

Sequence Models for words and pixels



© A.A. Efros

Many slides from Steve Seitz's wonderful [5 min Lectures](#)

CS194: Intro to Computer Vision & Comp. Photography
Alexei Efros, UC Berkeley, Fall 2022

Michel Gondry train video

<http://www.youtube.com/watch?v=0S43lwBF0uM>

“Amateur” by Lasse Gjertsen

<http://www.youtube.com/watch?v=JzqumbhfxRo>

similar idea:

<http://www.youtube.com/watch?v=MsBMG-p1HDM&feature=share&list=PLFFD733D0FF425290>

Weather Forecasting for Dummies™

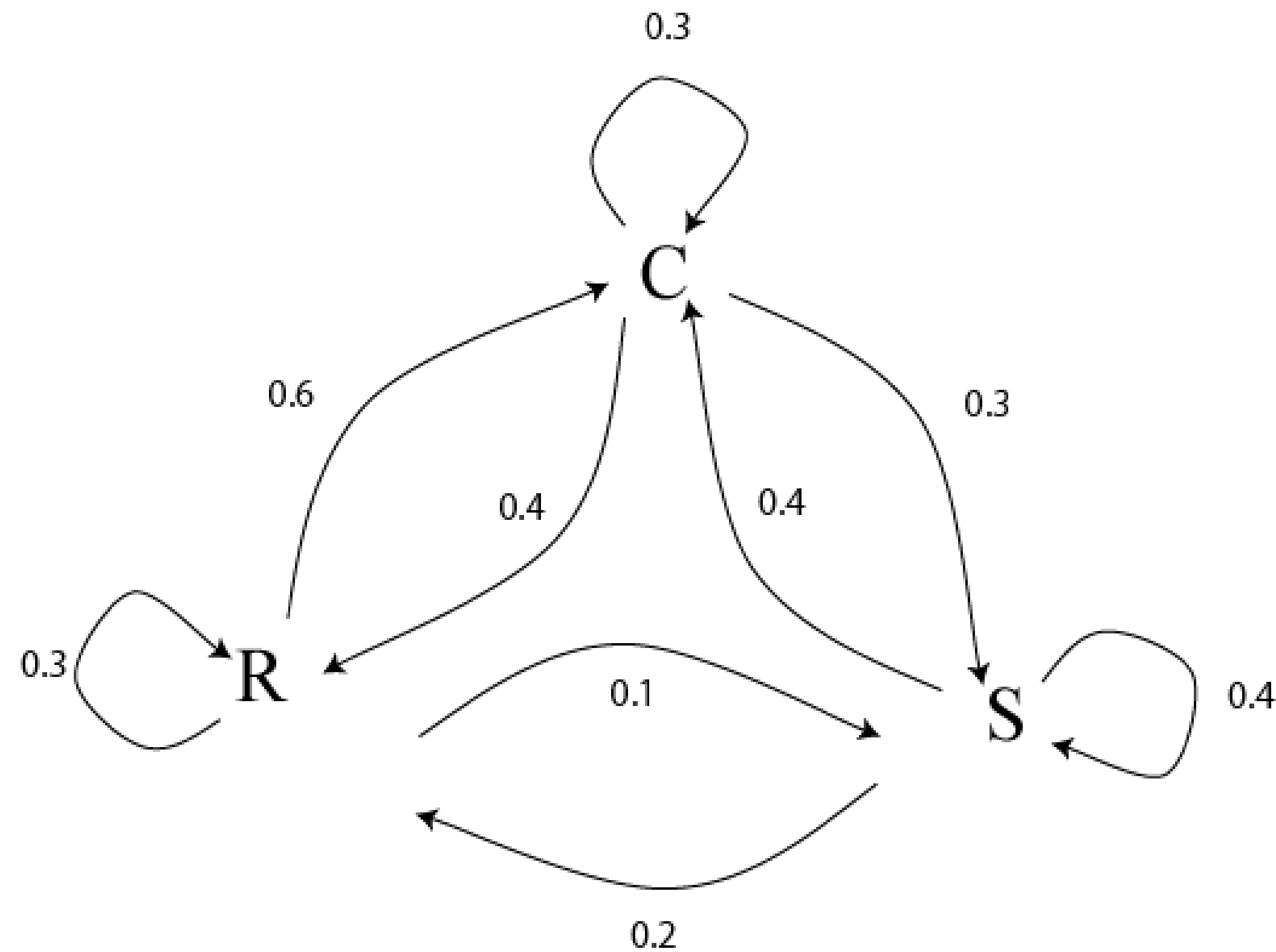
Let's predict weather:

- 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 yestarday'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*

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

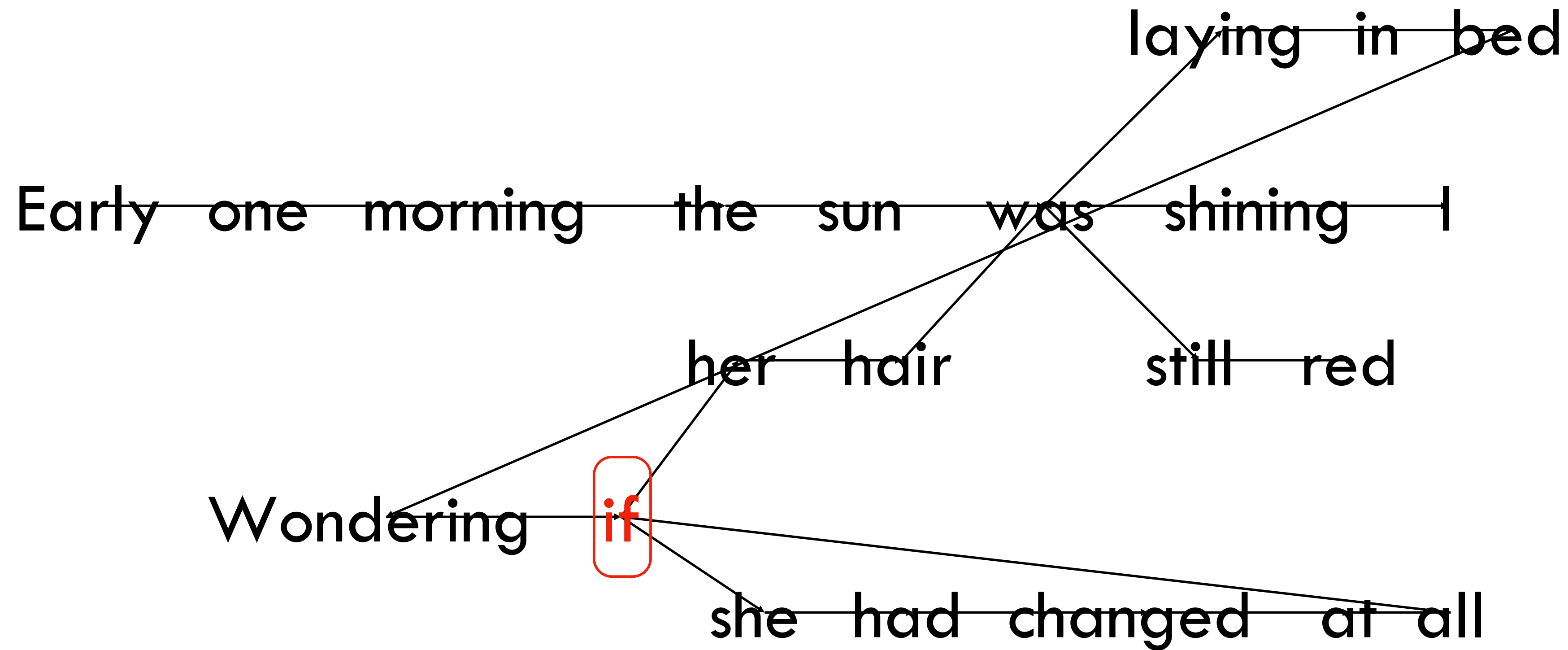
Early one morning the sun **was** shining I **was** laying in bed

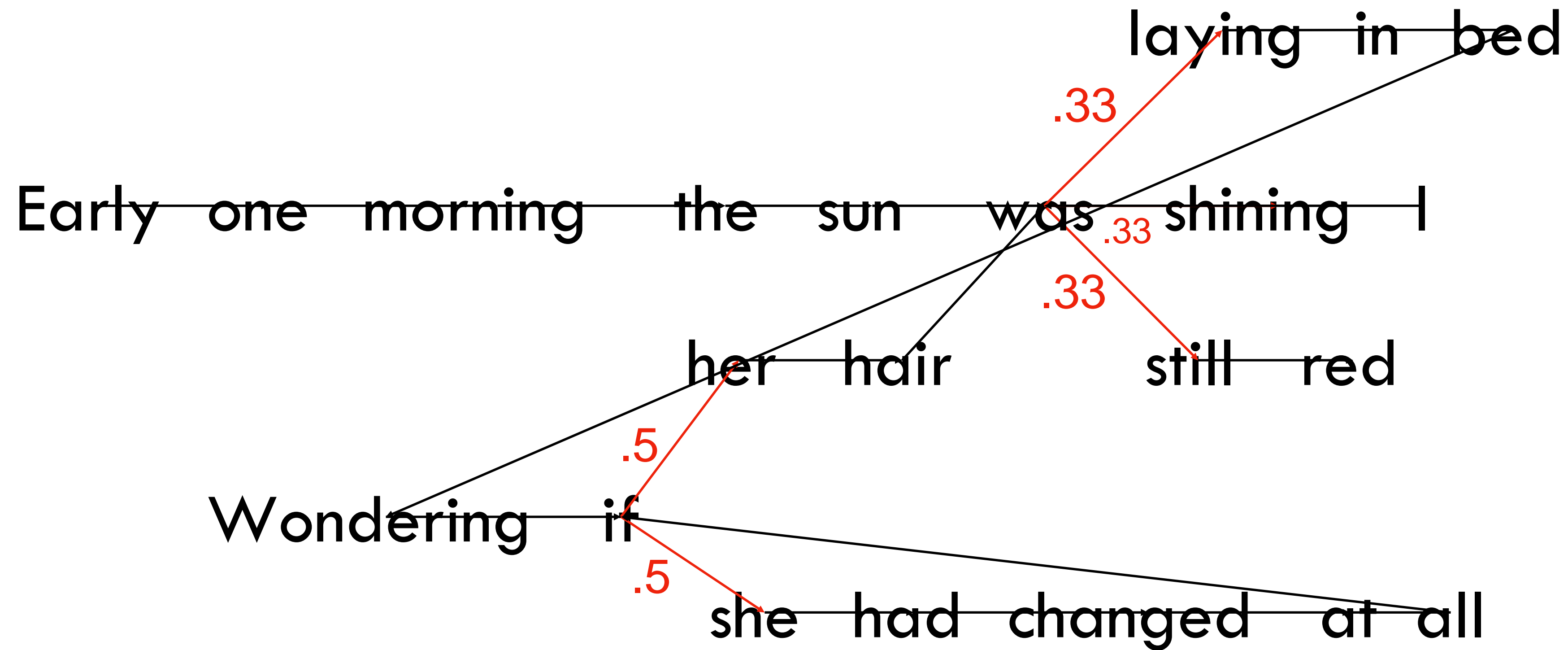
Wondering if she had changed at all if her hair **was** still red

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

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

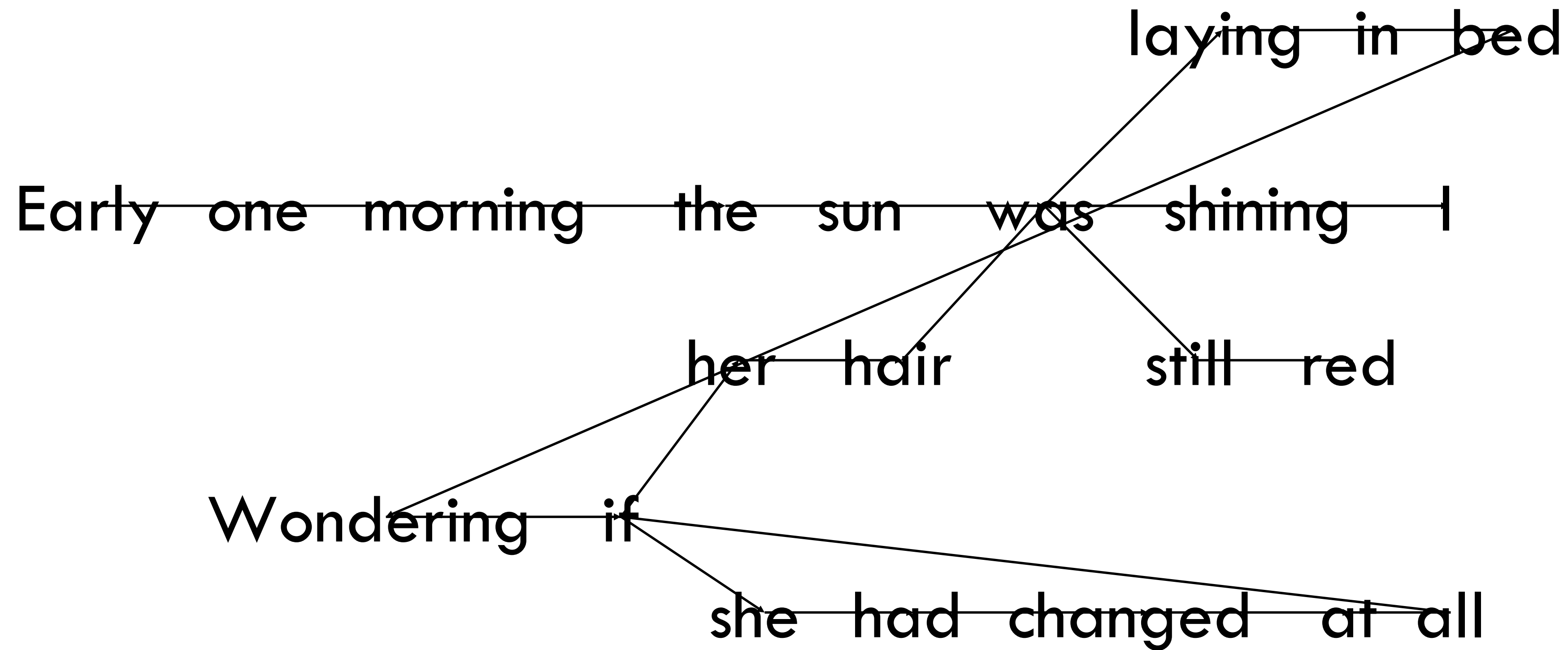
```
graph TD; W1[was] --- L1[laying]; W1 --- H1[her]; S1[still] --- I1[if];
```



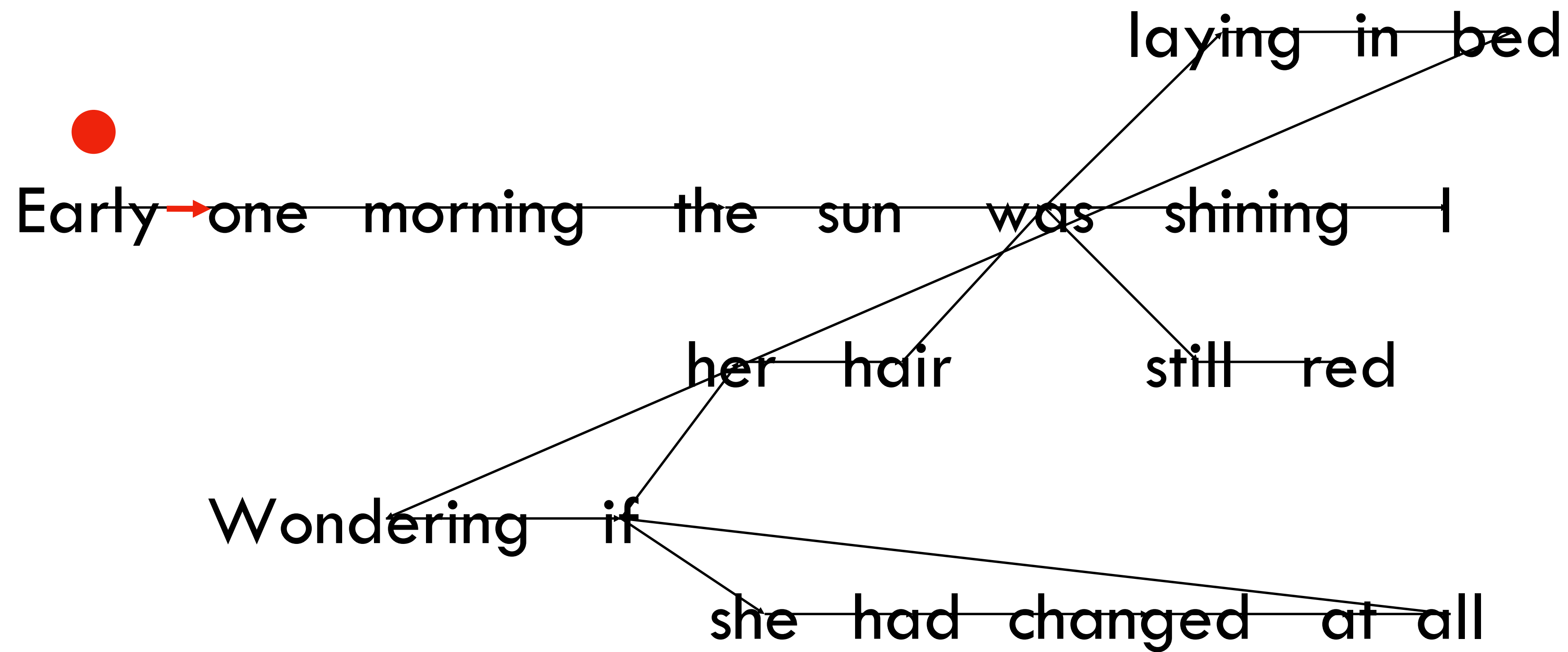


Language Model

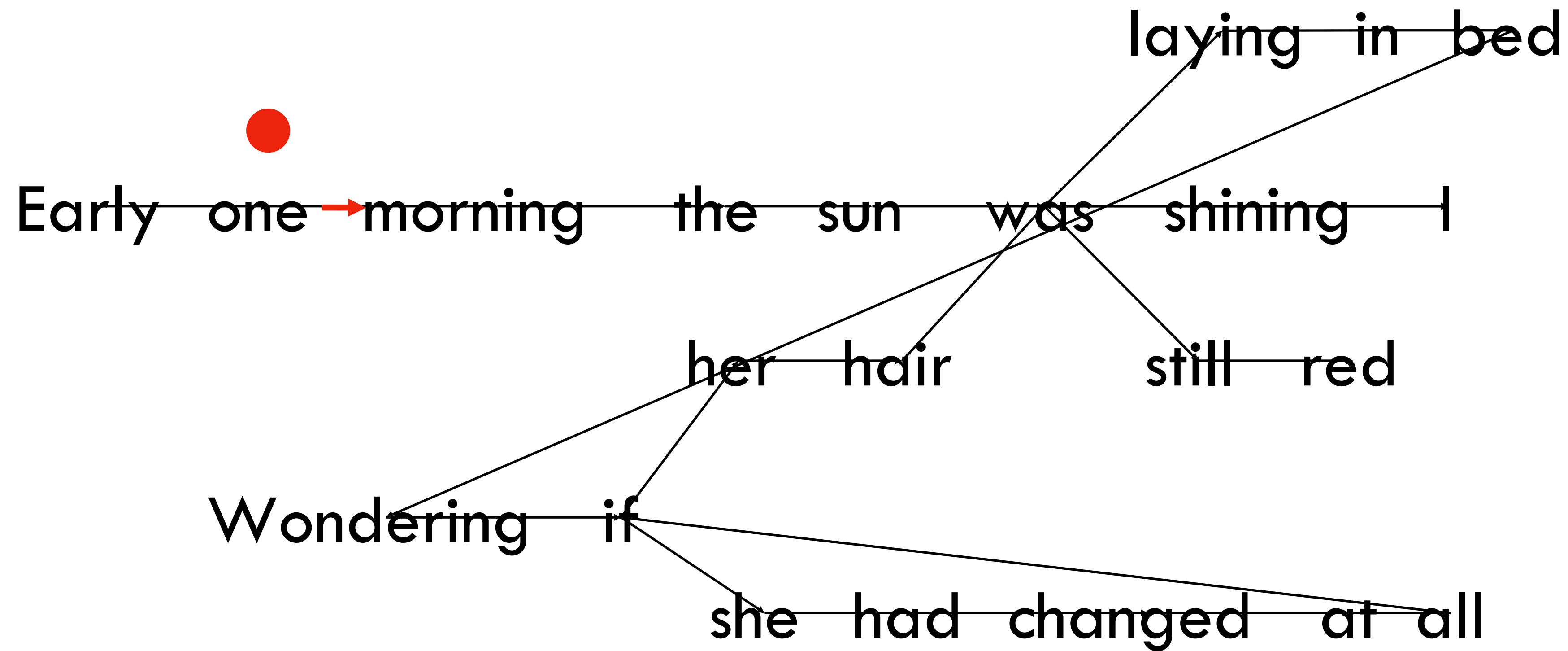
Early one morning the sun was shining I
laying in bed
her hair still red
Wondering if she had changed at all



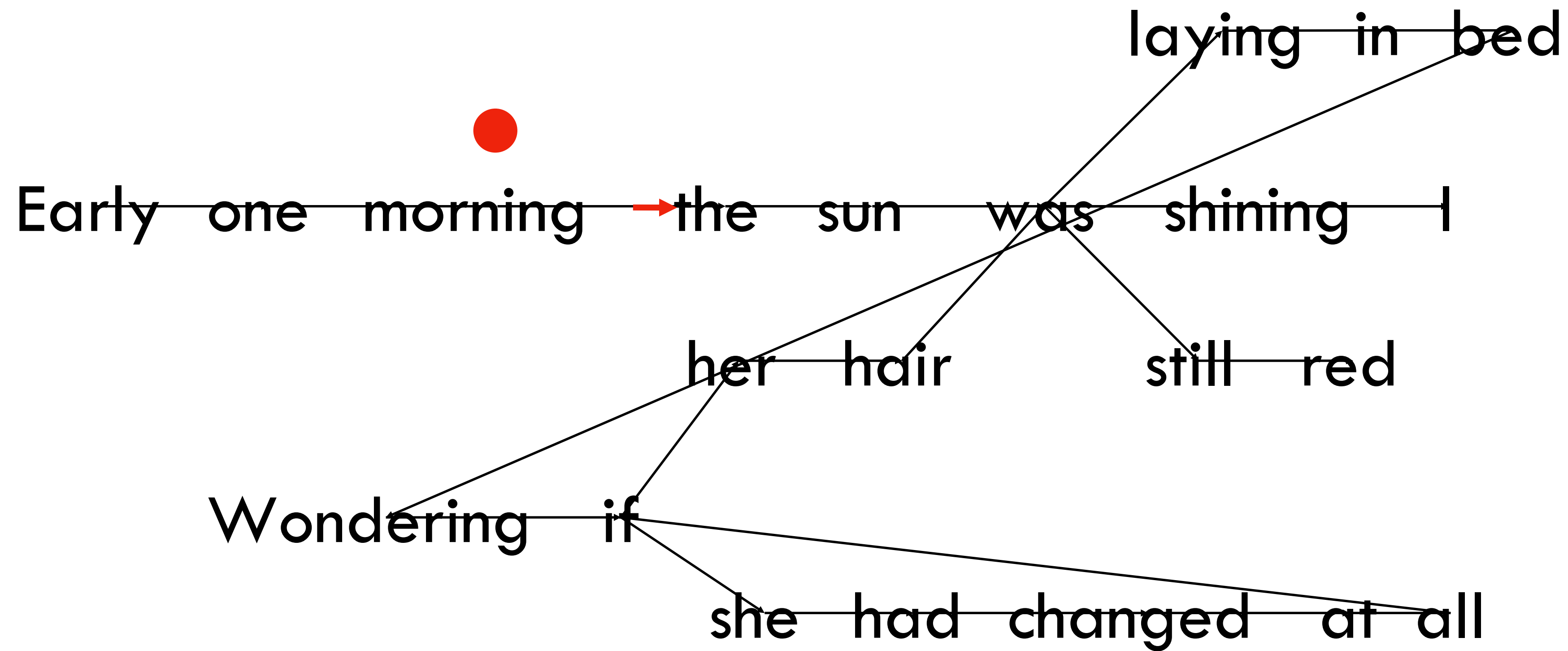
Early



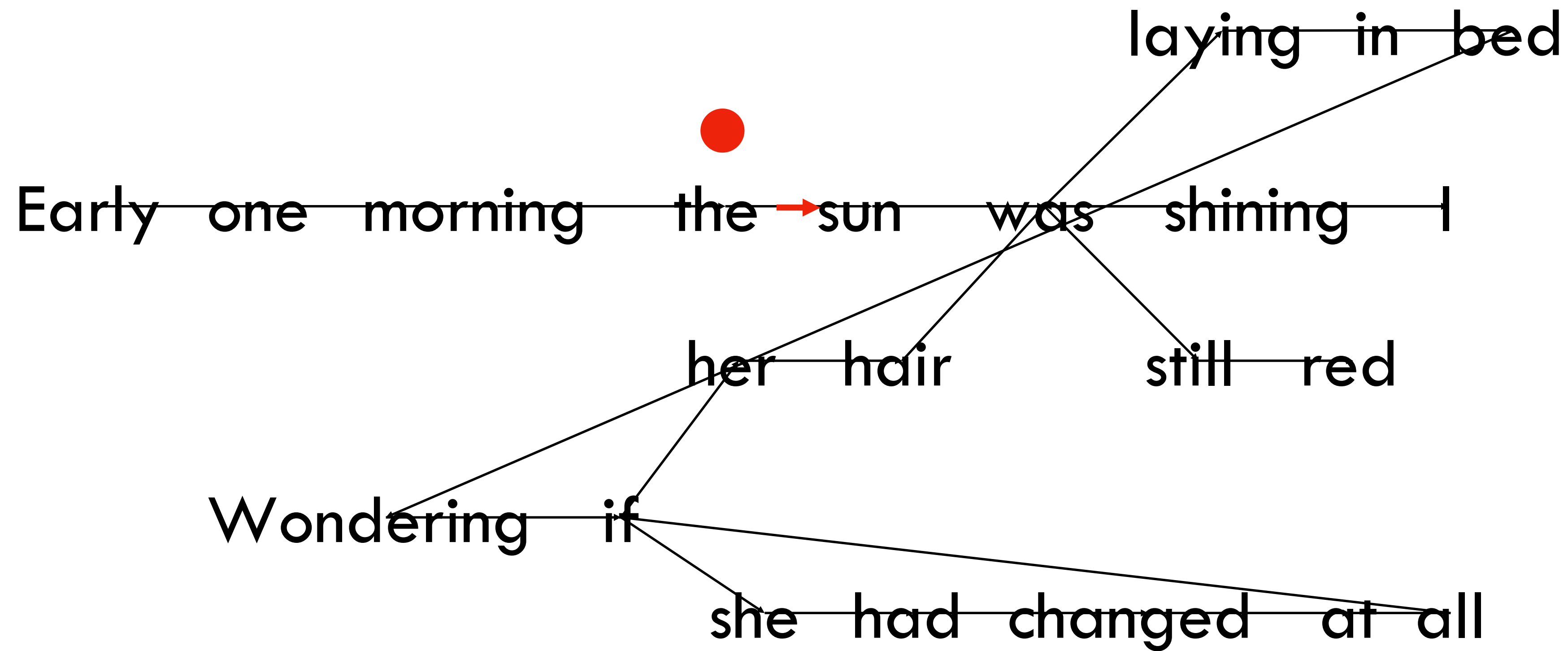
Early one



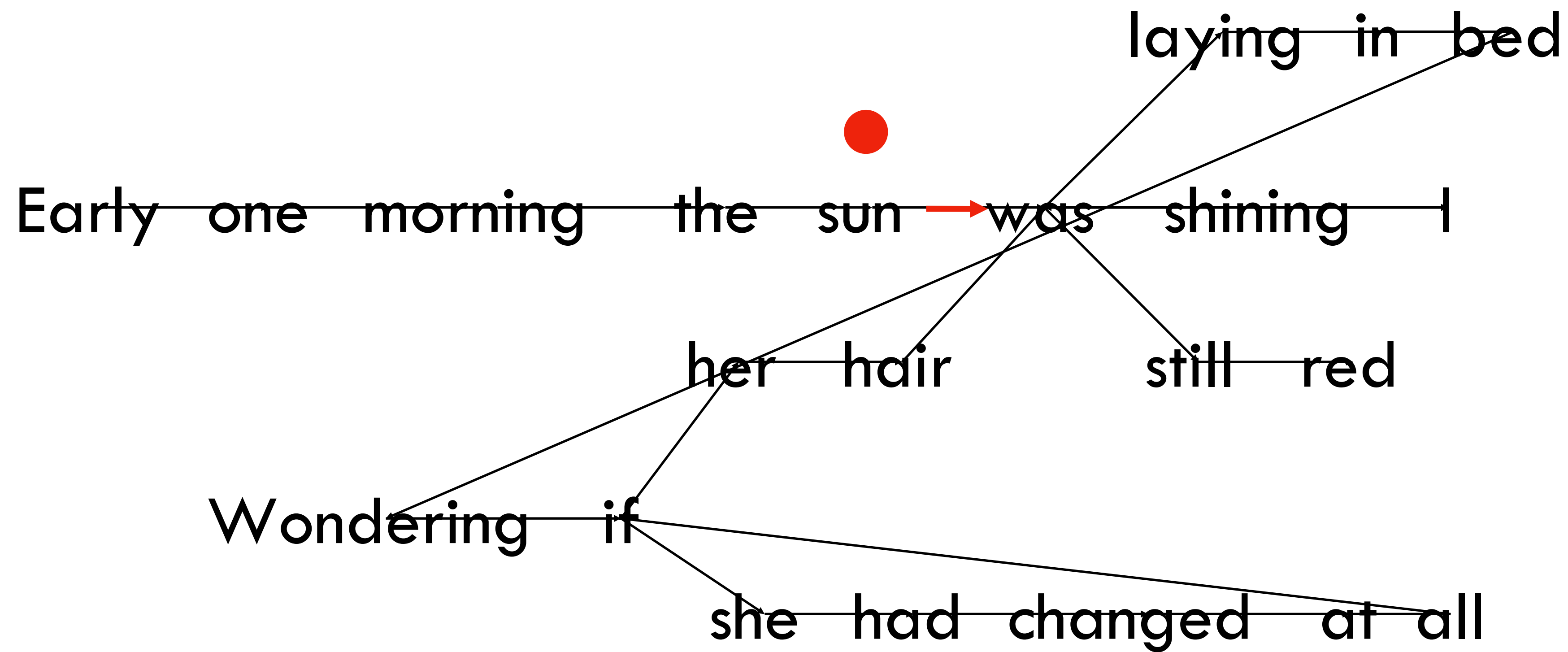
Early one morning



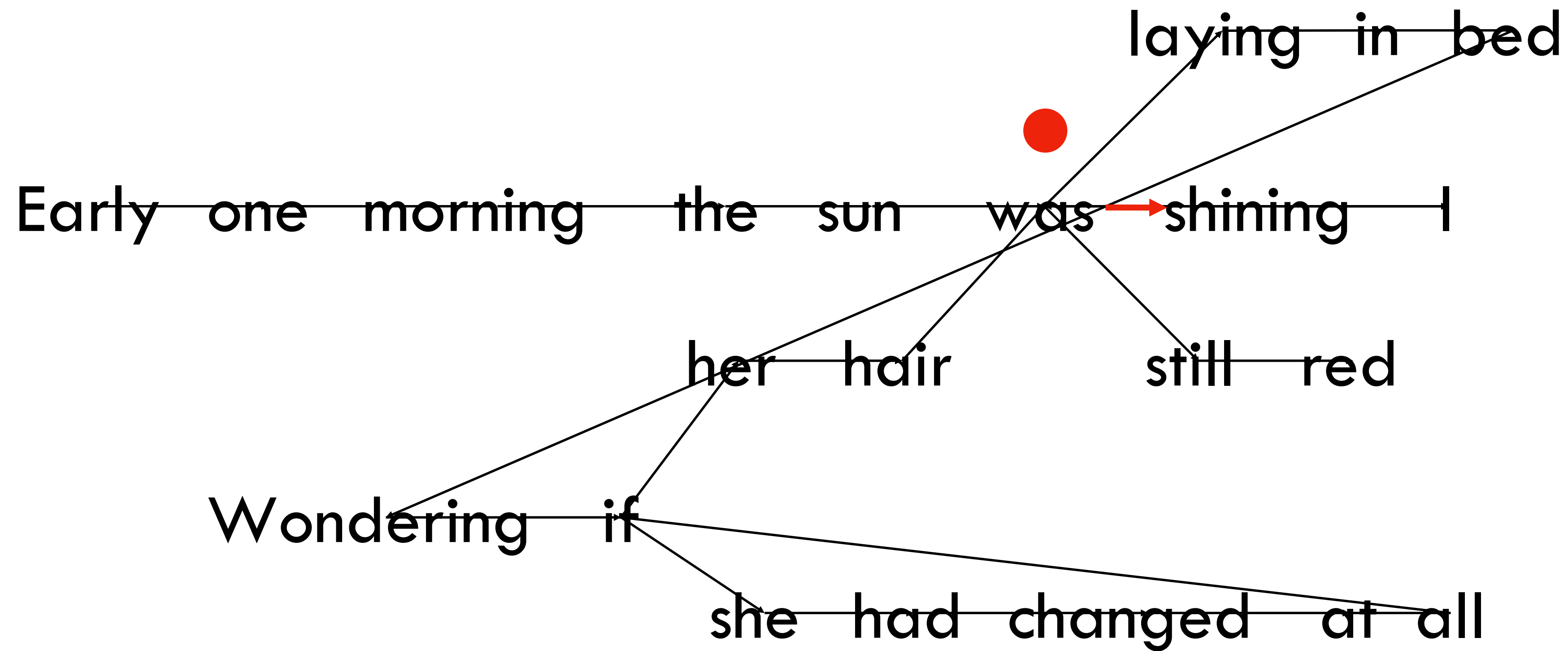
Early one morning the



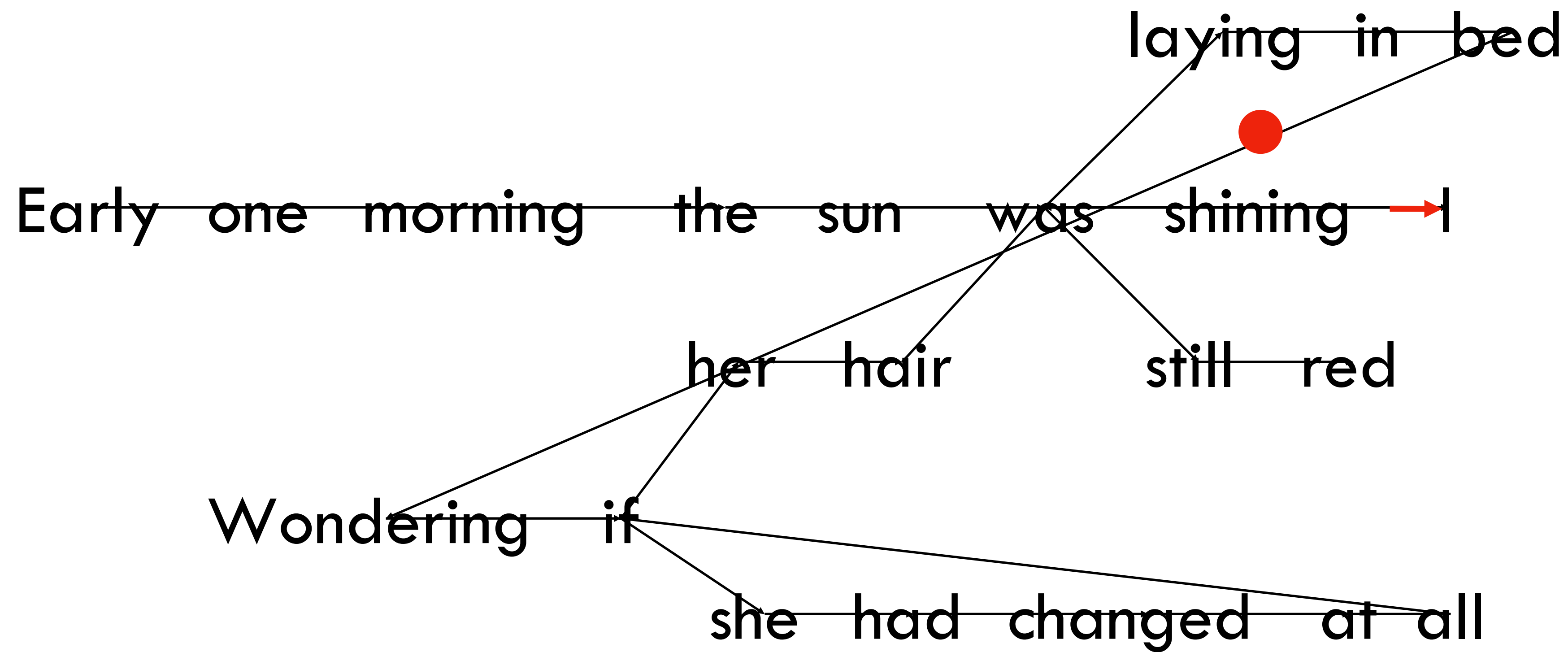
Early one morning the sun



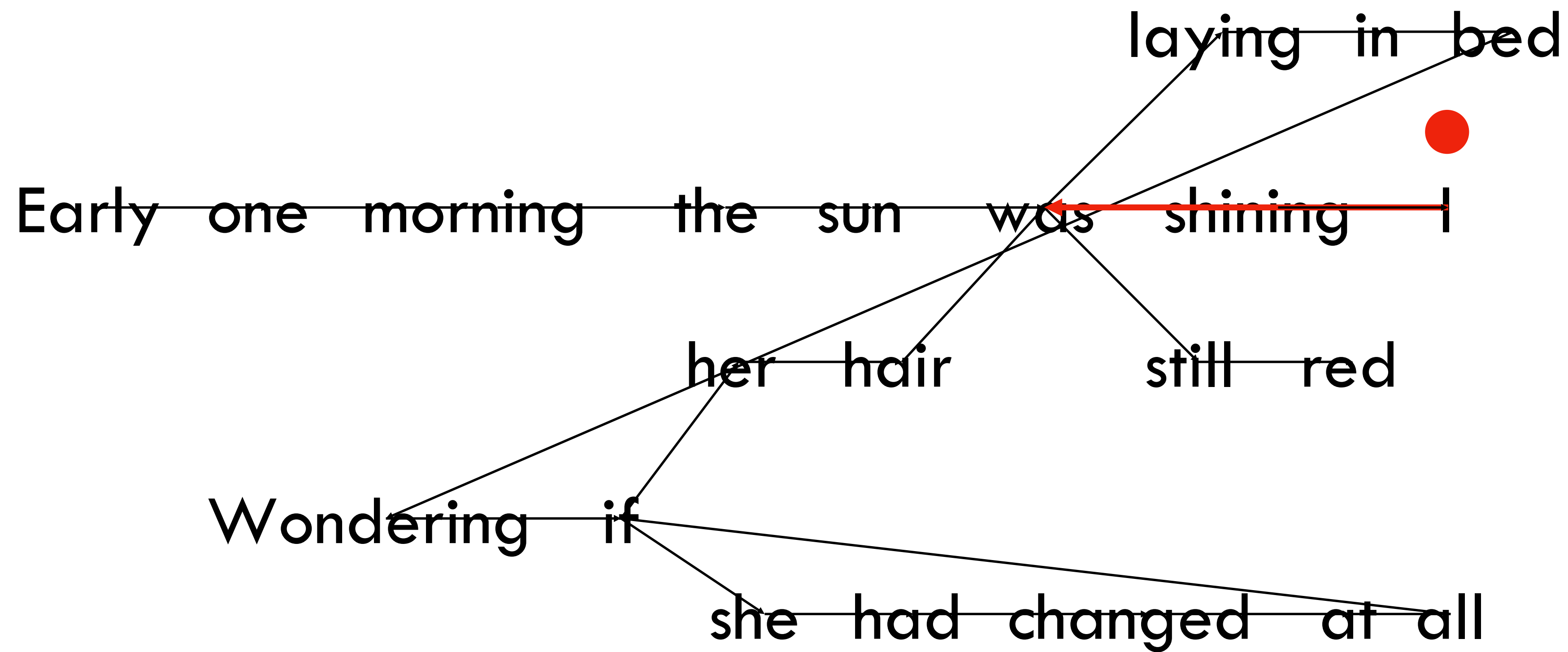
Early one morning the sun was



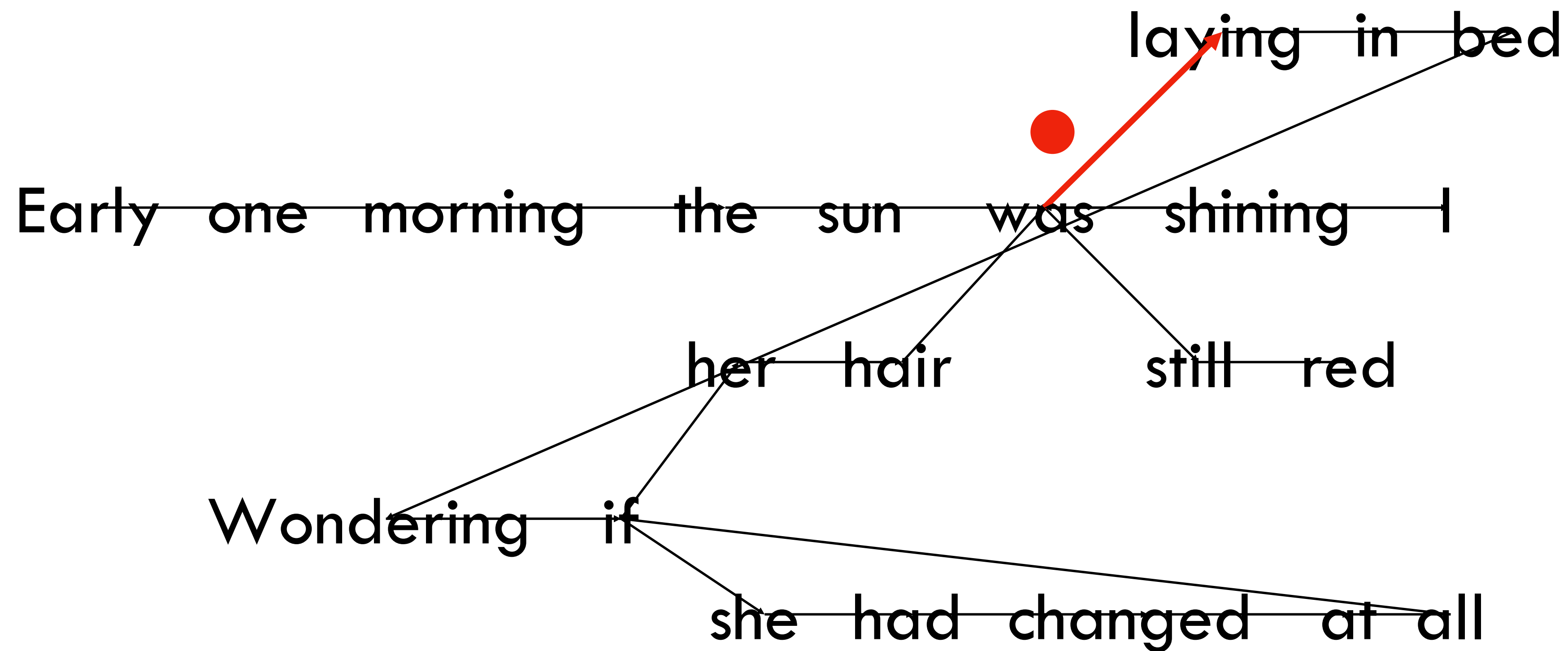
Early one morning the sun was shining



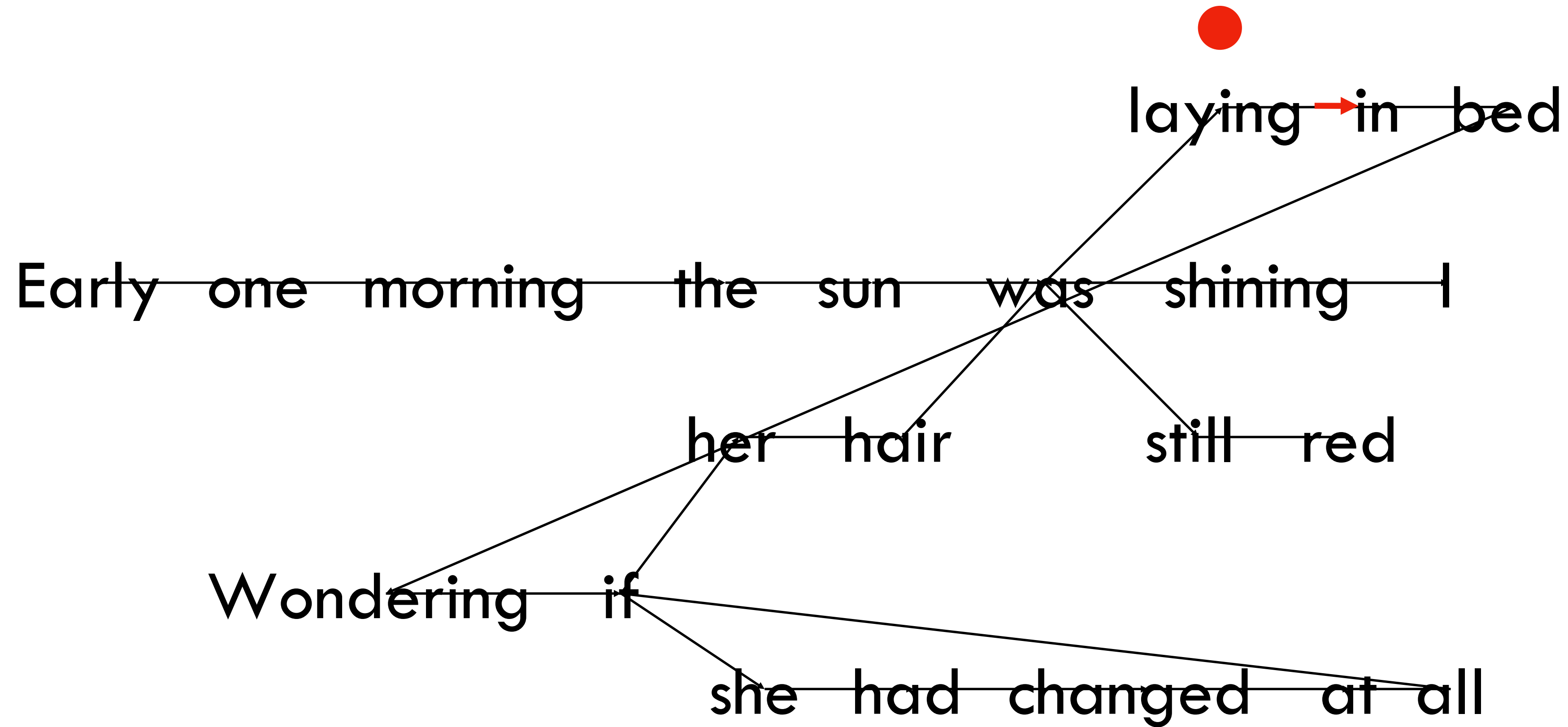
Early one morning the sun was shining I



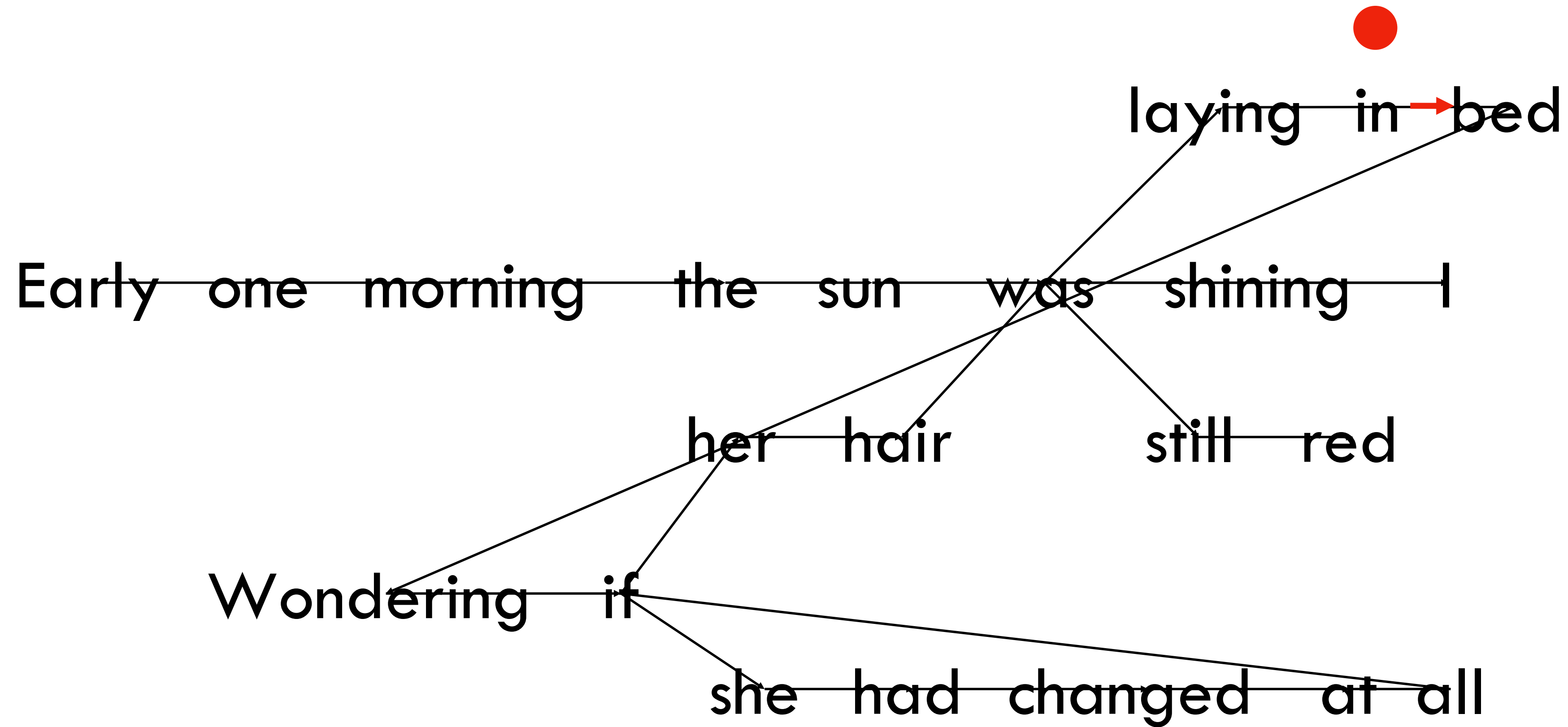
Early one morning the sun was shining I was



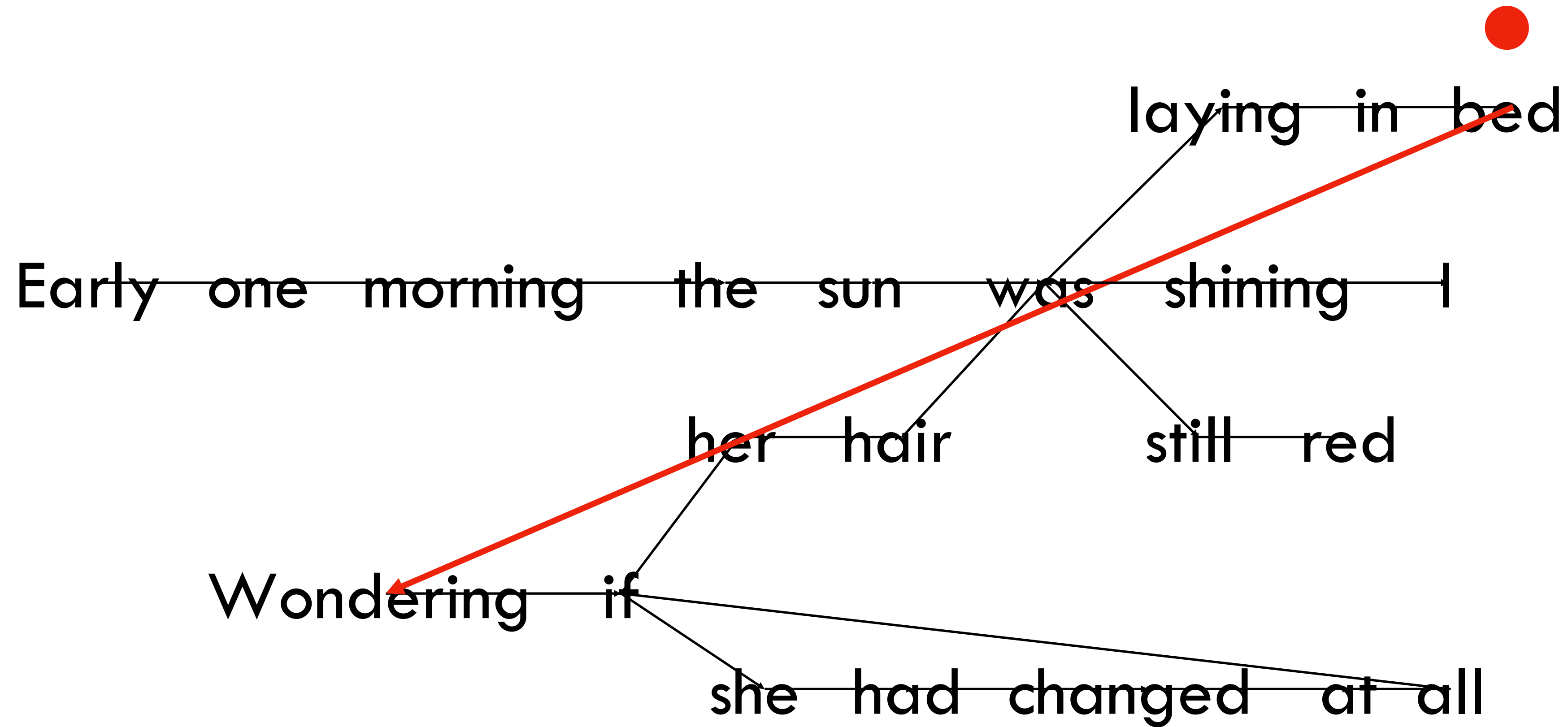
Early one morning the sun was shining I was laying



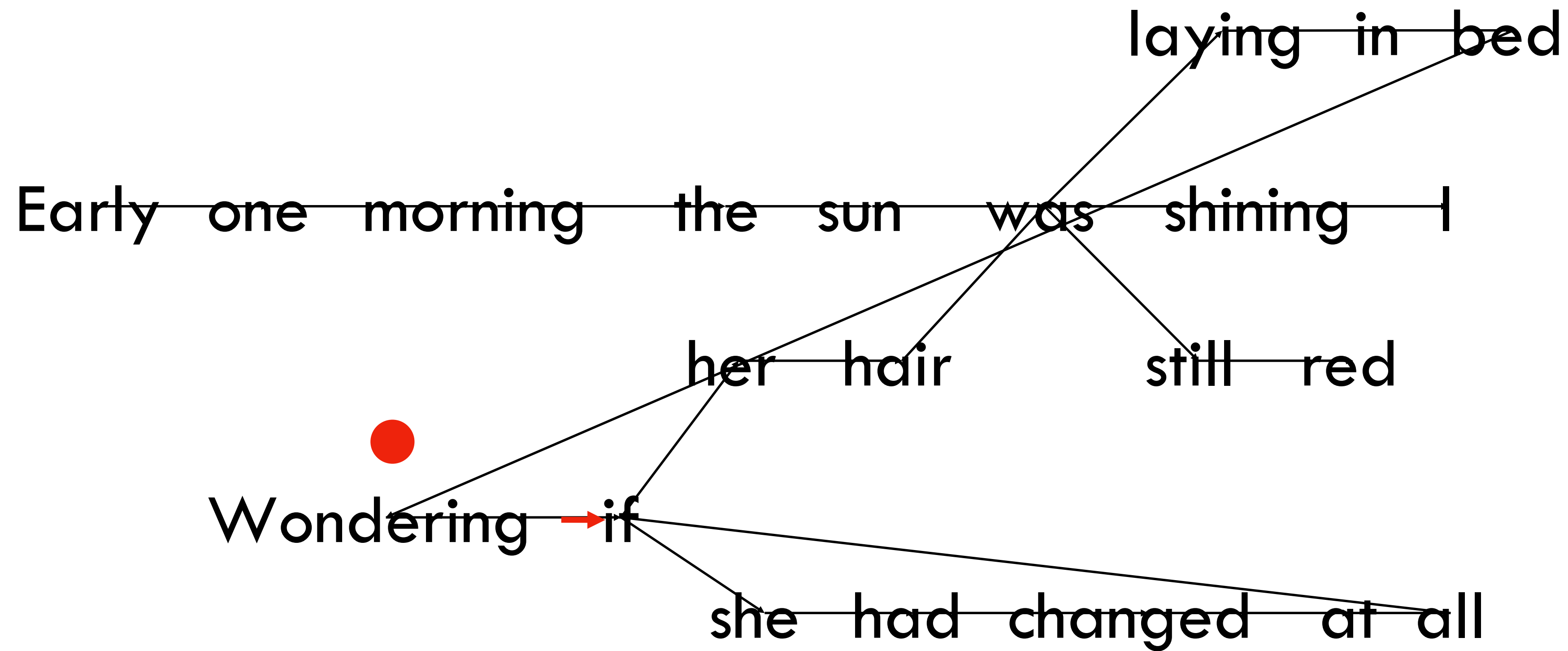
Early one morning the sun was shining I was laying in



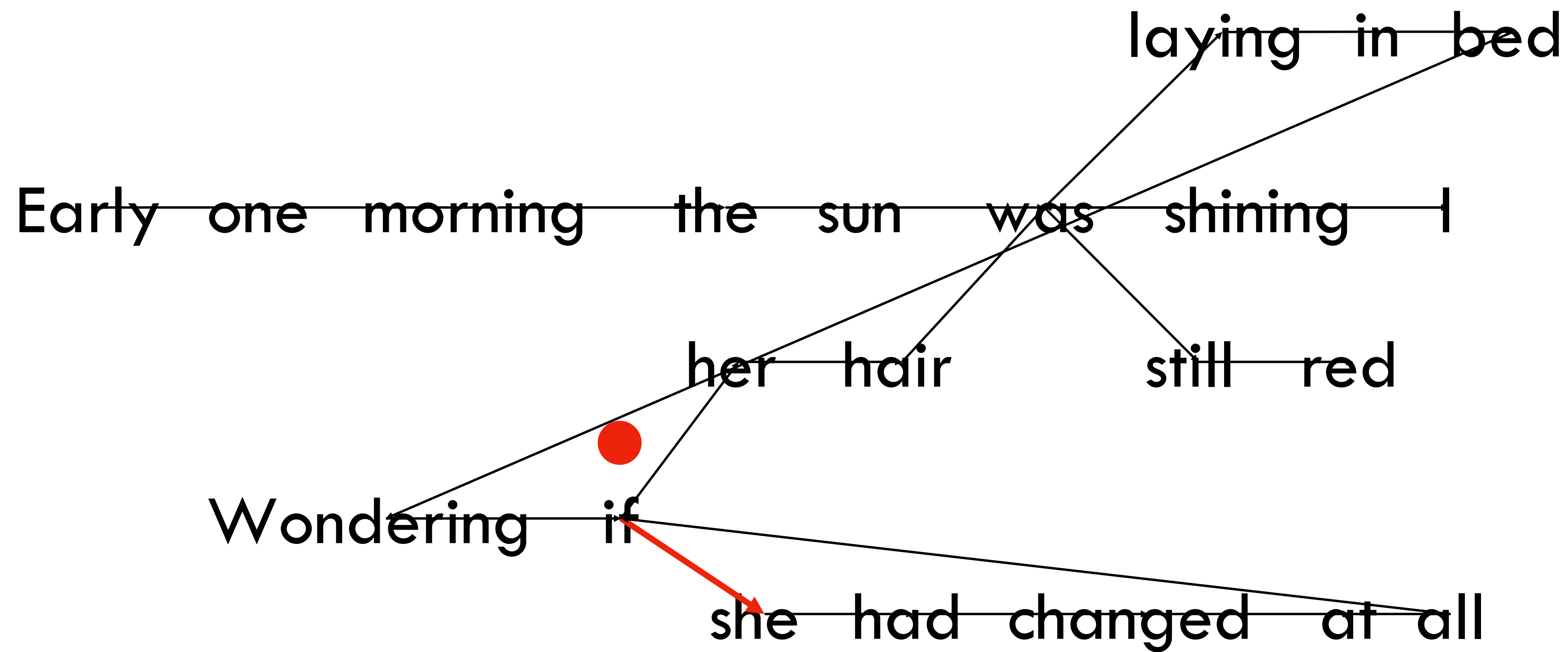
Early one morning the sun was shining I was laying in bed



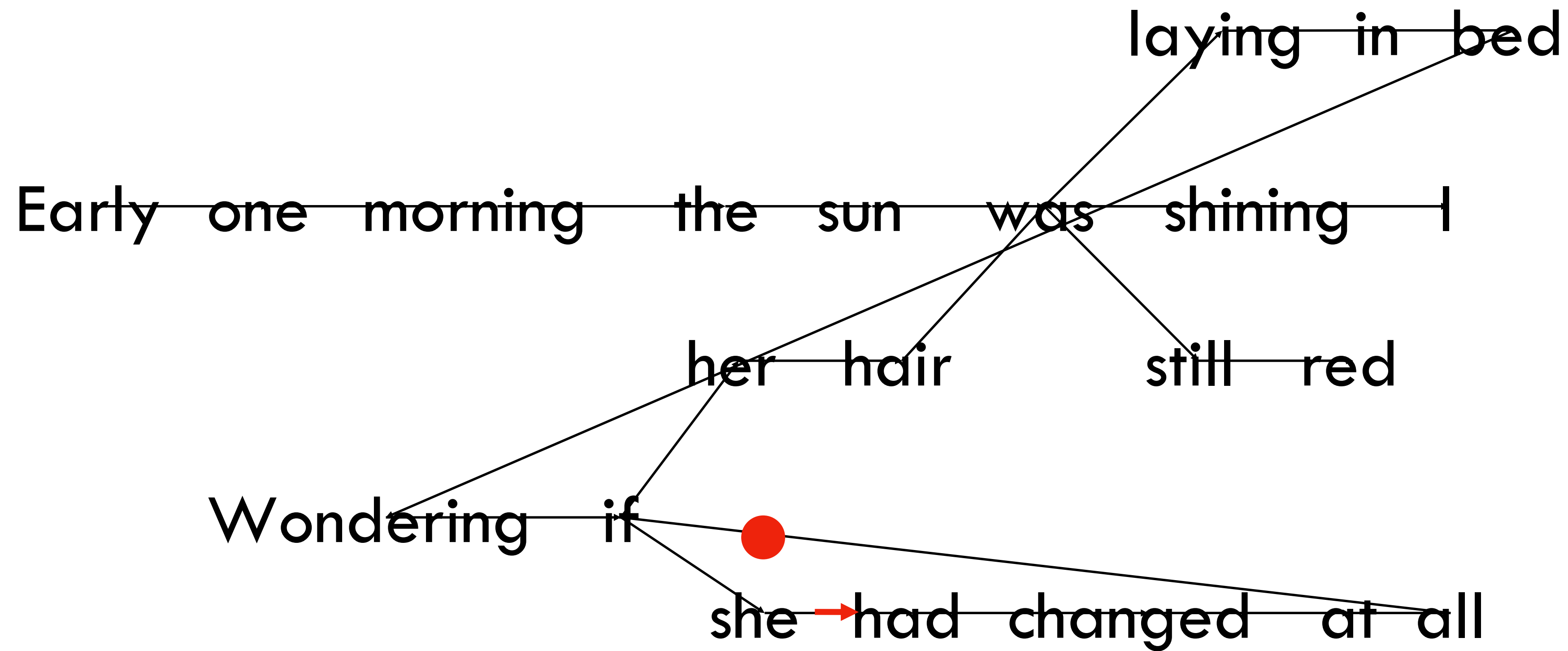
Early one morning the sun was shining I was laying in bed
Wondering



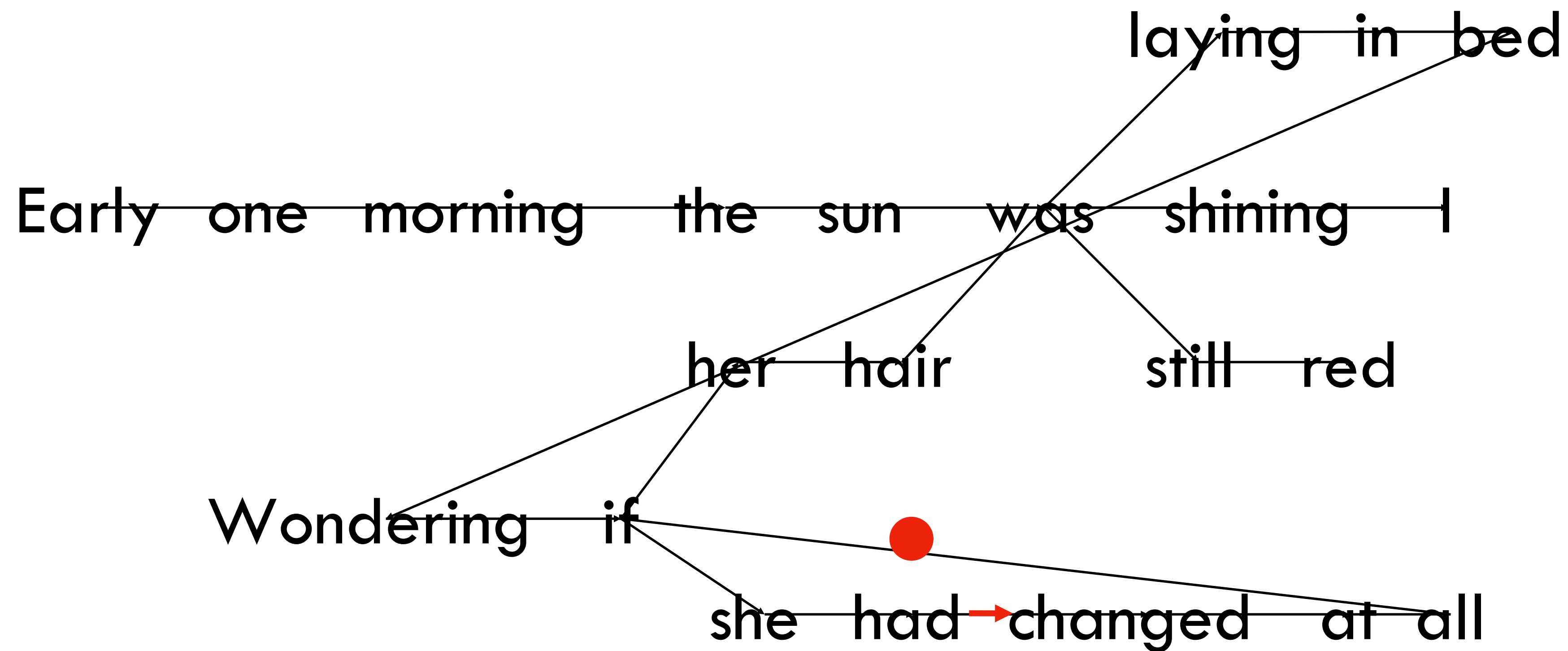
Early one morning the sun was shining I was laying in bed
Wondering if



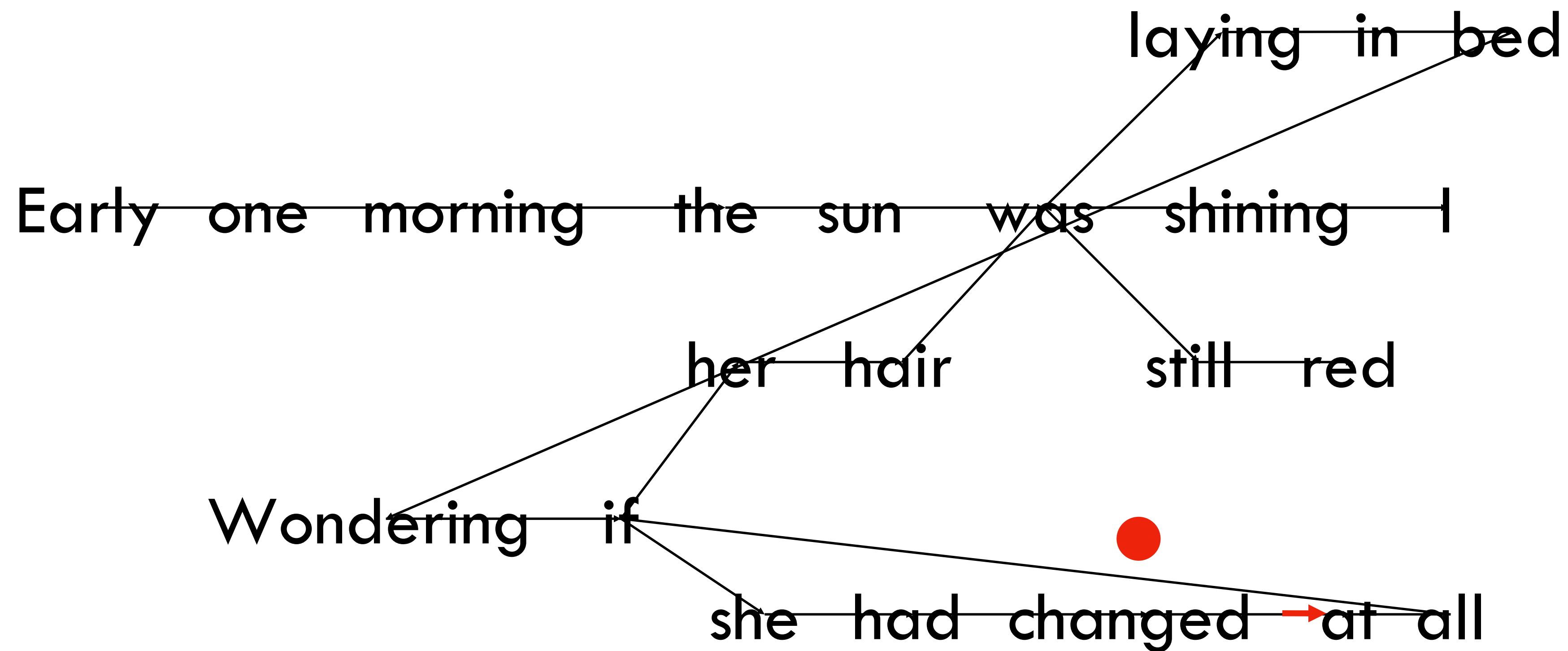
Early one morning the sun was shining I was laying in bed
Wondering if she



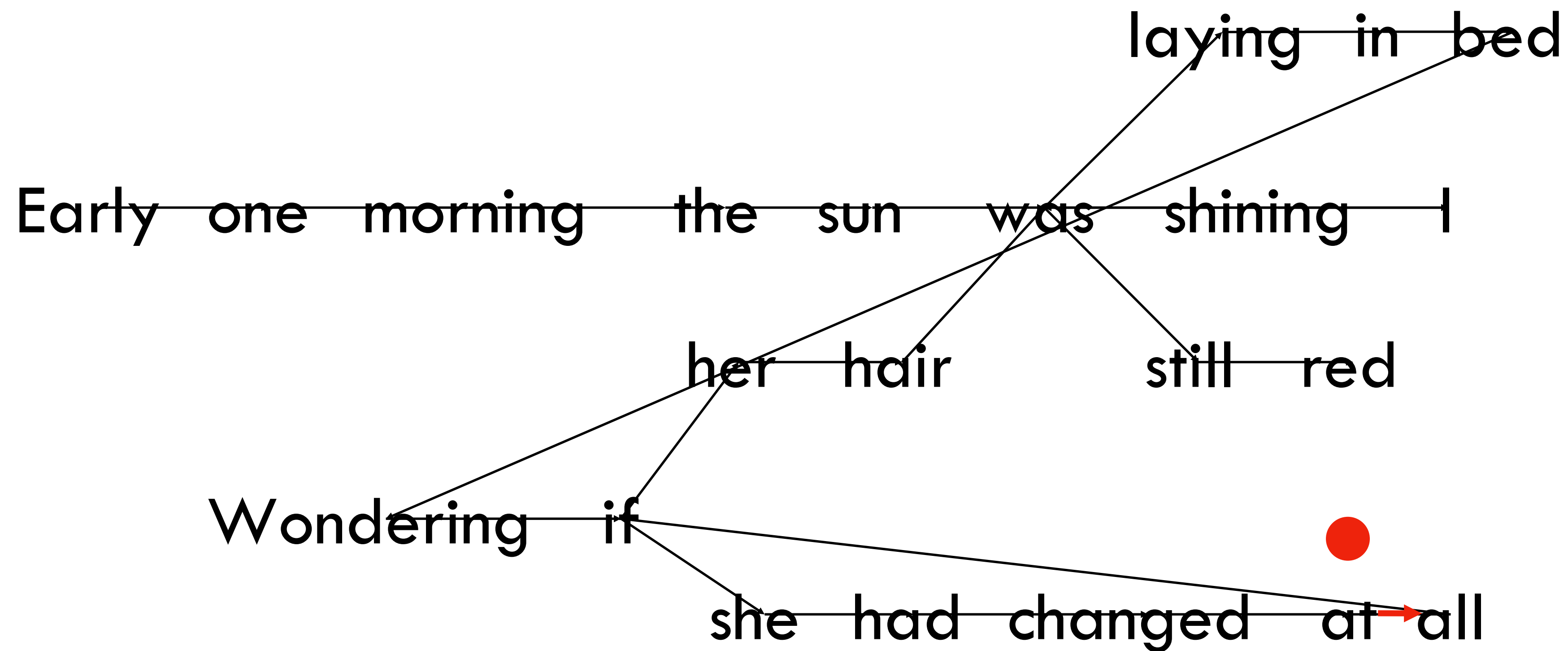
Early one morning the sun was shining I was laying in bed
Wondering if she had



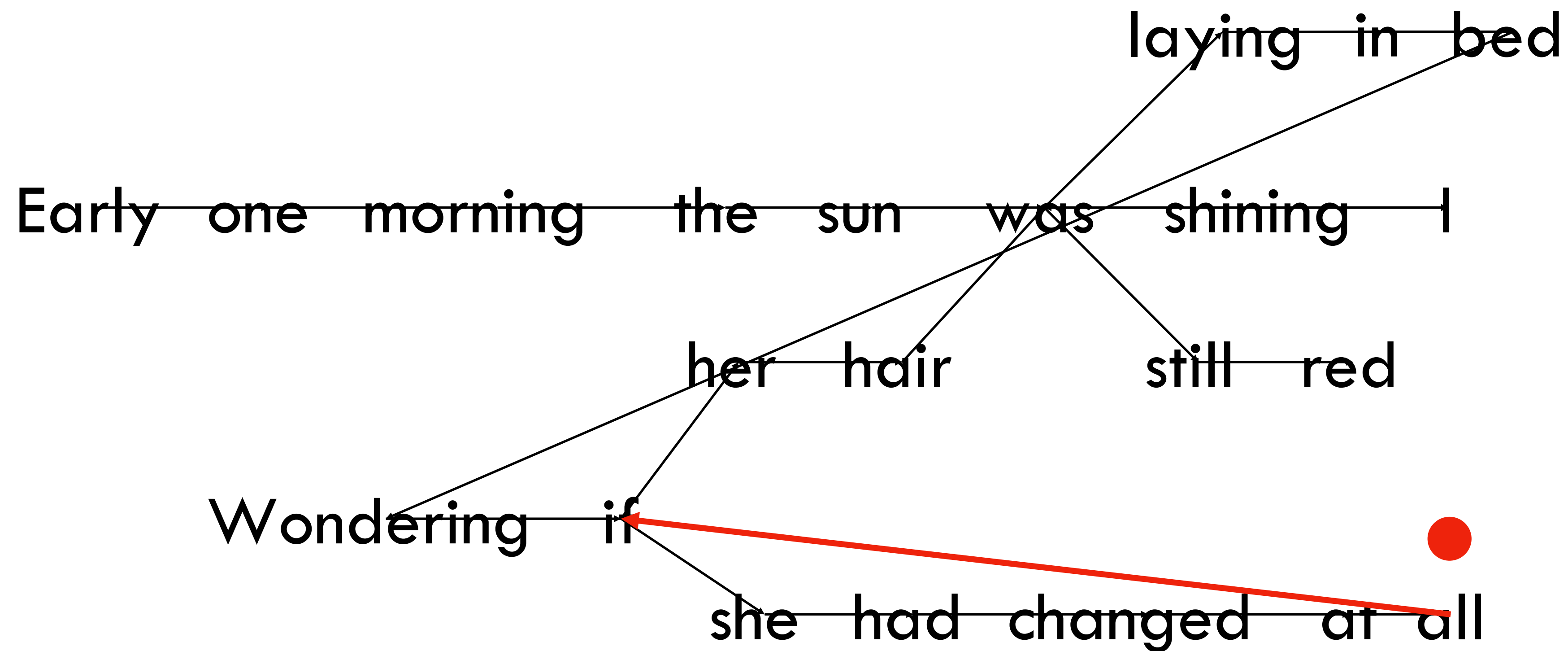
Early one morning the sun was shining I was laying in bed
Wondering if she had changed



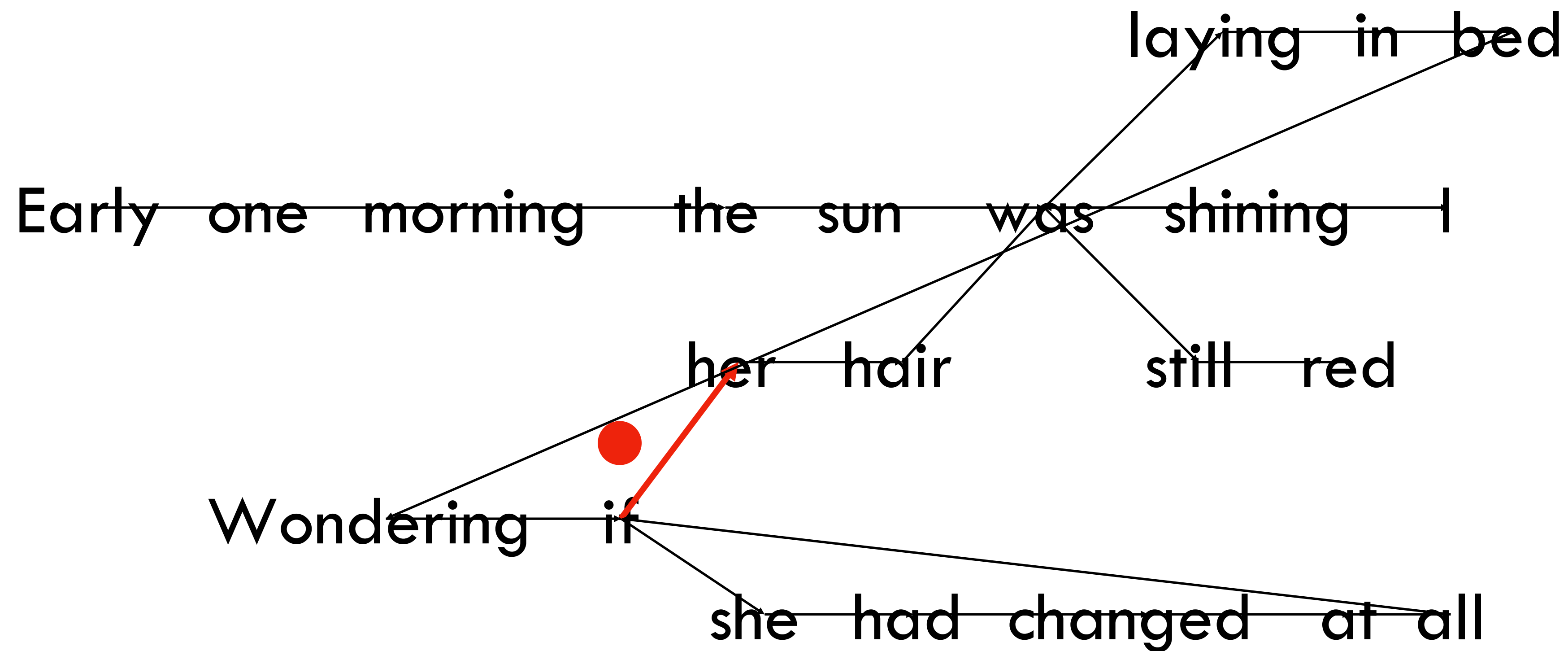
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at



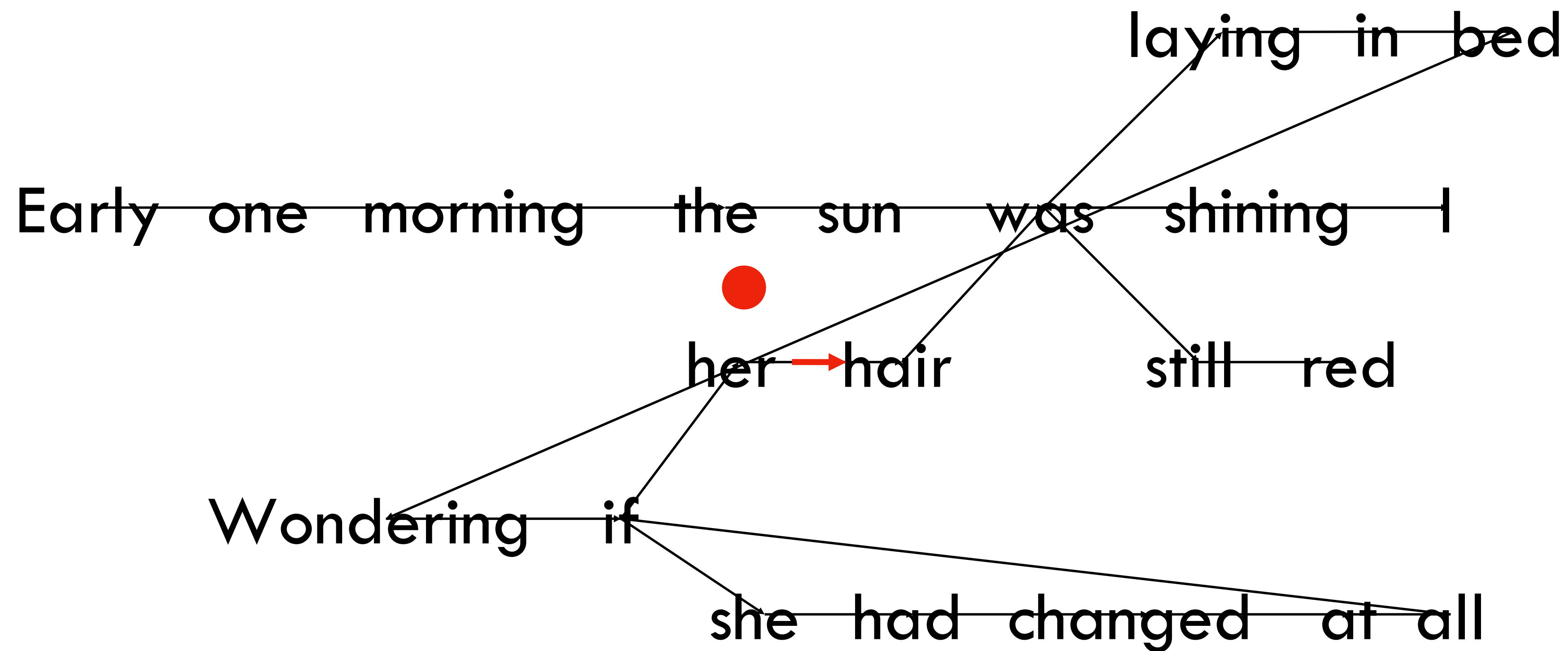
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all



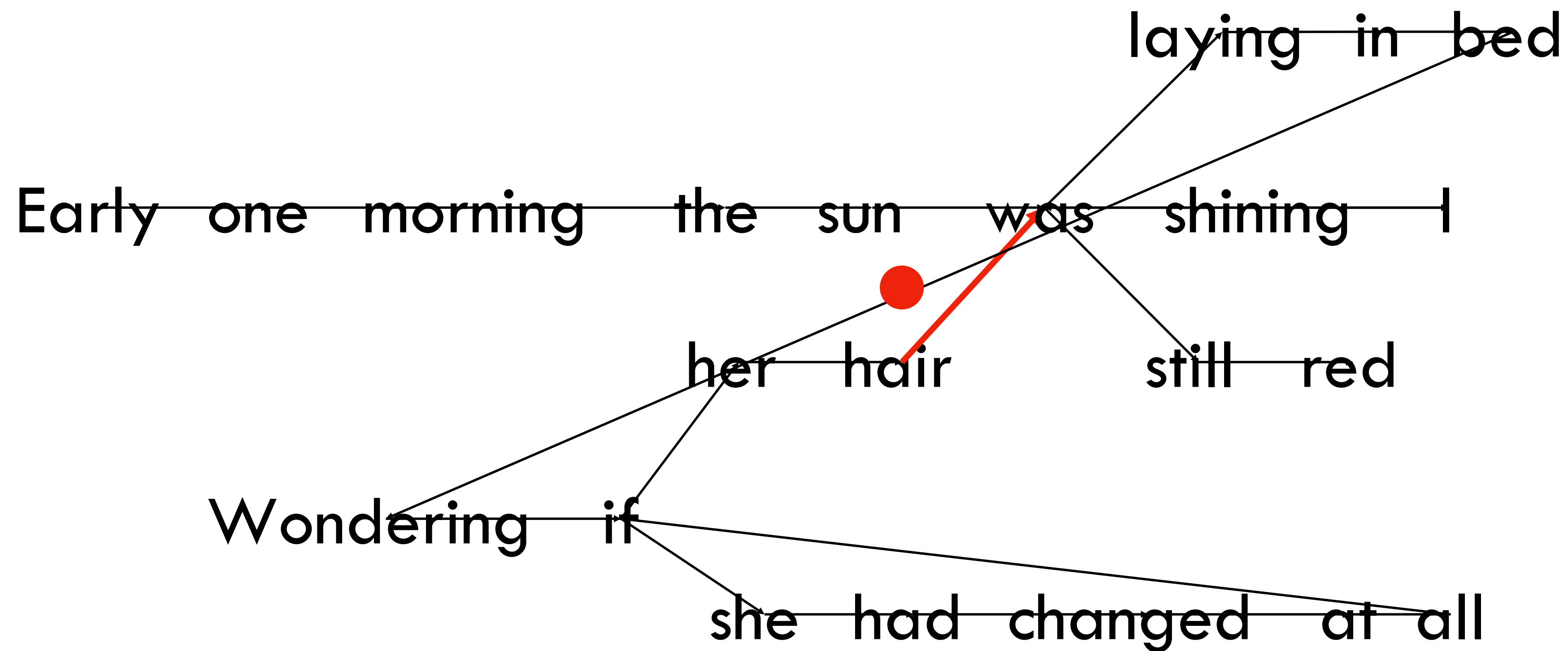
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if



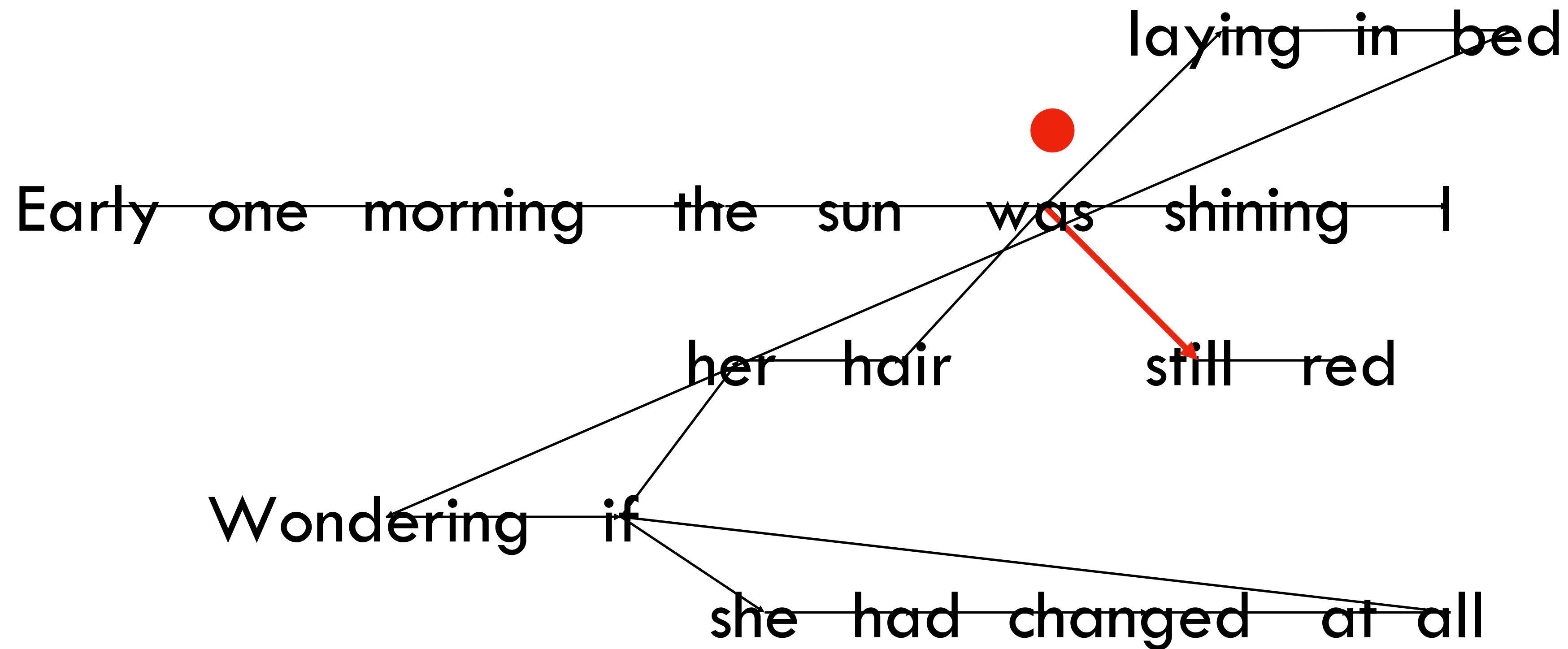
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her



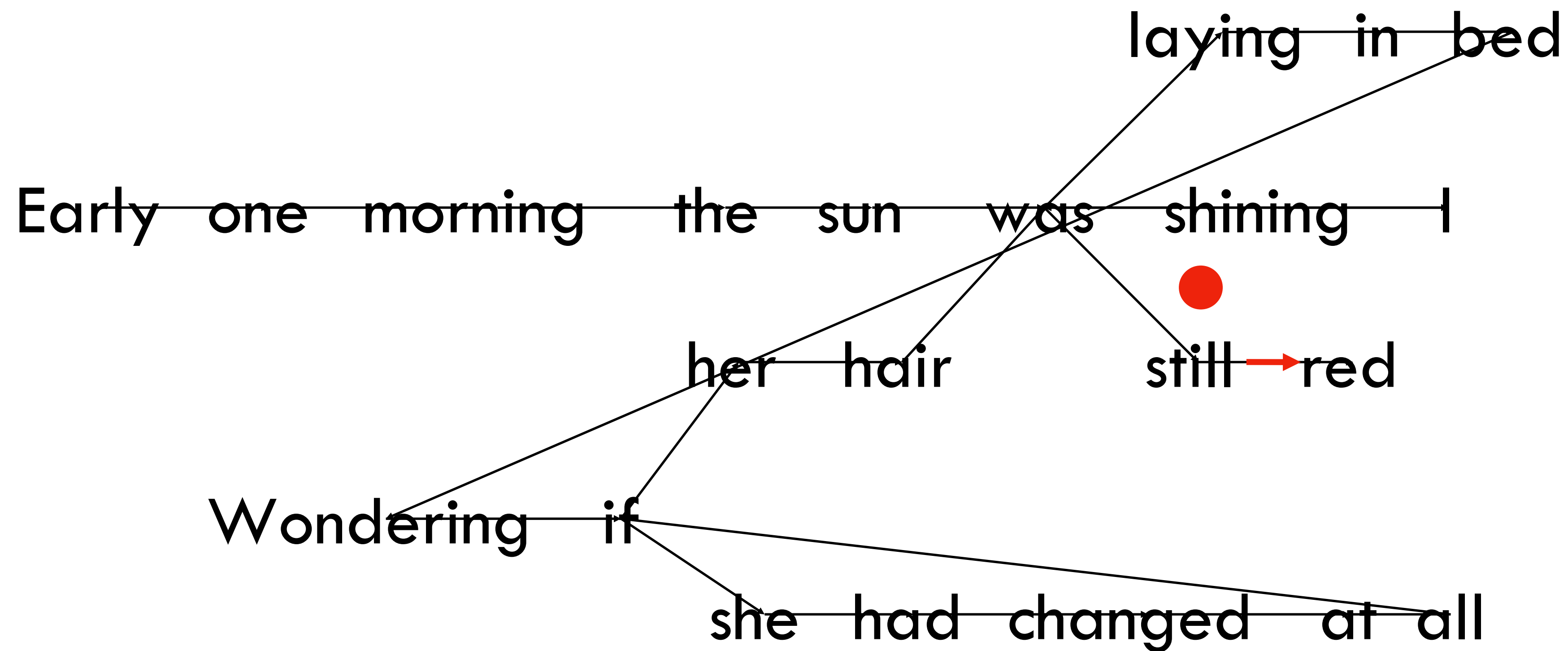
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair



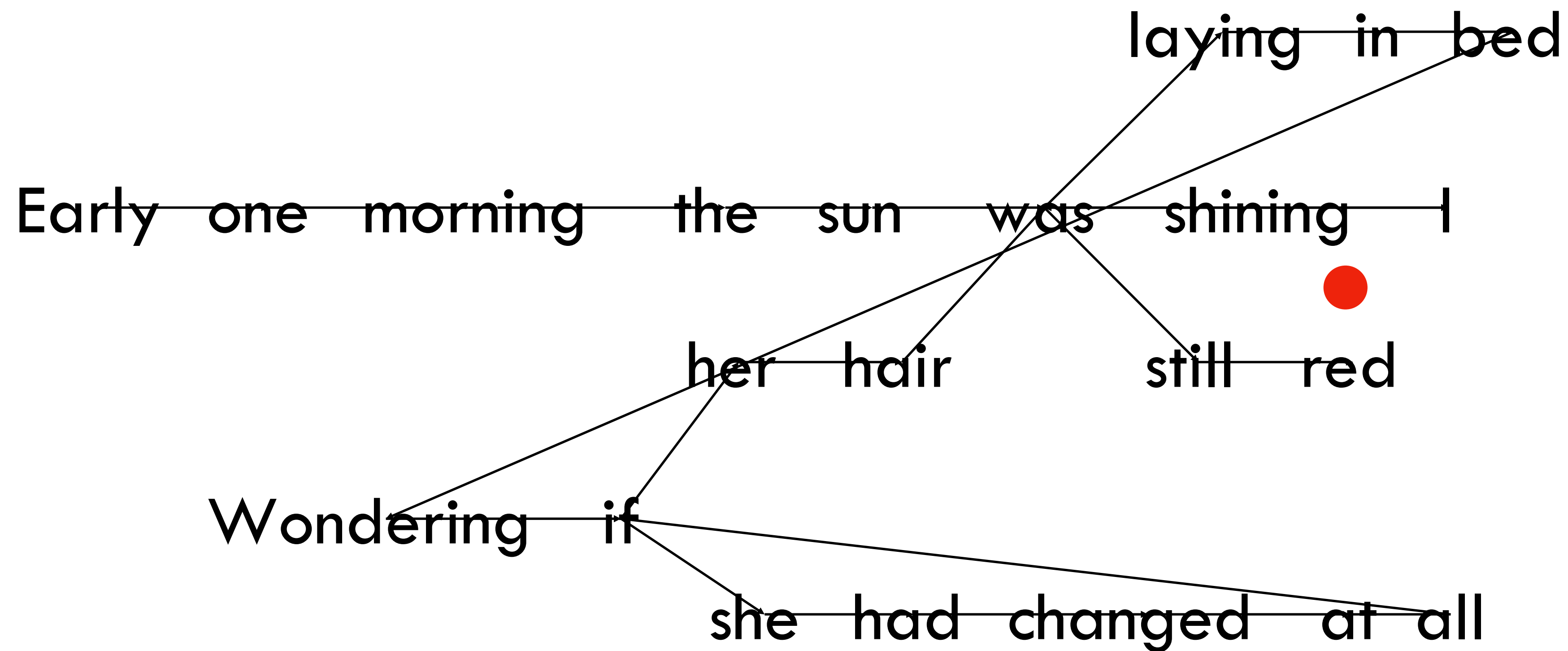
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was



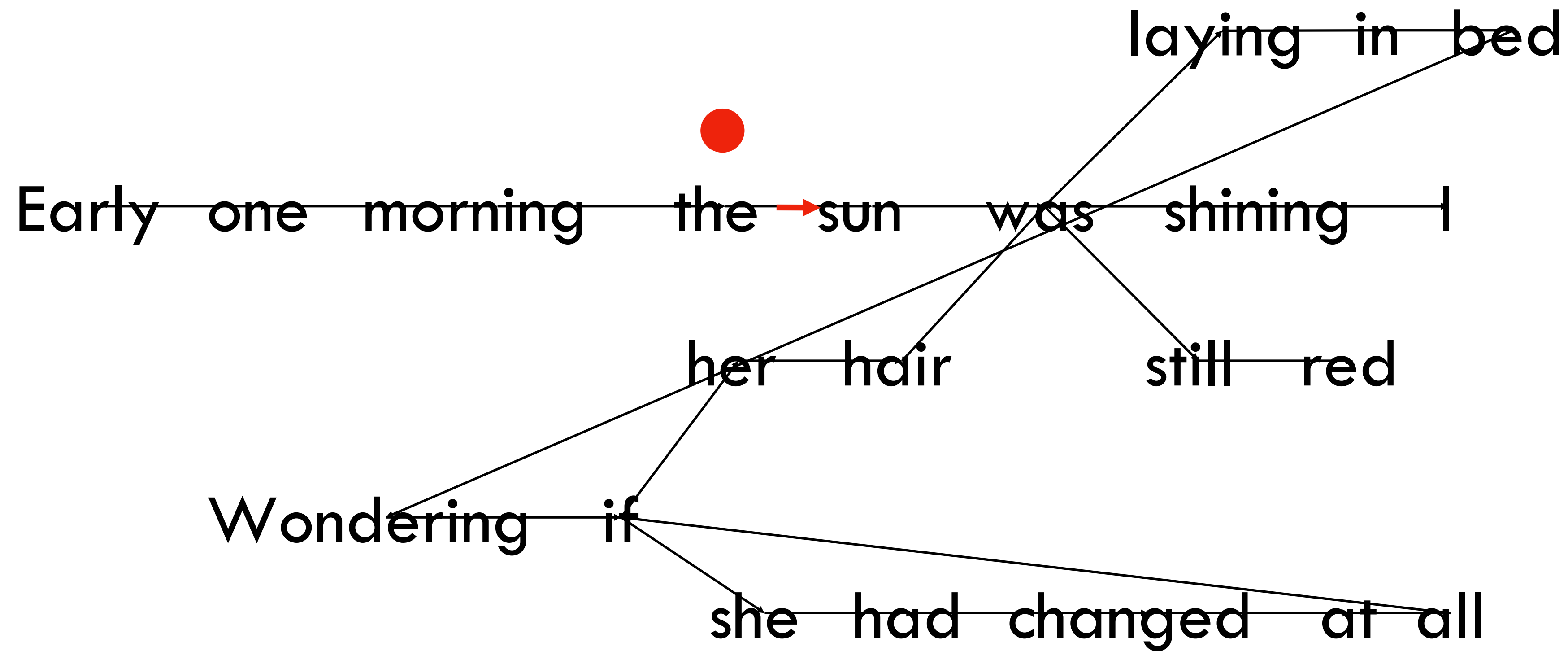
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still



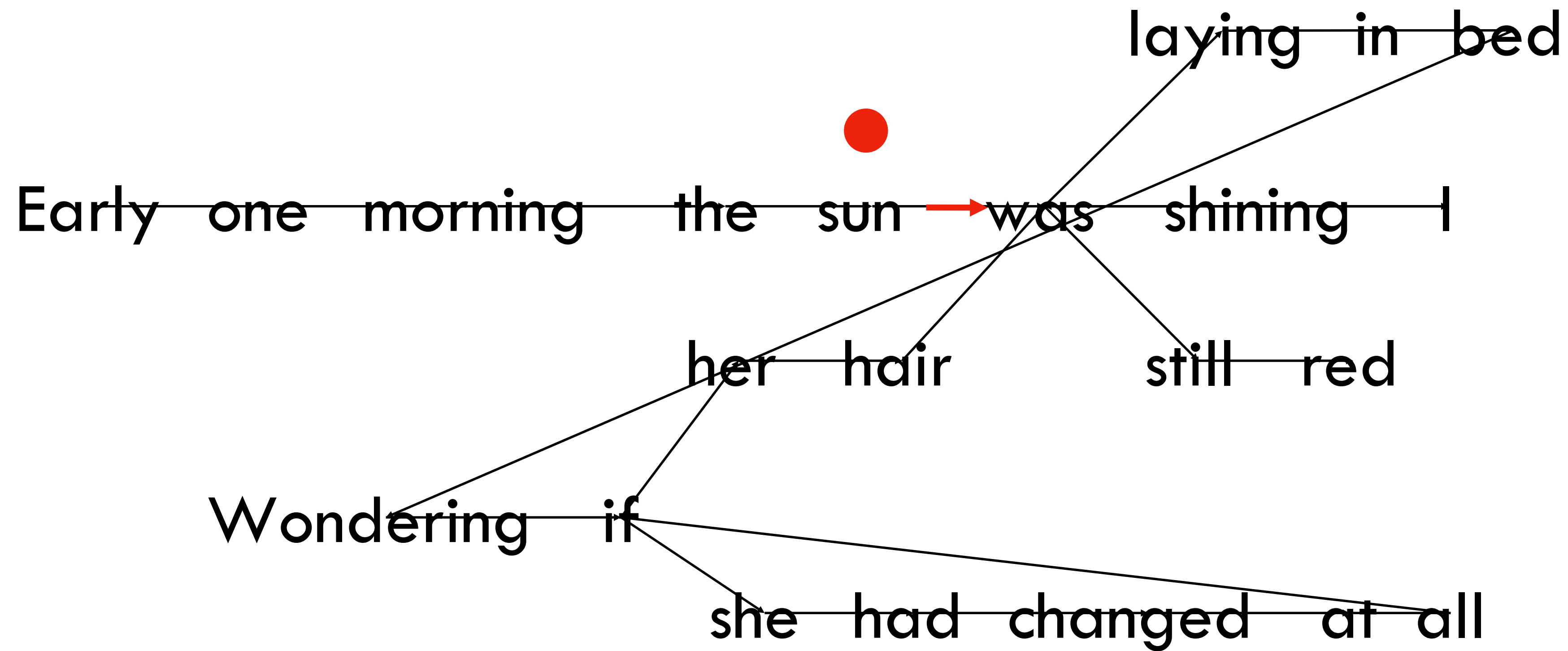
Early one morning the sun was shining I was laying in bed
Wondering if she had changed at all if her hair was still red



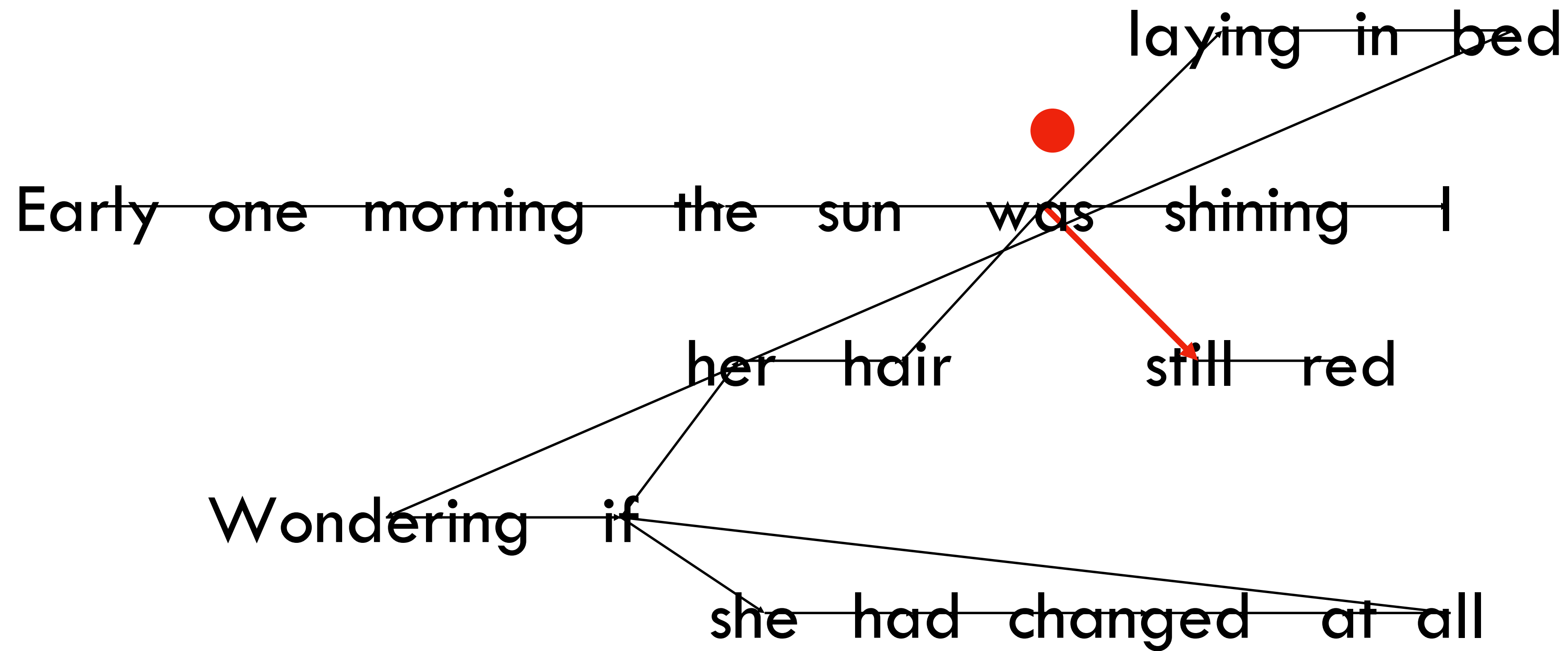
the



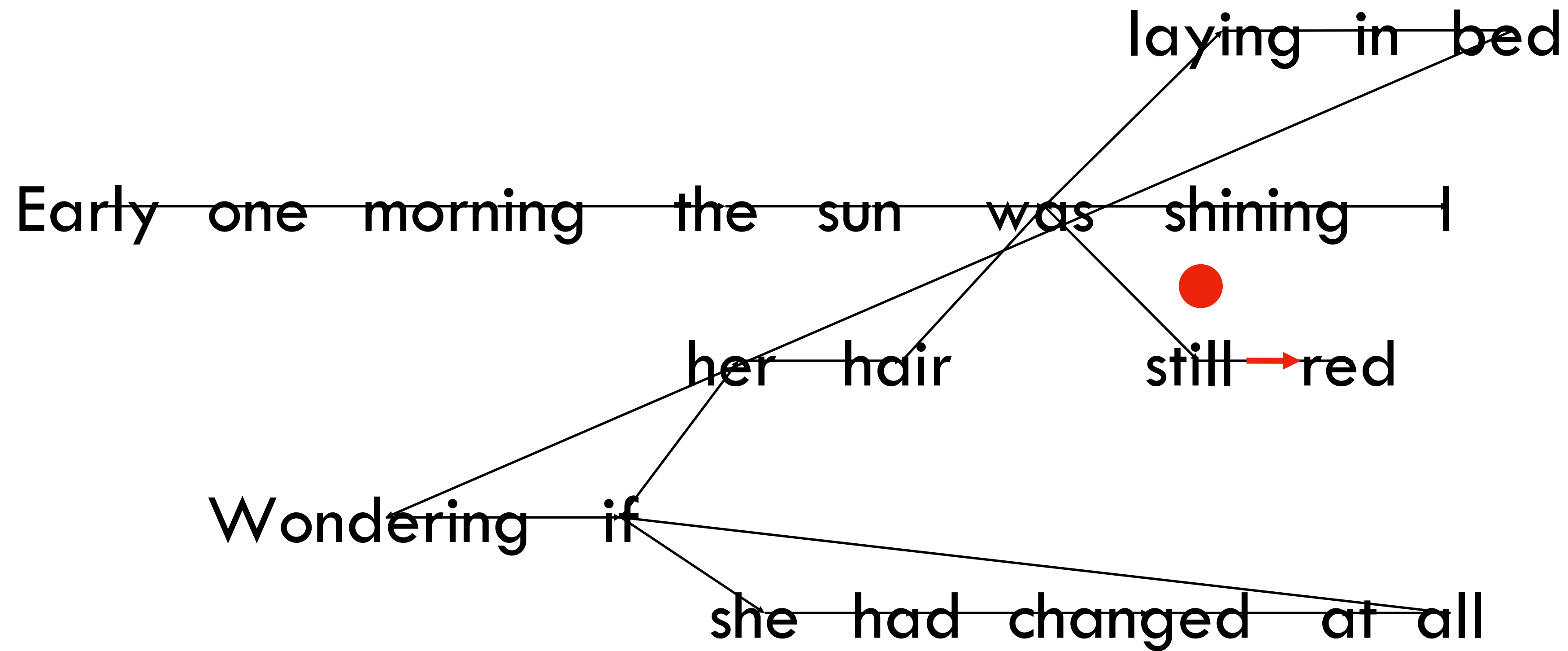
the sun



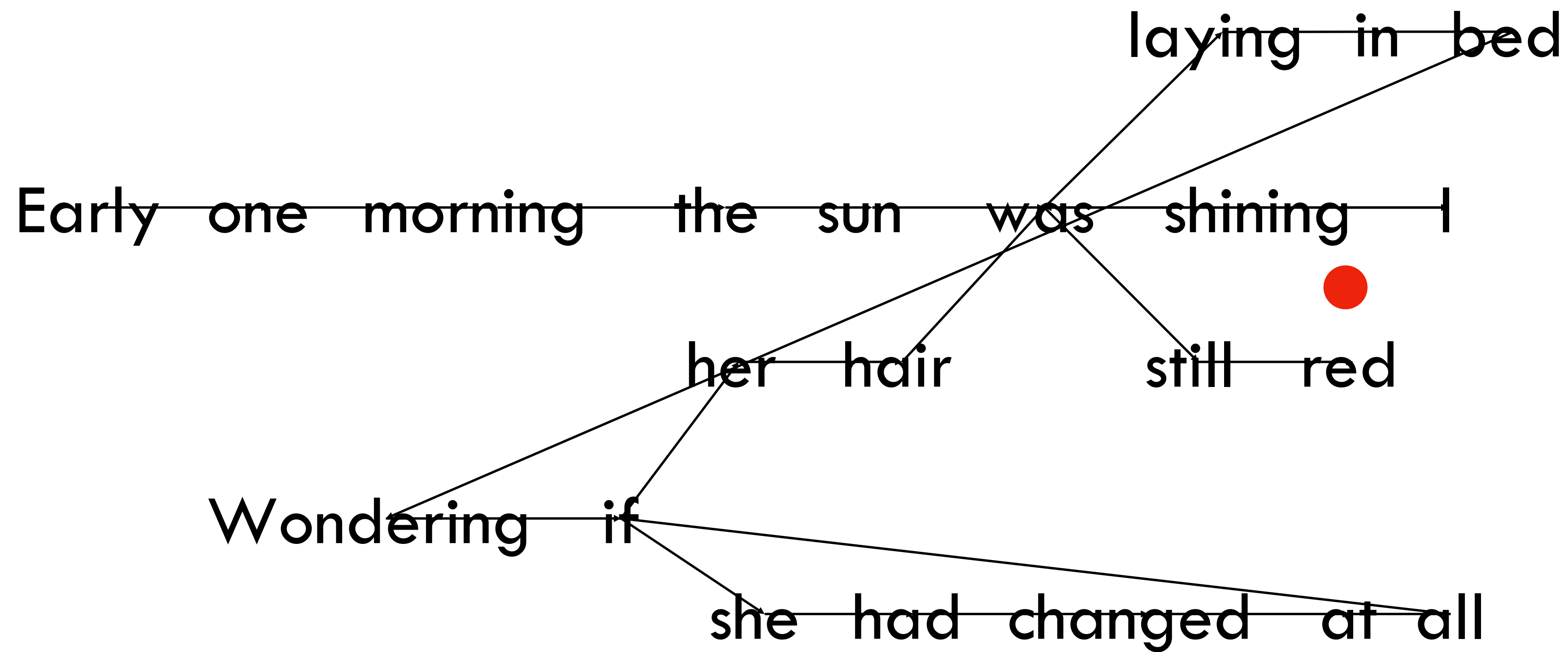
the sun was



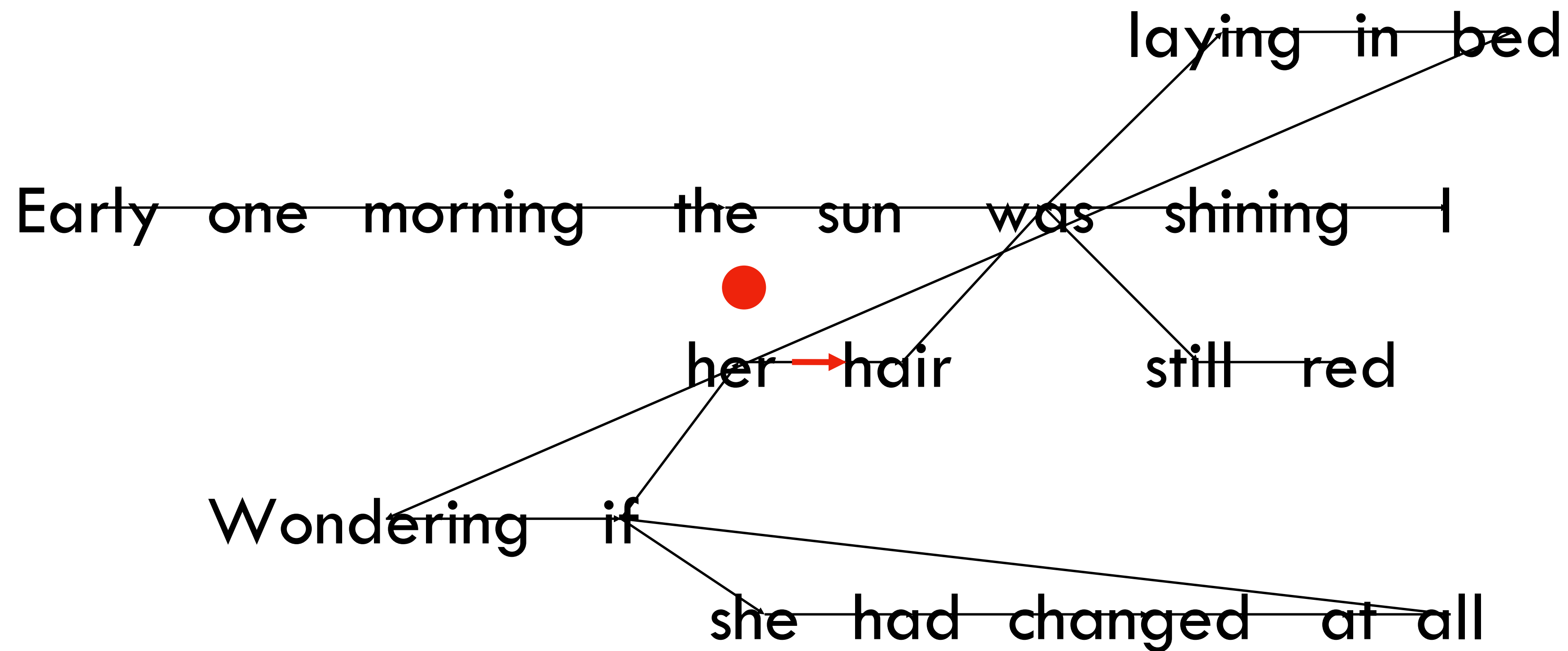
the sun was still



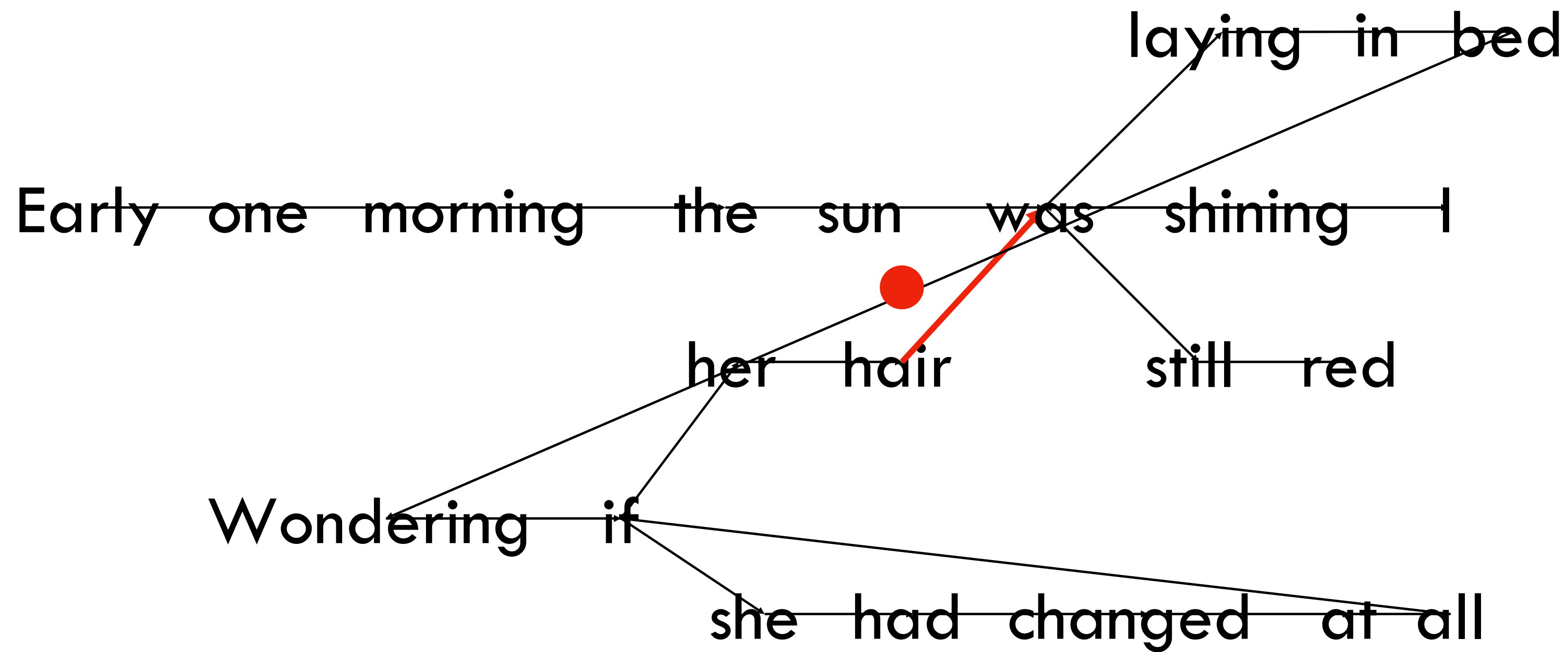
the sun was still red



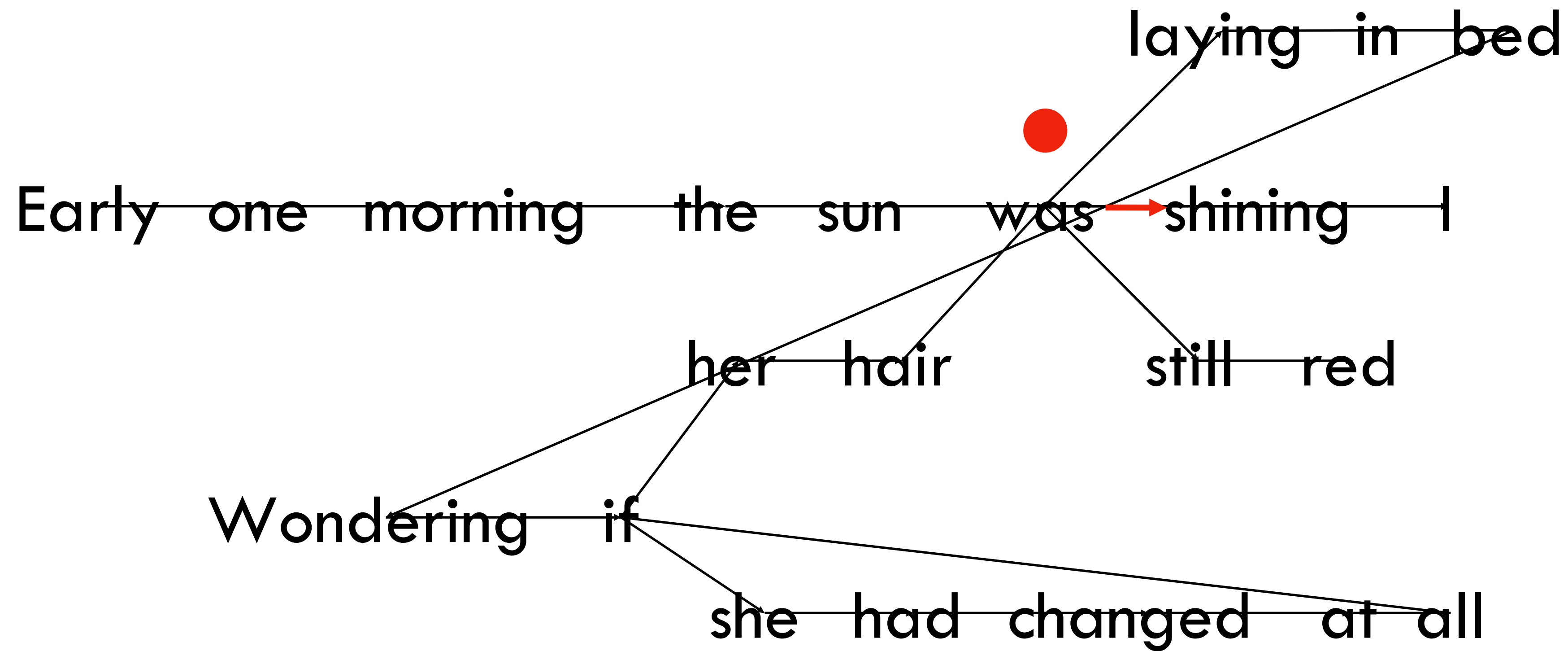
the sun was still red
her



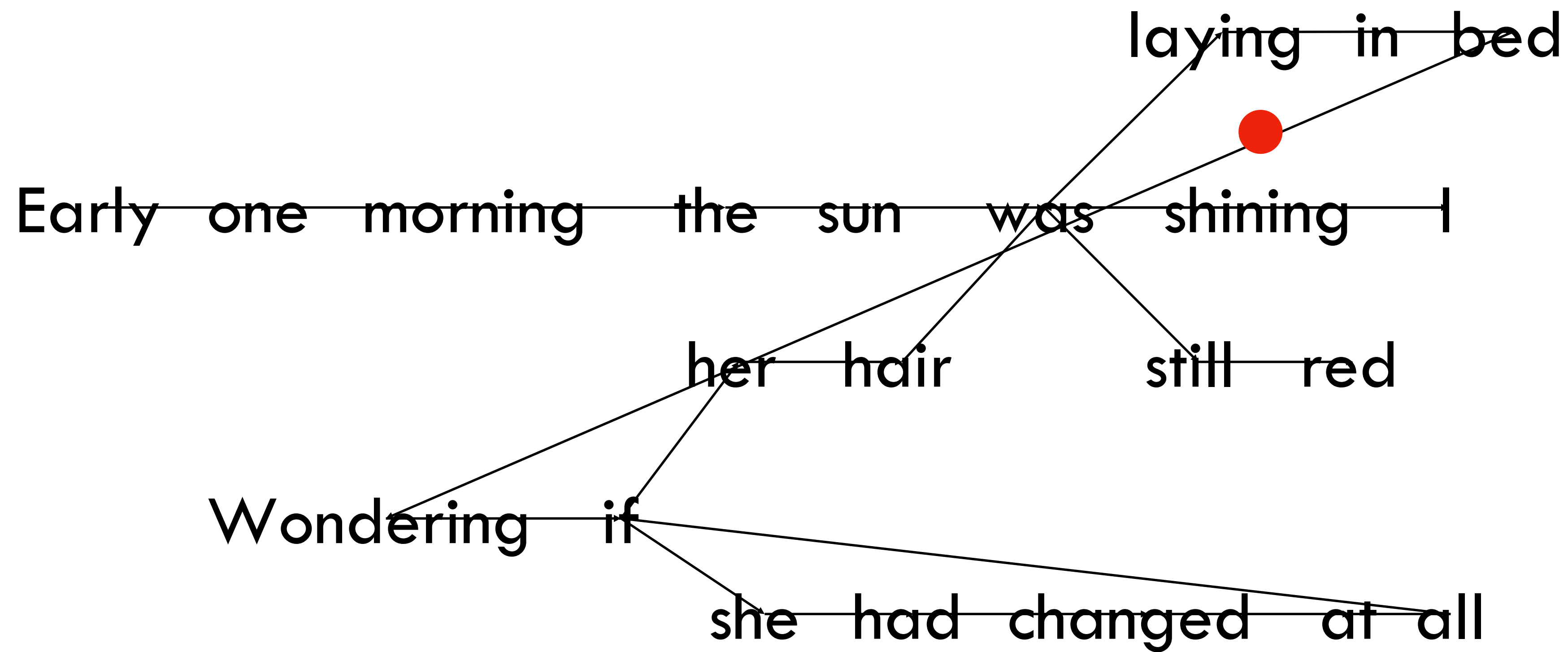
the sun was still red
her hair



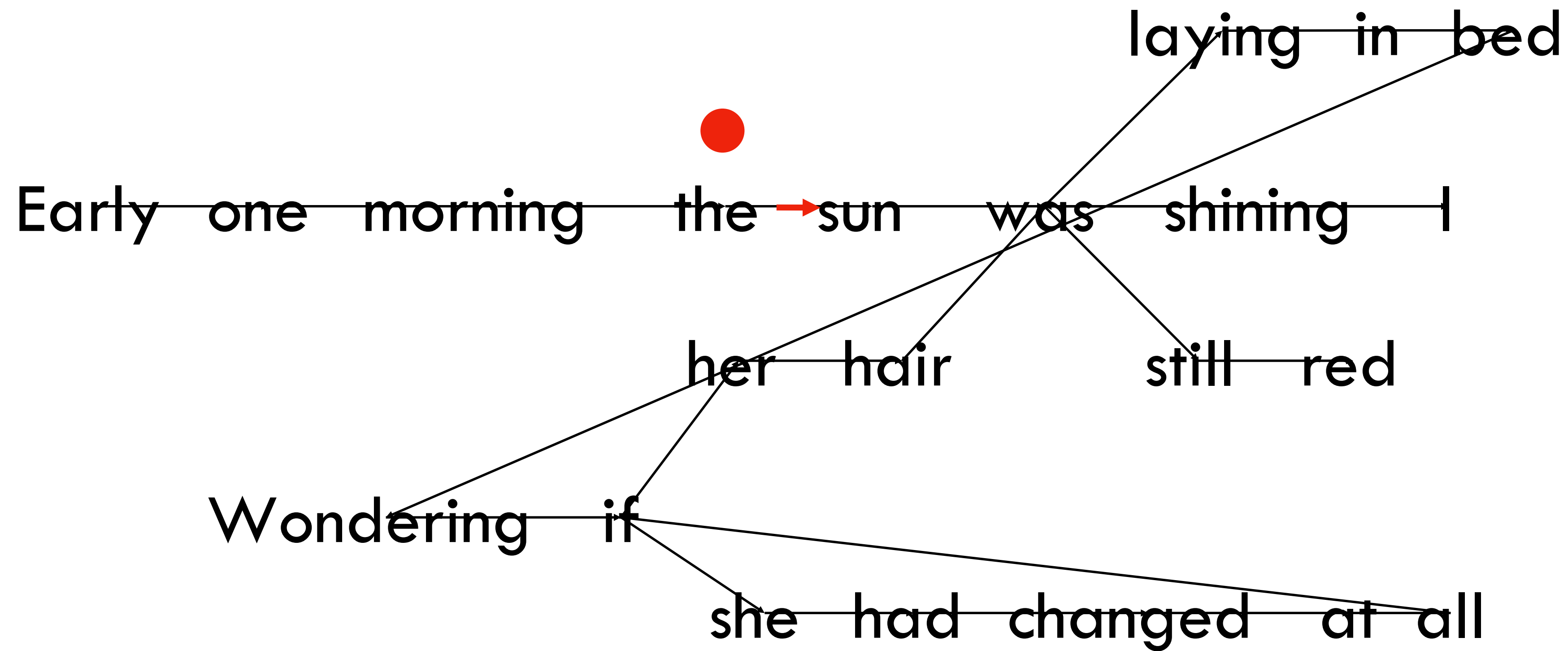
the sun was still red
her hair was



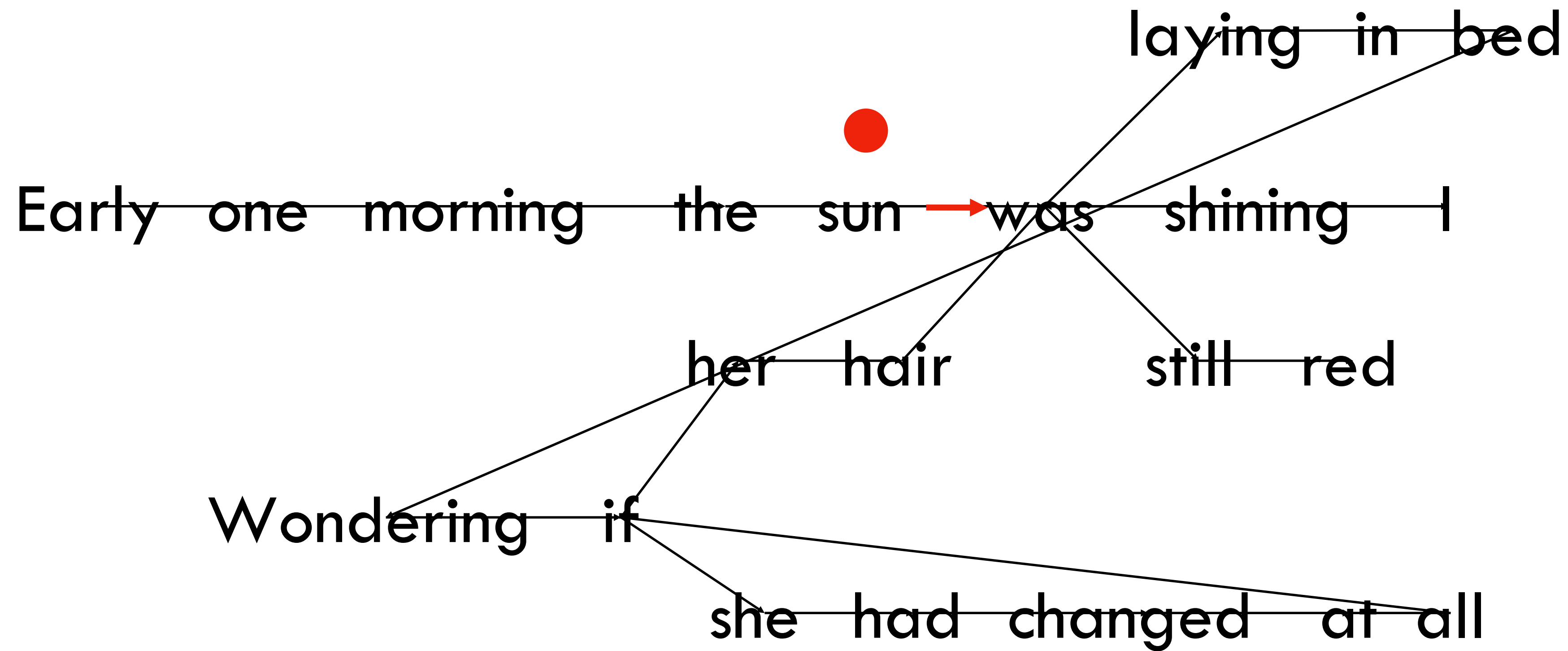
the sun was still red
her hair was shining



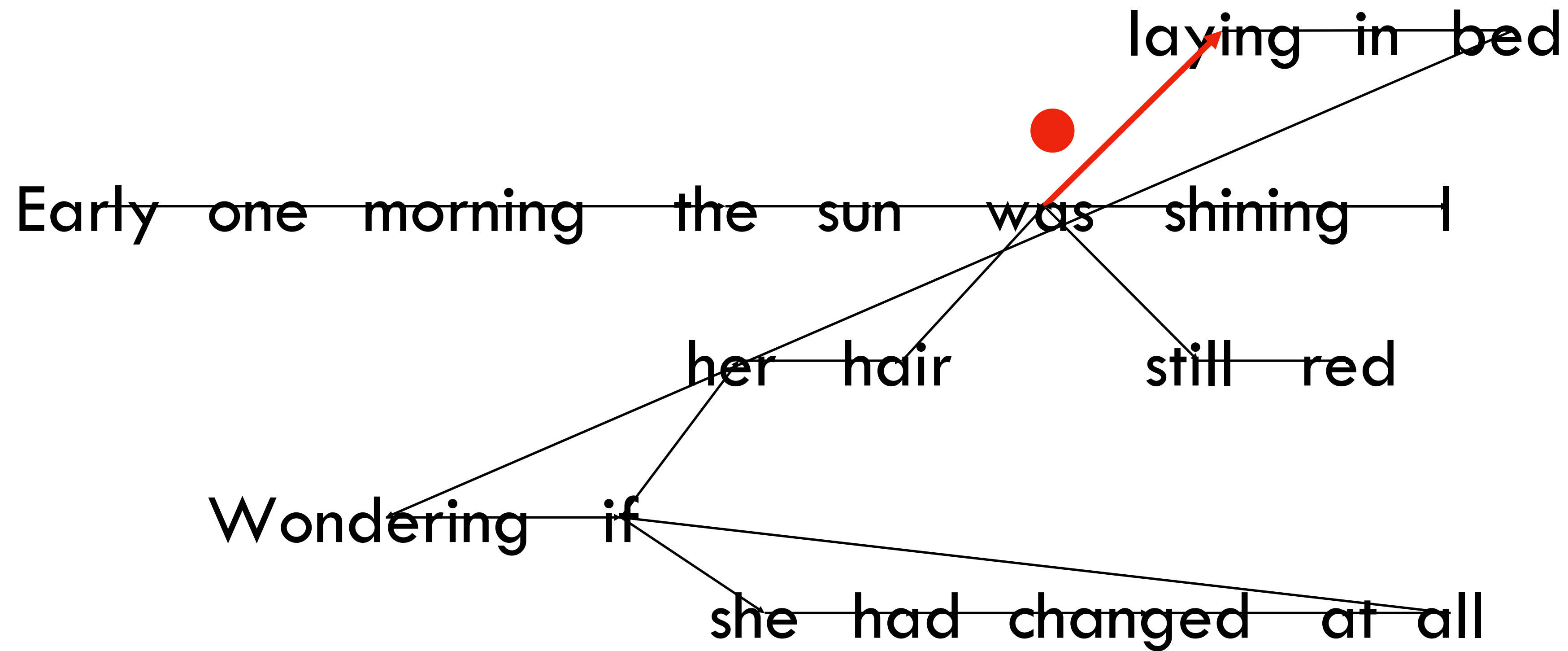
the



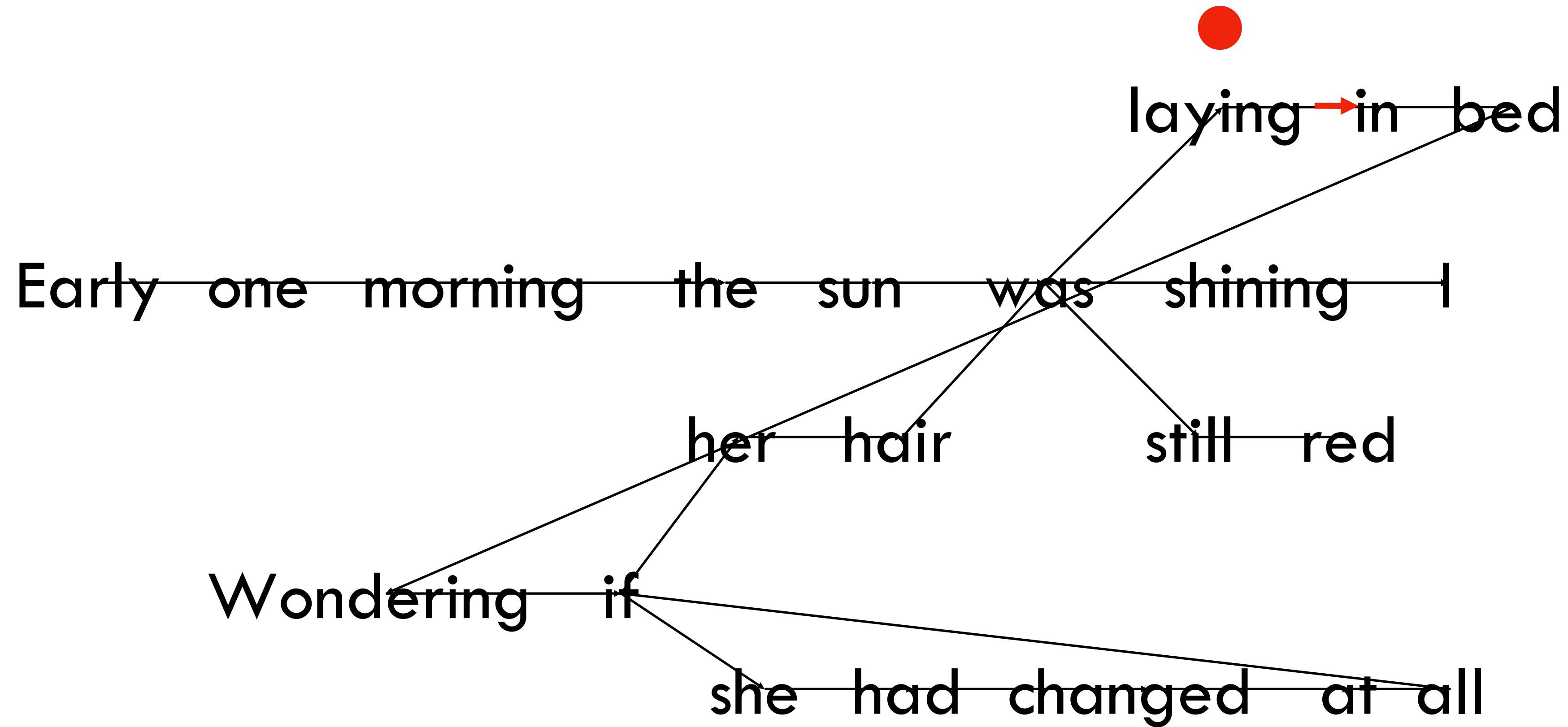
the sun



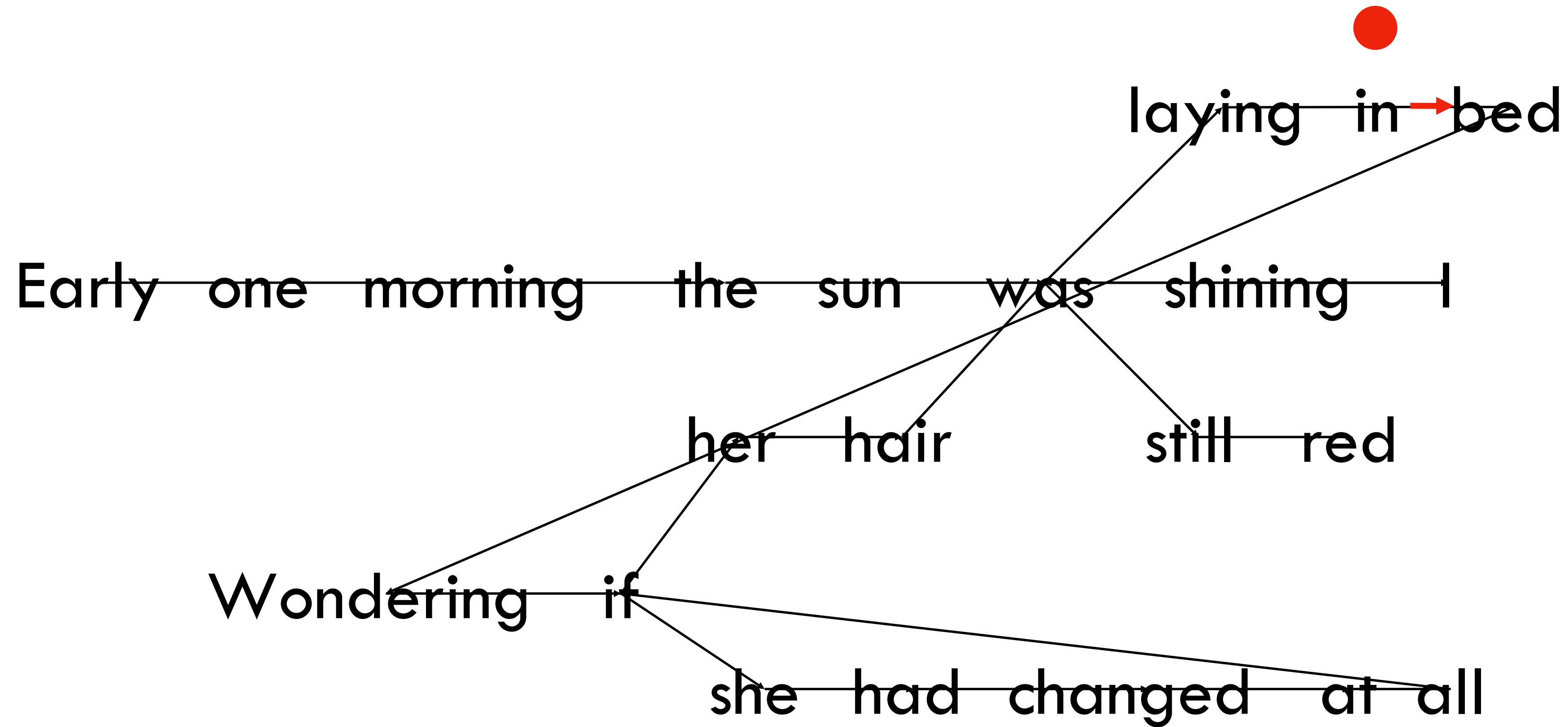
the sun was



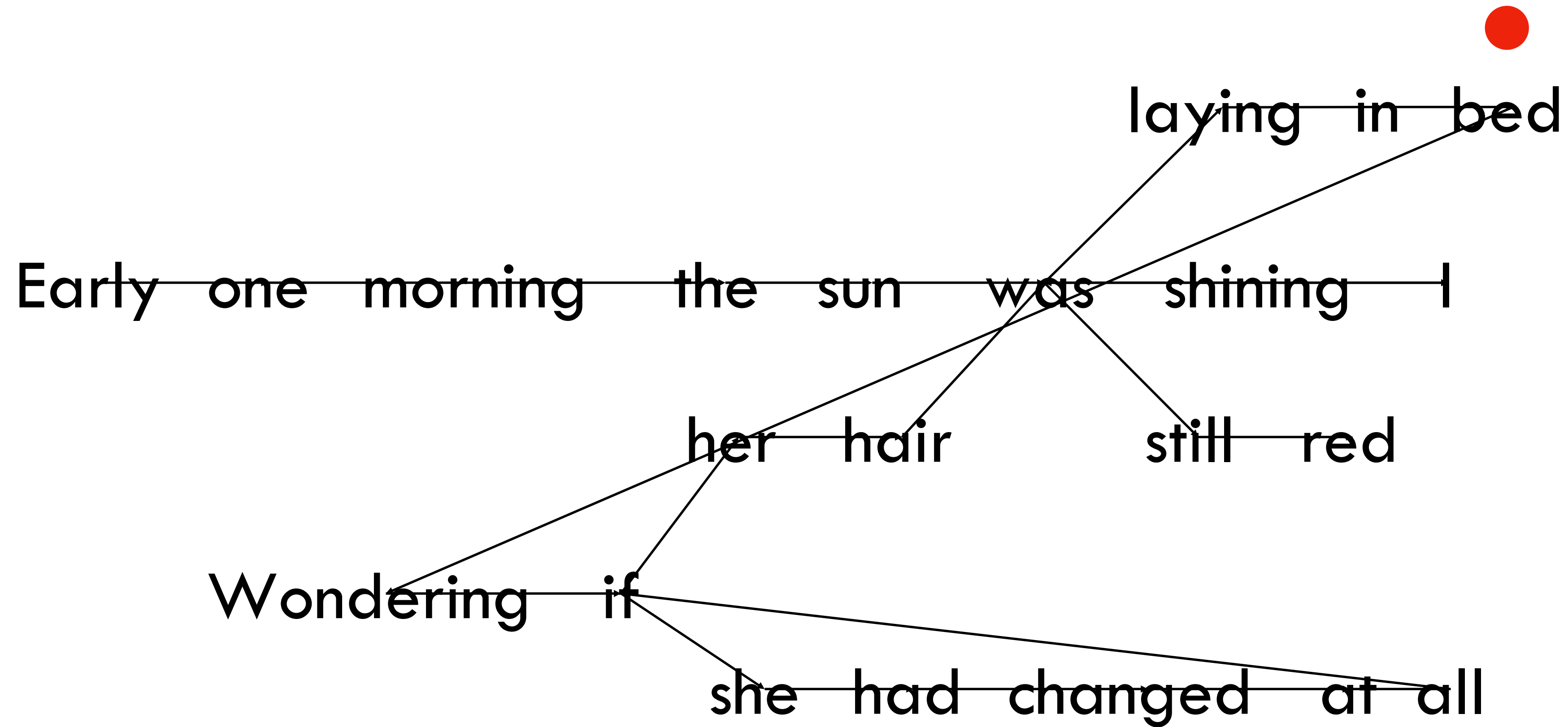
the sun was laying



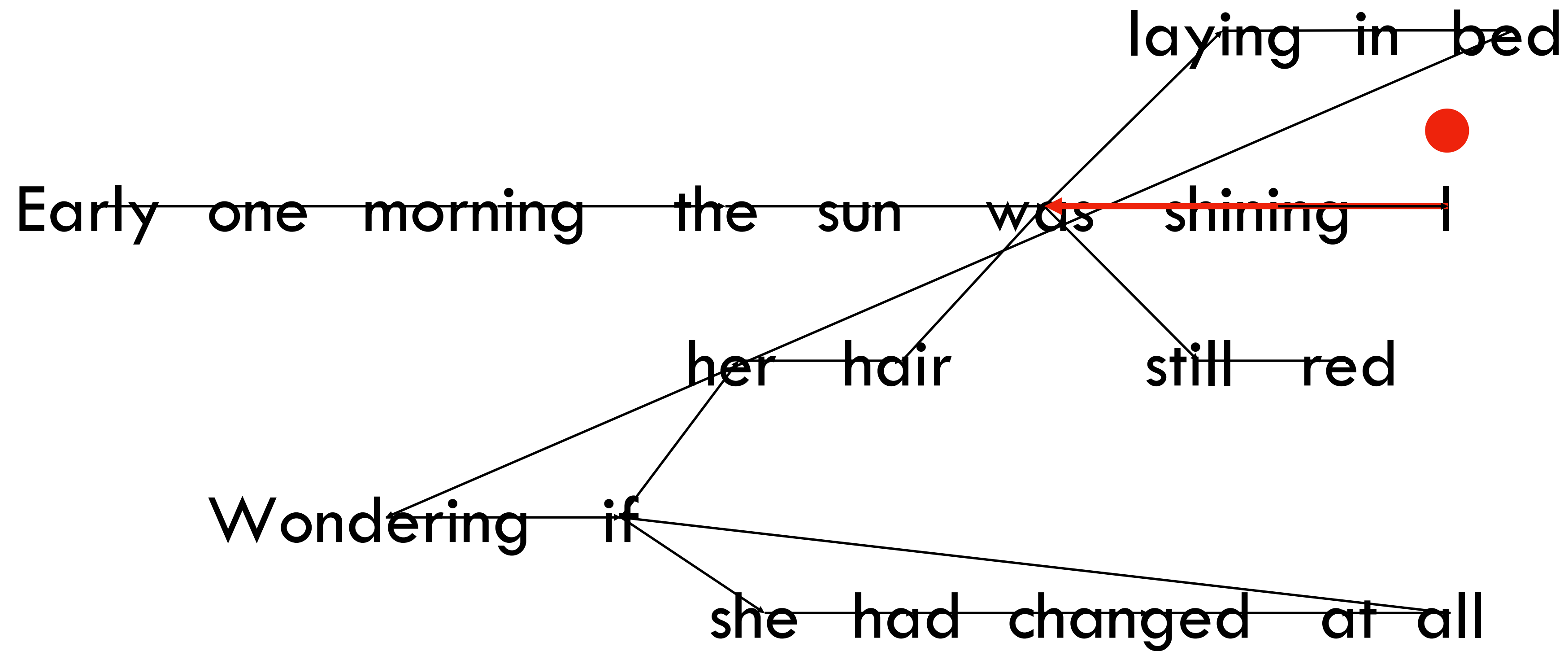
the sun was laying in



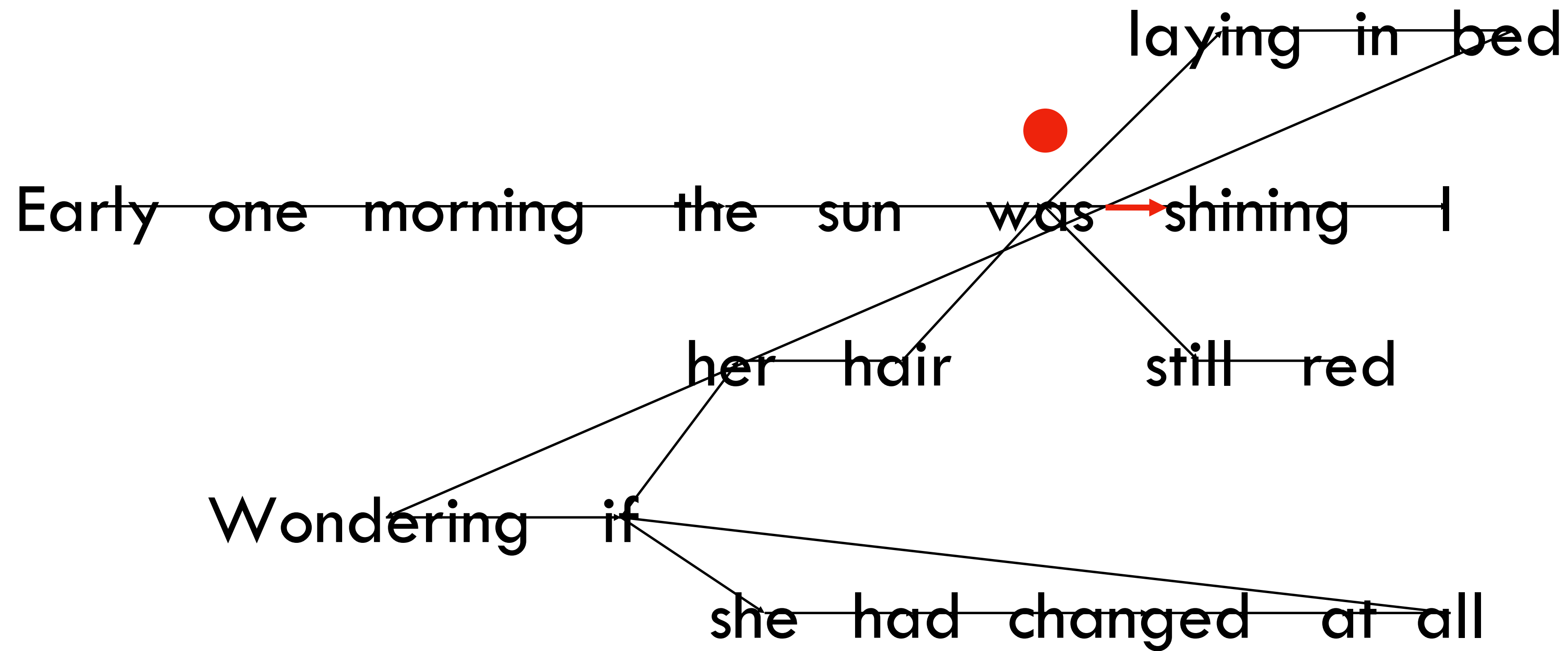
the sun was laying in bed



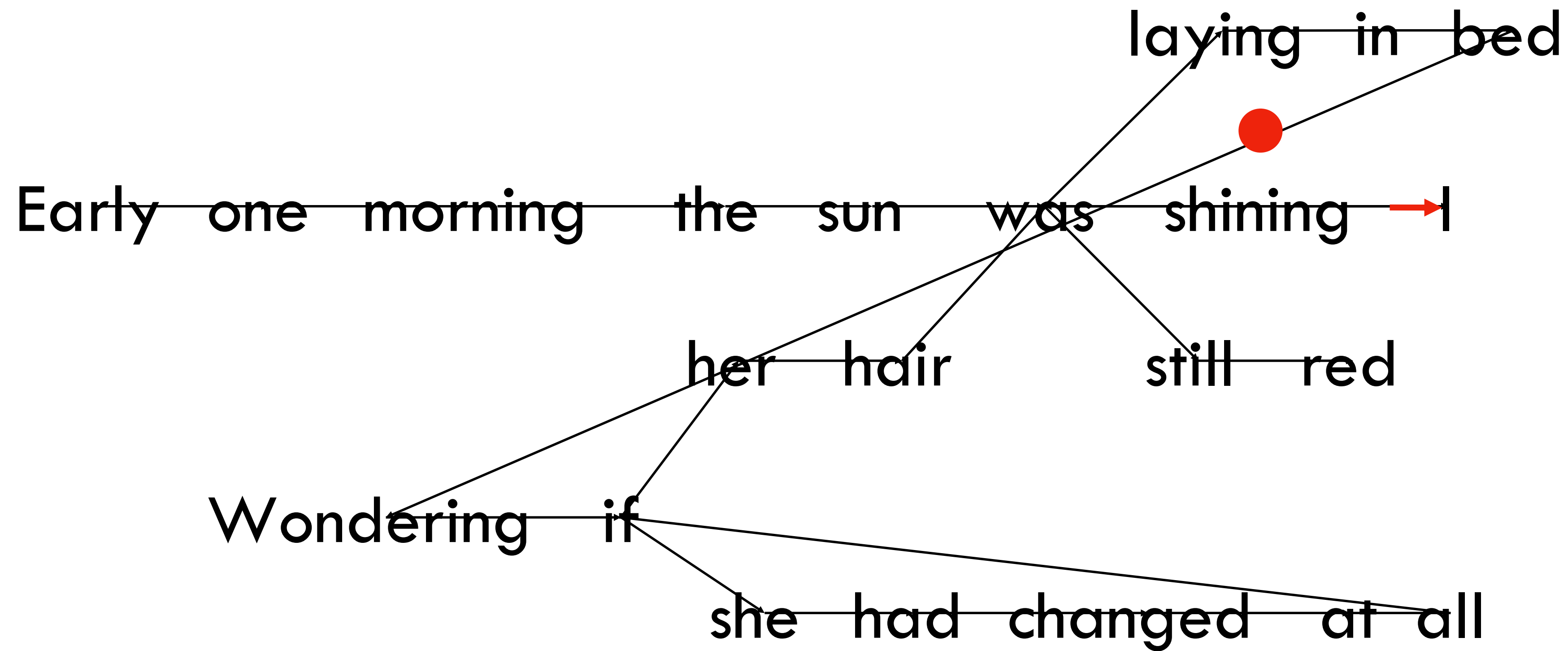
I



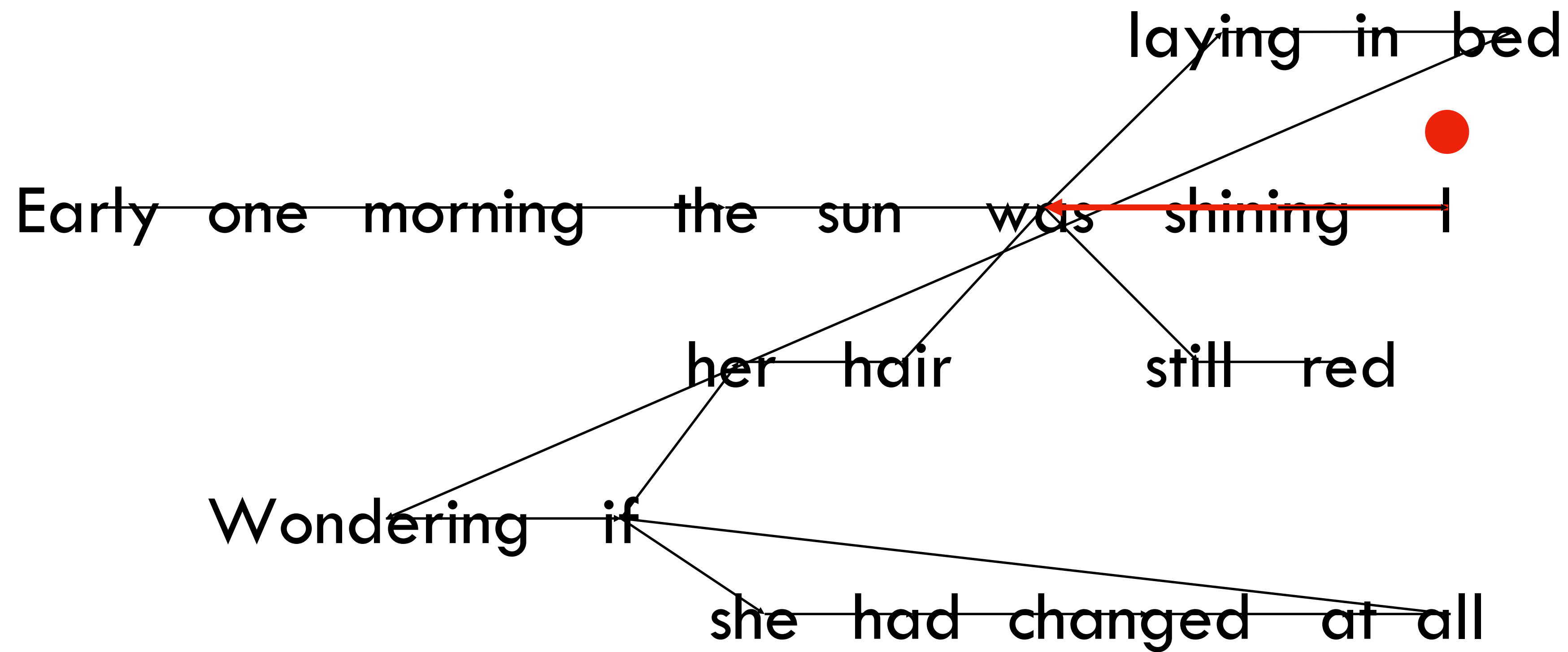
I was



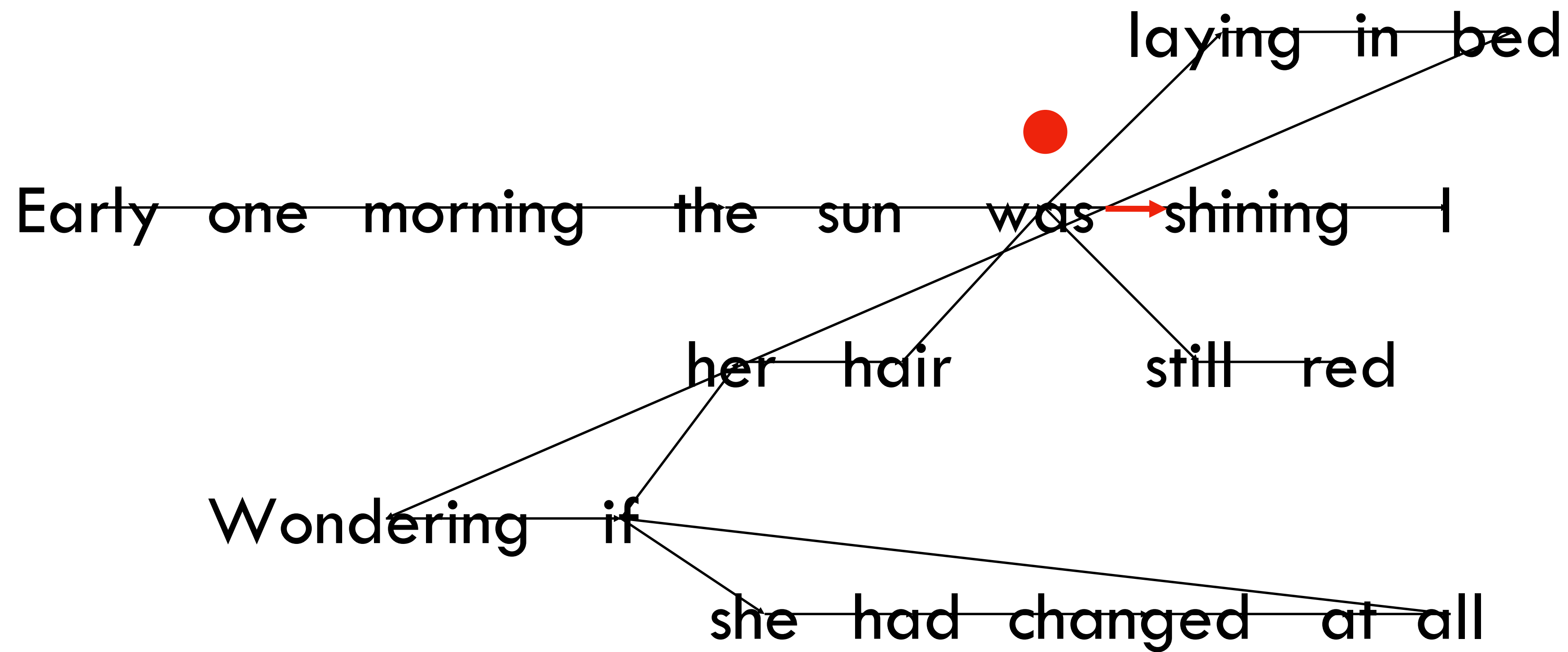
I was shining



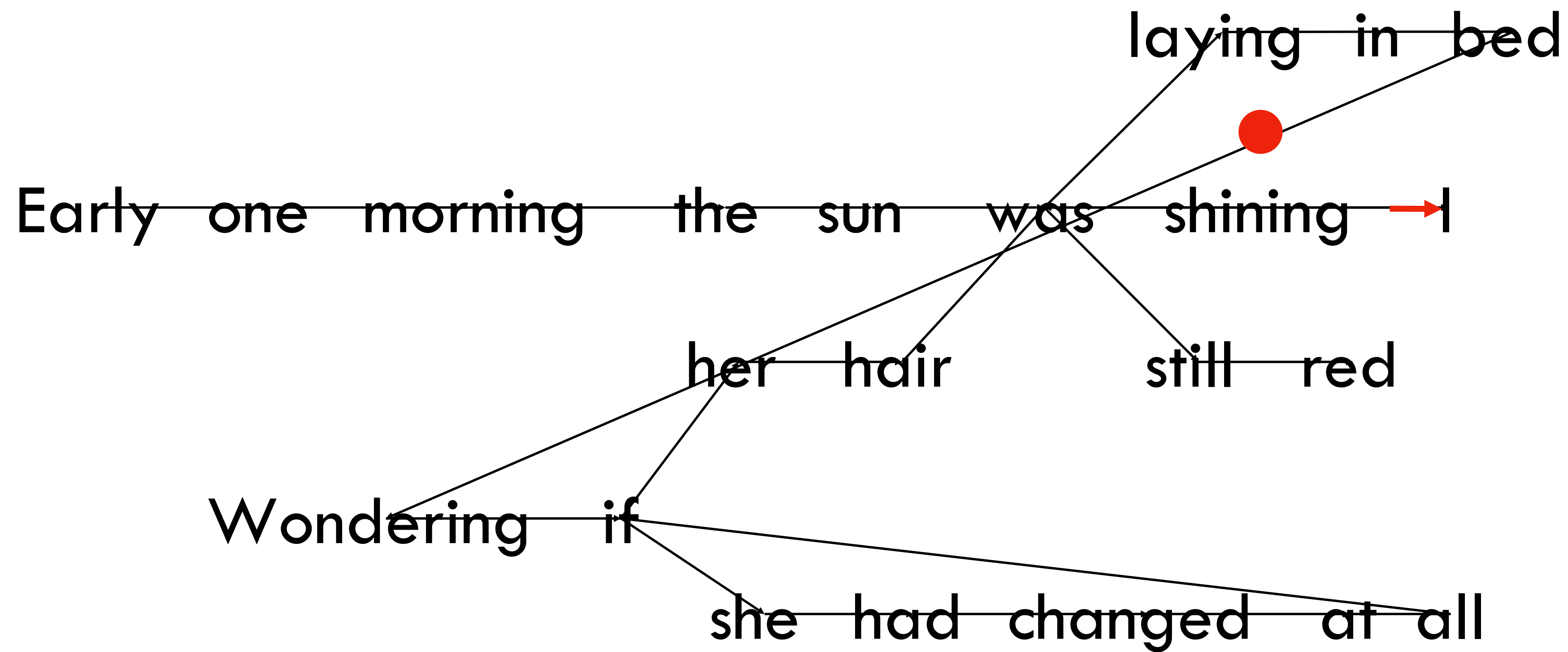
I was shining I



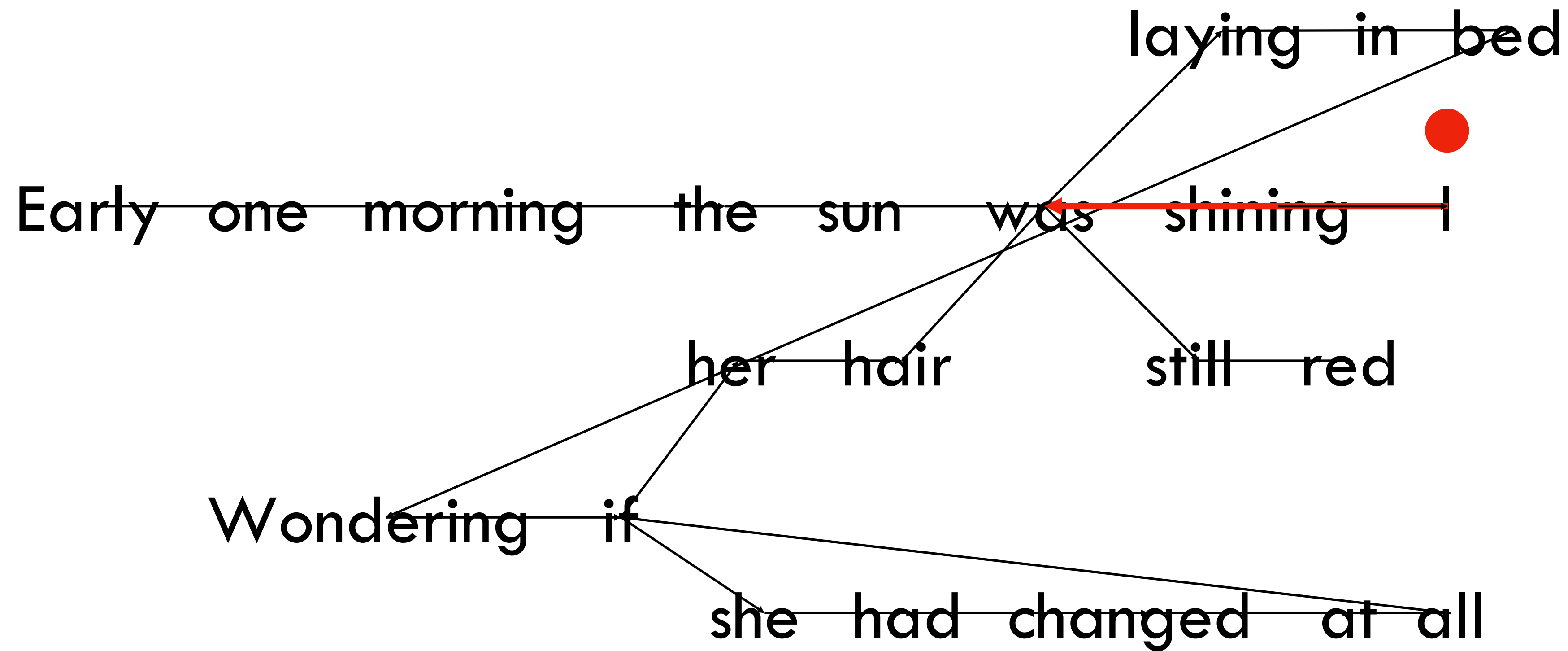
I was shining I was



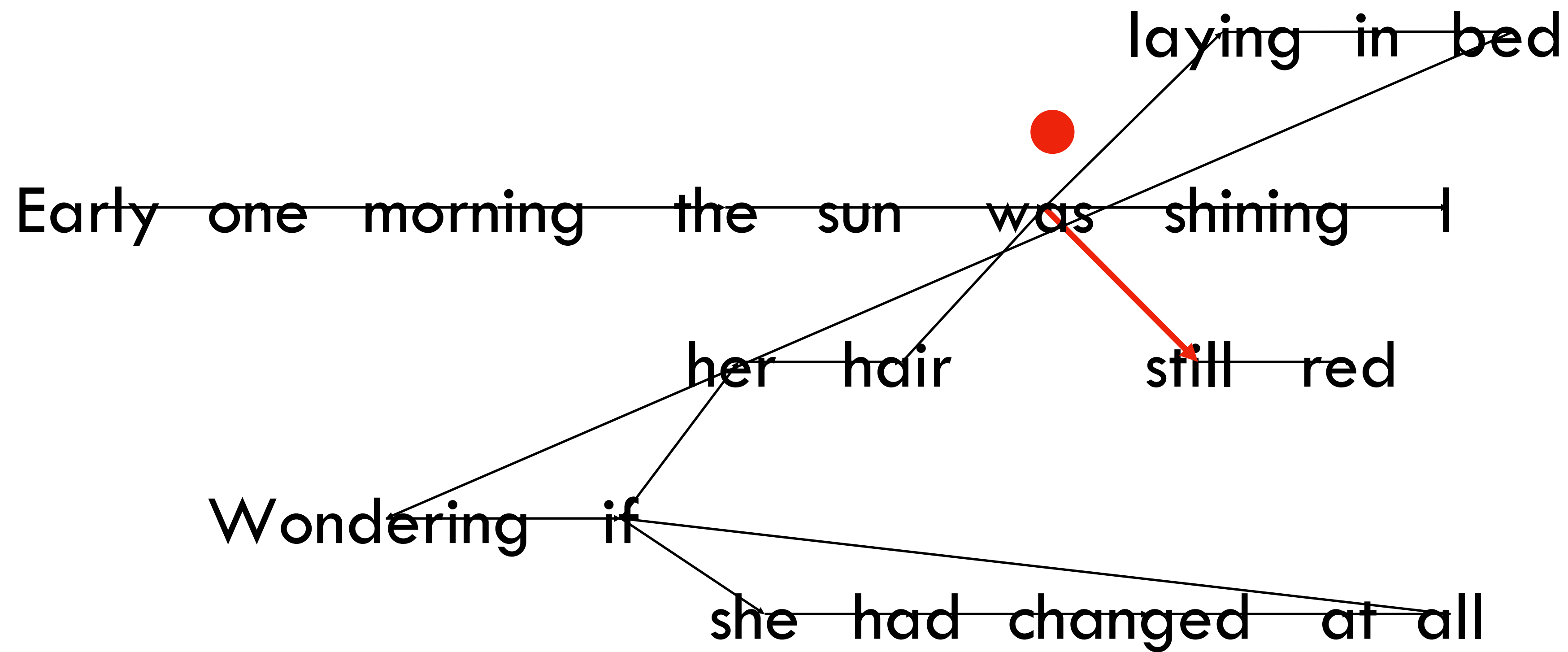
I was shining I was shining



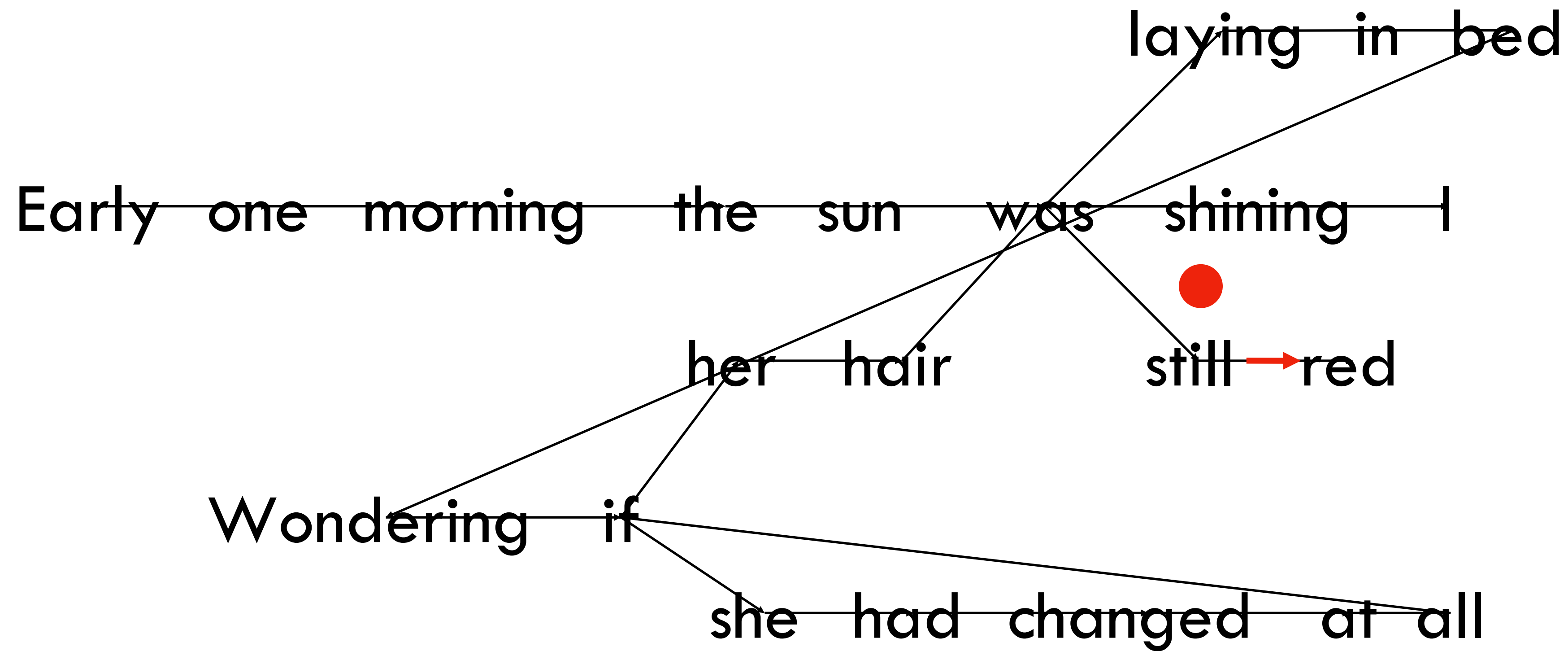
I was shining I was shining I



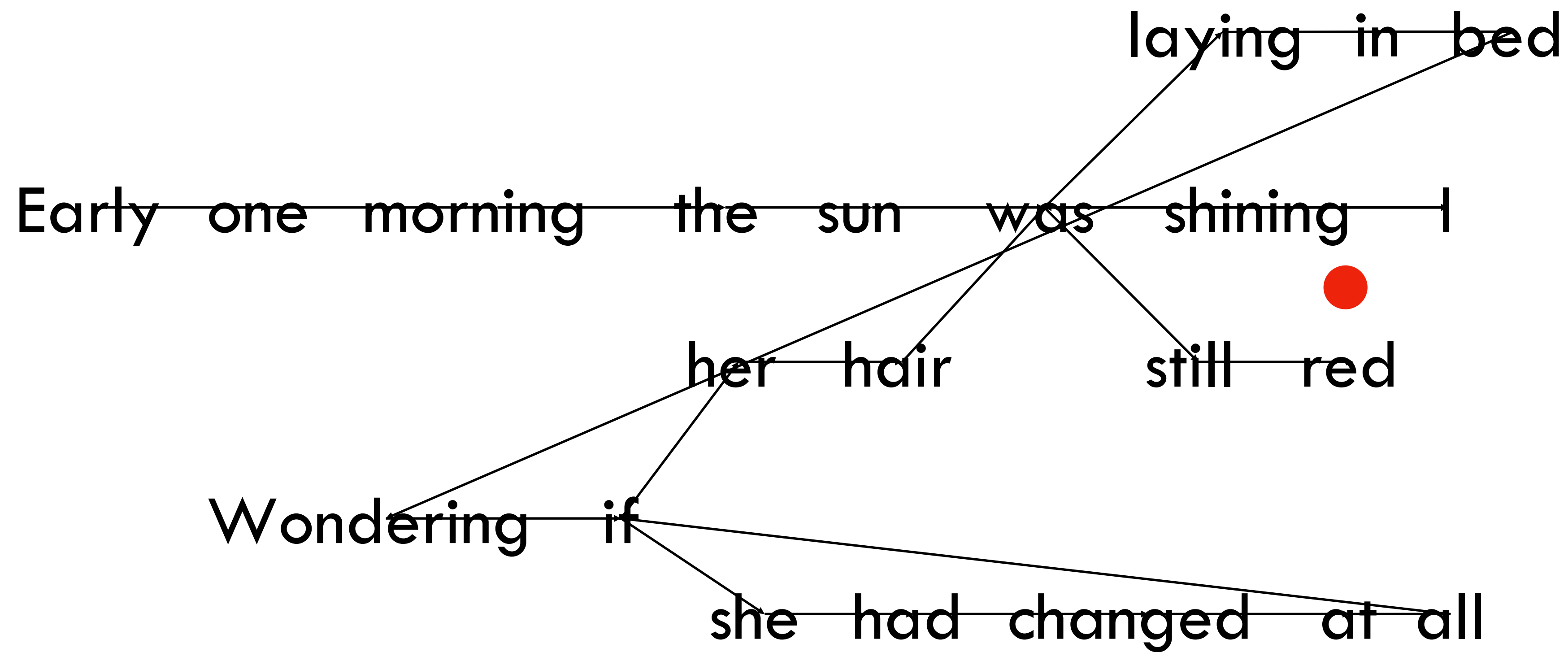
I was shining I was shining I was



I was shining I was shining I was still



I was shining I was shining I was still red



Early one morning the sun was shining
I was lying in bed
Wondering if time changed at all
If her hair was still red
Her hair they said our lives together
Sure was going to rough
They never did like Marsha's homemade dress
Papa's backhoe wasn't big enough
And I was standing on the side of the road
Rain falling on my shoes
Heading out for the East Coast
Lynn knows he just some those getting through
Tangled up in blue
She was married after we first met
Seem to be divorced
I stepped her out of a job - I guess
But I wasn't a little too much for me
We drove that car for as long as we could
Abandoned it out there
Apt up in a dark wet night
Both agreeing I was tired
She turned around to look at me
As I sat smiling away
I heard her say over my shoulder
That must again someday on the avenue
Tangled up in blue
I had a job in the great north woods
Working as a cook for a spell
But I never did that it all that much
And one day she just left
So I drifted down to New Orleans
Where I happened to be employed
Working for a while in a fishing boat
Right outside of Delcambre
But all the while I was alone
The past was closed behind
I seen a lot of women
But she never escaped my mind and I just grew
Tangled up in blue
She was working in a laptop place
And I stopped in for a beer
I just had looking at the side of her face
In the sunlight so clear
And later on as the crowd formed out
I was just about to do the same
She was standing there in back of my chair
Said to me Girl I know your name
I murmured something underneath my breath
She looked the lines on my face
I must admit I felt a little shrewy
When she bent down to tie the laces of my shoe
Tangled up in blue
She is a burner on the stove
And offered me a pipe
I thought you'd hear say baby she said
You look like the silent type
Then she opened up a book of poems
And handed it to me
Written by an Italian poet
From the fourteenth century
And every one of them sounds long true
And glossed the burning coal
Pouring off of every page
Like it was written in my soul from me to you
Tangled up in blue
I lived with them on Marquette Street
In a basement down the stairs
There was music in the walls at night
And revolution in the air
Then he started the dealing with power
And something inside of him said
She had to sell everything she owned
And broke up inside
And when finally the bottom fell out
I became withdrawn
The only thing I know how to do
Was to stay on top of my mind and I just grew
Tangled up in blue
So now I'm going back again
I got to get to her somehow
All the people are used to know
They're go down to me now
Some are professional
Some are independent wives
Don't know how it all got started
I don't know what they're doing with their lives
But me I'm still on the road
Heading for another point
We always did love the drive
We just take it from a different point of view
Tangled up in blue

she was standing on the side of my mind

...

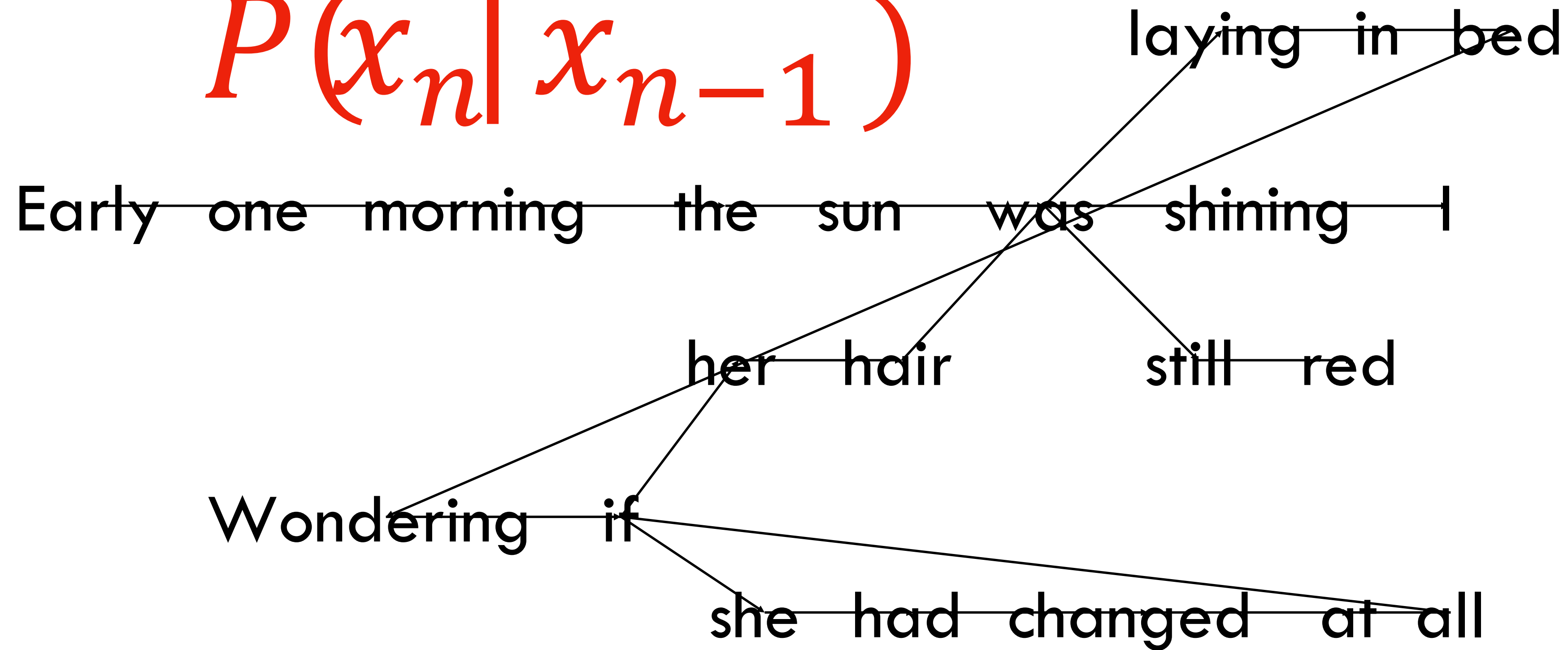
side of my shoes heading out of my face

...

one of my chair said our lives together

...

$$P(x_n | x_{n-1})$$



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

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

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

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

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

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

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

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

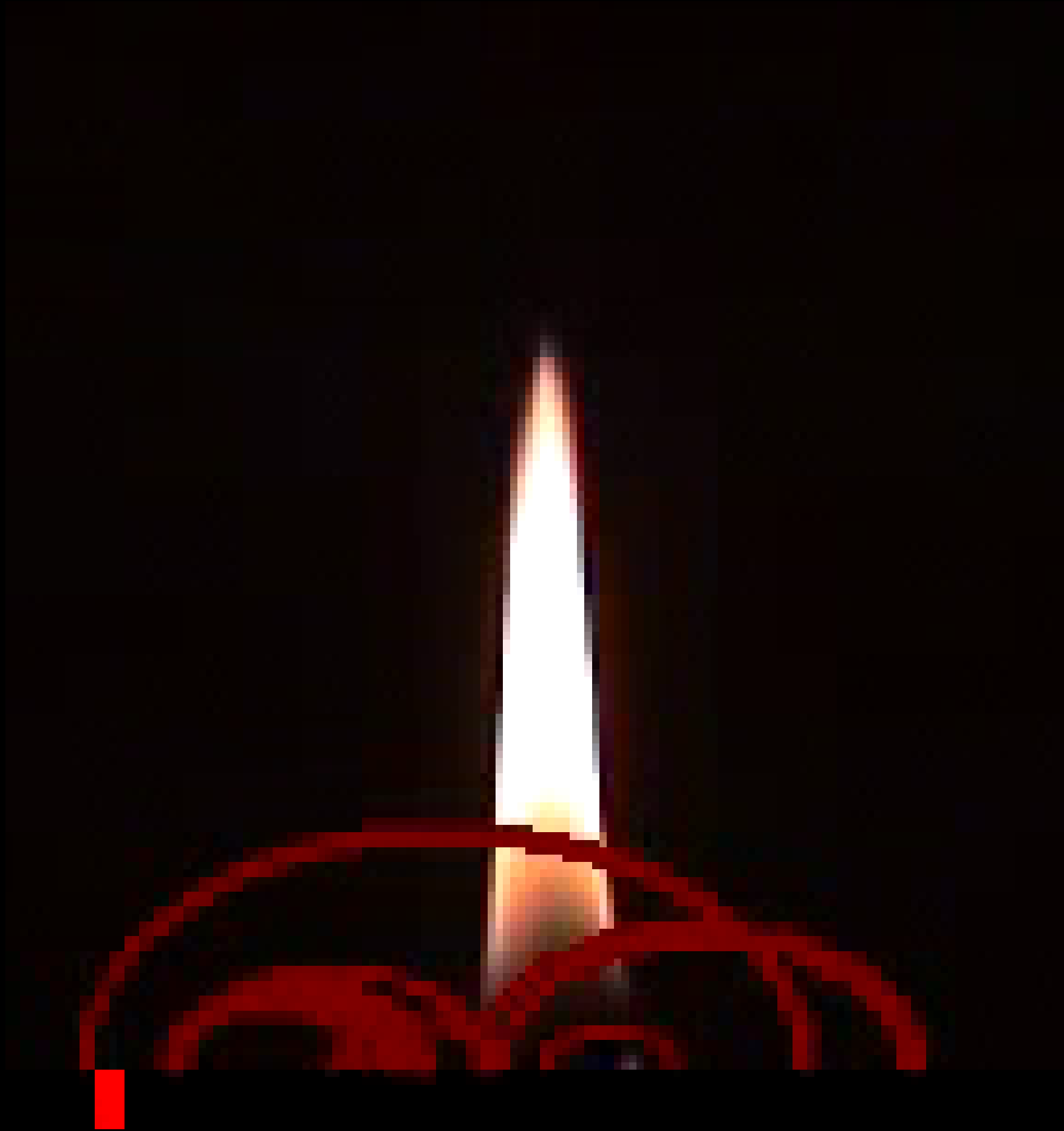


video clip



video texture

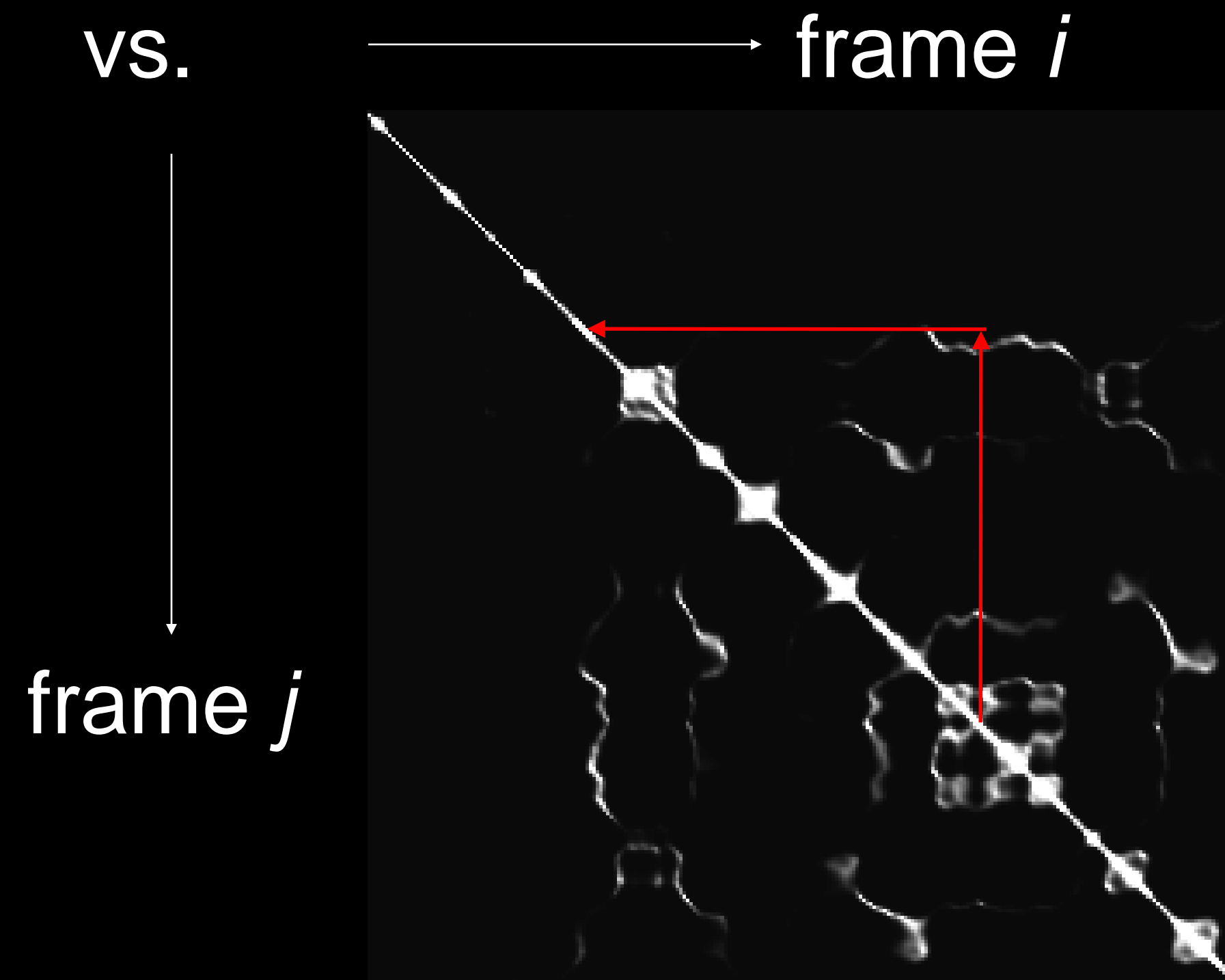
Our approach



- How do we find good transitions?

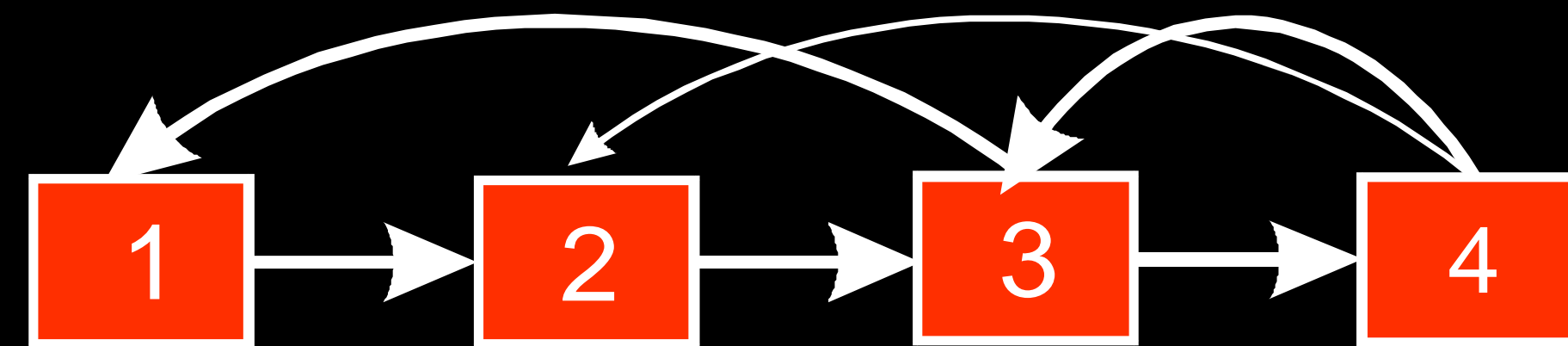
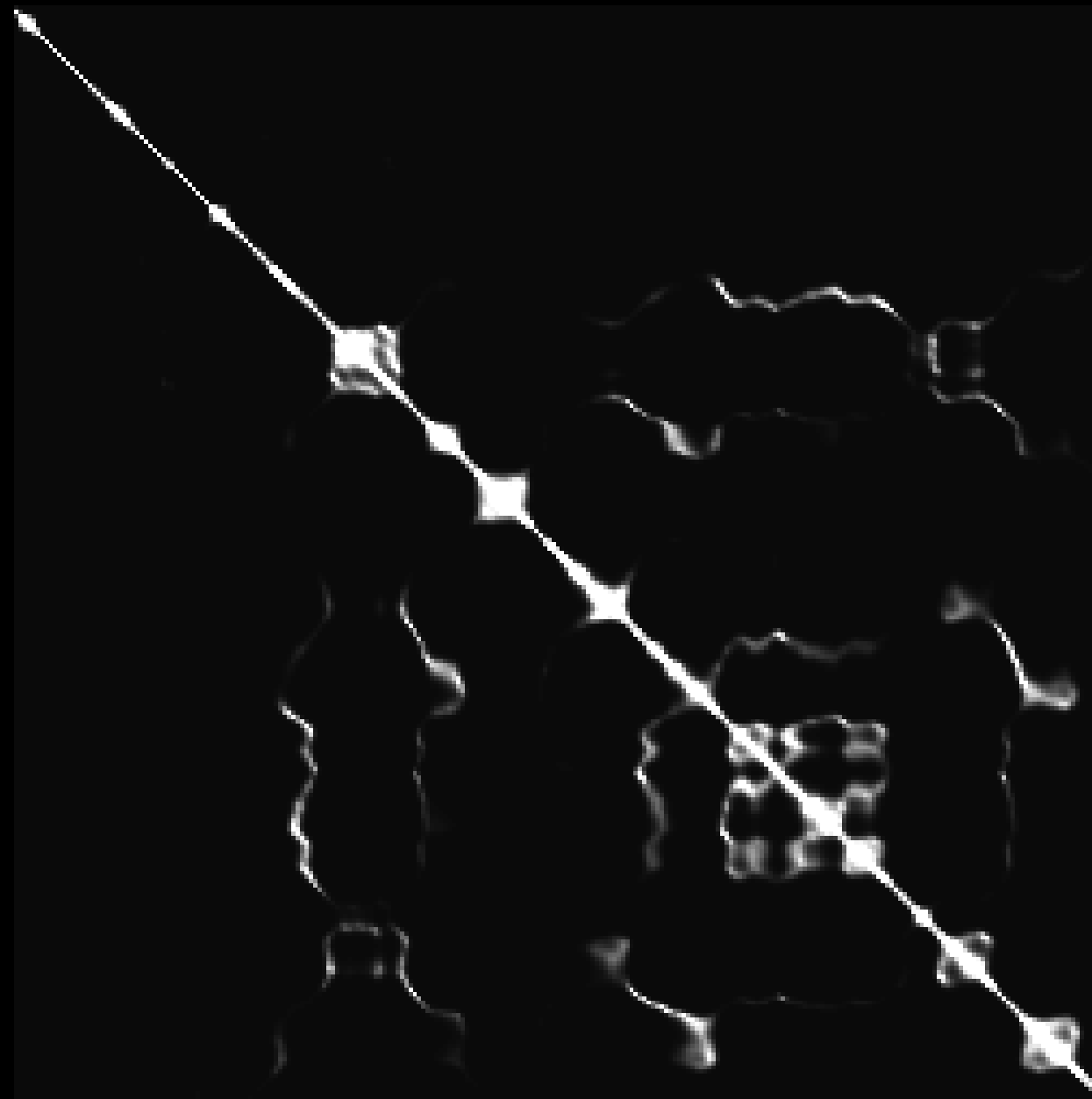
Finding good transitions

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



Similar frames make good transitions

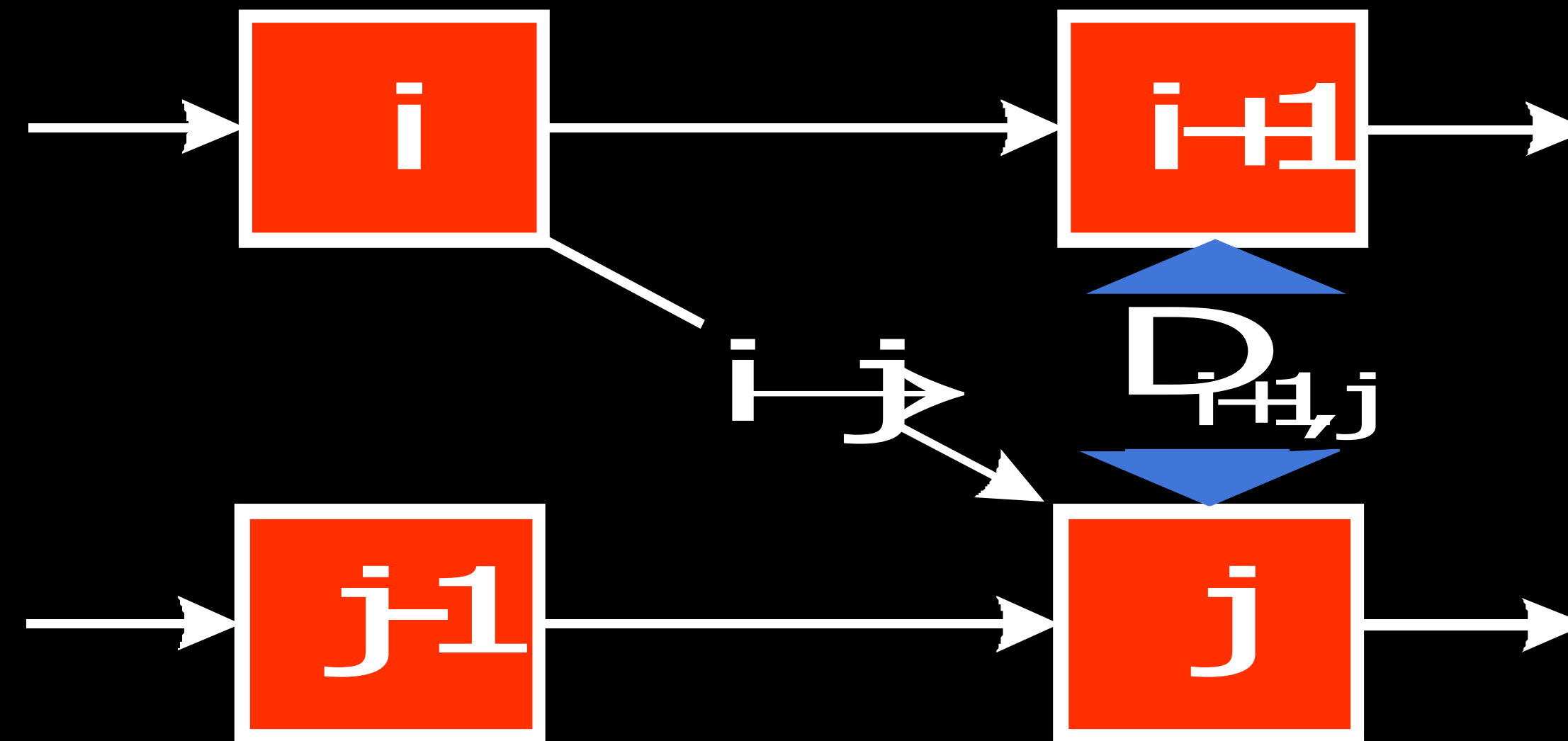
Markov chain representation



Similar frames make good transitions

Transition costs

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



Transition probabilities

- Probability for transition $P_{i \rightarrow j}$ inversely related to cost:

- $P_{i \rightarrow j} \sim \exp (- C_{i \rightarrow j} / \sigma^2)$



high σ

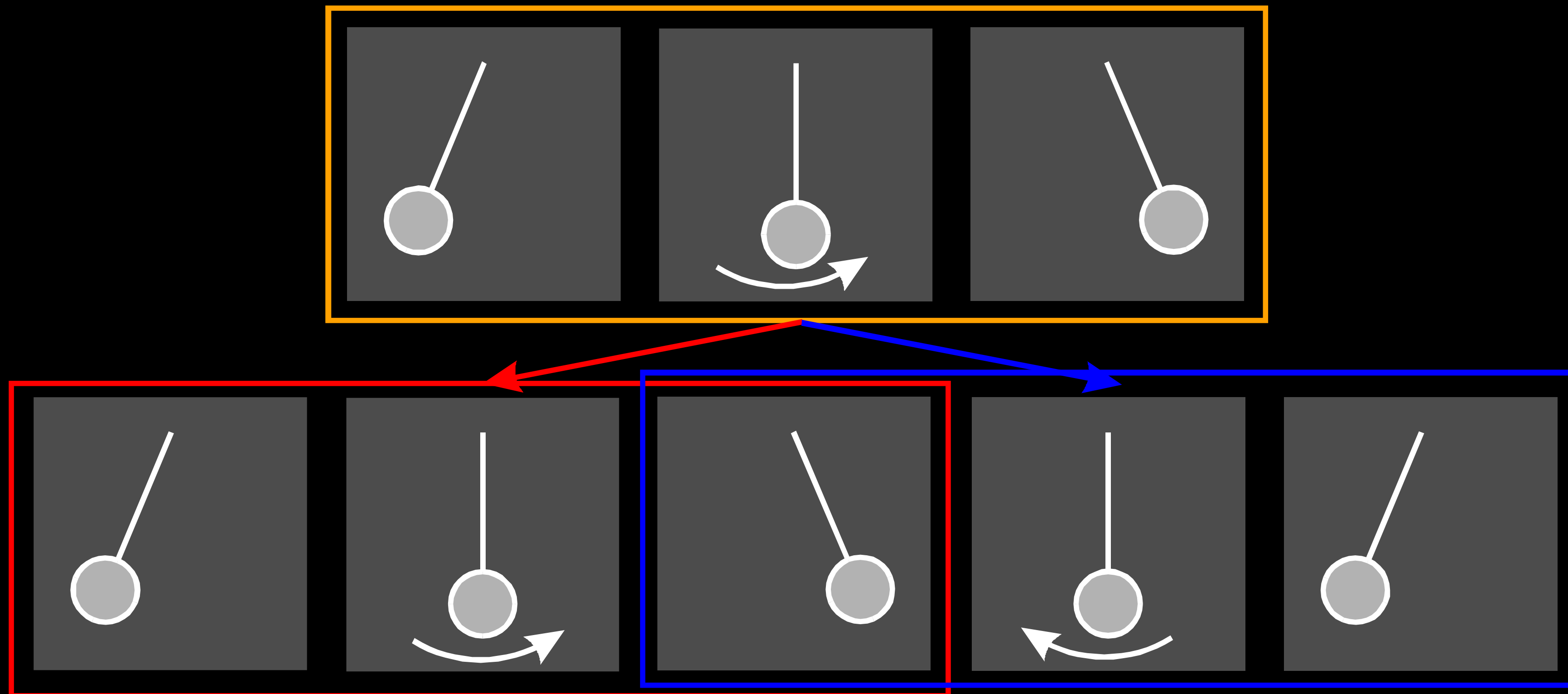


low σ

Preserving dynamics



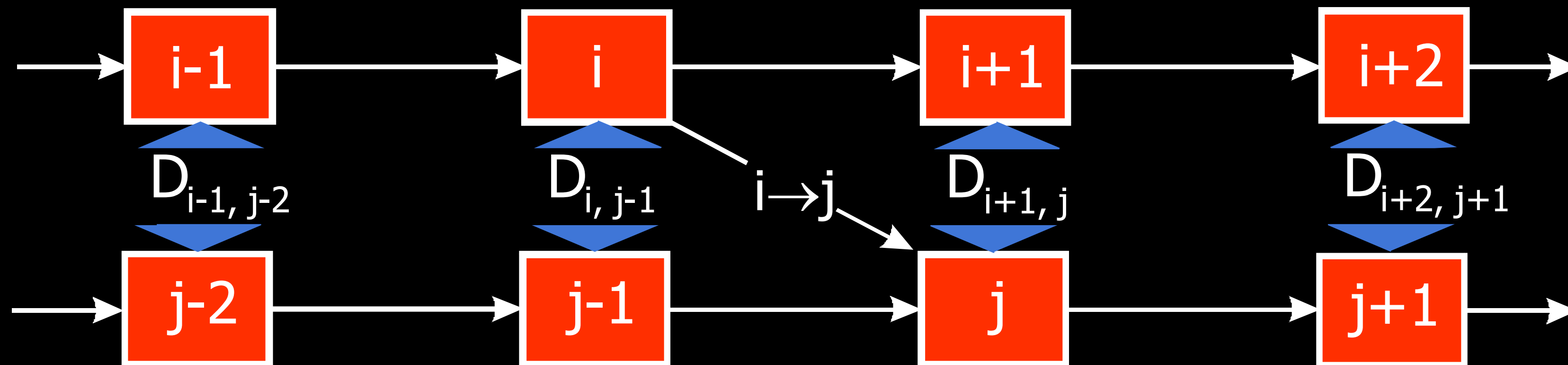
Preserving dynamics



Preserving dynamics

- Cost for transition $i \rightarrow j$

- $$C_{i \rightarrow j} = \sum_{k=-N}^{N-1} w_k D_{i+k+1, j+k}$$



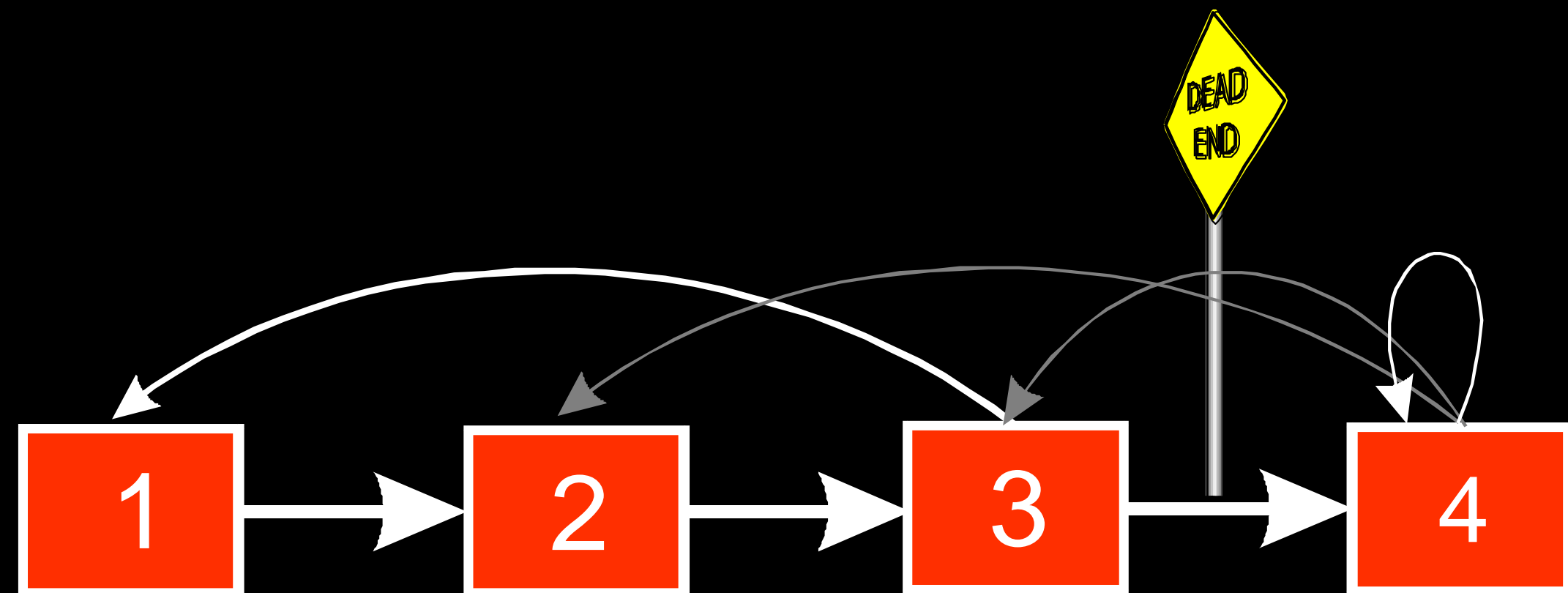
Preserving dynamics – effect

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



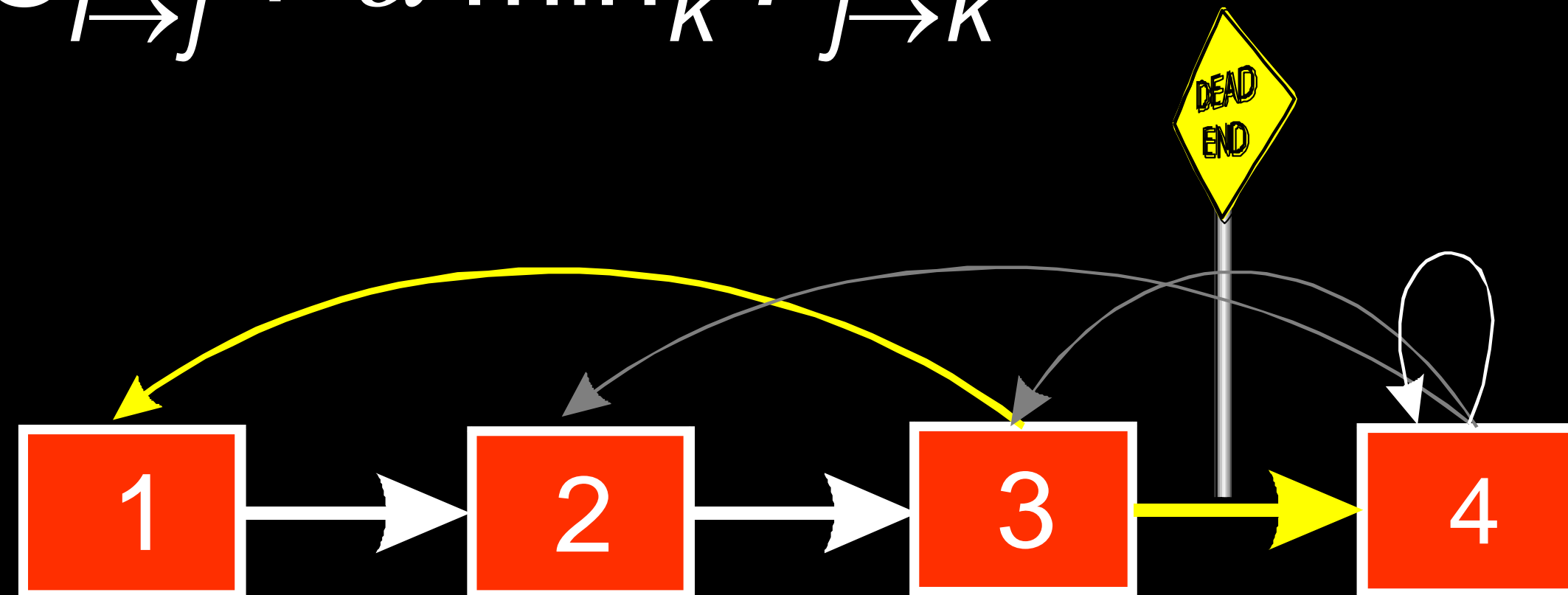
Dead ends

- No good transition at the end of sequence



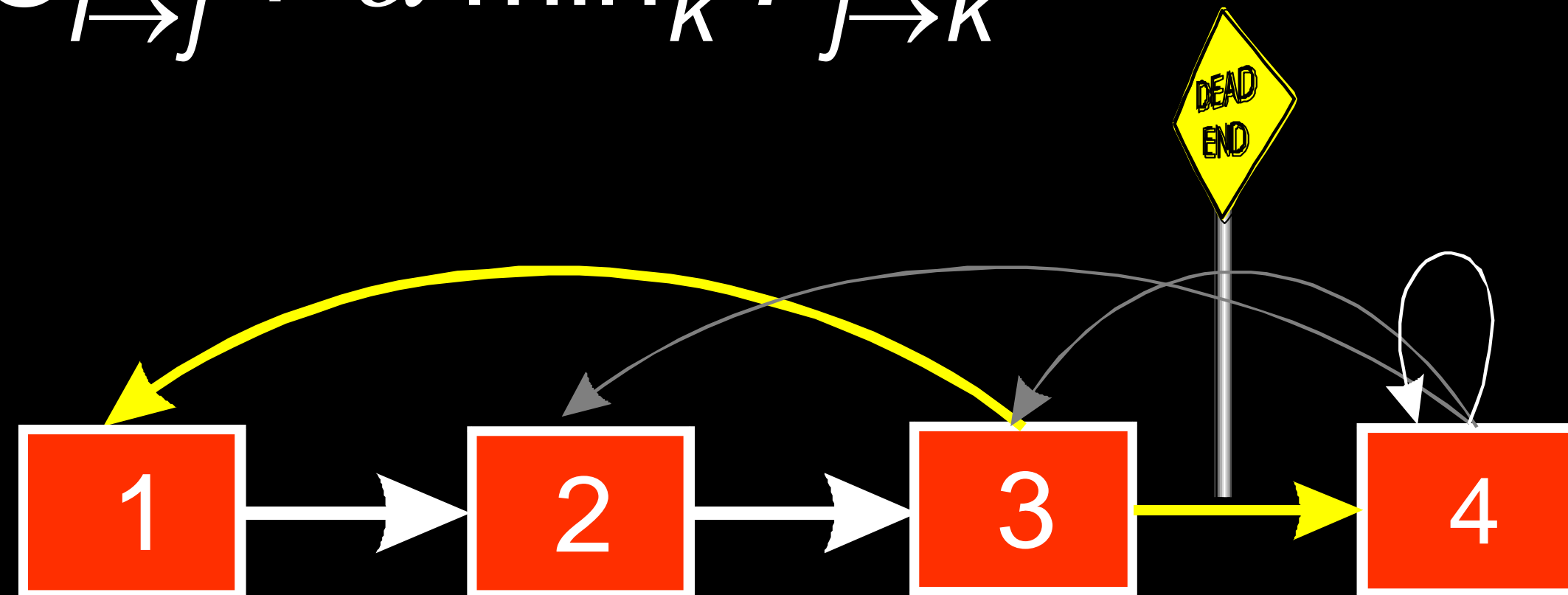
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost
 - $F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k}$



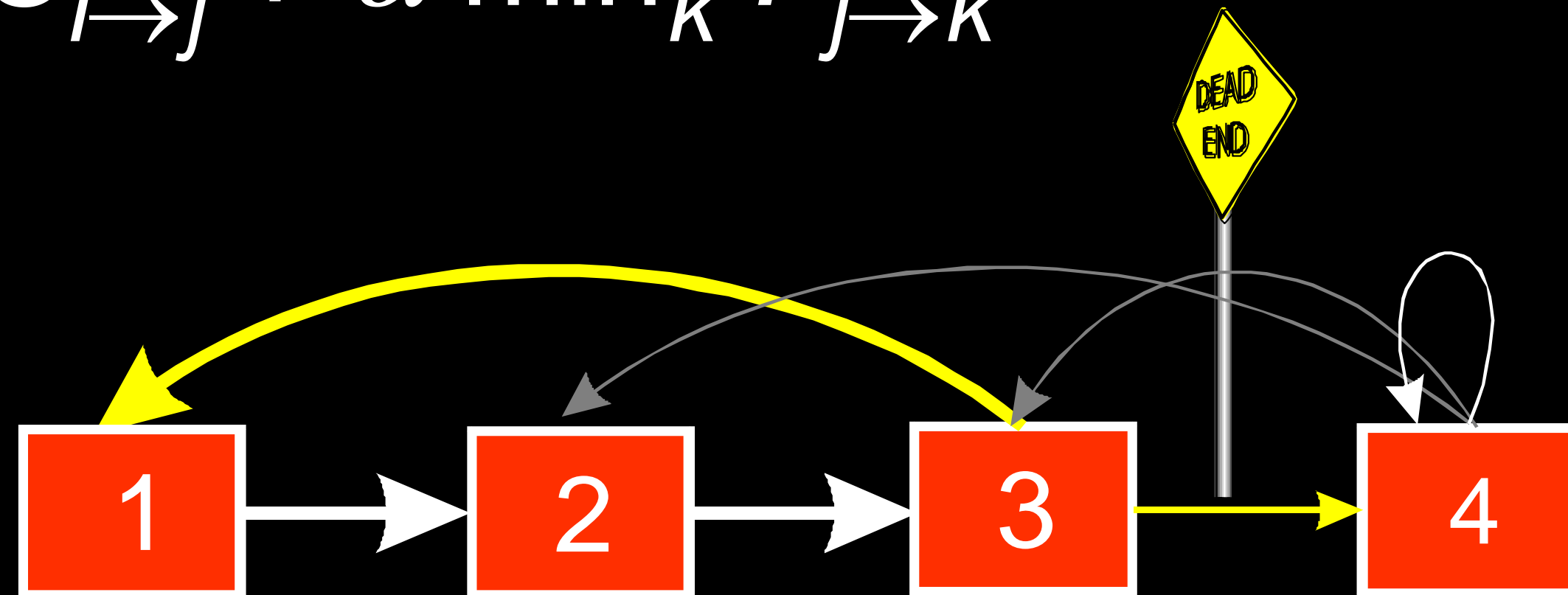
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost
 - $F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k}$



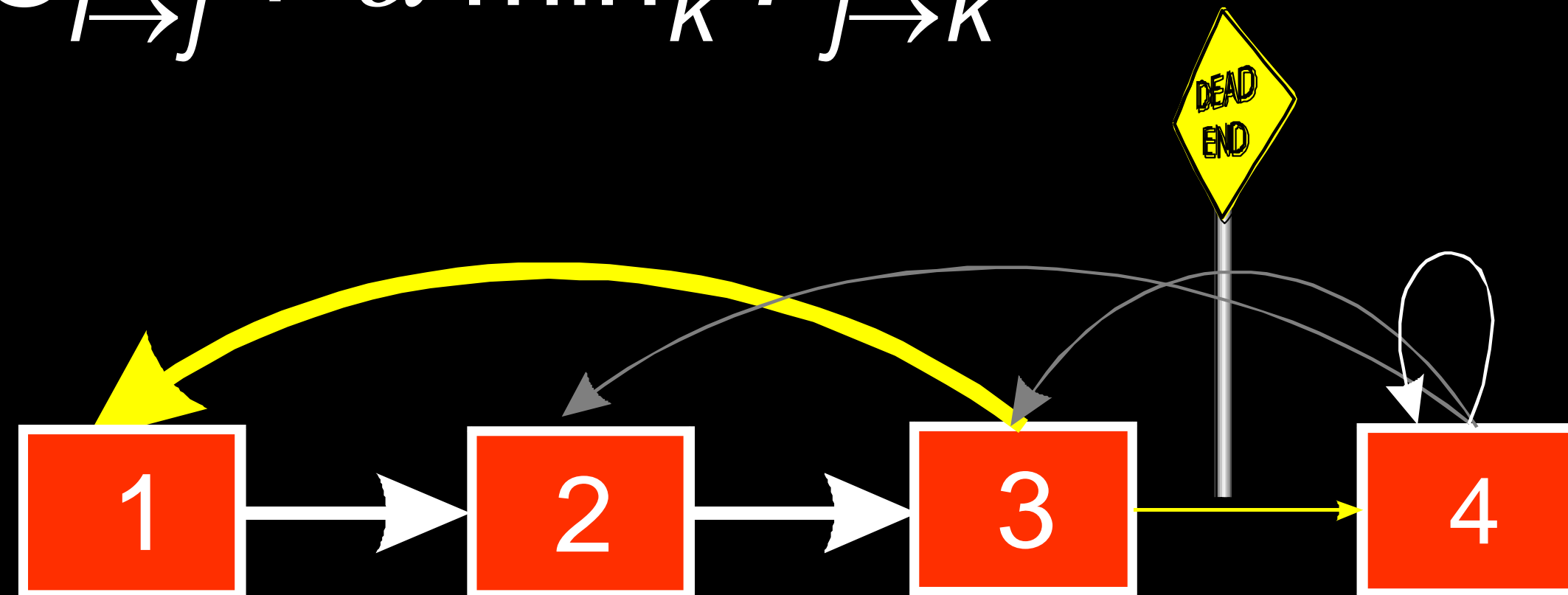
Future cost

- Propagate future transition costs backward
- Iteratively compute new cost
 - $F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k}$



Future cost

- Propagate future transition costs backward
- Iteratively compute new cost
 - $F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k}$

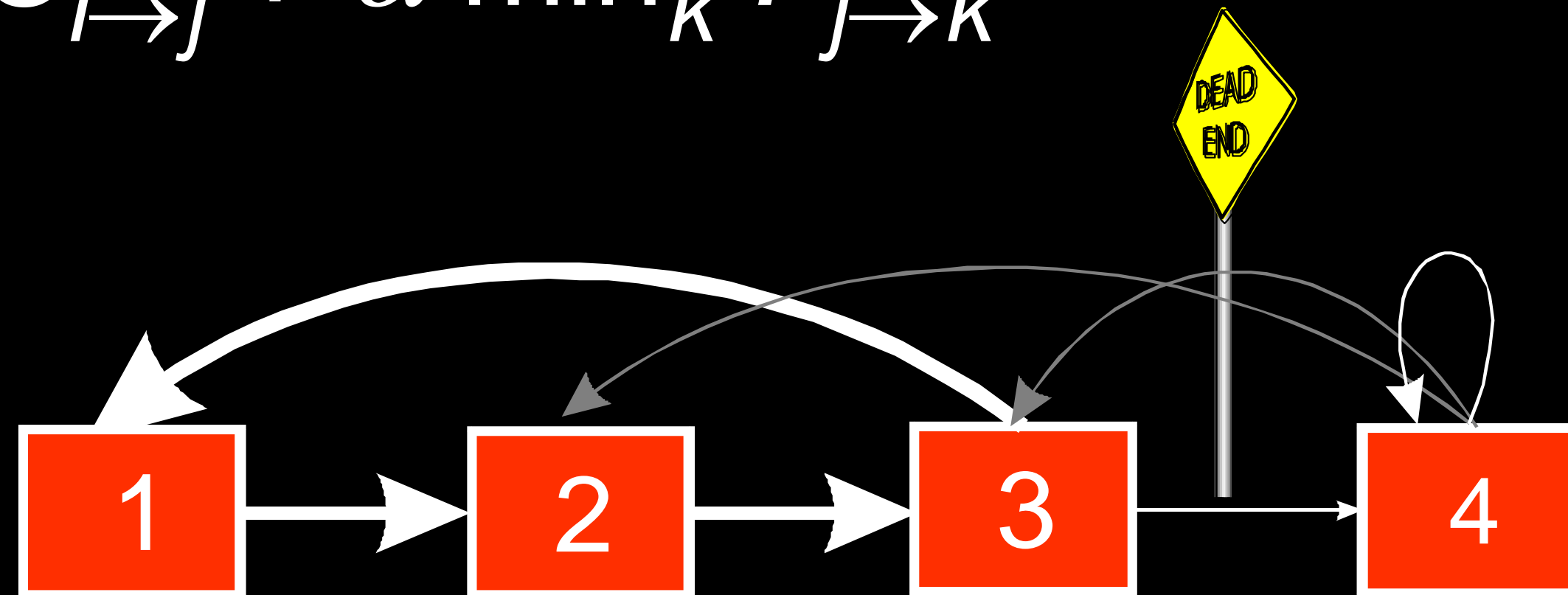


Future cost

- Propagate future transition costs backward
- Iteratively compute new cost

- $F_{i \rightarrow j} = C_{i \rightarrow j} + \alpha \min_k F_{j \rightarrow k}$

- Q-learning



Final result



Finding good loops

- Alternative to random transitions
- Precompute set of loops up front



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



slow



variable



fast

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

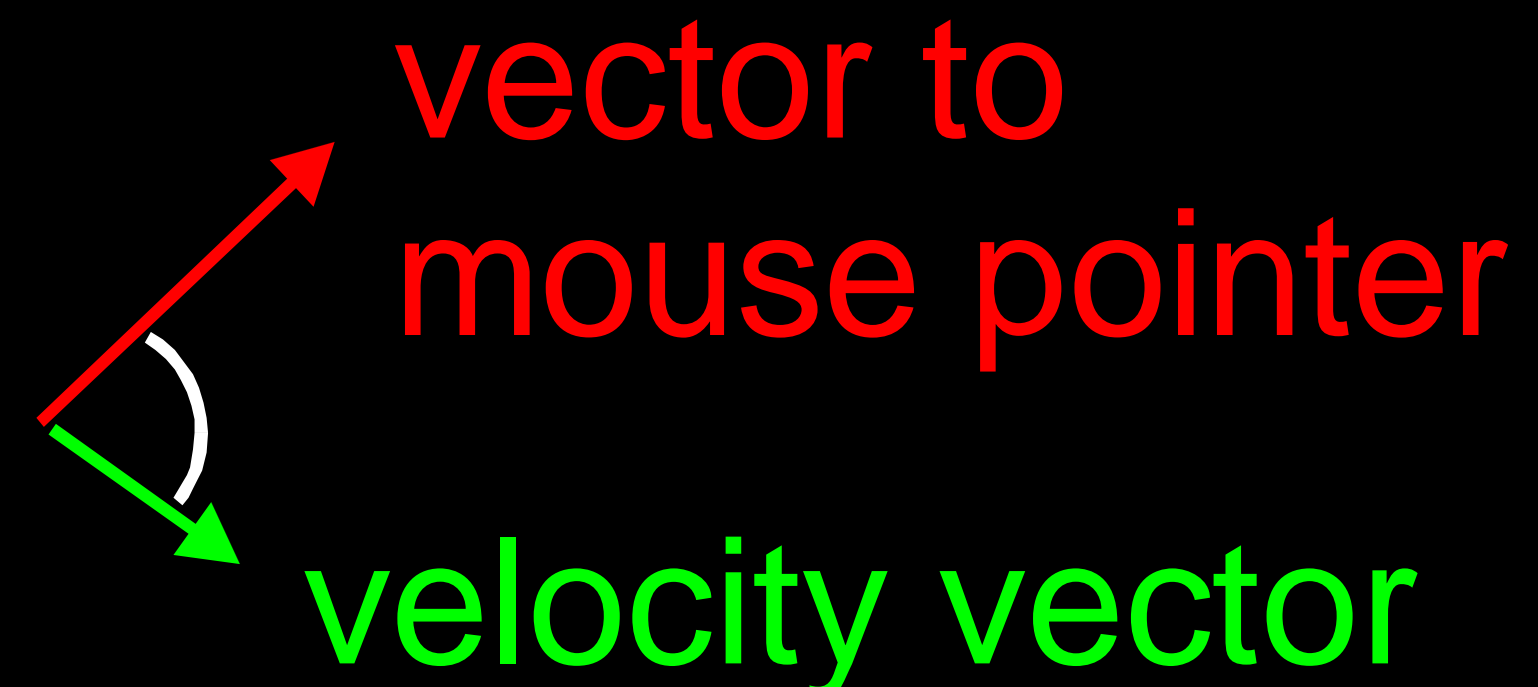


blue screen matting
and velocity estimation



Video sprite control

- Augmented transition cost:

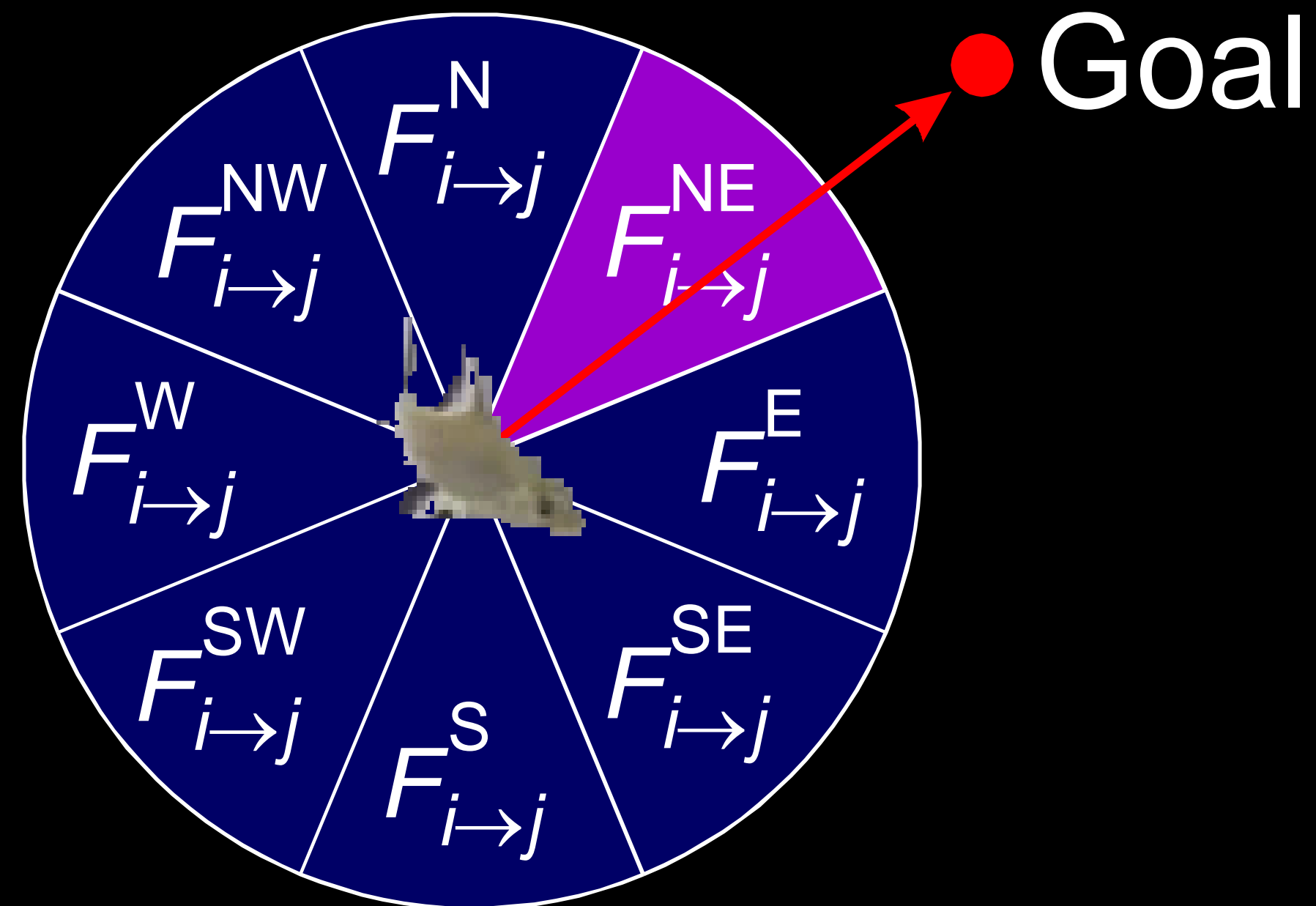
$$C_{i \rightarrow j}^{\text{Animation}} = \alpha \underbrace{C_{i \rightarrow j}}_{\text{Similarity term}} + \beta \underbrace{\text{angle}}_{\text{Control term}}$$


vector to mouse pointer

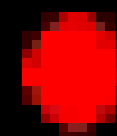
velocity vector

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



Texture

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



radishes



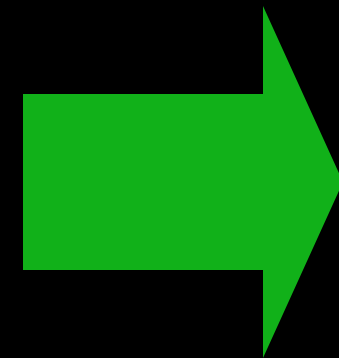
rocks



yogurt

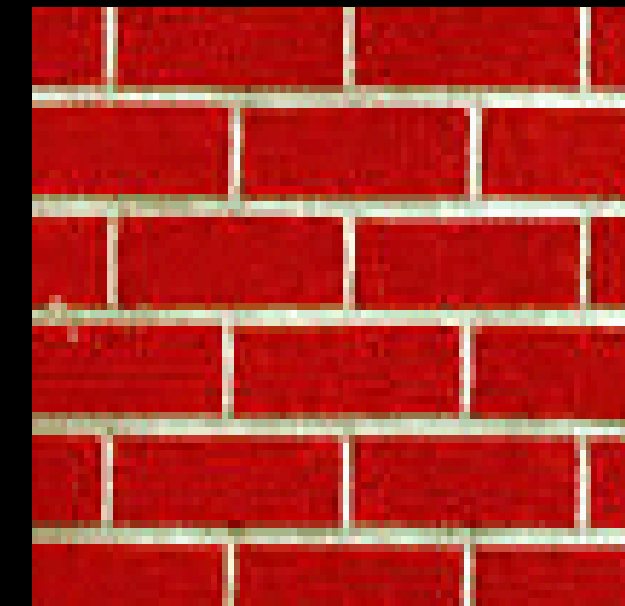
Texture Synthesis

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



The Challenge

- Need to model the whole spectrum: from repeated to stochastic texture



repeated

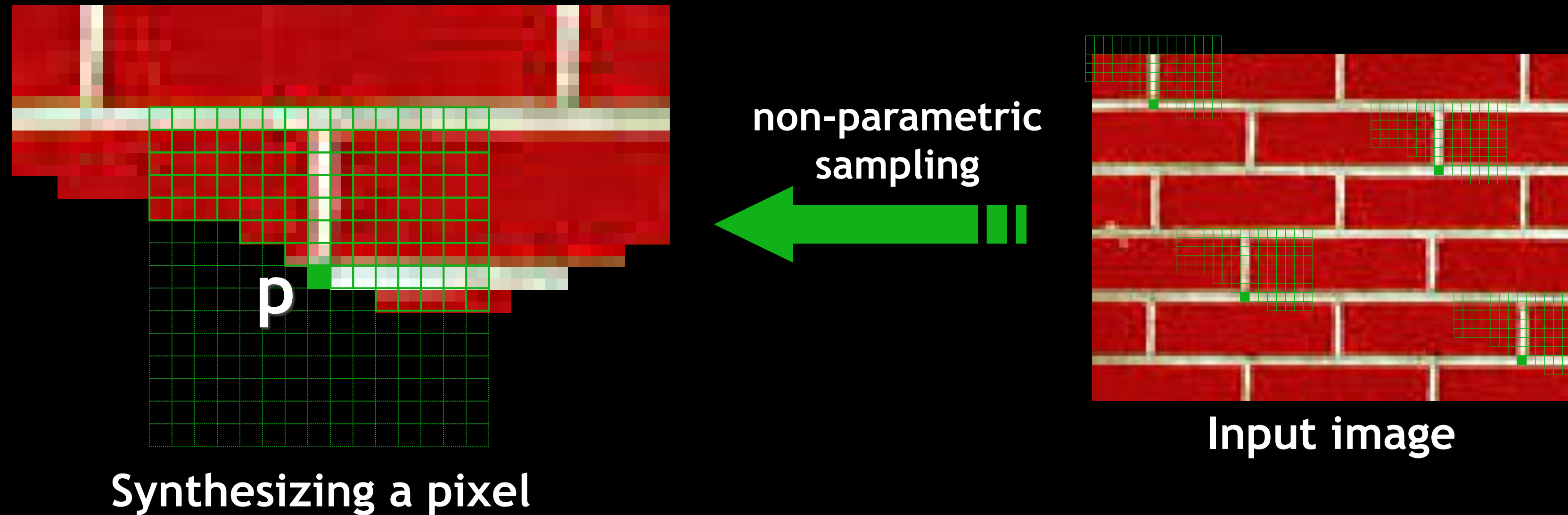


stochastic



Both?

Efros & Leung Algorithm (ICCV 1999)

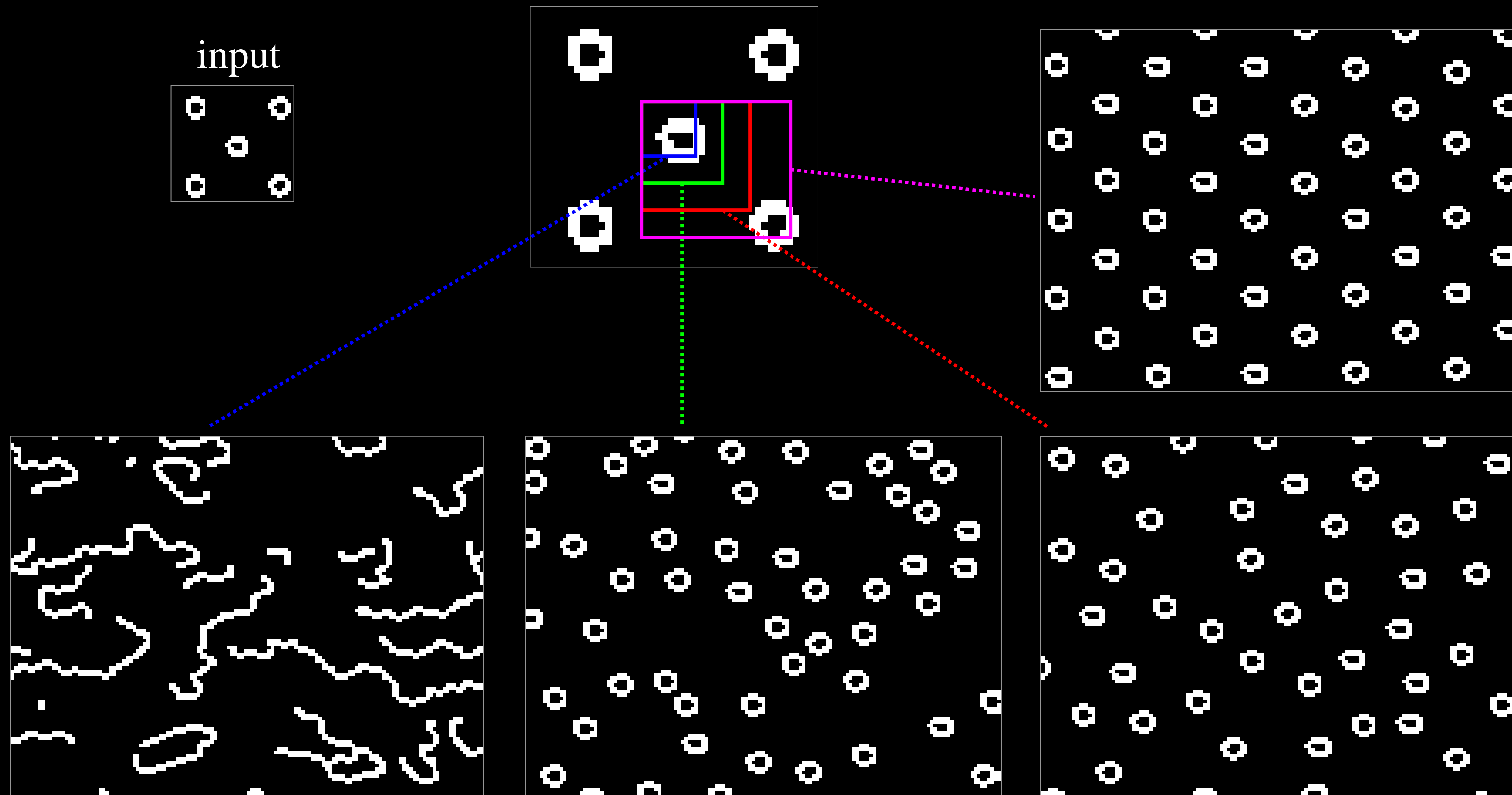


- Assuming Markov property, compute $P(\mathbf{p}|\mathbf{N}(\mathbf{p}))$
 - Building explicit probability tables infeasible
 - Instead, we *search the input image* for all similar neighborhoods — that's our pdf for \mathbf{p}
 - To sample from this pdf, just pick one match at random

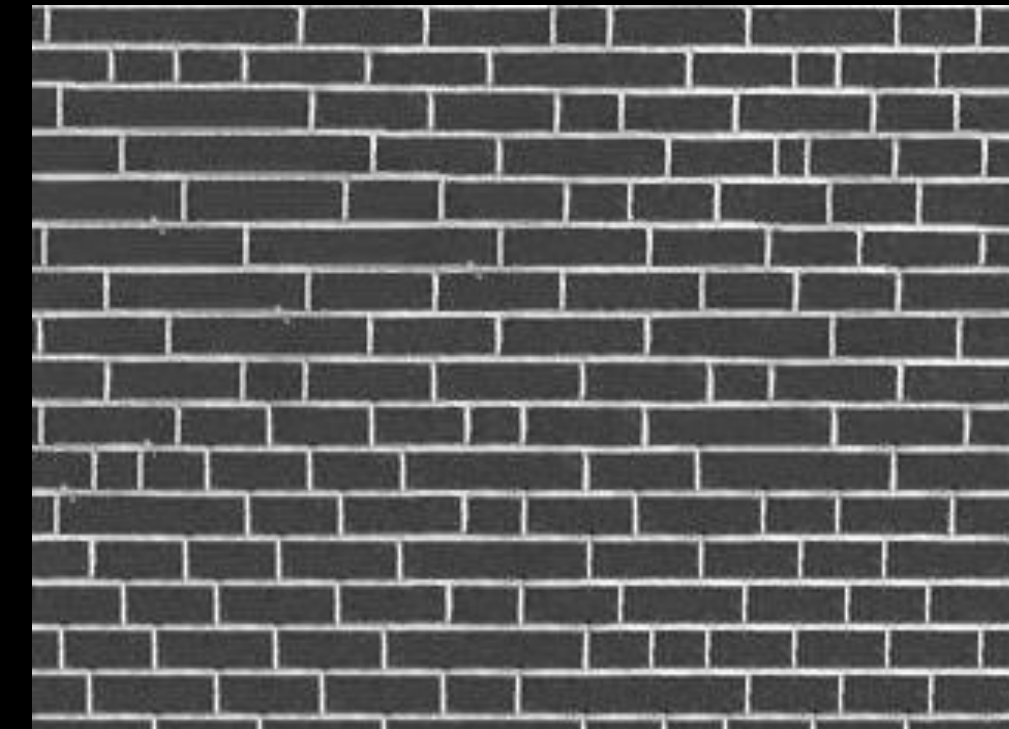
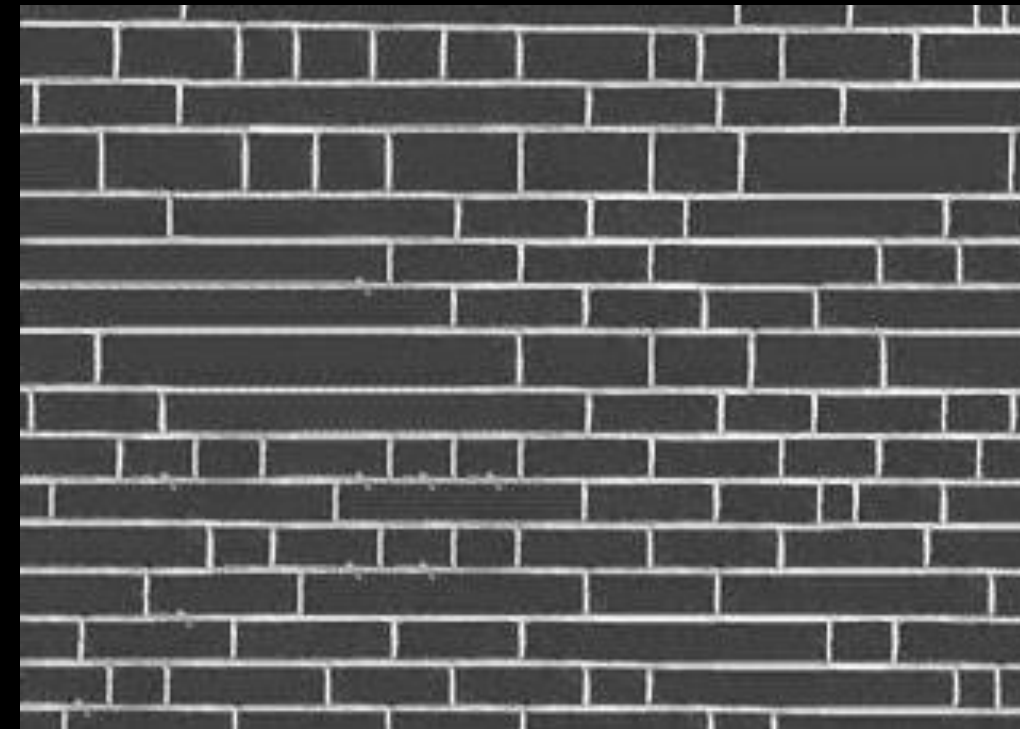
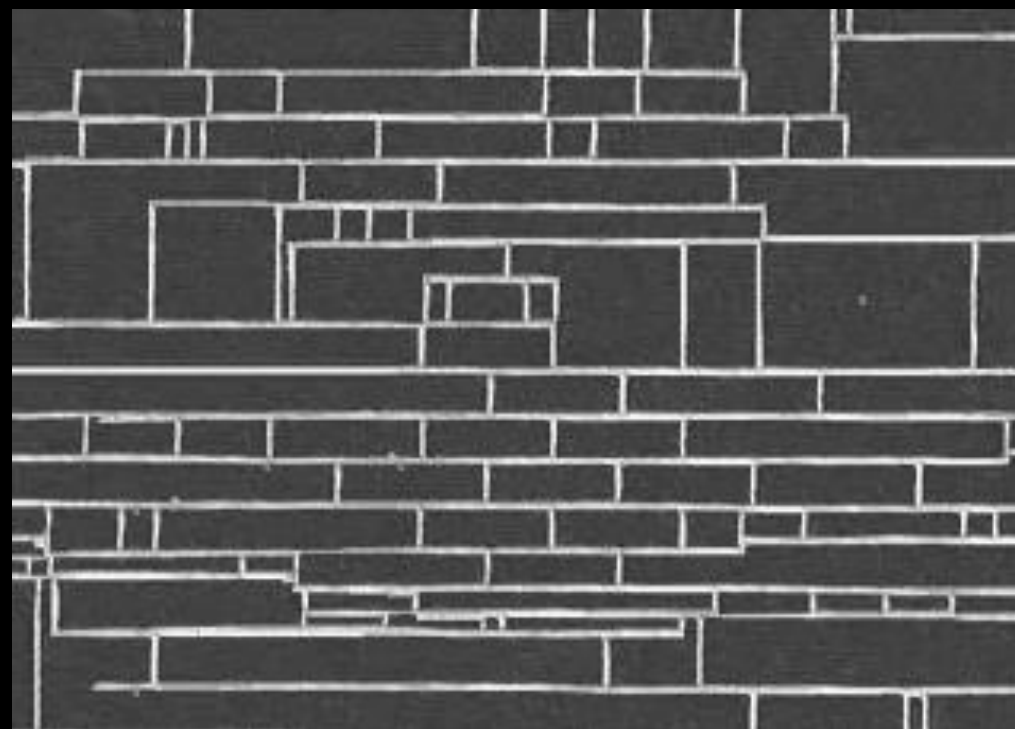
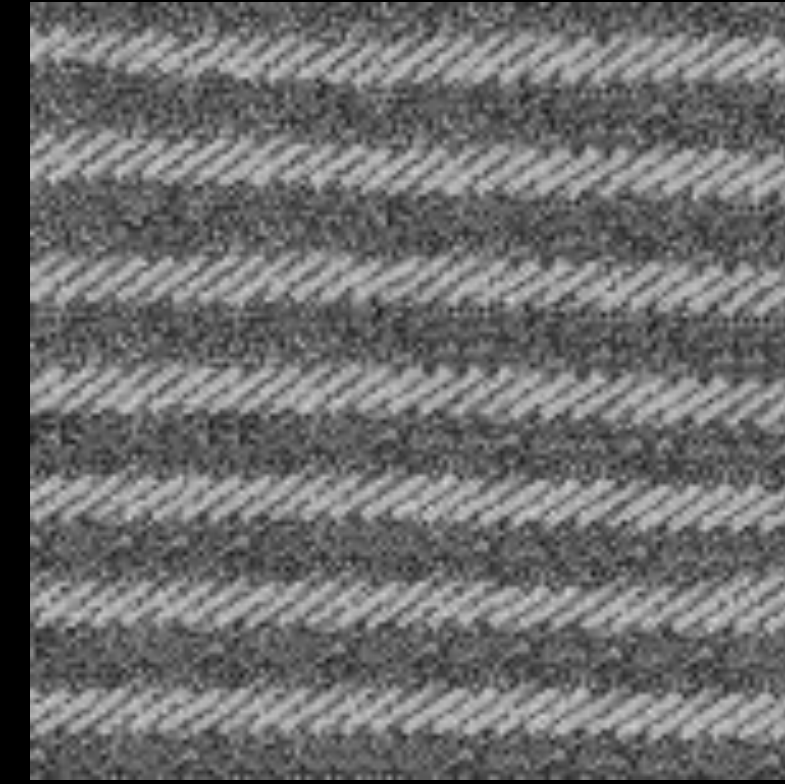
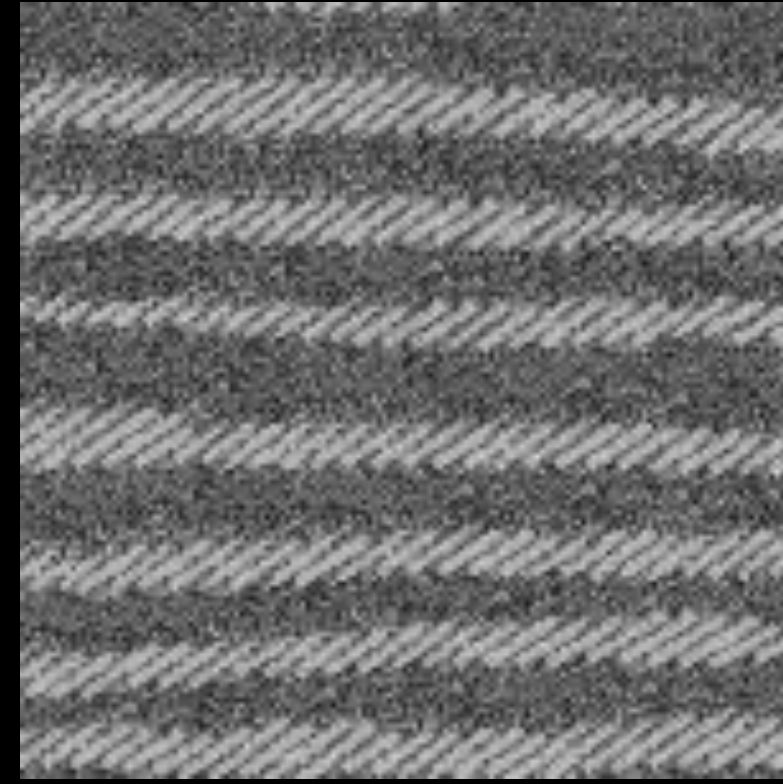
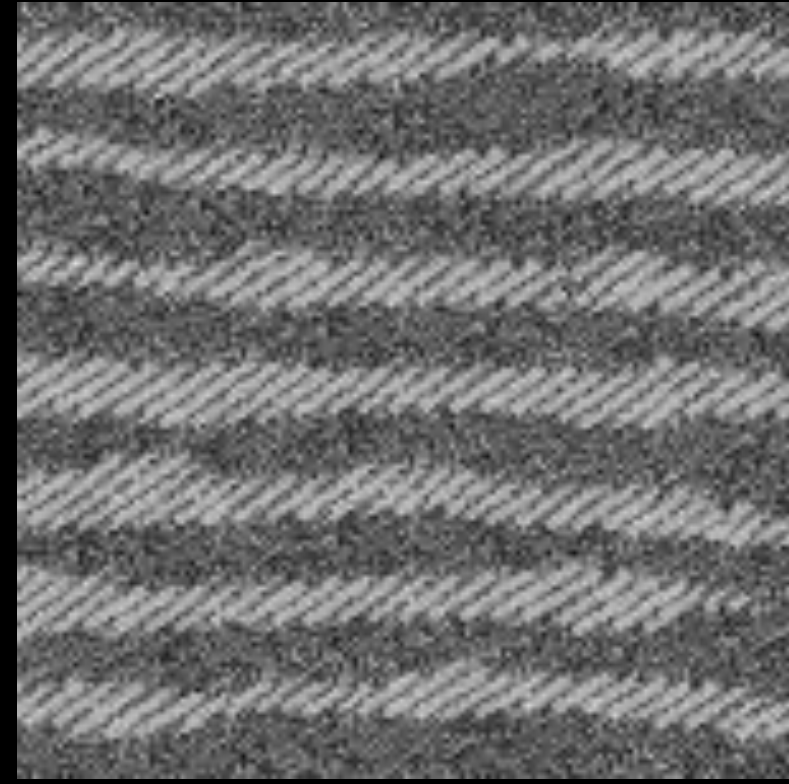
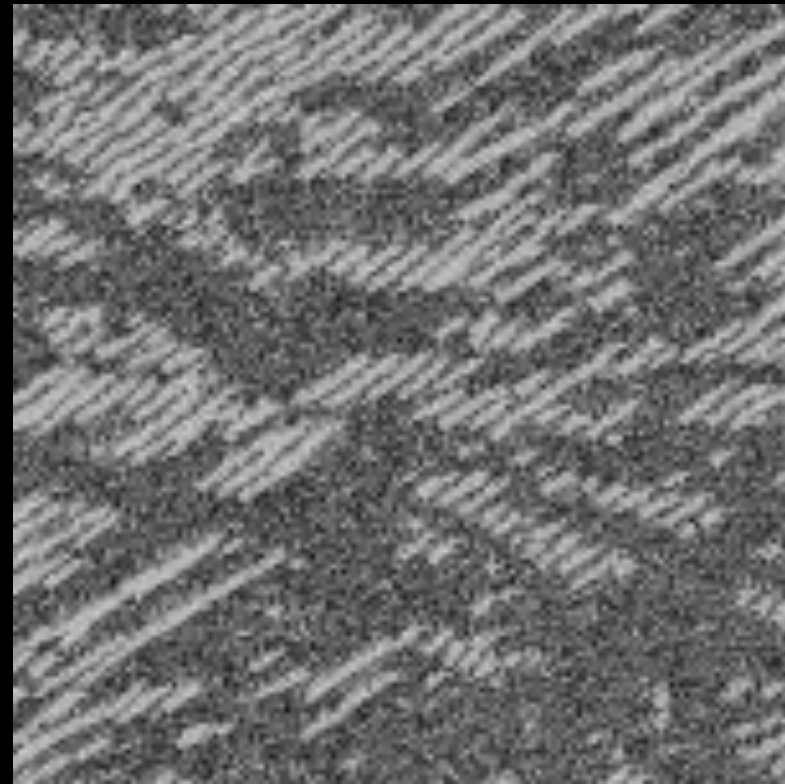
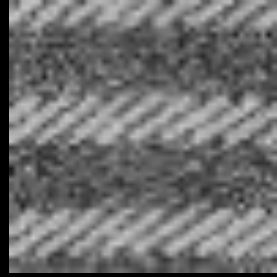
Some Details

- 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

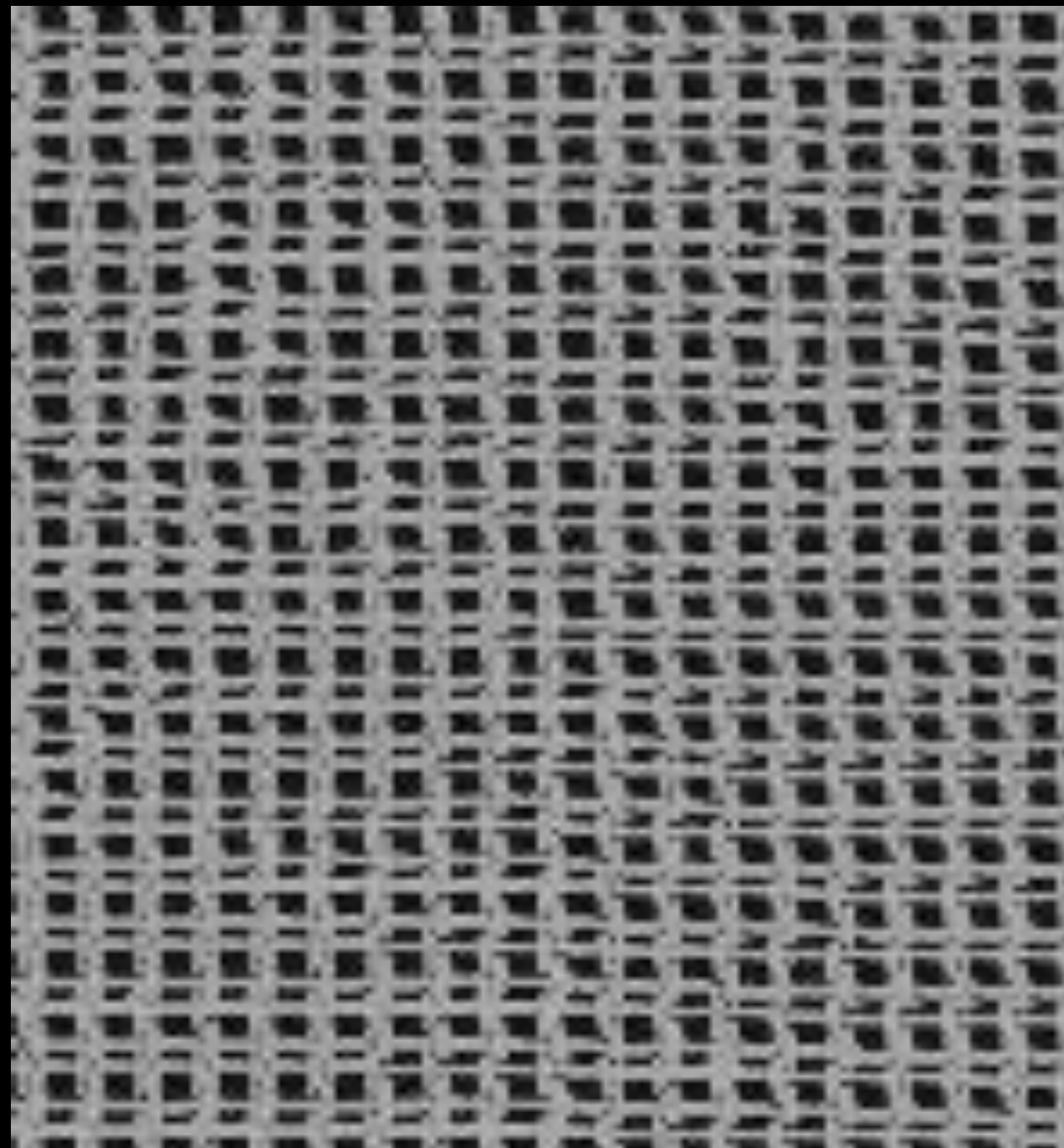
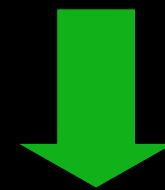
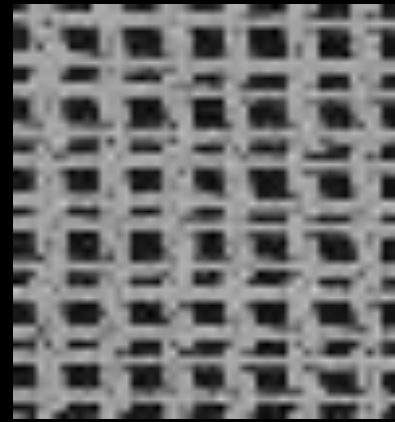


Increasing window size

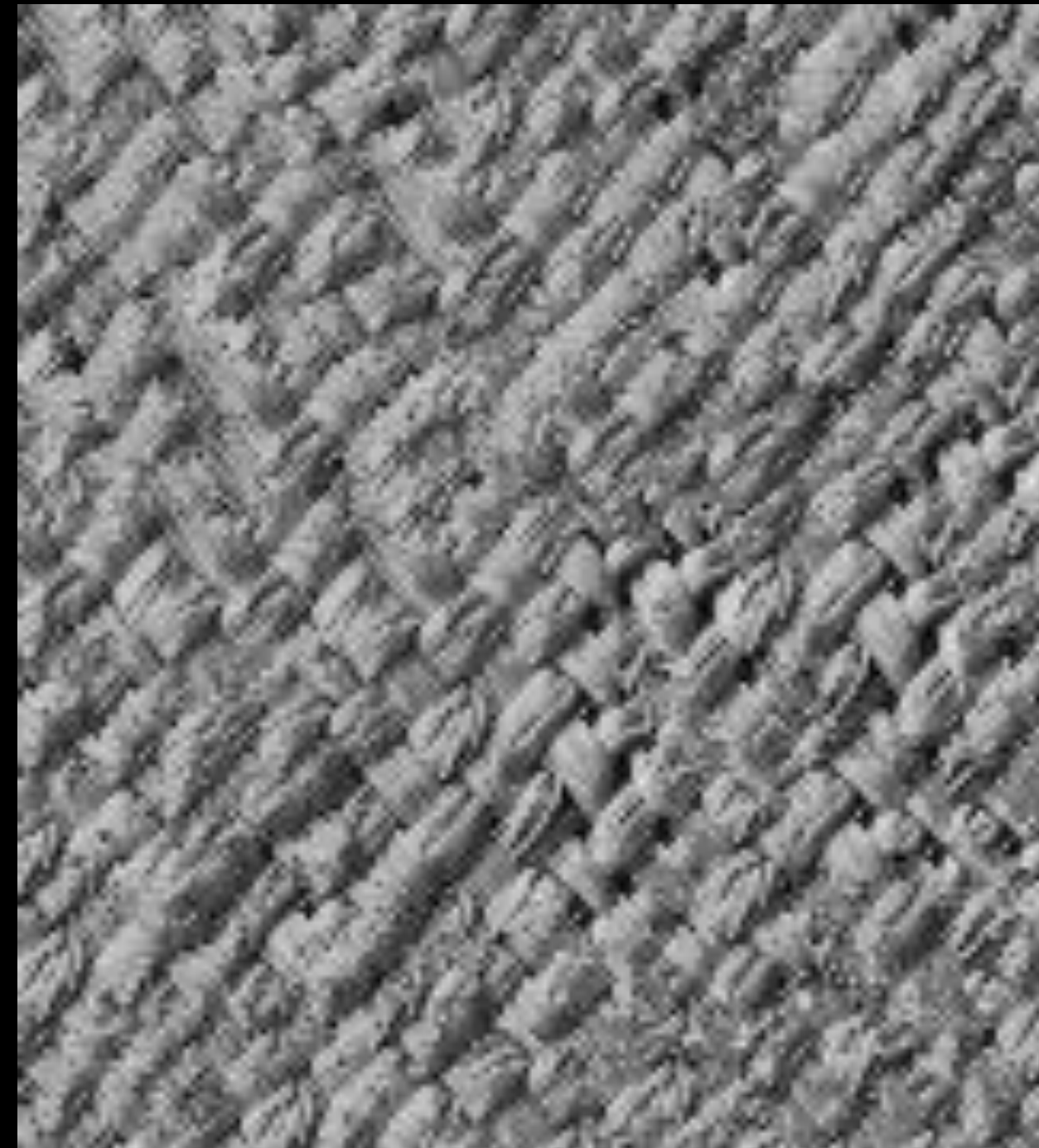


Synthesis Results

french canvas



rafia weave

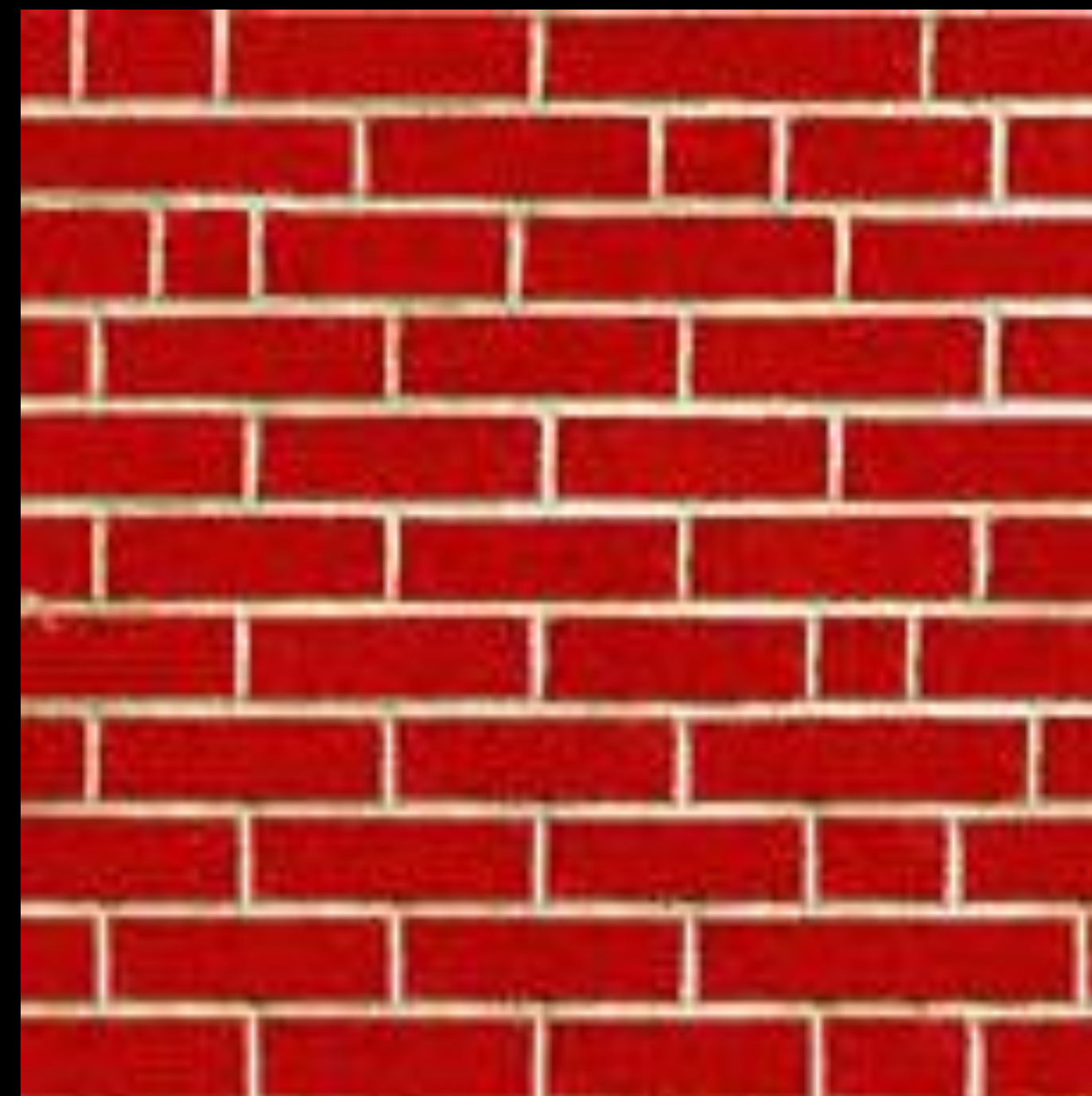
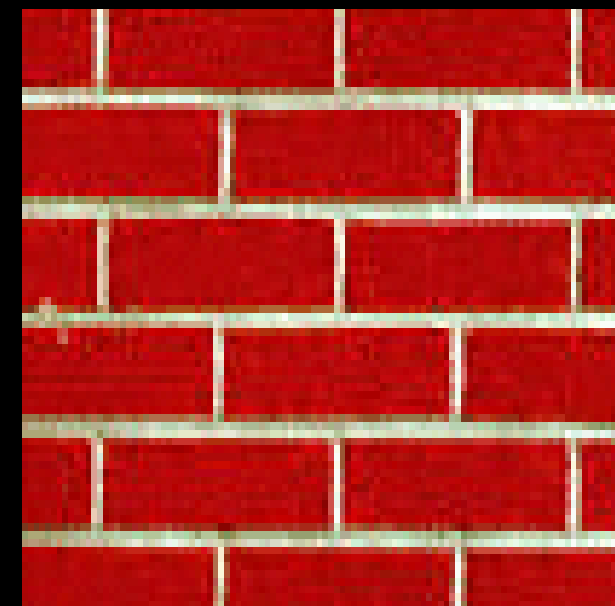


More Results

white bread



brick wall



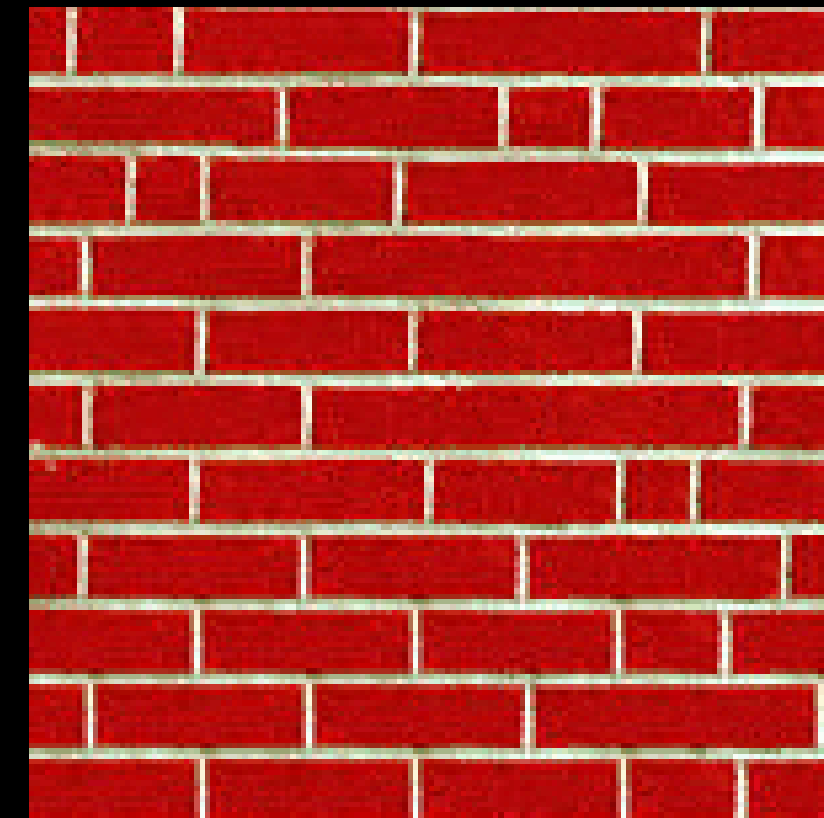
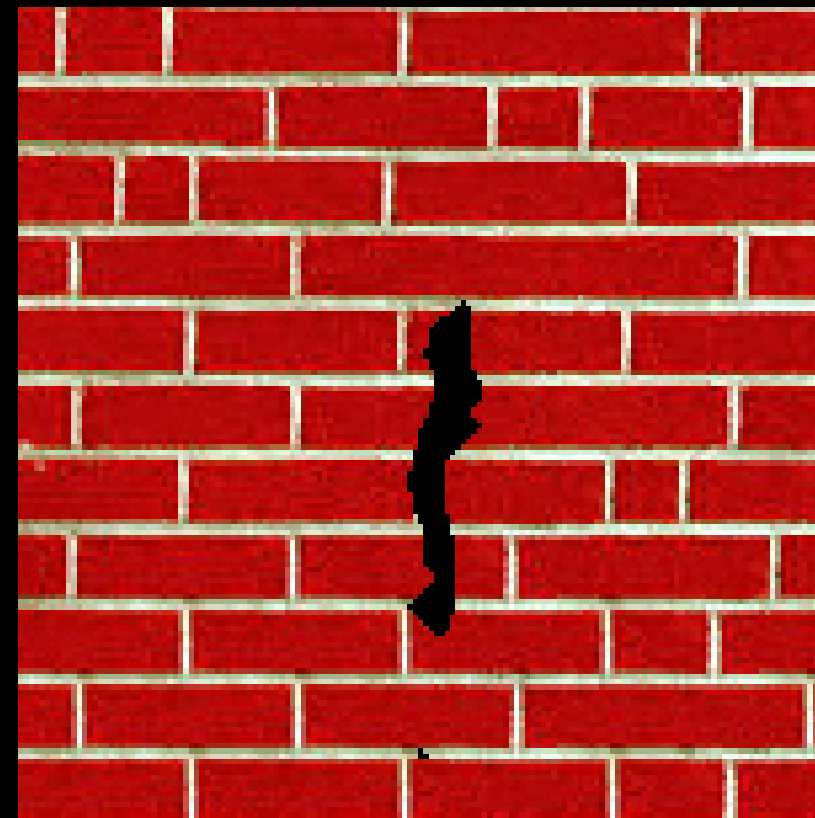
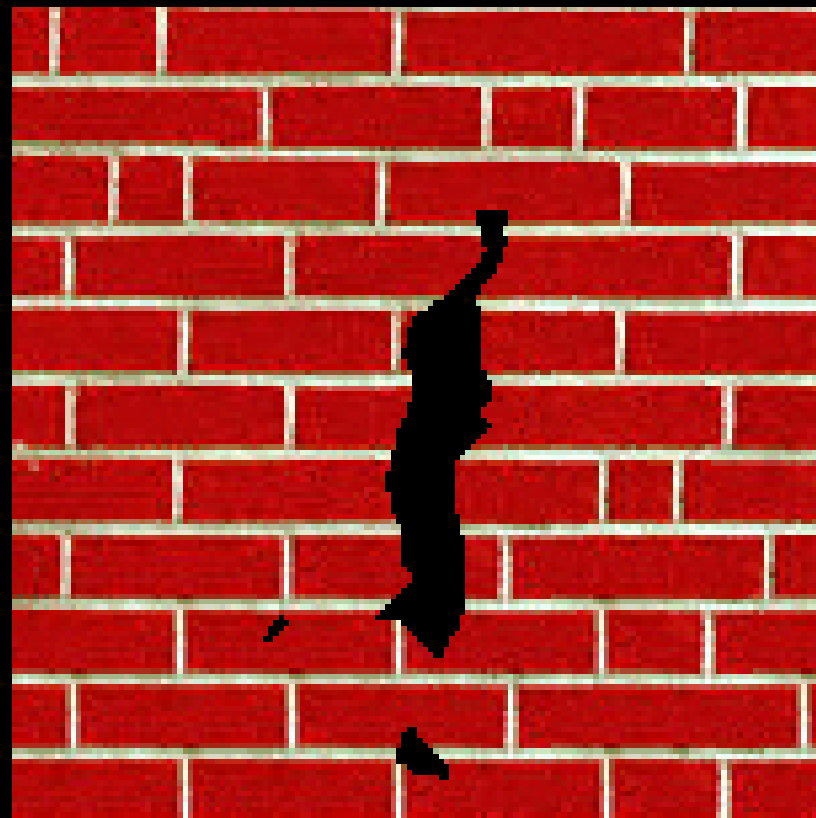
Homage to Shannon

coming in the unsensational
r Dick Gephardt was fai
rful riff on the looming
only asked, "What's your
tions?" A heartfelt sigh
story about the emergen
es against Clinton. "Boy
g people about continuin
ardt began, patiently obs
e, that the legal system h
g with this latest tanger

gle per
s otot
as "he
y f
ie "t
e f
n id
utonuc
qu
thenly
dthf
or
f
e
tin
s
s
th
adv
t
en
m
g
t

athaim. them . "Whnephartfe lartifelintomimer
fel ck Clirtioout omaim thartfelins.f out s aneto
the ry onst wartfe lck Gephtoomimeationl sigab
Chicoufit Clinut Clil riff on. hat's yordn, parut tly
ons yontonsteht wasked, paim t sahe loo riff on
nskoneploourtfeas leil A nst Clit, "Wleontongal s
k Clirtioouirtfepe ong pme abegal fartfenstemem
tiensteneltorydt telemephminsberdt was agemer
ff ons artientont Cling peme as rtf e atih, "Boui s
nal s fartfelt sig pedr thdt ske abounutie aboutioo
tfeone was yous abowonhardt thatins fain, ped, '
ains. them, pabout wasy arfnt courtly d, l n A h
ole emthrdngbooreme agas fa bontinsyst Clinut
ory about continst Clipeopinst Cloke agatiff out C
stome zinemen fly ardt beoraboul n, thenly as t C
cons fairme Dientont wat coutlyohgans as fan
ien, phrtfaul, "Wbaut cout congagal comininga
mifmst Clily abon al cocountha.emungairt tf oun
Whe loocrysta loontieph. intly on, theoplegatick C
oul tatieozntly atie Dientomt wal s f t begae ener
nthahgat's enenhhmas fan. "intchthorw ahons w

Hole Filling



Extrapolation

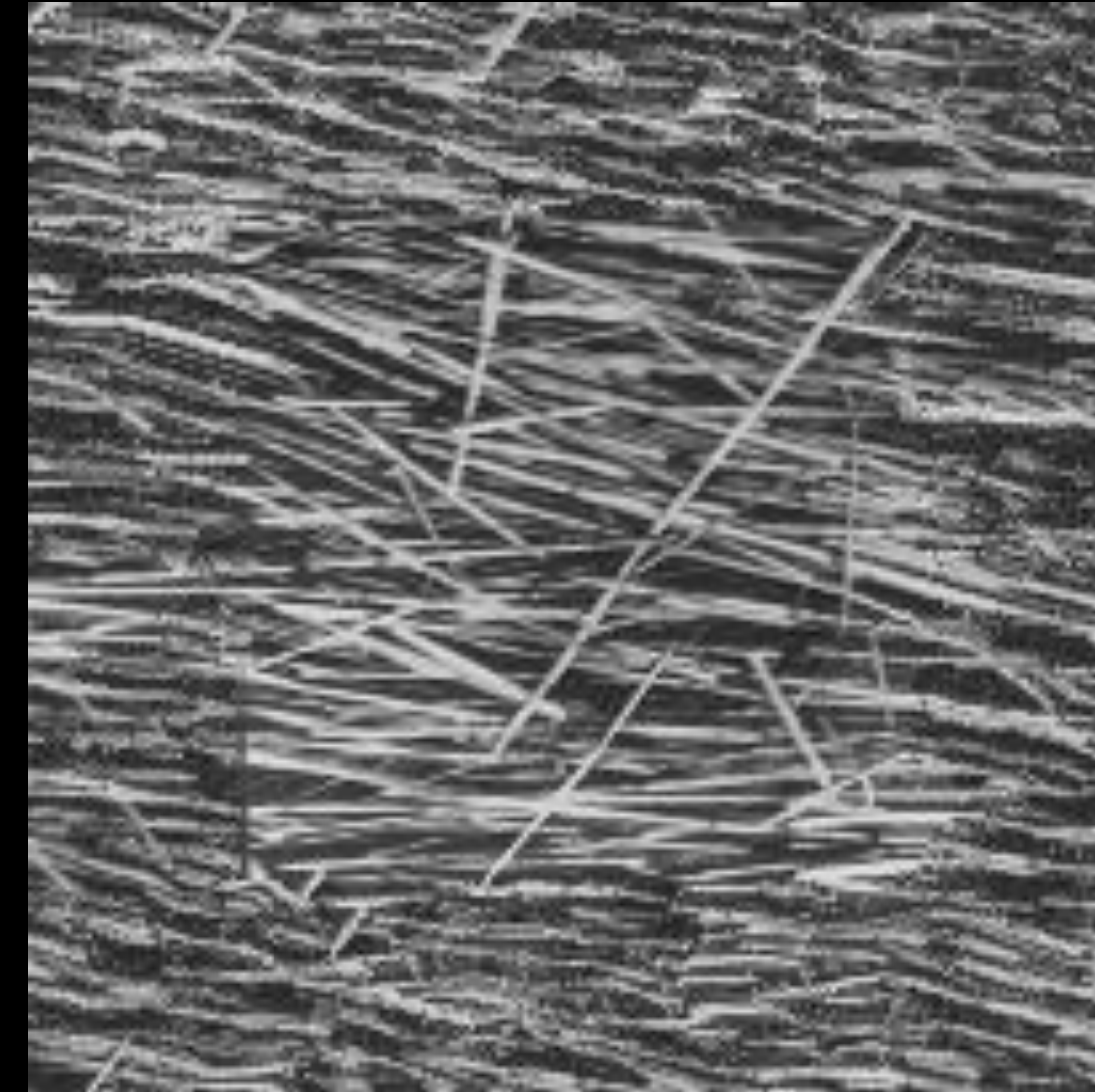
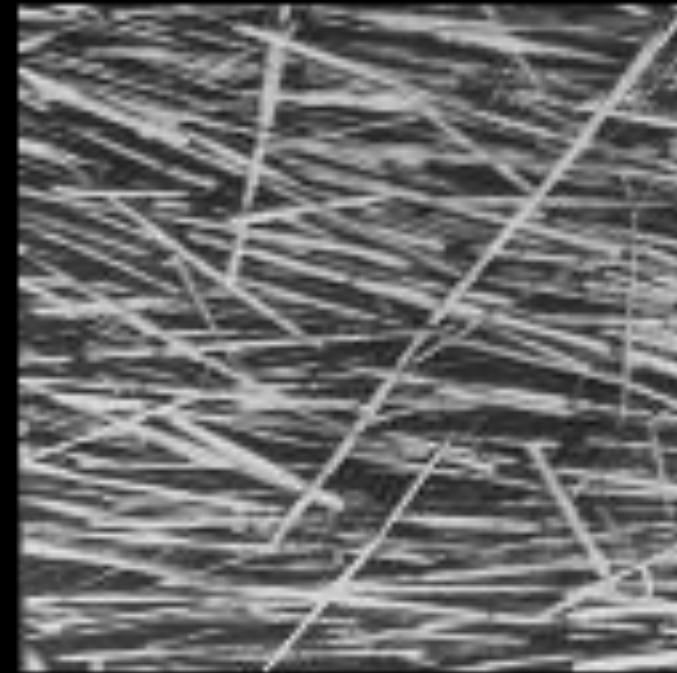


Image Analogies

Aaron Hertzmann^{1,2}

Chuck Jacobs²

Nuria Oliver²

Brian Curless³

David Salesin^{2,3}

¹New York University

²Microsoft Research

³University of Washington

Image Analogies



A



A'



B



B'



Image Analogies

Goal: Process an image by example



A

:



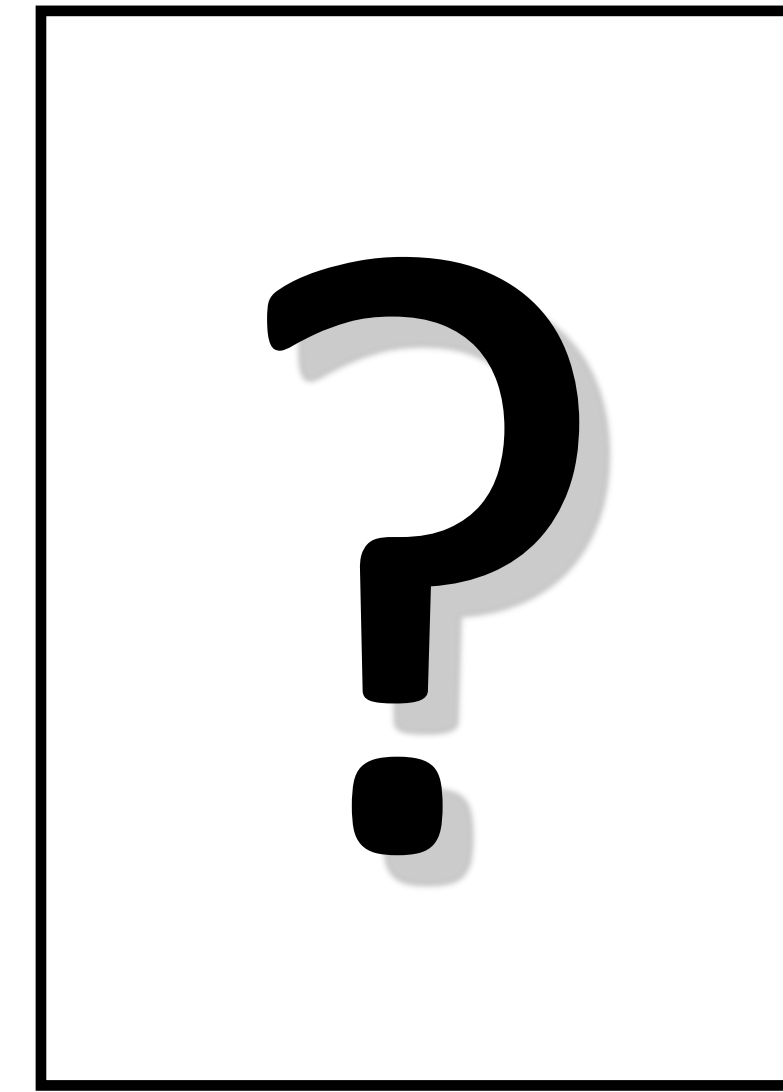
A'

::



B

:



B'

Non-parametric sampling



A



A'

⋮

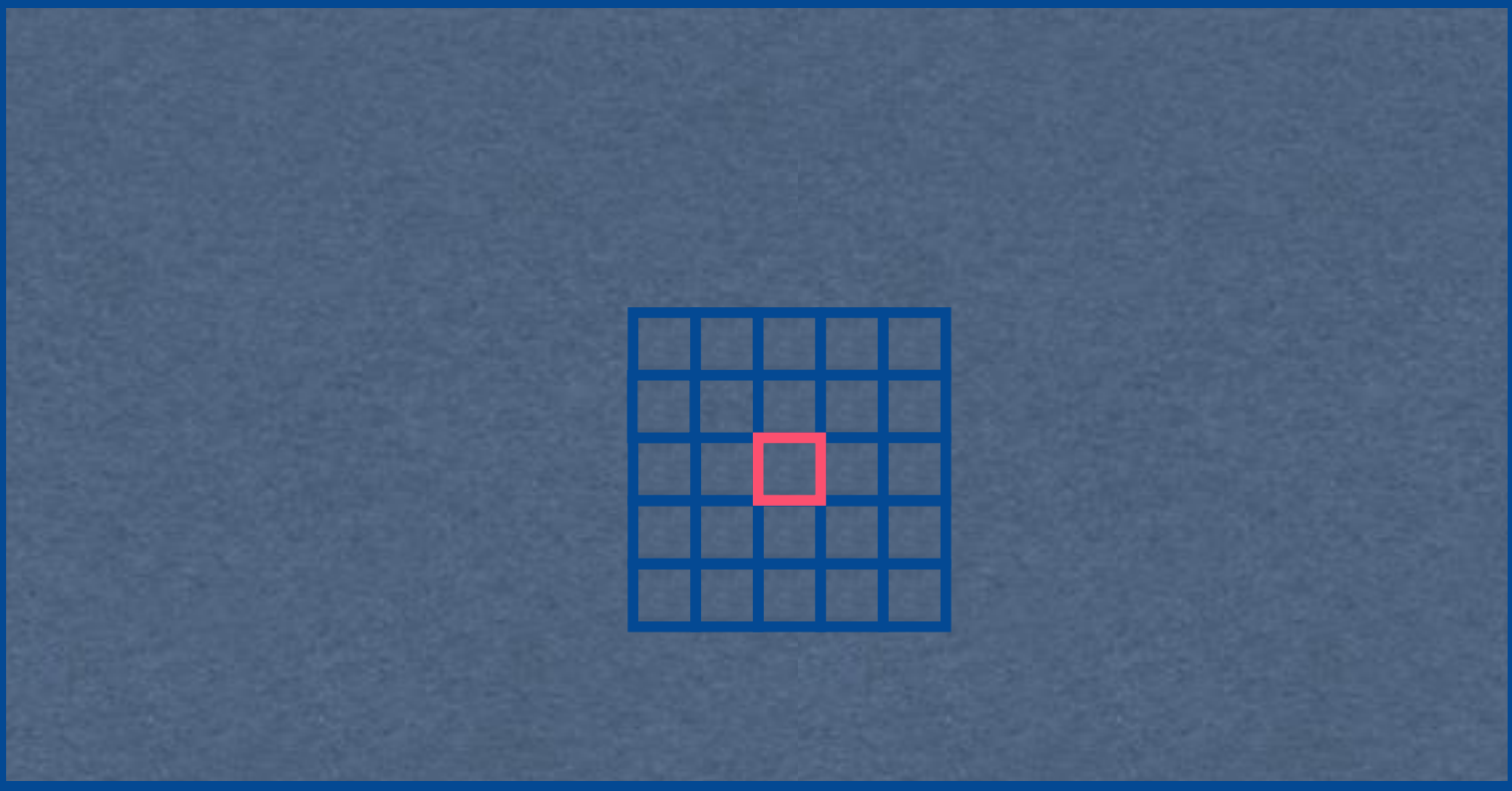


B

⋮

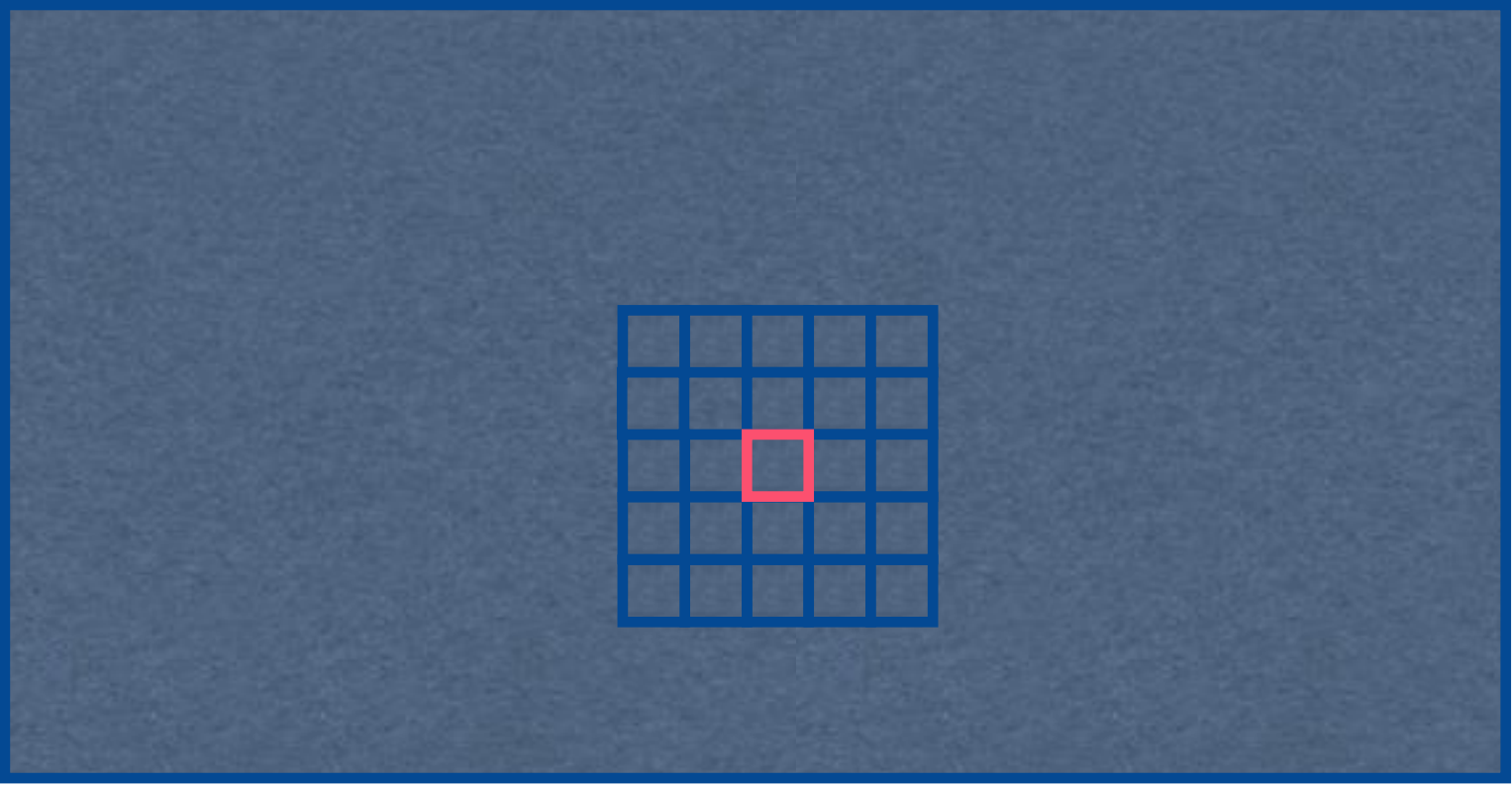


B'



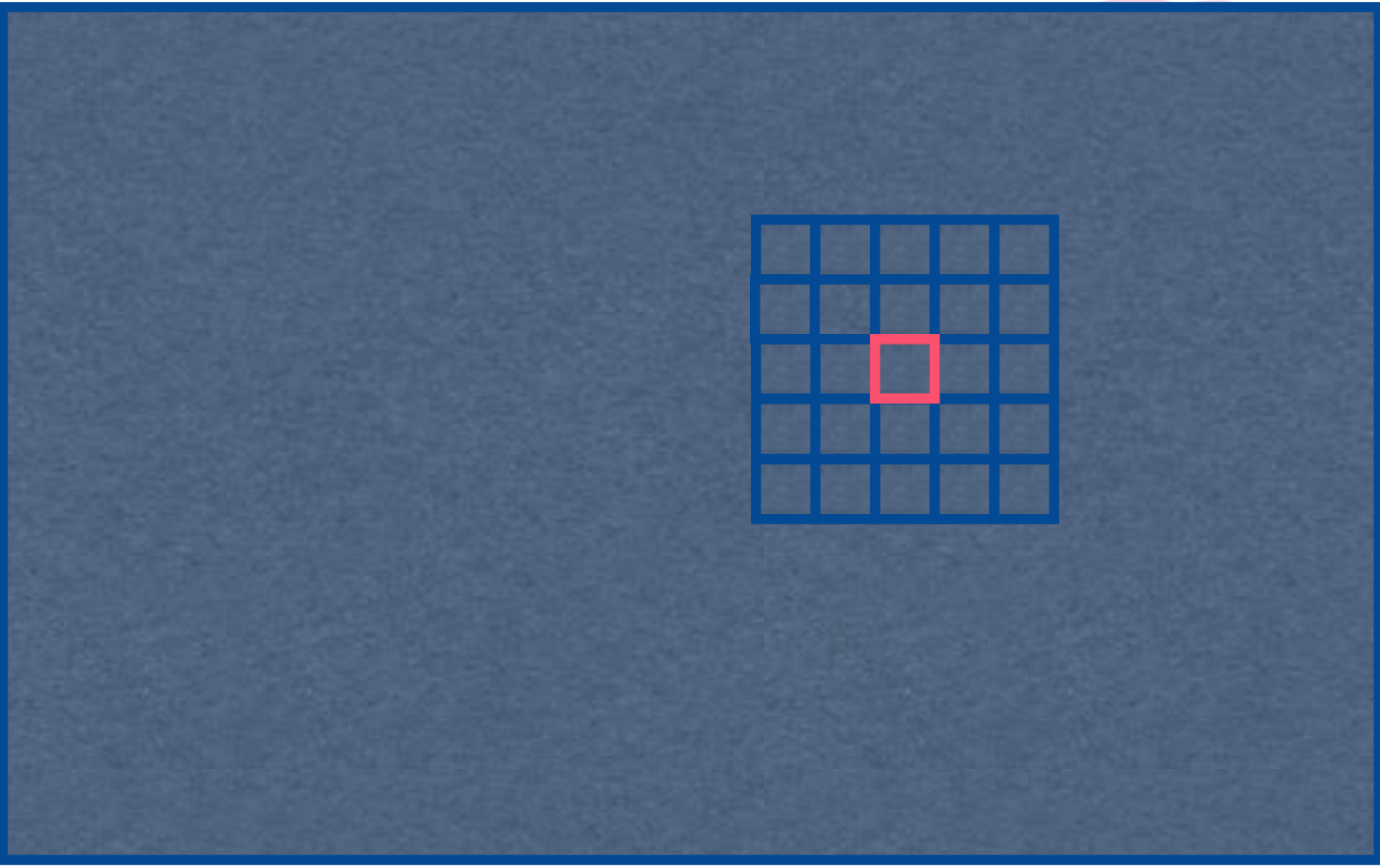
A

⋮



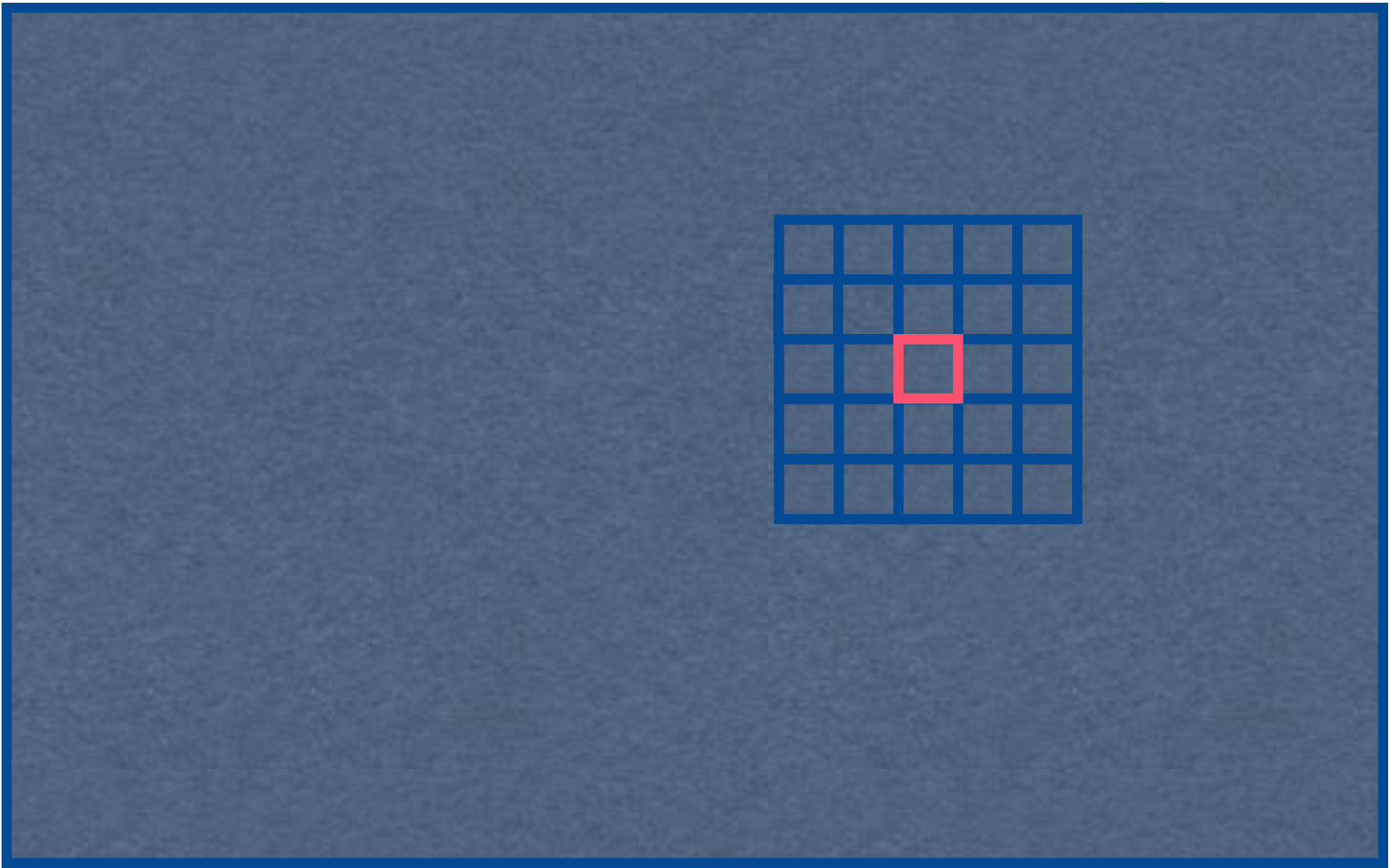
A'

⋮

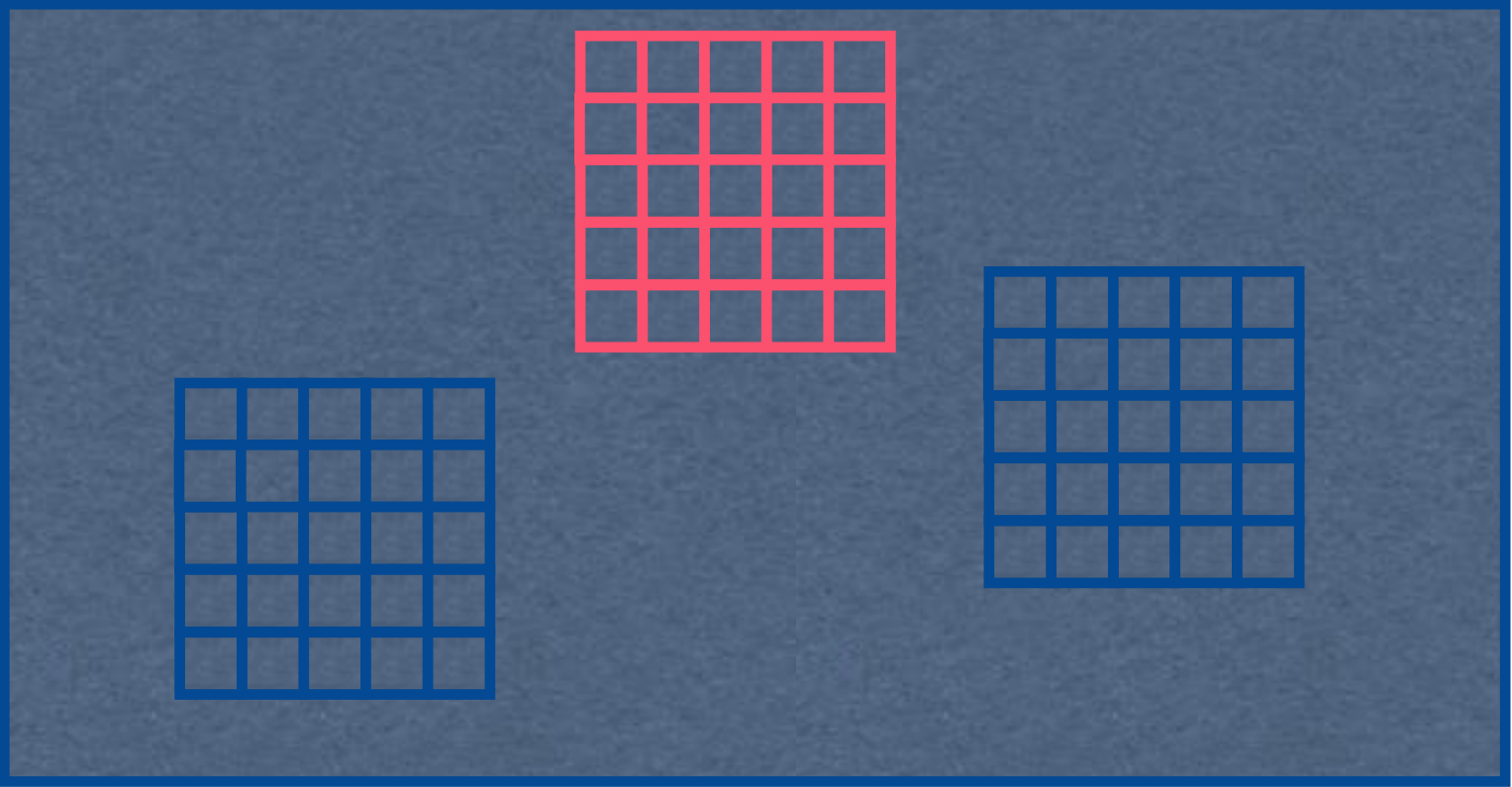


B

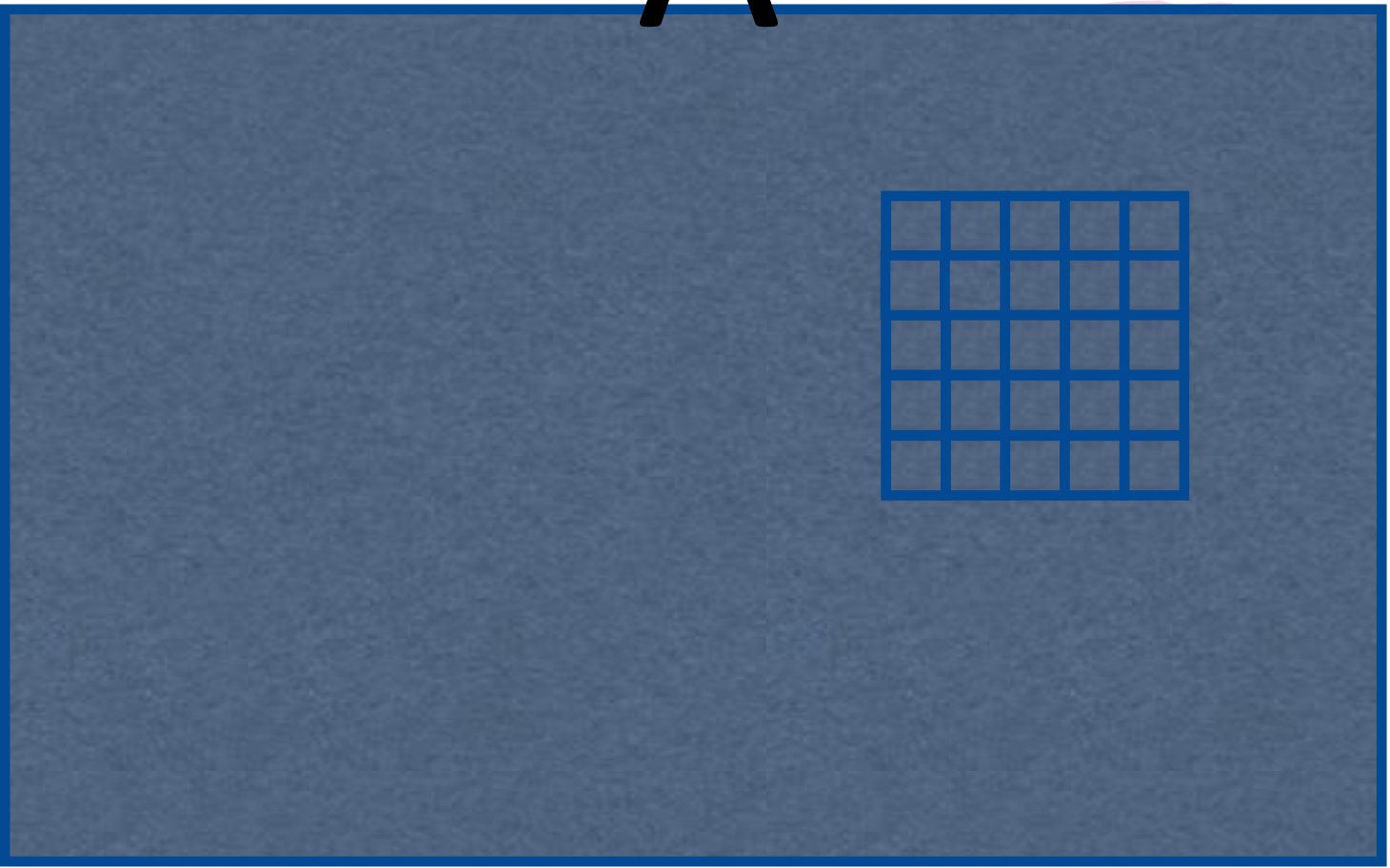
⋮



B'

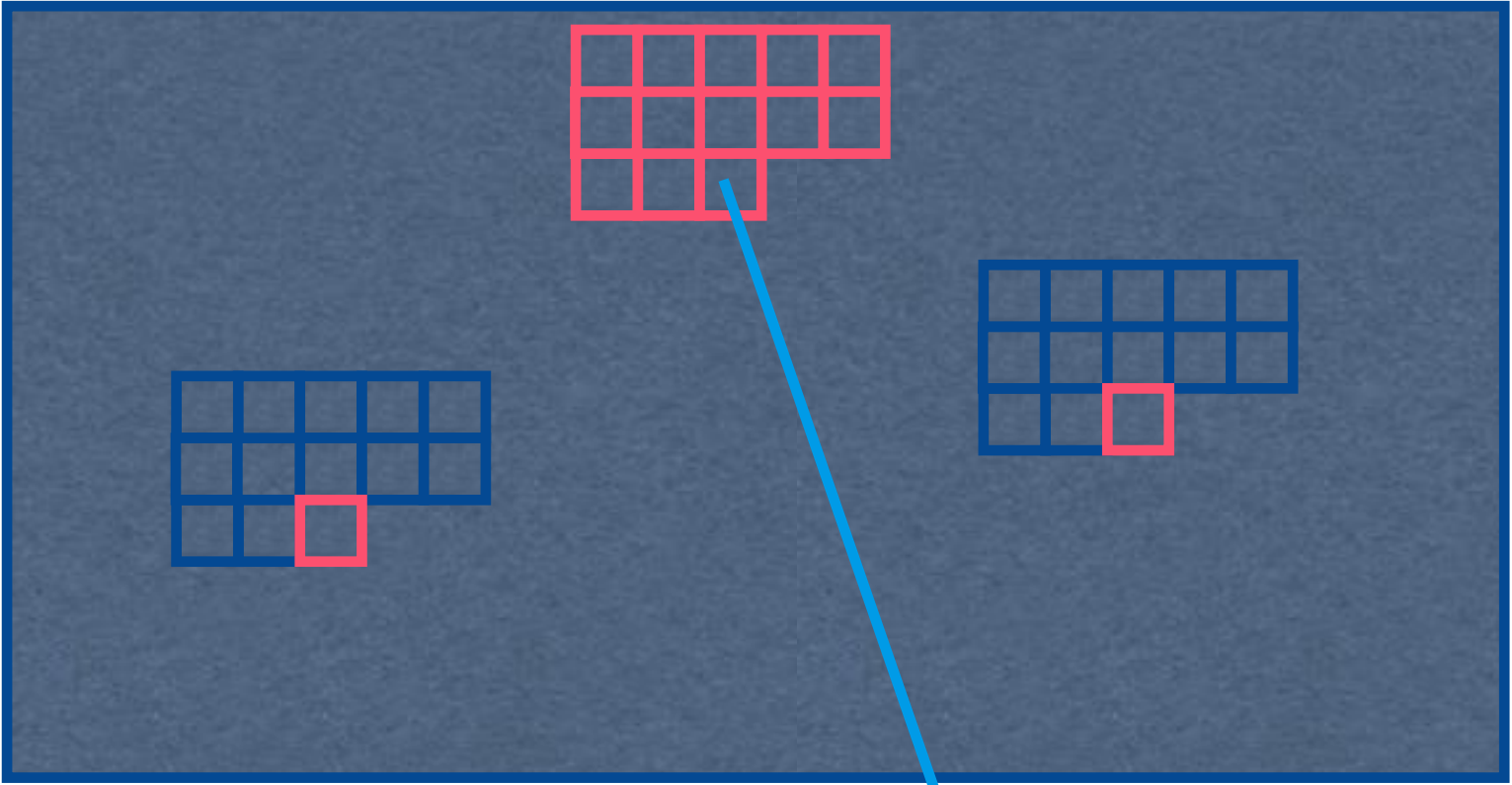


A

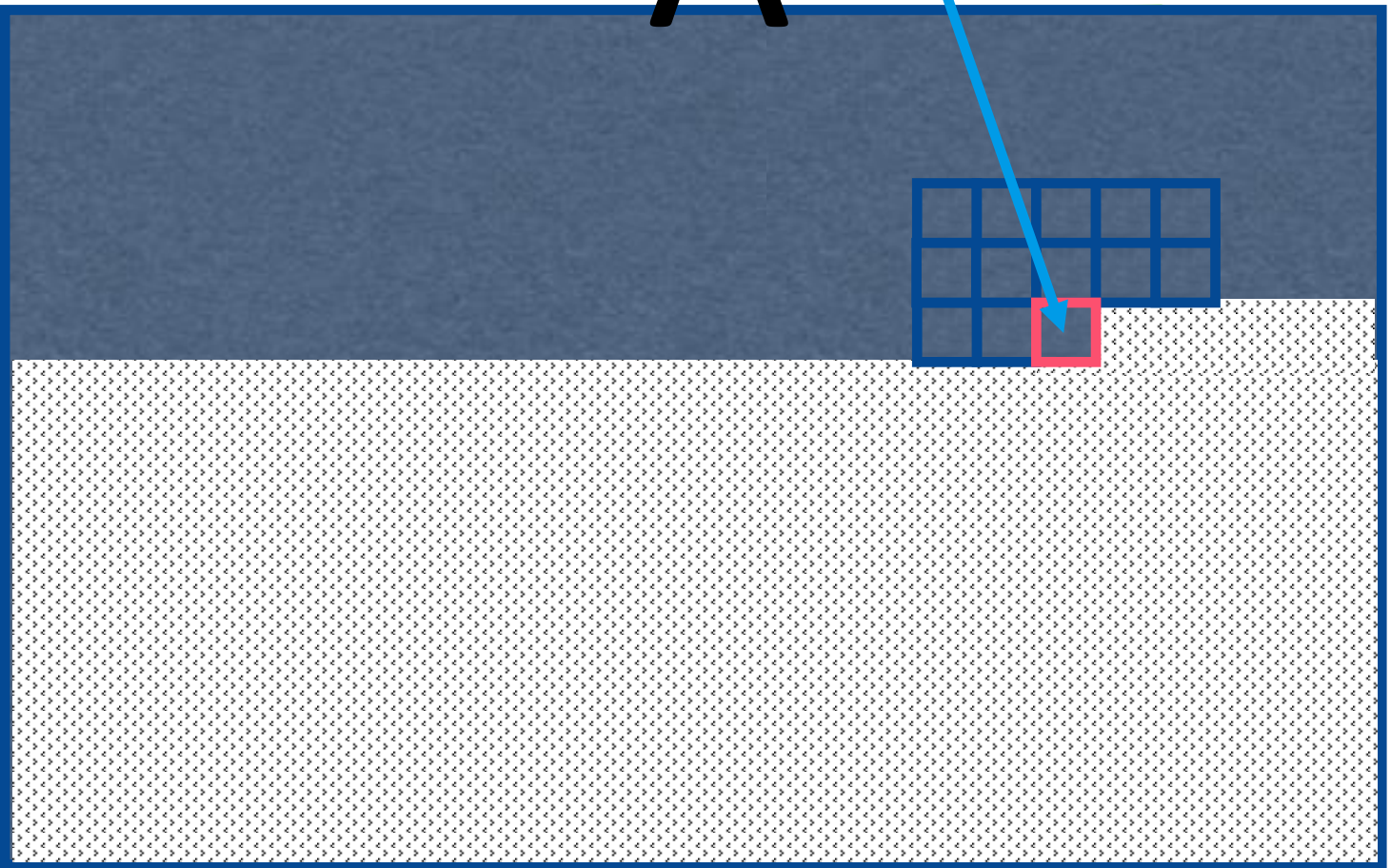


B

⋮



A'



B'

⋮

Blur Filter



Unfiltered source (A)



Filtered source (A')



Unfiltered target (B)

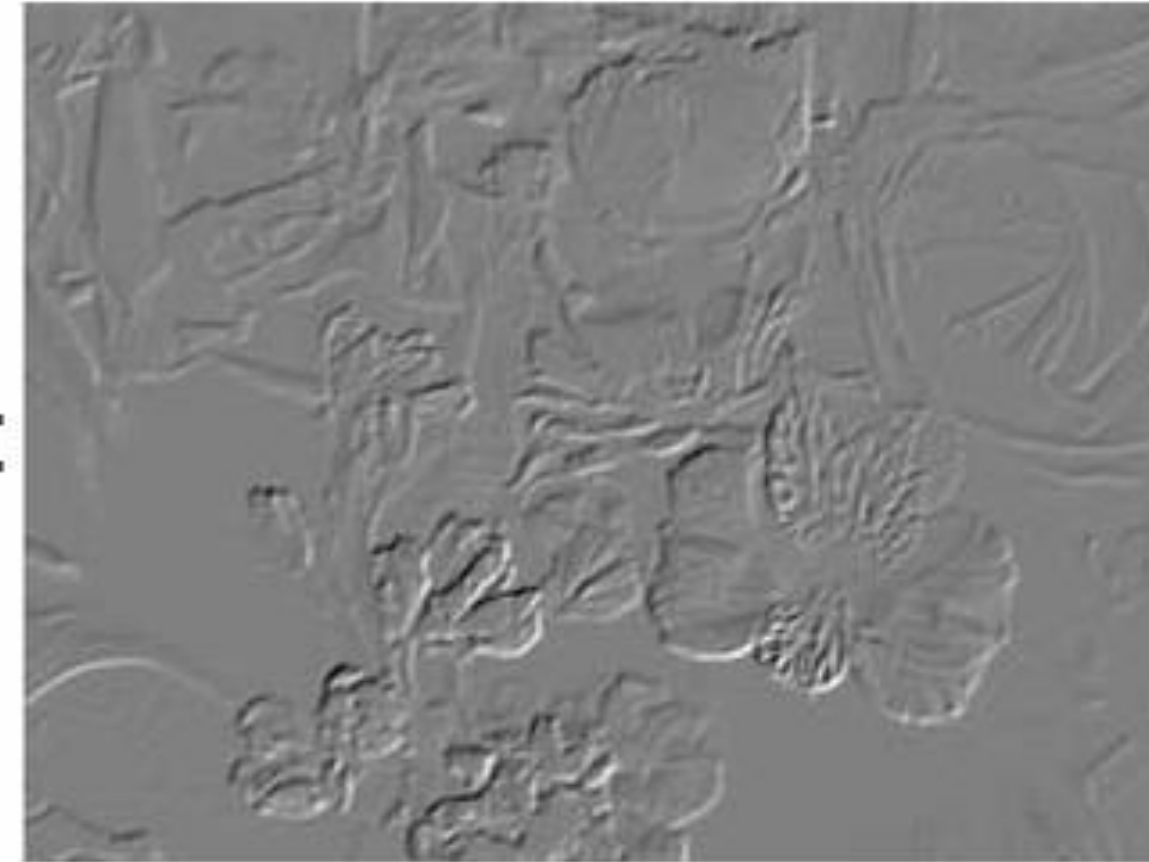


Filtered target (B')

Edge Filter



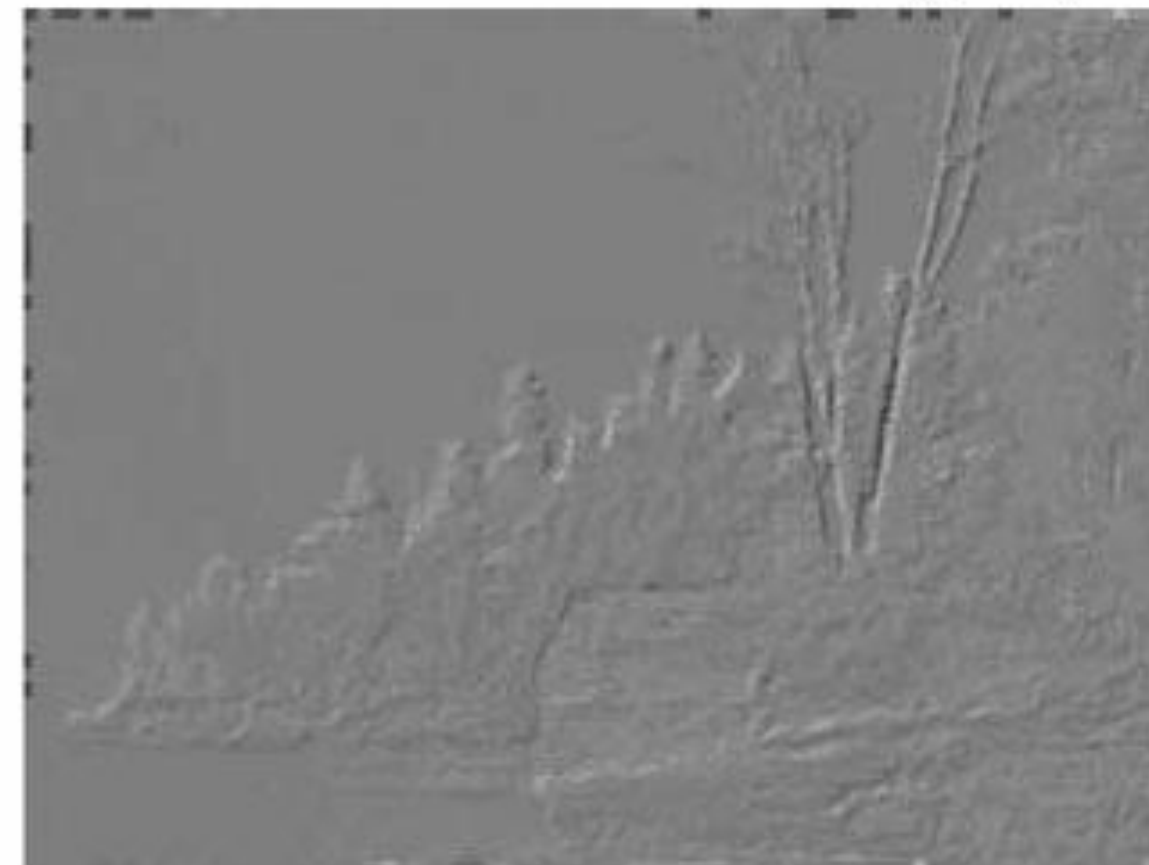
Unfiltered source (A)



Filtered source (A')



Unfiltered target (B)



Filtered target (B')

Artistic Filters



A



A'



B



B'

Colorization



Unfiltered source (A)

▪
▪



Filtered source (A')

▪ ▪
▪ ▪



Unfiltered target (B)

▪
▪



Filtered target (B')

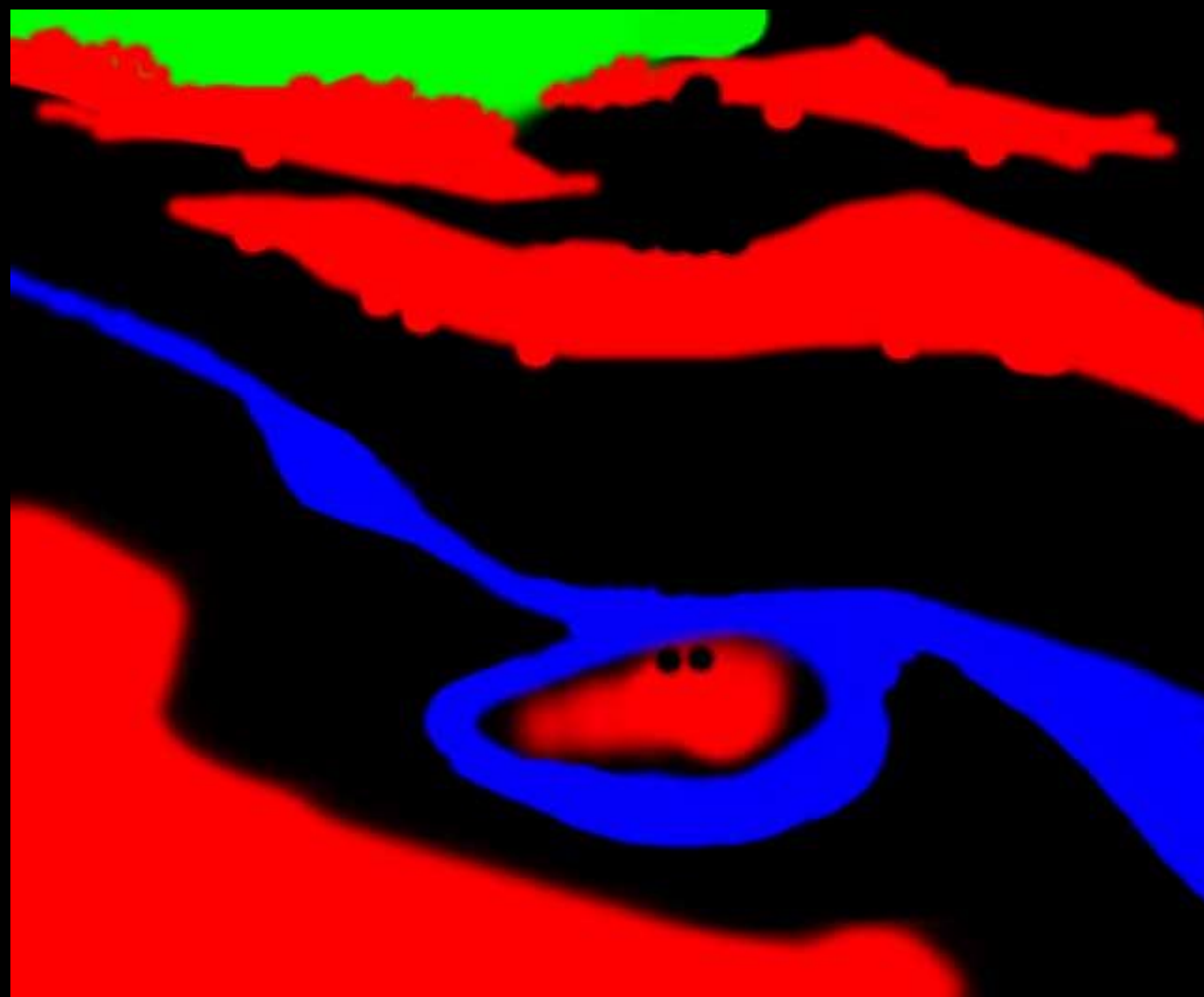
Texture-by-numbers



A



A'



B



B'

Visual Prompting via Image Inpainting

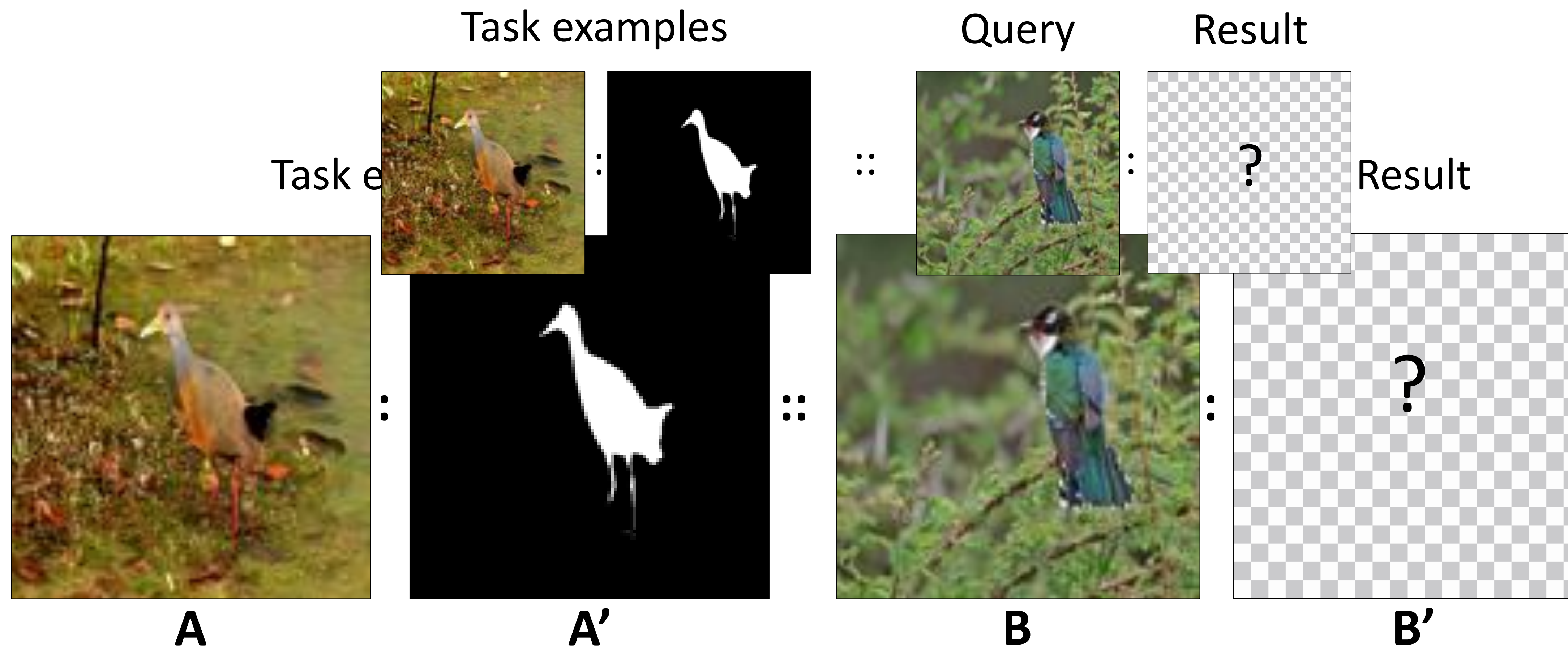
Amir Bar*, Yossi Gandelsman*,
Trevor Darrell, Amir Globerson, Alexei A Efros

NeurIPS 2022



* Equal contribution

Visual Prompting

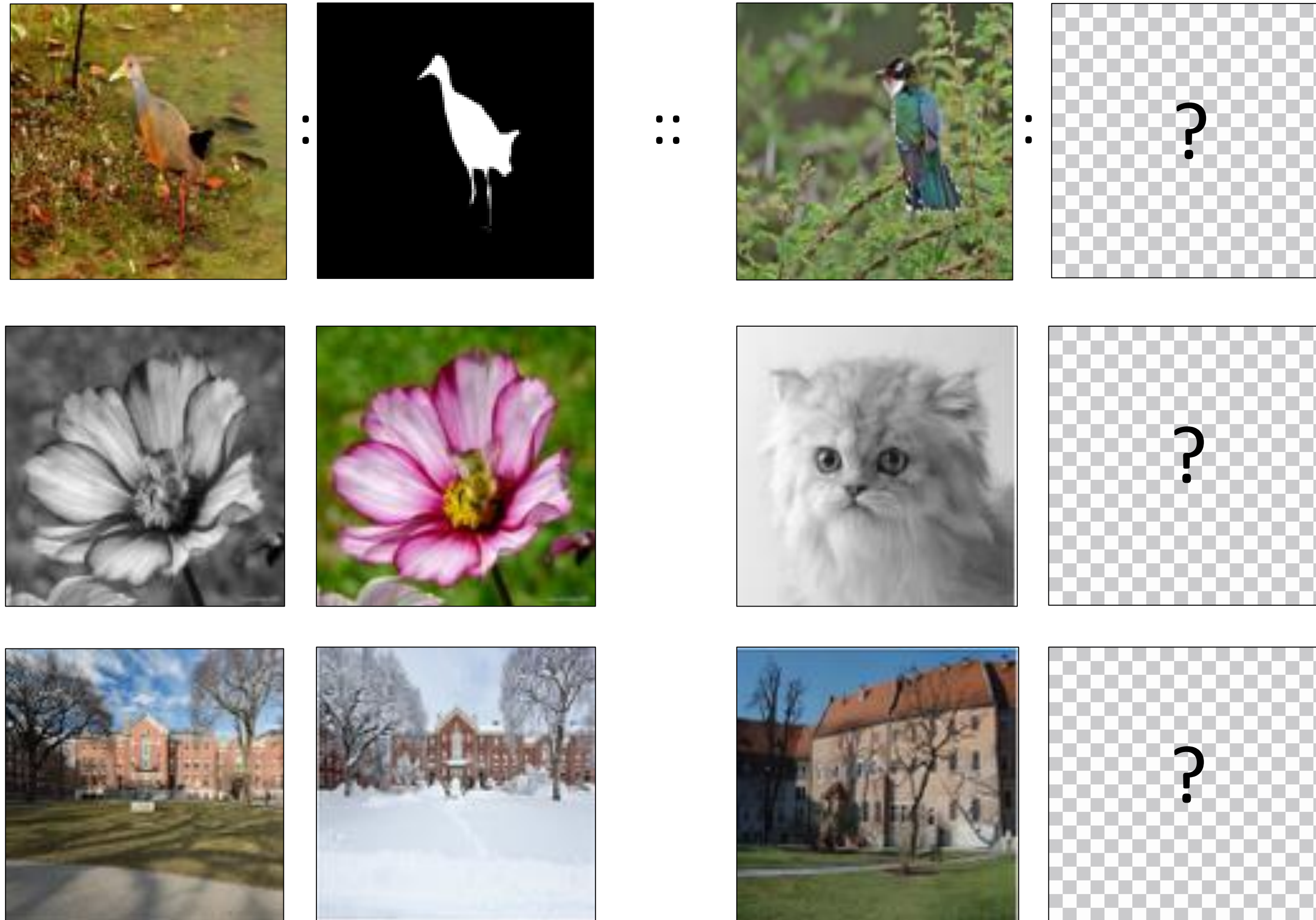


Visual Prompting

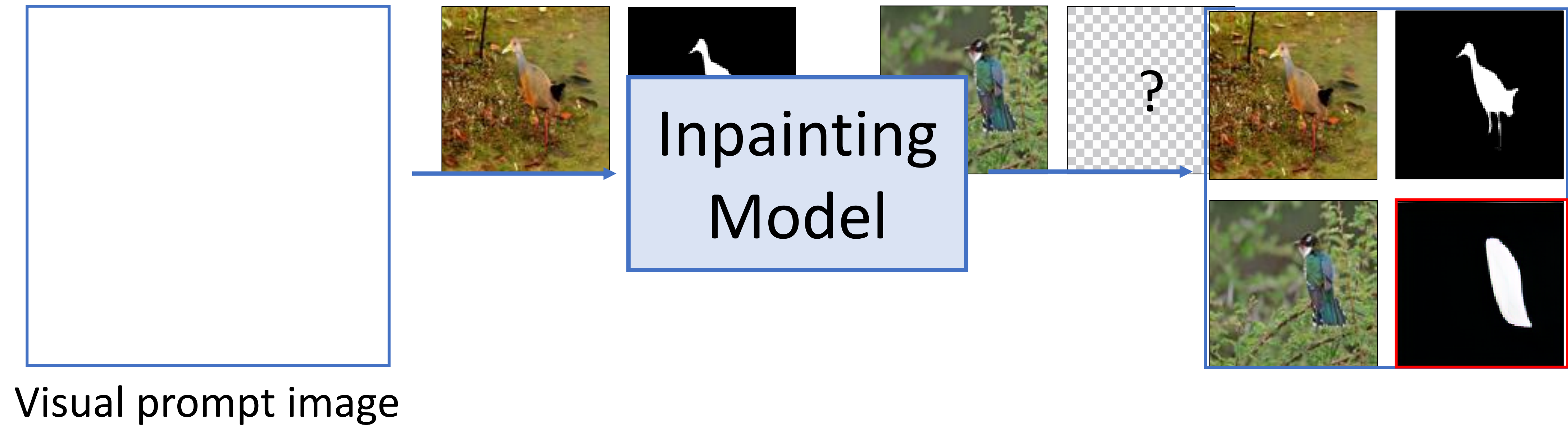
Task examples

Query

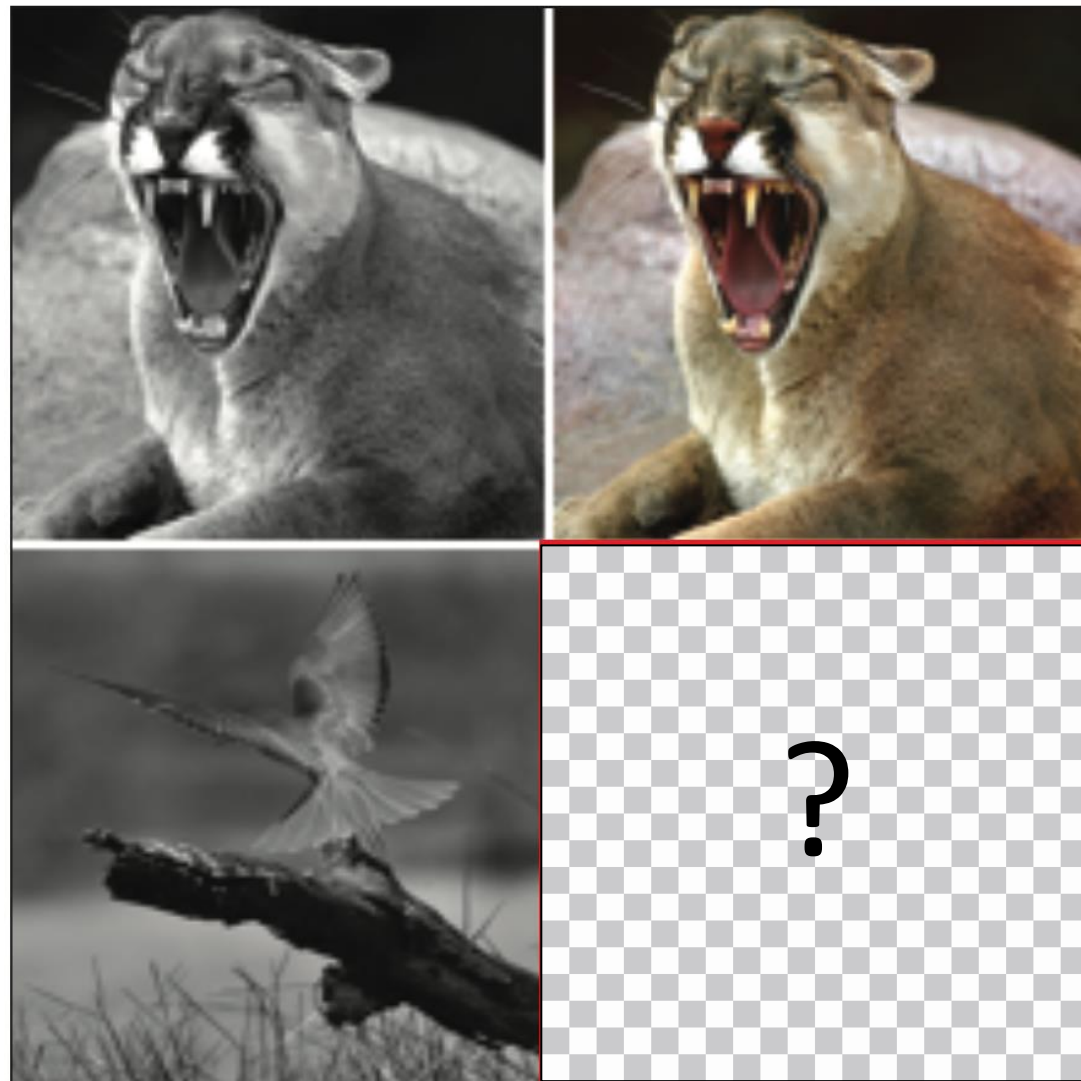
Result



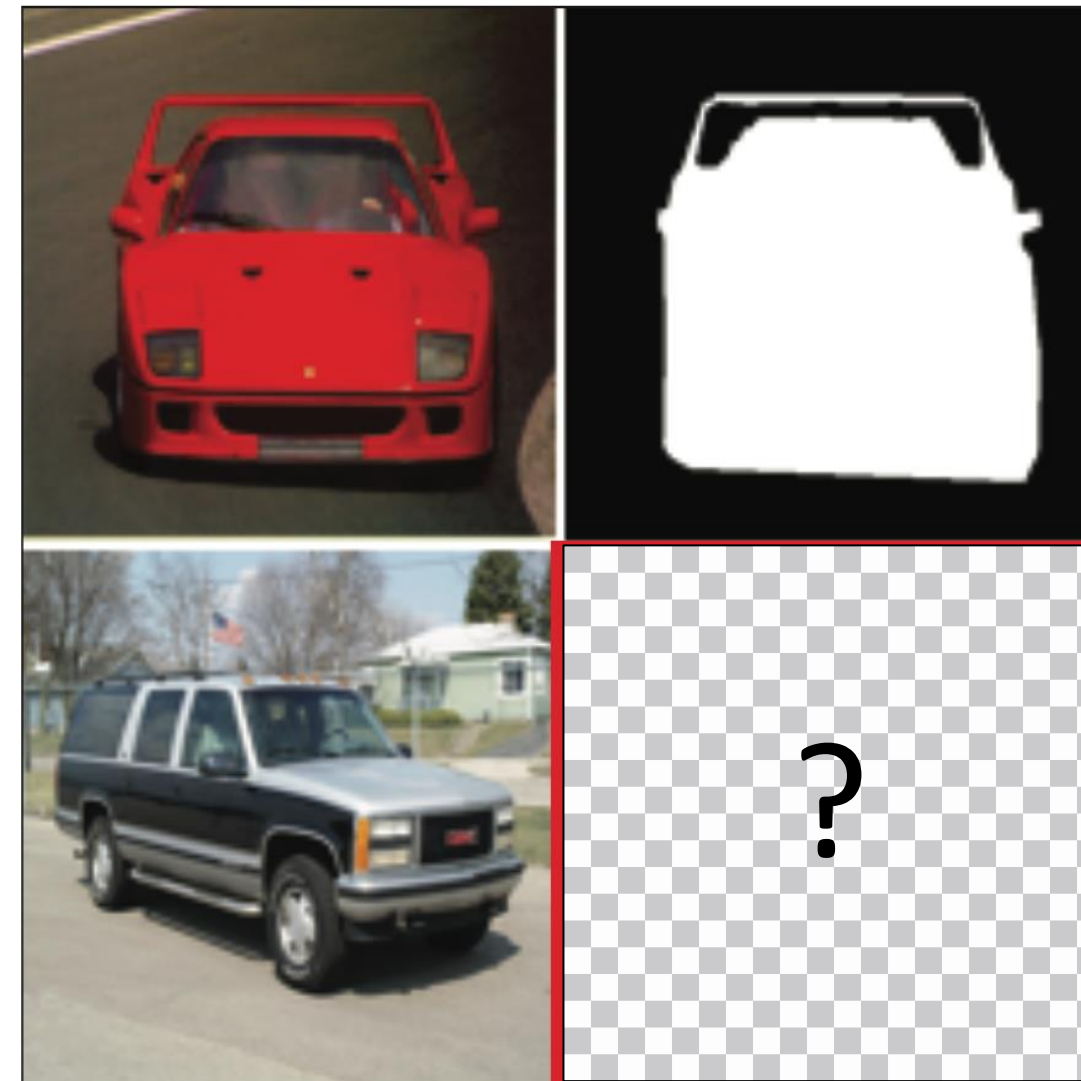
Inpainting models to the rescue!



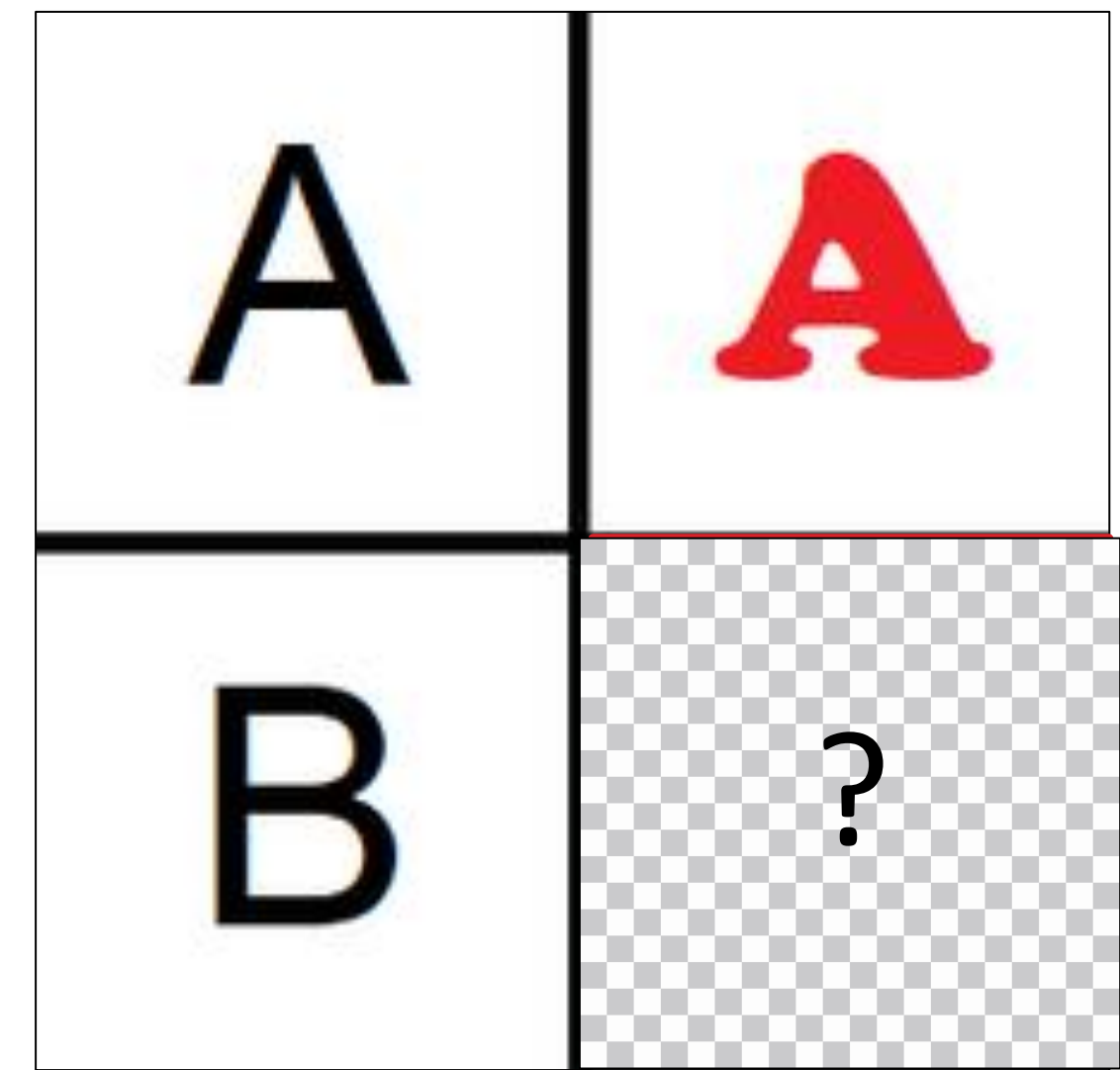
Wide range of tasks



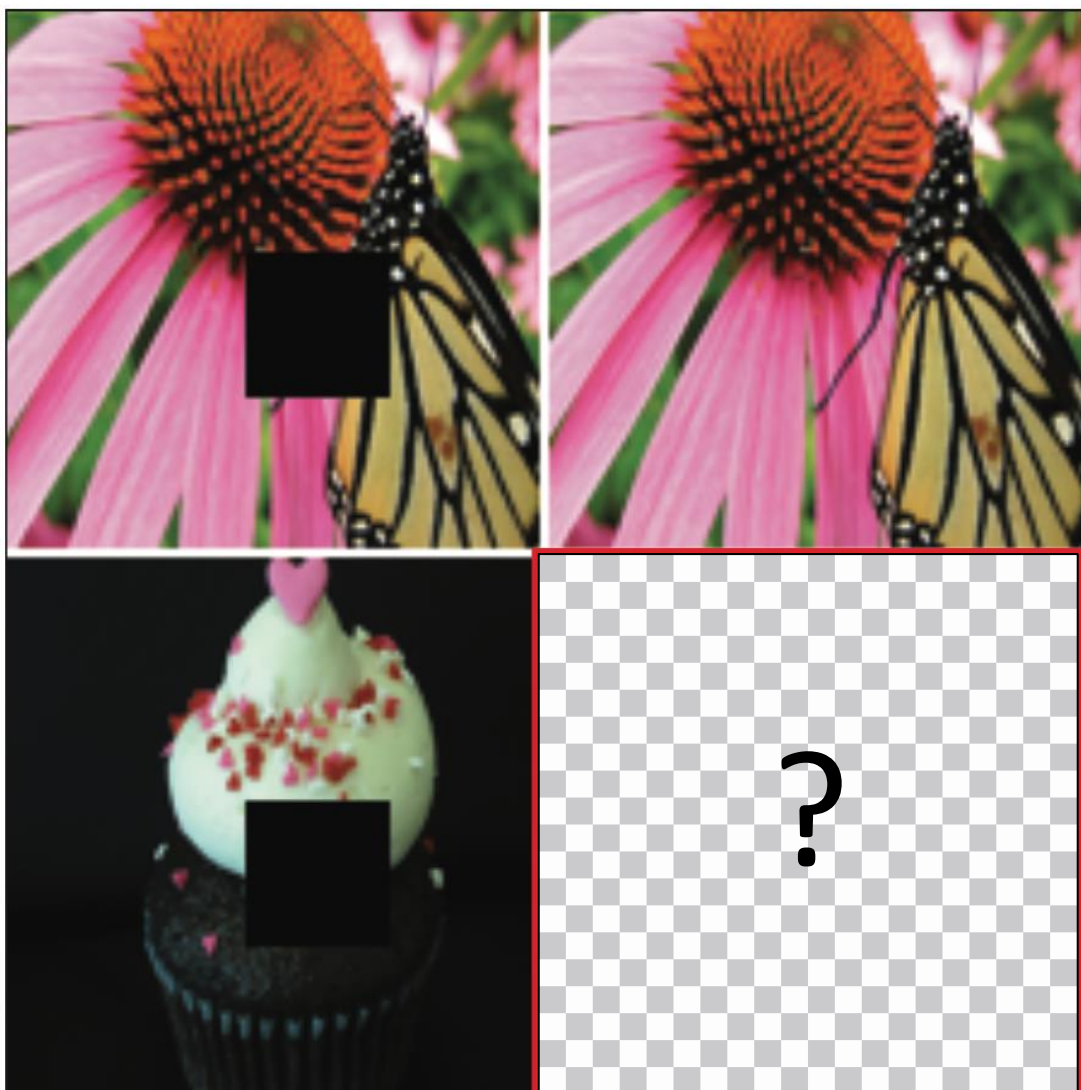
Colorization



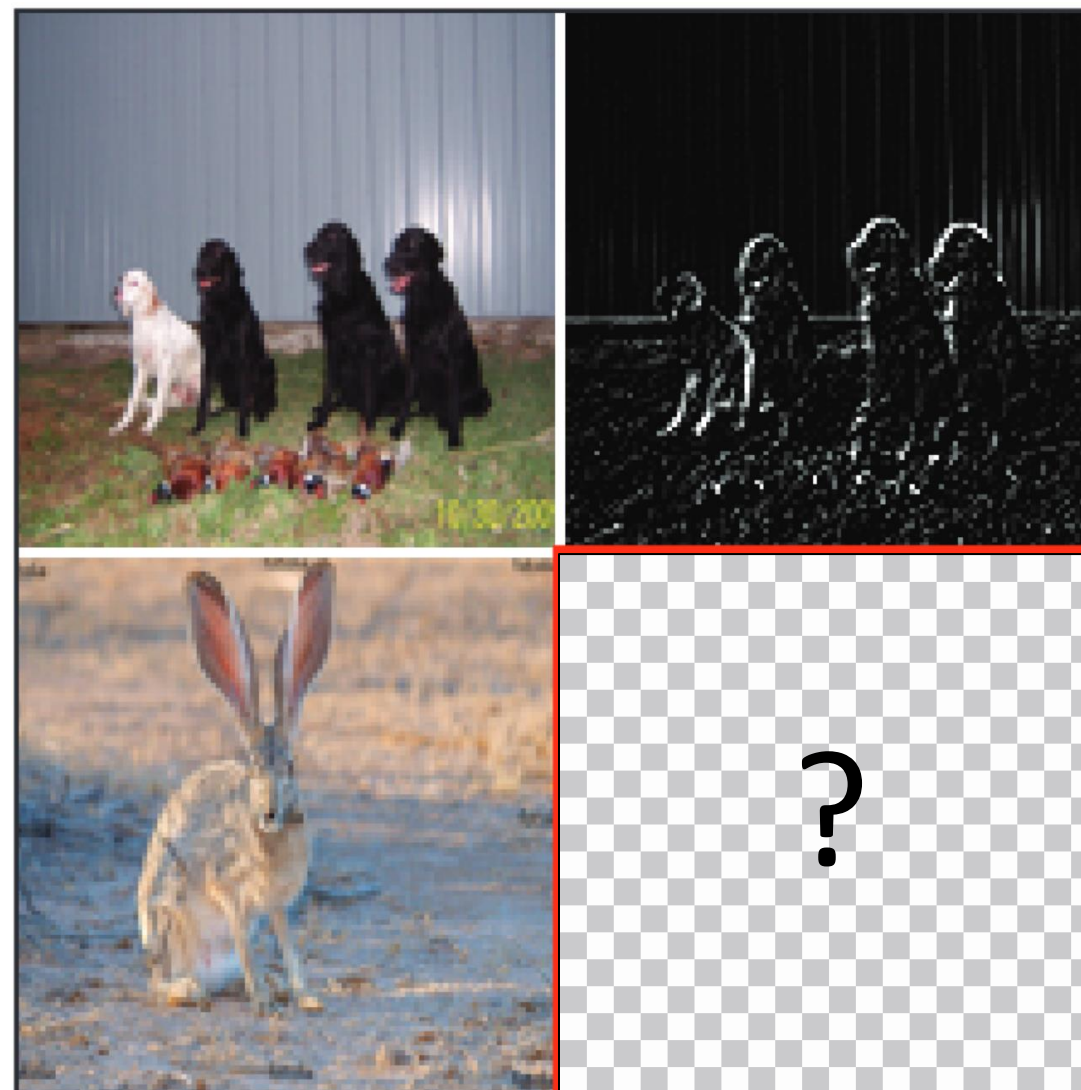
Segmentation



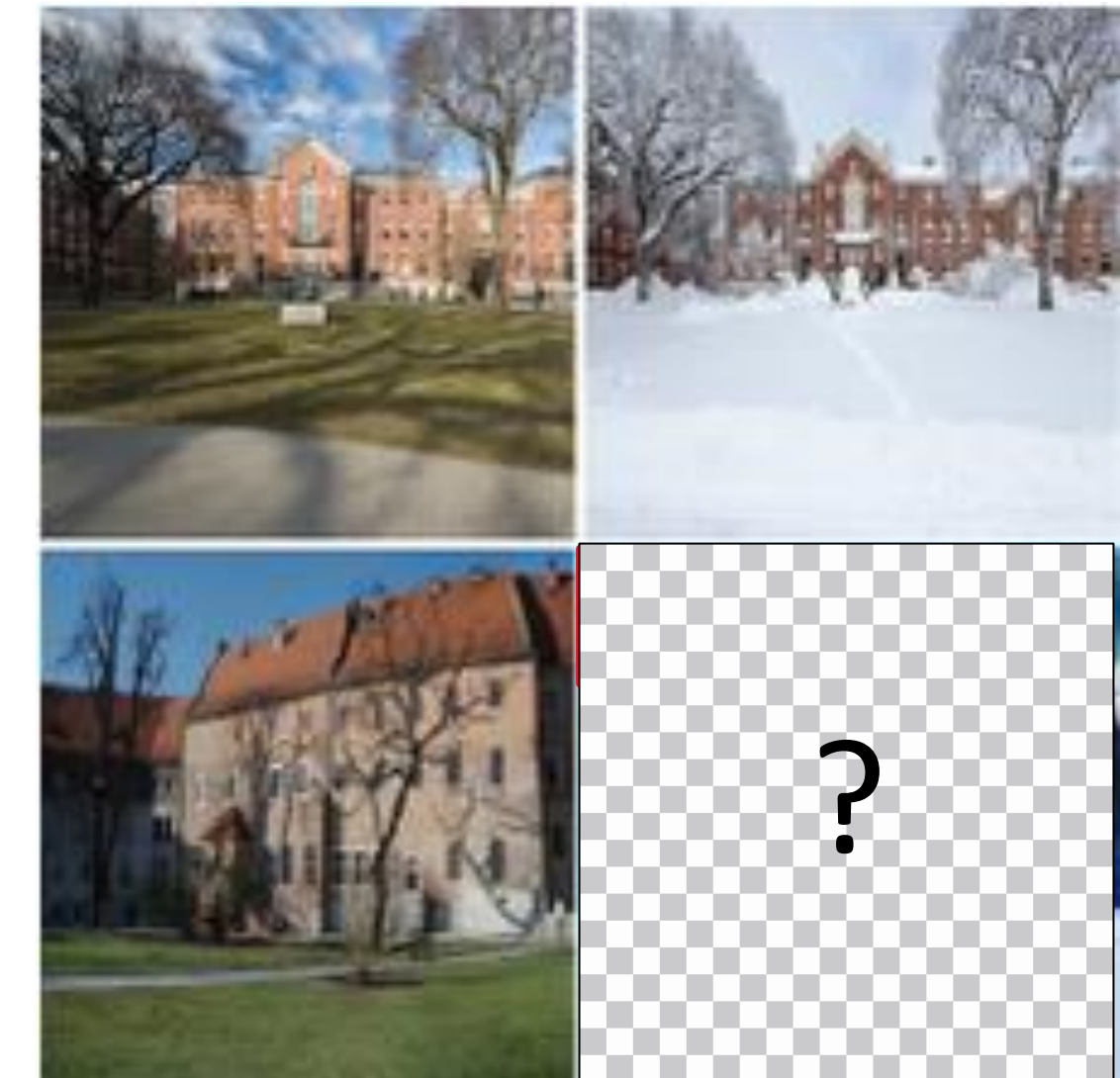
Font Style Transfer



Inpainting



Edge Detection



Style Transfer

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 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 the sun was shining I was laying in bed

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

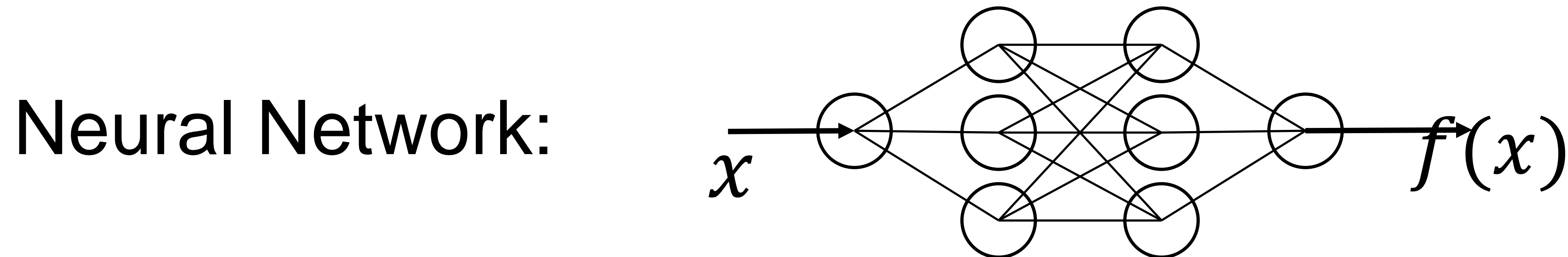
$$P(x_n | x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4}, x_{n-5}, x_{n-6}, x_{n-7}, x_{n-8}, x_{n-9}, x_{n-10}, x_{n-11}, x_{n-12}, x_{n-13})$$

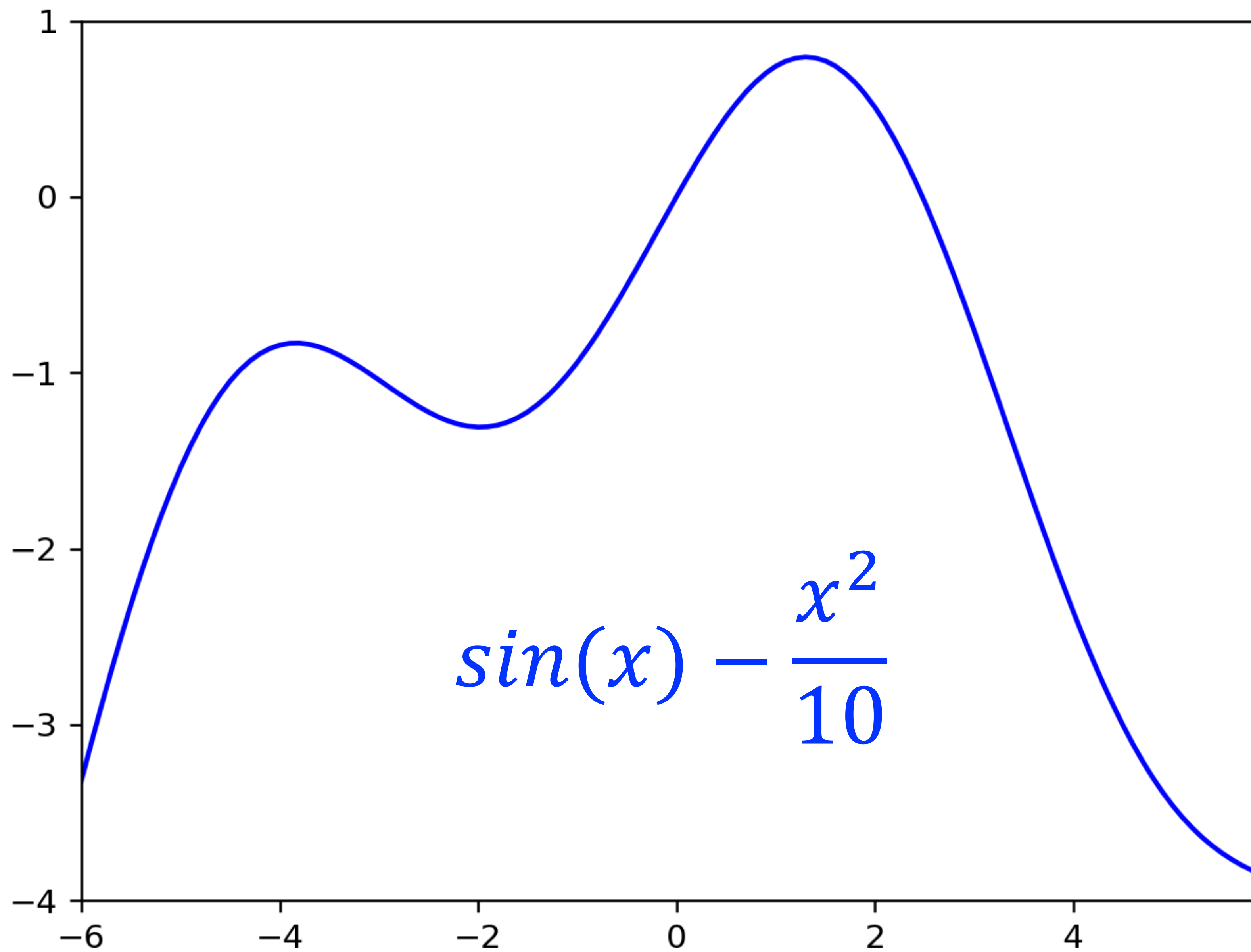
10^{70} combinations

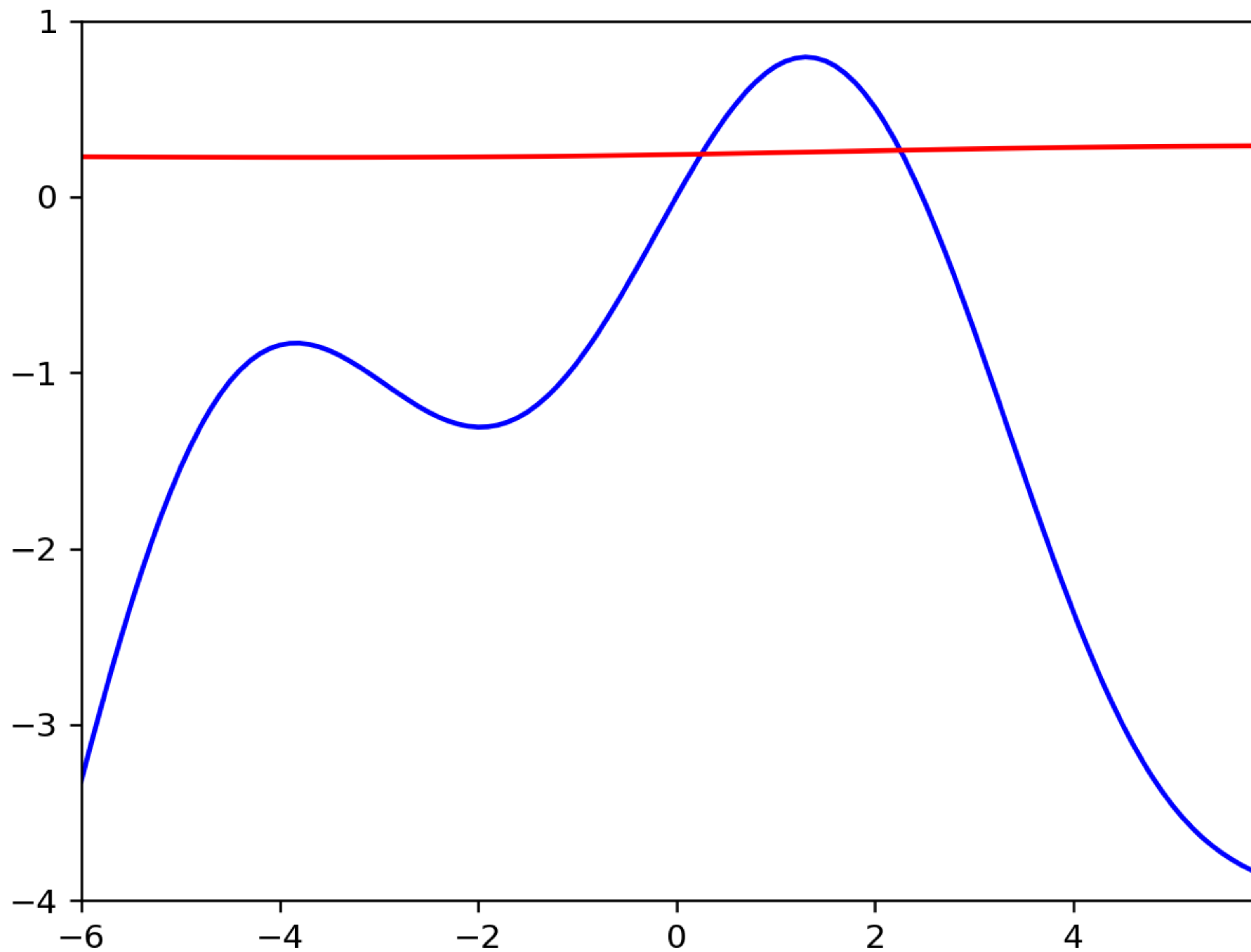
Function Approximation

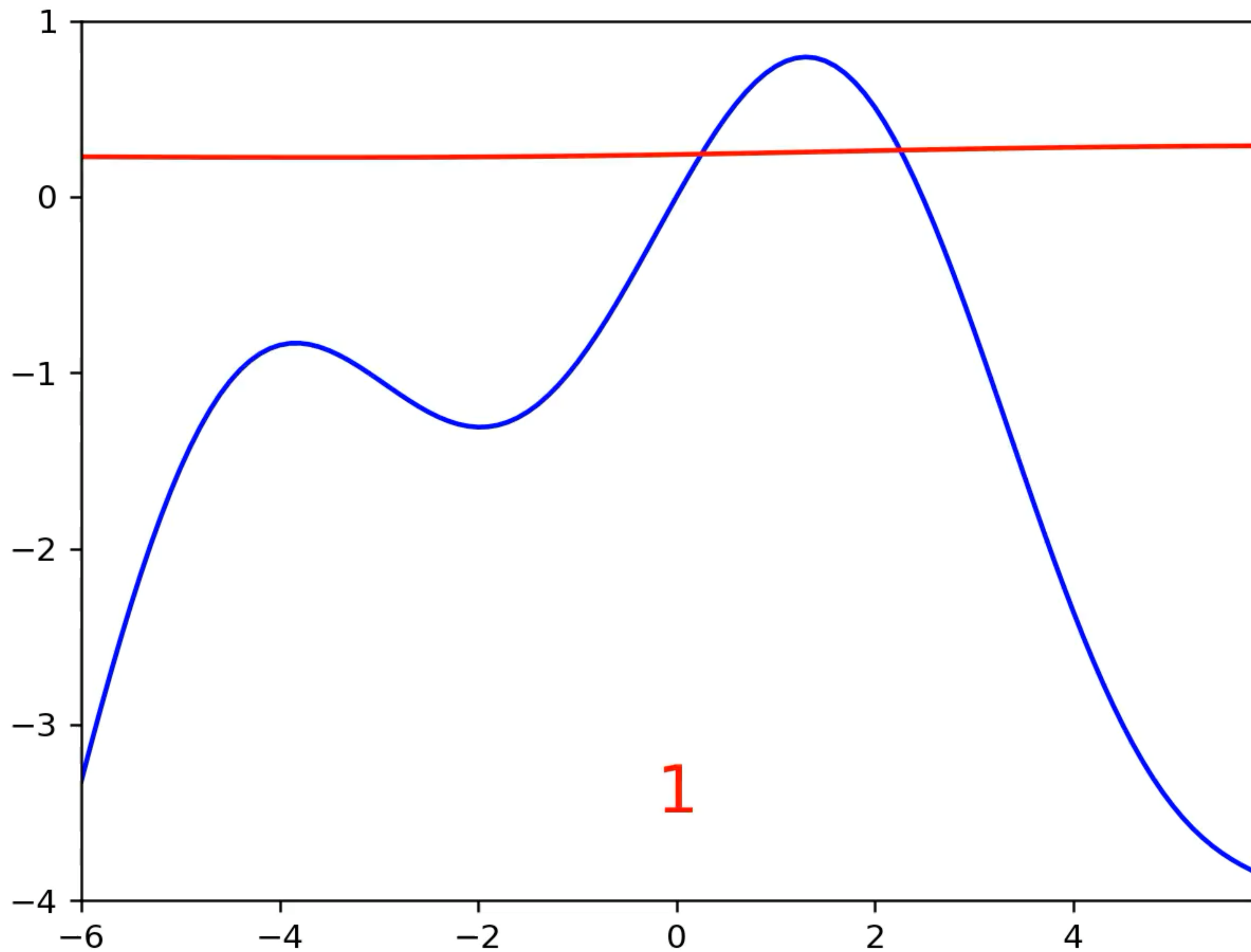
Fourier Series: $f(x) = \text{~} + \text{~} + \text{~} + \text{~} + \dots$

Taylor Series: $f(x) = \text{---} + \text{ / } + \text{ U } + \text{ ~ } + \dots$









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

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

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

red

neural network

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

0	a
1	able
2	about
	⋮
39	apex
	⋮
56,356	zenith
	⋮
92,487	zygote

word2vec

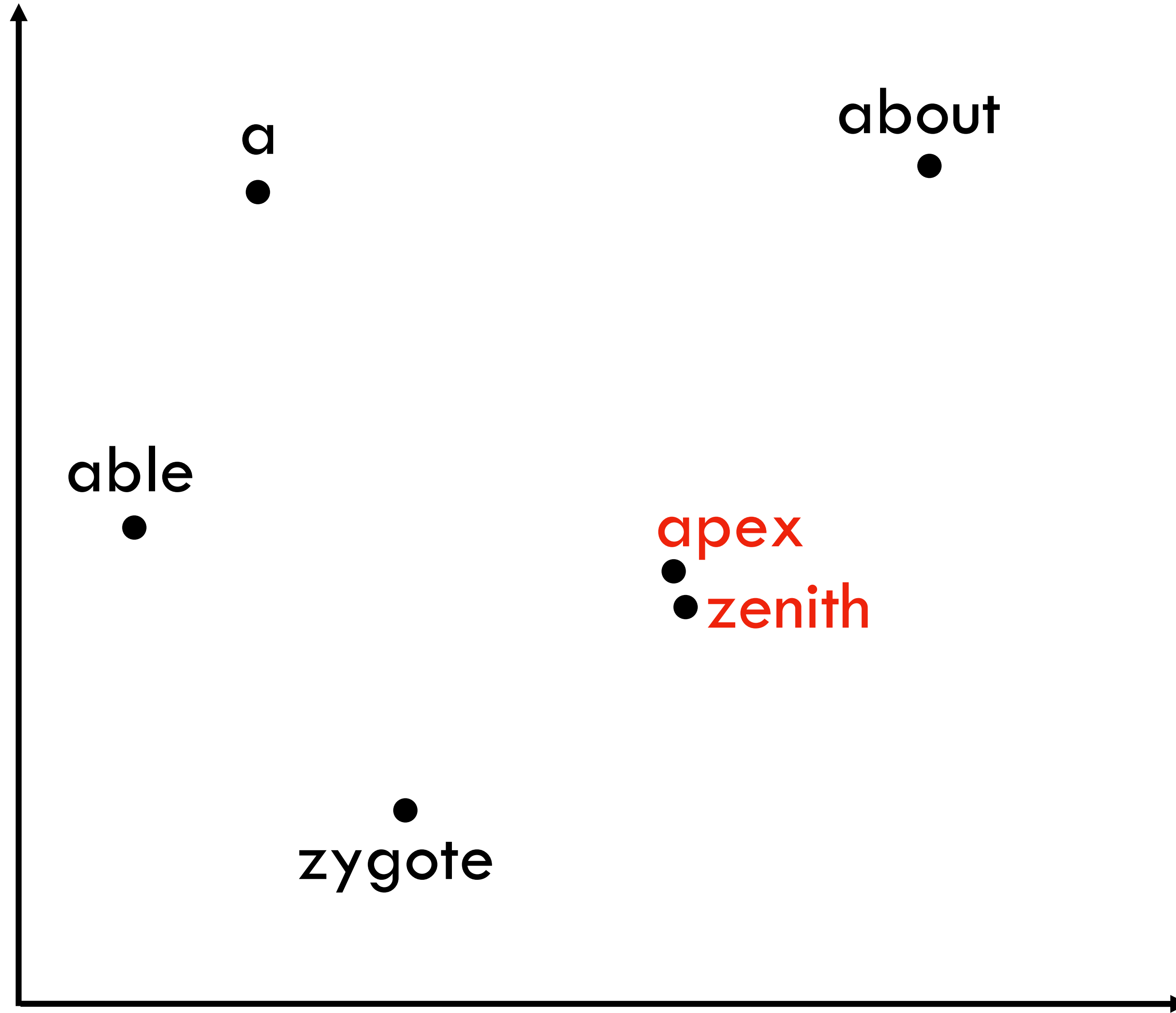
[Collobert & Weston 2008; Mikolov et al. 2013]

house, where the professor lived without his wife and child; or so he said jokingly sometimes: "Here's where I live. My house." His daughter often added, without resentment, for the visitor's information, "It started out to be for me, but it's really his." And she might reach in to bring forth an inch-high table lamp with fluted shade, or a blue dish the size of her little fingernail, marked "Kitty" and half full of eternal milk; but she was sure to replace these, after they had been admired, pretty near exactly where they had been. The little house was very orderly, and just big enough for all it contained, though to some tastes the bric-à-brac in the parlor might seem excessive. The daughter's preference was for the store-bought gimmicks and appliances, the toasters and carpet sweepers of Lilliput, but she knew that most adult visitors would



Deep
Net

The diagram illustrates word embeddings using the words 'Deep' and 'Net' as an example. A semi-transparent red rectangle is centered over the text. The words 'Deep' and 'Net' are written in red inside this rectangle. Five red arrows originate from a single point at the top center of the rectangle and point to the words 'added', 'without', 'resentment', 'for', and 'information' in the text above. Another red arrow points from the bottom center of the rectangle to the word 'daughter's' in the text below. This visualizes how a model can learn the semantic relationship between words by analyzing their context in a sentence.



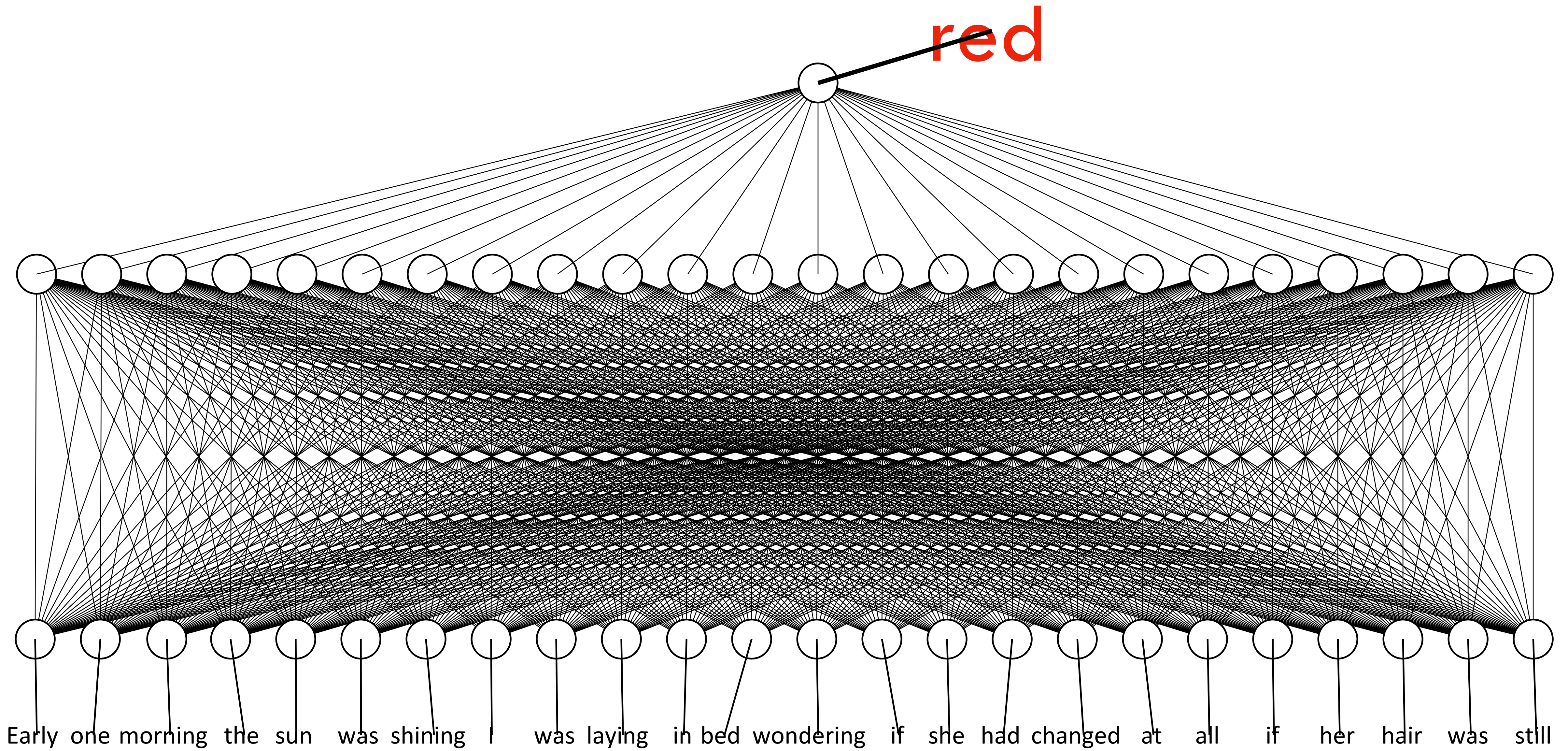
Word Embedding (e.g., word2Vec, GloVe)

slide from Steve Seitz's [video](#)

red

neural network

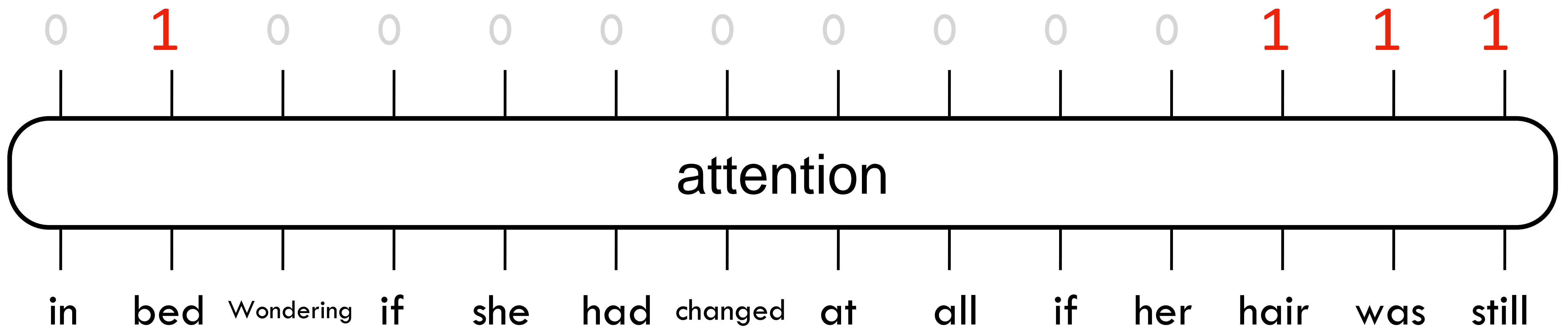
Early one morning the sun was shining I was laying in bed wondering if she had changed at all if her hair was still



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

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

_____ bed
_____ hair was still red



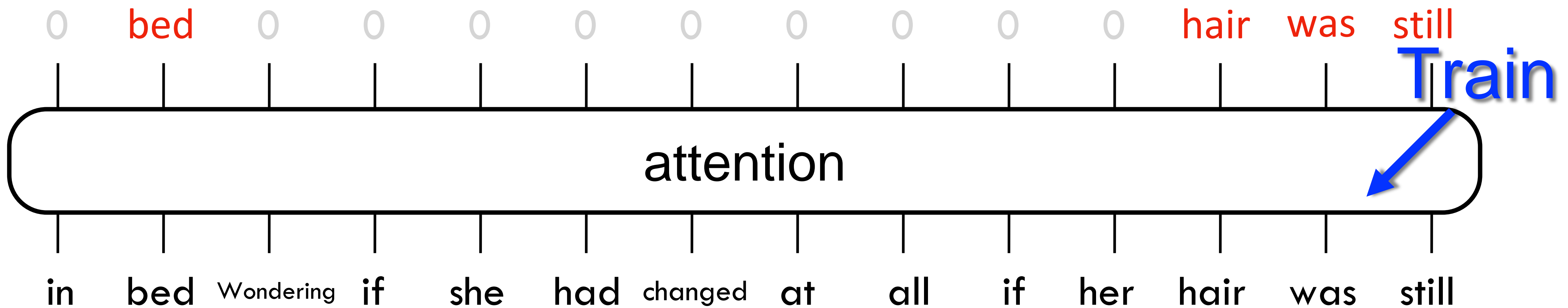
red

next word prediction

0 bed 0 0 0 0 0 0 0 0 0 hair was still

attention

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



Two roads diverted in a yellow wood
And sorry I could not travel both
And be one traveler, long I stood
And looked down as far as I could
To where it bent in the undergrowth;

Robert Frost, *Road Not Taken*

slide from Steve Seitz's [video](#)

Train



red

next word prediction

0 bed 0 0 0 0 0 0 0 0 0 hair was still

attention

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

Train

brown

next word prediction

0 bed 0 0 0 0 0 0 0 0 0 hair was still

attention

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

Train

brown

next word prediction

0 bed 0 0 0 0 0 0 0 0 0 hair was still

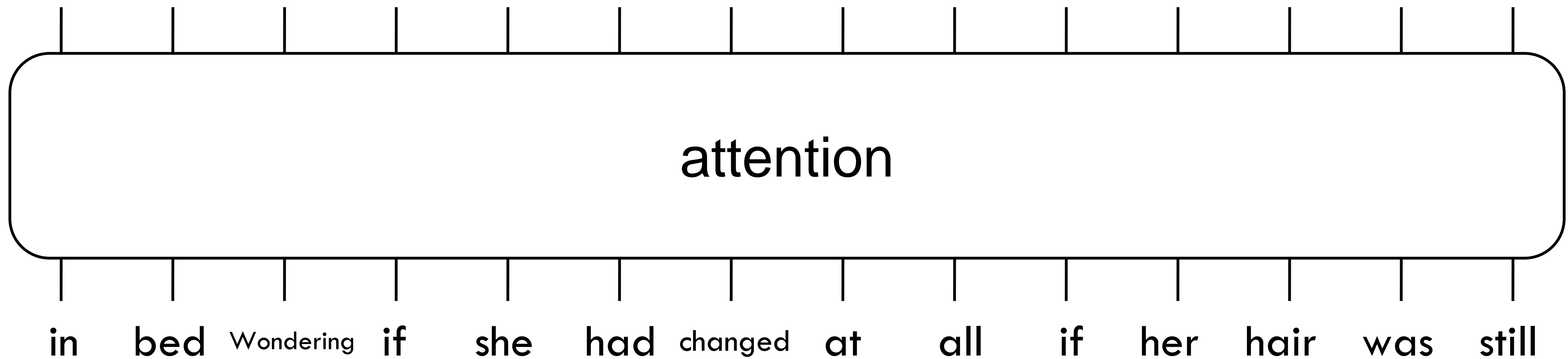
attention

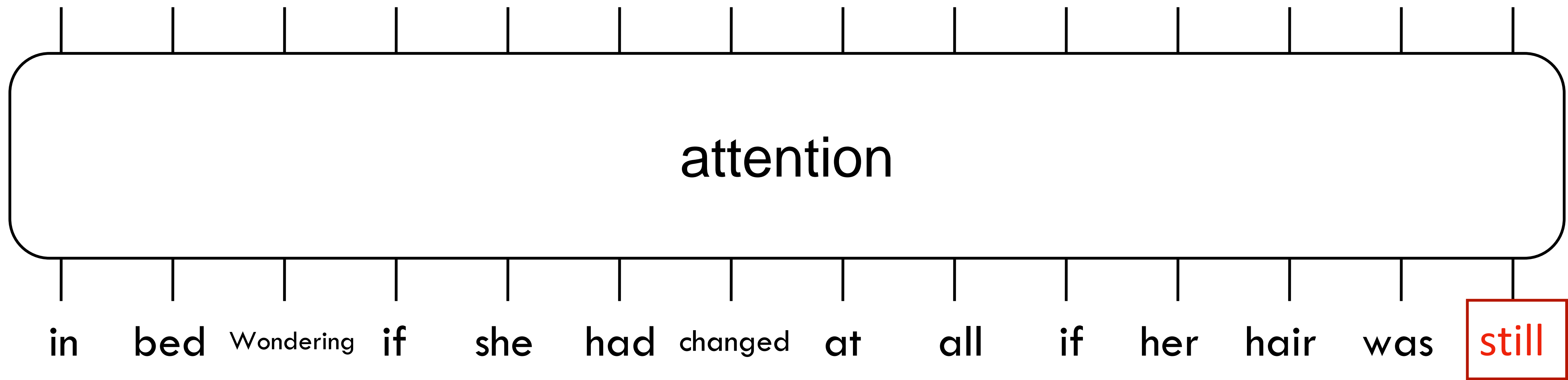
in bed Wondering if she had changed at all if her hair was still

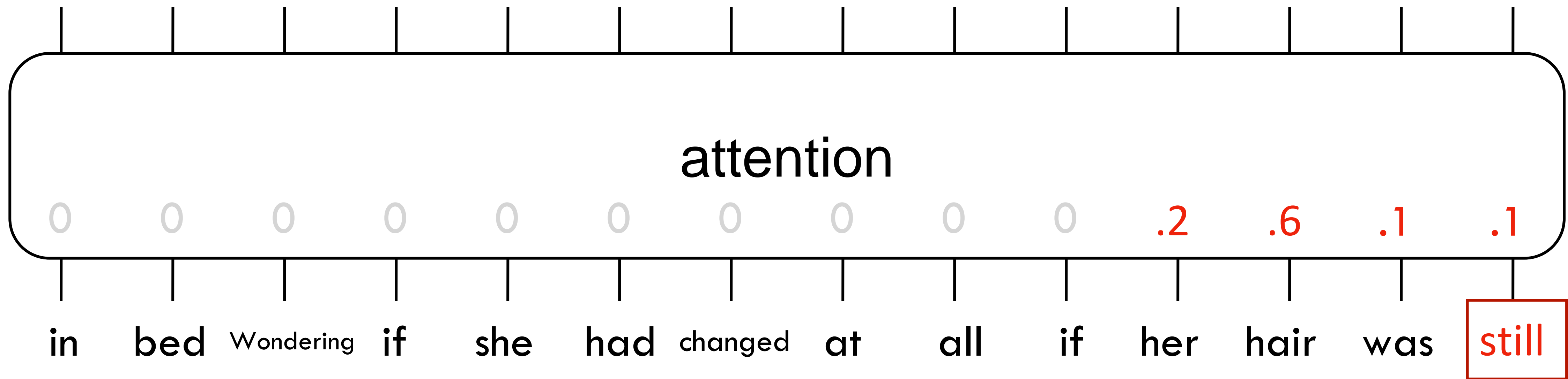
red

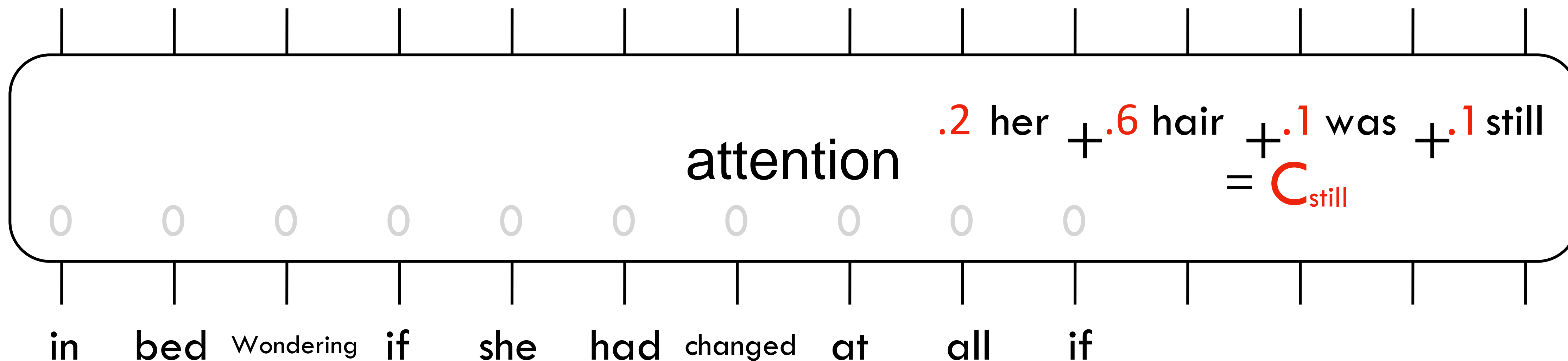
Transformer

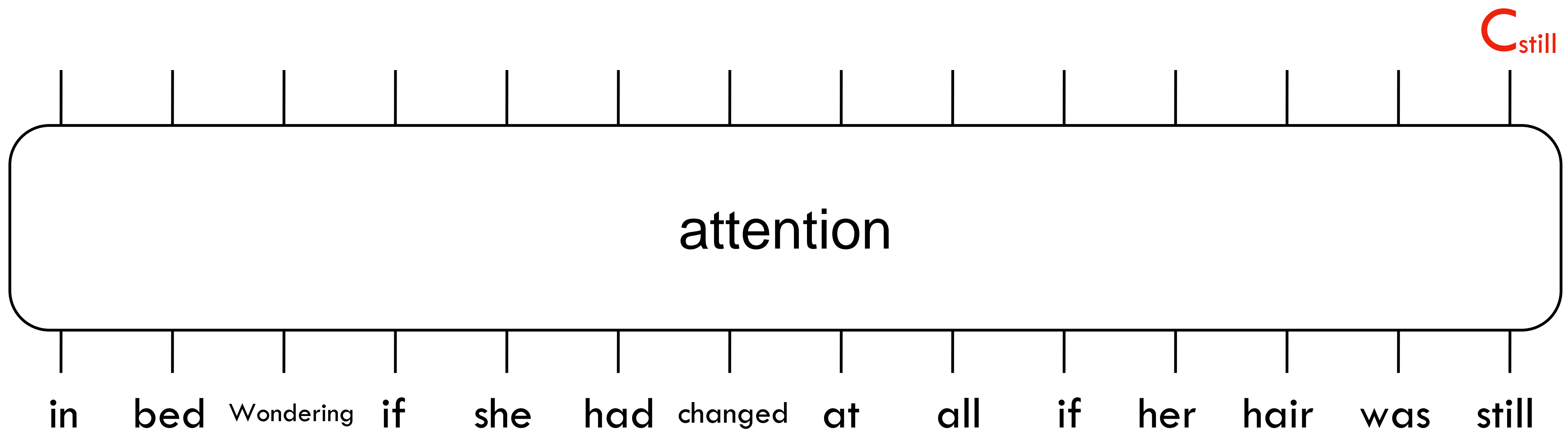
in bed Wondering if she had changed at all if her hair was still

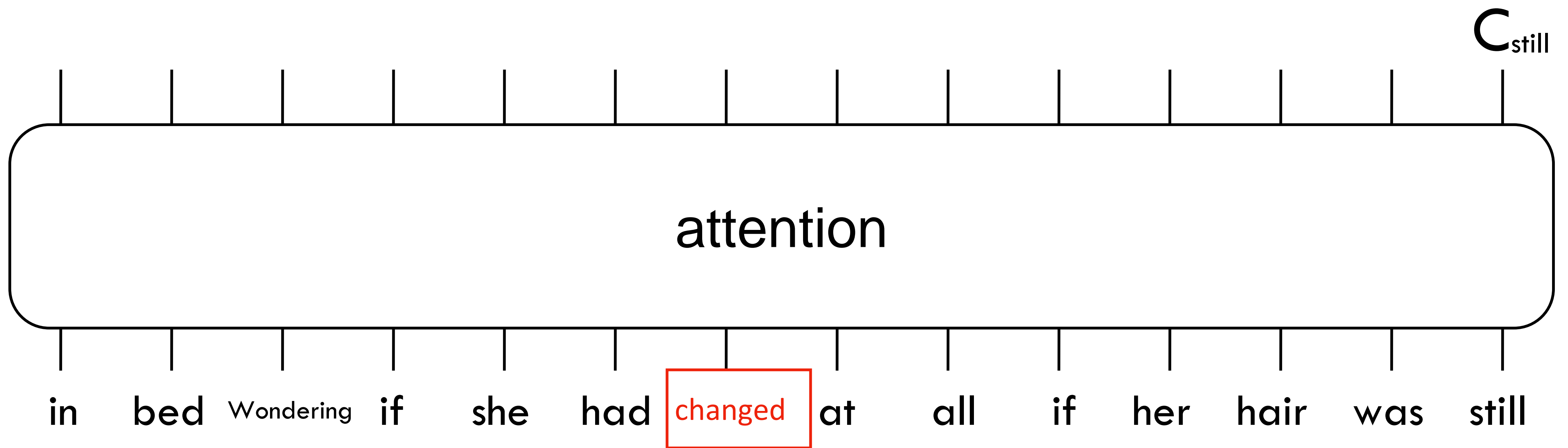


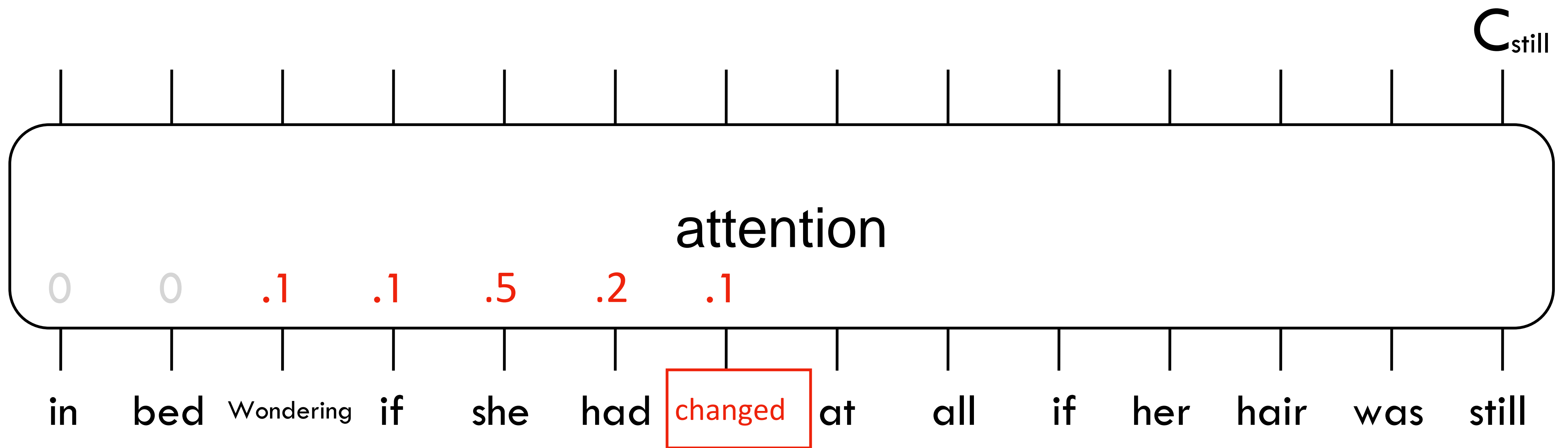












C_{still}

.1 Wondering + .1 if + .5 she + .2 had + .1 changed = C_{changed}

0

0

in

bed

at

all

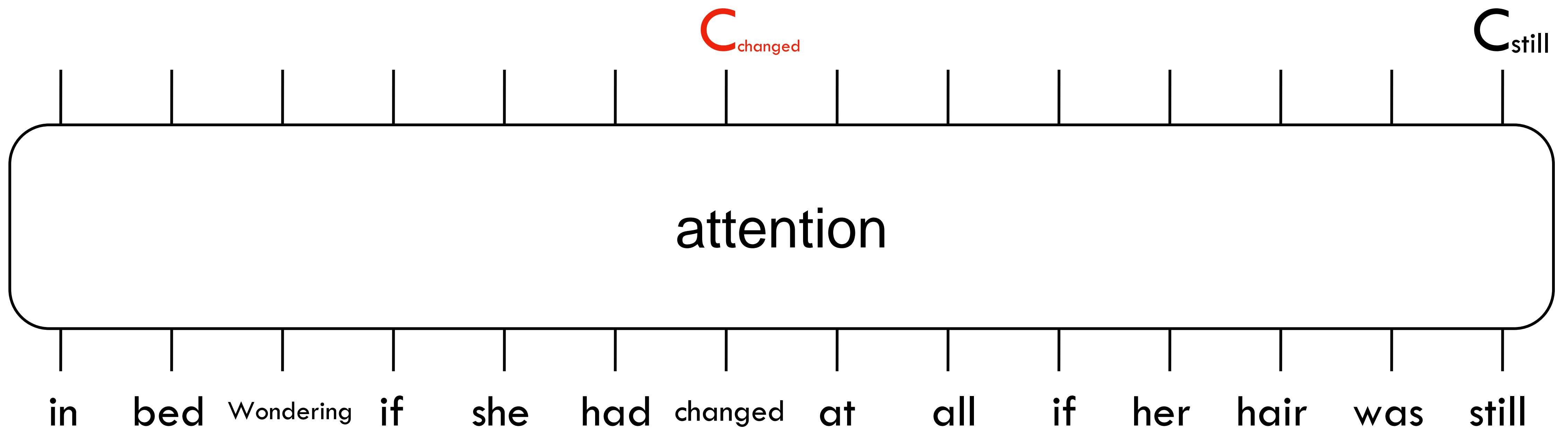
if

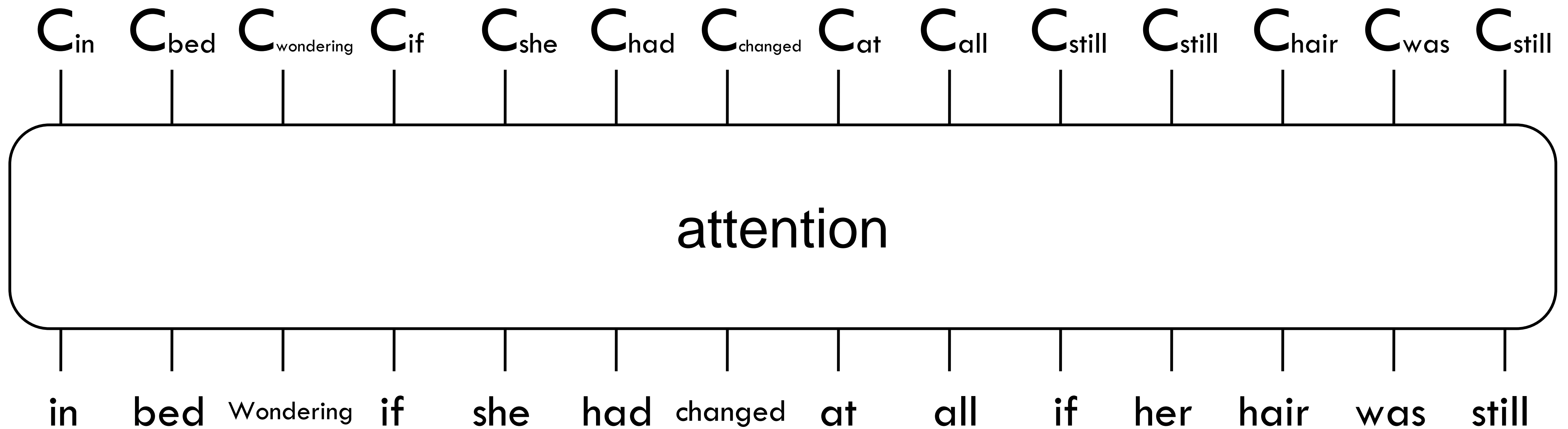
her

hair

was

still





prediction

C_{in} C_{bed} C_{wondering} C_{if} C_{she} C_{had} C_{changed} C_{at} C_{all} C_{still} C_{still} C_{hair} C_{was} C_{still}

attention

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

prediction

C_{in} C_{bed} C_{wondering} C_{if} C_{she} C_{had} C_{changed} C_{at} C_{all} C_{still} C_{still} C_{hair} C_{was} C_{still}
in bed Wondering if she had changed at all if her hair was still

attention

a

prediction

attention

It's

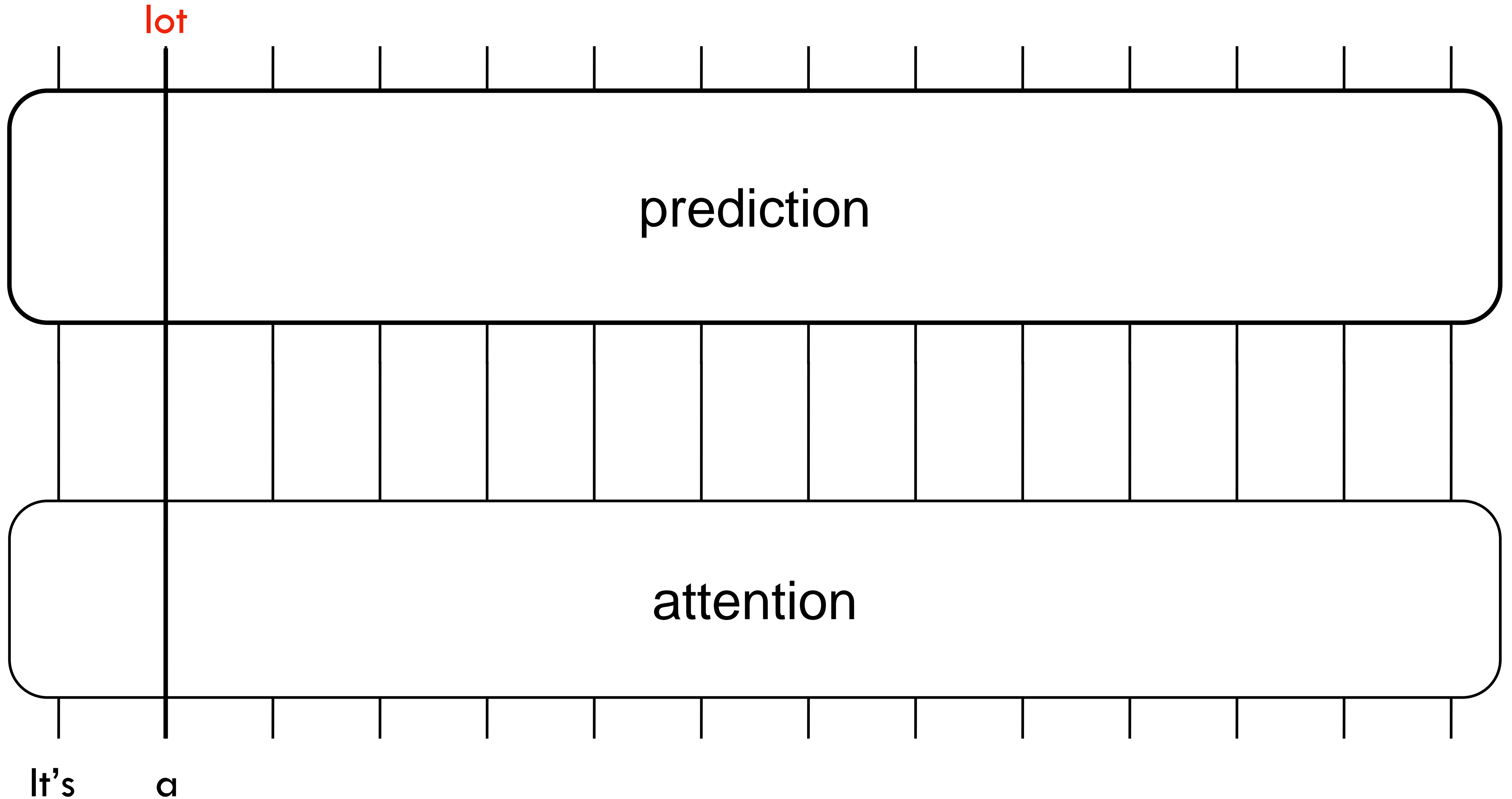
a	the	looking	possible	getting
0.4	0.3	0.1	0.1	0.1

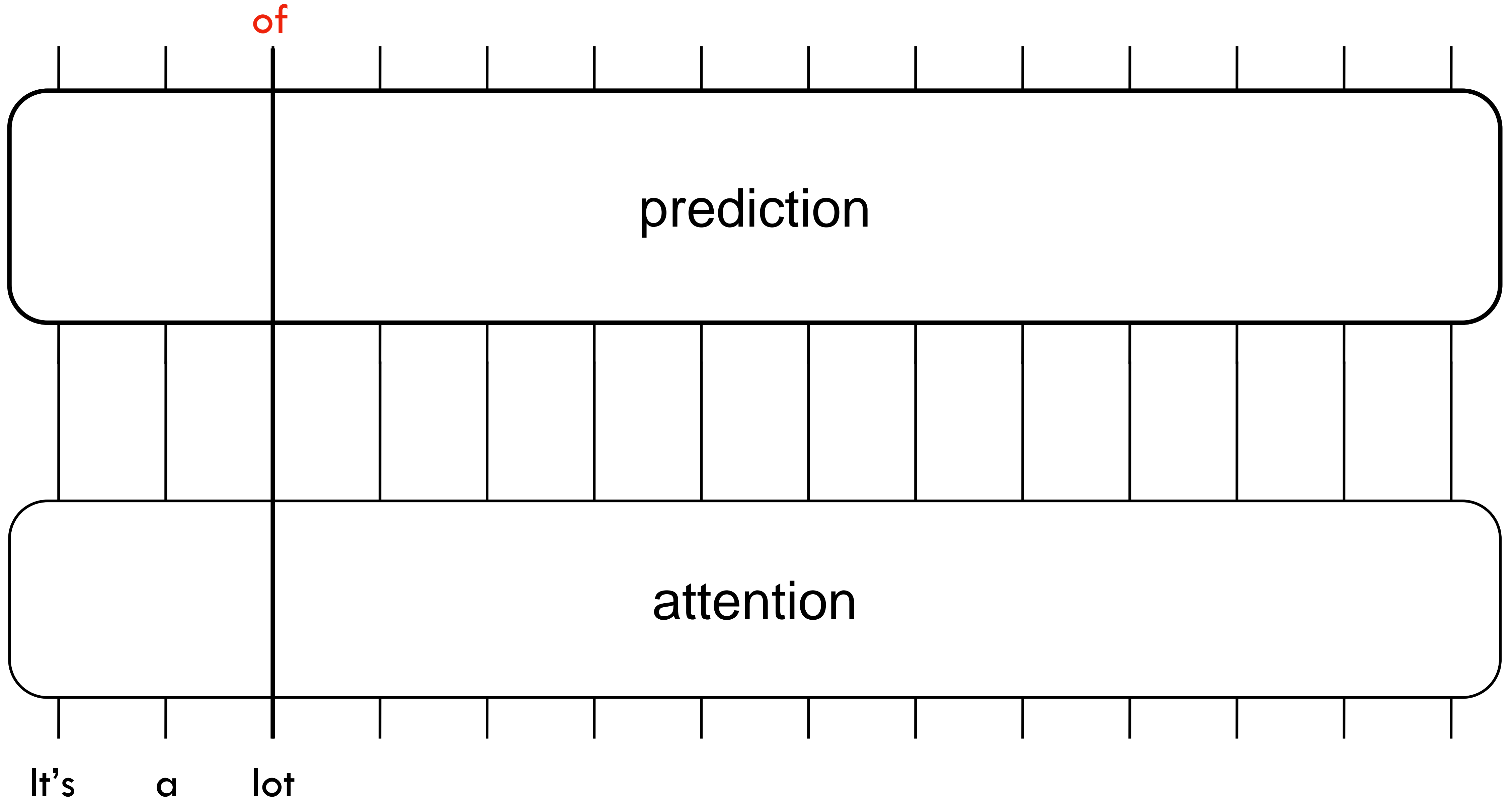
a

prediction

attention

It's



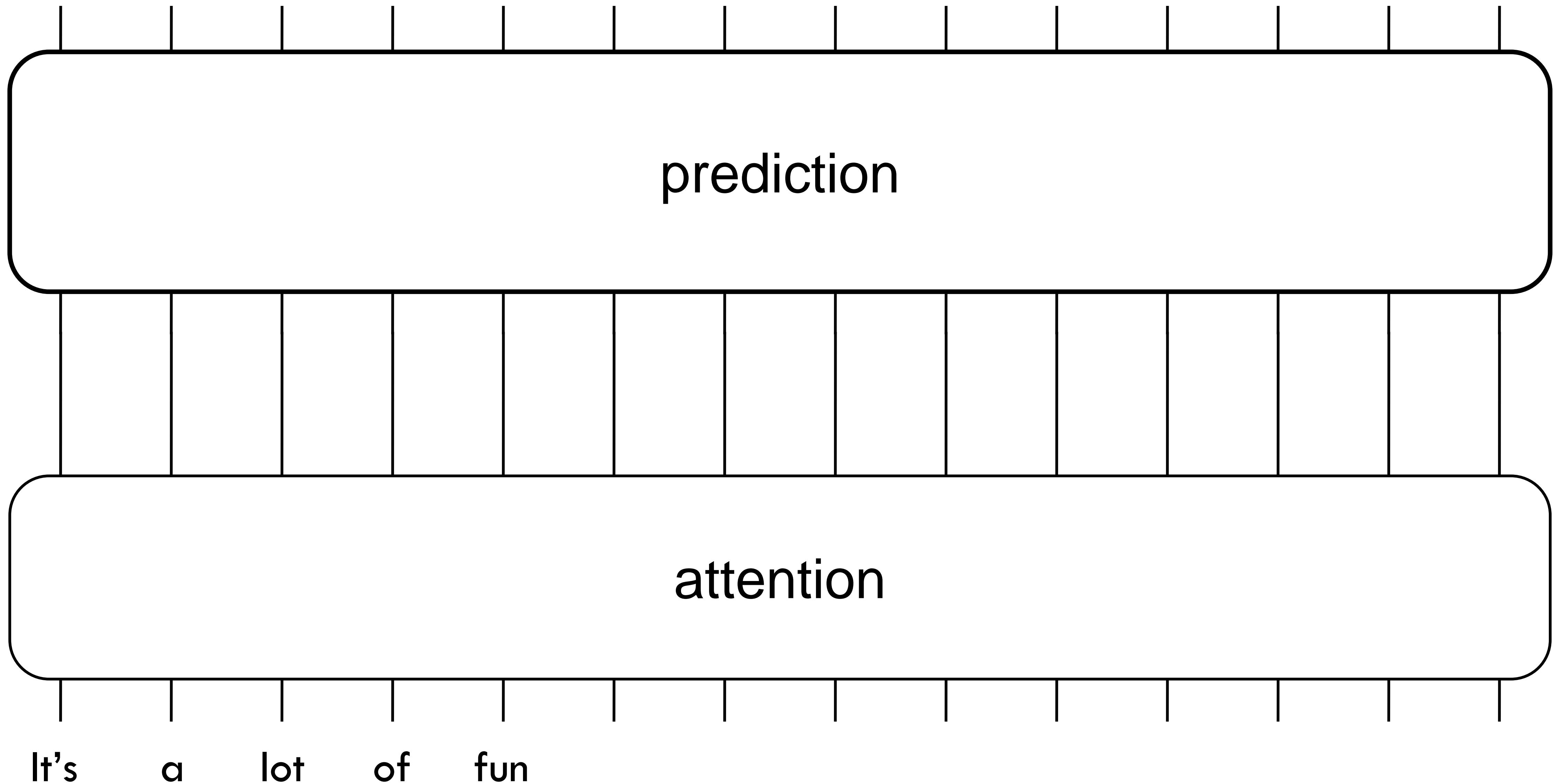


fun

prediction

attention

It's a lot of



Abraham

prediction

attention

The 16th
president was

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 ?

prediction

attention

⋮

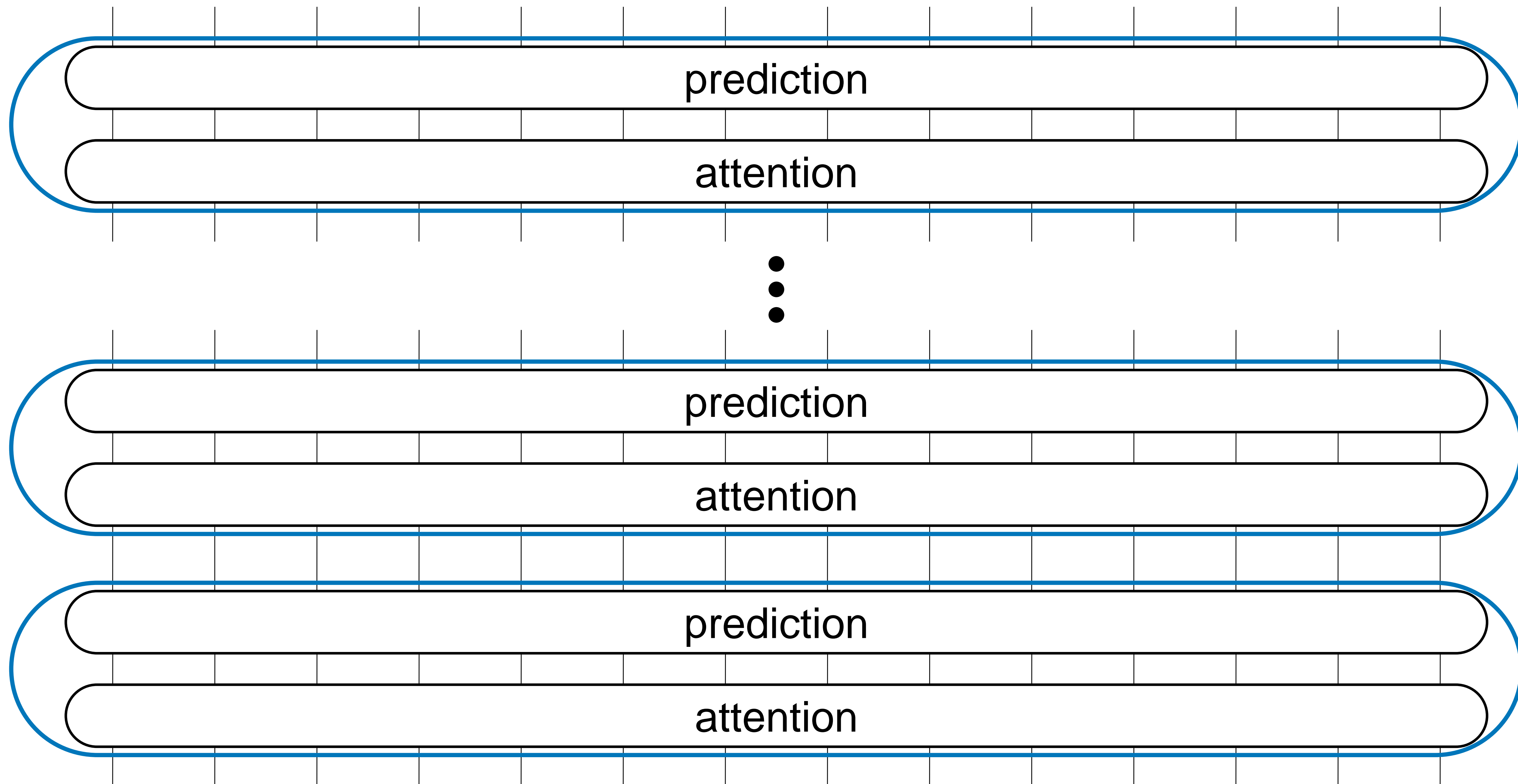
96 (GPT-3) **118** (Palm)

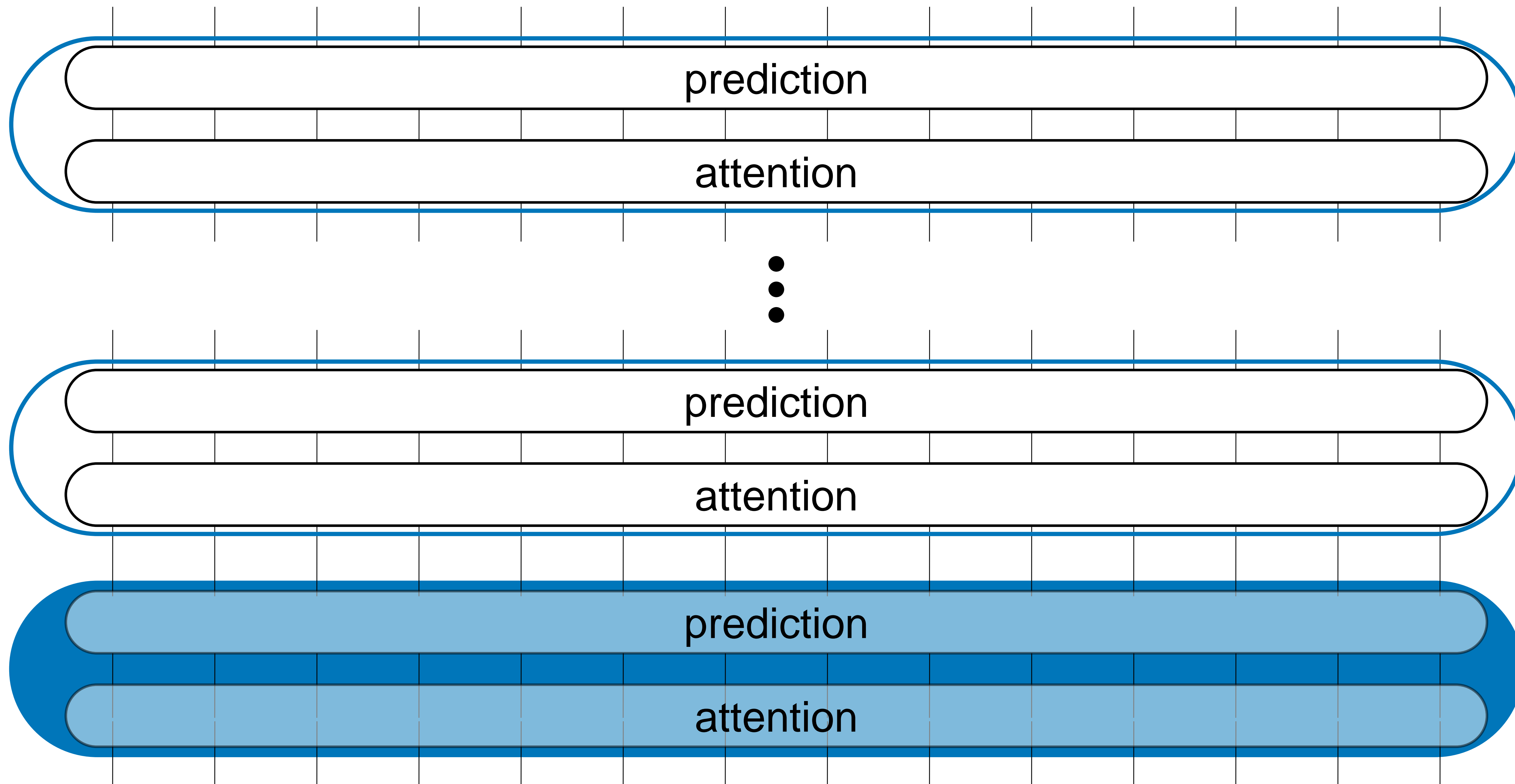
prediction

attention

prediction

attention

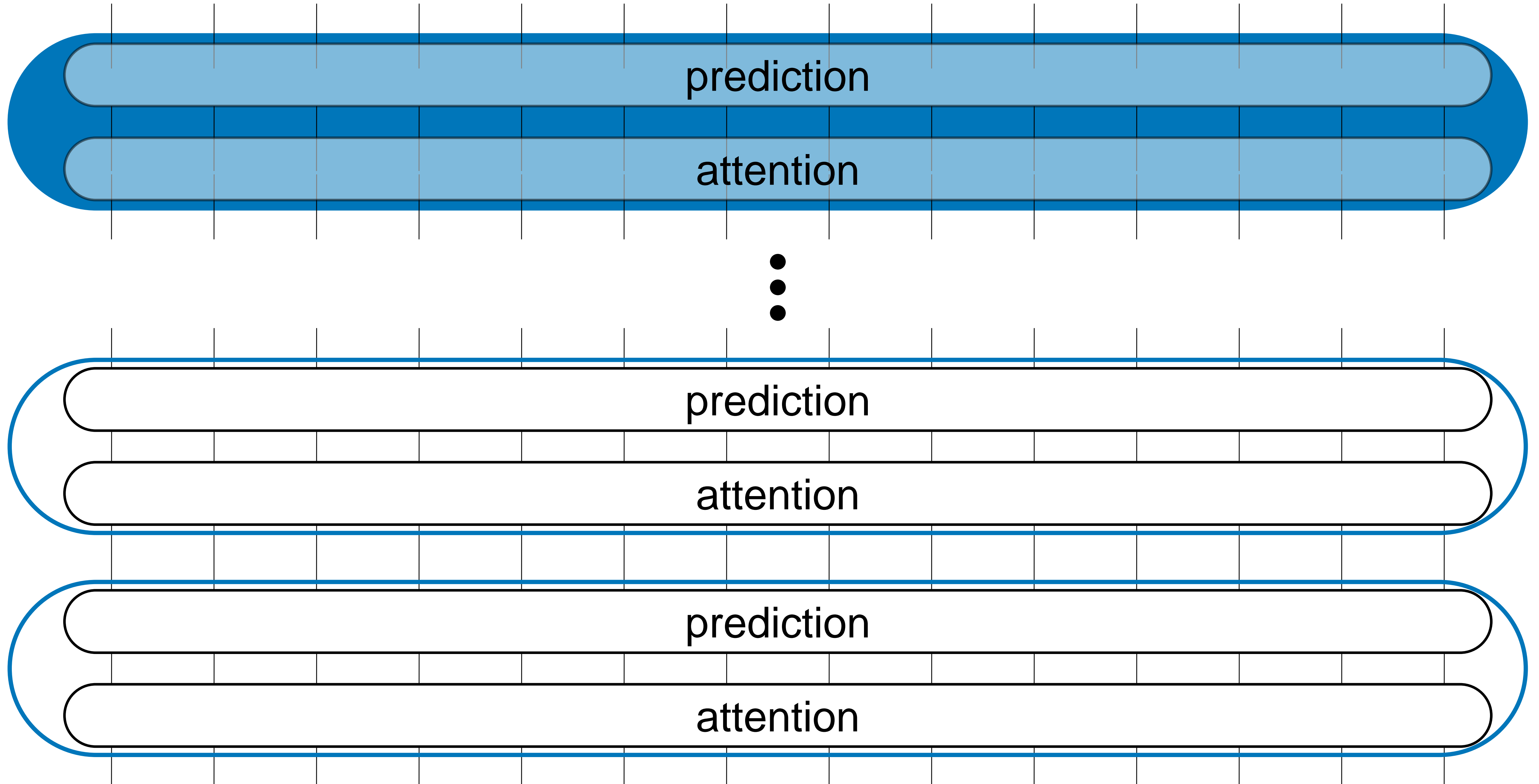


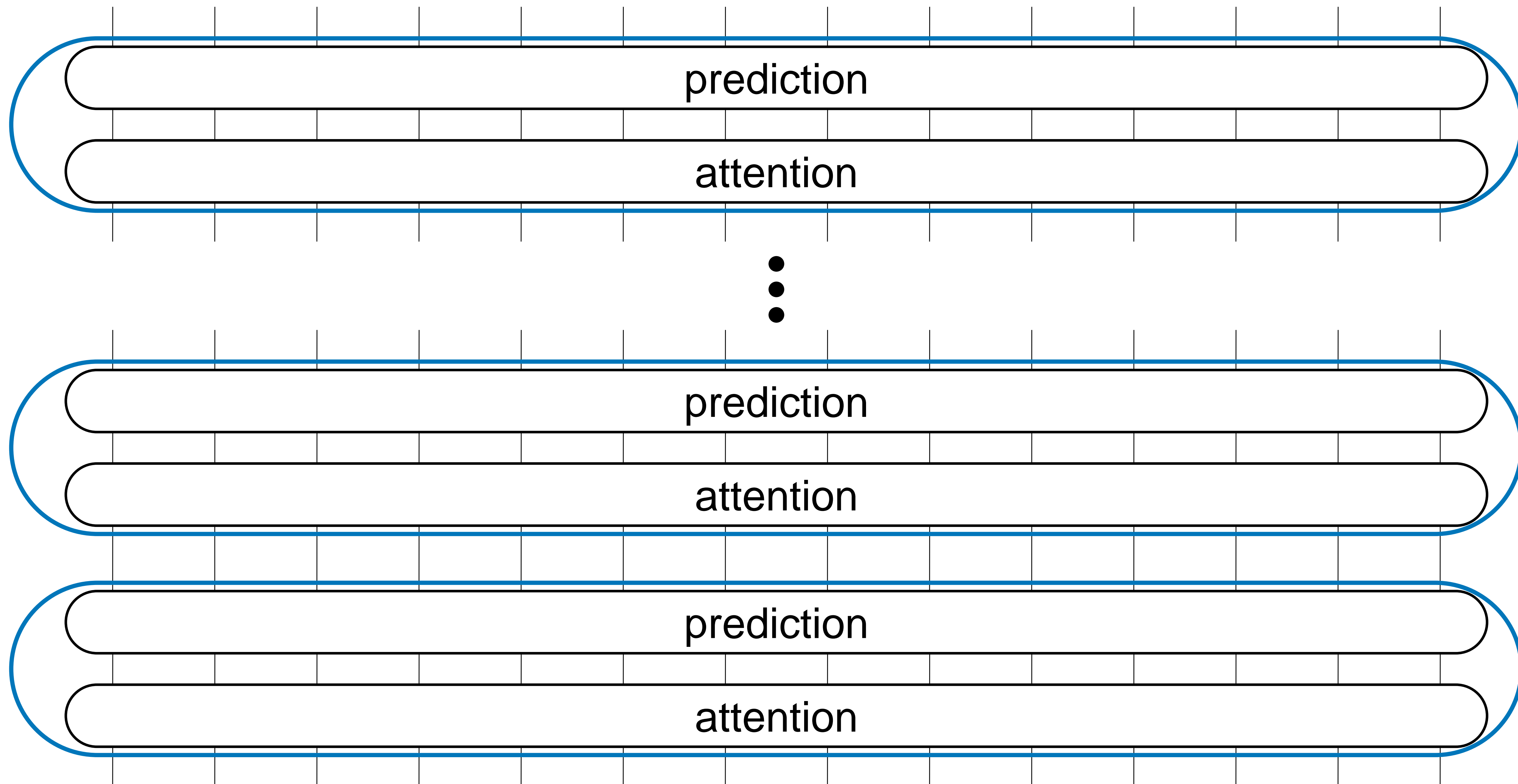


Syntax

slide from Steve Seitz's [video](#)

Semantics





How much data
to train?

All of it...

355 years

a month

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

The 16th President was Abraham Lincoln

The capital of Zimbabwe is Harare

Frank Zappa's middle name is Vincent

Napoleon was born on this date 1769

The prime factorization of 19456721434 is $2 \times 3 \times 3 \times 17$

Queen Victoria's maiden name was Alexandrina Victoria

US per-capita income in 1957 was \$2,974

The lat long coordinates of Rome are 41.894722, 12.48

a haiku about a turtle and a stop sign

a haiku about a turtle and a stop sign

Turtle at the stop sign
waiting for the perfect moment
to cross the busy road

translate “tangled up in blue” to 5 languages

translate “tangled up in blue” to 5 languages

1. Spanish: Enredado en azul
2. French: Emmêlé dans le bleu
3. Italian: Intrecciato nel blu
4. German: Verwickelt in blau
5. Japanese: 青い糸に絡まって

the number after the number after 37 is 74

I'm heading north. I turn right. Then I turn left.

I'm now heading south

the recipe for chocolate guacamole pancakes

Ingredients:

- 1 ripe avocado, mashed
- 1/4 cup cocoa powder
- 1/4 cup all-purpose flour
- 1 teaspoon baking powder
- 1/4 teaspoon salt
- 1/2 cup milk
- 1 egg
- 1 tablespoon vegetable oil
- 1/4 cup chocolate chips

Directions:

1. In a medium bowl, combine the avocado, cocoa powder, flour, baking powder, and salt.
2. In a separate bowl, whisk together the milk, egg, and vegetable oil.
3. Add the wet ingredients to the dry ingredients and mix until well combined.
4. Fold in the chocolate chips.
5. Heat a large skillet over medium heat.
6. Scoop 1/4 cup batter onto the skillet for each pancake.
7. Cook for 2-3 minutes per side, or until the pancakes are golden brown.
8. Serve with your favorite toppings.

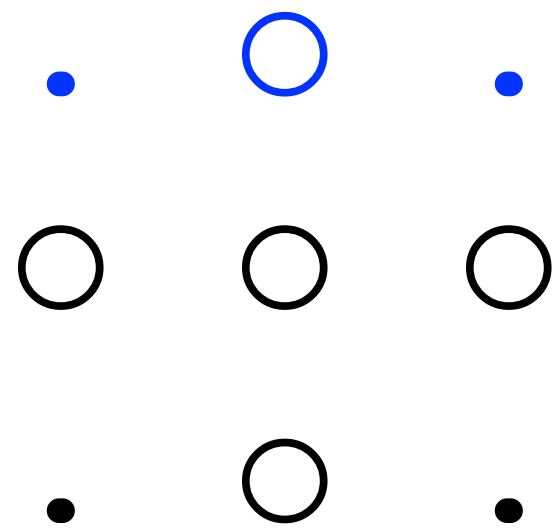


For more details:

<https://jalammar.github.io/illustrated-transformer/>
<https://jalammar.github.io/illustrated-gpt2/>

slide from Steve Seitz's [video](#)

a pattern of characters that looks like a star



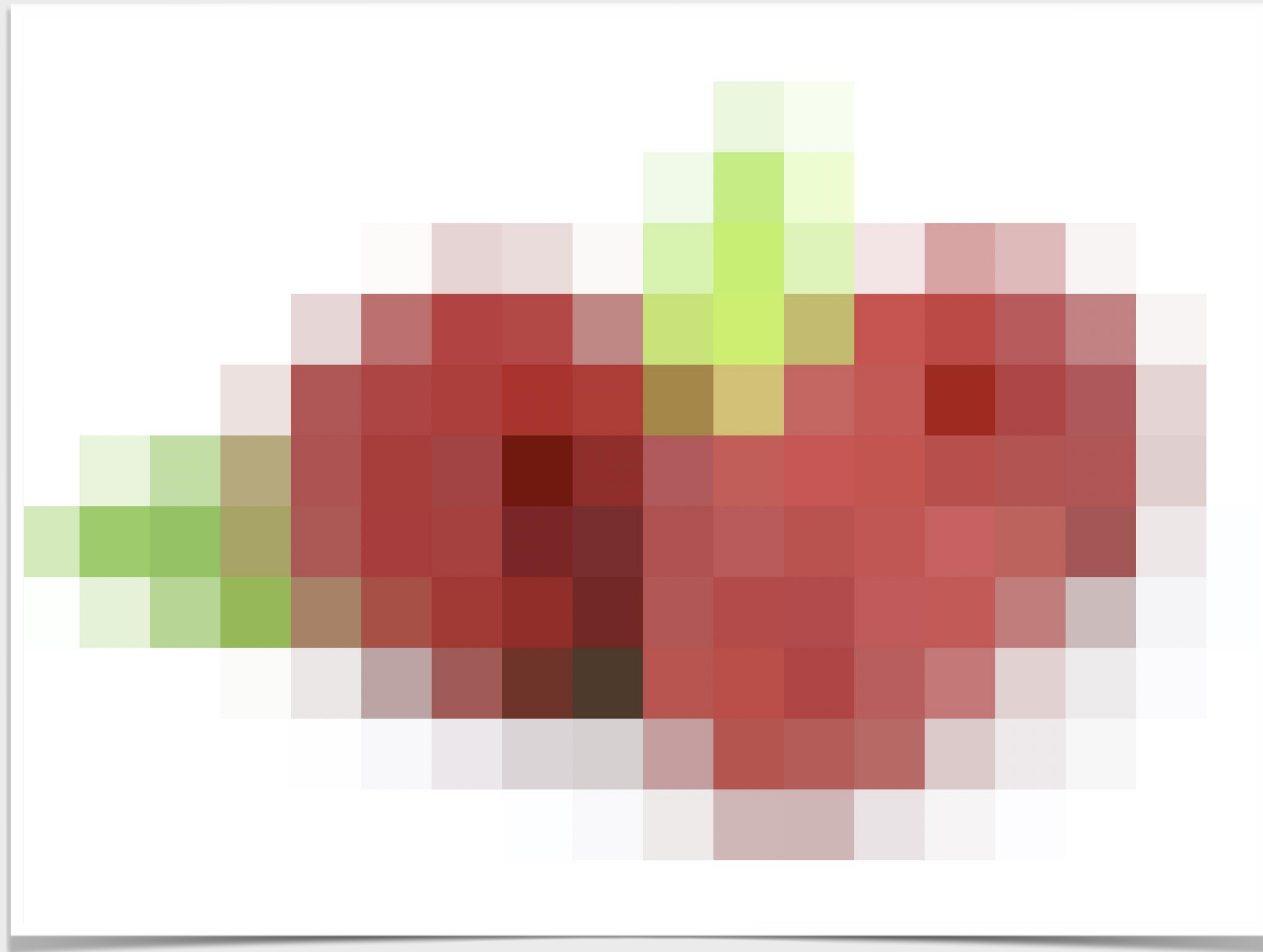
a pattern of characters that looks like a vertical line

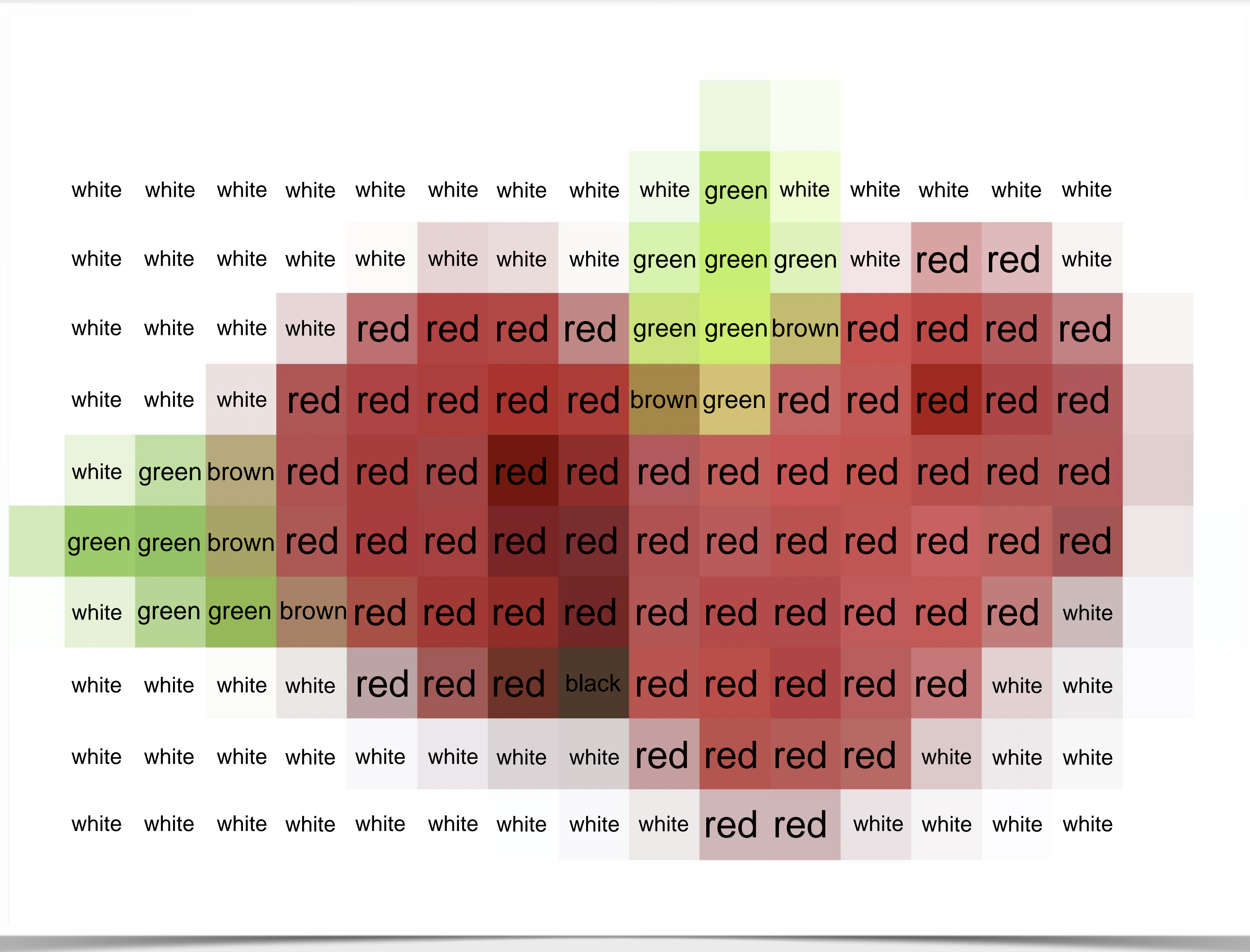
• ○ •
• ○ •
• ○ •
• ○ •
• ○ •

a pattern of characters that looks like a triangle

• ○ •
• ○ ○ •
• ○ ○ ○ •
• ○ ○ ○ ○ •
• ○ ○ ○ ○ ○ •







white white white white white white white white white green white white white white white
white white white white white white white white green green green white red red white
white white white white red red red red green green brown red red red red
white white white red red red red red brown green red red red red red red
white green brown red red red red red red red red red red red red red
green green brown red red red red red red red red red red red red red
white green green brown red red red red red red red red red red red white
white white white white red red red black red red red red red white white
white white white white white white white white red red red red white white white
white white white white white white white white white white red red white white white white

(255,0,0)

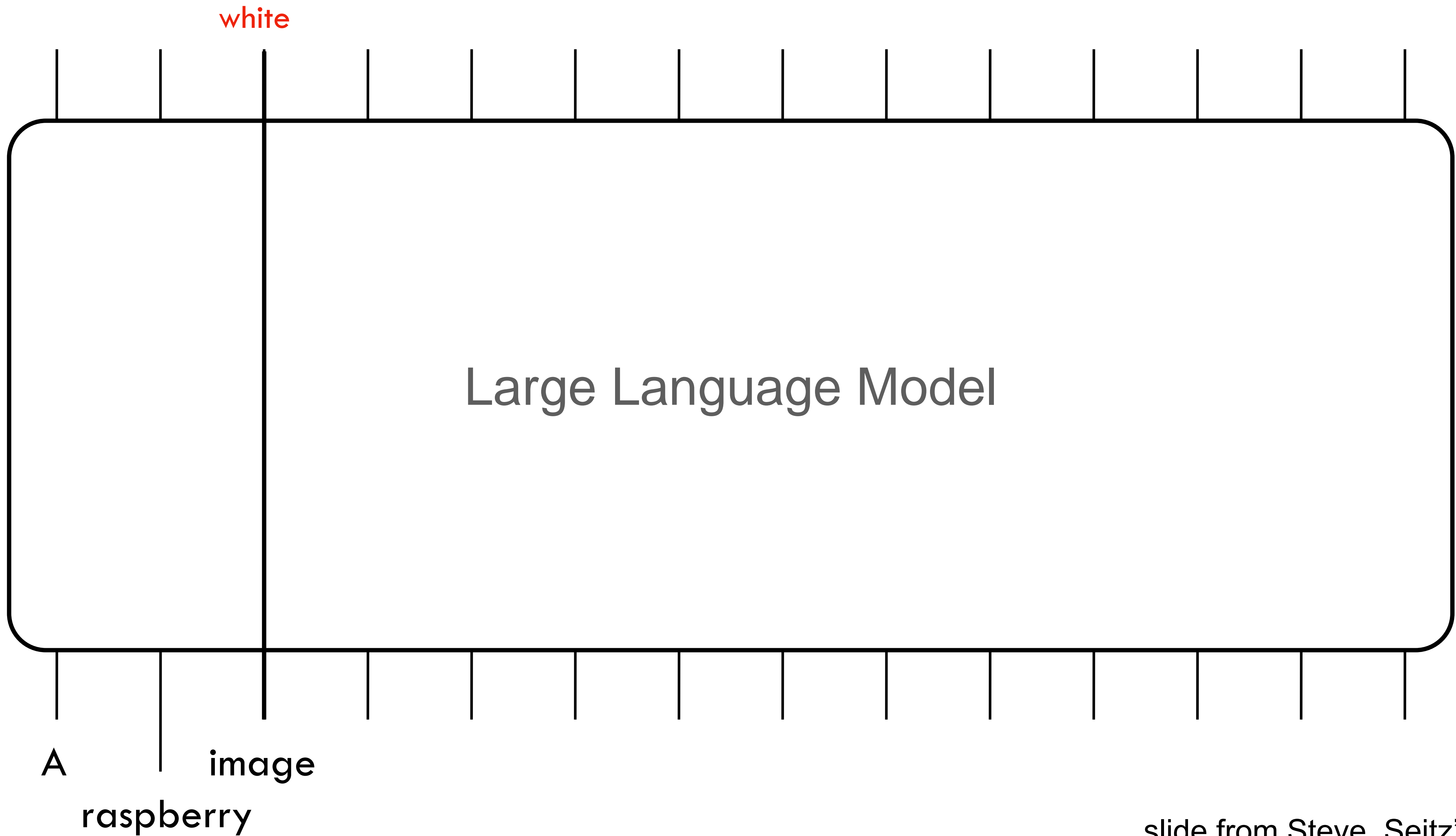
white white white white white white white white white green white white white white white
white white white white white white white white green green green white red red white
white white white white red red red red green green brown red red red red
white white white red red red red red brown green red red red red red red
white green brown red red red red red red red red red red red red red red
green green brown red red red red red red red red red red red red red red
white green green brown red red red red red red red red red red red white
white white white white red red red black red red red red red white white
white white white white white white white white red red red red white white white
white white white white white white white white white white red red white white white white



raspberries pancakes sunsets

1 Billion

slide from Steve Seitz's [video](#)



The diagram illustrates the architecture of a Large Language Model (LLM) within a grid-like frame. The frame is divided into four main sections by a vertical line and a horizontal line. The top-left section is labeled 'Input' and contains a sequence of tokens: 'The', 'cat', 'sat', 'on', 'the', 'mat'. The top-right section is labeled 'Encoder' and contains a sequence of tokens: 'The', 'cat', 'sat', 'on', 'the', 'mat'. The bottom-left section is labeled 'Decoder' and contains a sequence of tokens: 'The', 'cat', 'sat', 'on', 'the', 'mat'. The bottom-right section is labeled 'Output' and contains a sequence of tokens: 'The', 'cat', 'sat', 'on', 'the', 'mat'. The text 'Large Language Model' is centered in the middle of the diagram.

The diagram illustrates the architecture of a Large Language Model (LLM). It is divided into two main horizontal sections. The top section, labeled "Input Tokens", shows a sequence of 12 tokens: "The", "cat", "sat", "on", "the", "mat", "and", "the", "dog", "was", "barking", and "loudly". The bottom section, labeled "Hidden States", shows a corresponding sequence of 12 hidden states, represented by rounded rectangles. Each hidden state is connected to its corresponding input token by a vertical arrow pointing downwards. The hidden states are arranged in a sequence, with the first state connected to "The" and the last state connected to "loudly". The entire diagram is enclosed in a black border.

red

Large Language Model

A

image white white

raspberry

Large Language Model

A

image

white

white

red

red

red

white

white

green

green

green

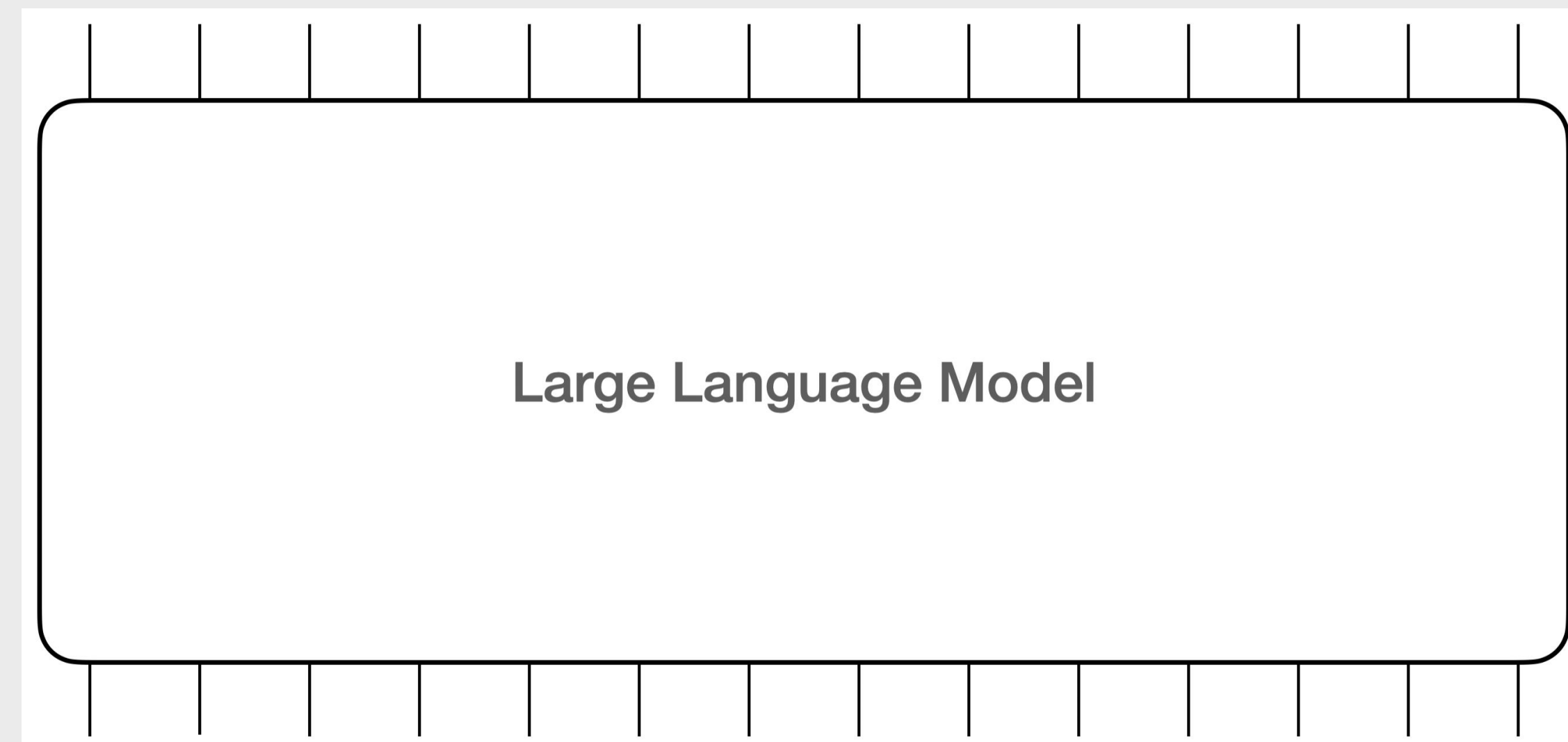
white

raspberry

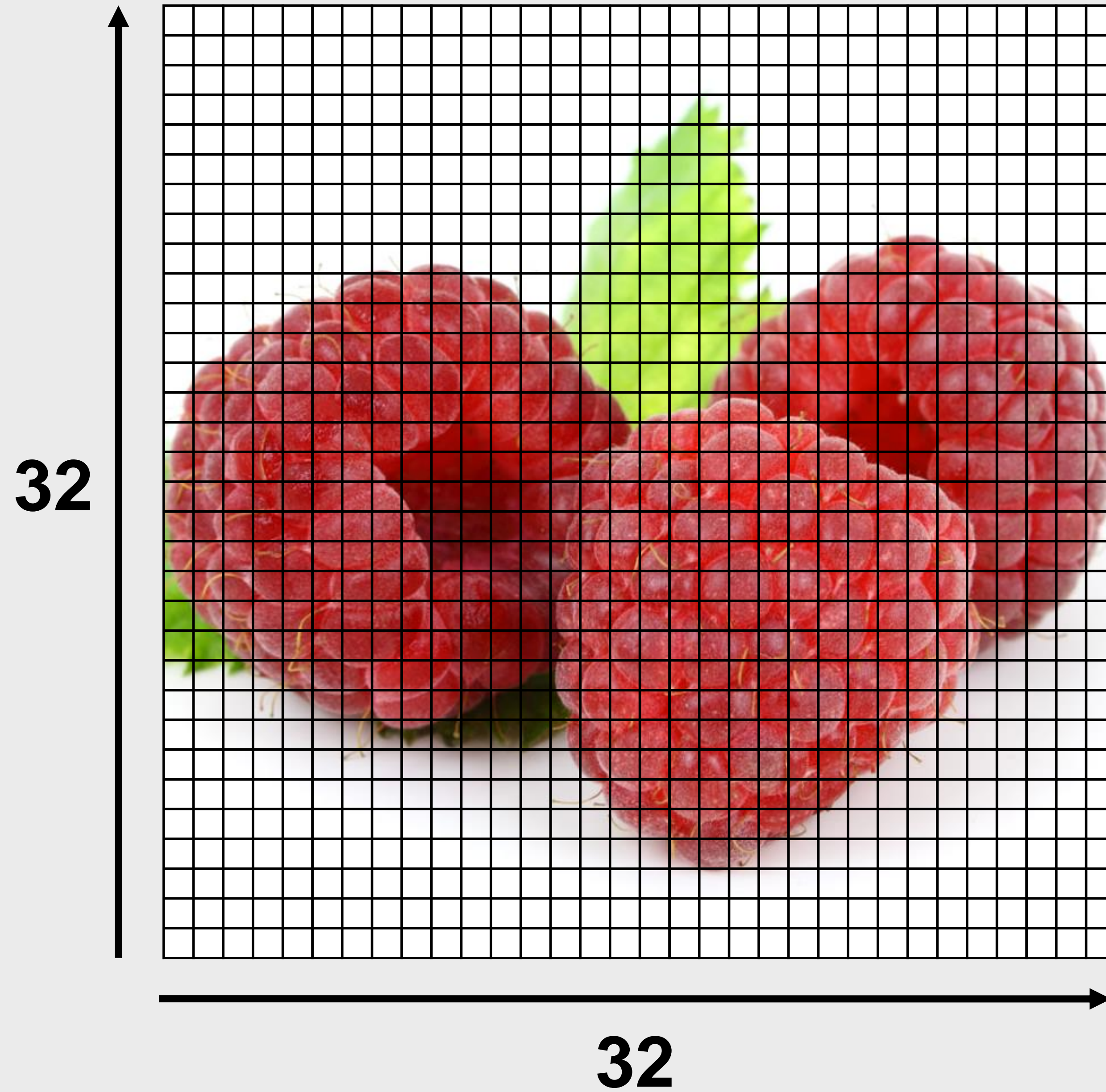
slide from Steve Seitz's [video](#)



1,000,000s of pixels



1,000s of words



$$32 \times 32 = 1024$$

Visual words

[illegible]

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

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	<div></div>															

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	<div></div>															

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	1	1	1	1	1	1	1	1	1	1	1	1	1	1	

squirrel reaching for a nut

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													

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															

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															

squirrel reaching for a nut

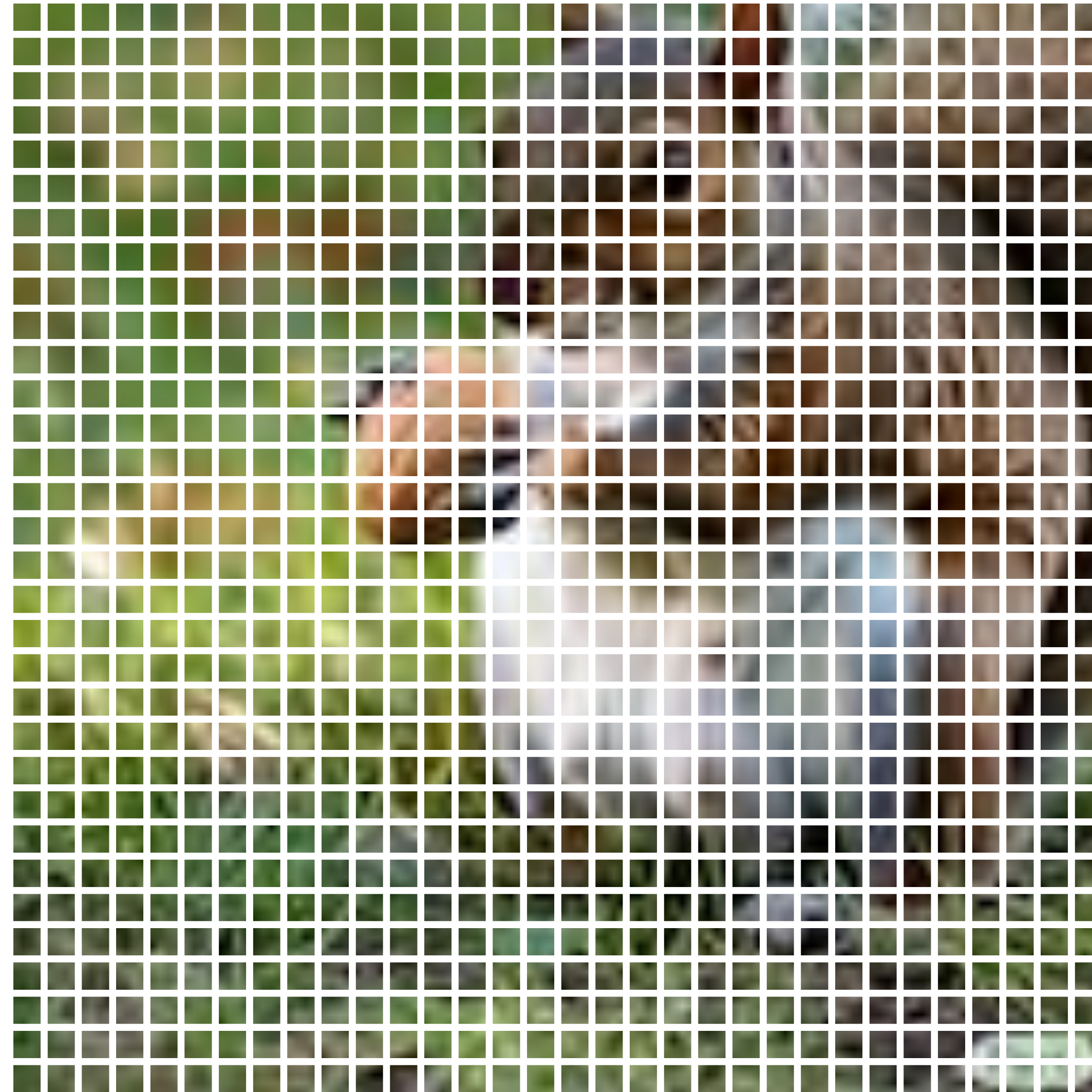
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 6 6 6 1 1 6 6 4 4 4 9 9 9 9 9 9
1 1 1 1 1 7 7 1 1 1 1 1 1 1 1 1 6

squirrel reaching for a nut

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	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	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				
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				
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				
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				
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	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	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	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	1	0	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	1	0	0	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	1	0	0	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	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	3	4	4	4	3	3	3	5	3	0	4	4	4	4	4	4	4	9	9	9	4	3	8	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	8		
3	3	8	3	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	3	8	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	8		
3	3	8	8	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	8	3	8	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	3	8	3	8	8	8	8	8	8		
3	3	8	3	3	3	3	8	3	8	8	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	8	8	8	8	8	8	8		

squirrel reaching for a nut

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



squirrel reaching for a nut



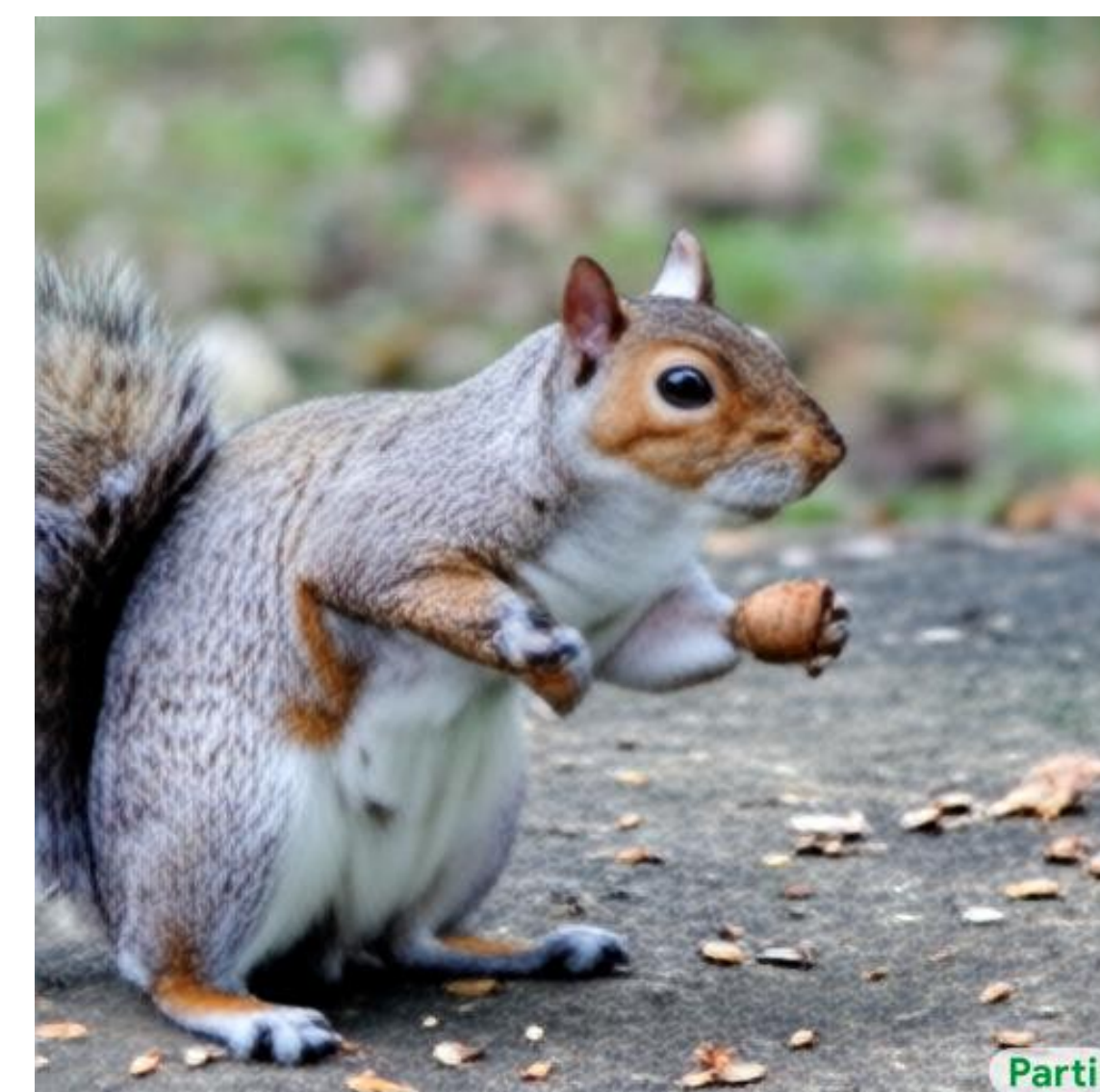
squirrel reaching for a nut



**Up-sampled
4x**

squirrel reaching for a nut

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](#)

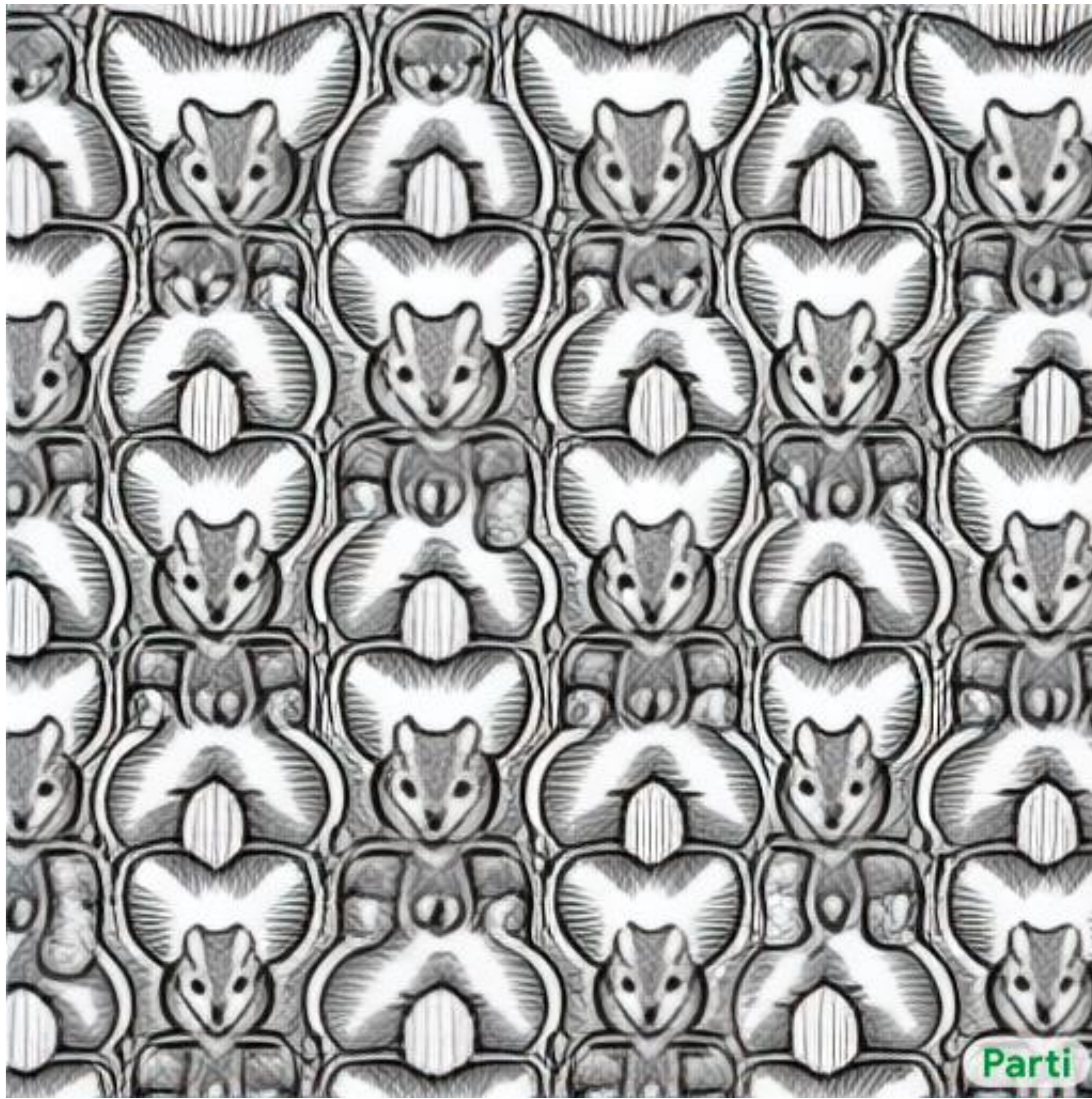


Squirrel reaching for a nut. Van Gogh painting

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](#)



Imagen



Dall-E 2

next time: diffusion