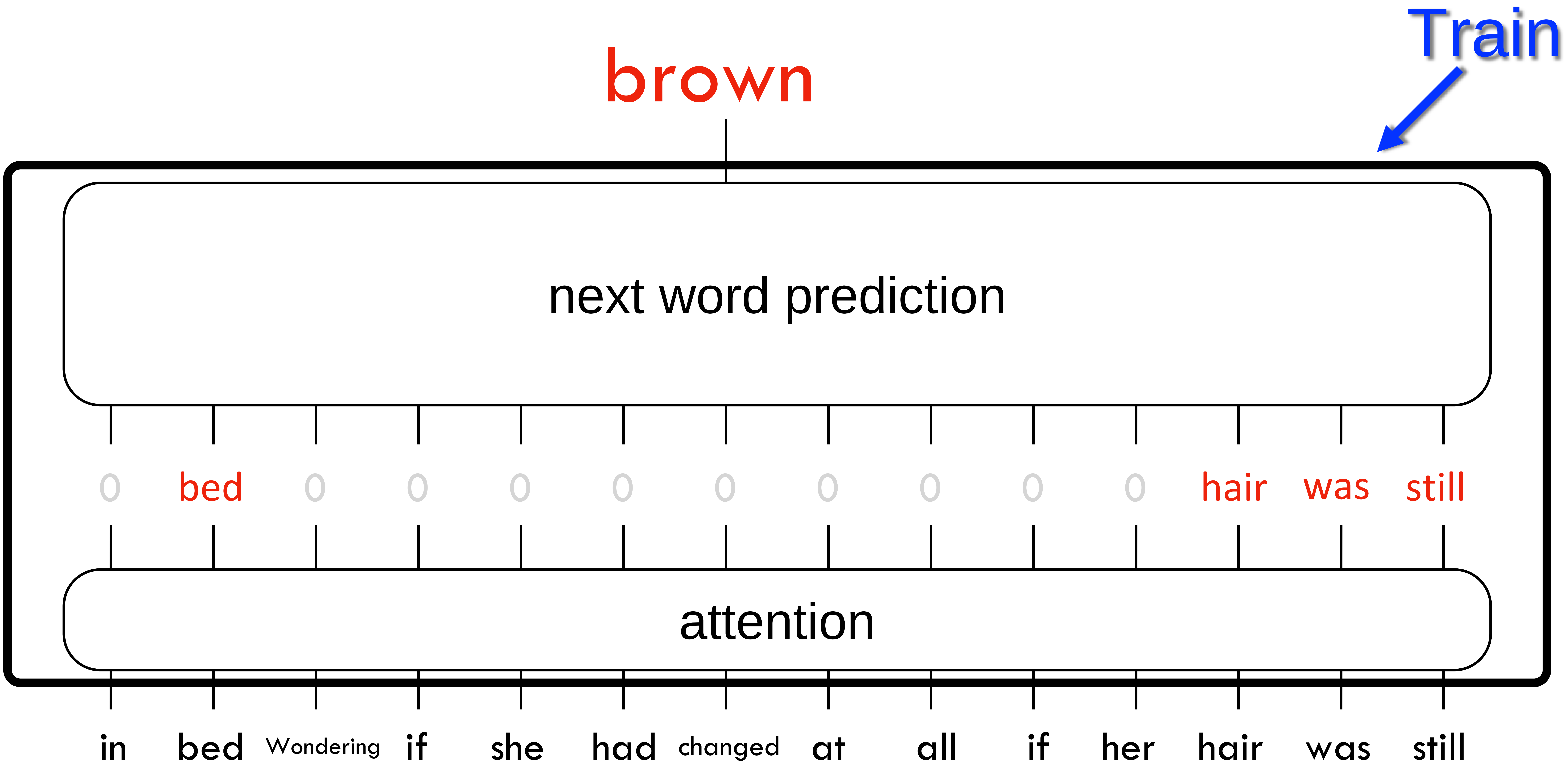
slide from Steve Seitz's [video]



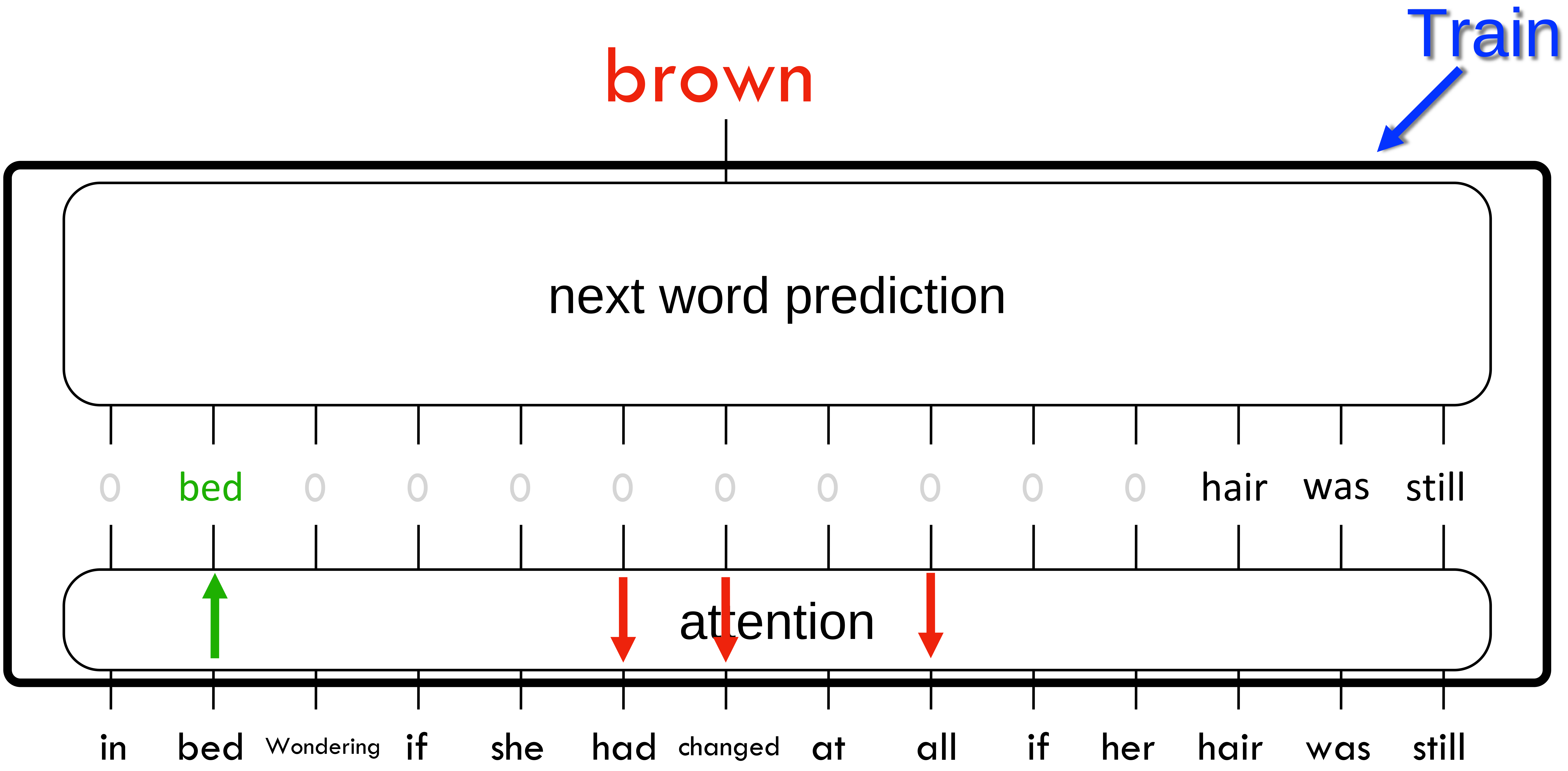
slide from Steve Seitz's video



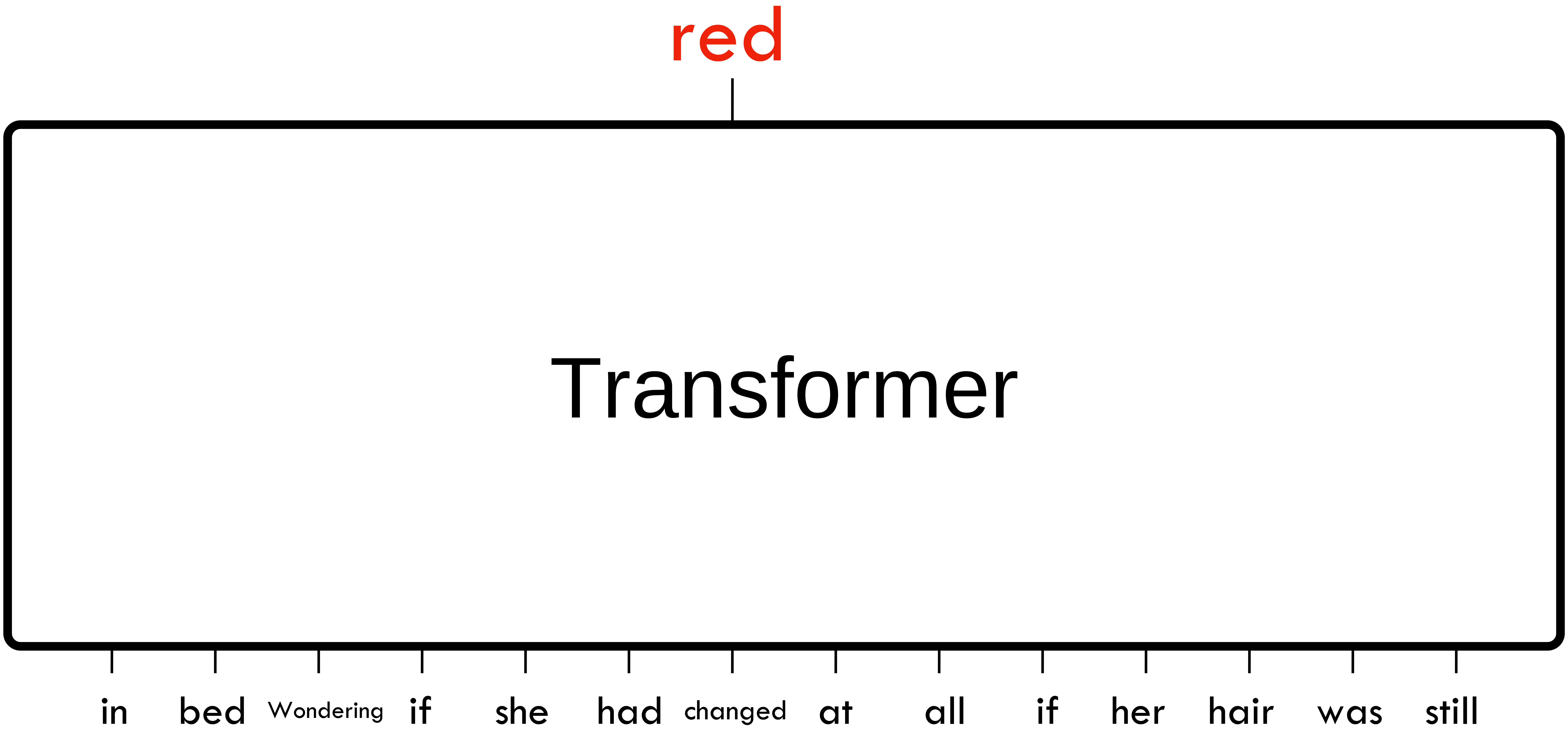
slide from Steve Seitz's video



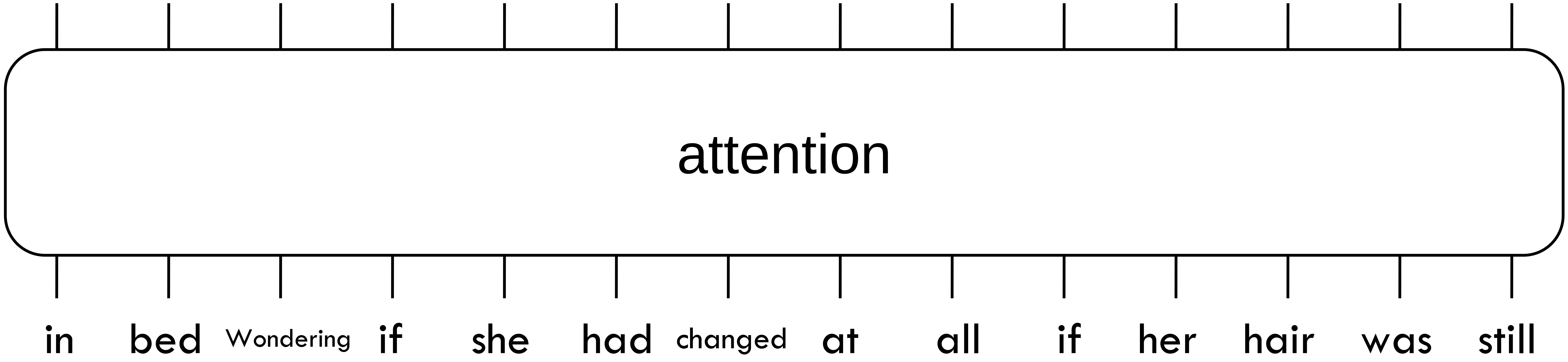
slide from Steve Seitz's [video]



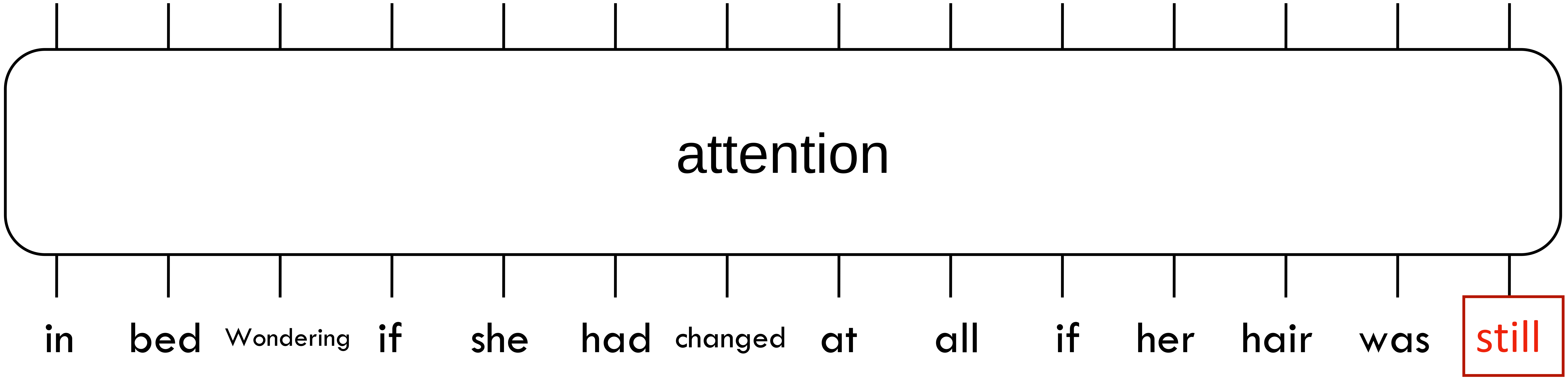
slide from Steve Seitz's [video]



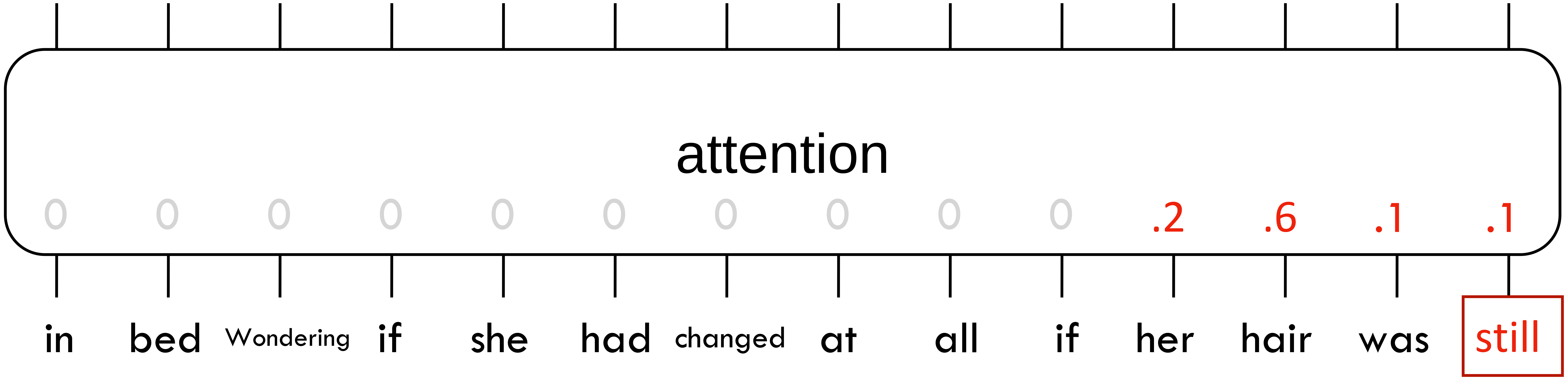
slide from Steve Seitz's [video]



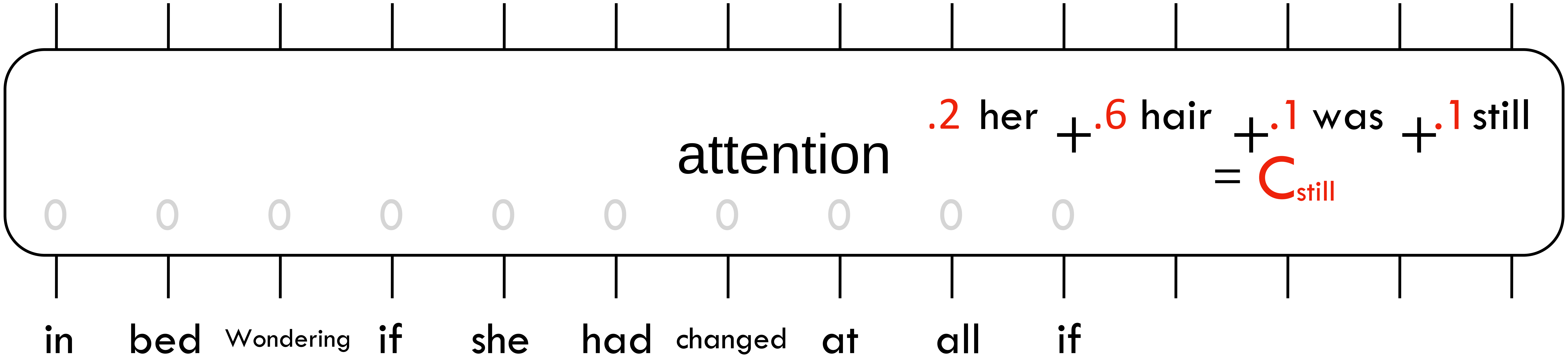
slide from Steve Seitz’s [video]



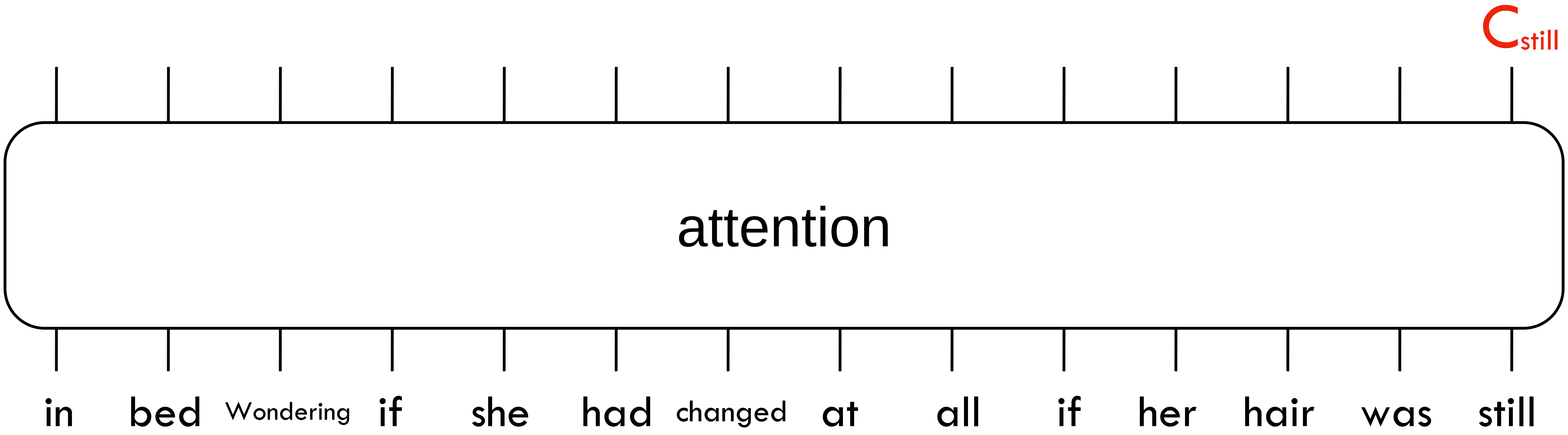
slide from Steve Seitz's [video]



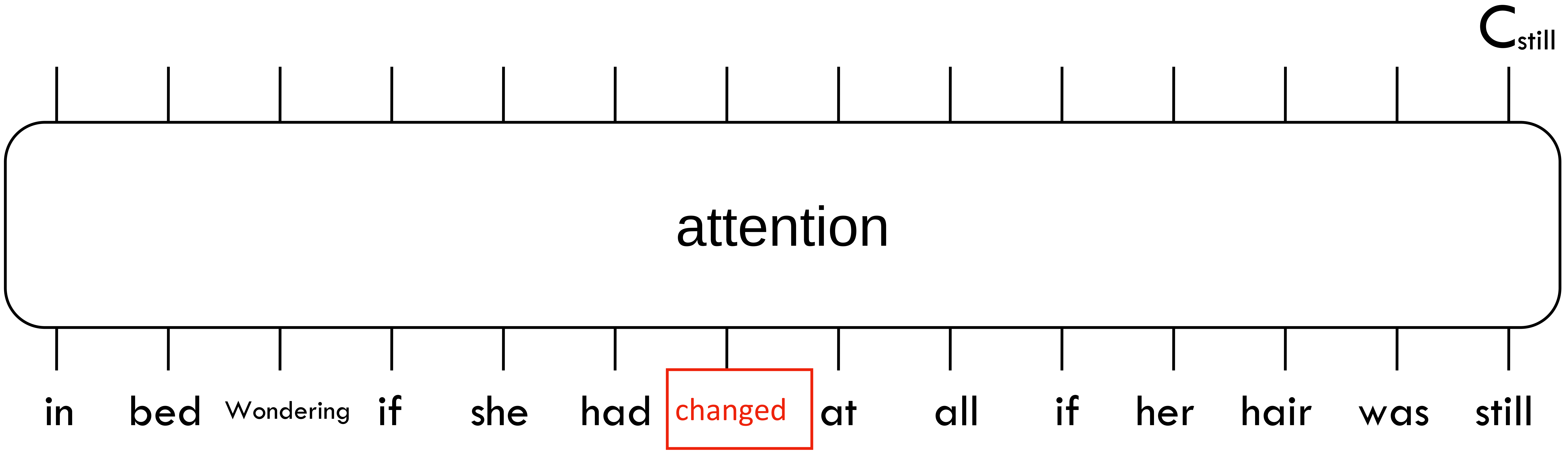
slide from Steve Seitz's [video]()



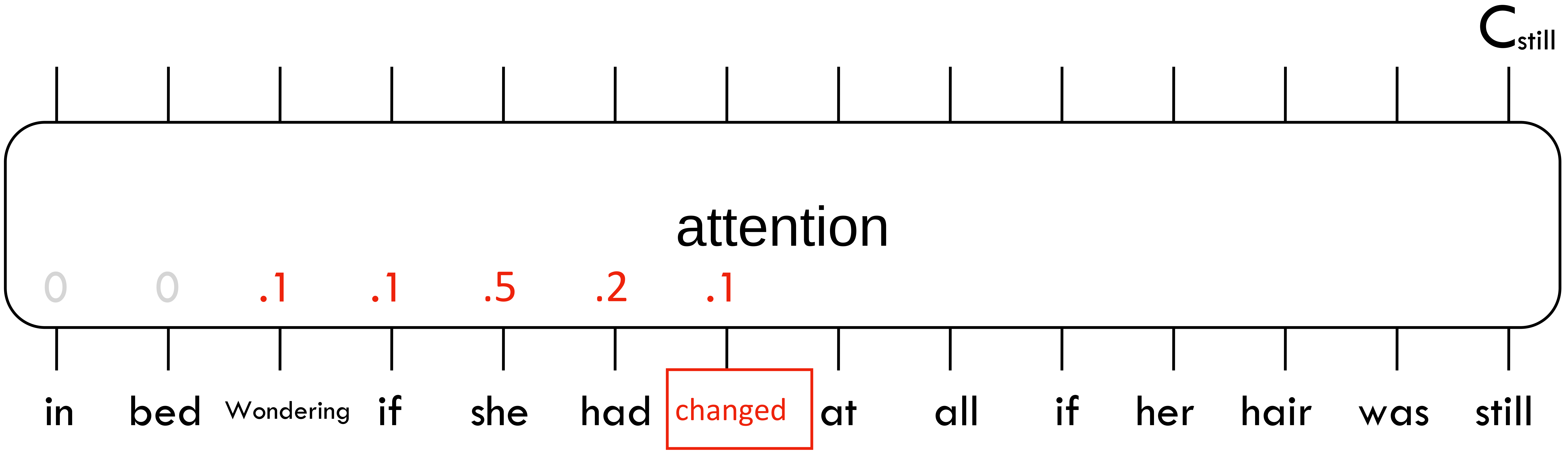
slide from Steve Seitz's video



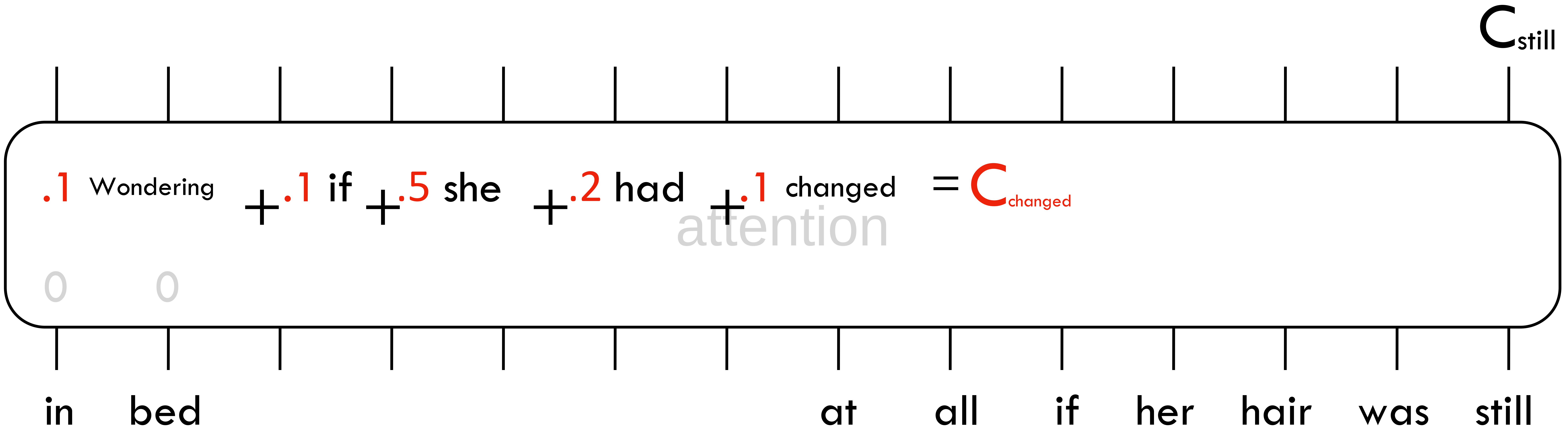
slide from Steve Seitz's video



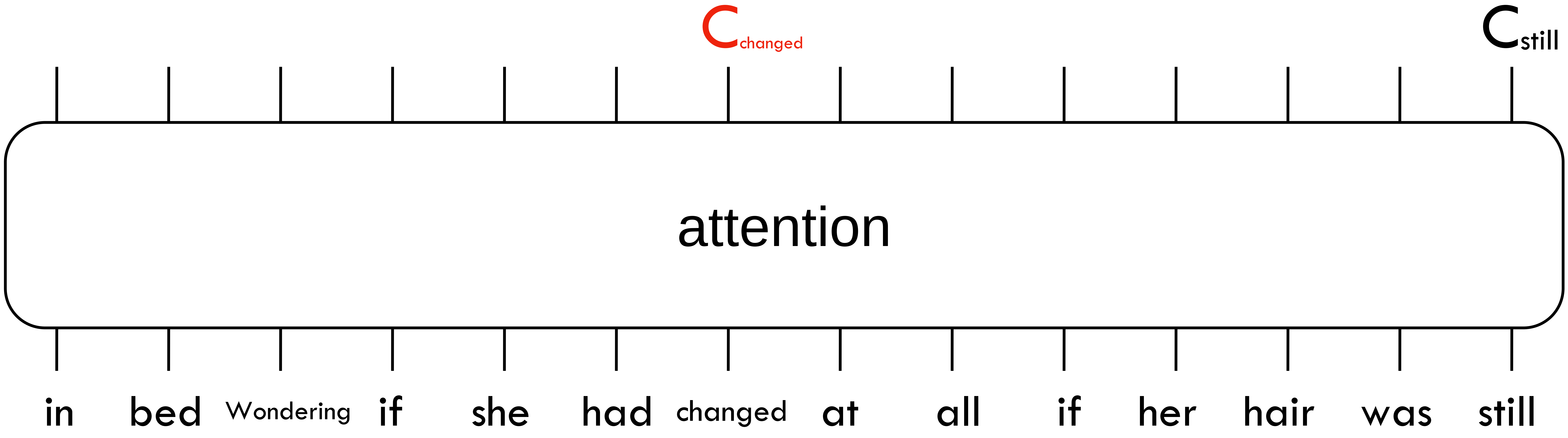
slide from Steve Seitz's video



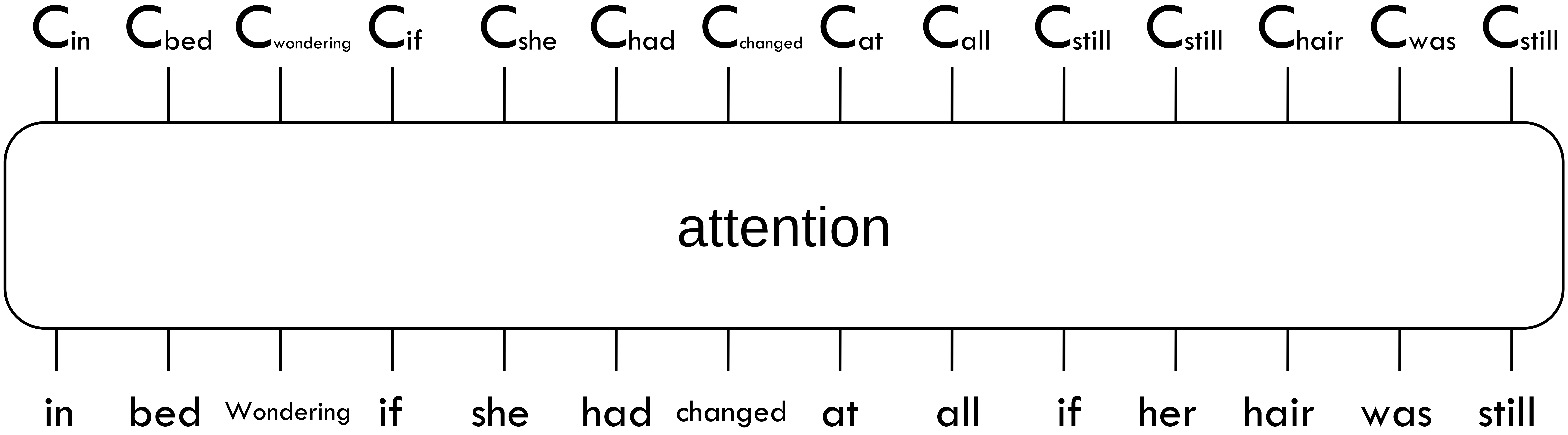
slide from Steve Seitz's [video]



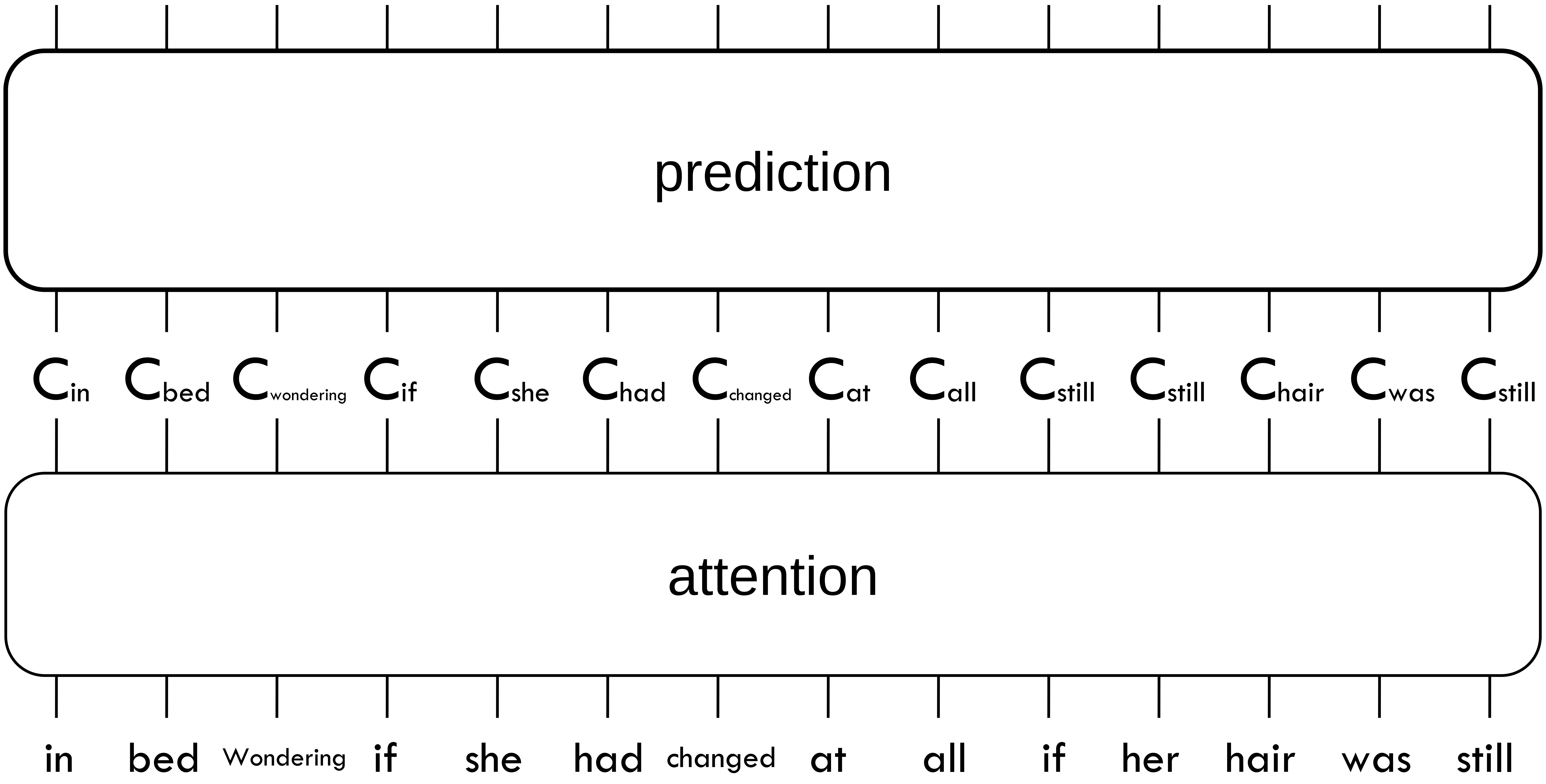
slide from Steve Seitz's video



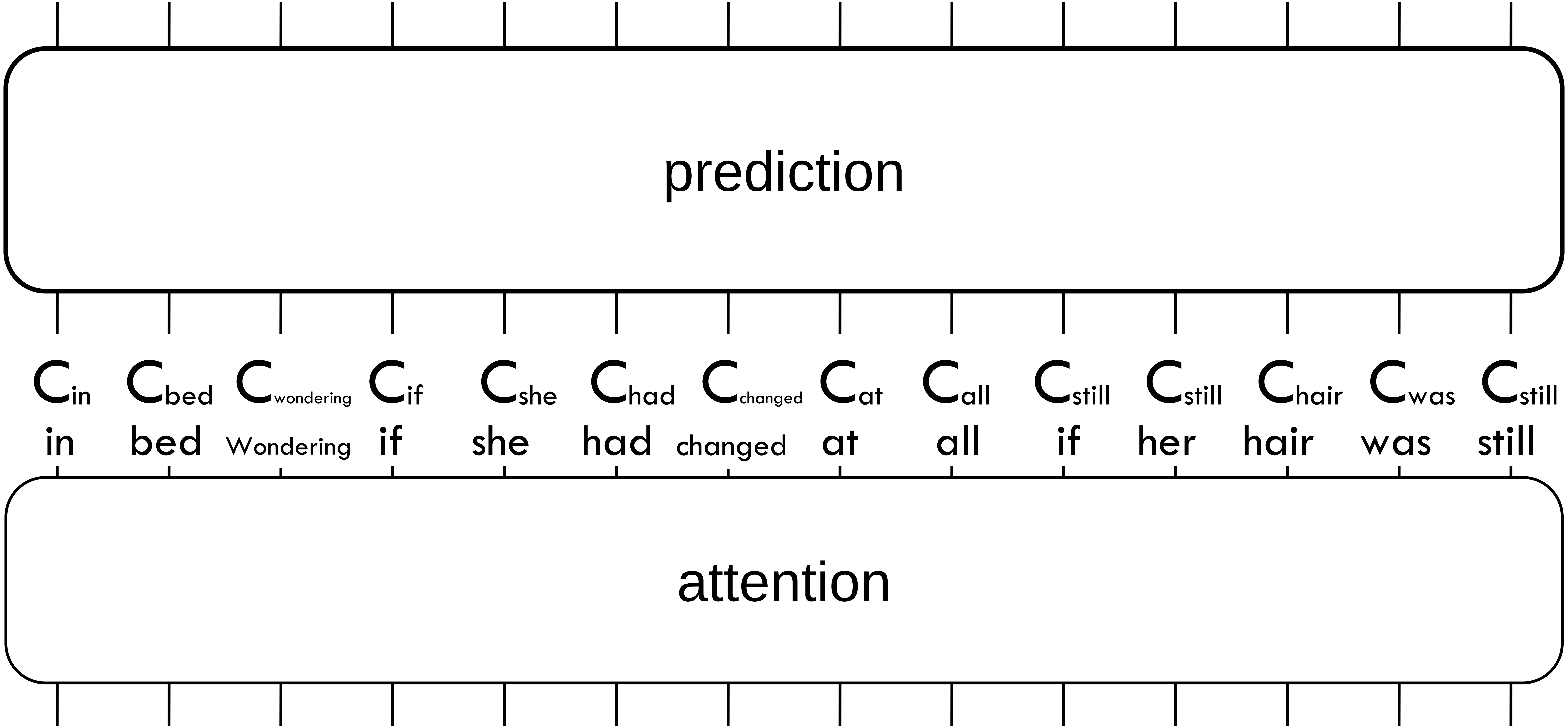
slide from Steve Seitz's [video]



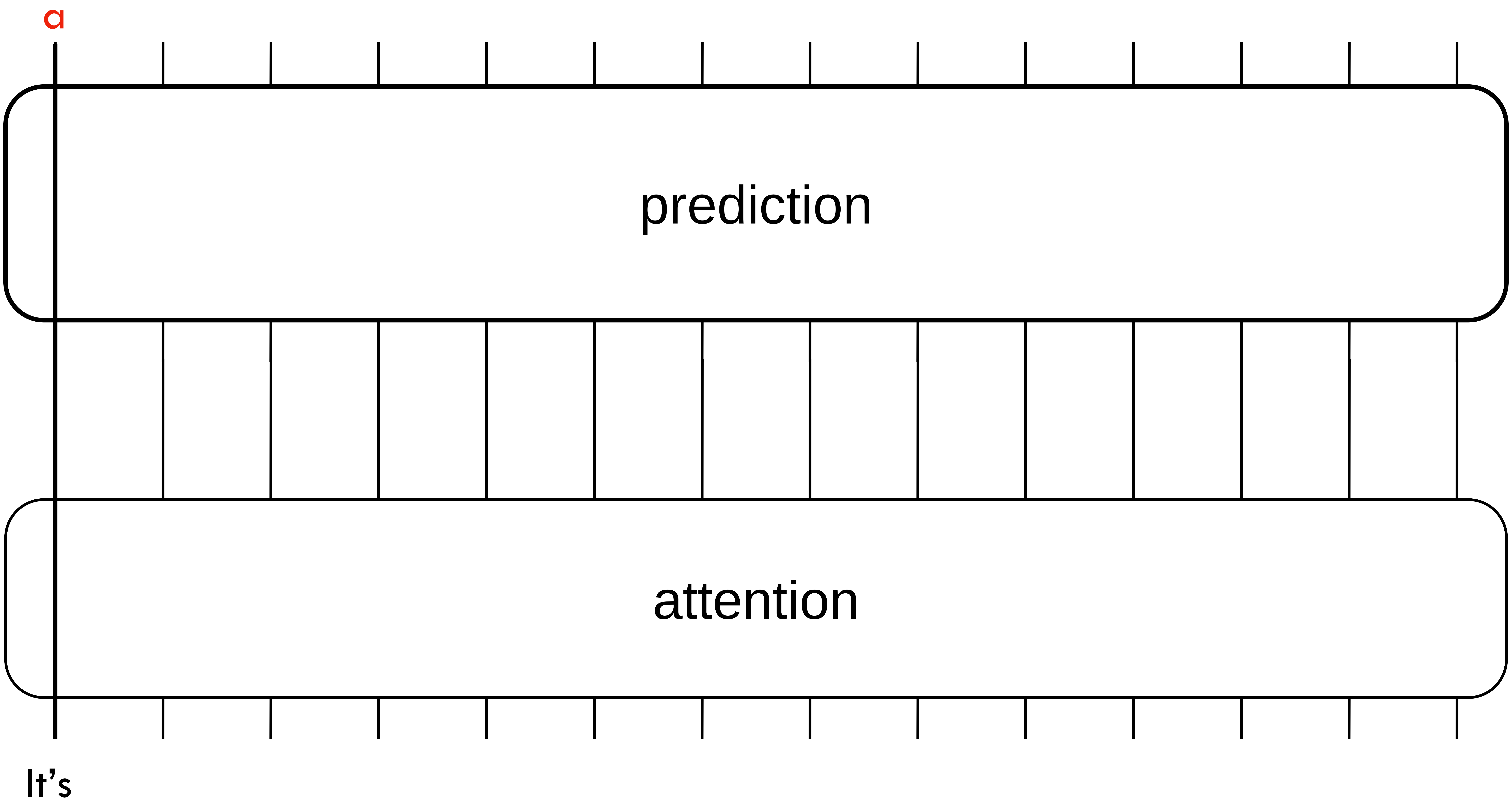
slide from Steve Seitz's video



slide from Steve Seitz's video



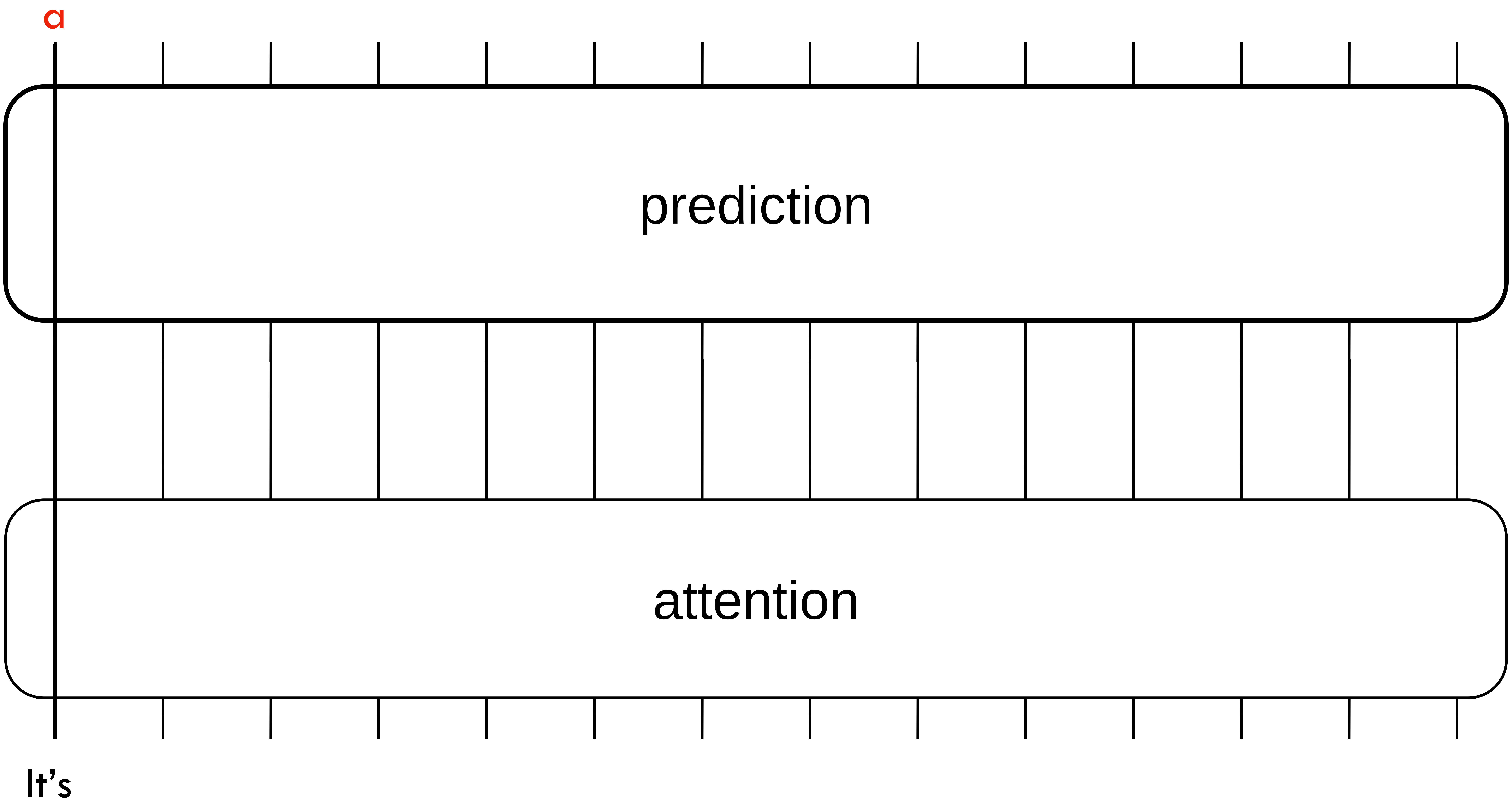
slide from Steve Seitz's [video]



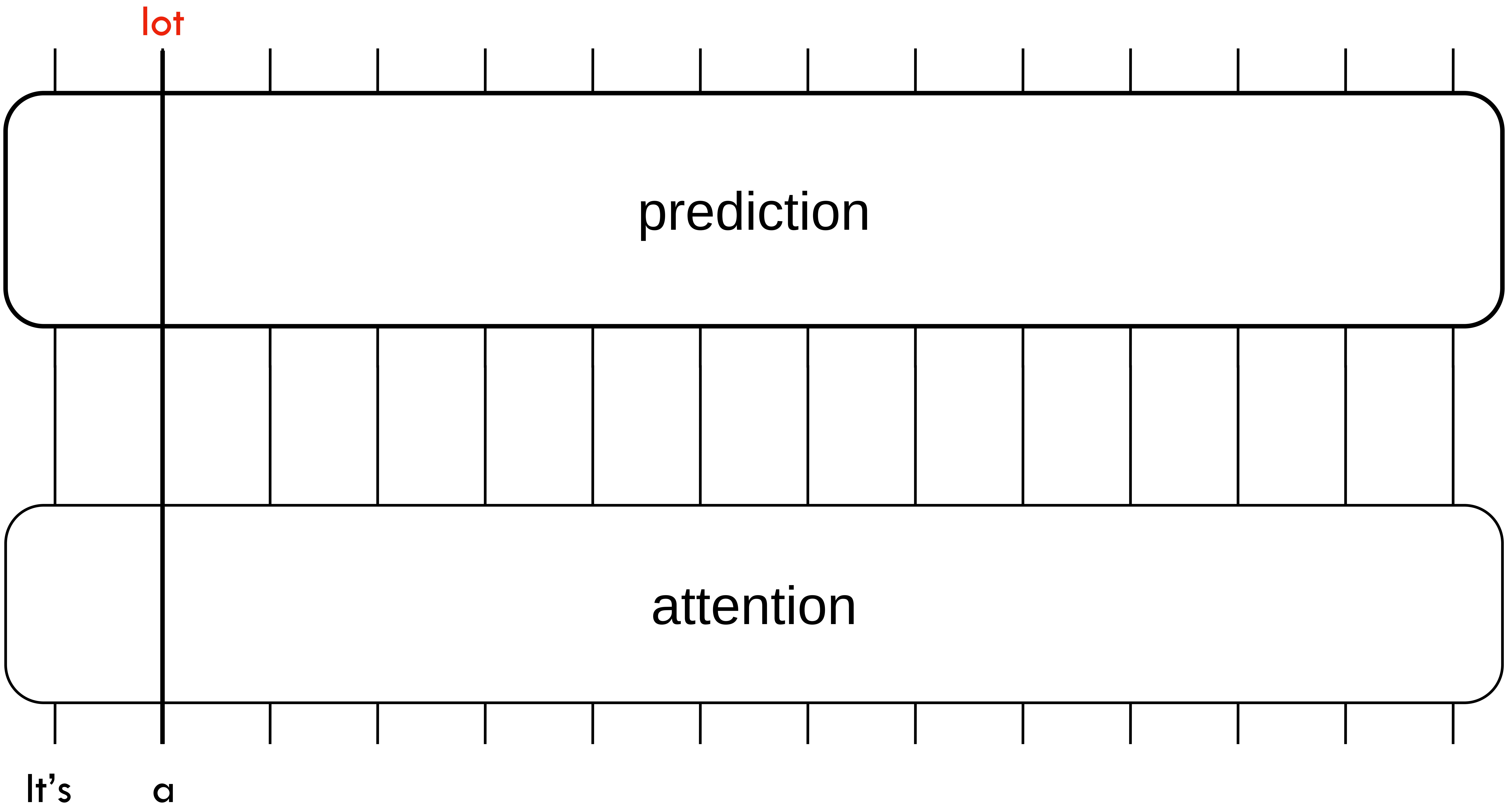
slide from Steve Seitz’s video

a	the	looking	possible	getting
0.4	0.3	0.1	0.1	0.1

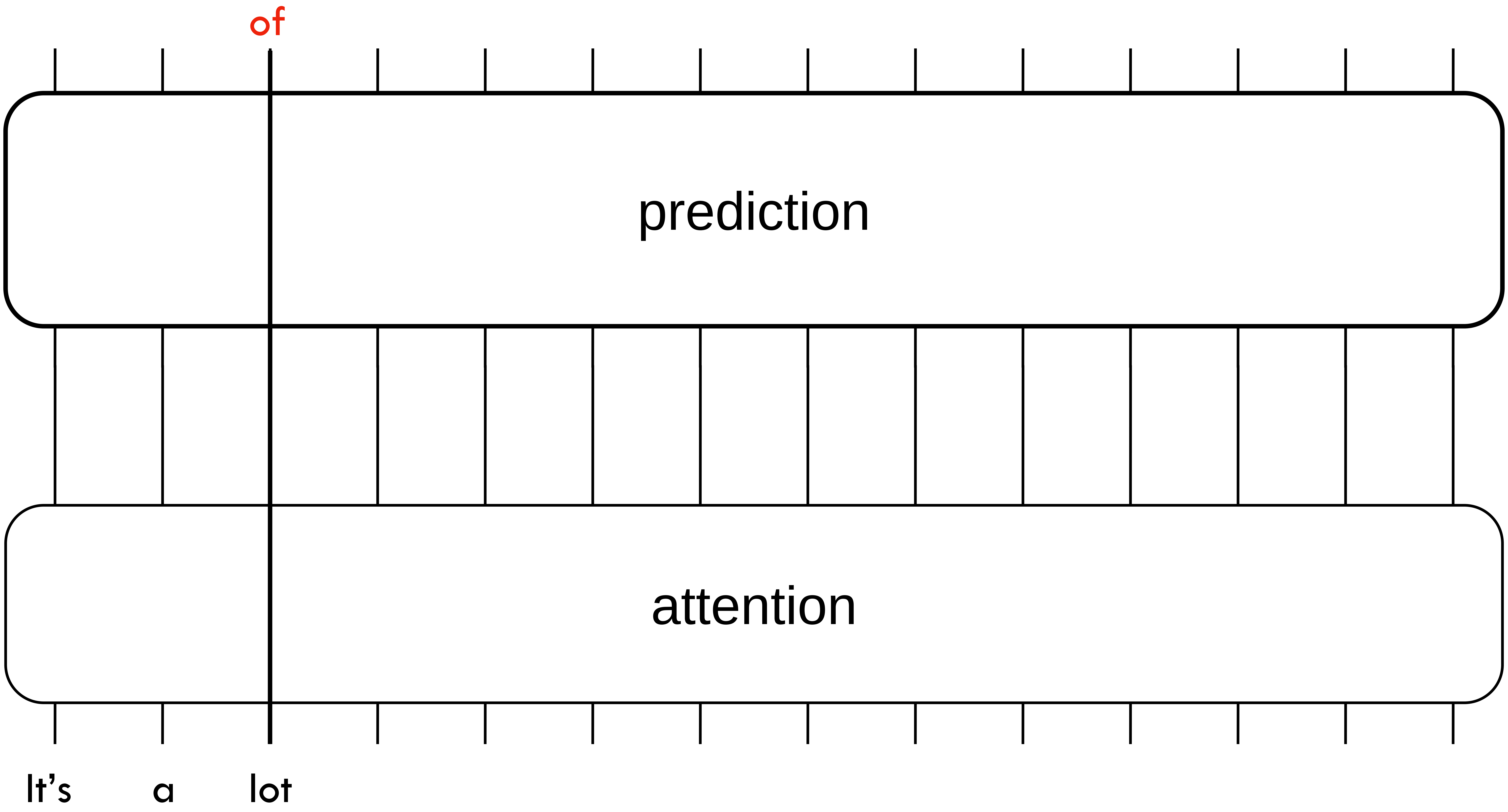
slide from Steve Seitz’s [video]



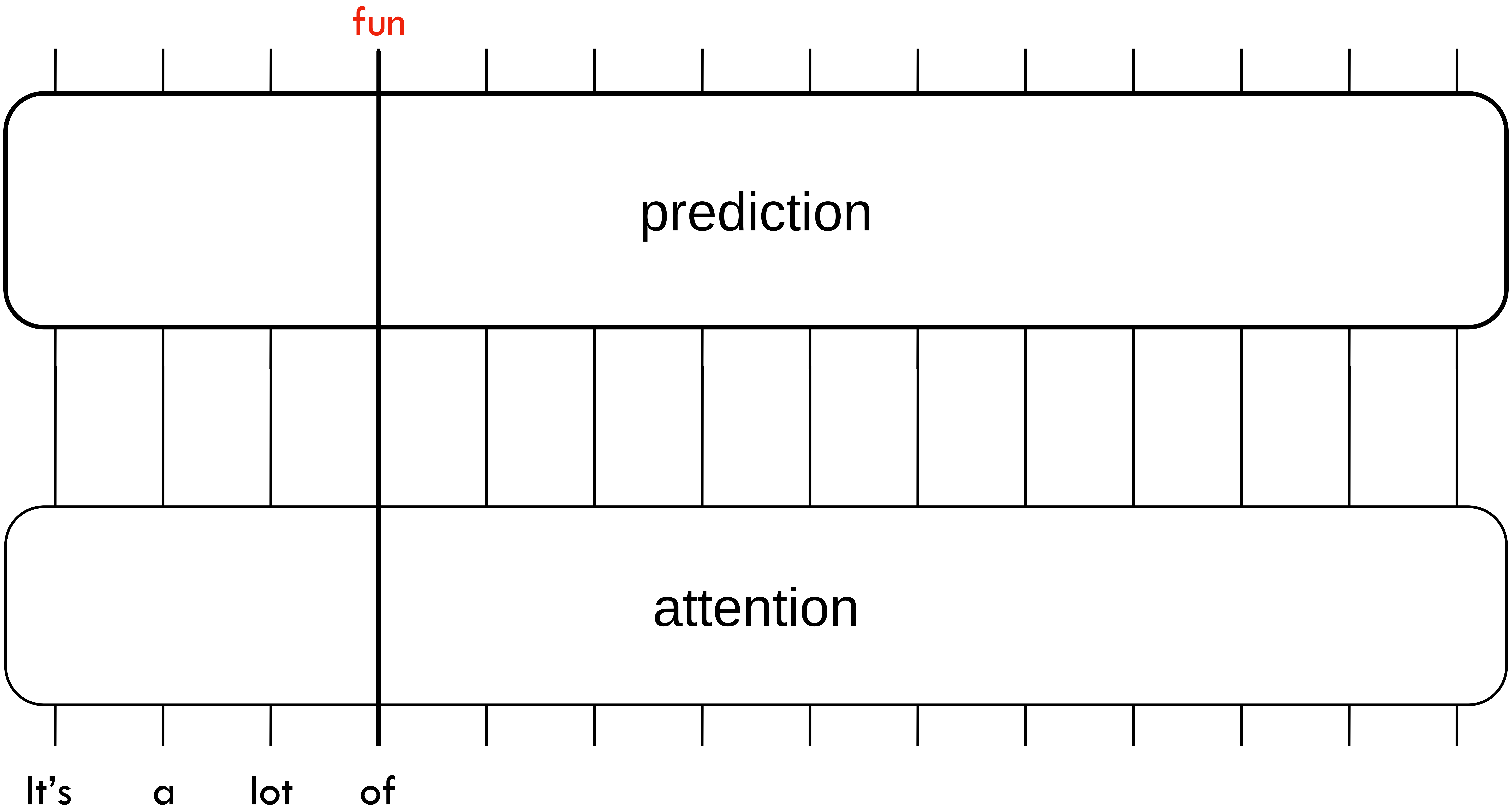
slide from Steve Seitz's [video]



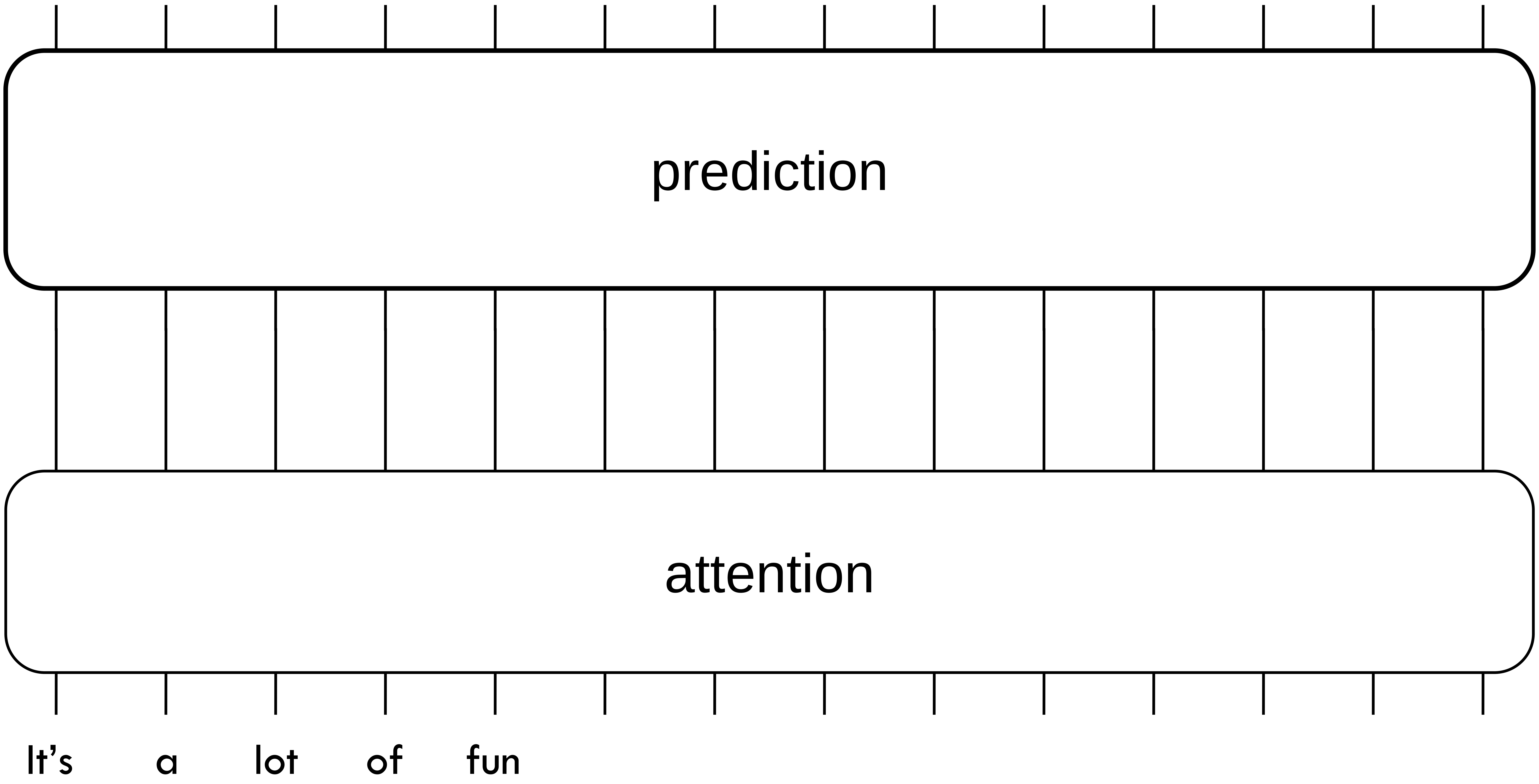
slide from Steve Seitz’s [video]



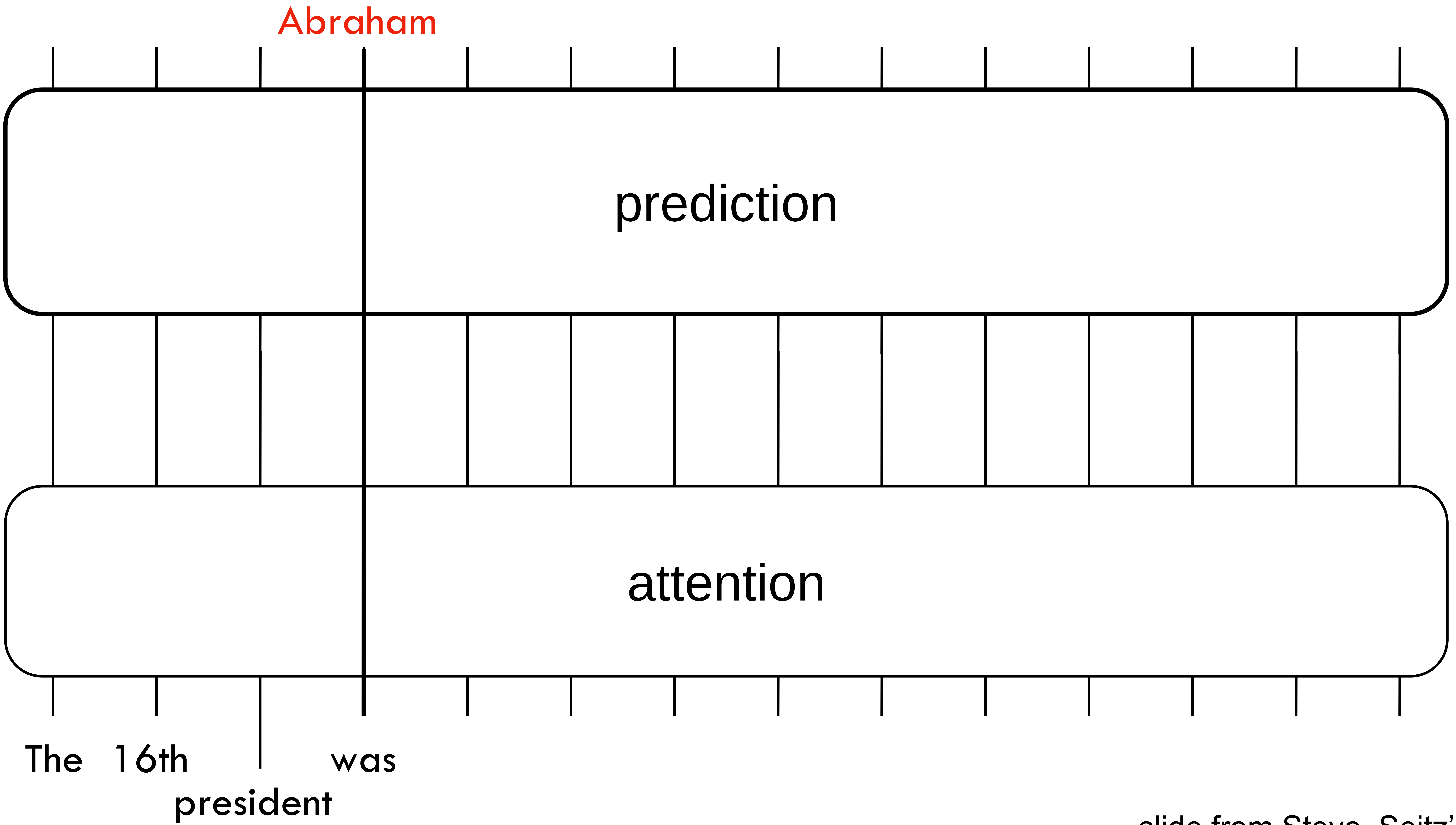
slide from Steve Seitz's [video]



slide from Steve Seitz's video



slide from Steve Seitz’s video



slide from Steve Seitz's [video]

The 16th President was ?

The capital of Zimbabwe is ?

Frank Zappa's middle name is ?

Napoleon was born on this date ?

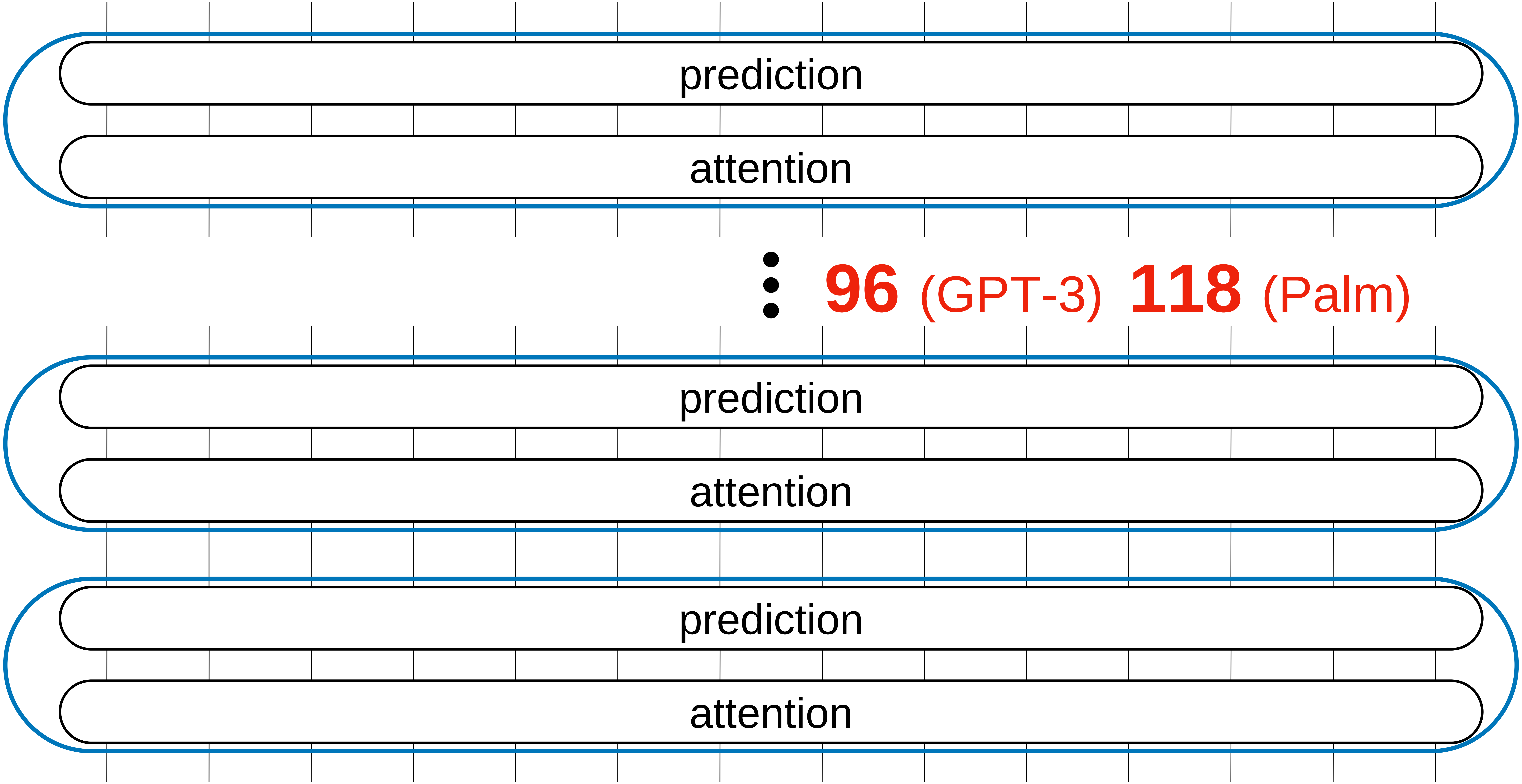
The prime factorization of 19456721434 is ?

Queen Victoria's maiden name was ?

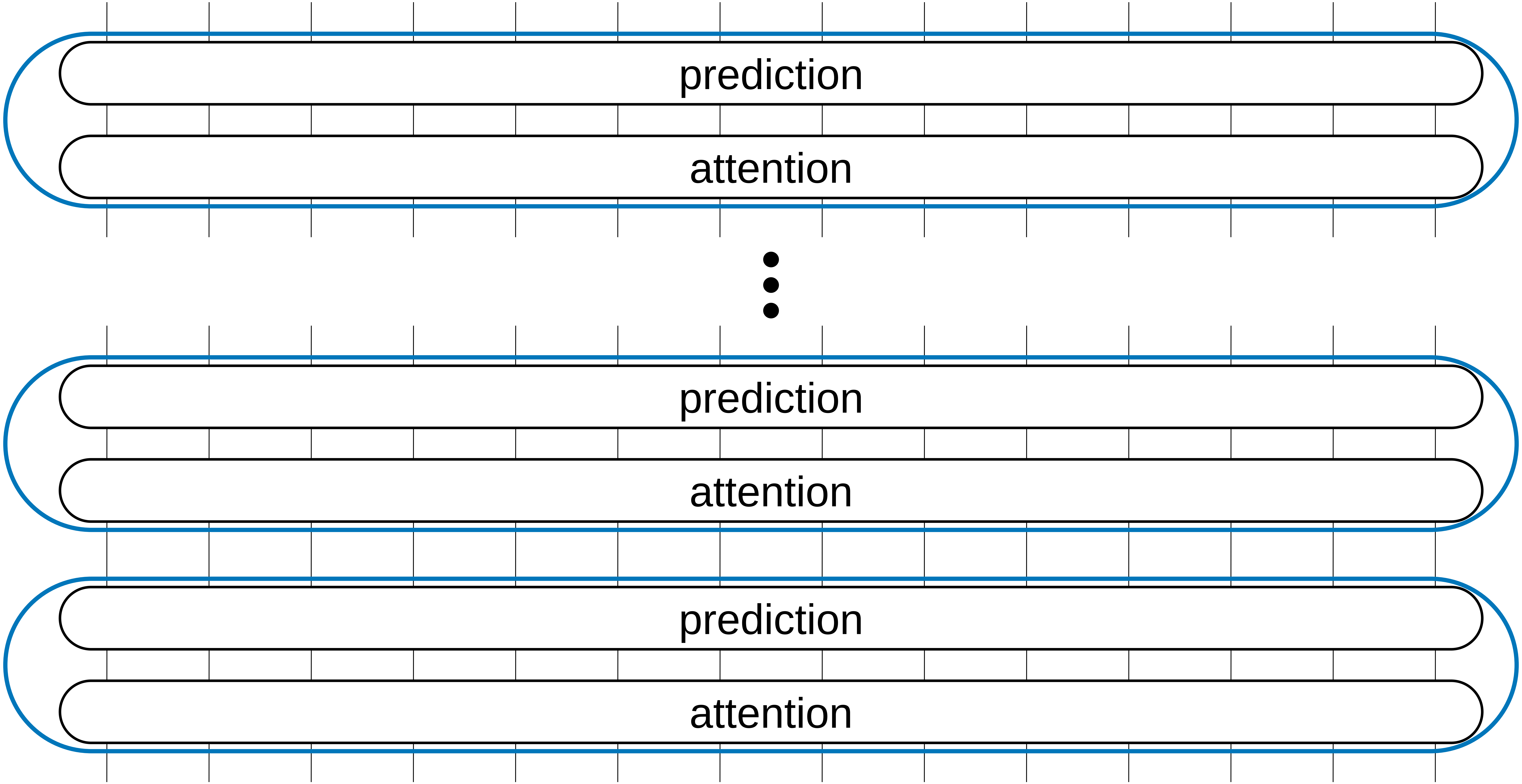
US per-capita income in 1957 was ?

The lat long coordinates of Rome are ?

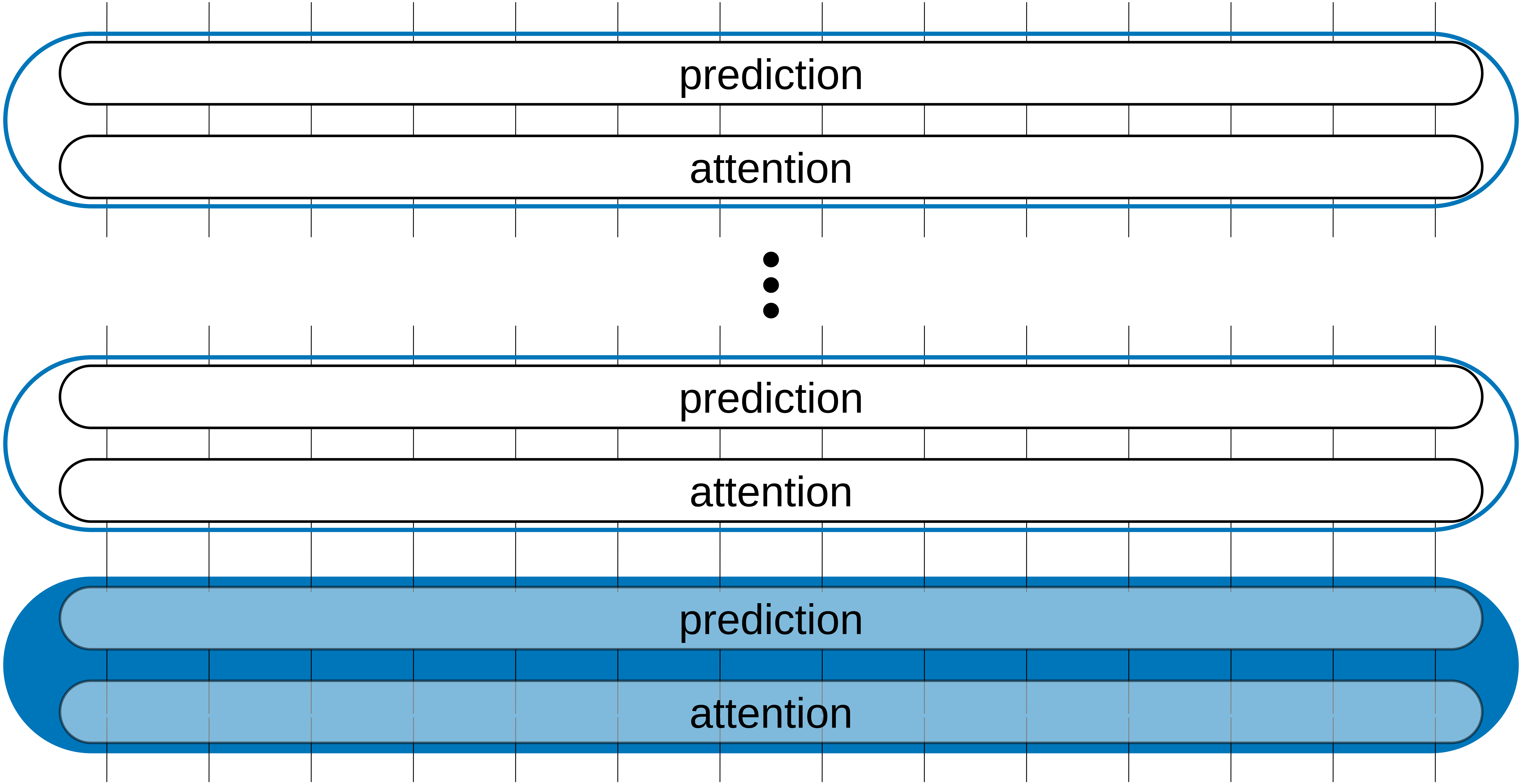
slide from Steve Seitz's video



slide from Steve Seitz's [video]



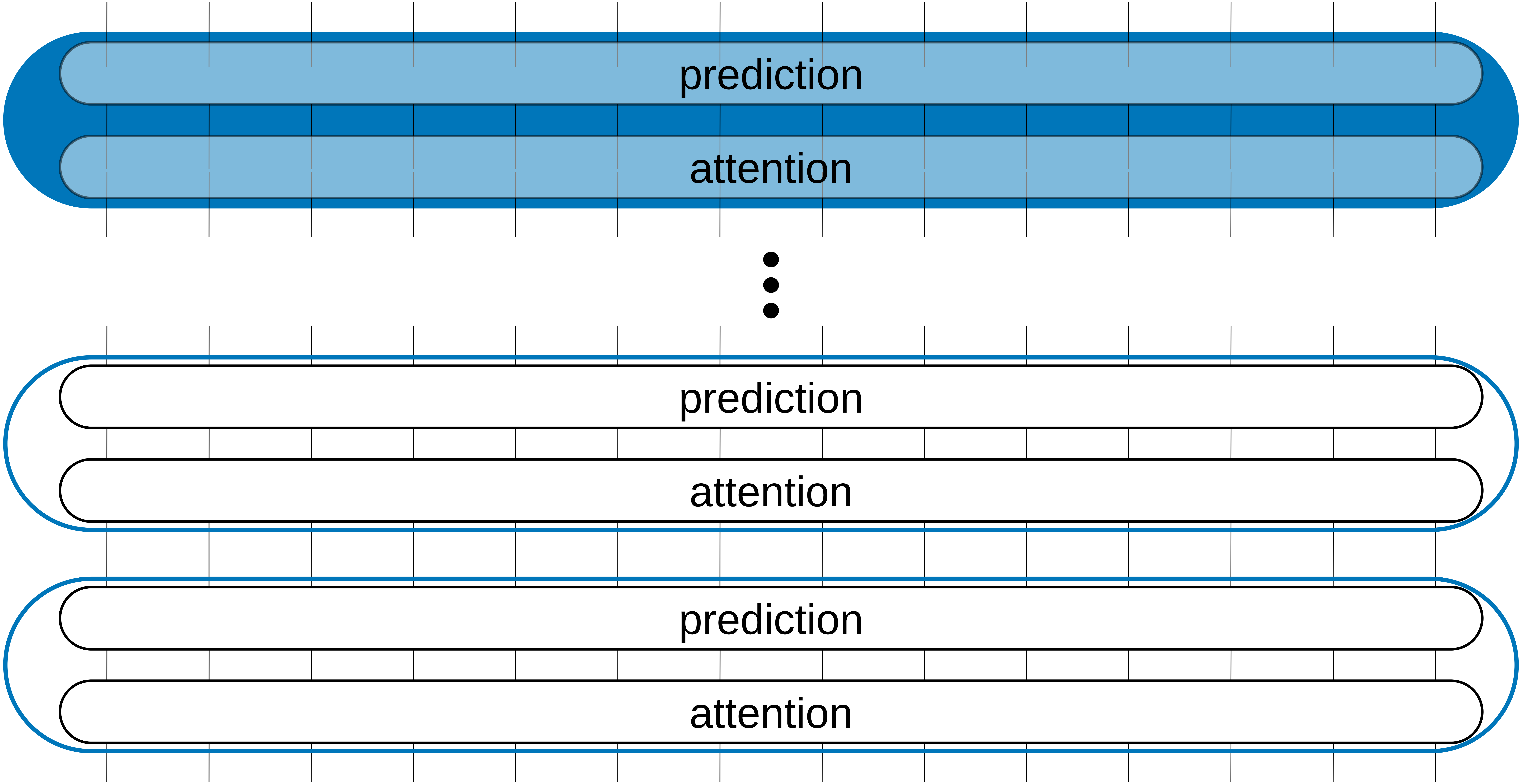
slide from Steve Seitz's [video]



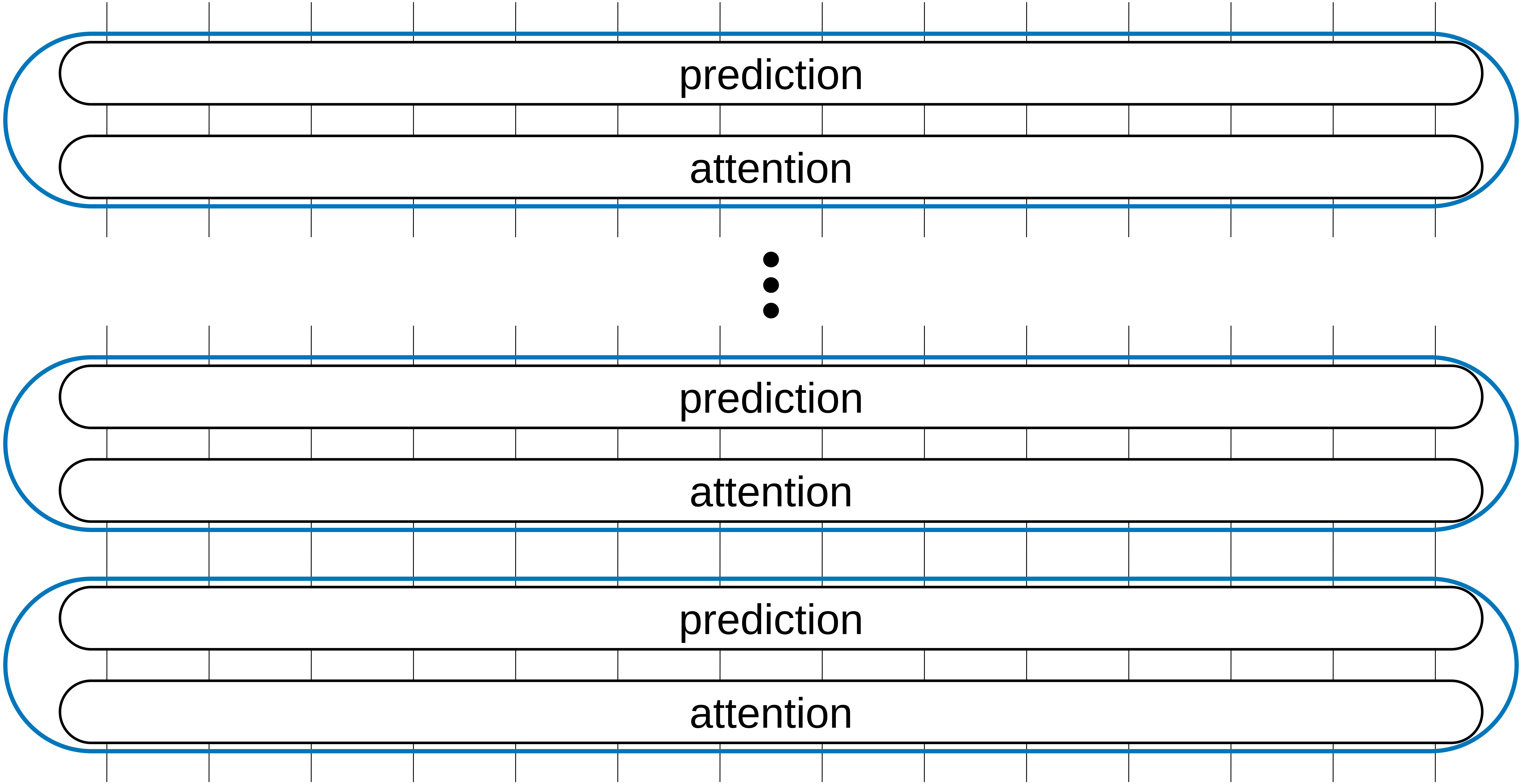
Syntax

slide from Steve Seitz's [video]

Semantics



slide from Steve Seitz's [video]



slide from Steve Seitz's [video]

How much data to train?

All of it…

a pattern of characters that looks like a star

```
.  O  .
O  O  O
.  O  .
```

slide from Steve Seitz’s [video]

a pattern of characters that looks like a vertical line

```
.   ○   .
.   ○   .
.   ○   .
.   ○   .
.   ○   .
```

slide from Steve Seitz’s [video]

a pattern of characters that looks like a triangle

. ○ .

. ○ ○ .

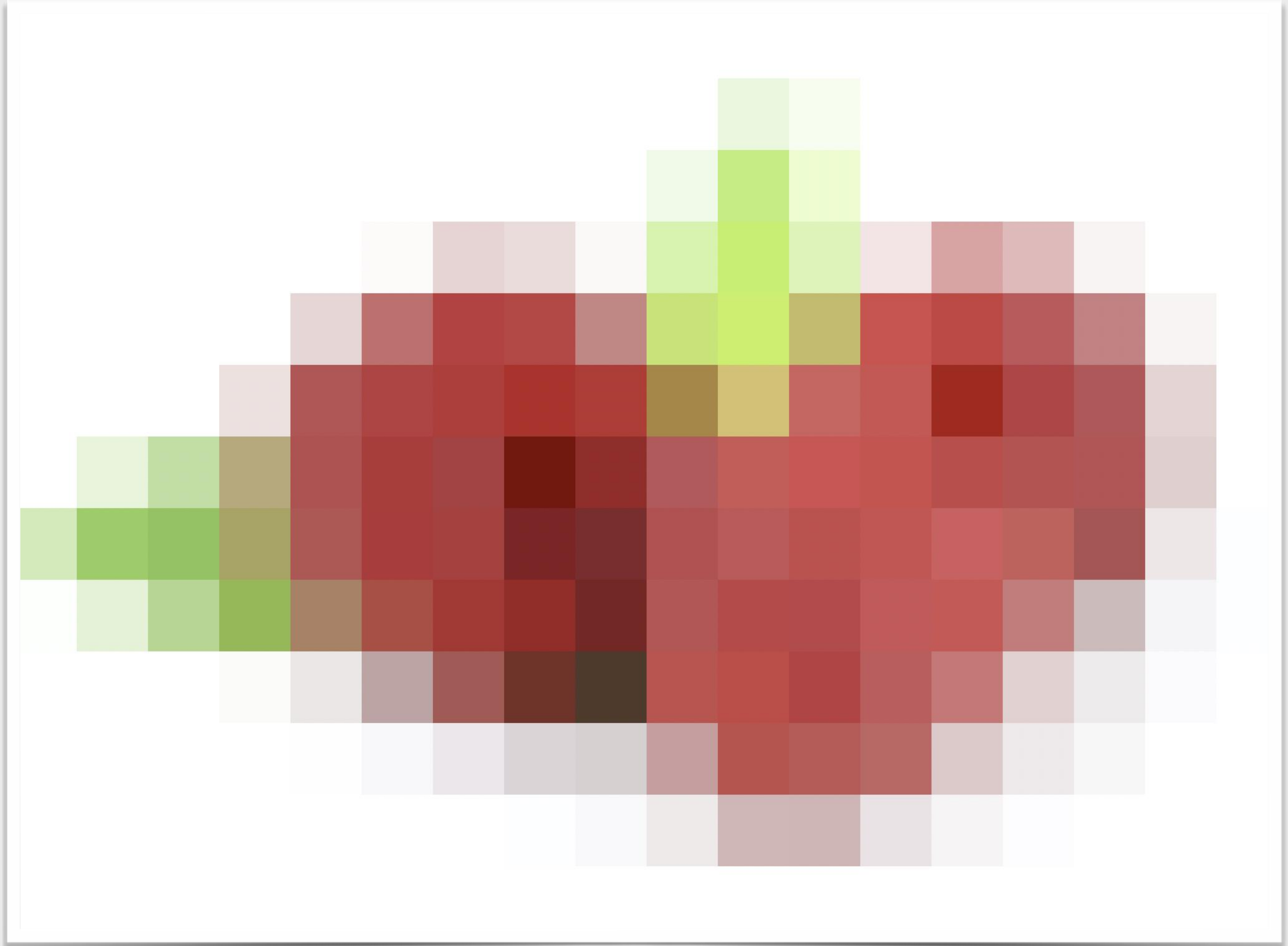
. ○ ○ ○ .

. ○ ○ ○ ○ .

. ○ ○ ○ ○ ○ .

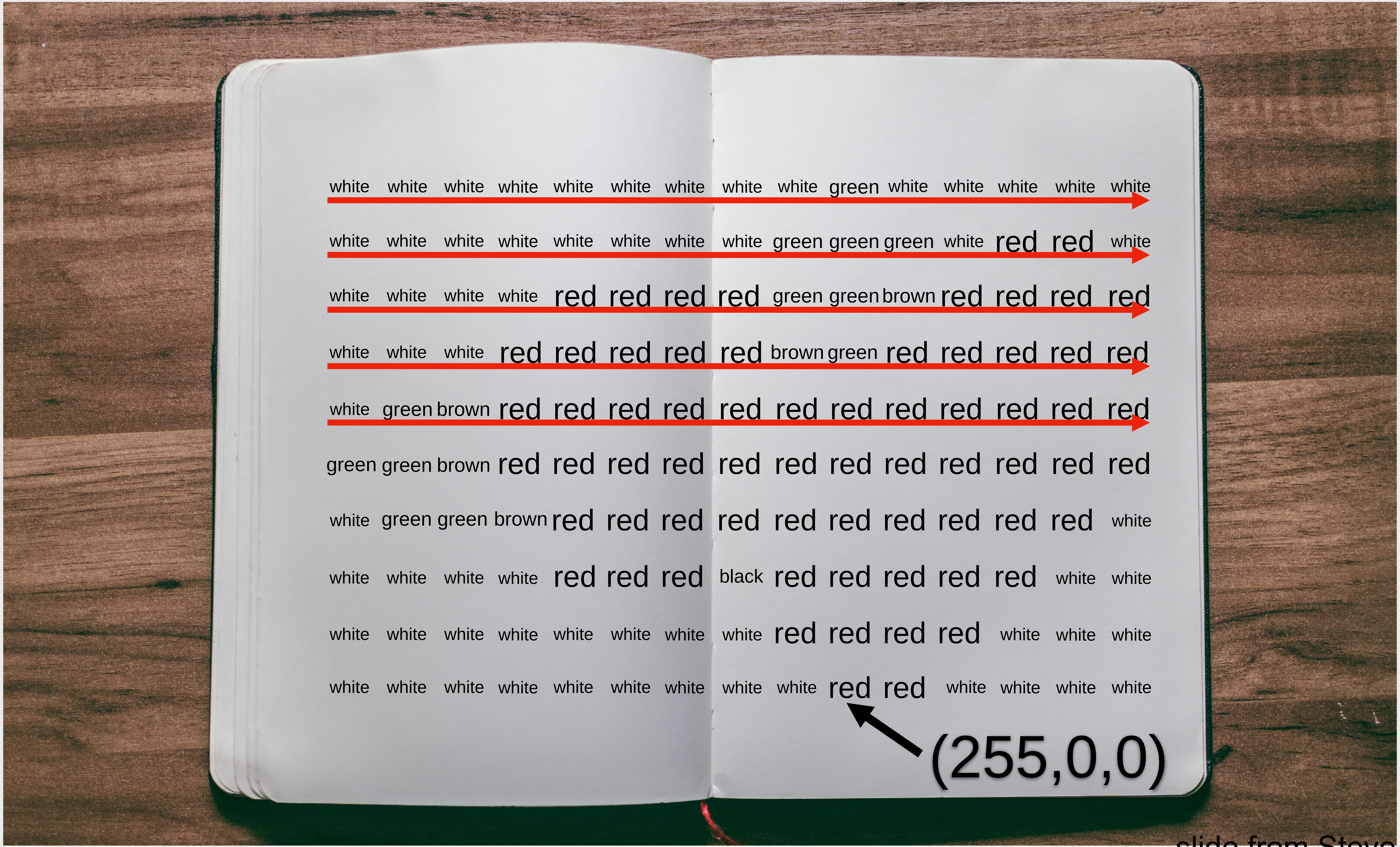
slide from Steve Seitz’s video

slide from Steve Seitz’s [video]

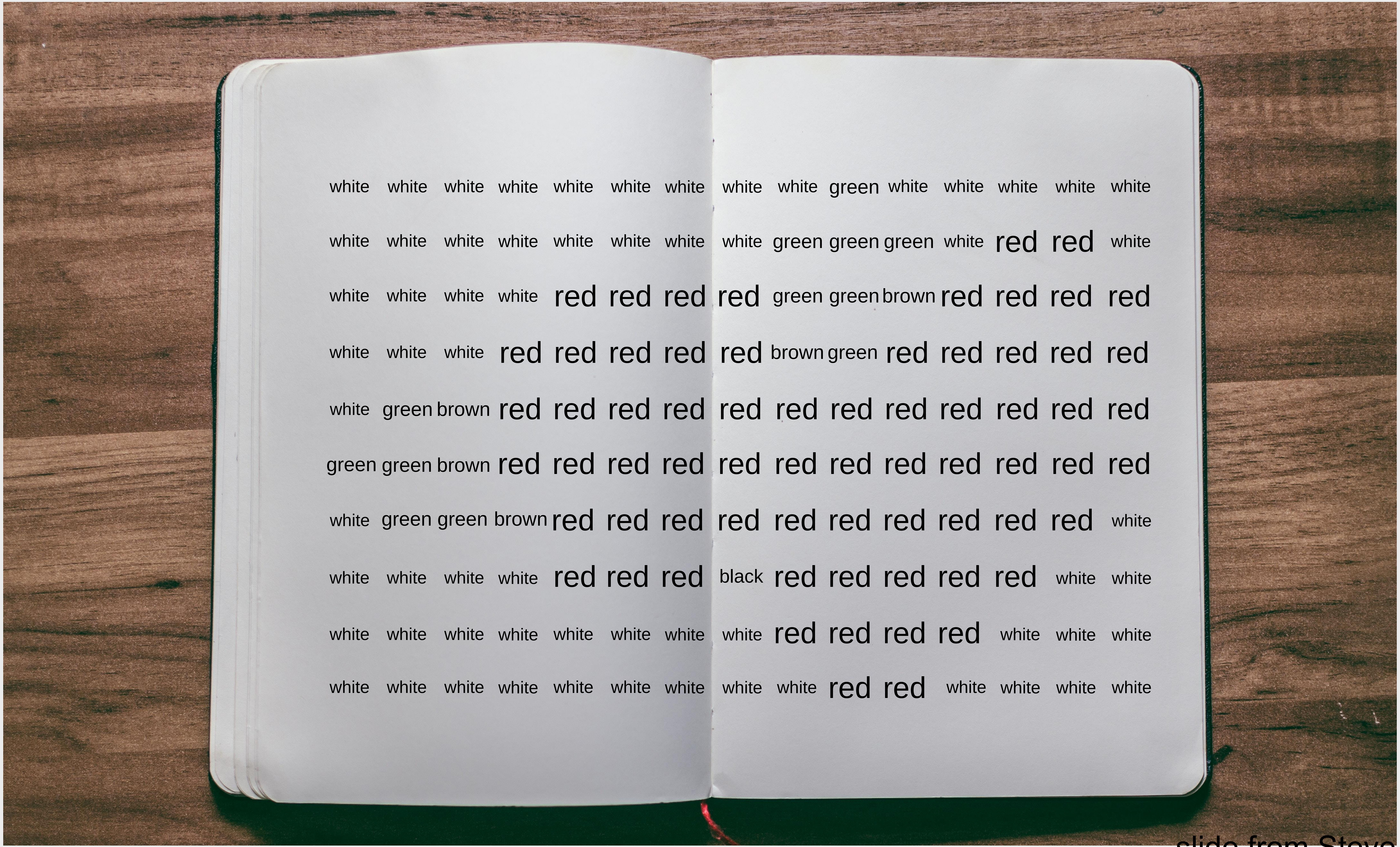


slide from Steve Seitz's video

slide from Steve Seitz's [video]



slide from Steve Seitz's [video]()



slide from Steve Seitz's video

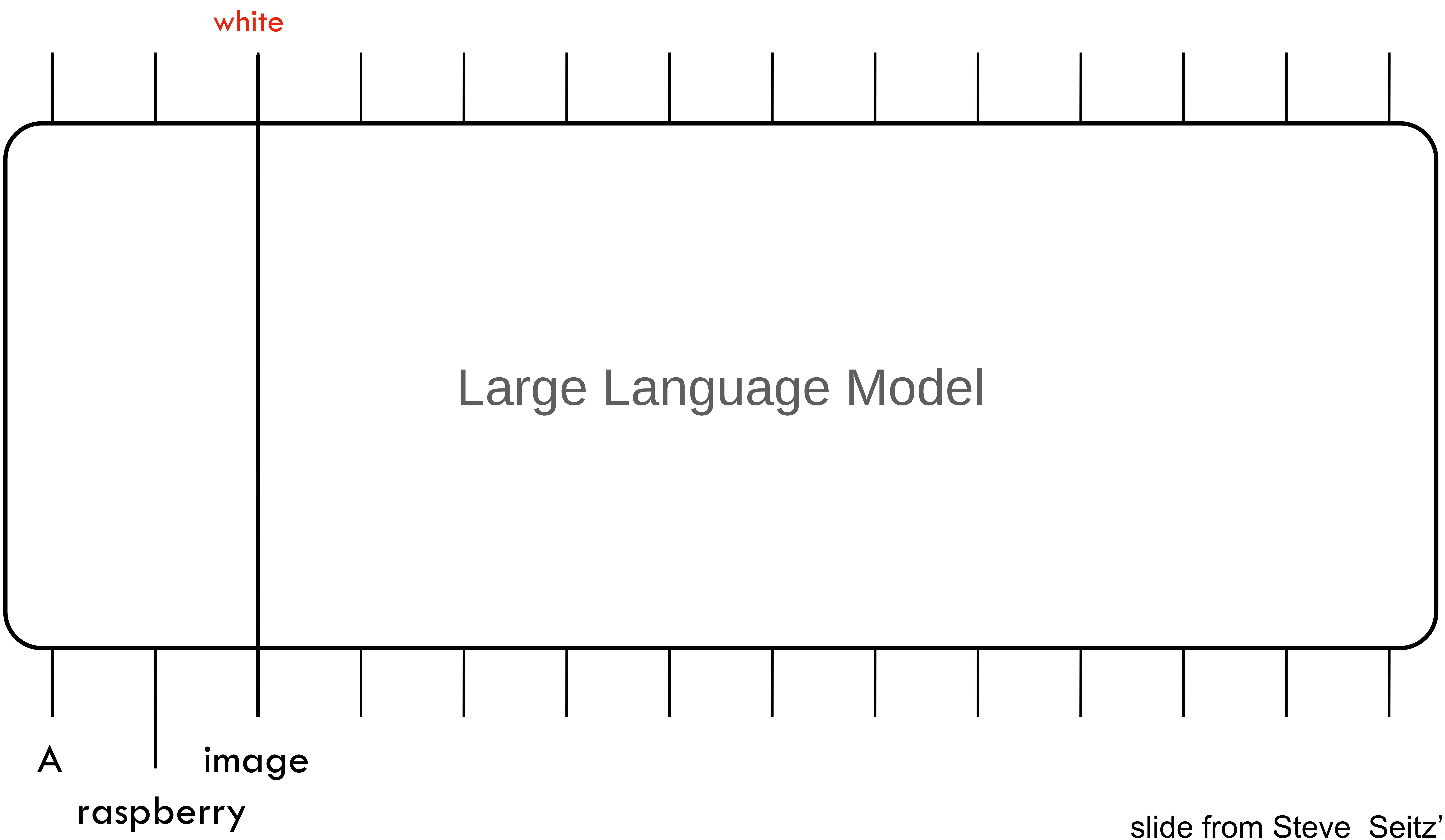
raspberri

an akes

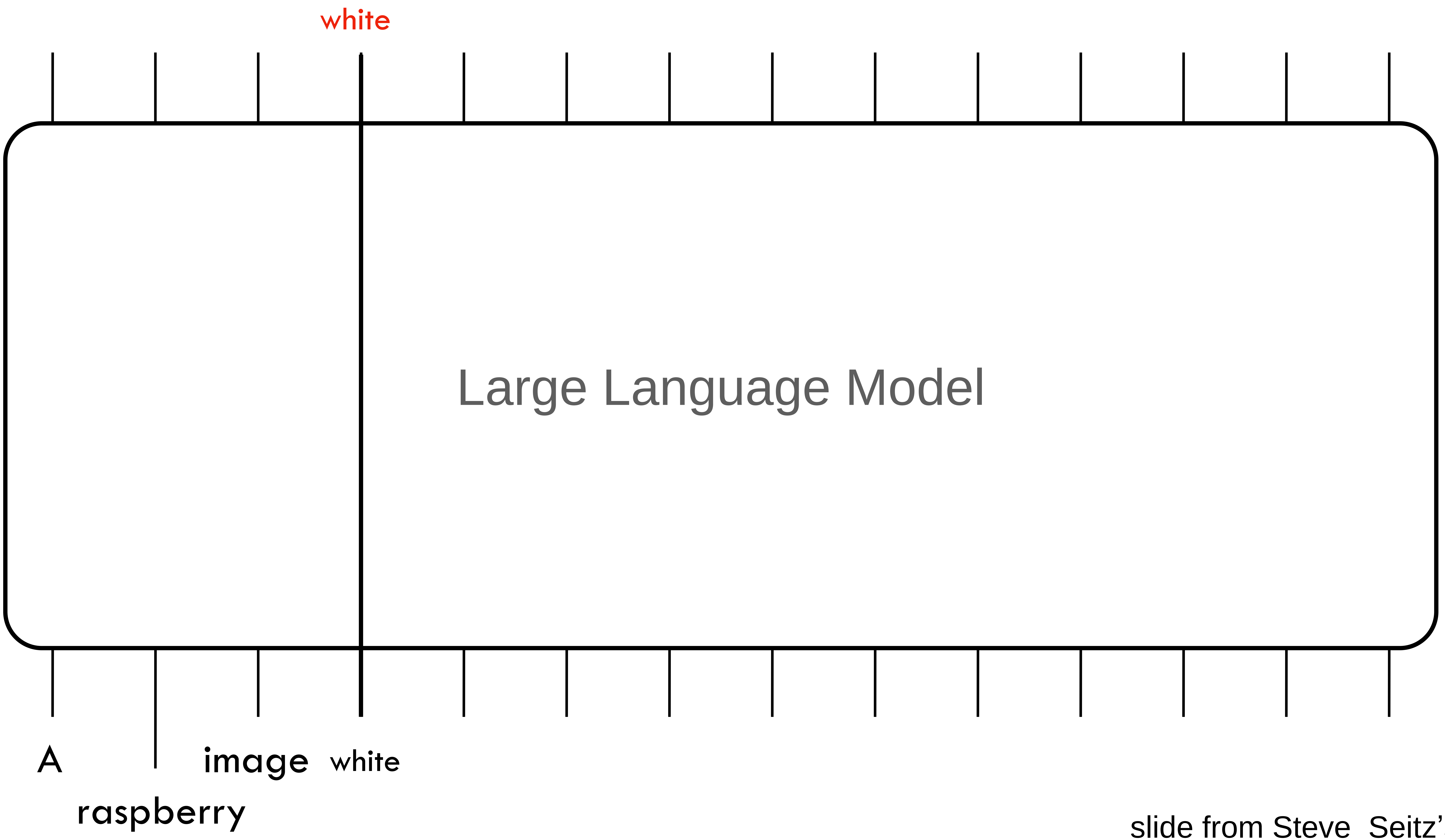
sunsets

1 Billion

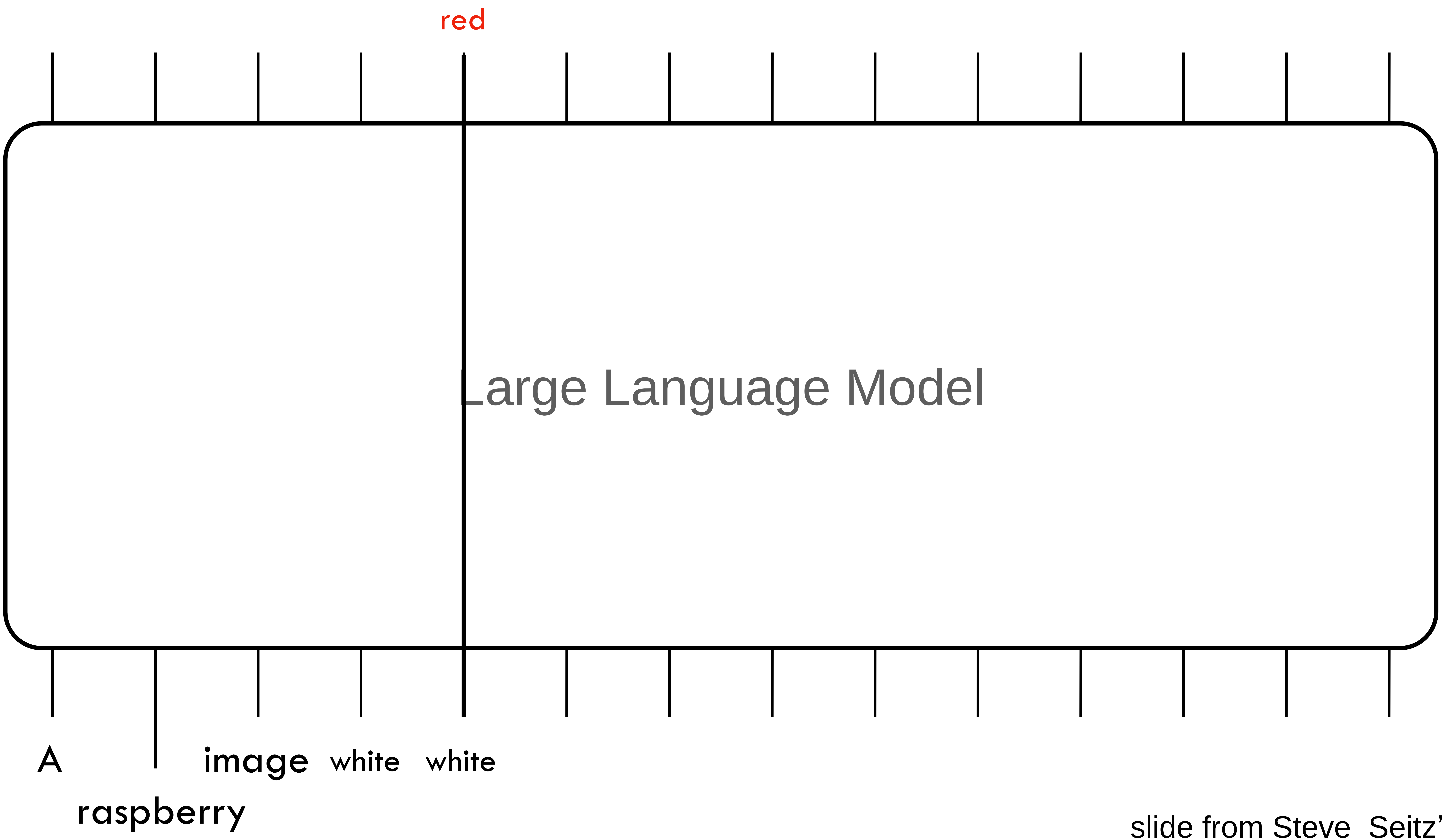
slide from Steve Seitz's [video]



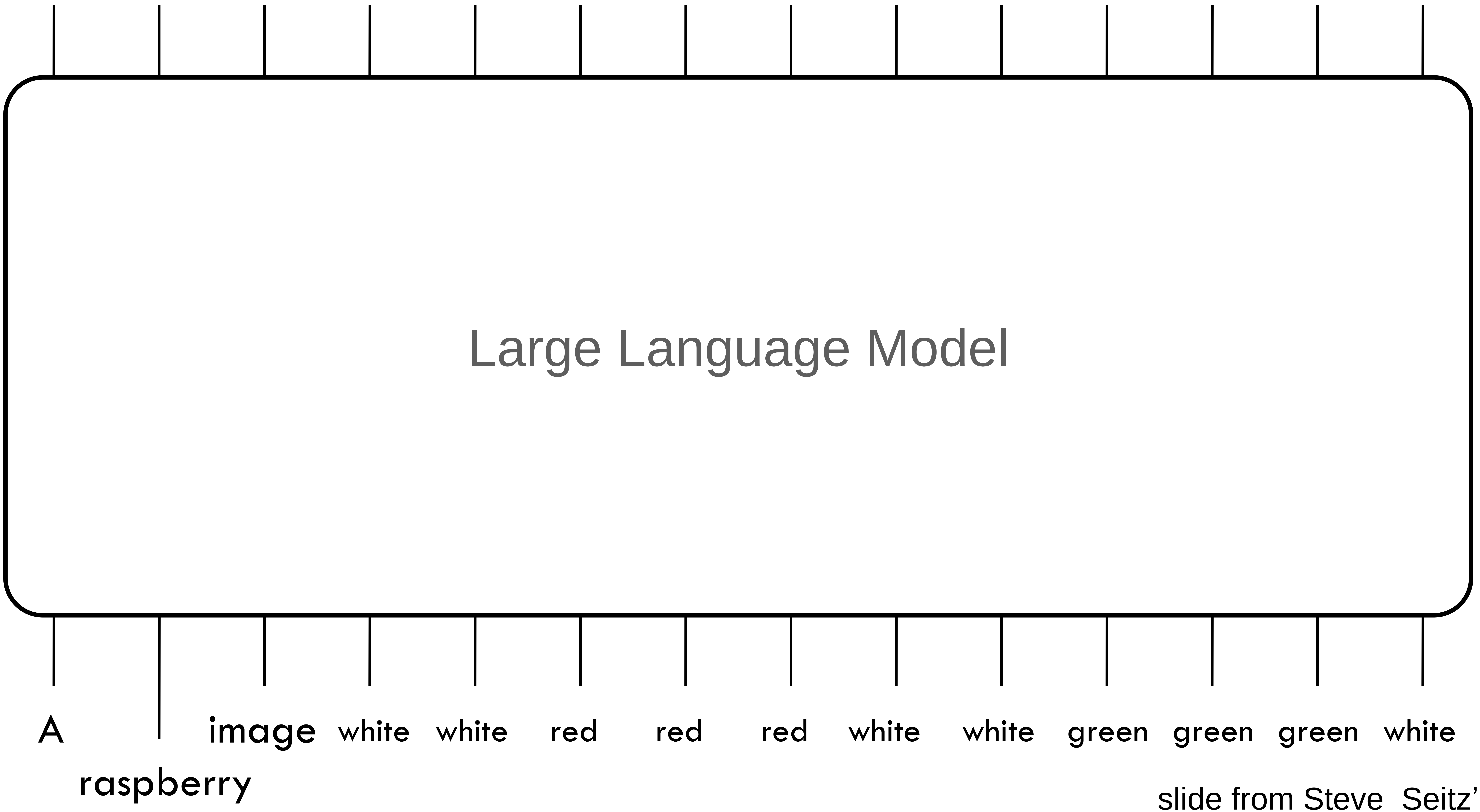
slide from Steve Seitz's video



slide from Steve Seitz's [video]

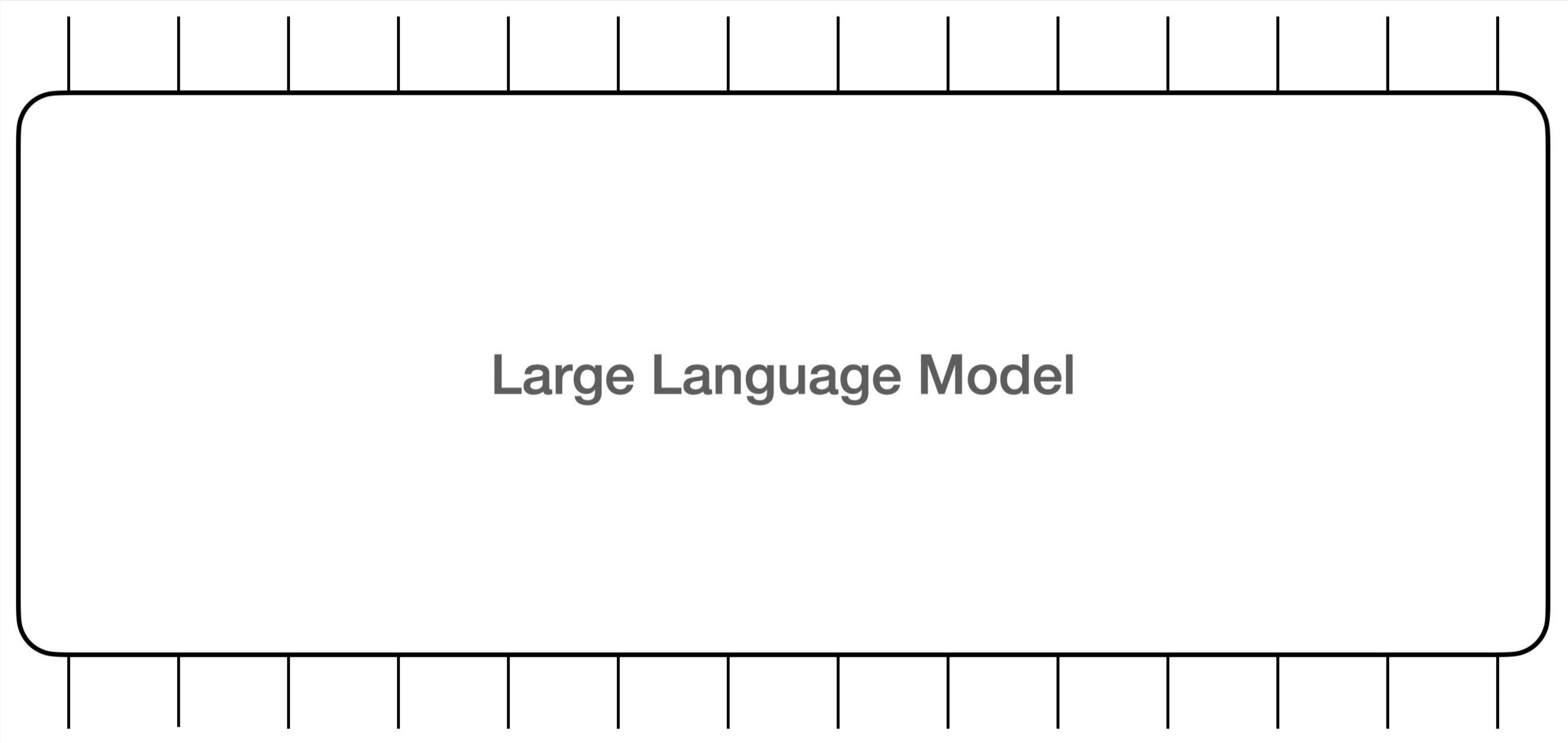


slide from Steve Seitz's [video]



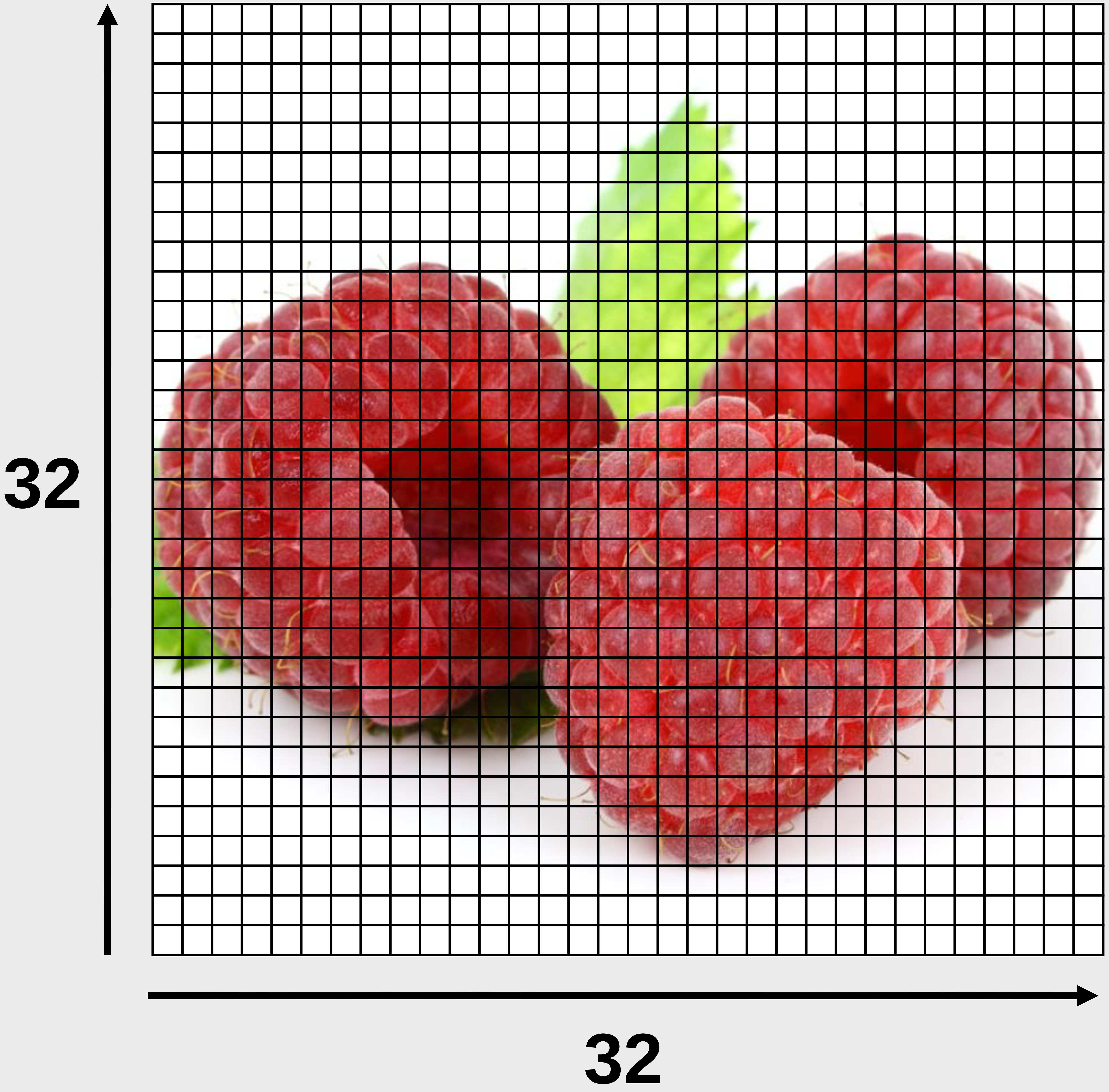
slide from Steve Seitz's video

1,000,000s of pixels



1,000s of words

slide from Steve Seitz’s video



slide from Steve Seitz's video

Torralba et al, 80 Million Tiny Images (2006)

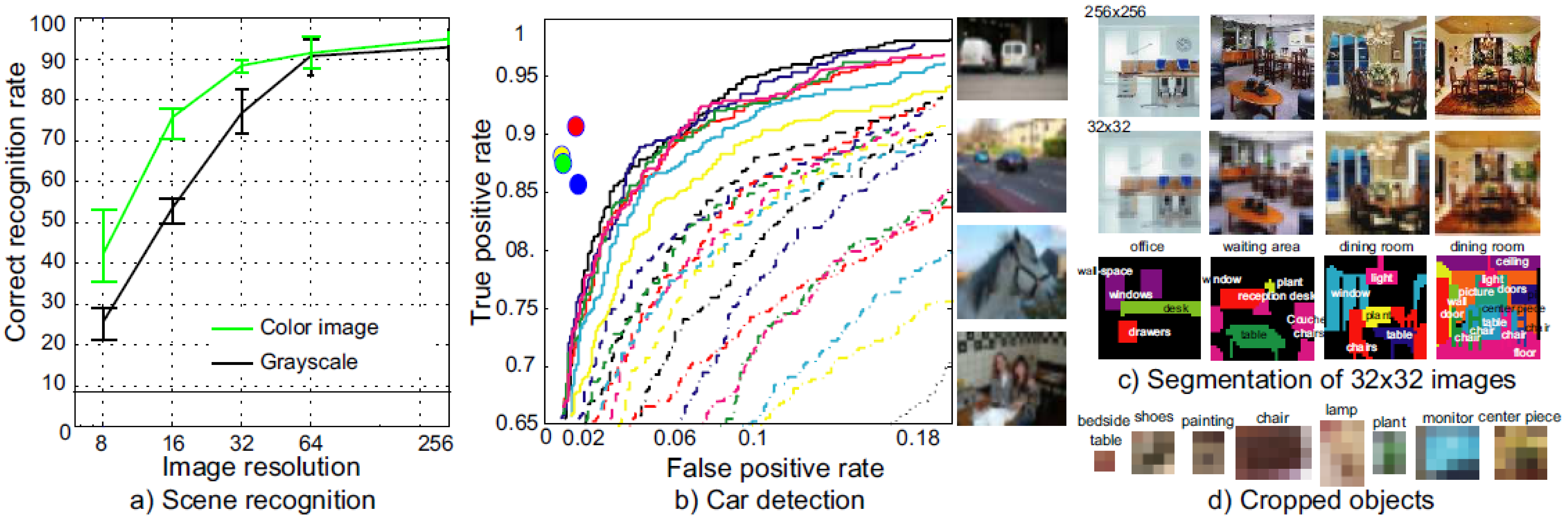


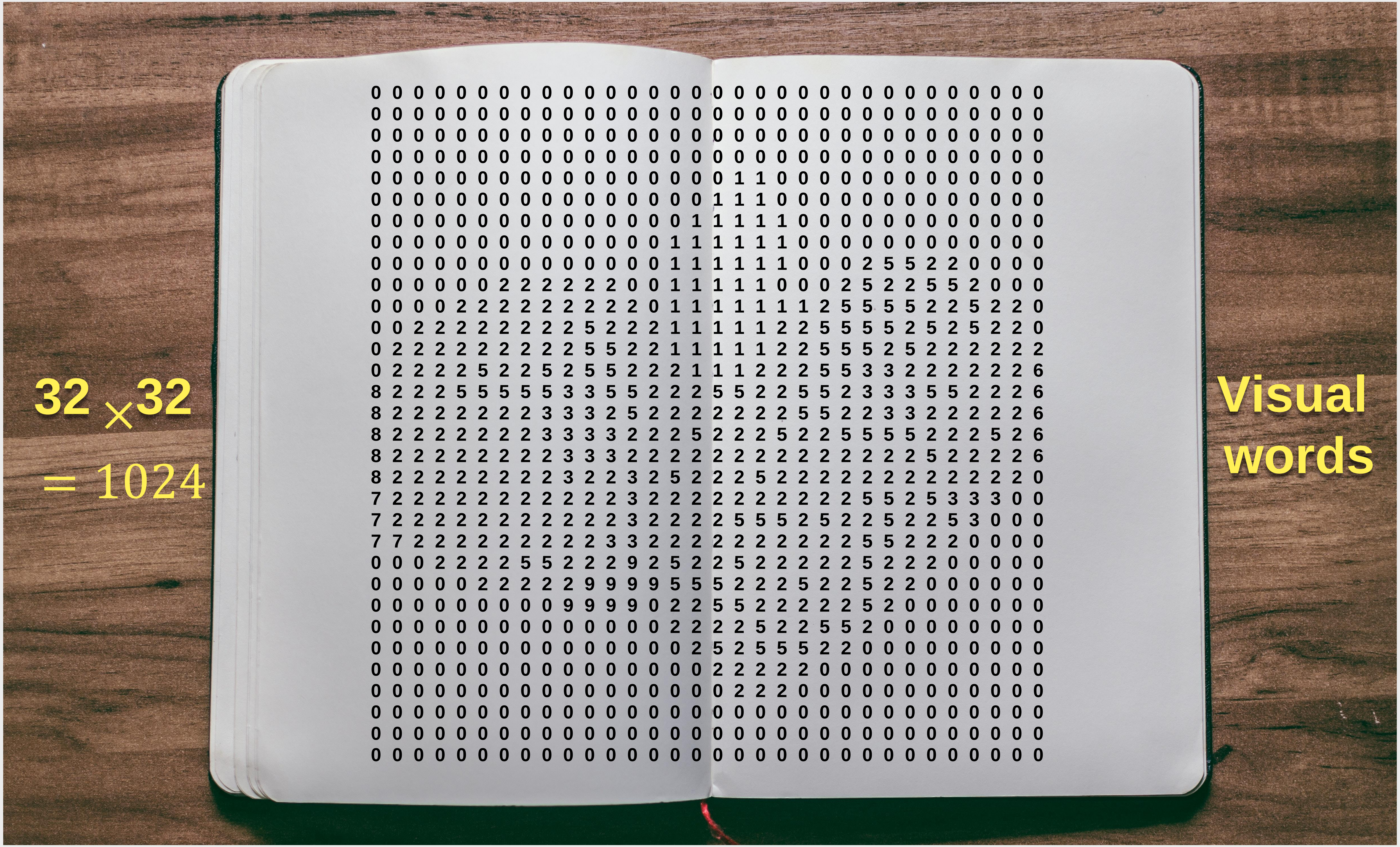
Fig. 1. a) Human performance on scene recognition as a function of resolution. The green and black curves show the performance on color and gray-scale images respectively. For color 32×32 images the performance only drops by 7% relative to full resolution, despite having 1/64th of the pixels. b) Car detection task on the PASCAL 2006 test dataset. The colored dots show the performance of four human subjects classifying tiny versions of the test data. The ROC curves of the best vision algorithms (running on full resolution images) are shown for comparison. All lie below the performance of humans on the tiny images, which rely on none of the high-resolution cues exploited by the computer vision algorithms. c) Humans can correctly recognize and segment objects at very low resolutions, even when the objects in isolation can not be recognized (d).

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	2	5	5	2	2	0	0	0	0
0	0	0	0	0	0	2	2	2	2	2	2	0	0	1	1	1	1	1	0	0	0	2	5	2	2	5	5	2	0	0	0
0	0	0	0	2	2	2	2	2	2	2	2	2	0	1	1	1	1	1	1	1	2	5	5	5	5	2	2	5	2	2	0
0	0	2	2	2	2	2	2	2	2	5	2	2	2	1	1	1	1	1	2	2	5	5	5	5	2	5	2	5	2	2	0
0	2	2	2	2	2	2	2	2	2	5	5	2	2	1	1	1	1	1	2	2	5	5	5	2	5	2	2	2	2	2	2
0	2	2	2	2	5	2	2	5	2	5	5	2	2	2	1	1	1	2	2	2	5	5	3	3	2	2	2	2	2	2	6
8	2	2	2	5	5	5	5	5	3	3	5	5	2	2	2	5	5	2	2	5	5	2	3	3	3	5	5	2	2	2	6
8	2	2	2	2	2	2	2	3	3	3	2	5	2	2	2	2	2	2	2	5	5	2	2	3	3	2	2	2	2	2	6
8	2	2	2	2	2	2	2	3	3	3	3	2	2	2	5	2	2	2	5	2	2	5	5	5	5	2	2	2	5	2	6
8	2	2	2	2	2	2	2	2	3	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	5	2	2	2	2	6
8	2	2	2	2	2	2	2	2	3	2	2	3	2	5	2	2	2	5	2	2	2	2	2	2	2	2	2	2	2	2	0
7	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2	2	2	2	5	5	2	5	3	3	3	0	0
7	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	2	5	5	5	2	5	2	2	5	2	2	5	3	0	0	0
7	7	2	2	2	2	2	2	2	2	2	3	3	2	2	2	2	2	2	2	2	2	2	5	5	2	2	2	0	0	0	0
0	0	0	2	2	2	2	5	5	2	2	2	9	2	5	2	2	5	2	2	2	2	2	5	2	2	2	0	0	0	0	0
0	0	0	0	0	2	2	2	2	2	9	9	9	9	5	5	5	2	2	2	5	2	2	5	2	2	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	9	9	9	9	0	2	2	5	5	2	2	2	2	2	5	2	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	2	5	2	2	5	5	2	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5	2	5	5	5	2	2	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

32

32

slide from Steve Seitz's video



slide from Steve Seitz's [video]

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1

slide from Steve Seitz's [video]()

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 $\boxed{}$

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 $\boxed{}$

slide from Steve Seitz's video

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 **1**

slide from Steve Seitz's [video]

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 **6**

slide from Steve Seitz's [video]

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 **6**

slide from Steve Seitz's video

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 **6**

slide from Steve Seitz's [video]

squirrel reaching for a nut

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 6

slide from Steve Seitz's [video]

squirrel reaching for a nut

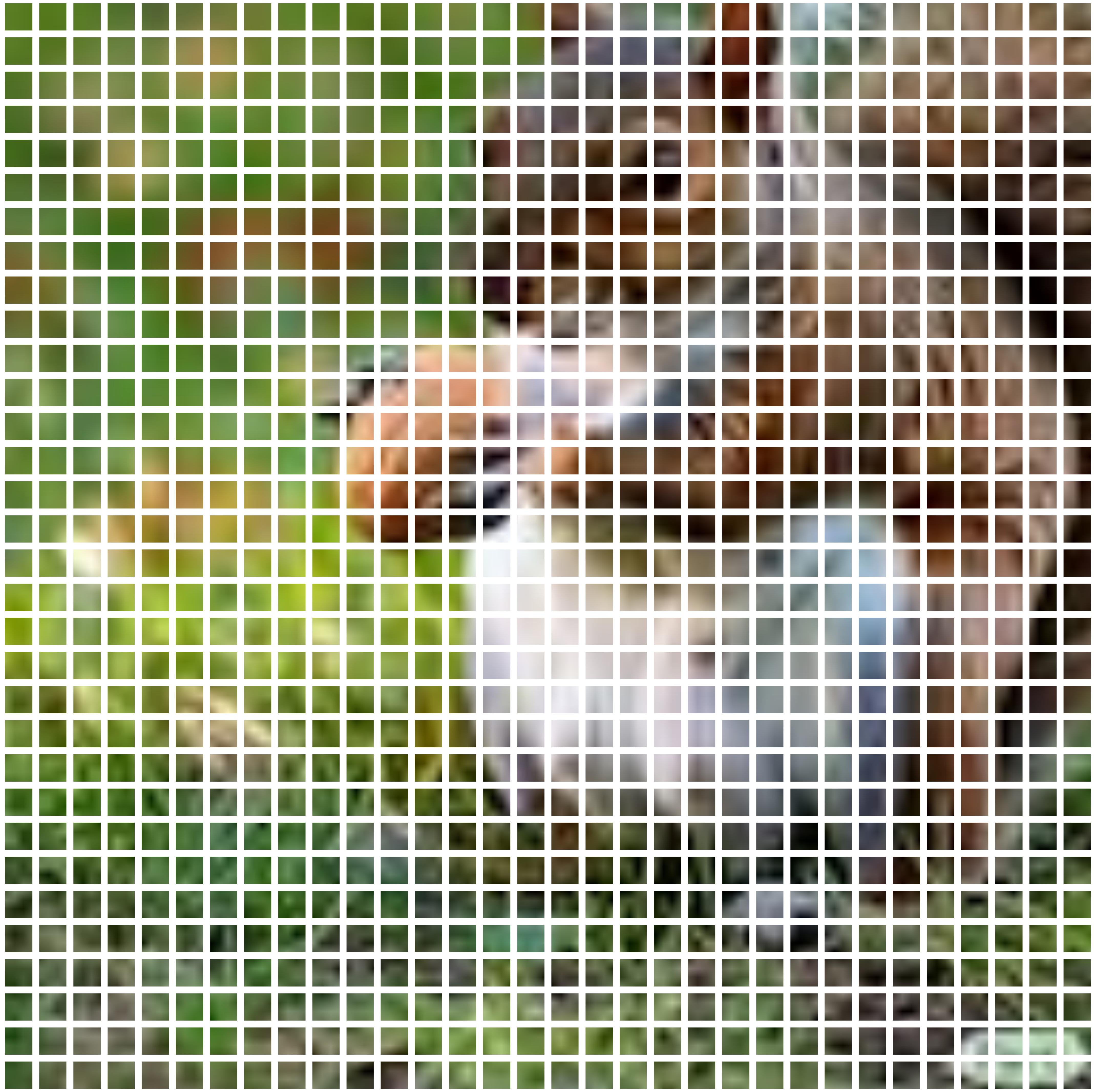
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	6	6	1	1	6	6	4	4	4	9	9	9	9	9	9
1	1	1	1	1	7	7	1	1	1	1	1	1	1	1	1	6	6	6	6	6	6	6	4	1	4	9	9	9	9	9	9
1	1	7	1	1	1	7	1	1	1	1	1	1	1	1	6	6	6	6	6	6	6	6	4	1	4	9	9	9	9	9	9
1	1	7	1	1	1	1	1	1	1	1	1	1	1	2	6	2	6	6	2	2	6	2	9	9	9	9	9	9	9	9	9
1	1	1	7	7	1	1	1	1	1	1	1	1	1	6	6	6	6	2	5	2	2	4	9	9	9	9	9	9	9	9	9
1	1	1	7	1	1	1	1	1	1	1	1	1	1	2	6	6	6	2	5	2	2	0	9	9	9	9	9	9	9	9	9
1	1	1	1	1	1	2	2	2	2	2	2	1	2	6	6	6	2	6	2	6	2	0	9	9	9	9	9	9	9	9	9
2	1	1	1	1	2	2	1	2	2	2	2	1	2	6	6	6	6	6	2	6	4	9	9	9	9	9	9	9	8	8	8
2	2	1	1	1	1	1	1	1	1	1	1	1	1	6	6	2	6	6	6	4	4	9	9	9	9	9	9	9	9	8	8
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	6	0	0	0	0	0	4	9	9	9	9	9	9	9	9	8
1	1	1	1	1	1	1	1	1	7	1	5	2	2	2	0	0	0	0	0	0	4	9	9	9	9	9	9	9	9	9	8
1	1	1	1	1	1	1	1	1	1	5	2	2	2	2	0	0	0	0	4	4	6	9	9	9	9	9	9	9	9	9	9
1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	5	2	0	5	4	6	9	9	9	9	9	9	9	9	9	9	9
1	1	1	1	7	1	1	1	1	1	2	2	6	6	5	5	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
1	1	1	1	7	2	2	2	1	1	2	2	6	5	5	0	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
1	1	1	7	2	2	2	2	1	1	2	6	6	5	5	0	0	4	4	4	9	9	0	0	0	0	9	9	9	9	9	9
1	1	0	1	2	1	2	1	1	1	1	1	1	1	0	0	0	0	4	0	0	0	4	0	0	0	0	9	9	9	9	8
1	1	1	1	1	1	1	1	1	1	1	7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	8
1	1	1	1	1	1	1	1	1	7	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	8	8
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	4	9	9	9	8	9
1	1	1	1	1	7	7	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	4	9	9	5	5	9
1	1	1	1	1	7	7	7	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	4	9	9	5	4	3
1	1	1	1	1	1	7	7	3	3	3	3	3	3	3	4	0	4	0	0	4	0	0	0	0	4	4	9	9	5	4	4
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	5	0	0	0	0	0	0	0	0	4	9	9	9	4	3	3
3	3	3	3	3	3	3	3	3	3	4	4	4	3	3	3	5	3	0	4	4	4	4	4	4	4	9	9	9	4	3	8
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	4	4	9	3	9	4	8	8
3	3	8	3	3	3	3	3	3	3	3	3	3	8	3	8	8	3	3	8	3	8	4	4	8	4	8	3	8	3	3	3
8	8	3	8	8	3	8	3	3	3	8	3	8	3	3	3	3	3	3	8	3	8	4	8	8	8	3	3	3	3	3	3
3	3	8	3	3	3	3	3	3	3	8	8	8	8	3	3	8	8	3	3	8	8	8	8	8	8	8	8	3	8	3	8
3	3	8	8	3	3	3	3	8	3	3	3	3	3	3	3	8	8	3	3	8	3	3	3	8	3	8	8	8	8	3	8
3	3	8	8	3	3	3	8	8	3	8	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	8	8	8	8	8	8
3	3	8	3	3	3	3	3	8	3	8	8	3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	8	8	8	8	8

slide from Steve Seitz's video

squirrel reaching for a nut

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	6	6	1	1	6	6	4	4	4	9	9	9	9	9	9
1	1	1	1	1	7	7	1	1	1	1	1	1	1	1	1	6	6	6	6	6	6	6	4	1	4	9	9	9	9	9	9
1	1	7	1	1	1	7	1	1	1	1	1	1	1	1	6	6	6	6	6	6	6	6	4	1	4	9	9	9	9	9	9
1	1	7	1	1	1	1	1	1	1	1	1	1	1	2	6	2	6	6	2	2	6	2	9	9	9	9	9	9	9	9	9
1	1	1	7	7	1	1	1	1	1	1	1	1	1	6	6	6	6	2	5	2	2	4	9	9	9	9	9	9	9	9	9
1	1	1	7	1	1	1	1	1	1	1	1	1	1	2	6	6	6	2	5	2	2	0	9	9	9	9	9	9	9	9	9
1	1	1	1	1	1	2	2	2	2	2	2	1	2	6	6	6	2	6	2	6	2	0	9	9	9	9	9	9	9	9	9
2	1	1	1	1	2	2	1	2	2	2	2	1	2	6	6	6	6	6	2	6	4	9	9	9	9	9	9	9	8	8	8
2	2	1	1	1	1	1	1	1	1	1	1	1	1	6	6	2	6	6	6	4	4	9	9	9	9	9	9	9	9	8	8
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	6	6	0	0	0	0	0	4	9	9	9	9	9	9	9	9	8
1	1	1	1	1	1	1	1	1	7	1	5	2	2	2	0	0	0	0	0	0	4	9	9	9	9	9	9	9	9	9	8
1	1	1	1	1	1	1	1	1	1	5	2	2	2	2	0	0	0	0	4	4	6	9	9	9	9	9	9	9	9	9	9
1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	5	2	0	5	4	6	9	9	9	9	9	9	9	9	9	9	9
1	1	1	1	7	1	1	1	1	1	2	2	6	6	5	5	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
1	1	1	1	7	2	2	2	1	1	2	2	6	5	5	0	2	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9
1	1	1	7	2	2	2	2	1	1	2	6	6	5	5	0	0	4	4	4	9	9	0	0	0	0	9	9	9	9	9	9
1	1	0	1	2	1	2	1	1	1	1	1	1	1	0	0	0	0	4	0	0	0	4	0	0	0	0	9	9	9	9	8
1	1	1	1	1	1	1	1	1	1	1	7	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	9	8
1	1	1	1	1	1	1	1	1	7	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	9	9	9	8	8
1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	4	9	9	9	8	9
1	1	1	1	1	7	7	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	4	9	9	5	5	9
1	1	1	1	1	7	7	7	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	4	4	9	9	5	4	3
1	1	1	1	1	1	7	7	3	3	3	3	3	3	3	4	0	4	0	0	4	0	0	0	0	4	4	9	9	5	4	4
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	5	0	0	0	0	0	0	0	0	4	9	9	9	4	3	3
3	3	3	3	3	3	3	3	3	3	4	4	4	3	3	3	5	3	0	4	4	4	4	4	4	4	9	9	9	4	3	8
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	4	4	4	4	4	9	3	9	4	8	8
3	3	8	3	3	3	3	3	3	3	3	3	3	8	3	8	8	3	3	8	3	8	4	4	8	4	8	3	8	3	3	3
8	8	3	8	8	3	8	3	3	3	8	3	8	3	3	3	3	3	3	8	3	8	4	8	8	8	3	3	3	3	3	3
3	3	8	3	3	3	3	3	3	3	8	8	8	8	3	3	8	8	3	3	8	8	8	8	8	8	8	8	3	8	3	8
3	3	8	8	3	3	3	3	8	3	3	3	3	3	3	3	8	8	3	3	8	3	3	3	8	3	8	8	8	8	3	8
3	3	8	8	3	3	3	8	8	3	8	3	3	3	3	3	3	3	3	3	3	3	3	3	8	3	8	8	8	8	8	8
3	3	8	3	3	3	3	3	8	3	8	8	3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	8	8	8	8	8

slide from Steve Seitz's video



squirrel reaching for a nut

slide from Steve Seitz's video

squirrel reaching for a nut

slide from Steve Seitz's video

squirrel reaching for a nut

Up-sampled 4x

slide from Steve Seitz's video

squirrel reaching for a nut

Parti, https://parti.research.google/

squirrel reaching for a nut underwater

slide from Steve Seitz's video

fossil of a squirrel reaching for a nut

slide from Steve Seitz's [video]

squirrel made of toothpicks wearing sunglasses reaching for a nut

slide from Steve Seitz’s video

DLSR photograph of a whimsical fantasy house shaped like a squirrel
with windows and a door, in the forest

slide from Steve Seitz's [video]

Squirrel reaching for a nut. by Leonardo da Vinci

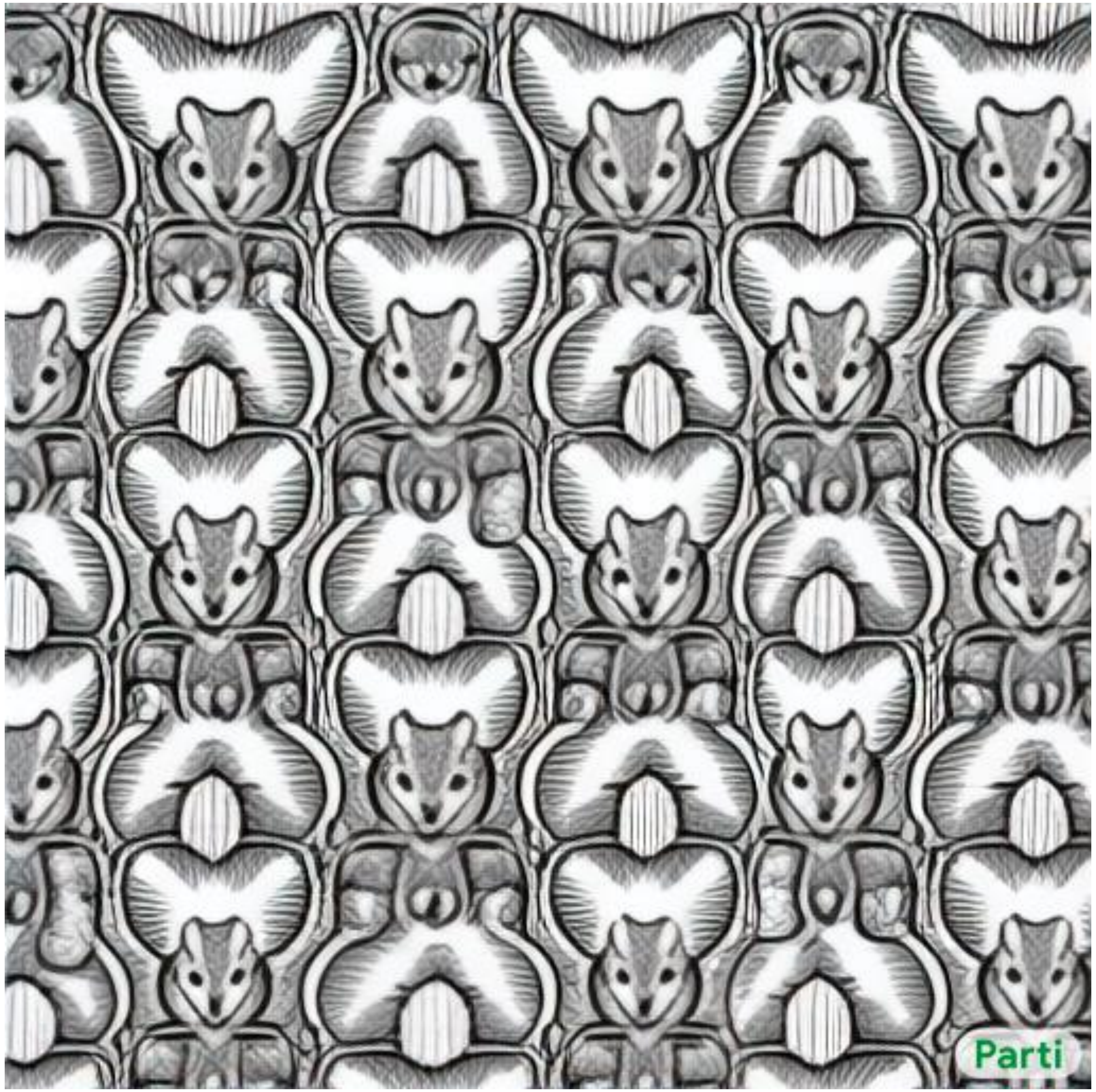
slide from Steve Seitz's video

Squirrel reaching for a nut. Van Gogh painting

slide from Steve Seitz's [video]

Intricately carved cathedral door of a squirrel reaching for a nut

slide from Steve Seitz's video



Squirrel reaching for a nut.   Woodcut tessellation pattern by M.C. Escher

slide from Steve Seitz's [video]

Squirrel reaching for a nut. Latte art

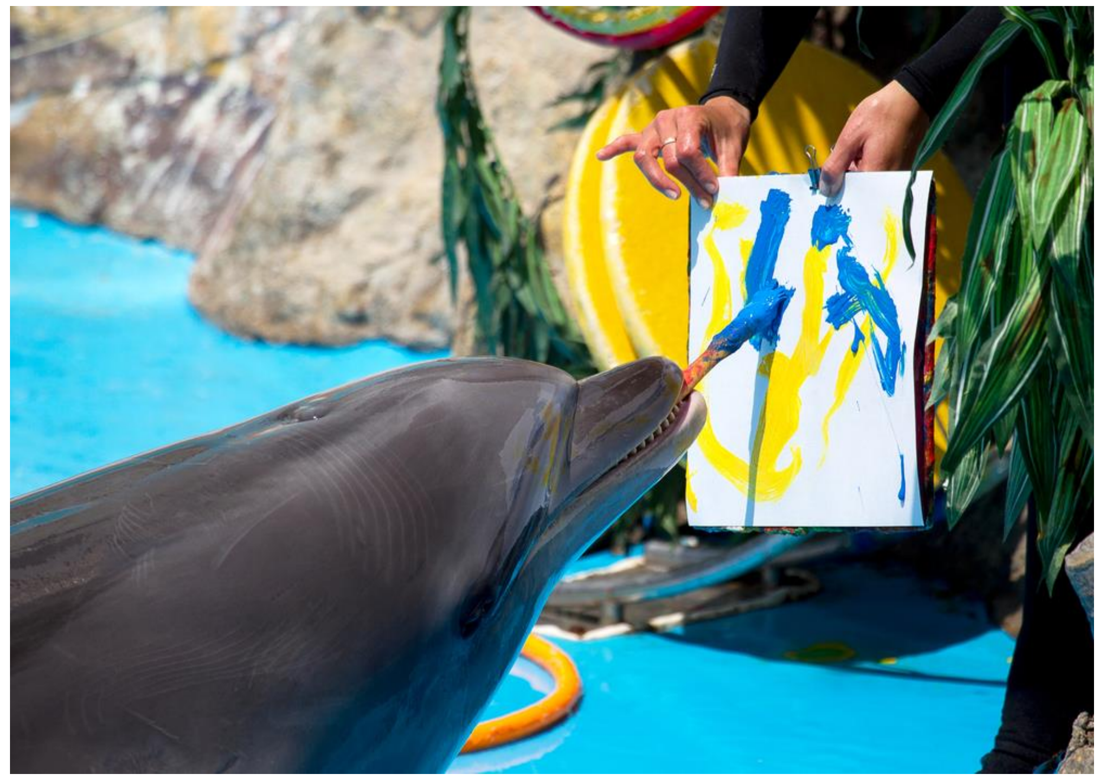
slide from Steve Seitz's [video]()

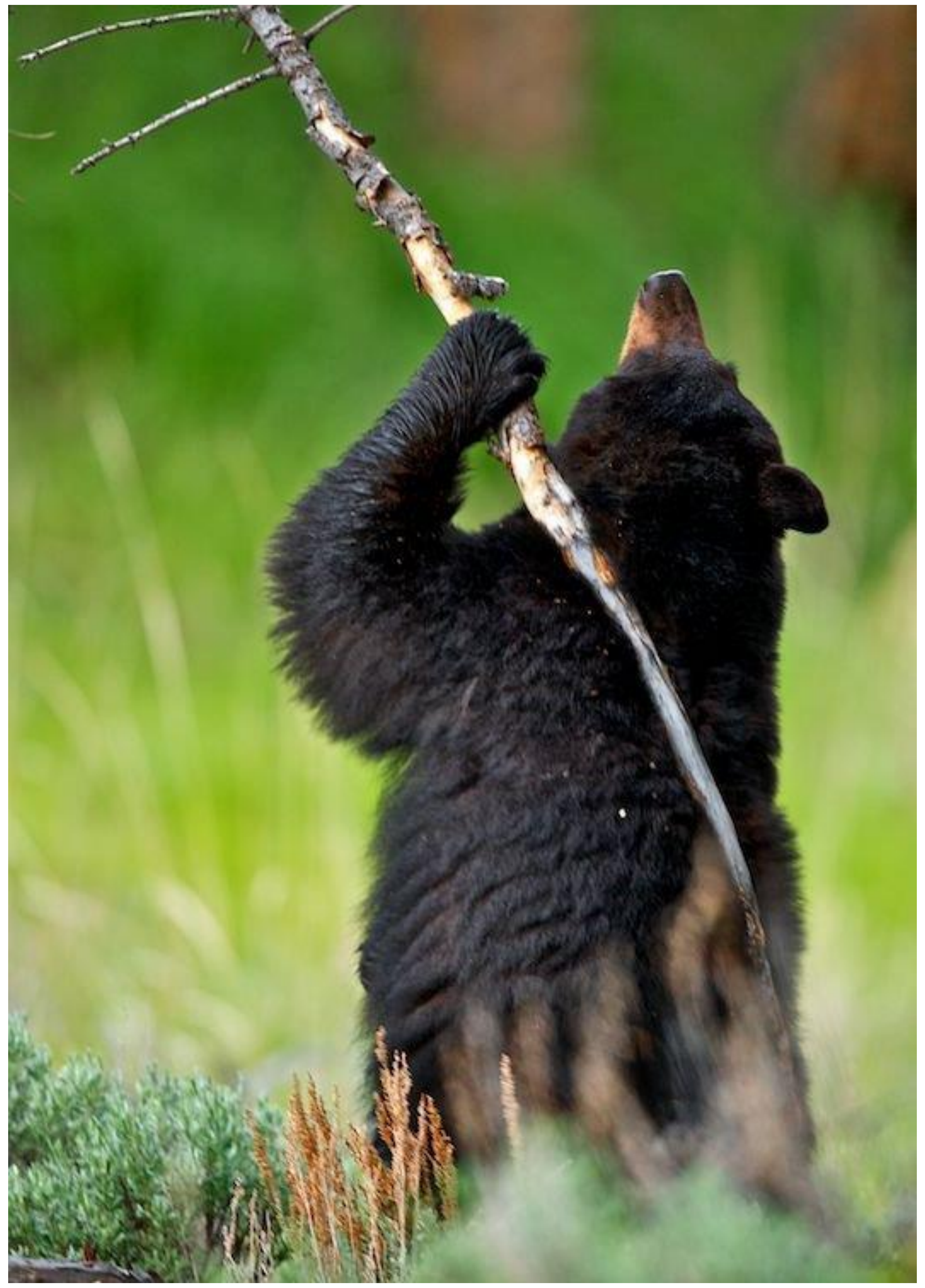
Sequential Modeling Enables Scalable Learning for Large Vision Models

Yutong Bai, Xinyang Geng,
Karttikeya Mangalam, Amir Bar,
Alan Yuille, Trevor Darrell, Jitendra Malik, Alexei A Efros

In CVPR’24

Top-down vs. bottom-up





Scientific Question:
How far can we go from pixels **alone**?

across:
images,
videos,
supervised / unsupervised
synthetic /real,
all kinds of tasks
2D / 3D / 4D data etc.

Dataset	Tokens (Millions)	Annotation Type	Annotation Source
Unpaired Image Data			
LAION 5B [71] (1.5B images subset)	380690	-	-
Images with Annotations			
ImageNet 1K [25]	1317.40	Image Classification	Ground Truth
COCO [54]	363	Object Detection	MMDetection [16]
ADE 20K [100], Cityscapes [22]	66.88	Semantic Segmentation	Ground Truth
COCO [54], ImageNet 1K [25]	2078.06	Semantic Segmentation	Mask2Former [19]
COCO [54], lvmhp [51], mpii [4], Unite [49]	950.79	Human Pose	MMPose[21]
COCO [54], ImageNet 1K [25]	1623.85	Depth Map Image	DPT [67]
Subset of InstructPix2Pix [34]	415.46	Style Transfer	InstructPix2Pix [34]
COCO[54], ImageNet 1K[25]	1623.85	Surface Normal Image	NLL-AngMF [7]
COCO [54], ImageNet 1K [25]	1623.85	Edge Detection	DexiNed [79]
DID-MDN [98]	35.06	Rainy and Clean Image Pairs	Ground Truth
SIDD [3]	245.76	Denoised Image	Ground Truth
LOL[89]	0.458	Light Enhanced Image	Ground Truth
ImageNet 1K [25]	1321.07	Grayscale and Colorized Image Pairs	Ground Truth
ImageNet 1K [25]	1321.07	Inpainting	Ground Truth
Kitti [34]	9.21	Stereo	Ground Truth
Videos			
UCF101 [78]	109.11	-	-
DAVIS [65]	0.36	-	-
HMDB [48]	55.41	-	-
ActivityNet [13]	380.63	-	-
Moments in Time [59]	2979.00	-	-
Multi-moments in Time [60]	4124.04	-	-
Co3D [69]	228.75	-	-
Charades v1 [76]	241.53	-	-
Something-something v2 [37]	904.57	-	-
YouCook [23]	3.14	-	-
Kinetics 700 [14]	7092.04	-	-
MSR-VTT [92]	57.34	-	-
Youtube VOS [93]	63.70	-	-
jester [57]	606.47	-	-
diving48 [52]	150.73	-	-
MultiSports [53]	78.44	-	-
CharadesEgo [77]	193.06	-	-
AVA [61]	117.96	-	-
Ego4D [38]	1152.12	-	-
Videos with Annotations			
VIPSeg [58]	64.47	Video Panoptic Segmentation	Ground Truth
Hand14K [32]	1.96	Hand Segmentation	Ground Truth
AVA [61]	122.88	Video Detection	Ground Truth
JHMDB [43]	19.00	Optical Flow	Ground Truth
JHMDB [43]	37.92	Video Human Pose	Ground Truth
Synthetic 3D Views			
Objaverse [24] Rendered Multiviews	217.85	-	-

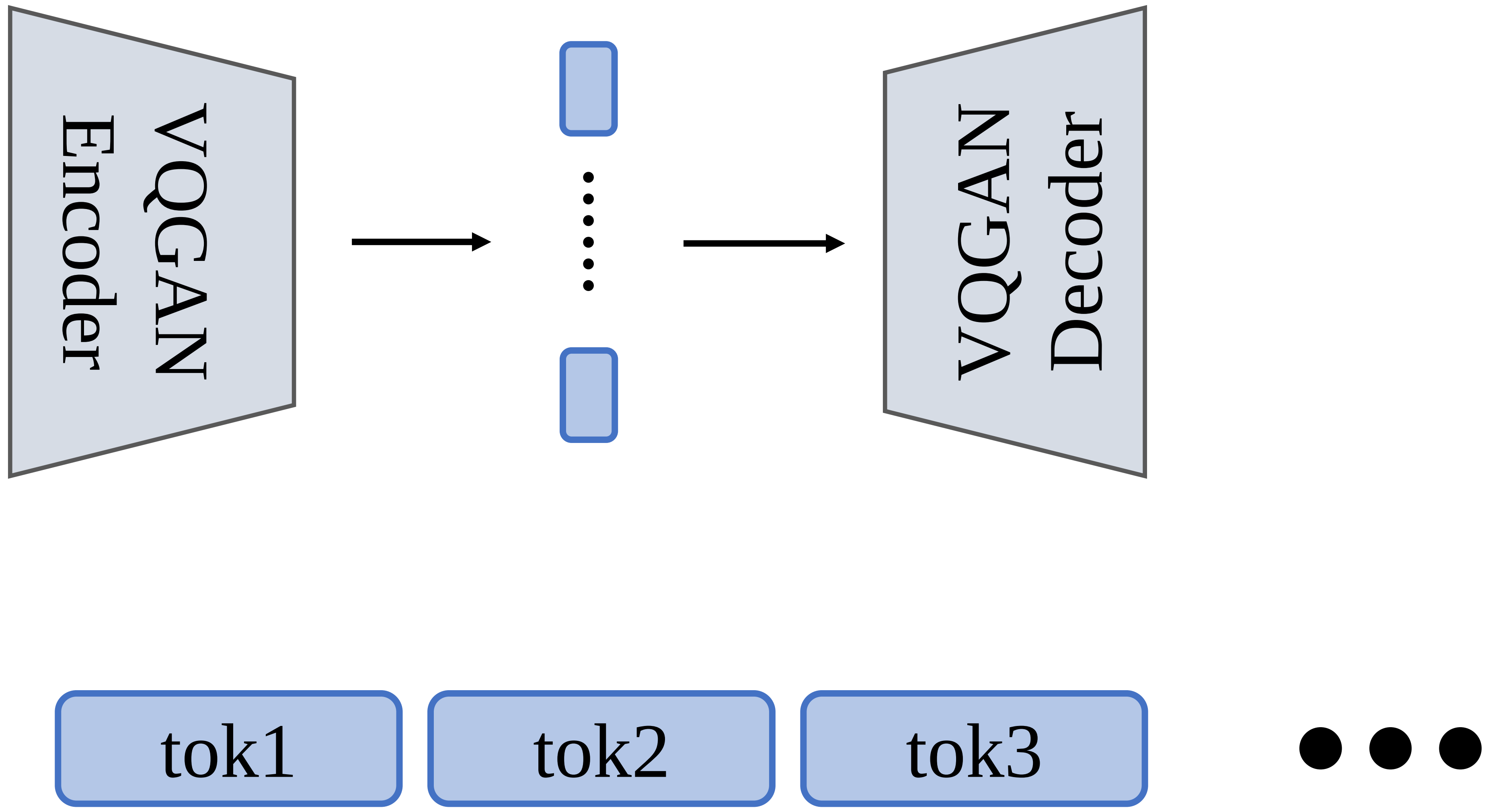
Sentence -> Visual Sentence

Single images

<BOS>

<EOS>

Tokenizer



Videos

<BOS>  ... <EOS>

Image sequences

<BOS>  **…** **<EOS>**

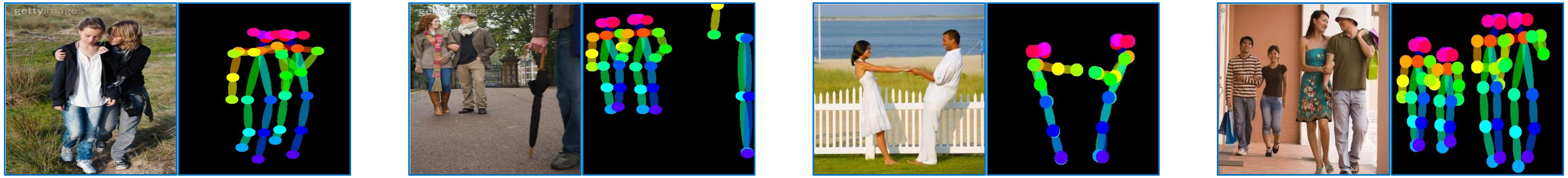
categories

<BOS>

… <EOS>

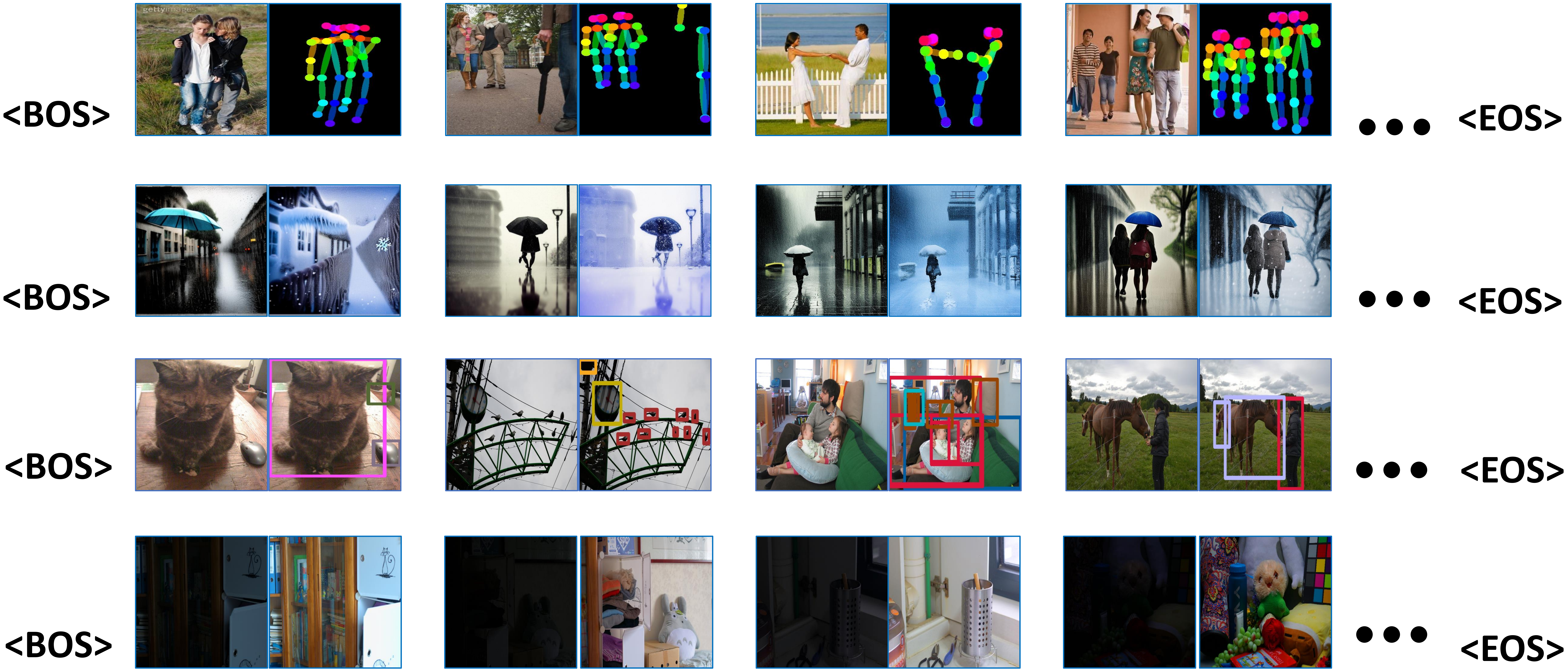
Images with annotation

<BOS>



… <EOS>

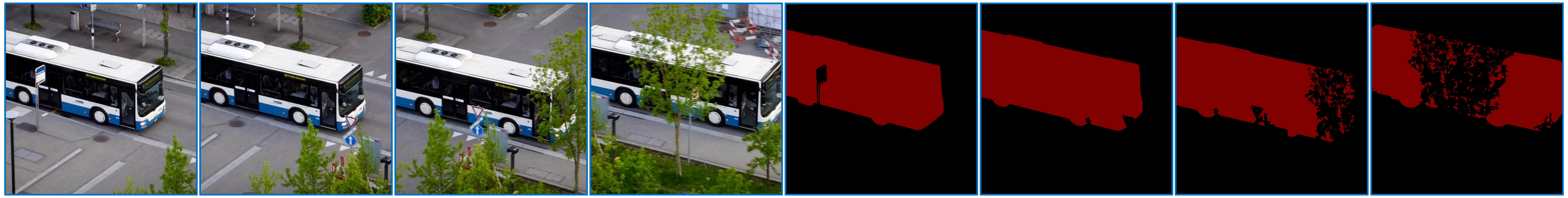
Images with annotation



Images with free form annotation

Videos with annotation

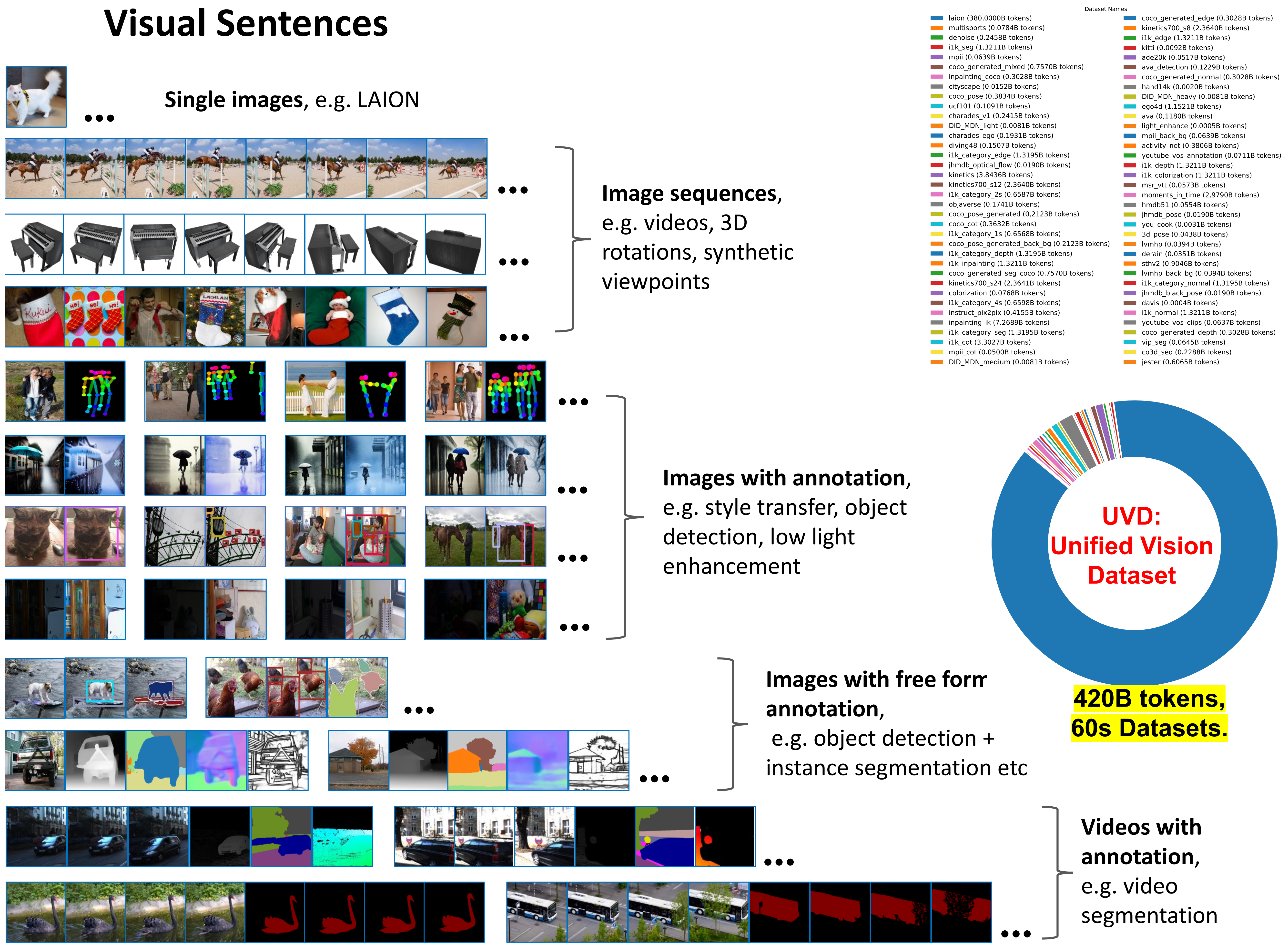
<BOS>

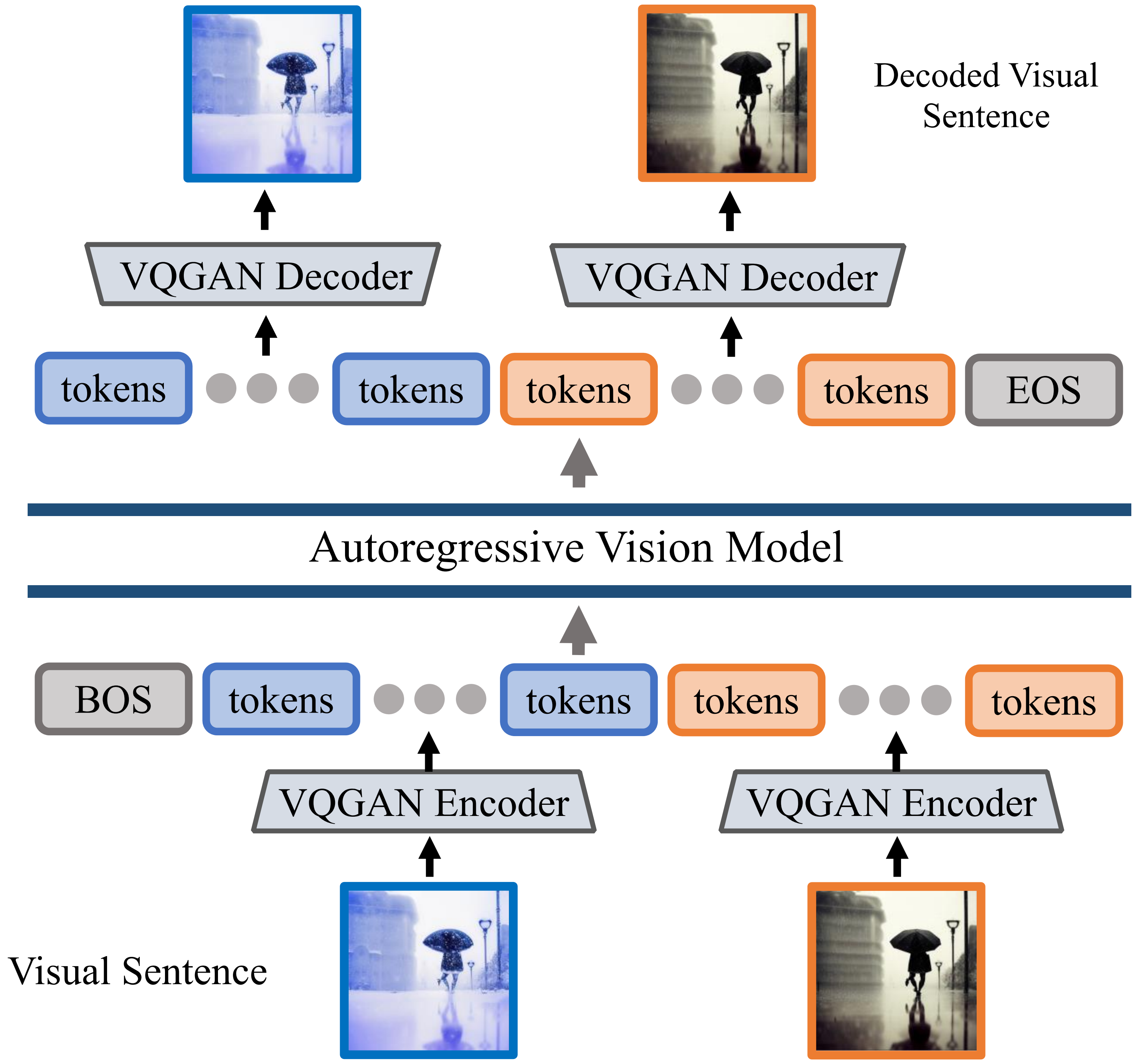


<EOS>

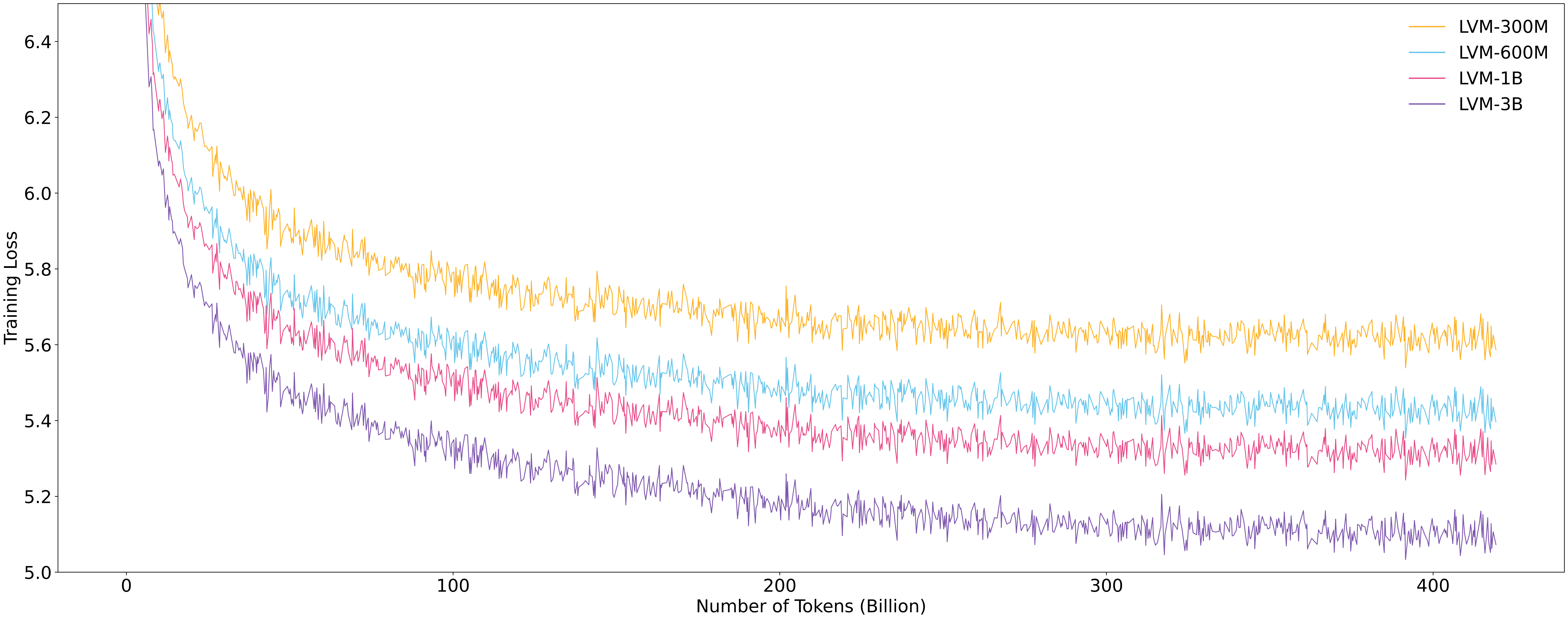
LVM: Large Vision Model

Visual Sentences



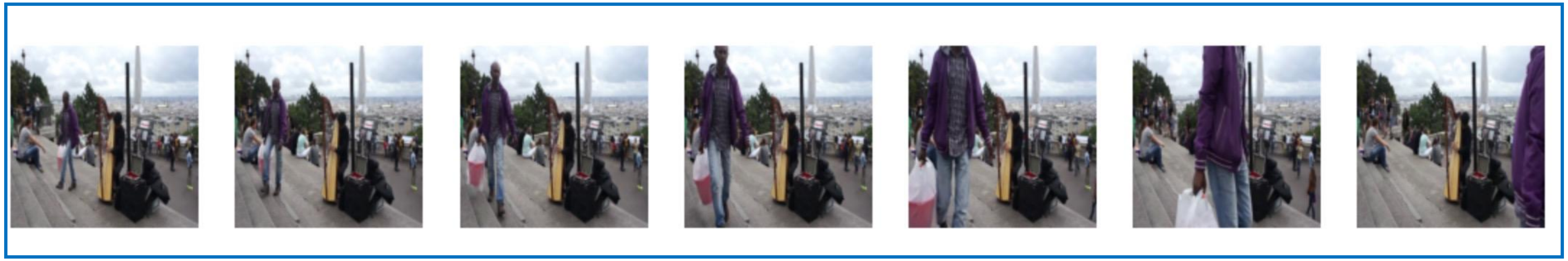


Training Loss (1 epoch) ~ Validation Loss

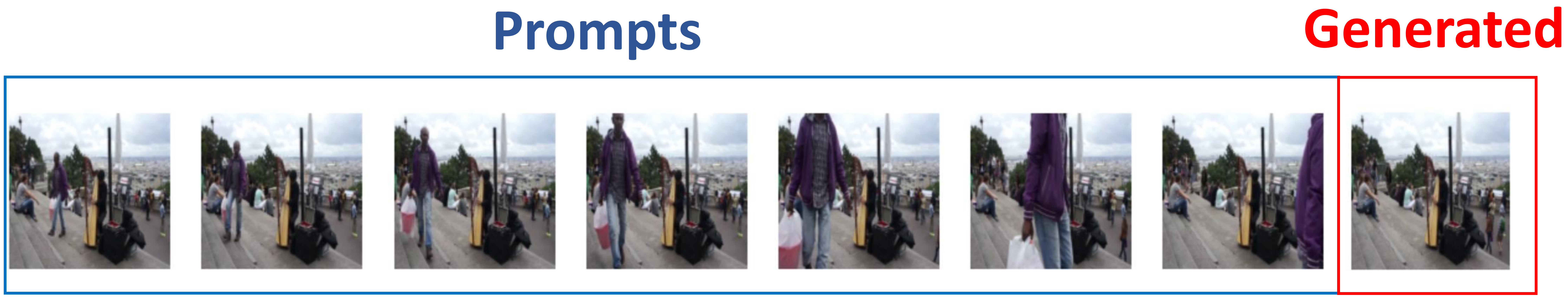


Sequential Prompting

Prompts



Sequential Prompting

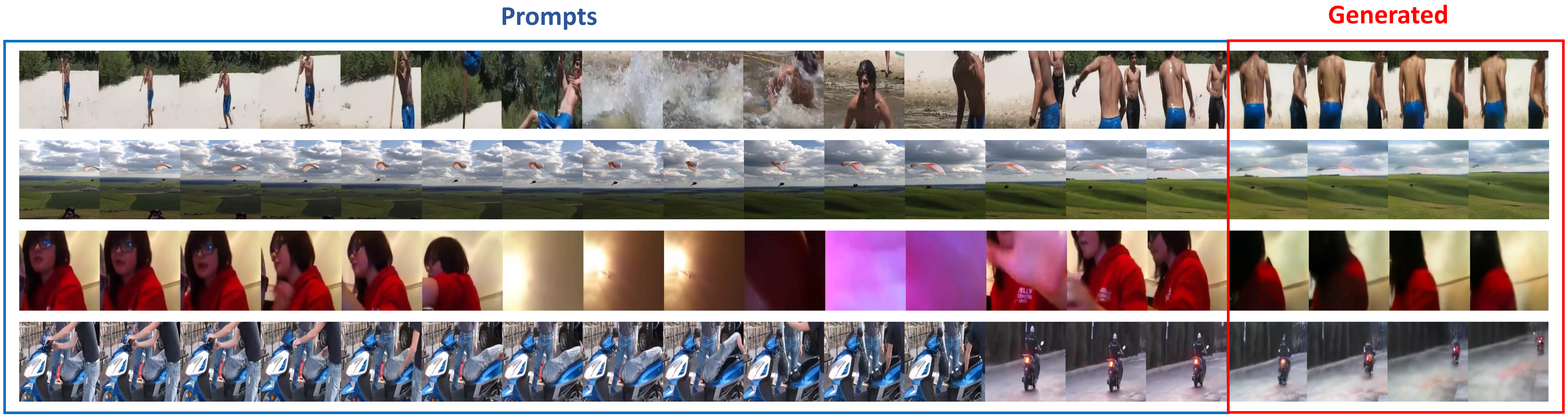


Sequential Prompting

Sequential Prompting

Longer contexts

Sequential Prompting



Sequential Prompting

Analogy Prompting

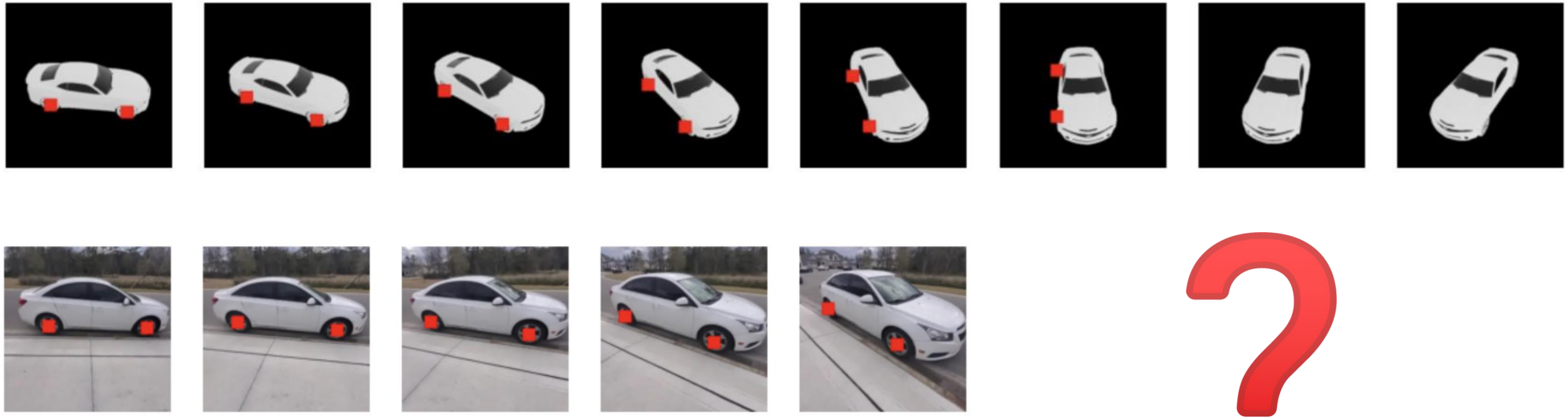
Analogy Prompting


Prompts
Generated

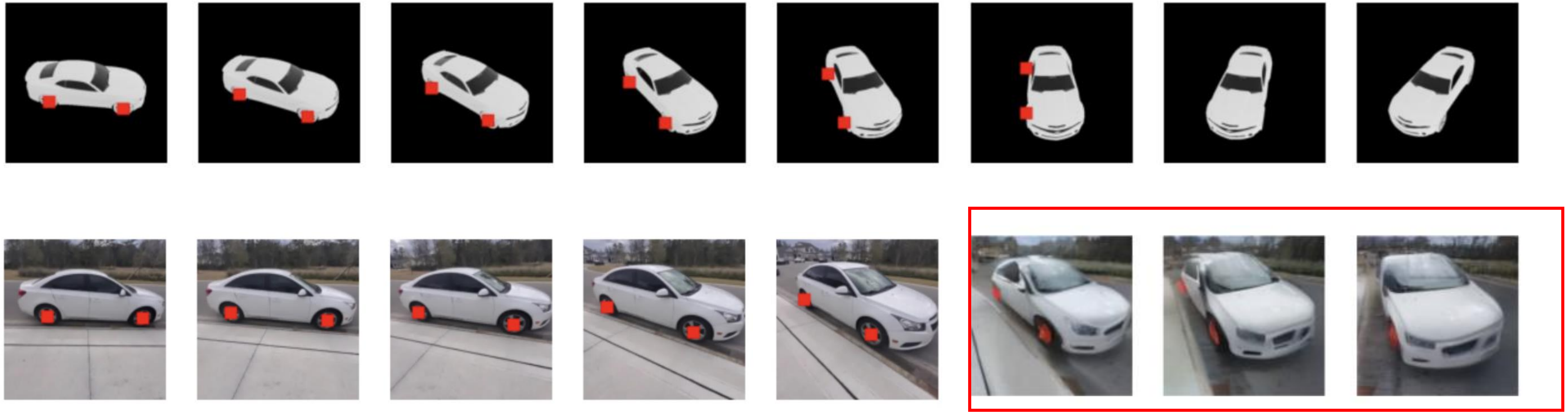
More complicated

More complicated

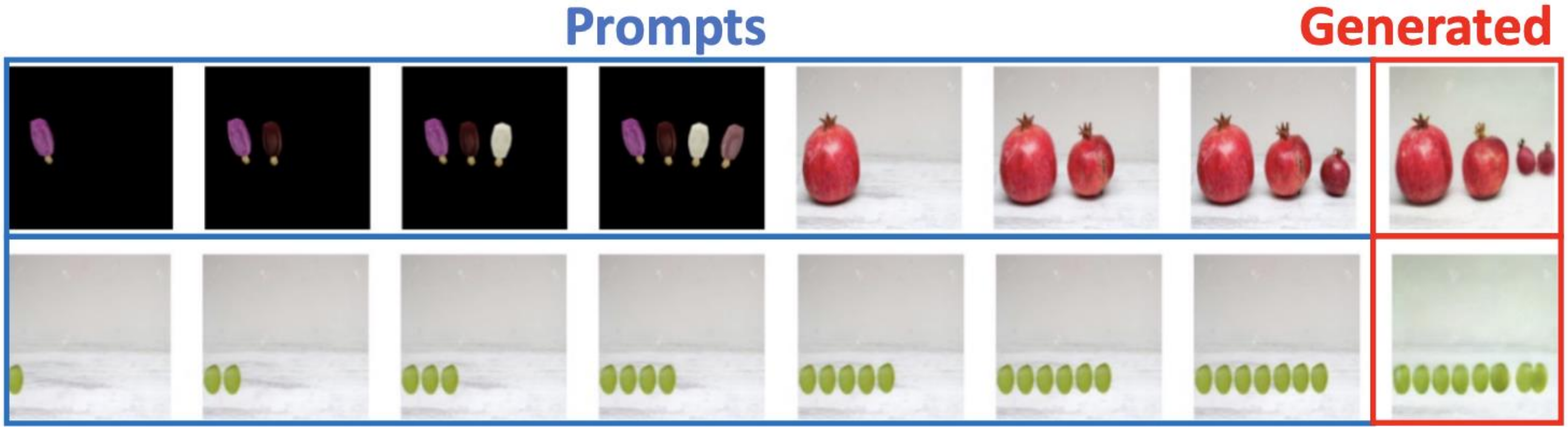
More complicated



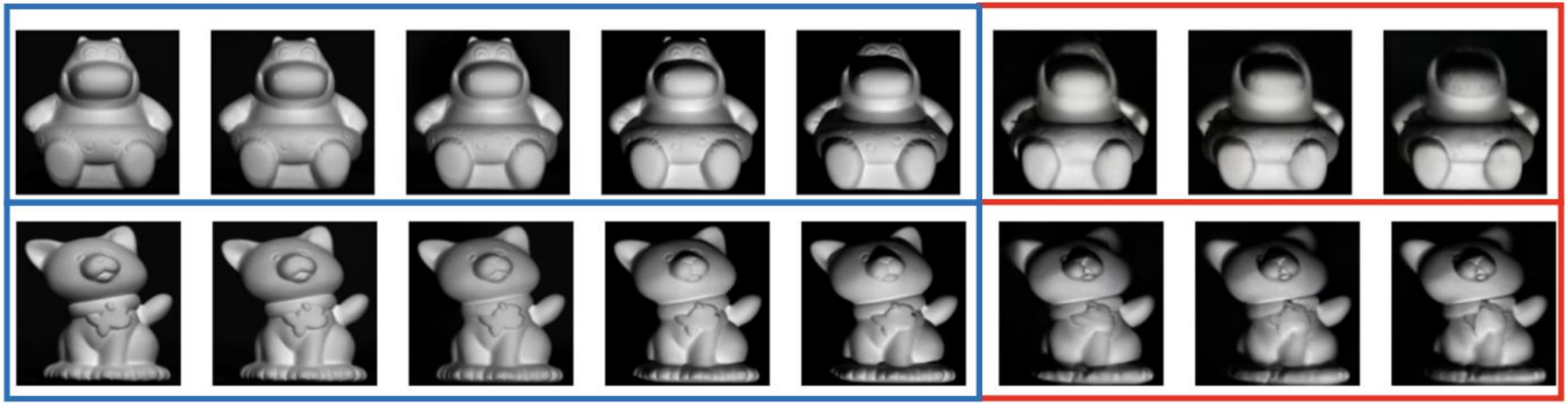
More complicated



Unseen tasks



Unseen tasks



Not easily describable

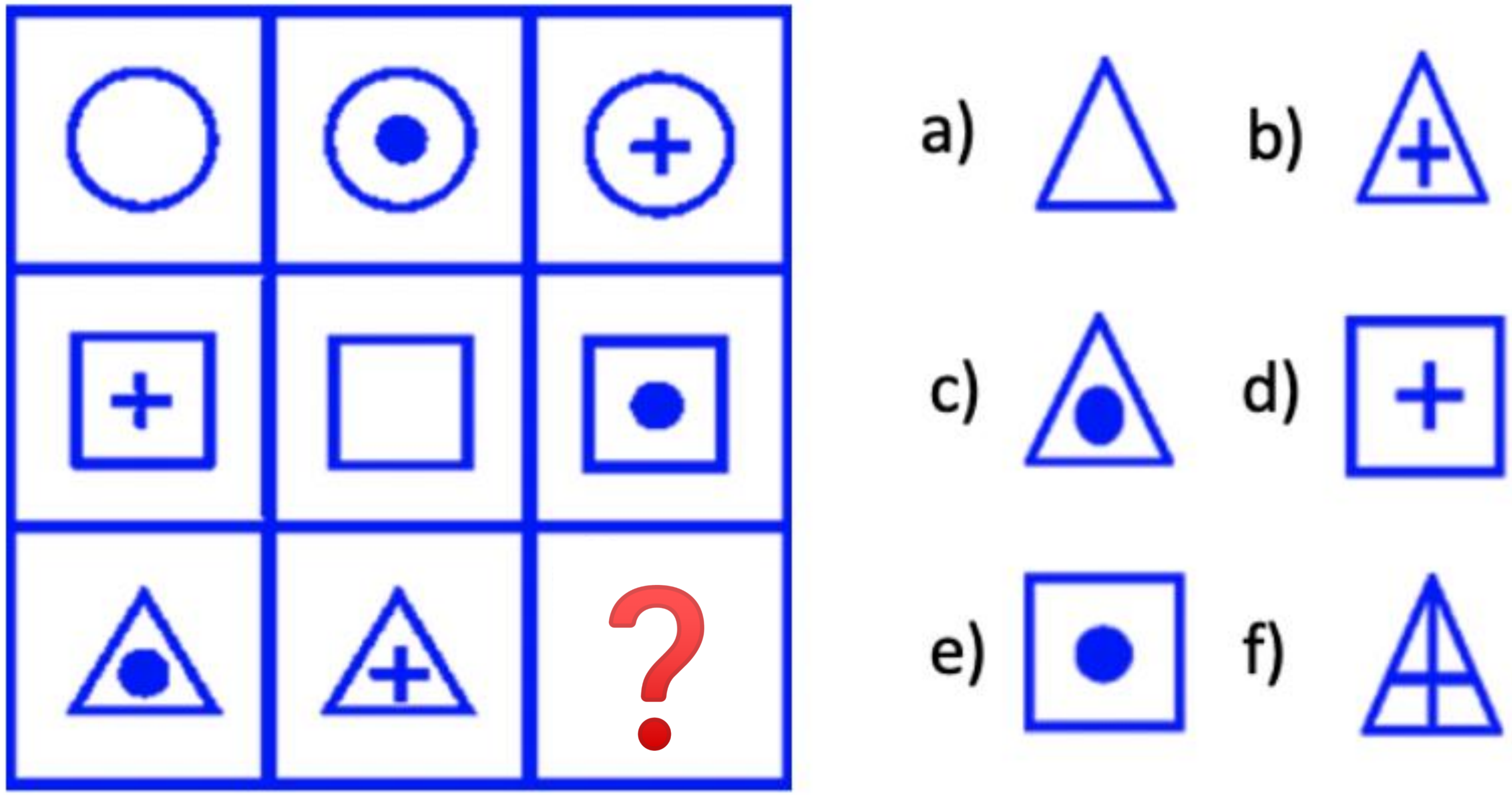
Not easily describable



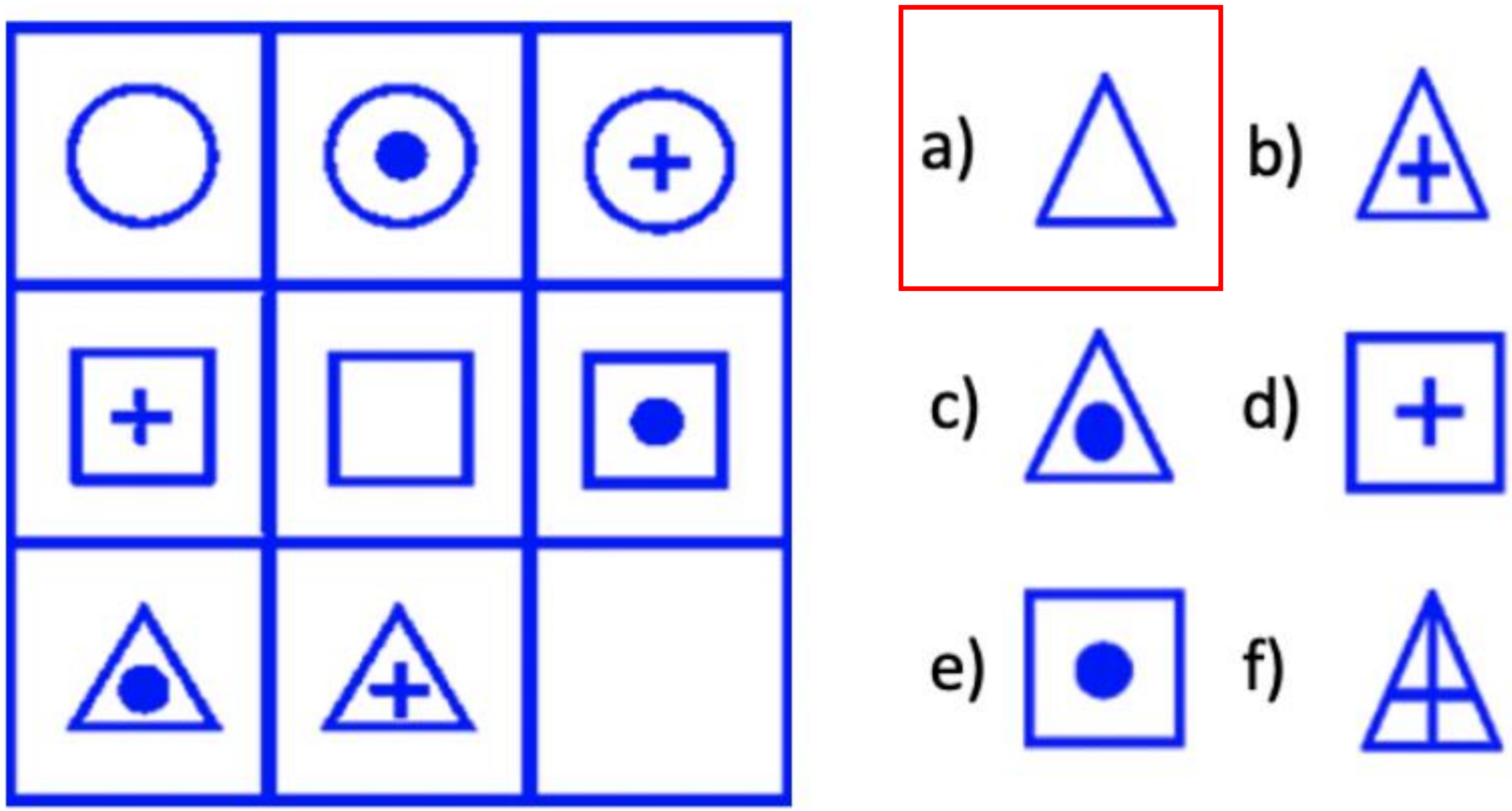
Not easily describable

Not easily describable

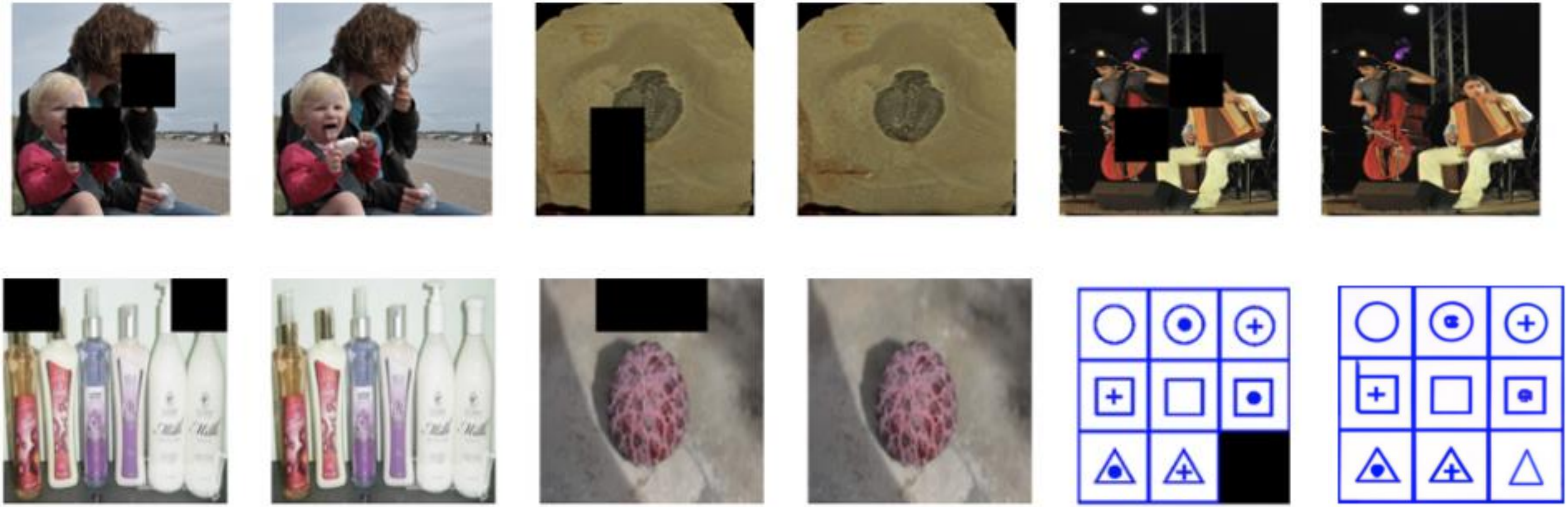
Raven's Progressive Test (Non-verbal IQ test)



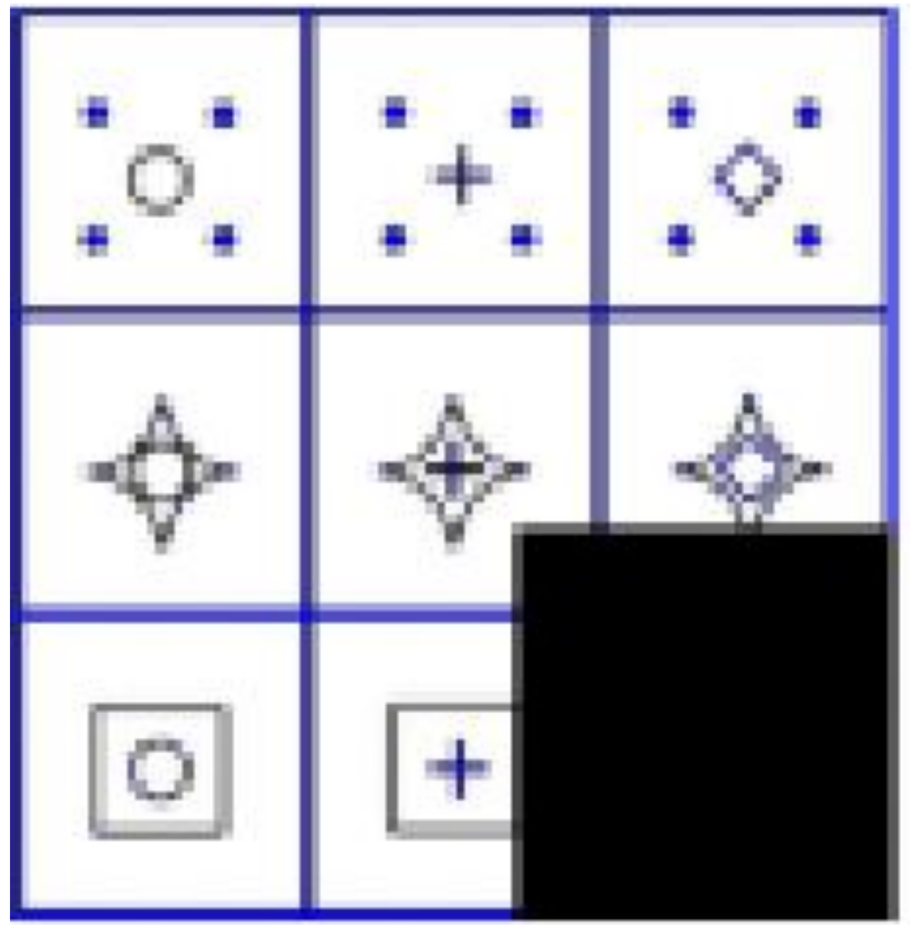
Raven’s Progressive Test (Non-verbal IQ test)



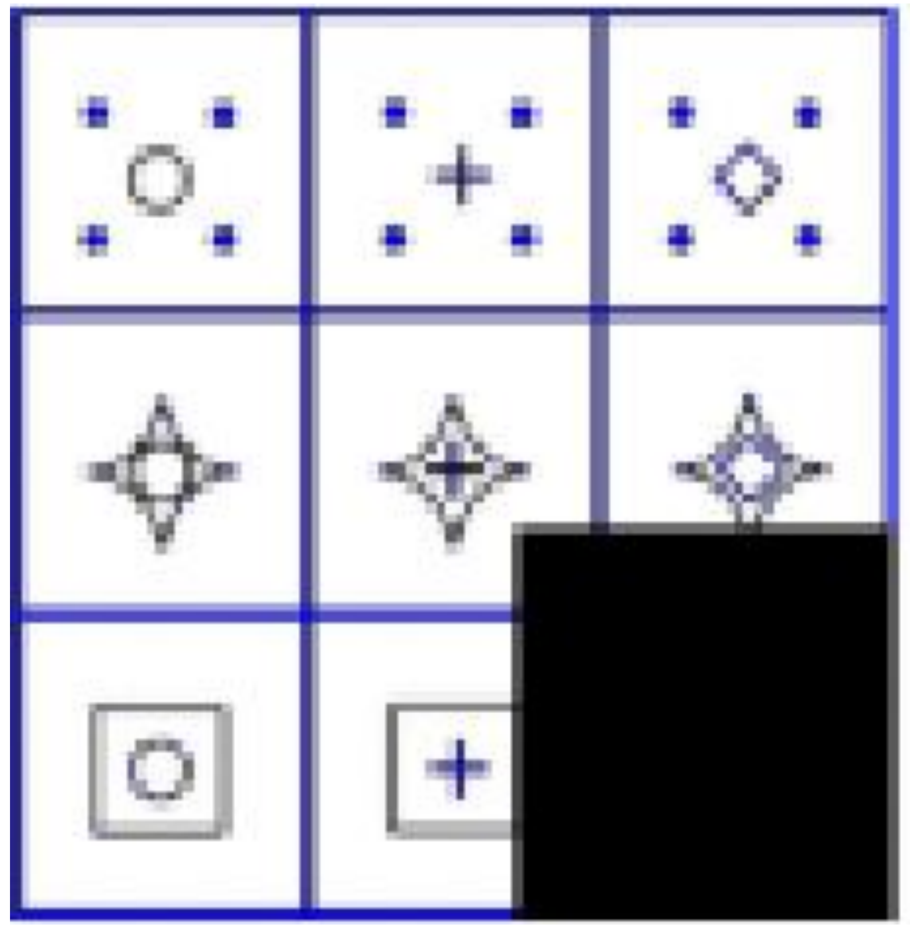
Raven’s Progressive Test (Non-verbal IQ test)



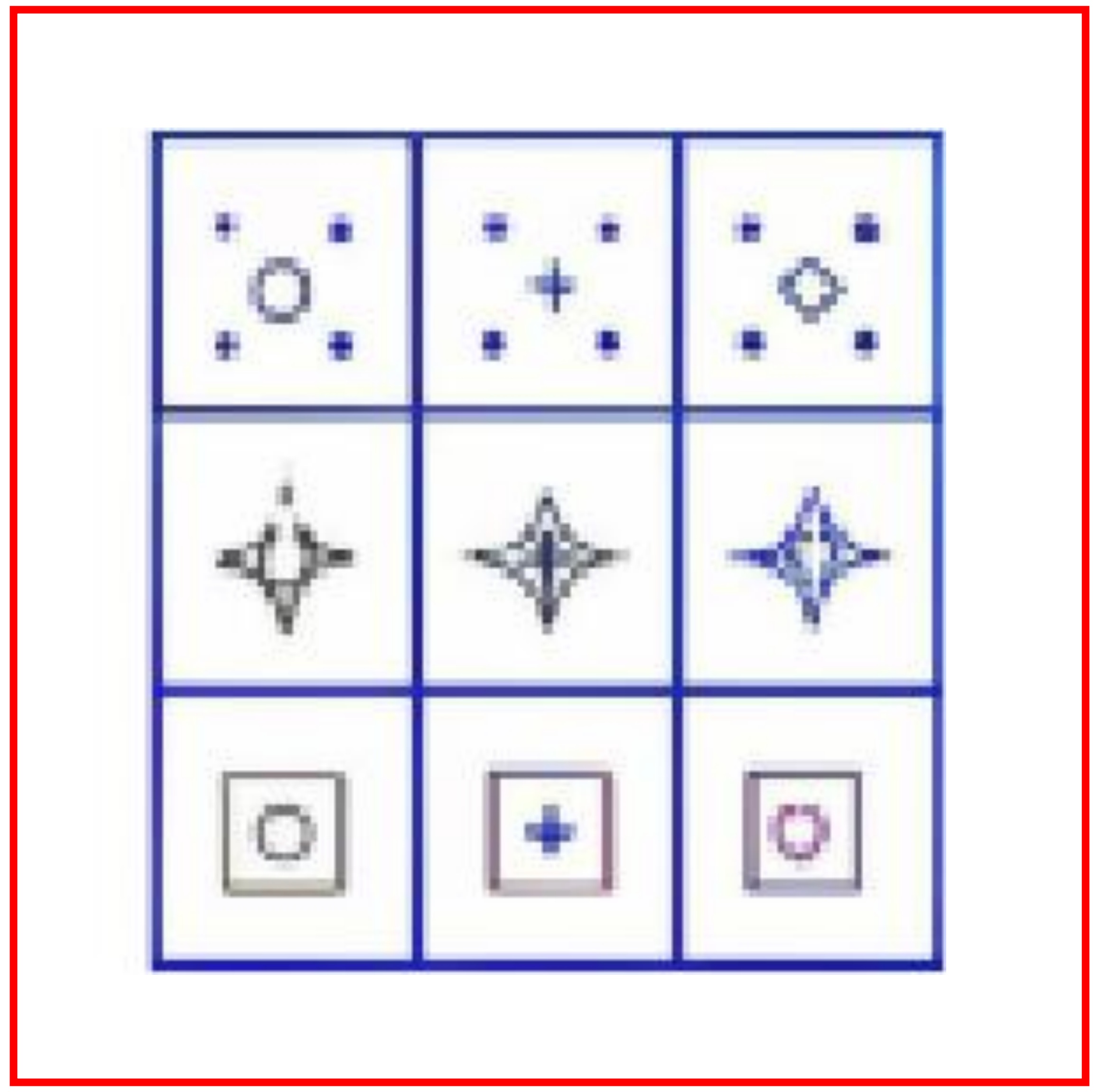
More Difficult?



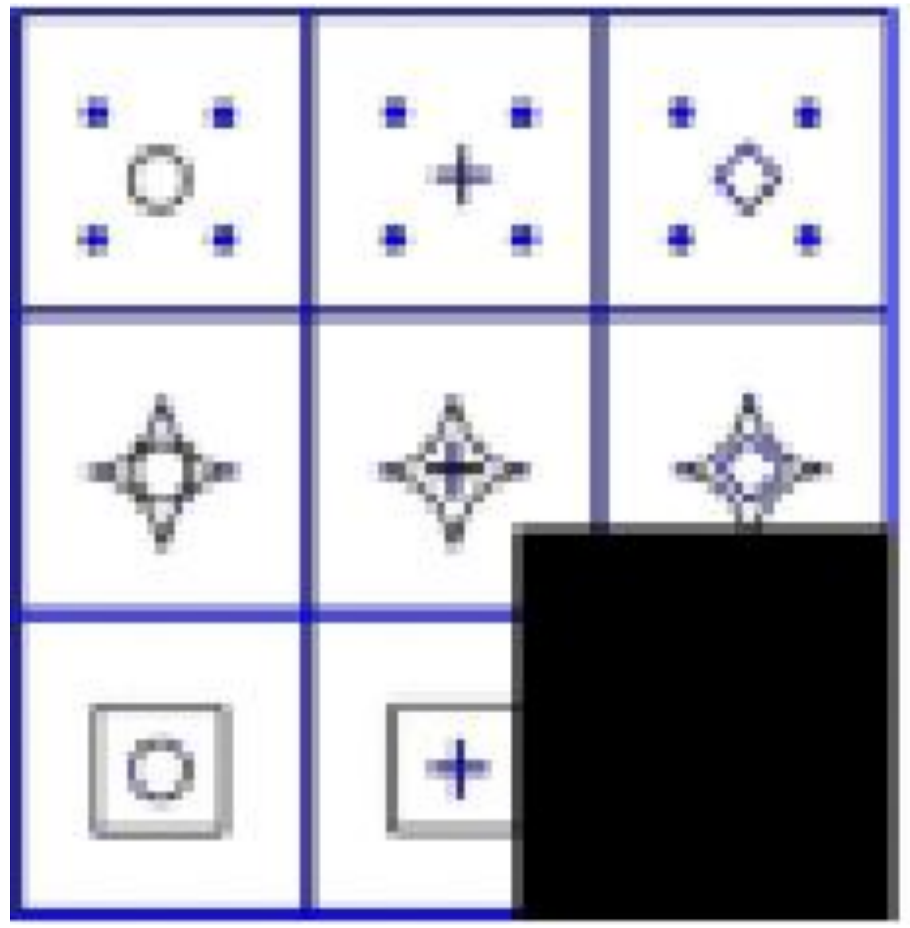
More Difficult?



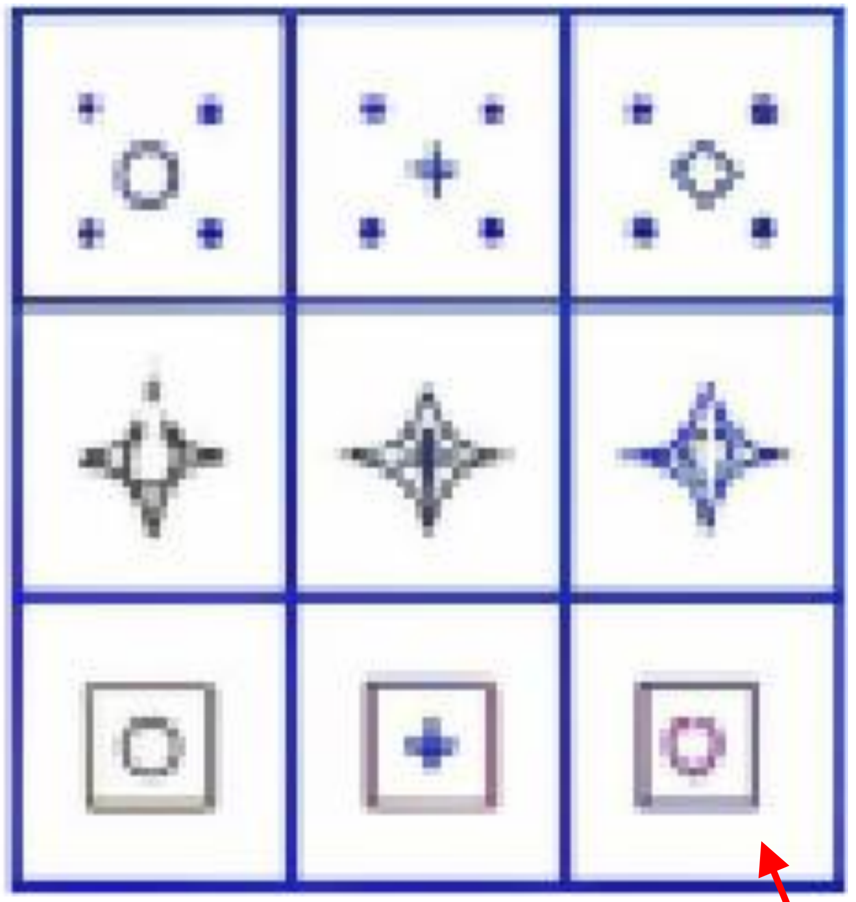
Generated



More Difficult?



Generated



Hard to tell if correct or not 🤔

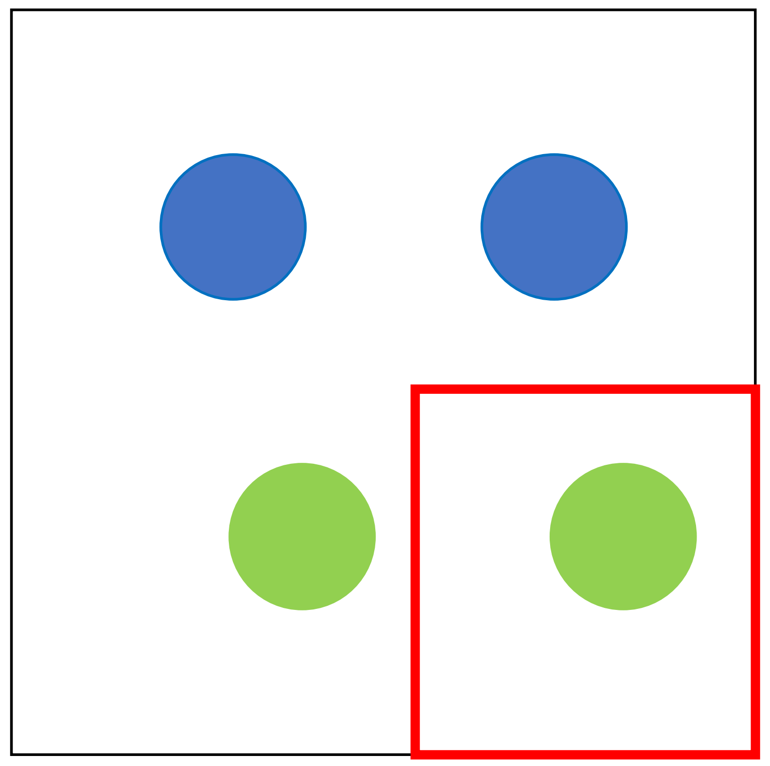
Perplexity

- Small Quantitative study (n=10):
  - perplexity analysis on classic Raven 5-way multiple-choice Matrices, choosing the answer with lowest perplexity.

	Raven's Progressive Matrices
Chance	20%
Ours	30%

Synthetic Reasoning

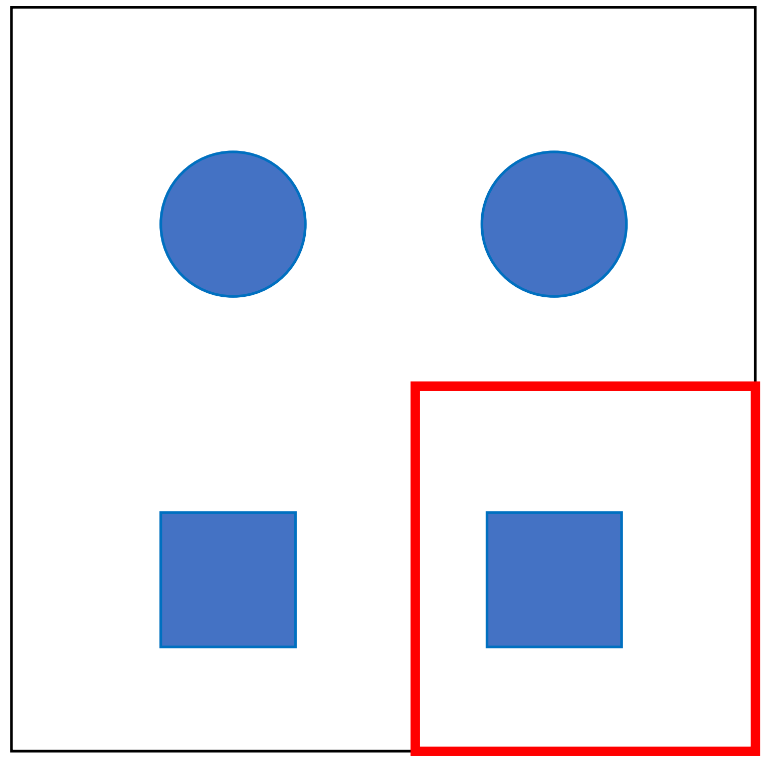
- **Color Change:** choose from 3 random generated colors.



	color
Chance	33%
Ours	**42%**

Synthetic Reasoning

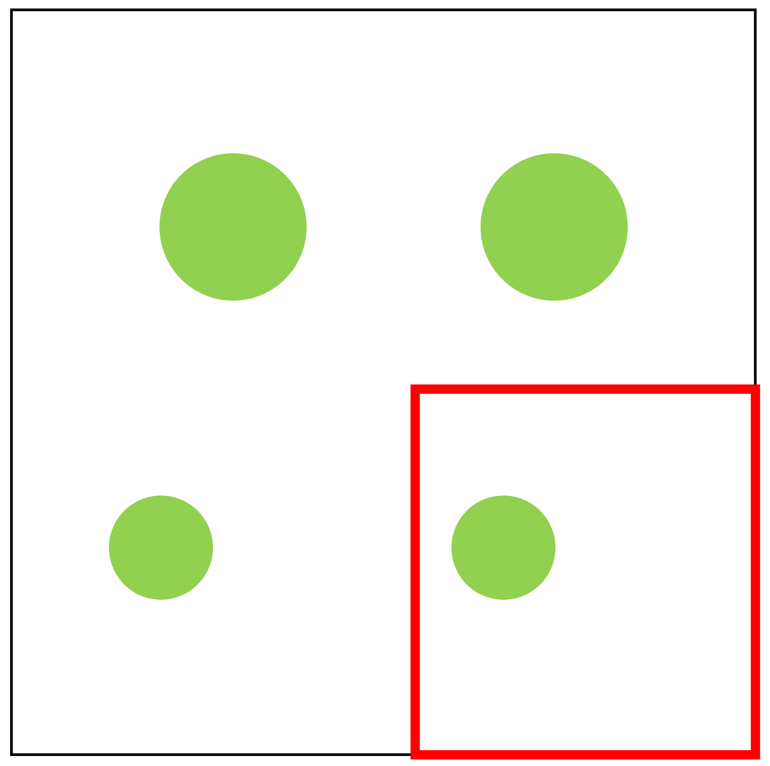
- **Shape Change:** choose from 3 random generated shapes.



	shape
Chance	33%
Ours	**45%**

Synthetic Reasoning

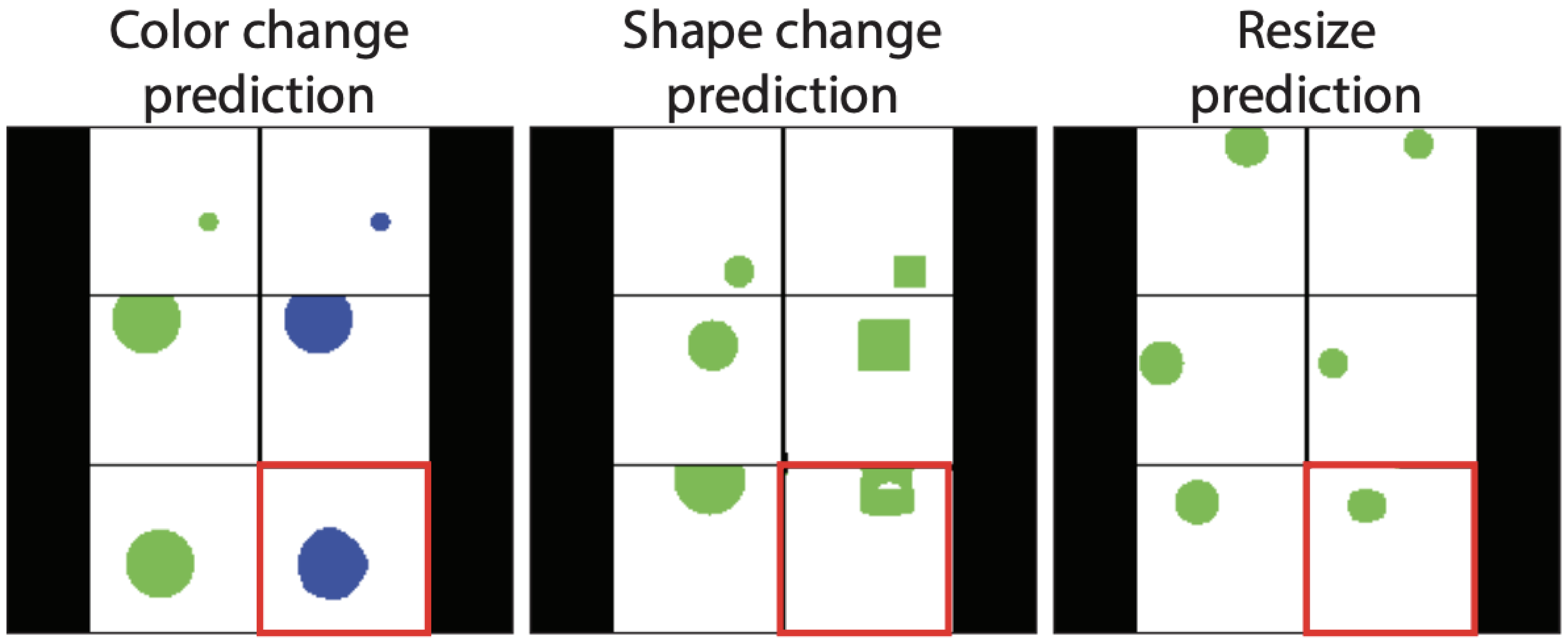
- **Size Change:** choose from 2 random generated sizes. (resolution)



	size
Chance	50%
Ours	**94%**

Synthetic Reasoning

- In total 900 experiments



	color	shape	size
Chance	33%	33%	50%
Ours	**42%**	**45%**	**94%**

Visual Prompting via Image Inpainting, Bar et al.

Summary

- The dream of self-supervising ourselves with **natural world data** (rather than text) is alive and well!
- But we are still at the beginning of the journey
- Plenty left to do!