Natural Language Processing

Berkeley

NLP

Historical Document Transcription

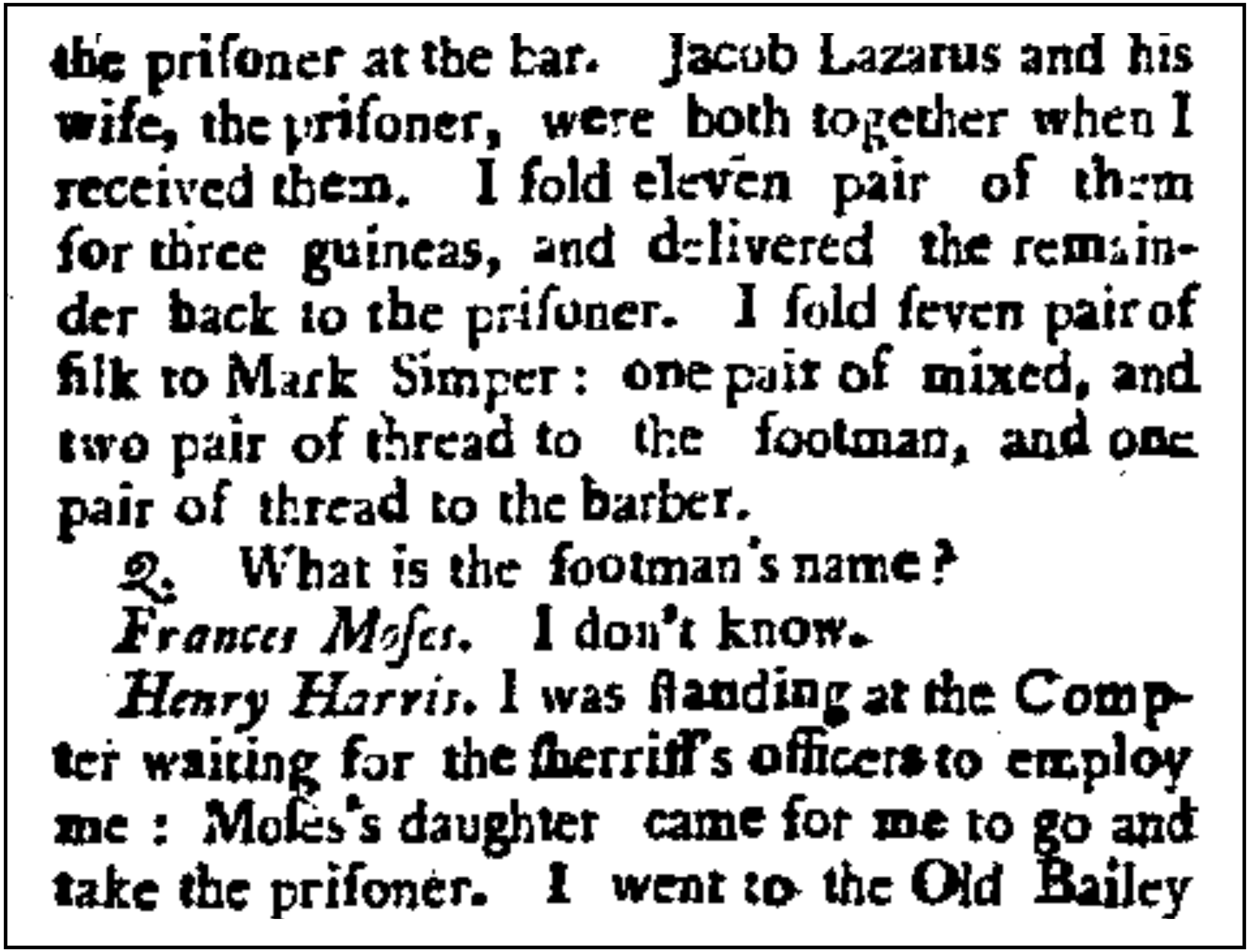
Dan Klein — UC Berkeley

Joint work with Taylor Berg-Kirkpatrick and Greg Durrett [ACL 2013]

Historical Document

Historical Document

Old Bailey Court Proceedings 1775



the priſoner at the bar. Jacob Lazarus and his wife, the priſoner, were both together when I received them. I ſold eleven pair of them for three guineas, and delivered the remainder back to the priſoner. I ſold ſeven pair of ſilk to Mark Simper: one pair of mixed, and two pair of thread to the footman, and one pair of thread to the barber.

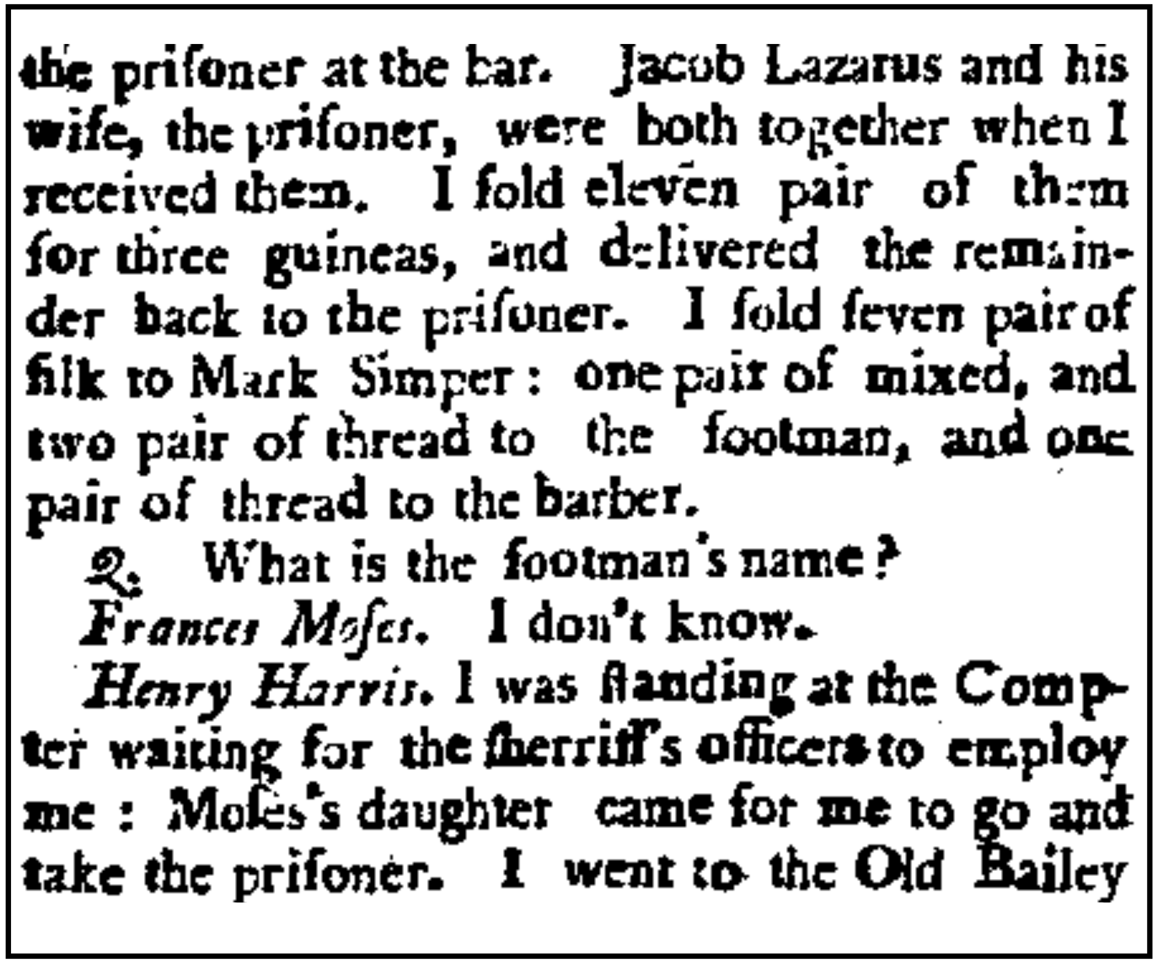
Q. What is the footman's name?

Frances Moſes. I don't know.

Henry Harris. I was ſtanding at the Compter waiting for the ſherriff's officers to employ me: Moſes's daughter came for me to go and take the priſoner. I went to the Old Bailey

Transcription

Document Image



the priſoner at the bar. Jacob Lazarus and his wife, the priſoner, were both together when I received them. I ſold eleven pair of them for three guineas, and delivered the remainder back to the priſoner. I ſold ſeven pair of ſilk to Mark Simper: one pair of mixed, and two pair of thread to the footman, and one pair of thread to the barber.

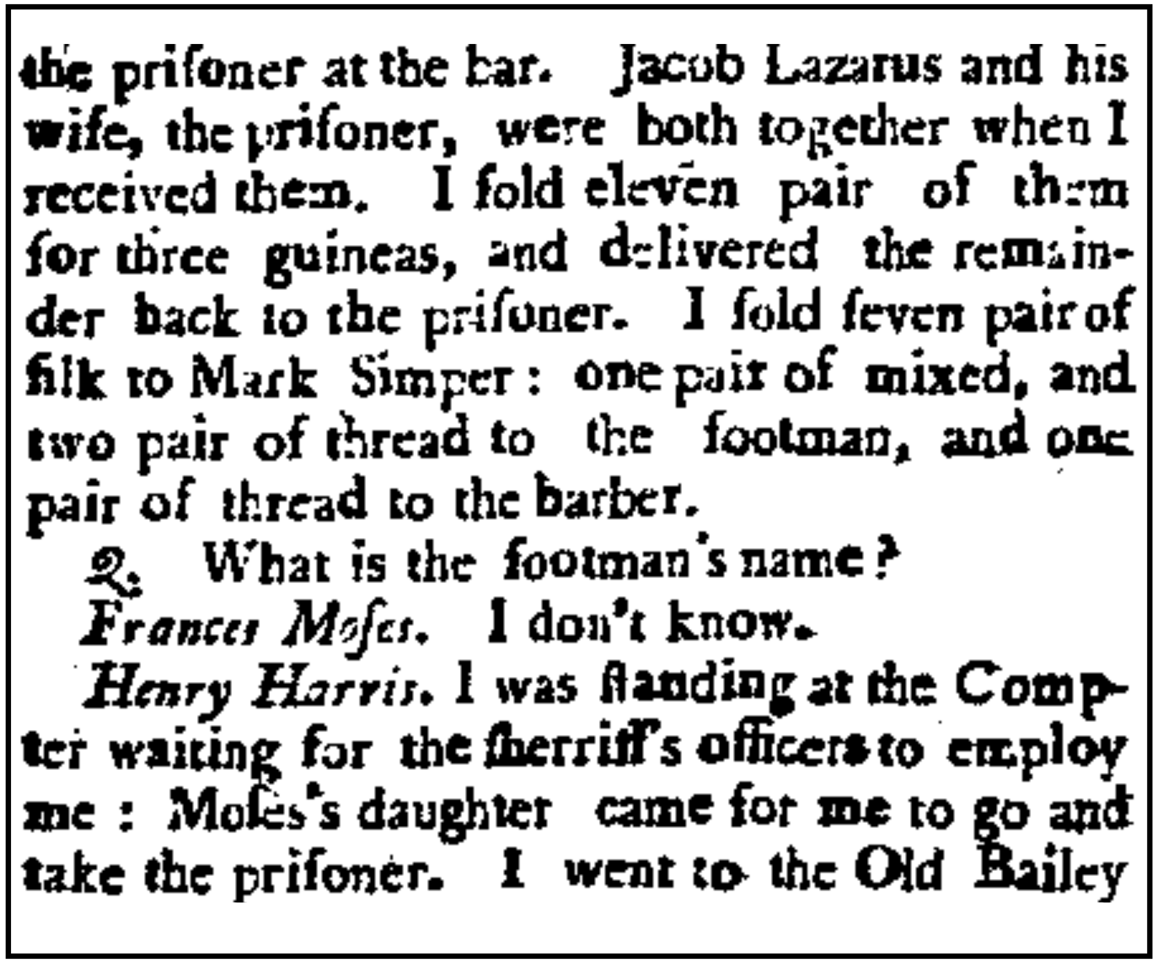
Q. What is the footman's name?

Frances Moſes. I don't know.

Henry Harris. I was ſtanding at the Compter waiting for the ſherriff's officers to employ me: Moſes's daughter came for me to go and take the priſoner. I went to the Old Bailey

Transcription

Document Image



the priſoner at the bar. Jacob Lazarus and his wife, the priſoner, were both together when I received them. I ſold eleven pair of them for three guineas, and delivered the remainder back to the priſoner. I ſold ſeven pair of ſilk to Mark Simper: one pair of mixed, and two pair of thread to the footman, and one pair of thread to the barber.

Q. What is the footman's name?

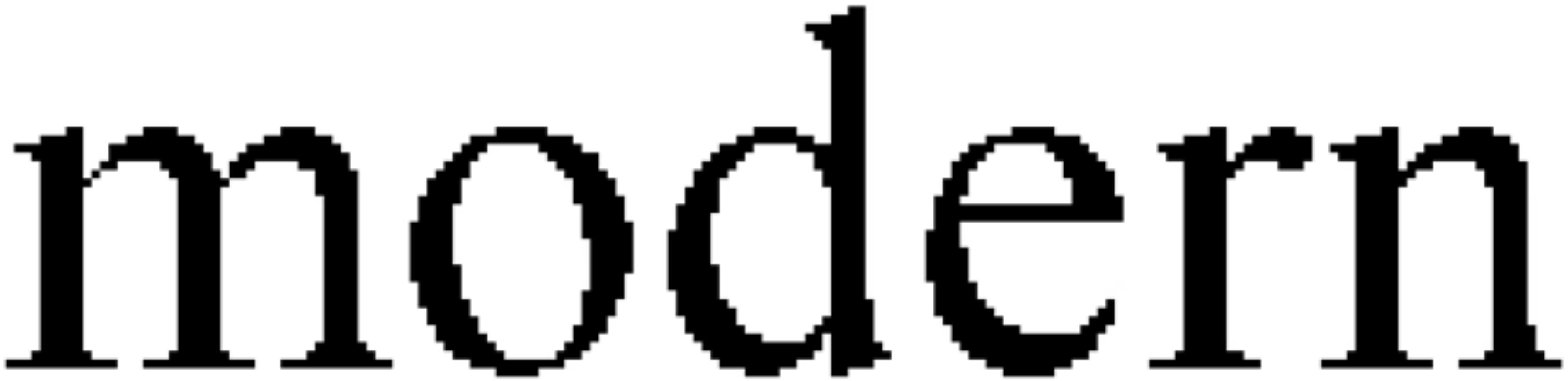
Frances Moſes. I don't know.

Henry Harris. I was ſtanding at the Compter waiting for the ſherriff's officers to employ me: Moſes's daughter came for me to go and take the priſoner. I went to the Old Bailey

Transcription (Google Tesseract)

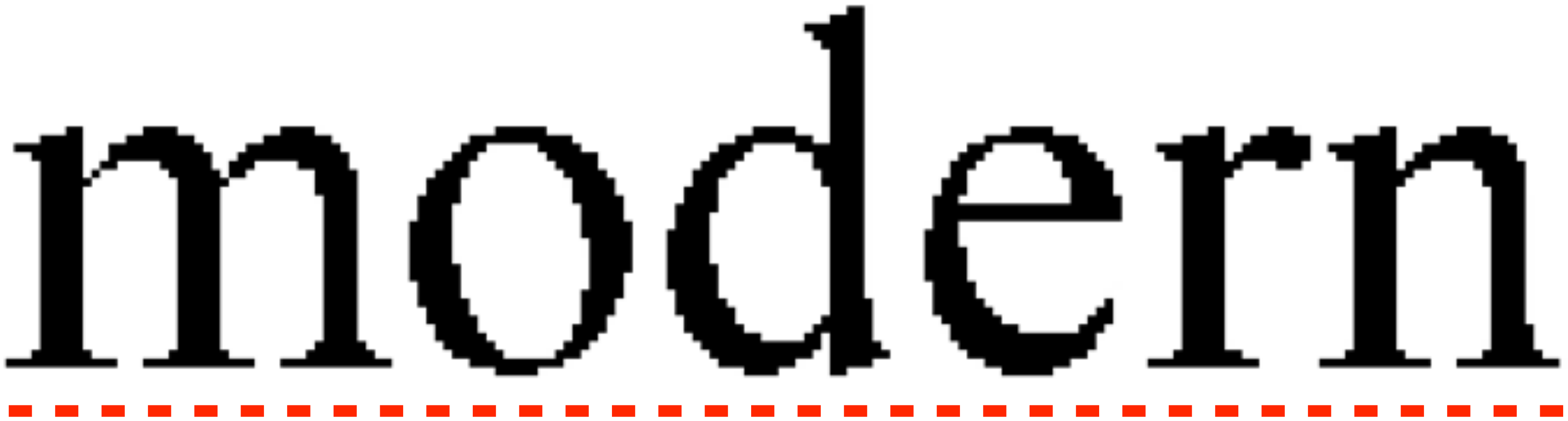
```
and Ch’: priftmer anhc bar. Jacob Lazarus and his
IHP1 uh: prifoner. were both together when!
rcccivcd lhczn. I fold eievén pair of than
for xiirce guincas, and dclivcrcd the rcll'l:.in-
d:r hack lo :11: prifuner. 1 fold ftvcn pairof
filk to Mark Simpcr : nncpuir of mixcd. and.
mo pair of Ifircad to lhz: foolnun, and on:
pair of zhrzad to lh: barber. '
  Q: What is the foolmarfs name?
  Fraum Mgfzr. I dun’: know.
  Hairy Hzrvir. l was flandingar the Camp
Icr waizin far the thcrrilfs ufliceruo employ
in: : Mo 3‘: daughter came for me to 0 am!
take the prifoncr. 1 Wm! to |hc Old aailcy
```

Pipelined Approach



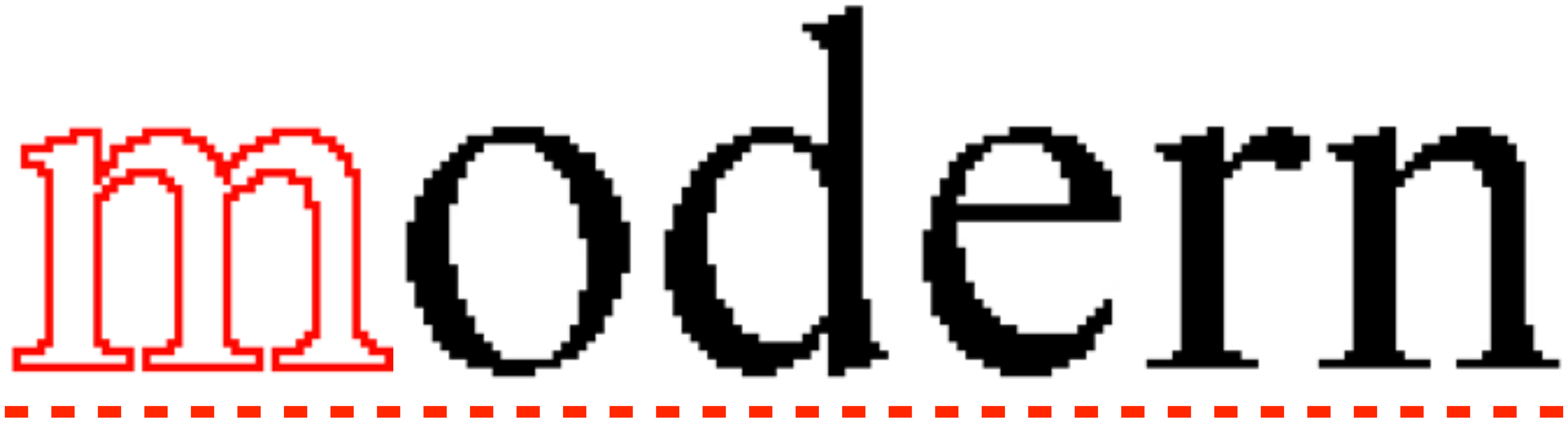
modern

Pipelined Approach

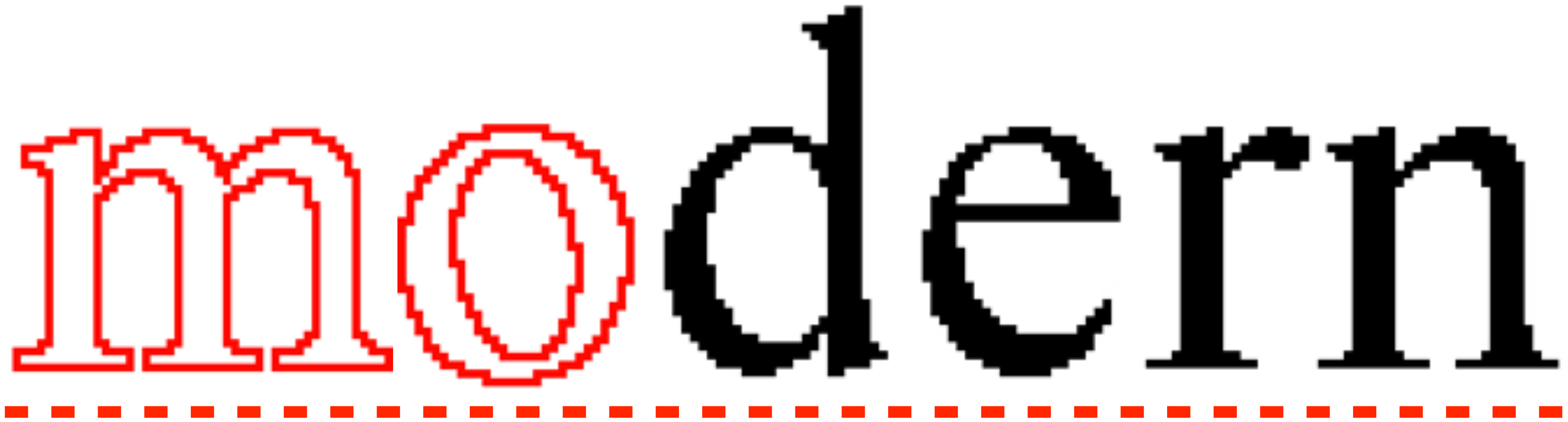


modern

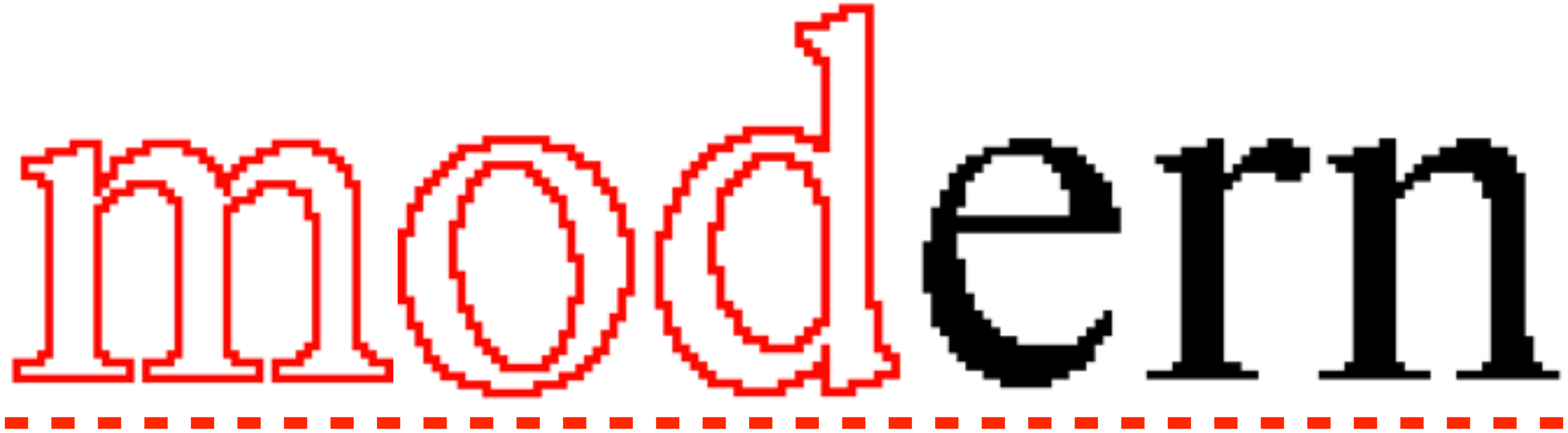
Pipelined Approach



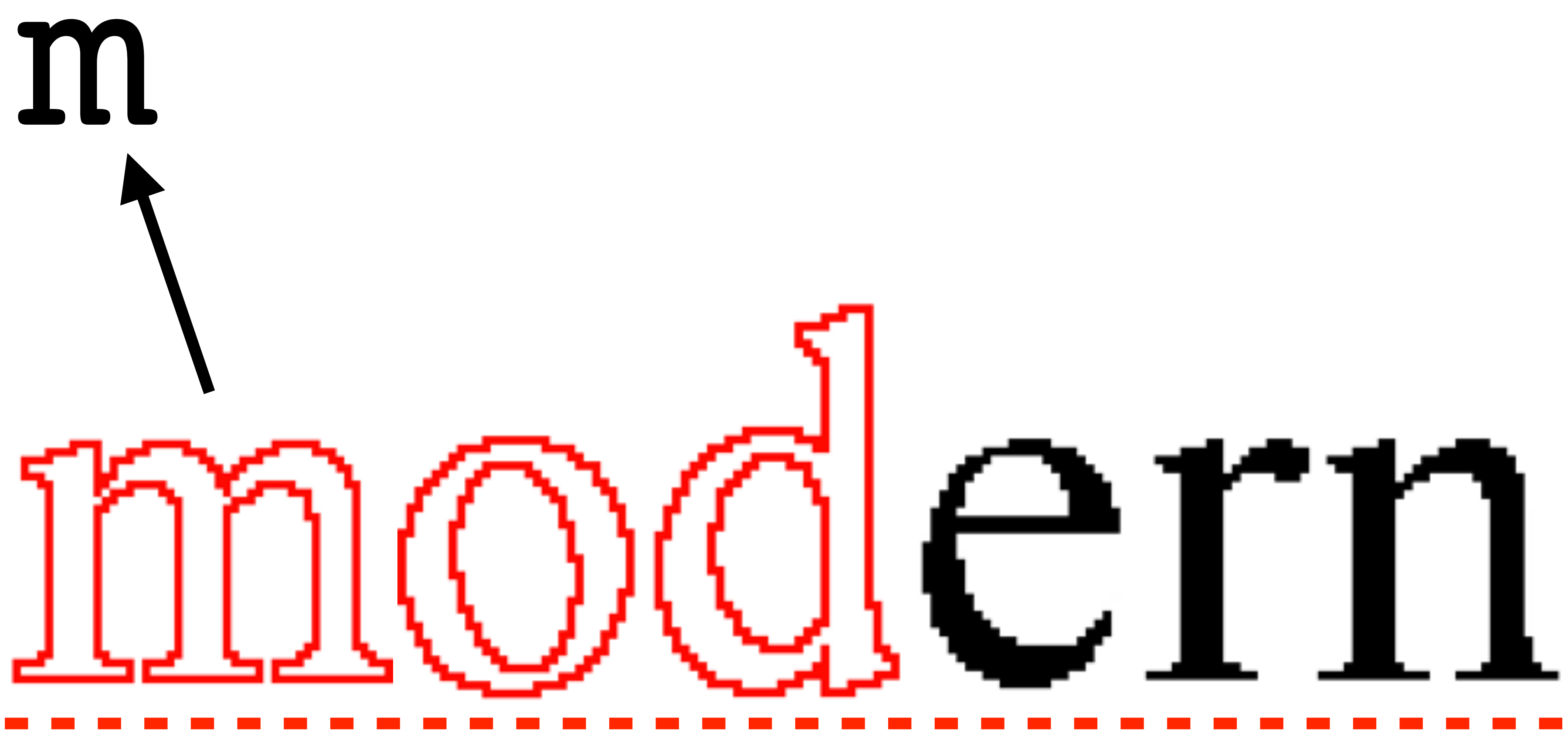
Pipelined Approach



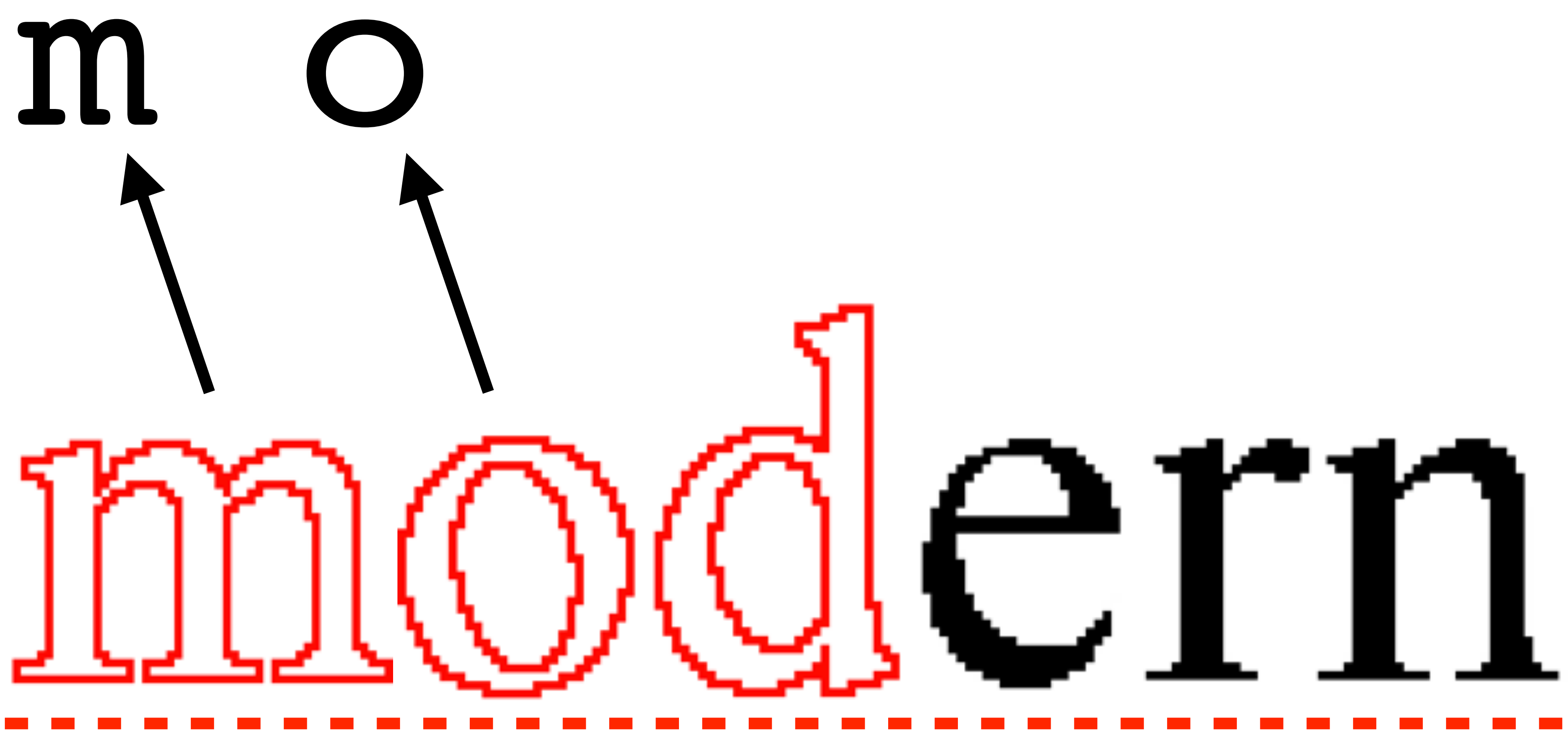
Pipelined Approach



Pipelined Approach

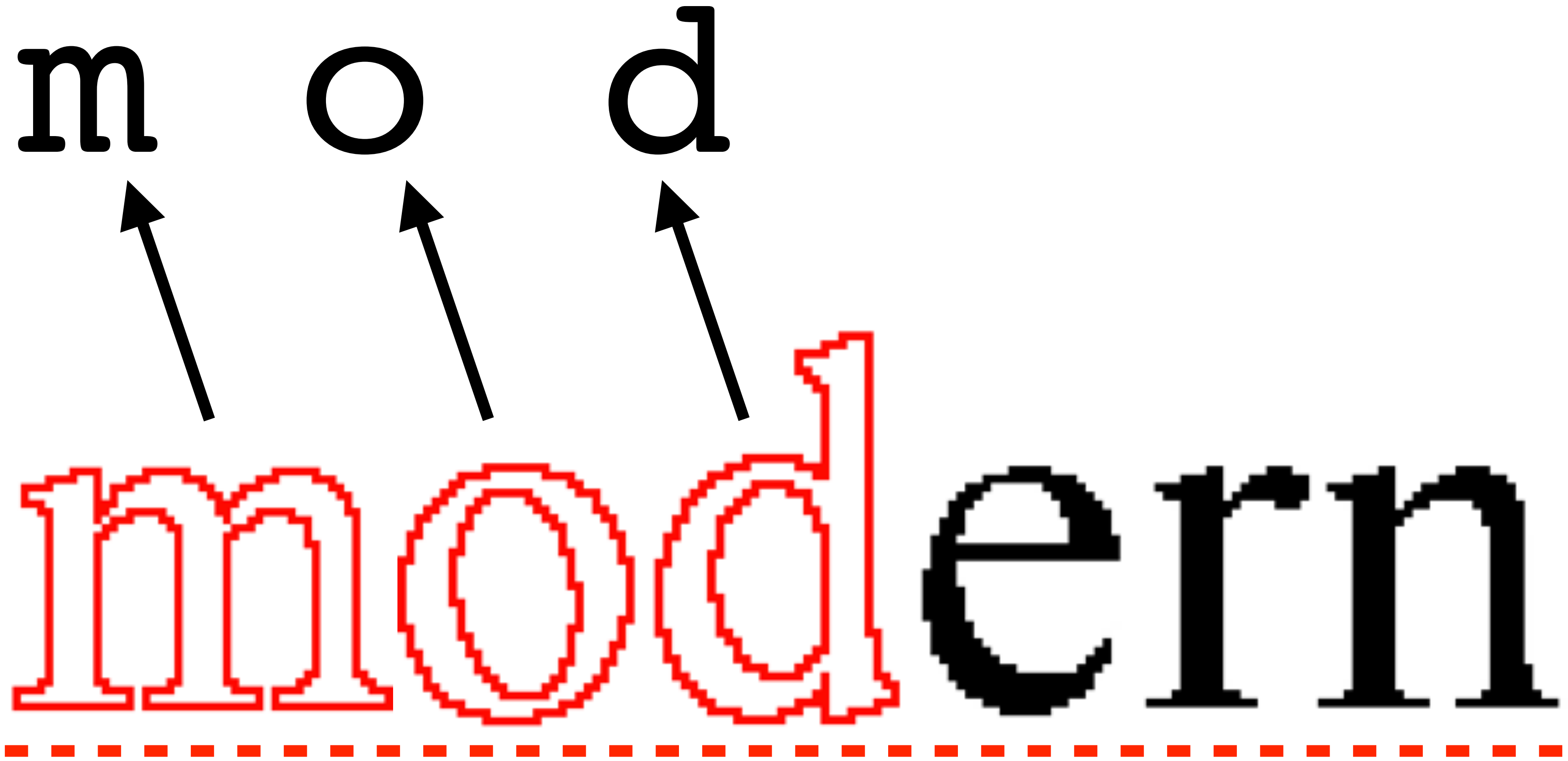


Pipelined Approach



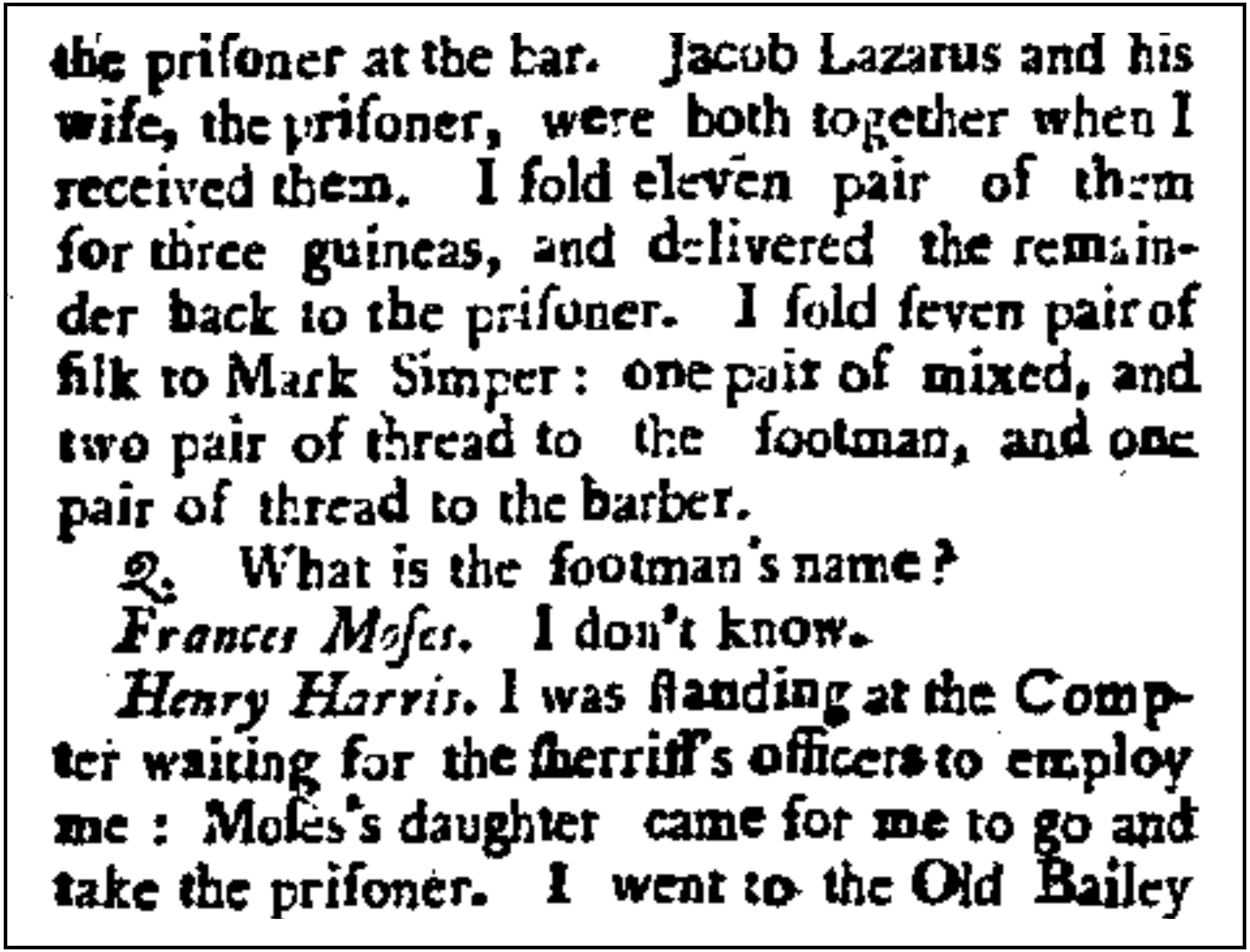
m o

Pipelined Approach



m o d

Historical Document



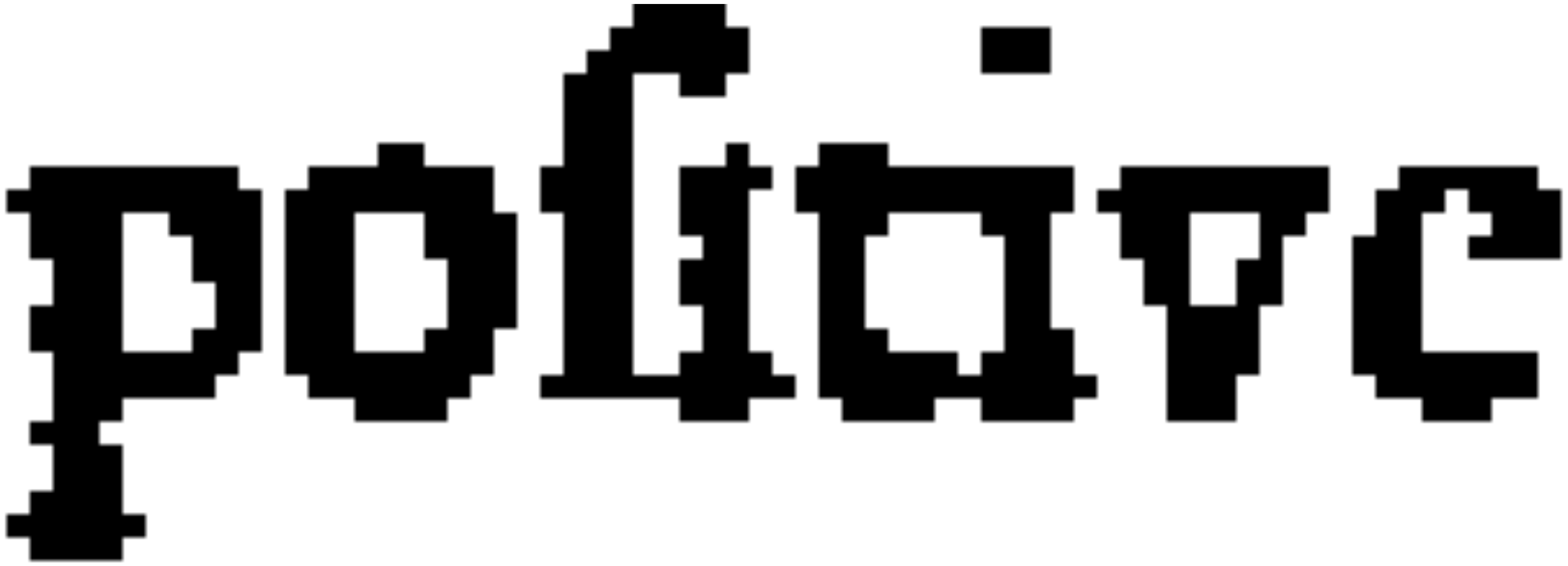
the priſoner at the bar. Jacob Lazarus and his wife, the priſoner, were both together when I received them. I ſold eleven pair of them for three guineas, and delivered the remainder back to the priſoner. I ſold ſeven pair of ſilk to Mark Simper: one pair of mixed, and two pair of thread to the footman, and one pair of thread to the barber.

Q. What is the footman's name?

Francis Moſes. I don't know.

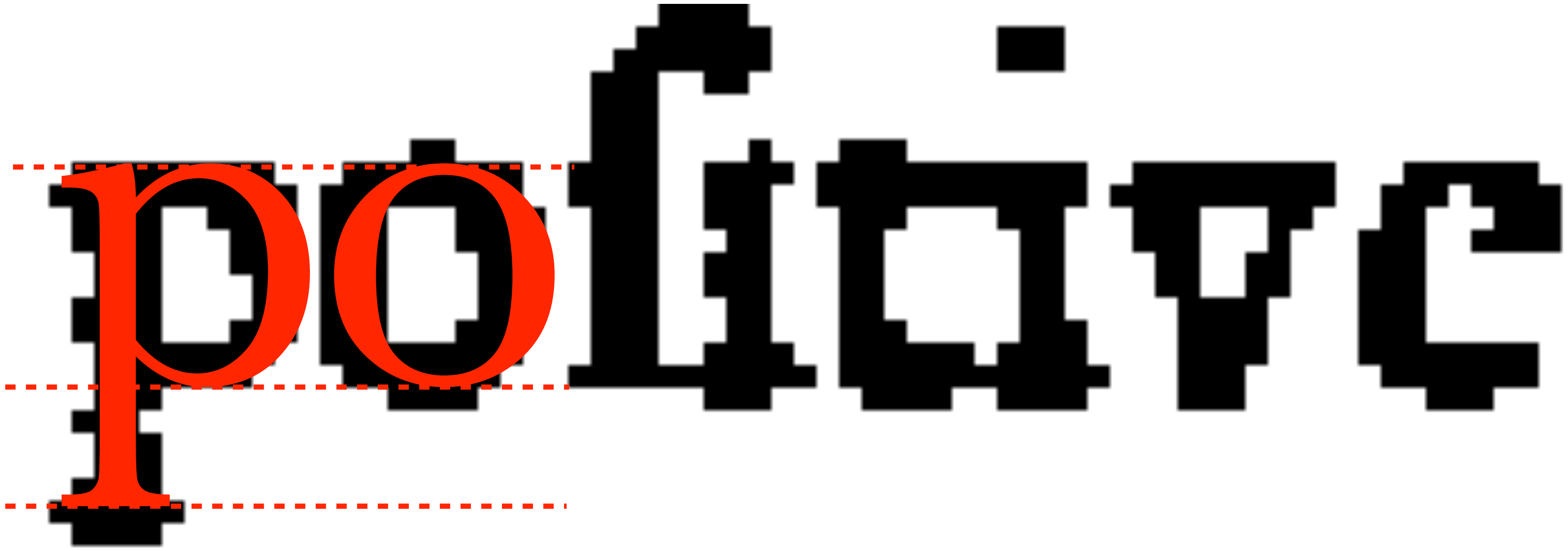
Henry Harris. I was ſtanding at the Compter waiting for the ſherriff's officers to employ me: Moſes's daughter came for me to go and take the priſoner. I went to the Old Bailey

Unknown Fonts

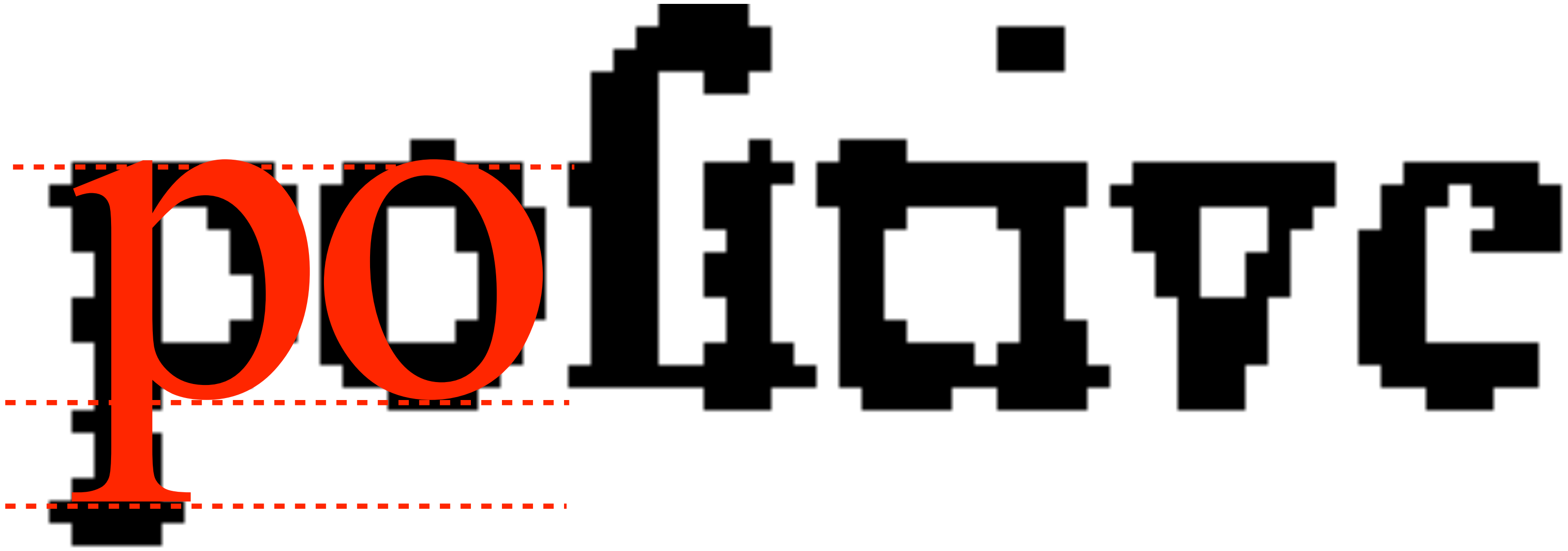


poſitive

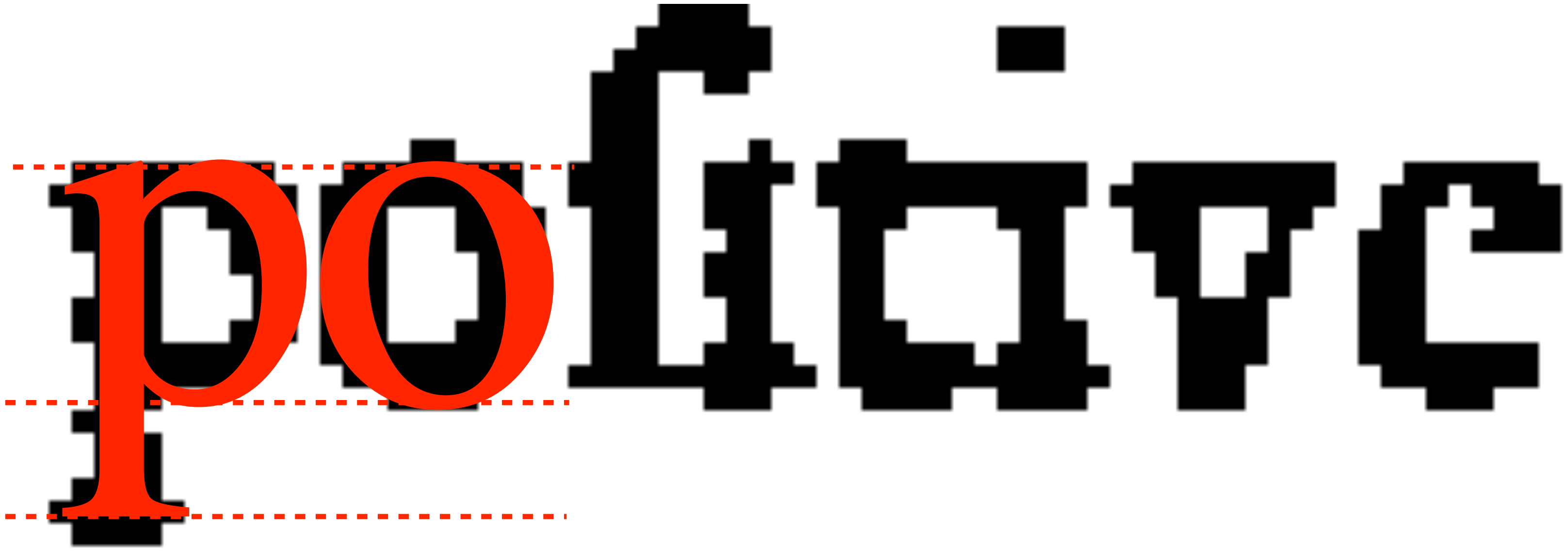
Unknown Fonts



Unknown Fonts

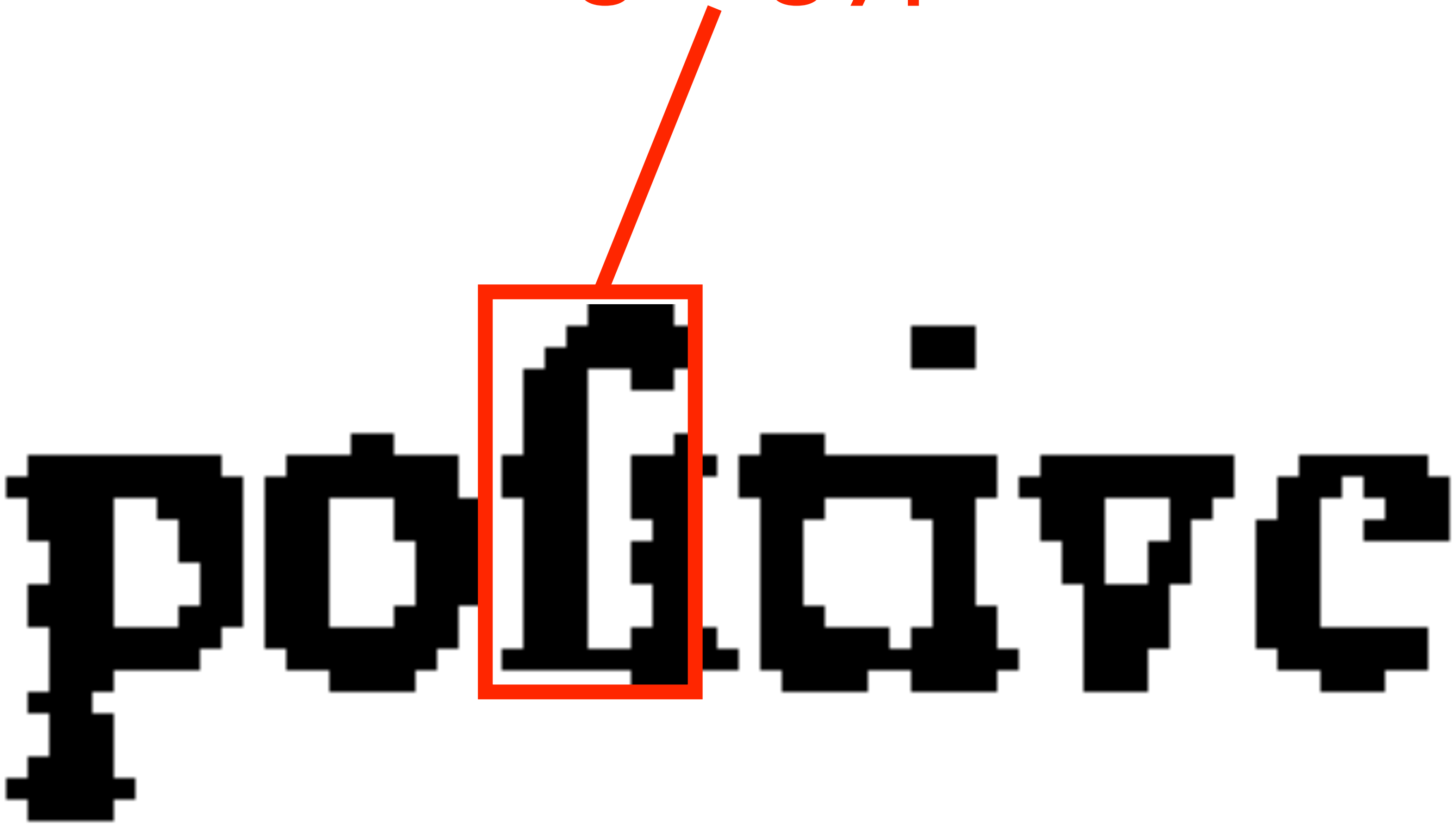


Unknown Fonts

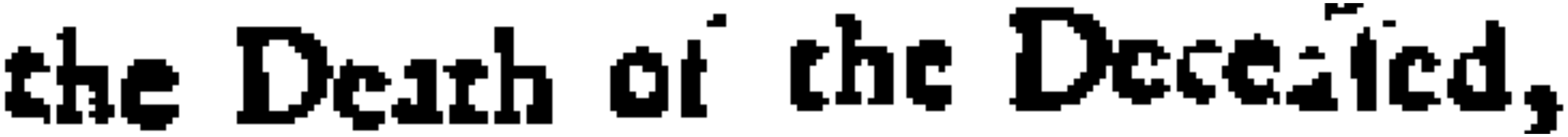


Unknown Fonts

long s glyph

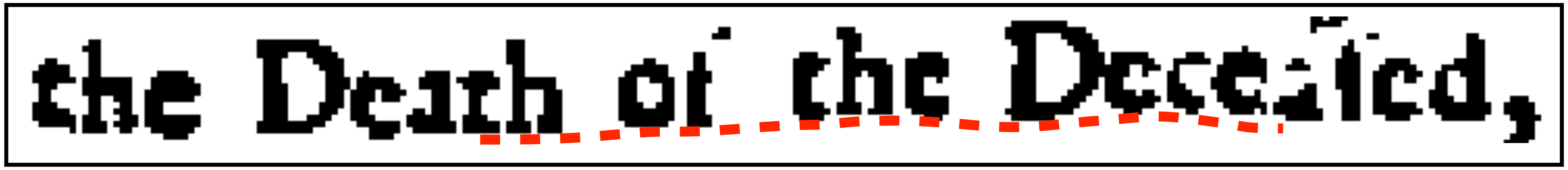


Wandering Baseline



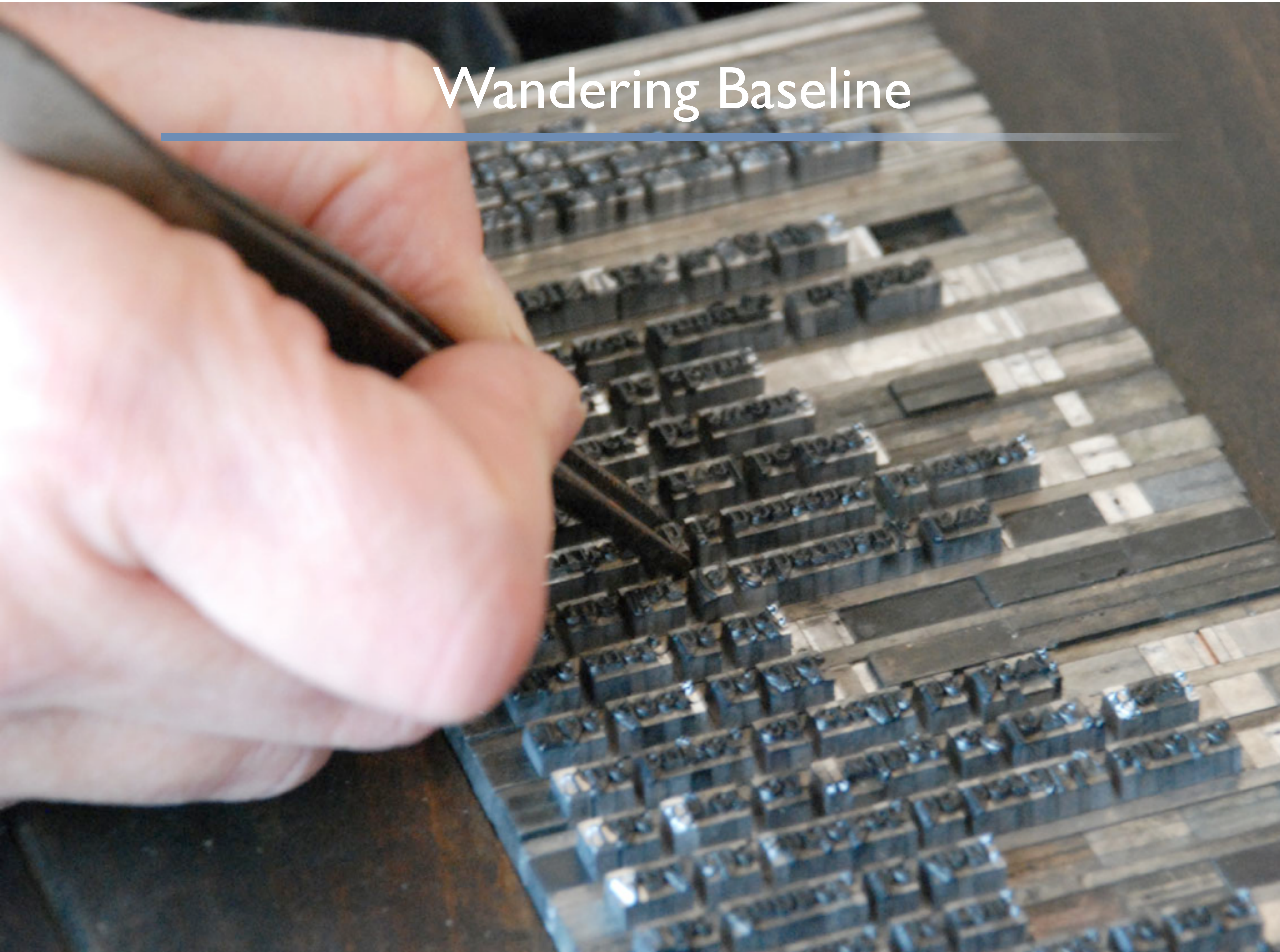
the Death of the Deceaſed,

Wandering Baseline

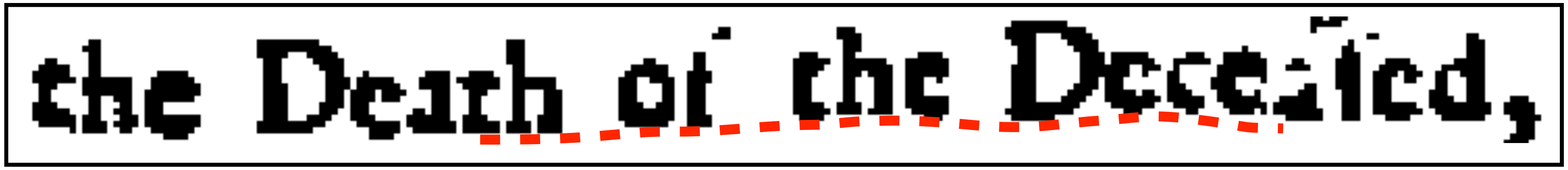


the Death of the Deceaſed,

Wandering Baseline

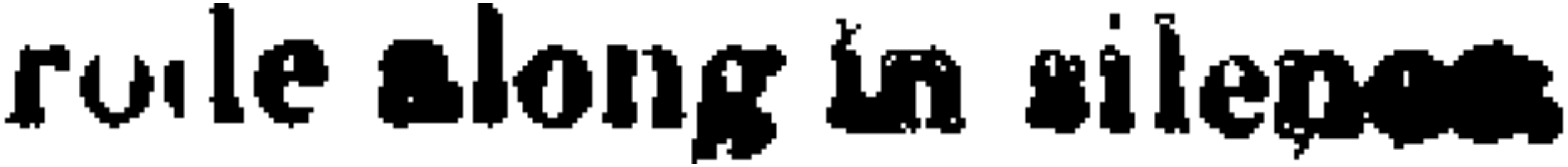


Wandering Baseline

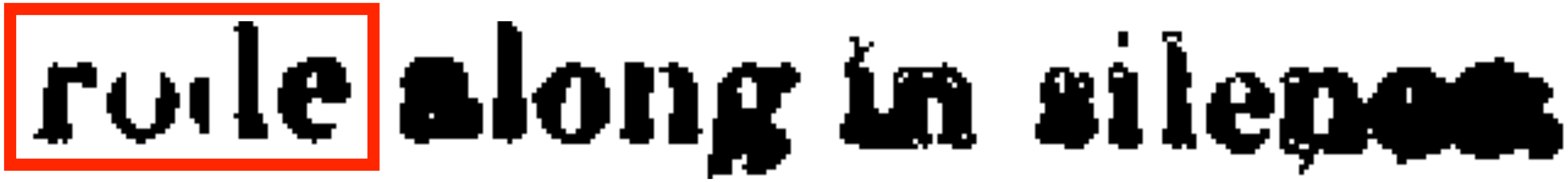


the Death of the Deceaſed,

Uneven Inking

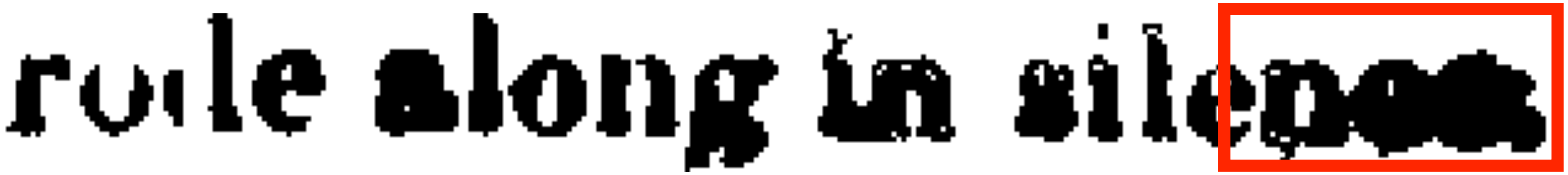

rode along in silence

Uneven Inking



rule along in silence

Uneven Inking

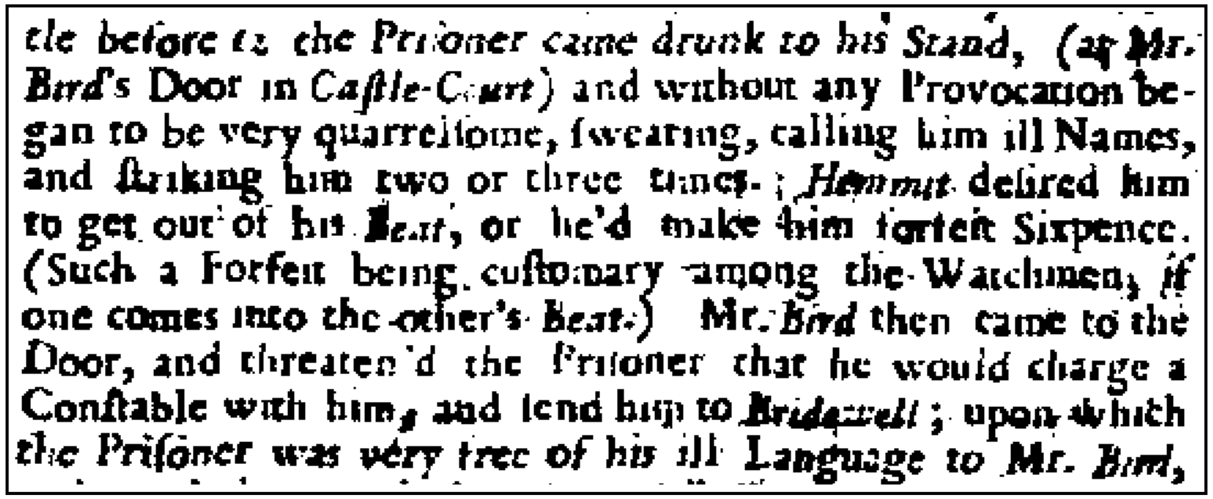


rode along in silence

Uneven Inking

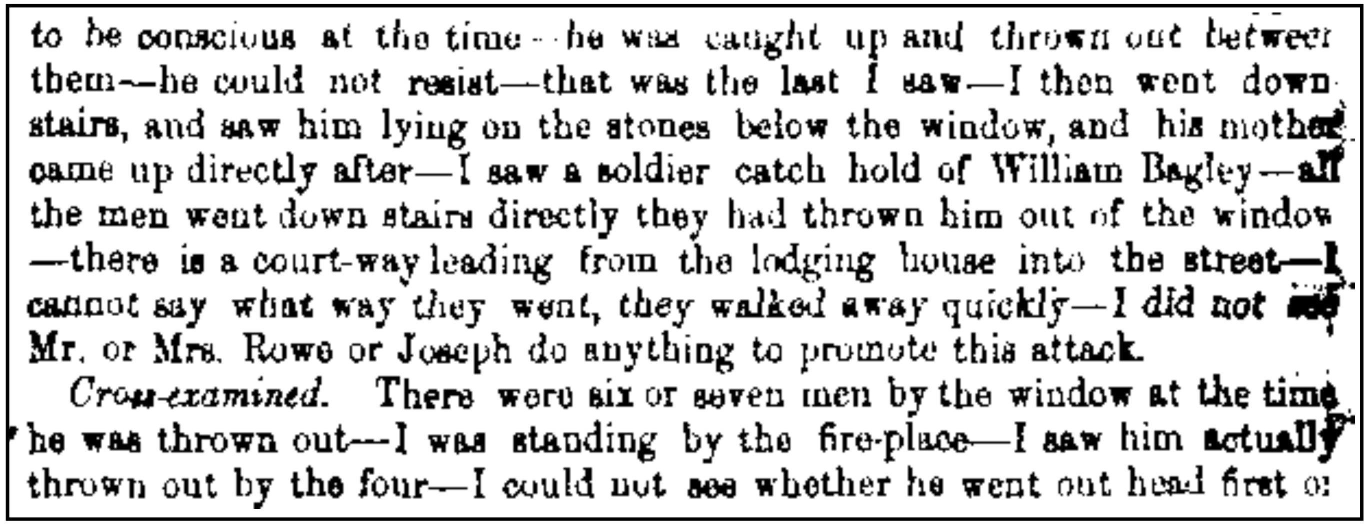
Various Historical Documents

1725



tle before t: the Priſoner came drunk to his Stand, (at Mr. Bird's Door in Caſtle-Court) and without any Provocation began to be very quarrelſome, ſwearing, calling him ill Names, and ſtriking him two or three times.; Hammit deſired him to get out of his Beat, or he'd make him forfeit Sixpence. (Such a Forfeit being cuſtomary among the Watchmen, if one comes into the other's Beat.) Mr. Bird then came to the Door, and threaten'd the Priſoner that he would charge a Conſtable with him, and ſend him to Bridewell; upon which the Priſoner was very free of his ill Language to Mr. Bird,

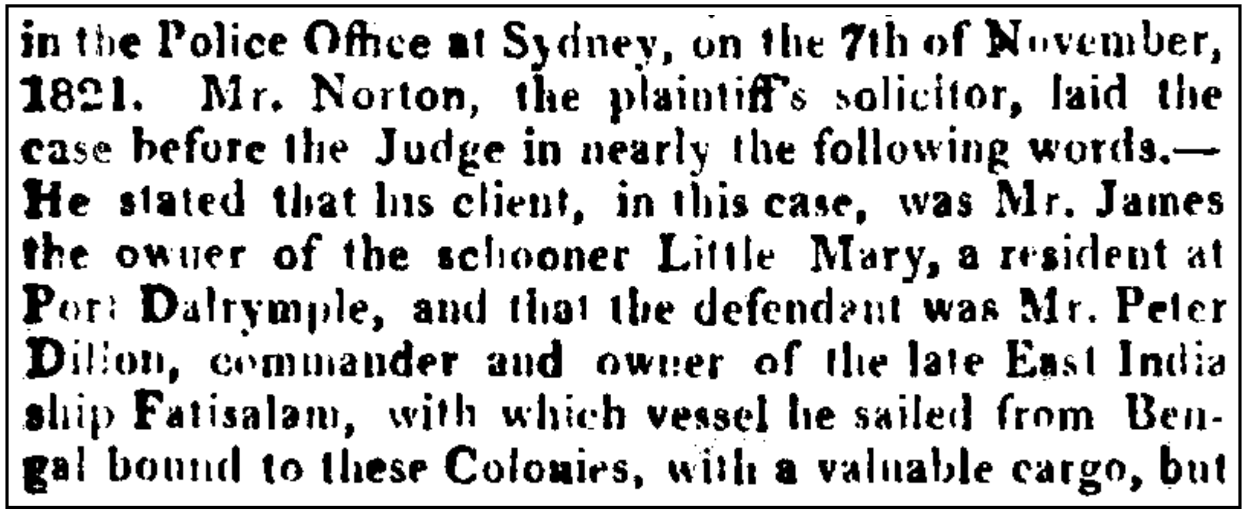
1875



to be conscious at the time—he was caught up and thrown out between them—he could not resist—that was the last I saw—I then went down stairs, and saw him lying on the stones below the window, and his mother came up directly after—I saw a soldier catch hold of William Bagley—all the men went down stairs directly they had thrown him out of the window—there is a court-way leading from the lodging house into the street—I cannot say what way they went, they walked away quickly—I did not see Mr. or Mrs. Rowe or Joseph do anything to promote this attack.

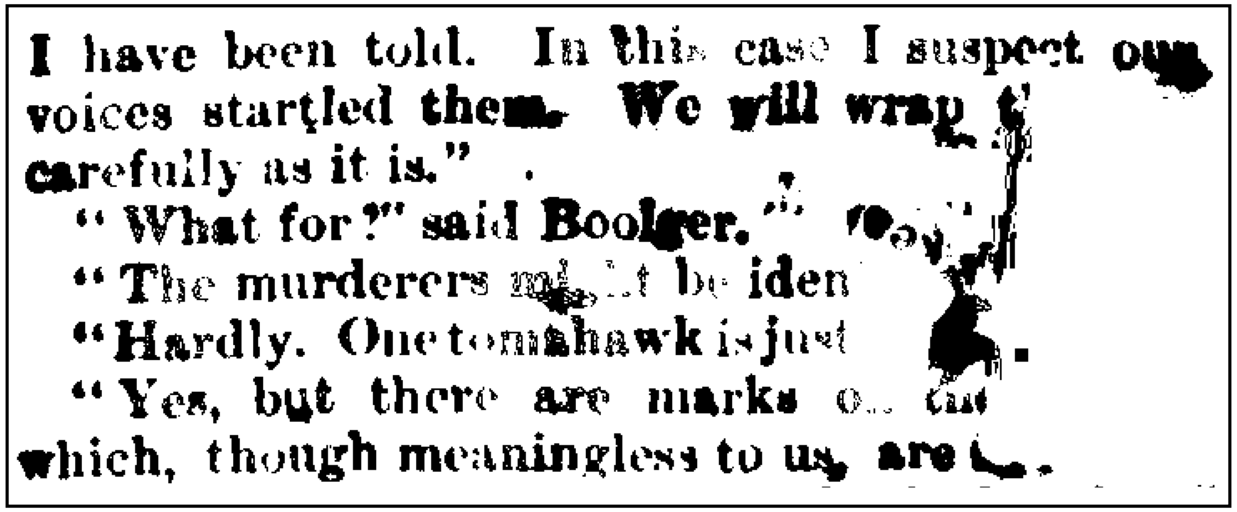
Cross-examined. There were six or seven men by the window at the time he was thrown out—I was standing by the fire-place—I saw him actually thrown out by the four—I could not see whether he went out head first or

1823



in the Police Office at Sydney, on the 7th of November, 1821. Mr. Norton, the plaintiff's solicitor, laid the case before the Judge in nearly the following words.—He stated that his client, in this case, was Mr. James the owner of the schooner Little Mary, a resident at Port Dalrymple, and that the defendant was Mr. Peter Dillon, commander and owner of the late East India ship Fatisalam, with which vessel he sailed from Bengal bound to these Colonies, with a valuable cargo, but

1883:



I have been told. In this case I suspect our voices startled them. We will wrap t' carefully as it is."

"What for?" said Boolger.

"The murderers might be iden

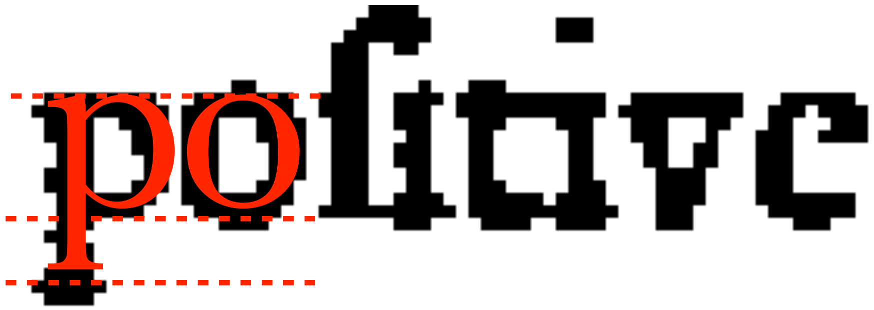
"Hardly. One tomahawk is just

"Yes, but there are marks o...

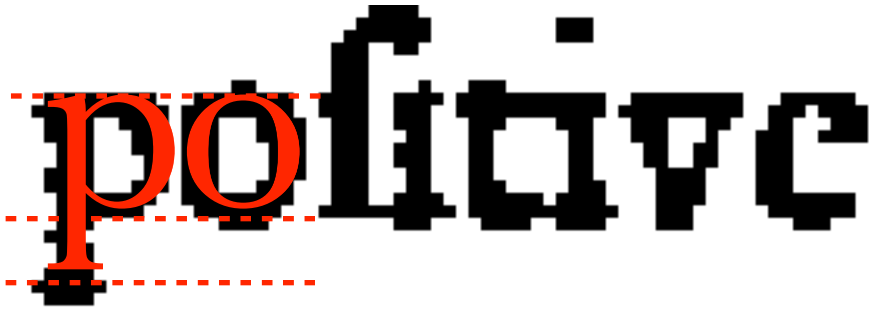
which, though meaningless to us, are

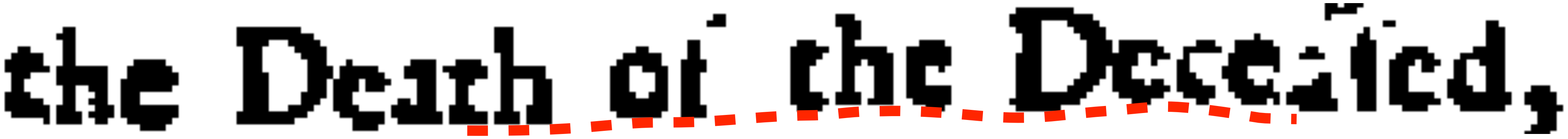
Our Approach

Our Approach

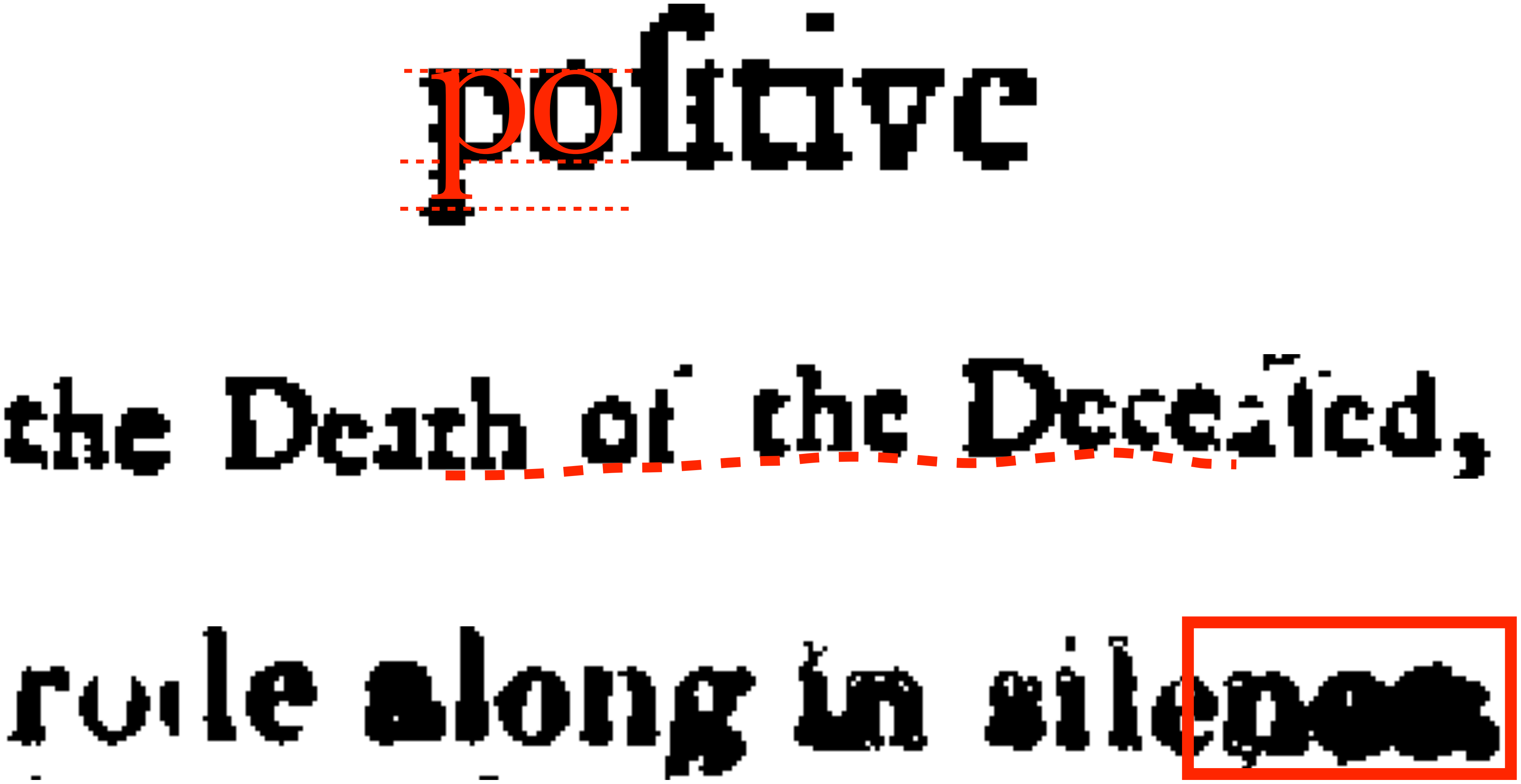


Our Approach





Our Approach



Generative Model

Generative Model

p

Generative Model

`p r`

Generative Model

`prisoner`

Generative Model

Language Model

prisoner

Generative Model

`prisoner`

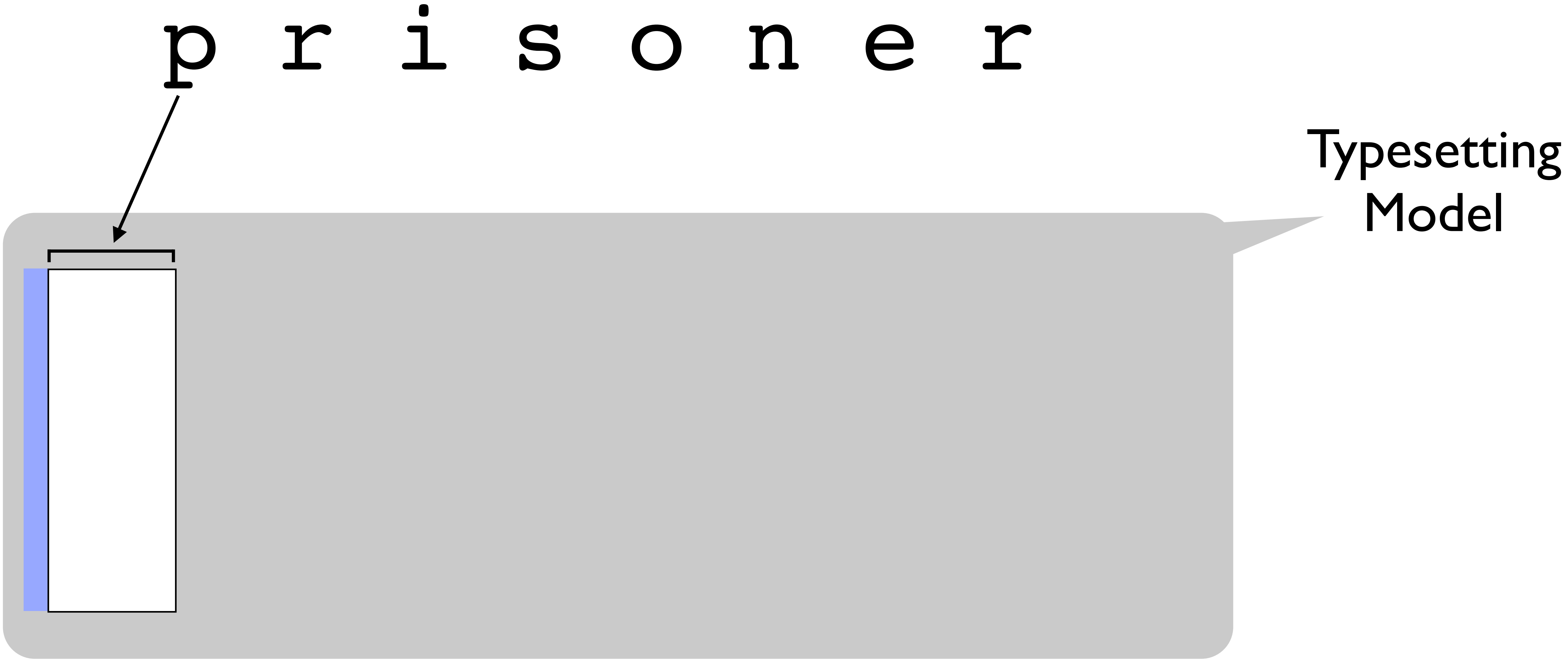
Typesetting Model

Generative Model

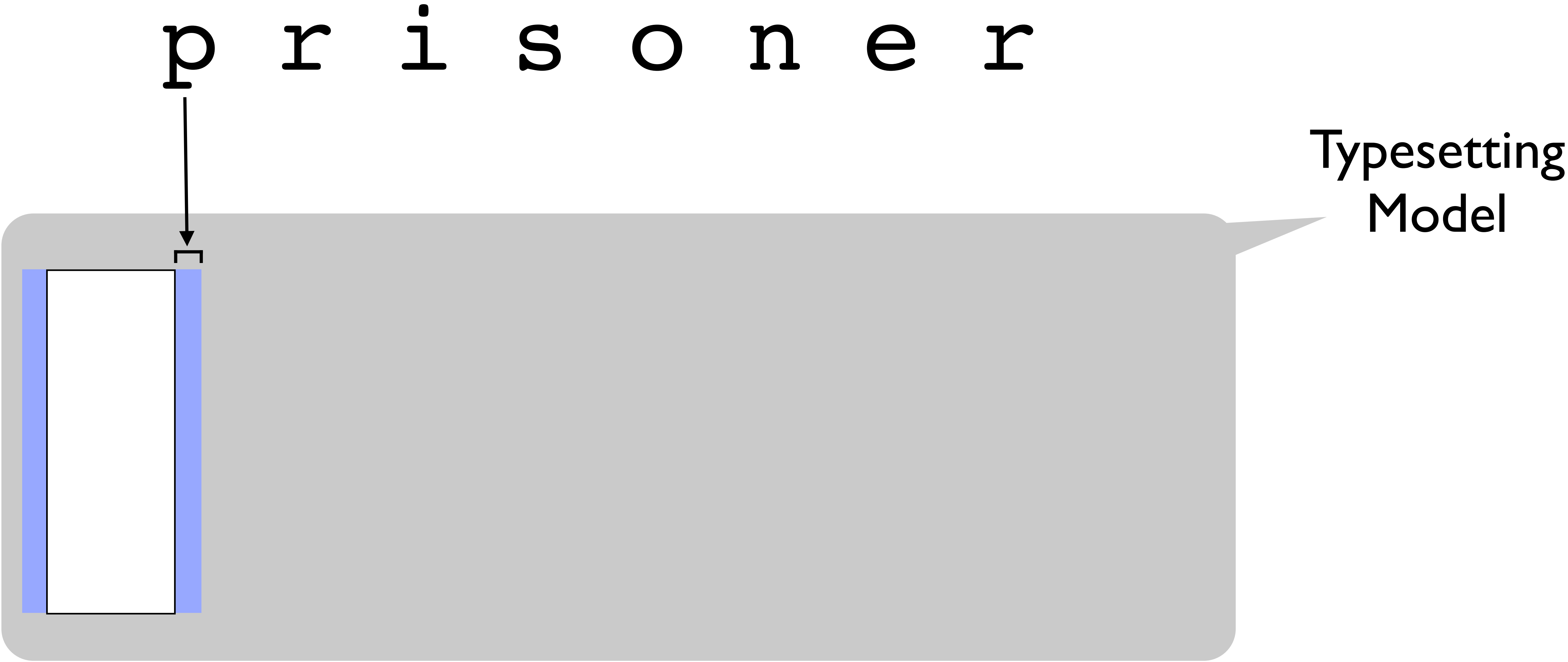
prisoner

Typesetting Model

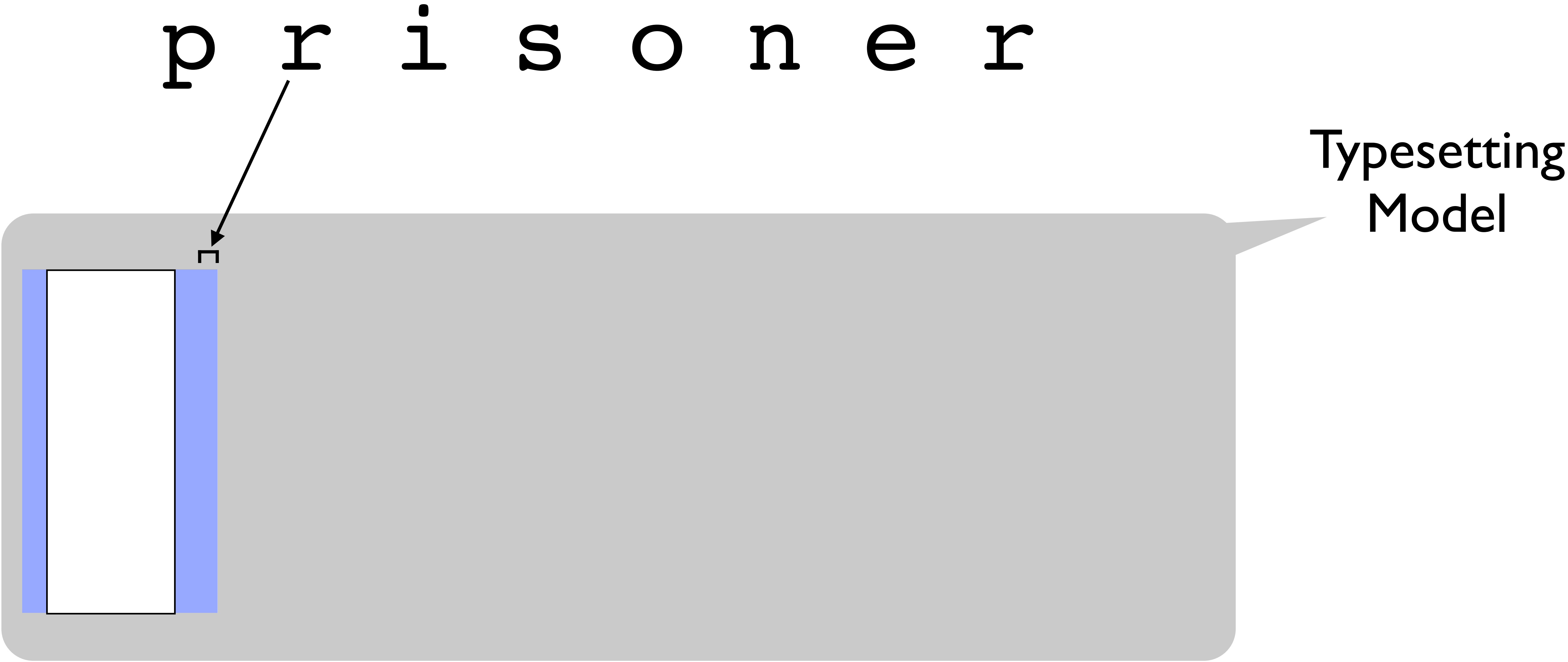
Generative Model



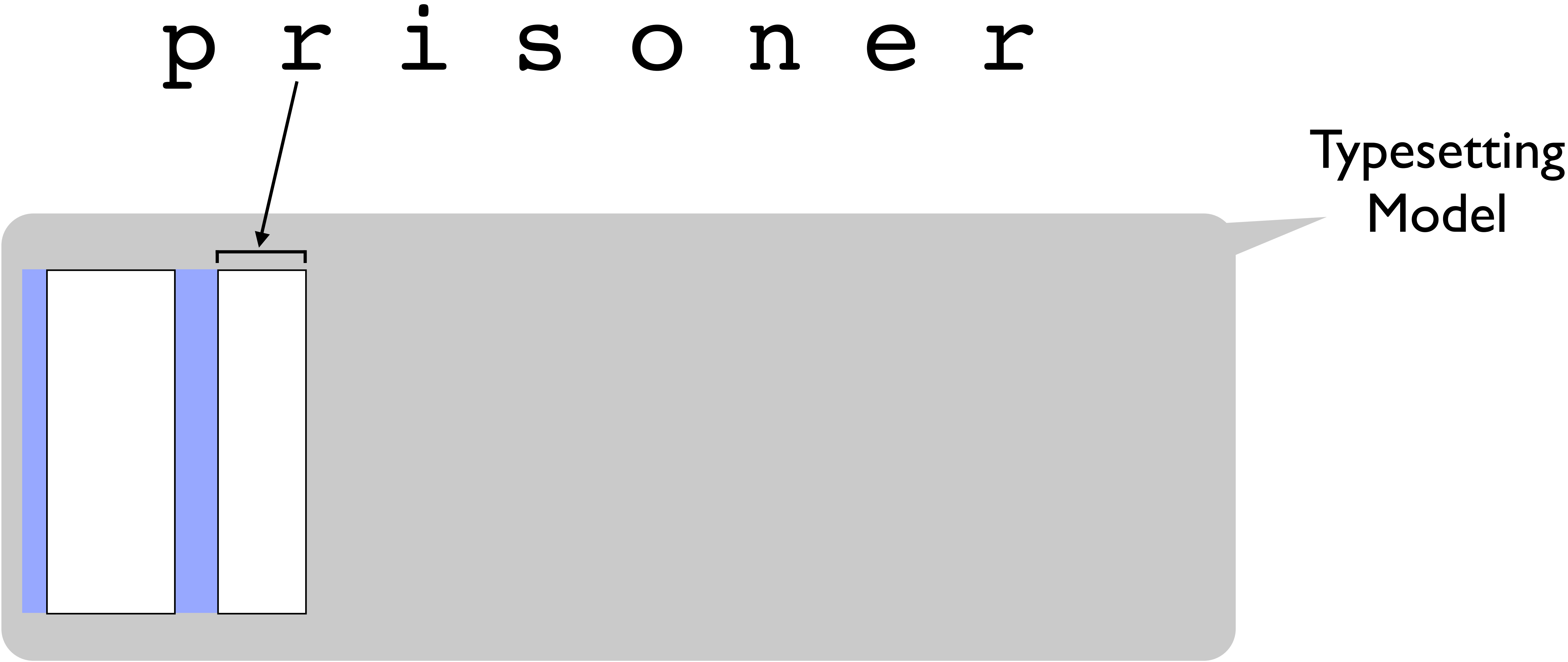
Generative Model



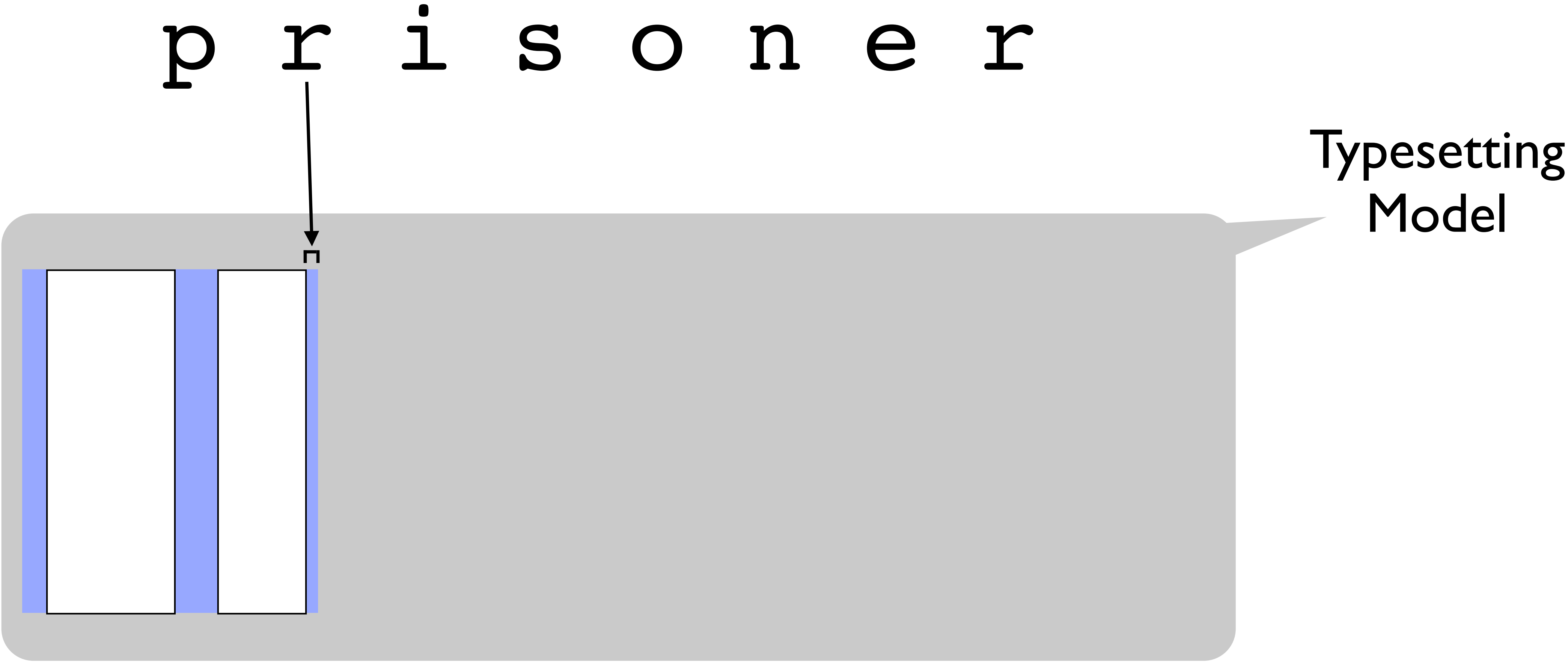
Generative Model



Generative Model

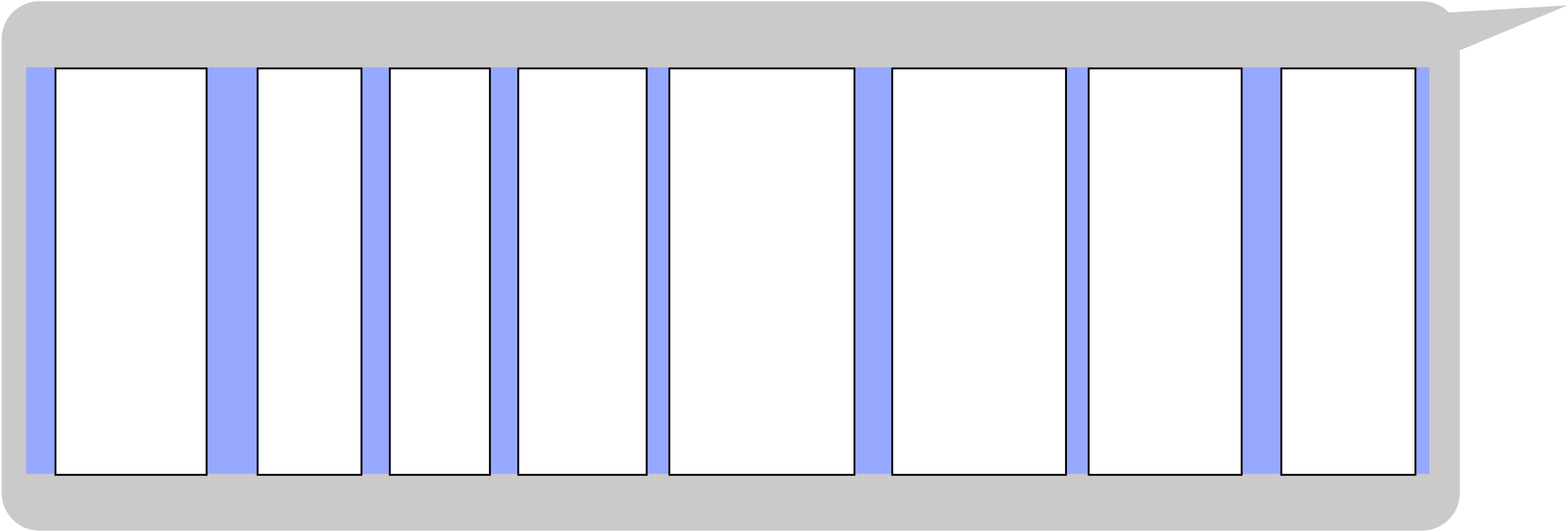


Generative Model



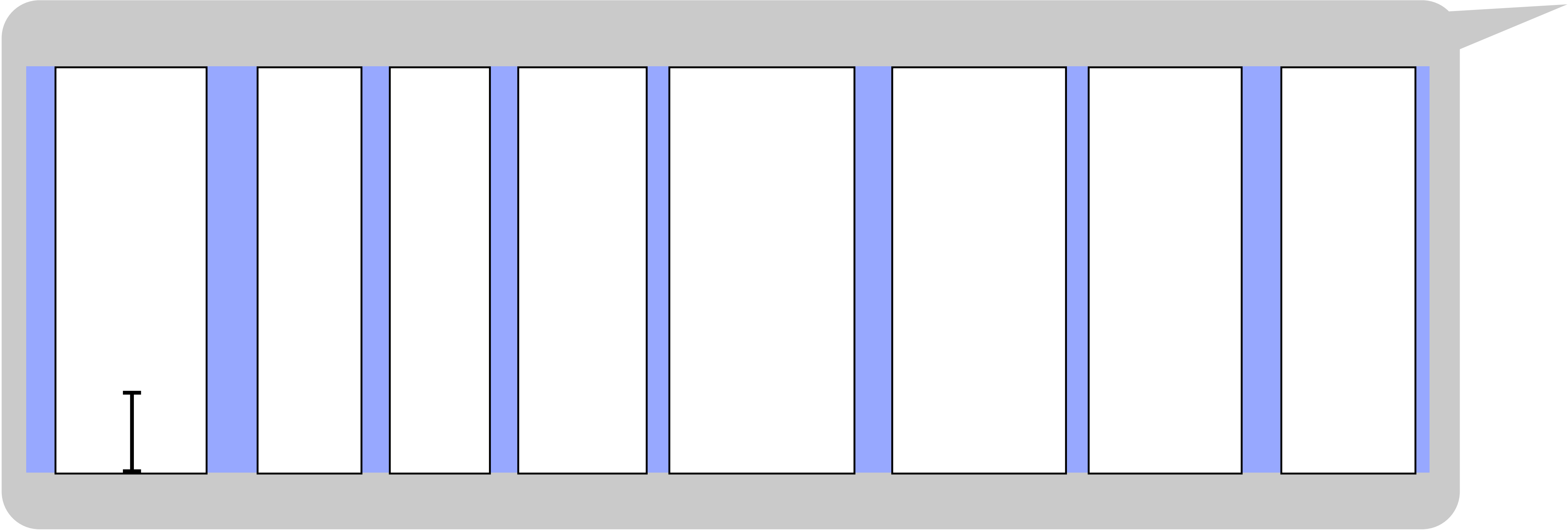
Generative Model

prisoner



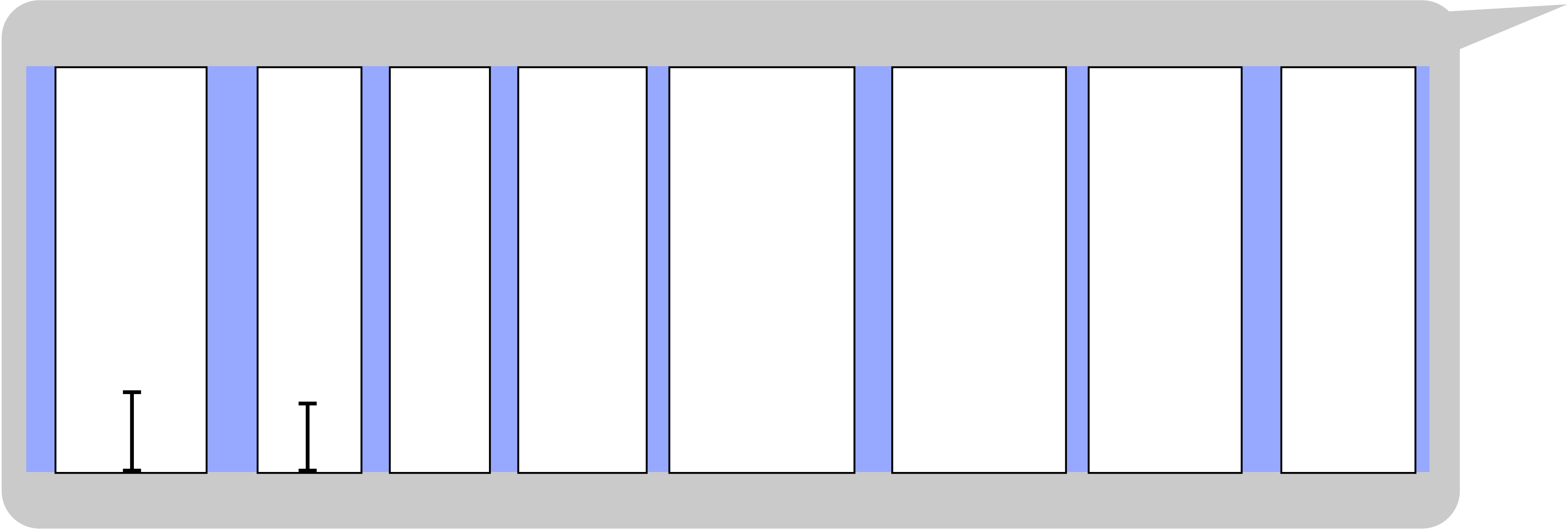
Typesetting Model

Generative Model


Typesetting Model

Generative Model

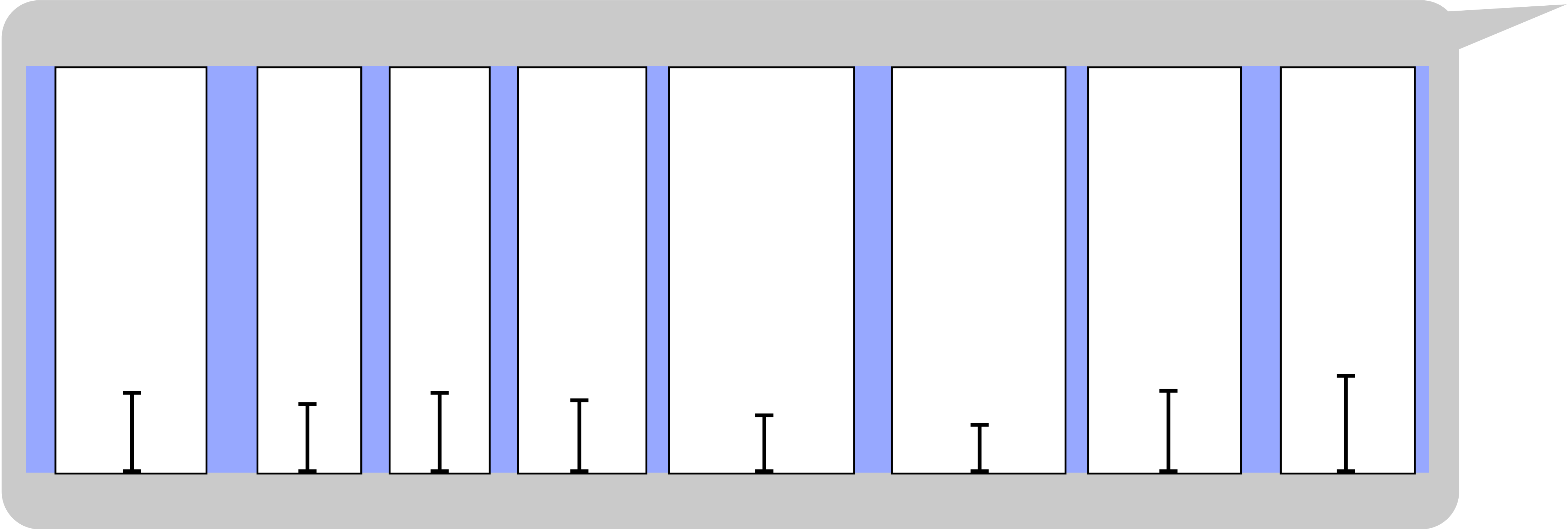
p r i s o n e r



Typesetting Model

Generative Model

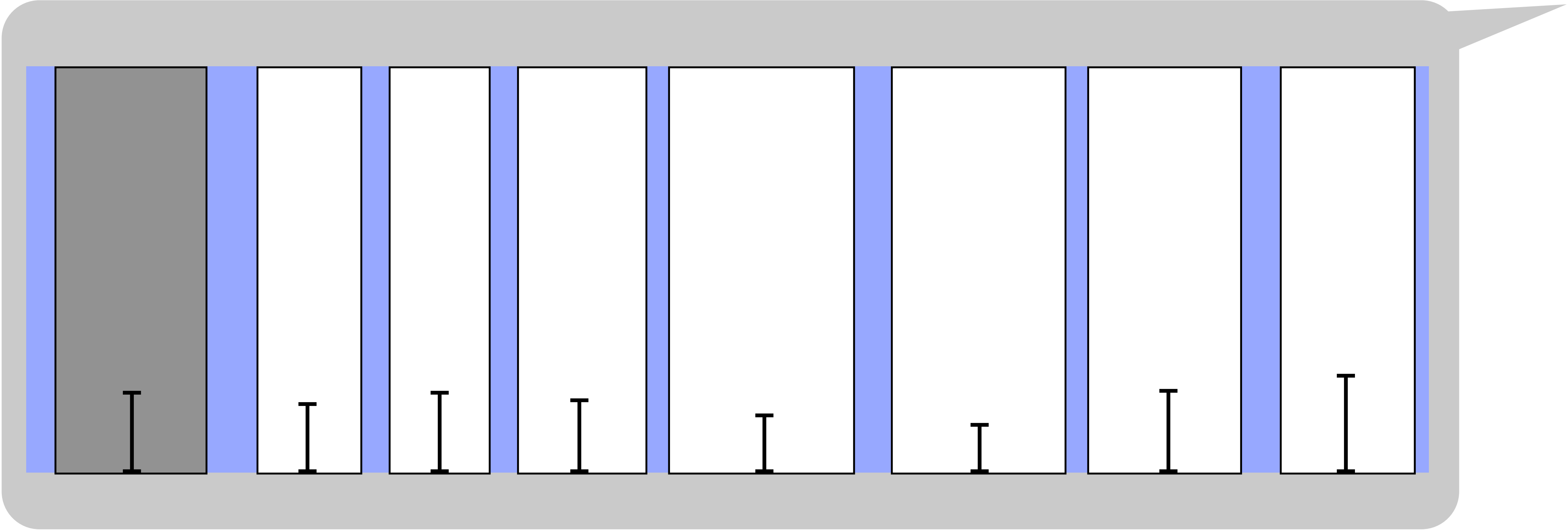
p r i s o n e r



Typesetting Model

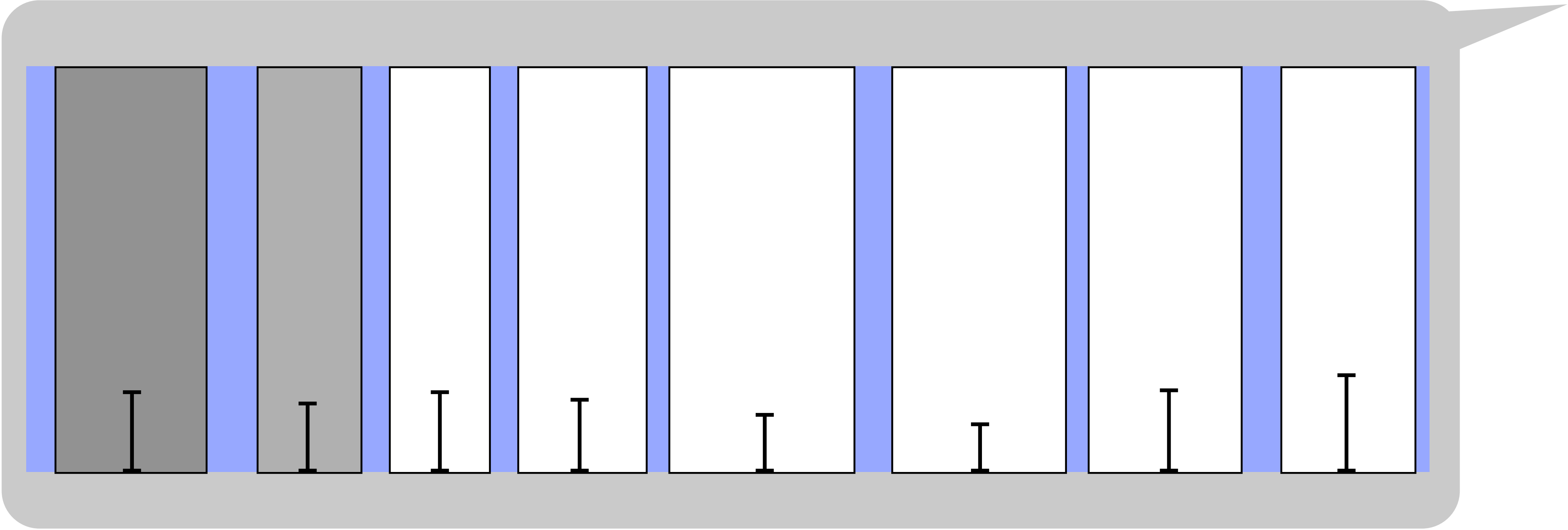
Generative Model

p r i s o n e r



Generative Model

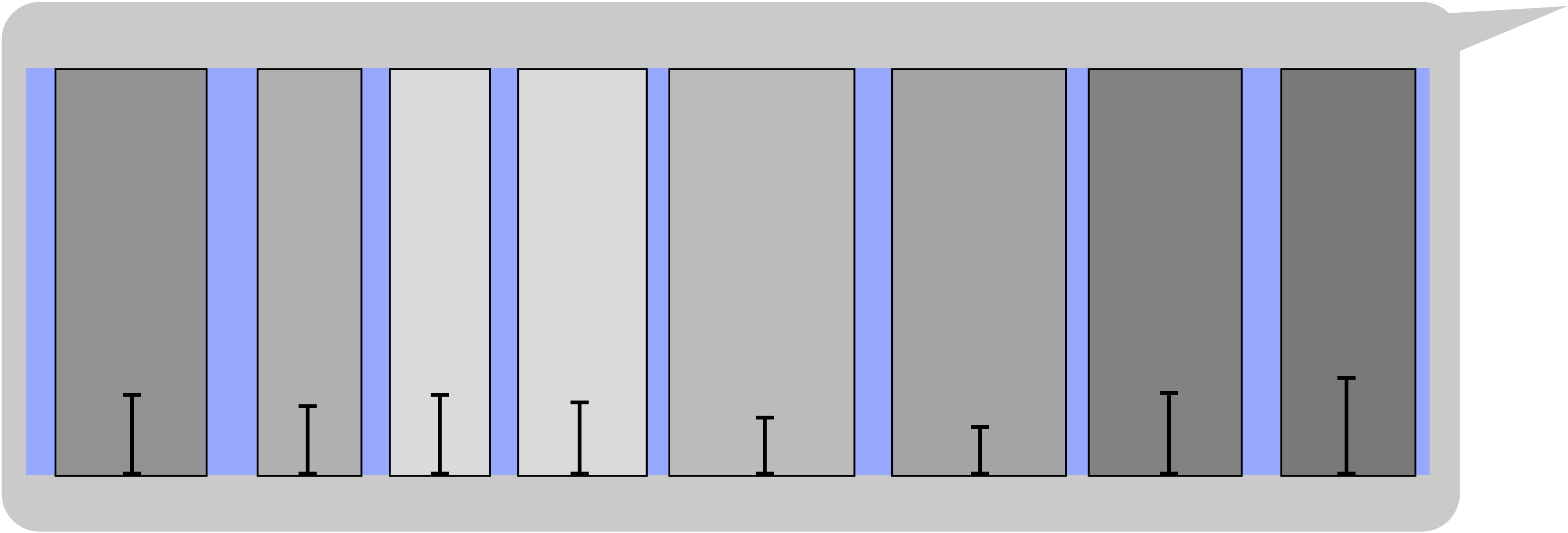
p r i s o n e r



Typesetting Model

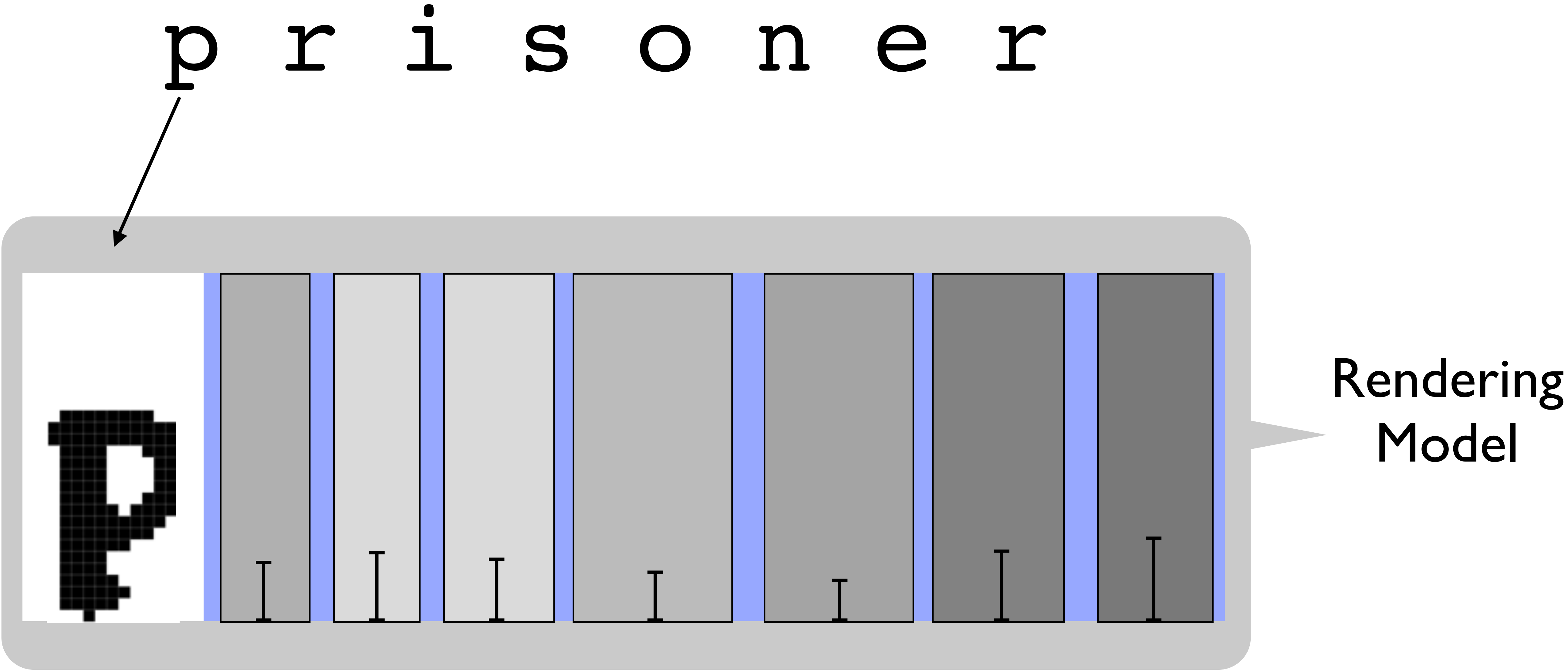
Generative Model

prisoner

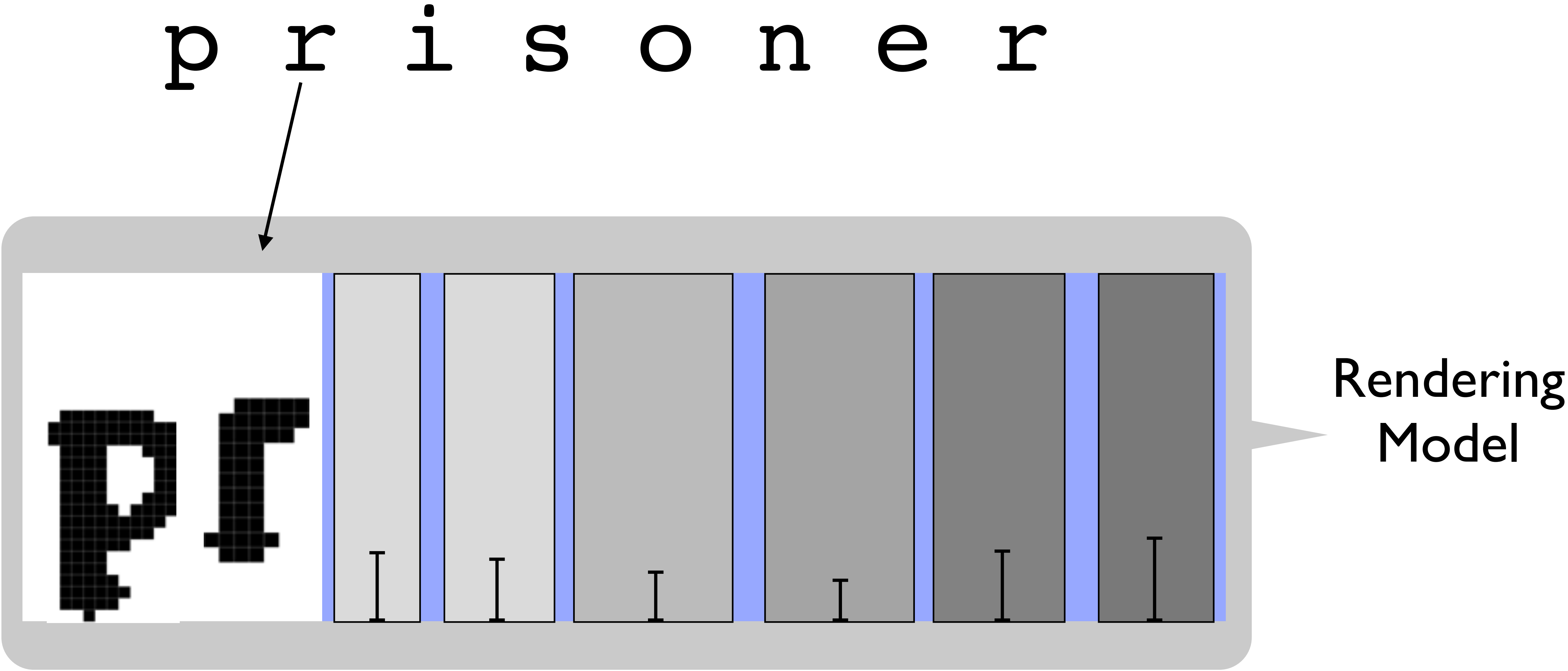


Typesetting Model

Generative Model

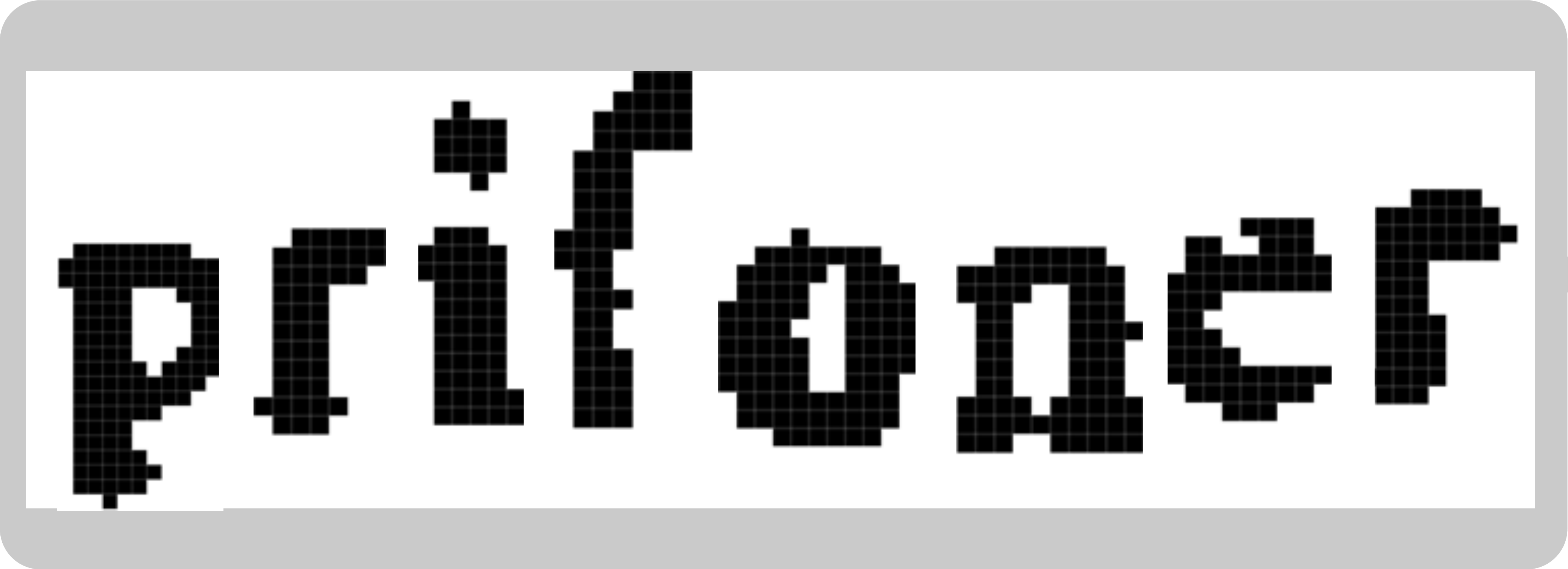


Generative Model



Generative Model

p r i s o n e r



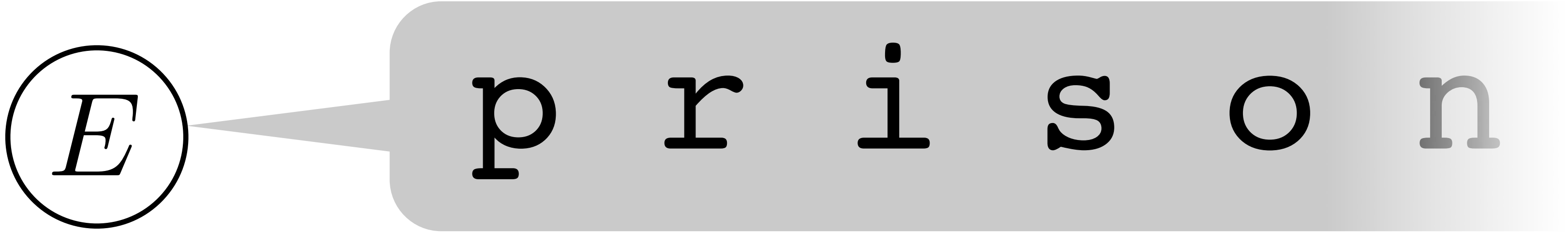
Rendering Model

Generative Model

prisoner

Generative Model

Language Model
$P(E)$



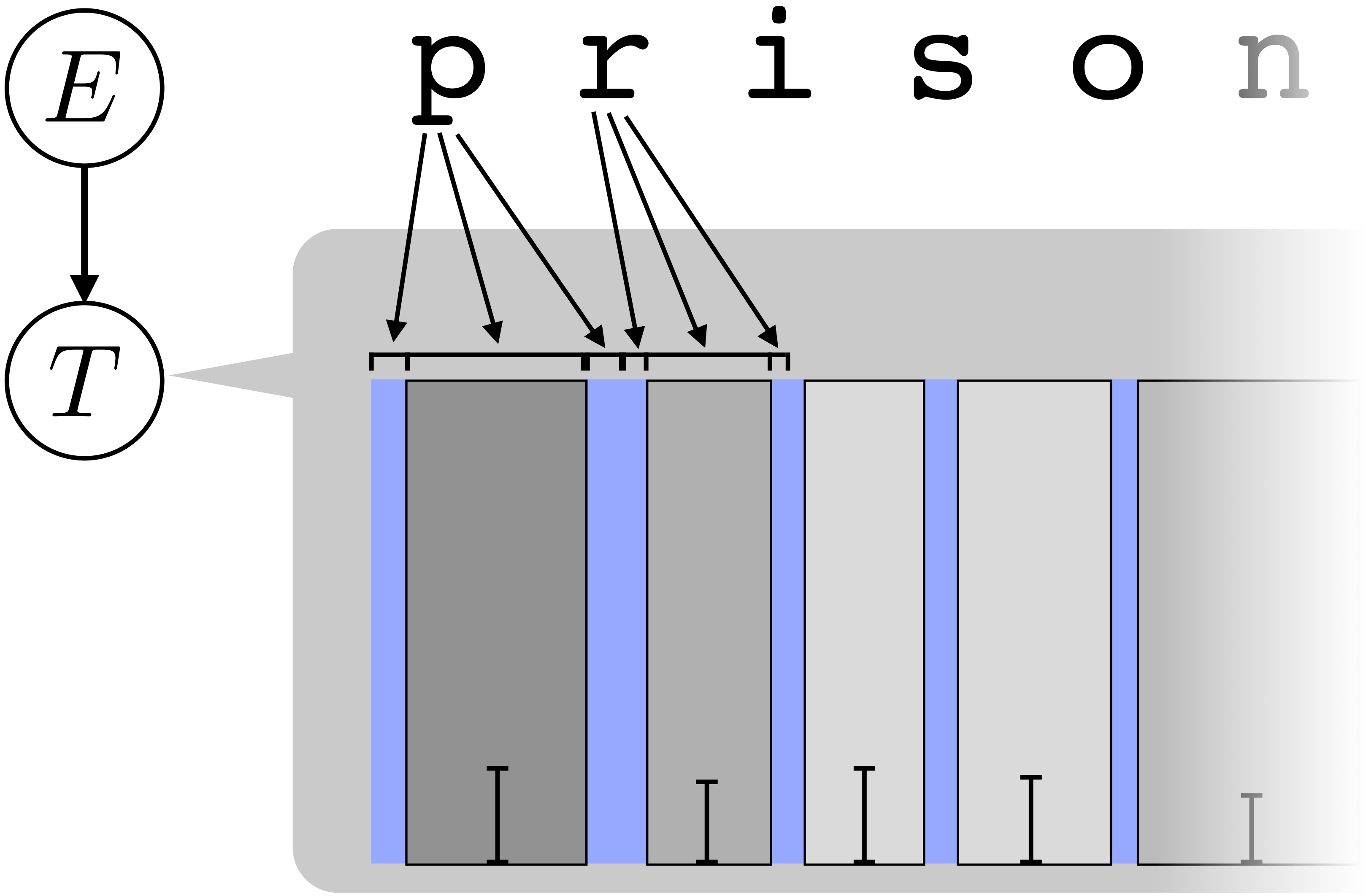
E

prison

Generative Model

Language Model
$P(E)$

Typesetting Model
$P(T|E)$

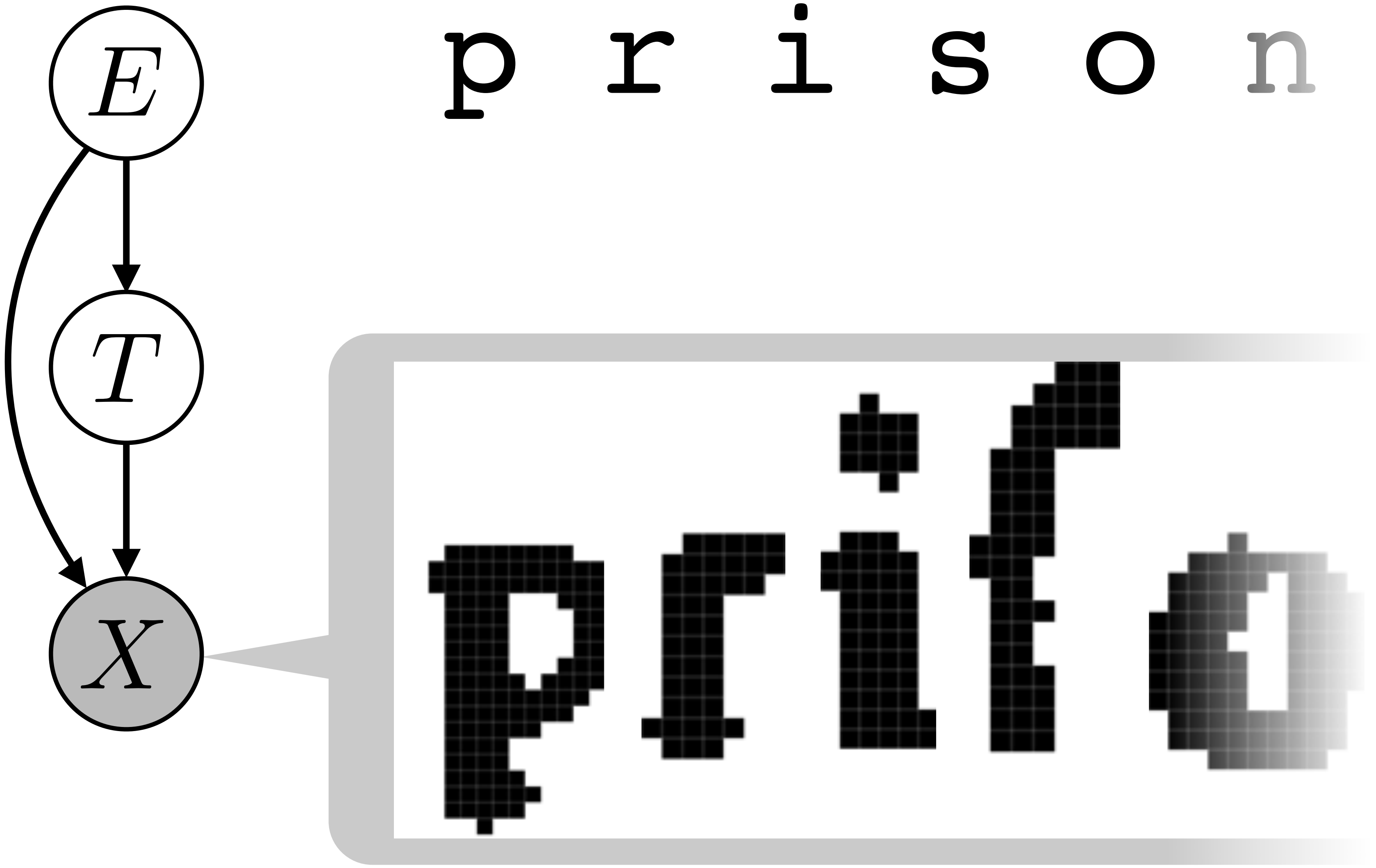


Generative Model

Language Model
$P(E)$

Typesetting Model
$P(T|E)$

Rendering Model
$P(X|E,T)$

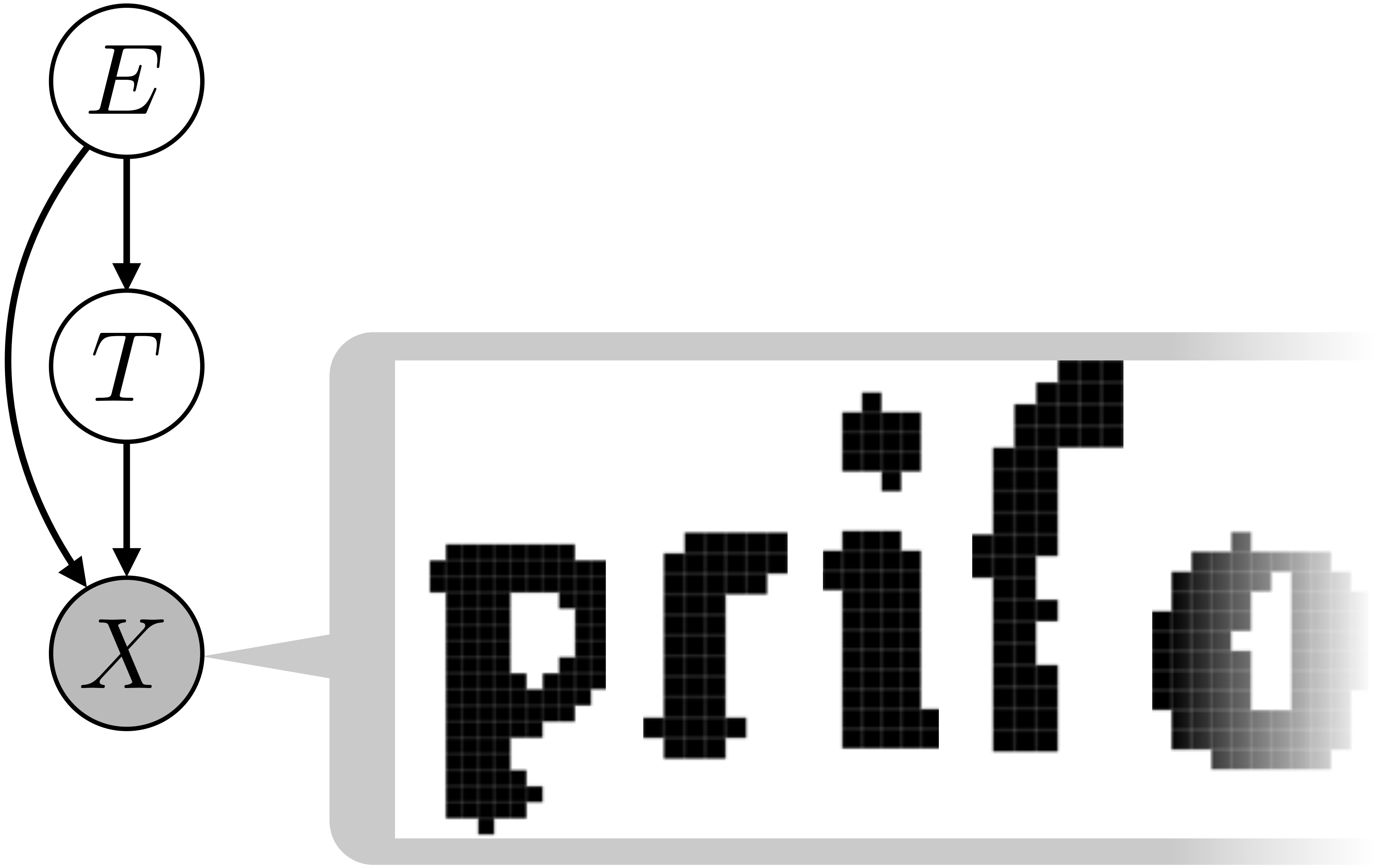


Generative Model

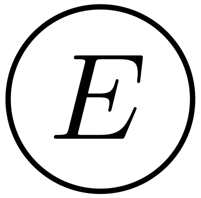
Language Model
$P(E)$

Typesetting Model
$P(T|E)$

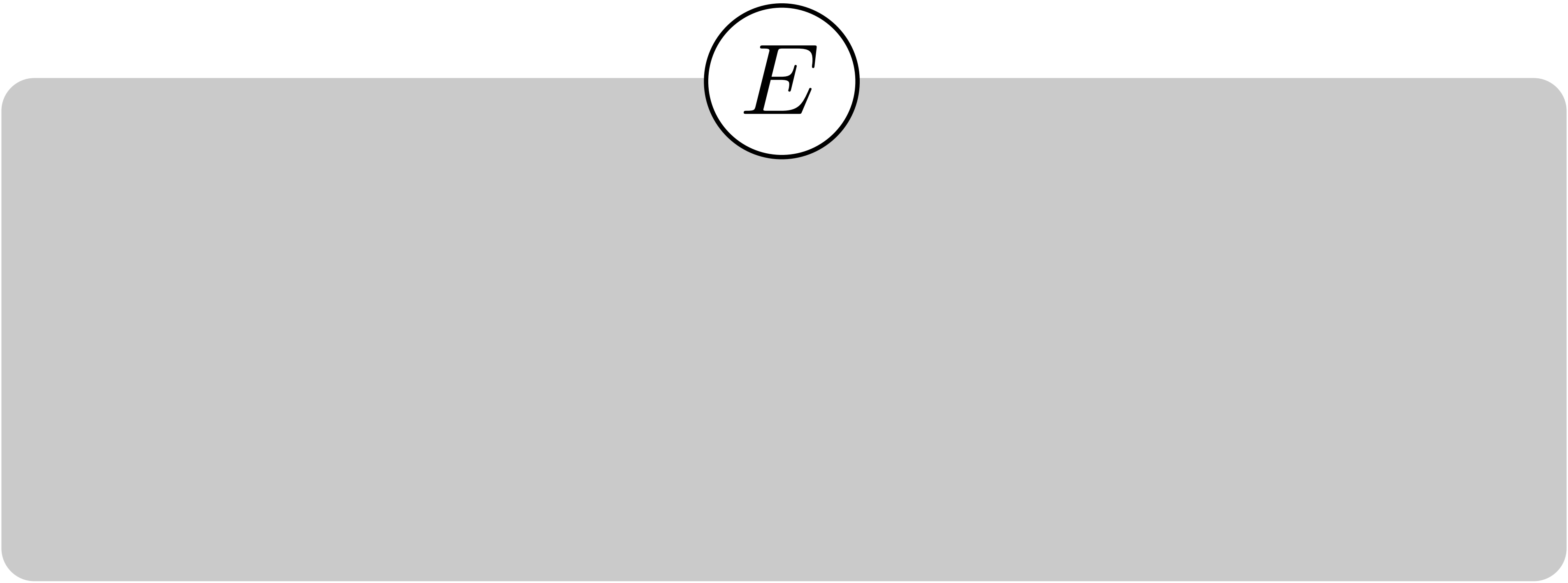
Rendering Model
$P(X|E,T)$



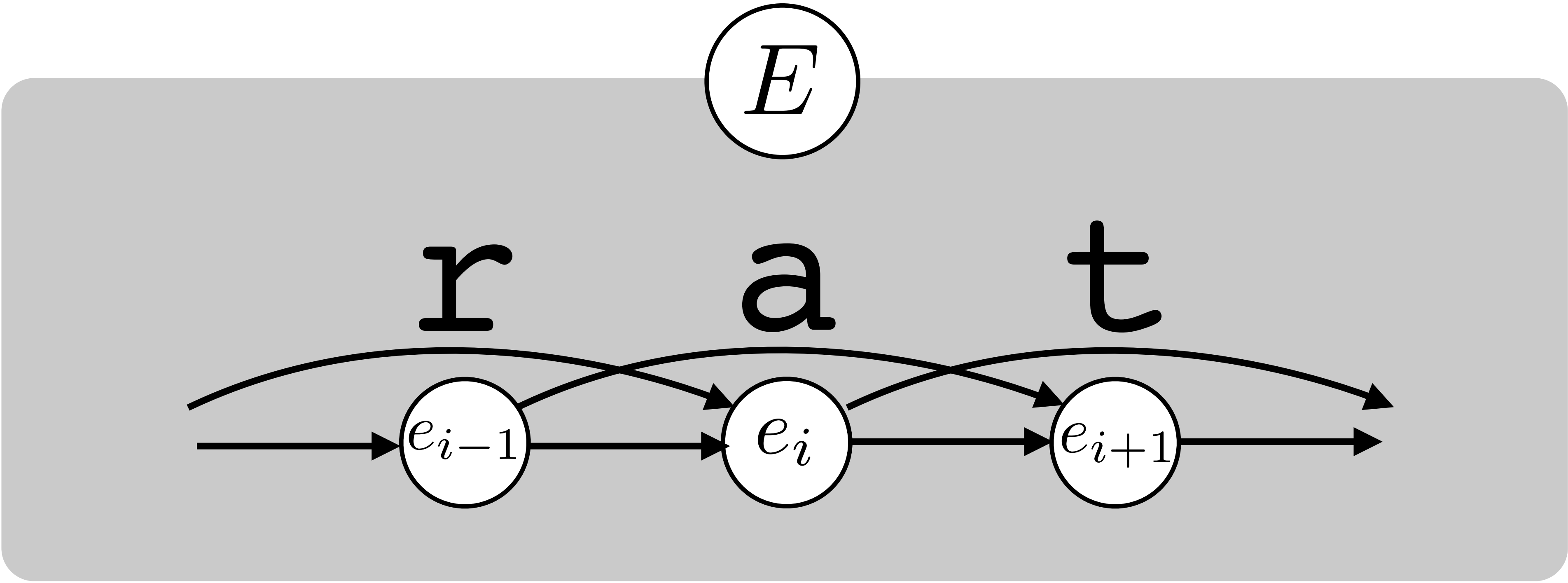
Language Model



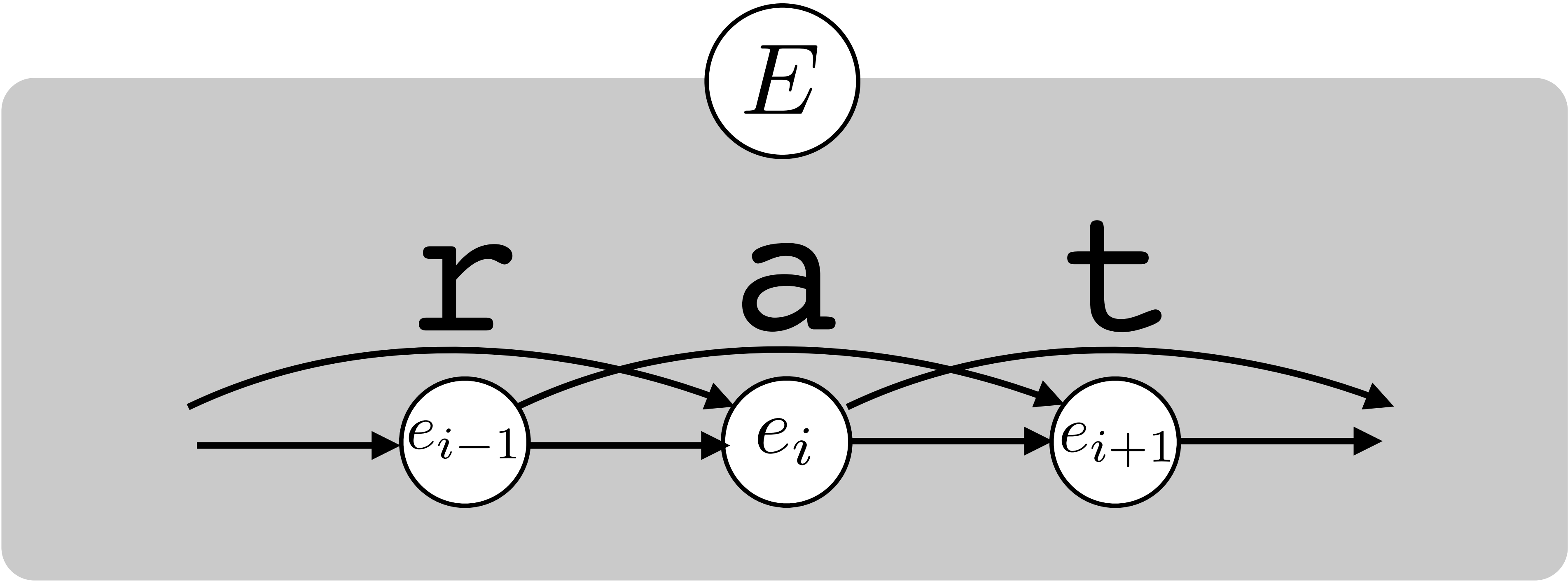
Language Model



Language Model

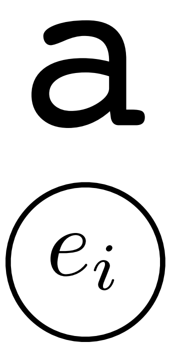


Language Model

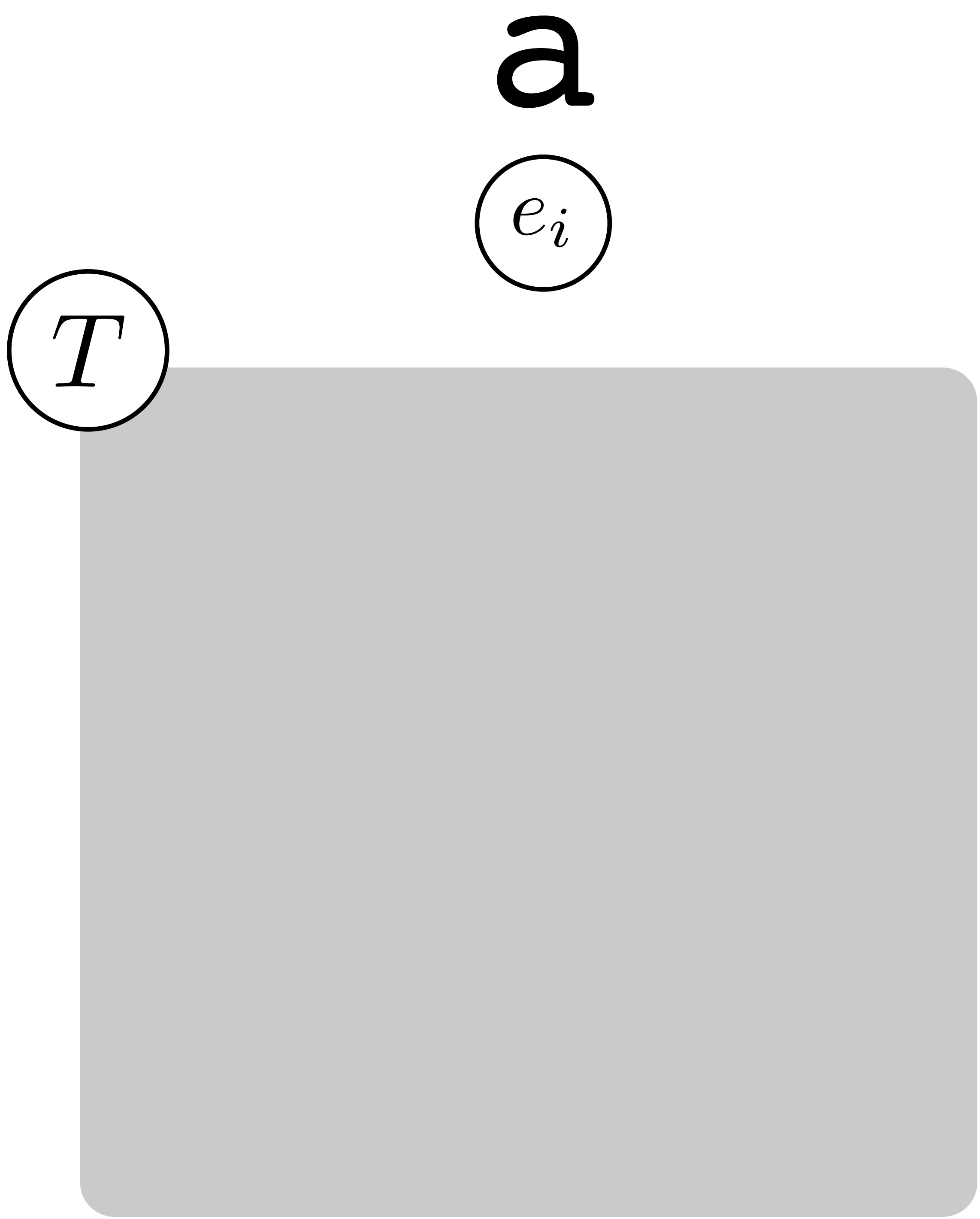


Kneser-Ney smoothed character 6-gram

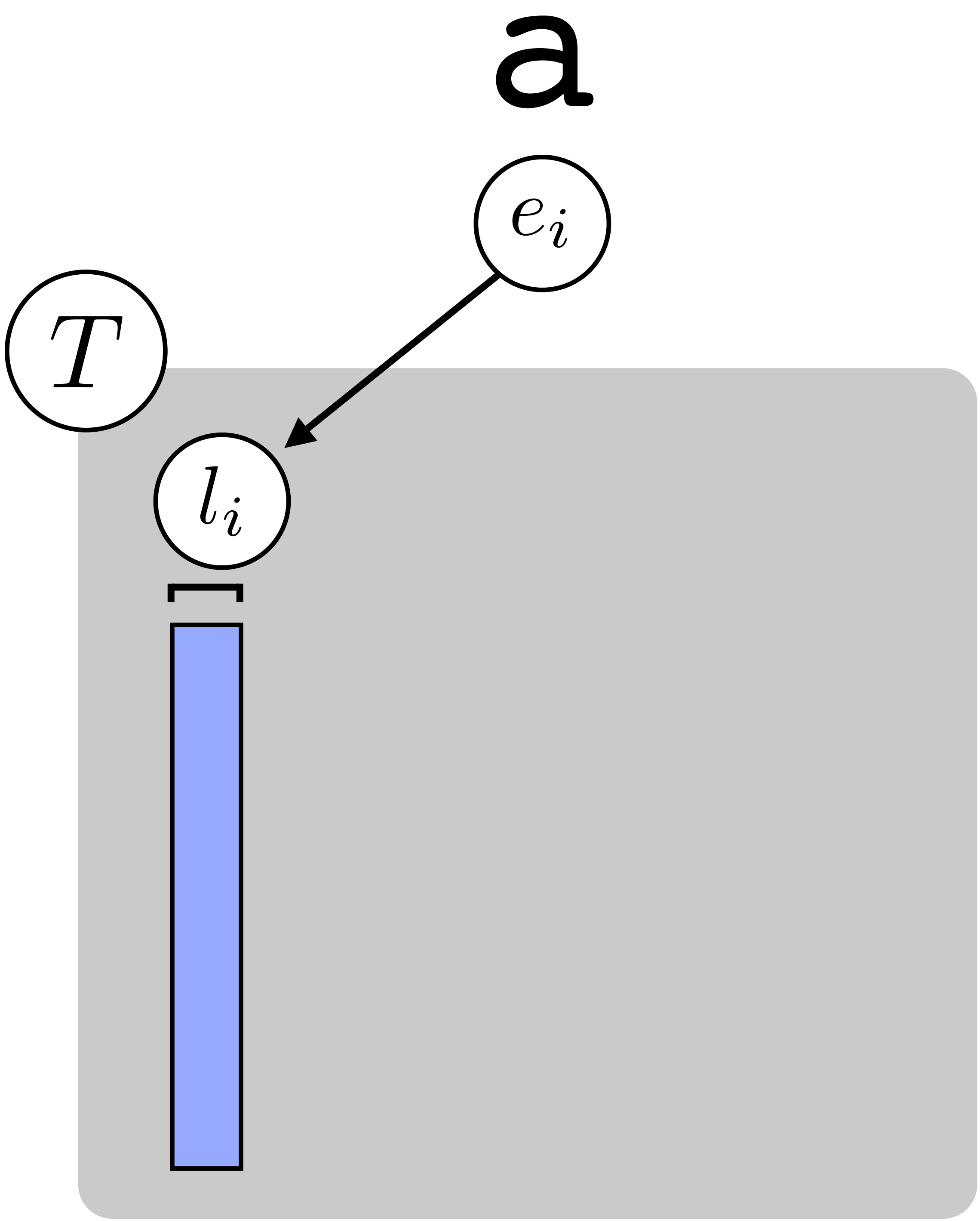
Typesetting Model

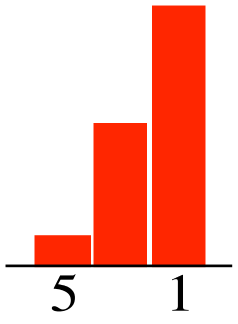


Typesetting Model



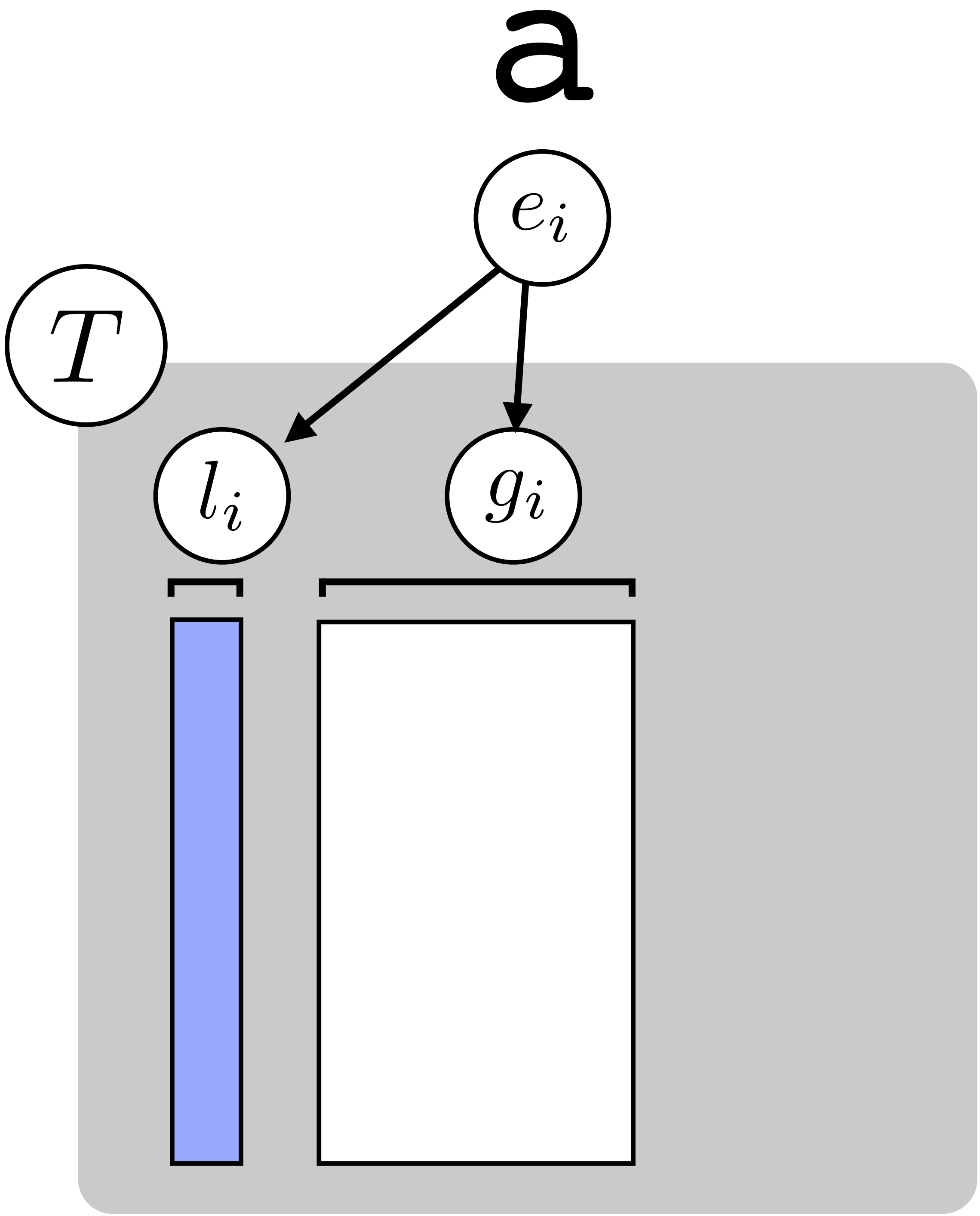
Typesetting Model

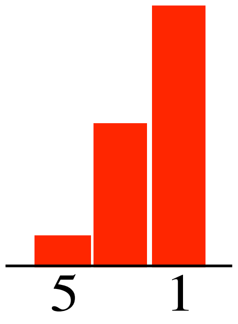




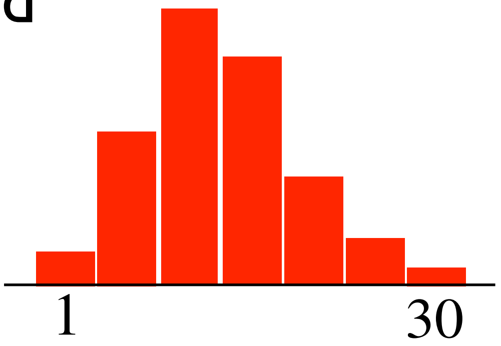
Left pad width

Typesetting Model



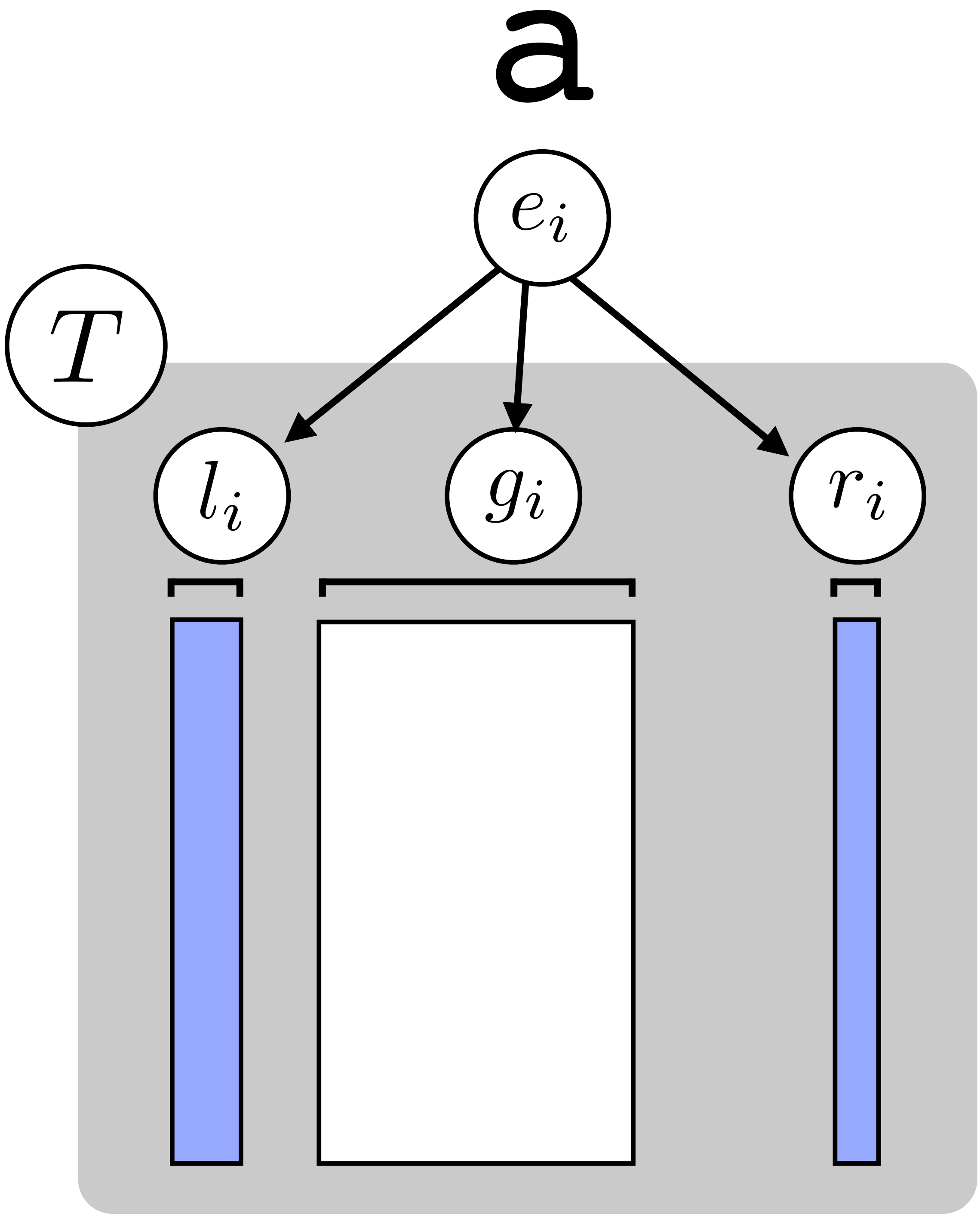


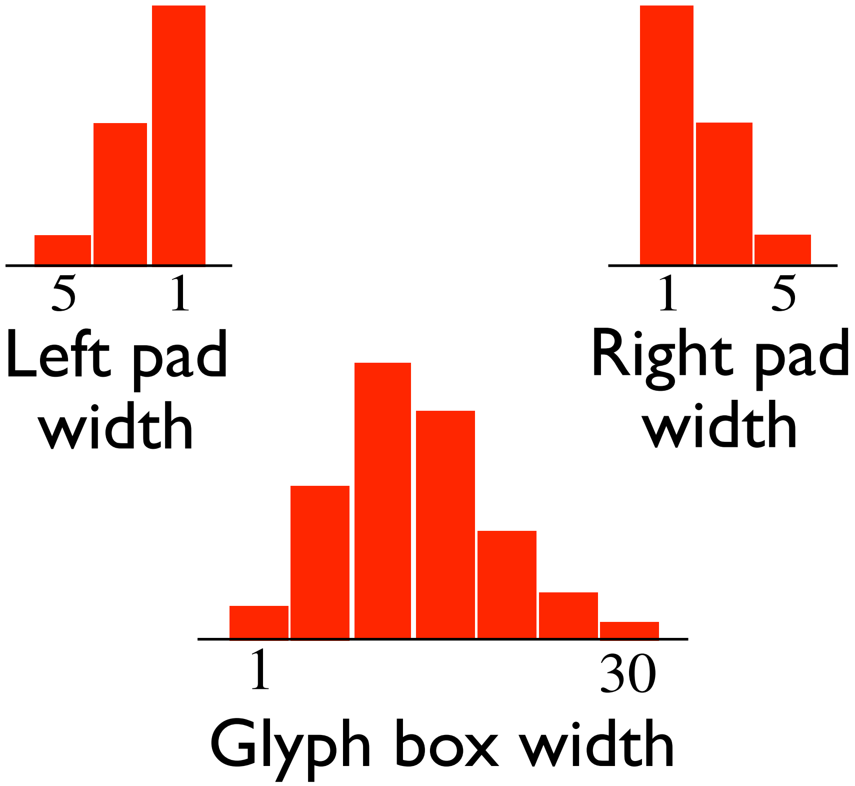
Left pad width



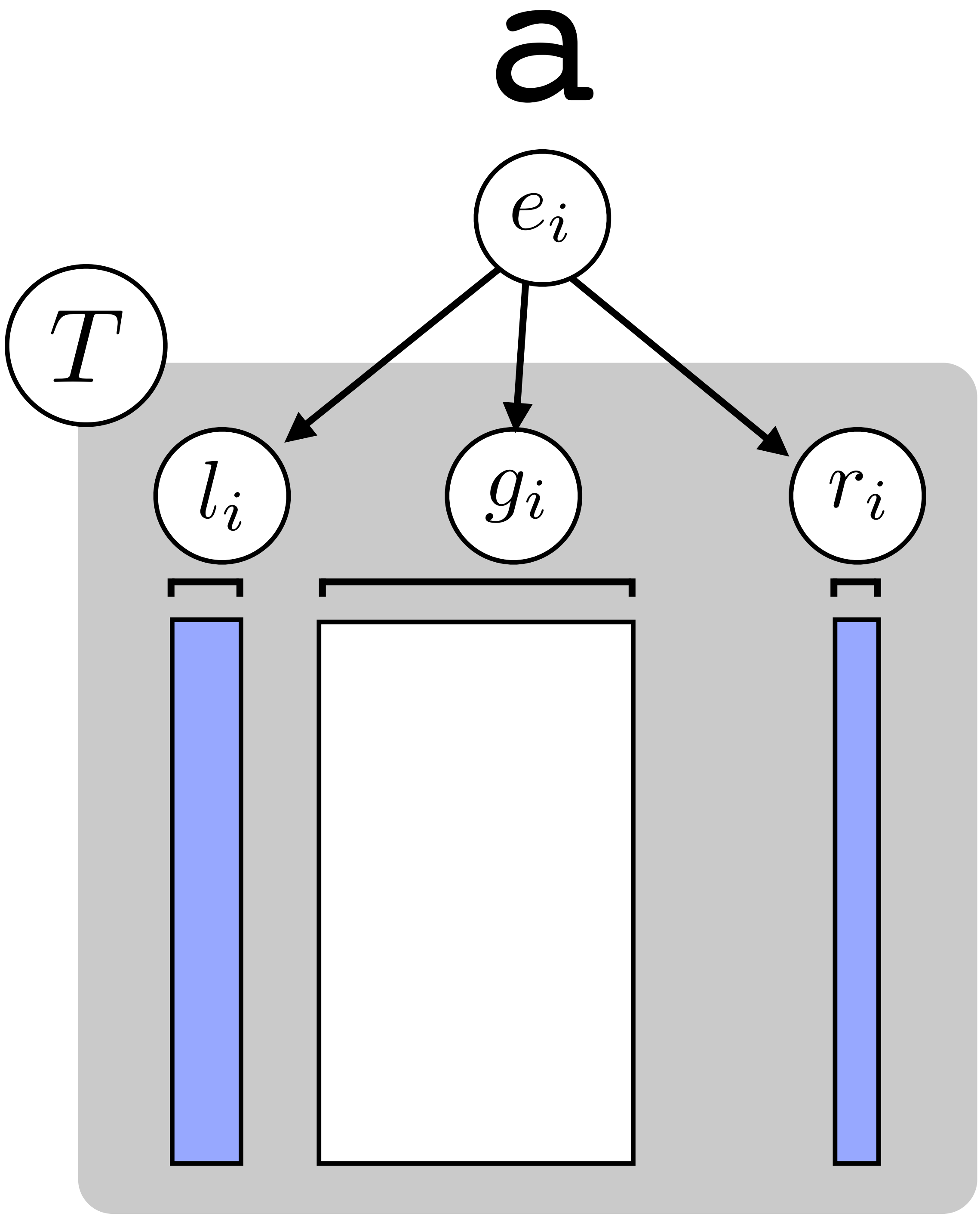
Glyph box width

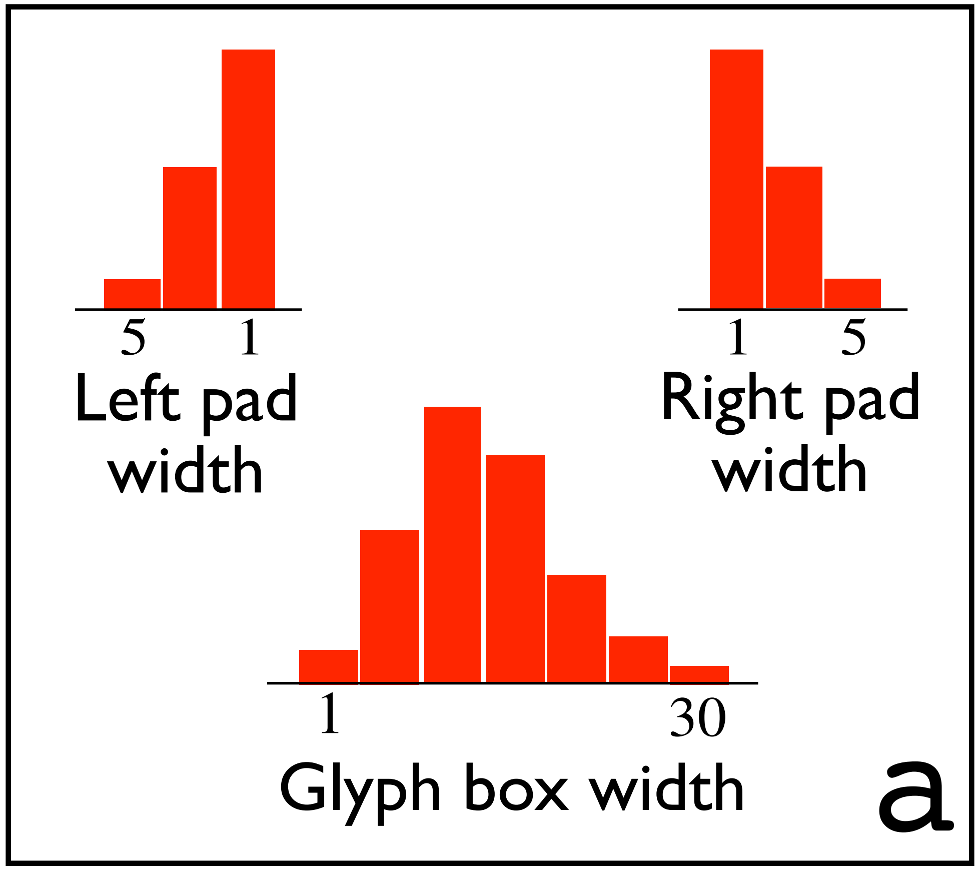
Typesetting Model



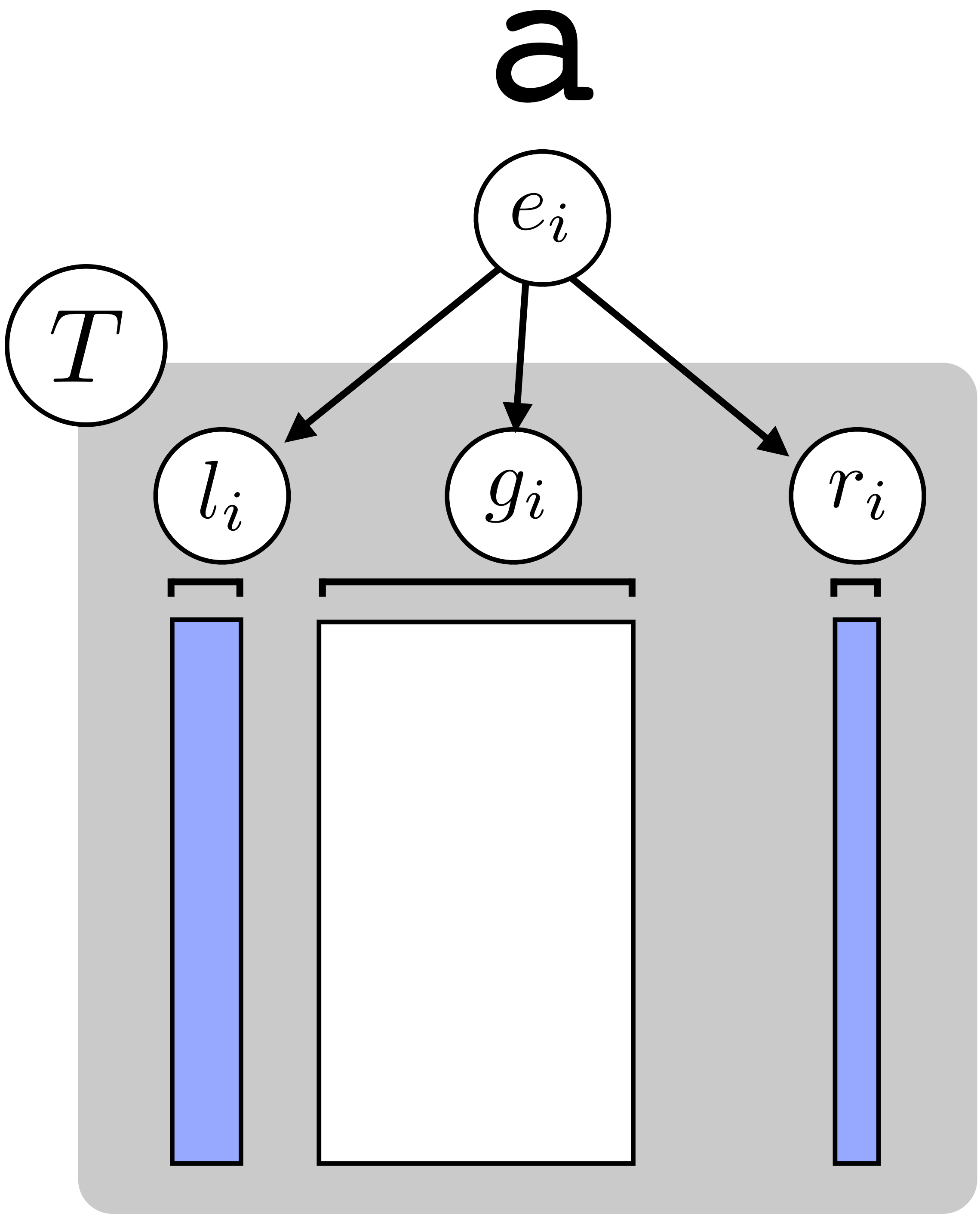


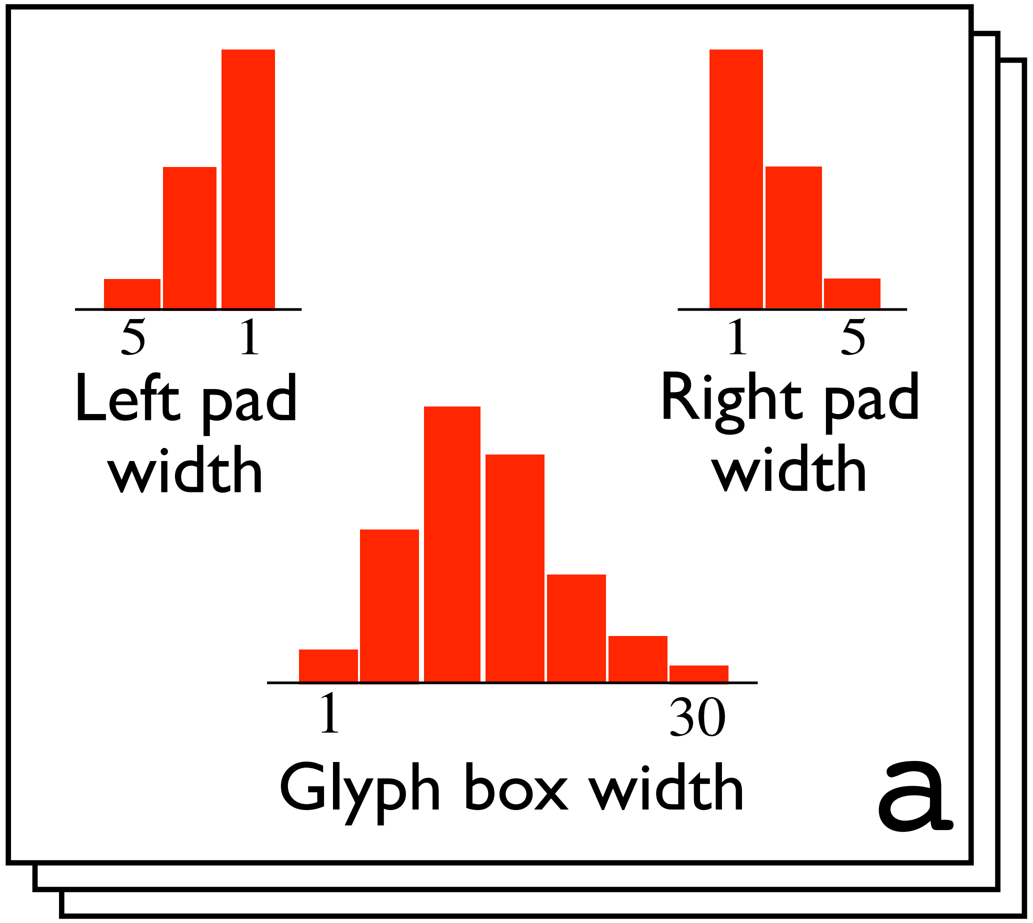
Typesetting Model



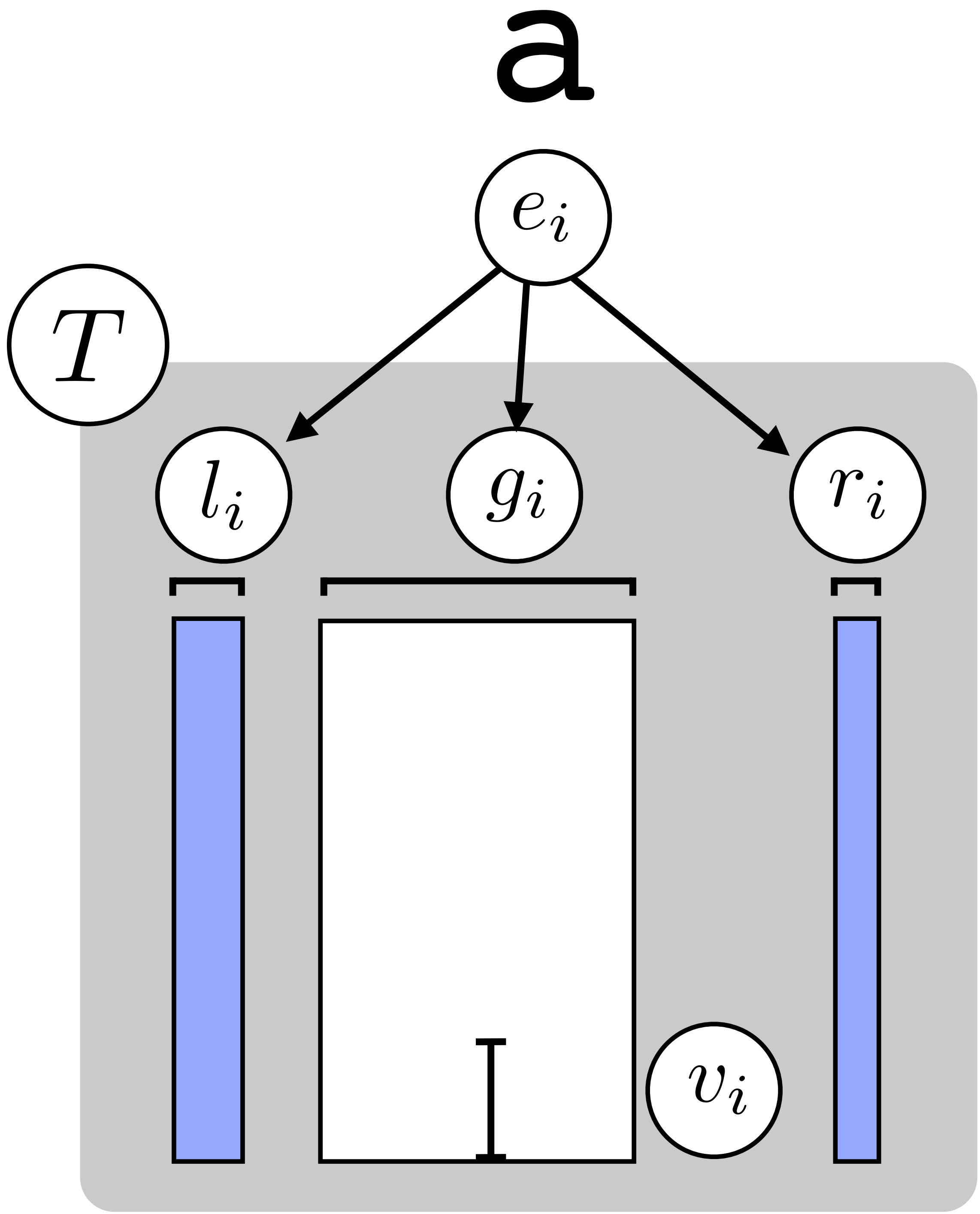


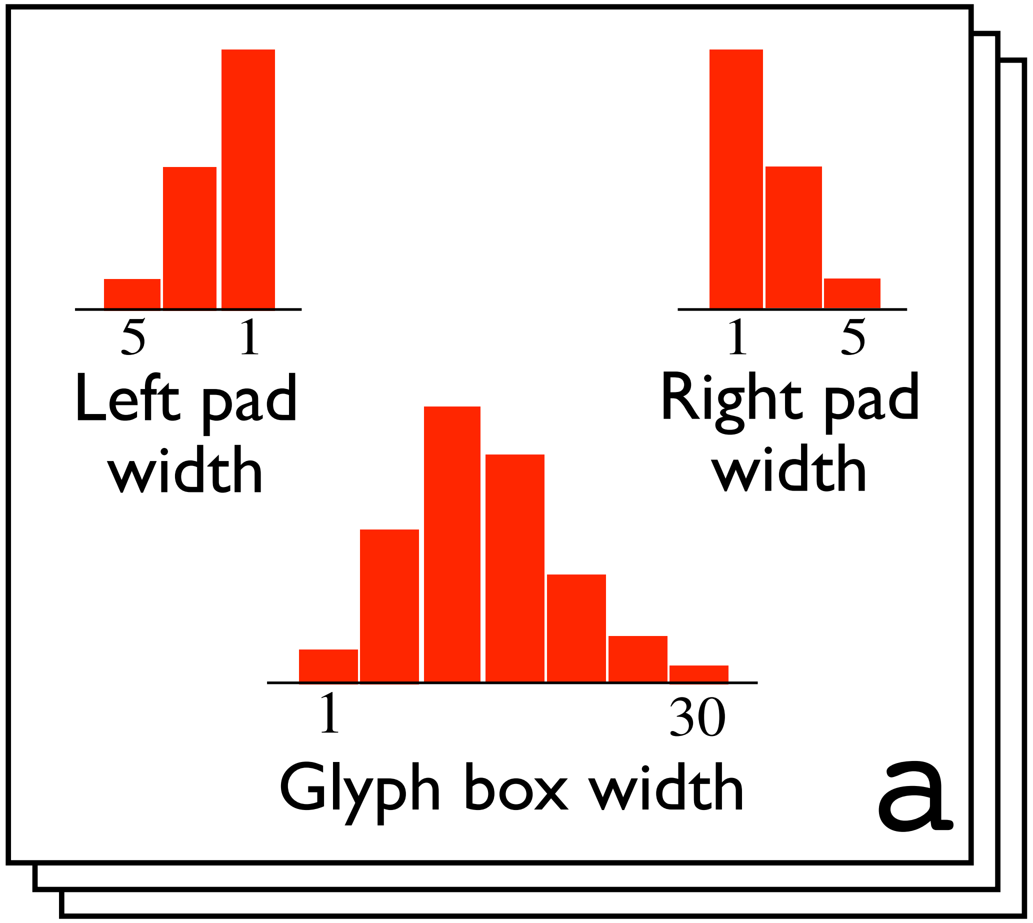
Typesetting Model

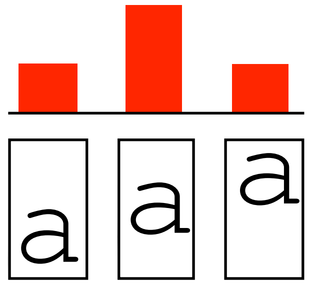




Typesetting Model

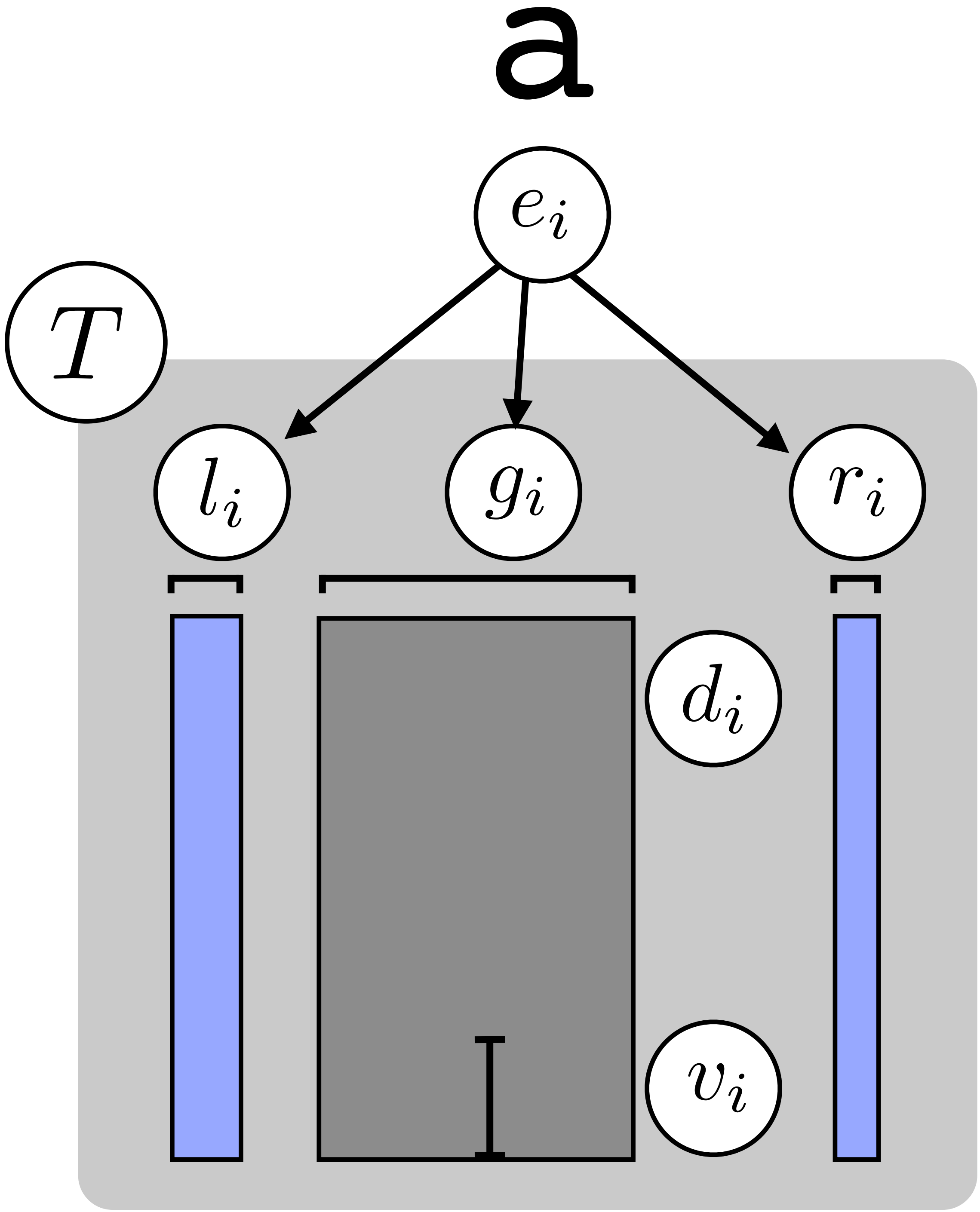


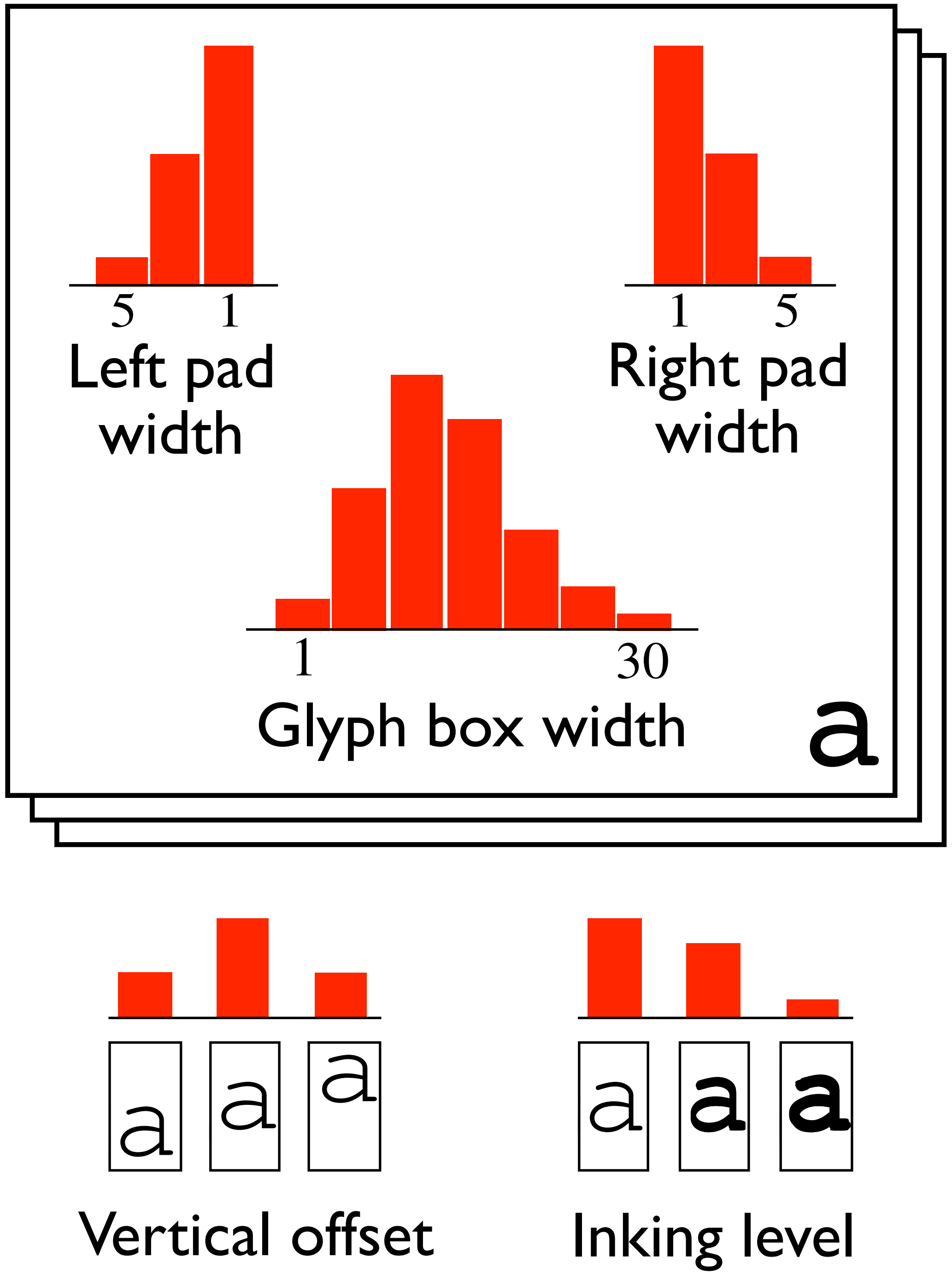




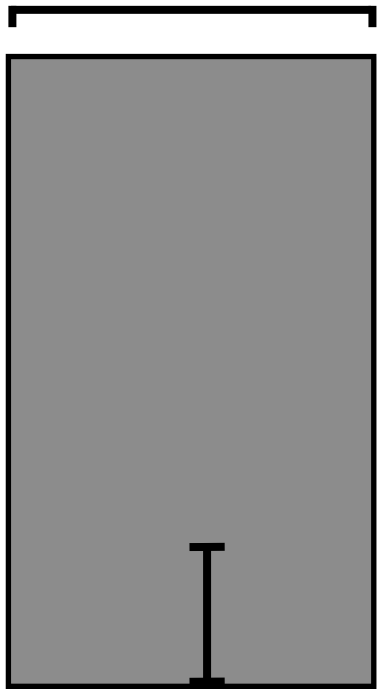
Vertical offset

Typesetting Model

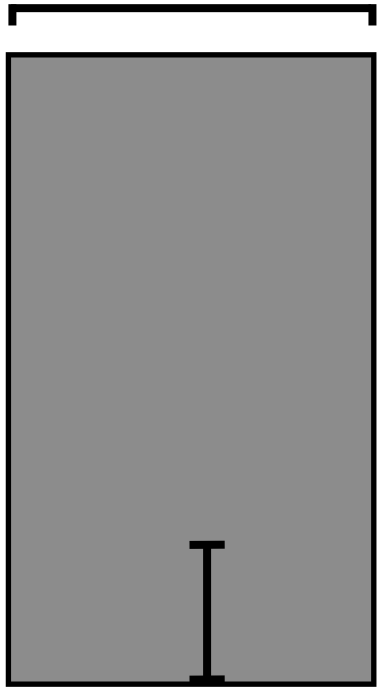




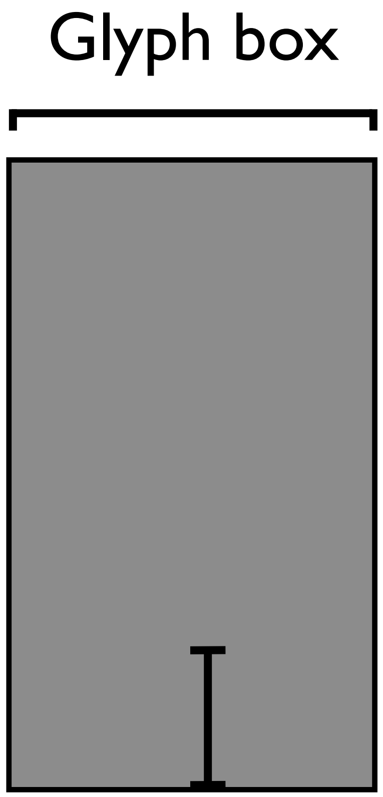
Rendering Model



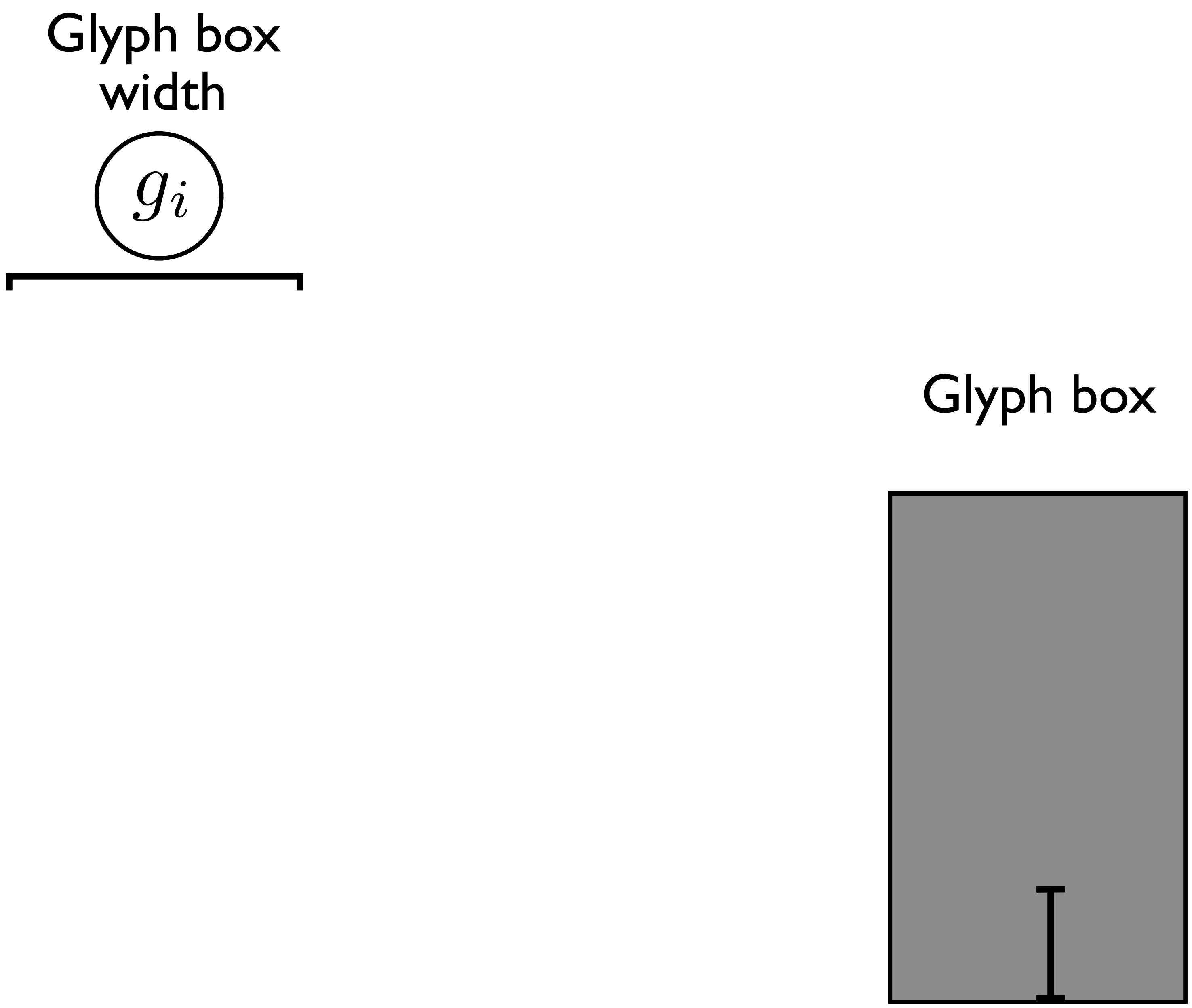
Rendering Model



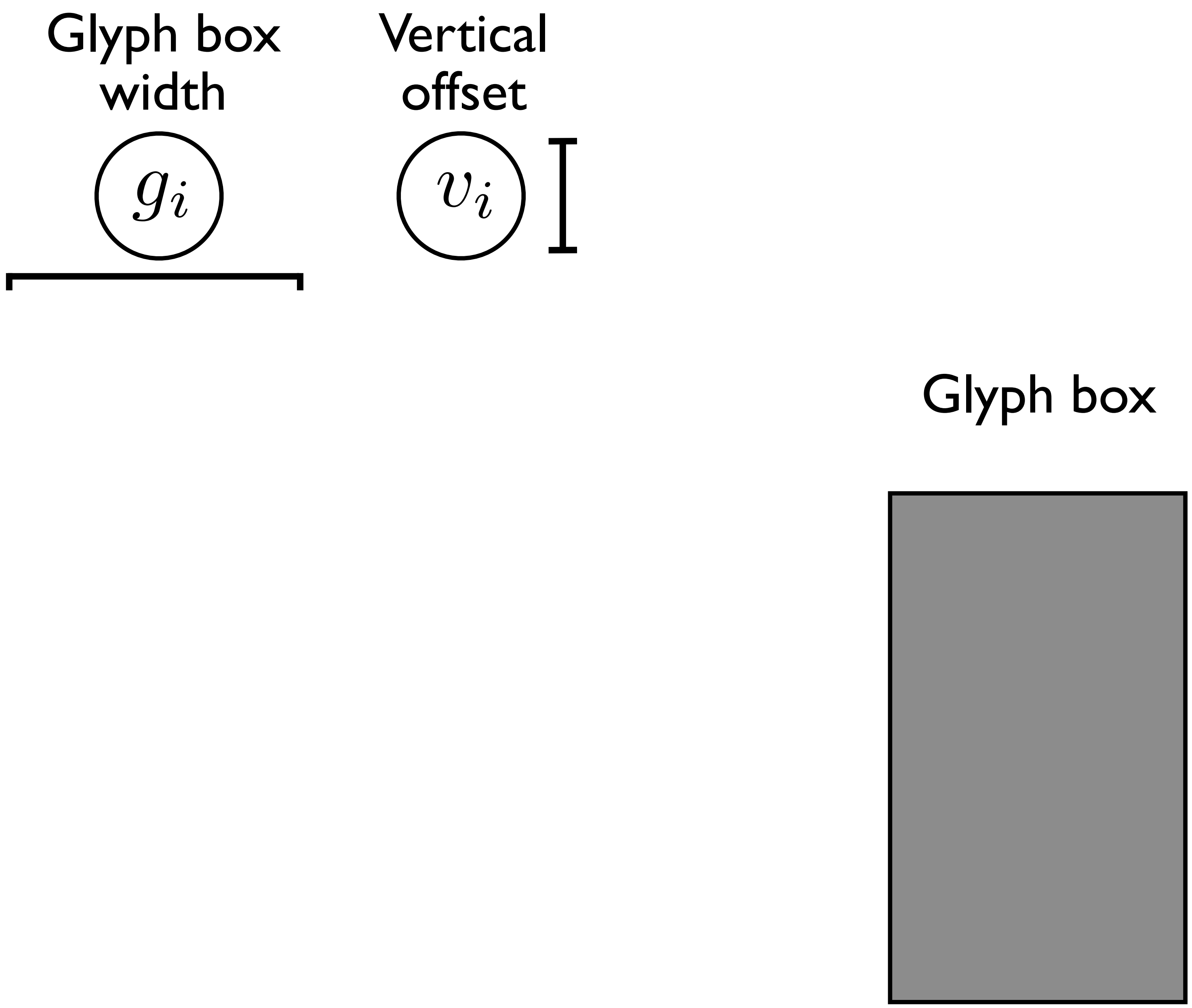
Rendering Model



Rendering Model



Rendering Model



Glyph box width

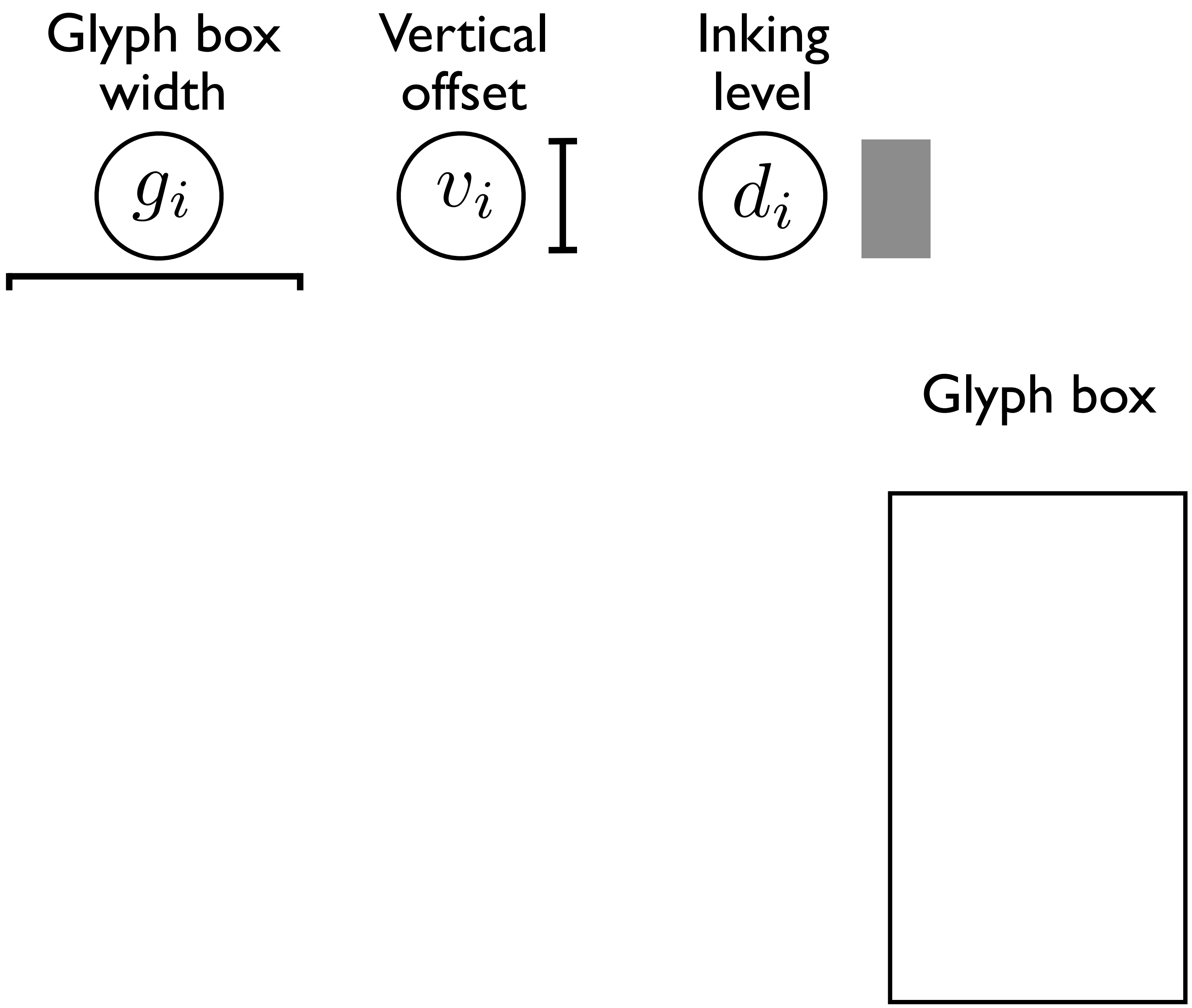
g_i

Vertical offset

v_i

Glyph box

Rendering Model



Glyph box width

g_i

Vertical offset

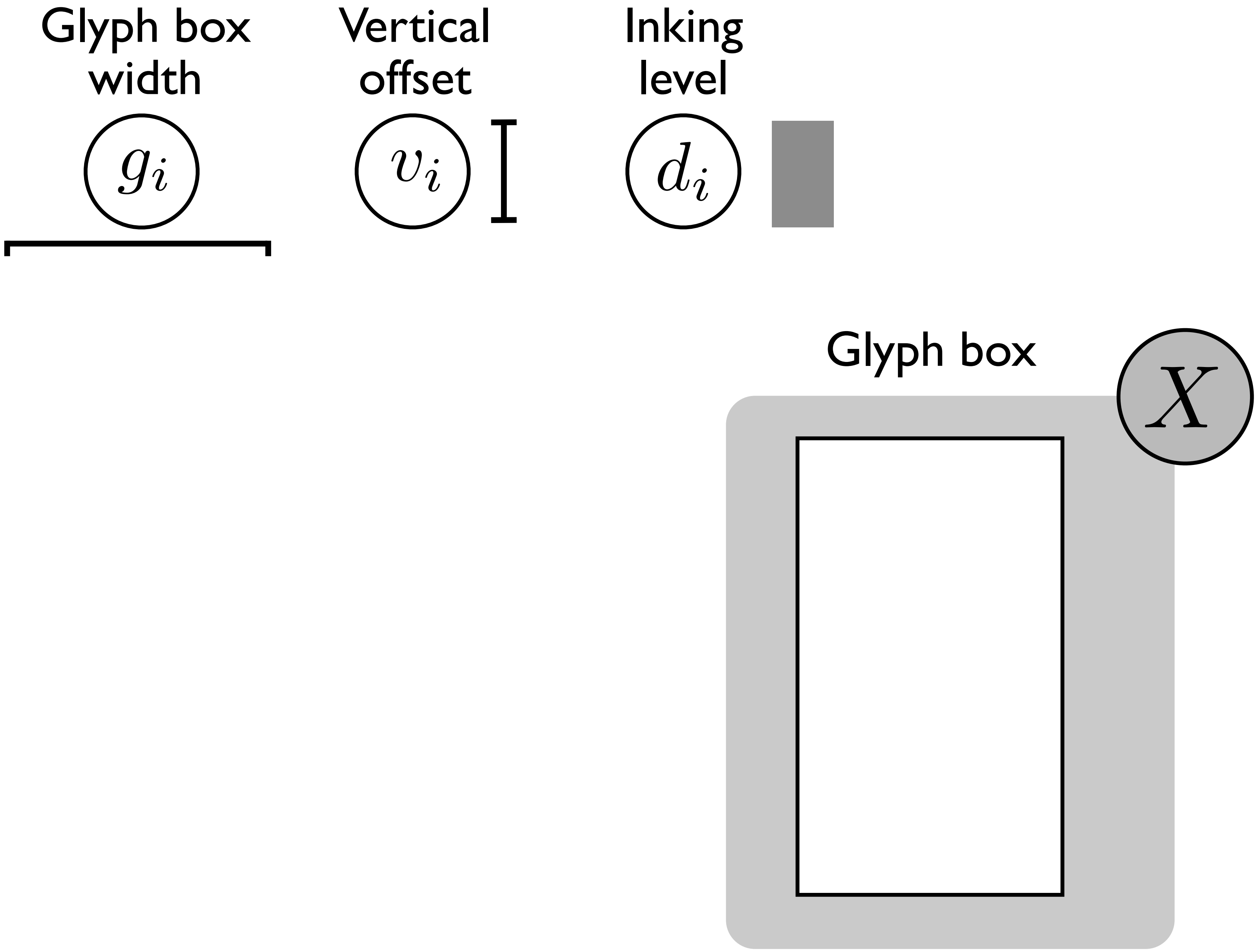
v_i

Inking level

d_i

Glyph box

Rendering Model



Glyph box width

g_i

Vertical offset

v_i

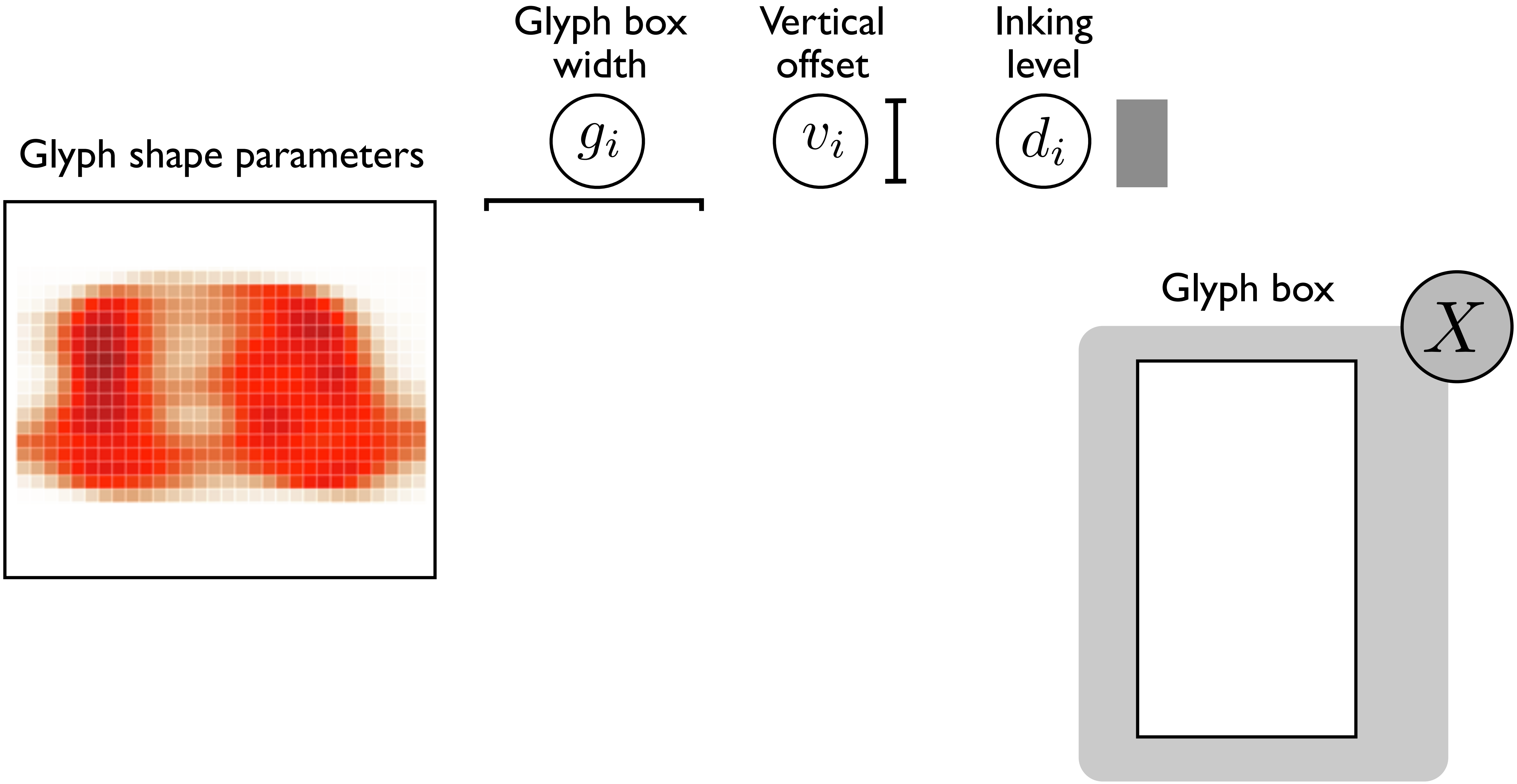
Inking level

d_i

Glyph box

X

Rendering Model



Glyph shape parameters

Glyph box width

g_i

Vertical offset

v_i

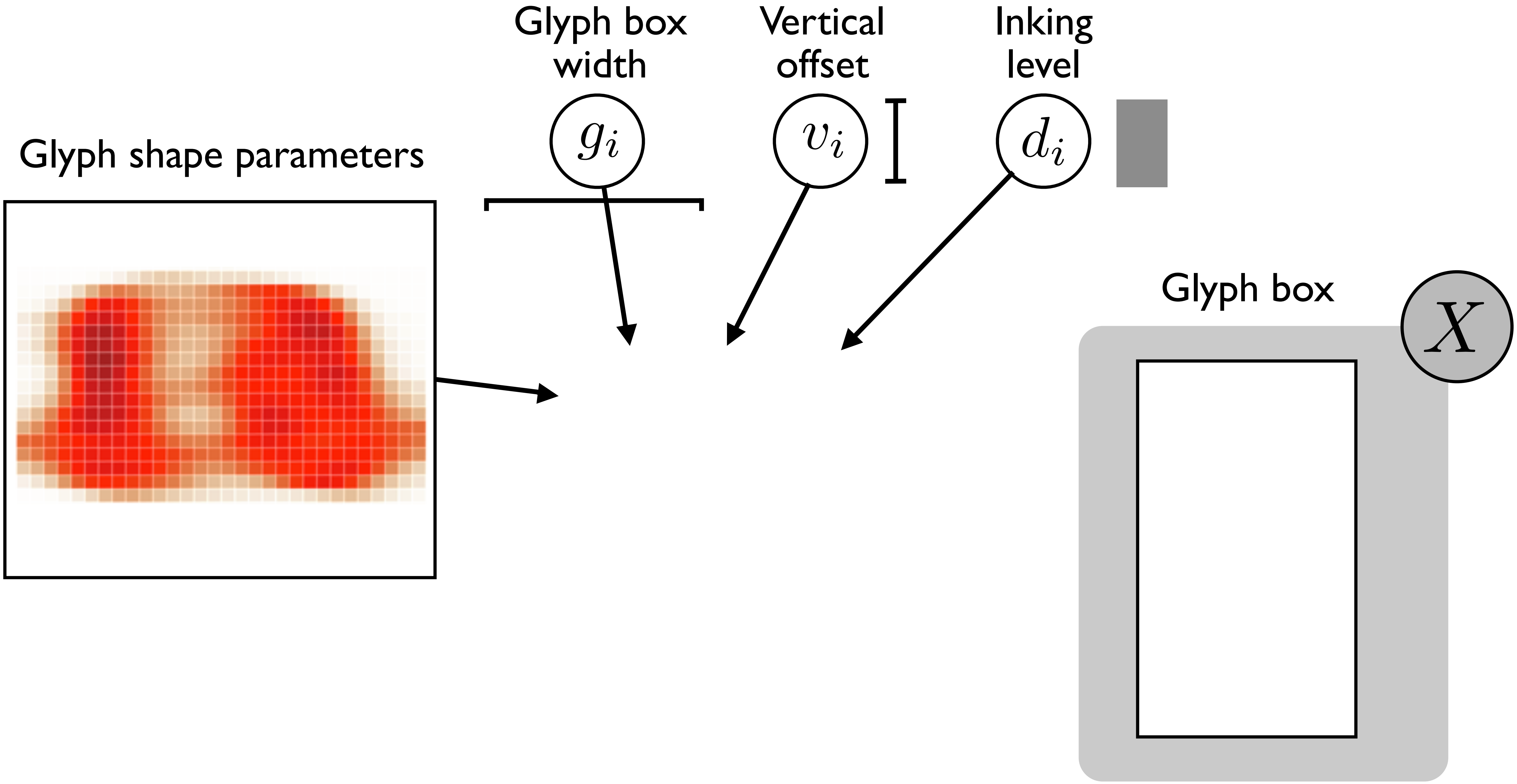
Inking level

d_i

Glyph box

X

Rendering Model



Glyph box width

g_i

Vertical offset

v_i

Inking level

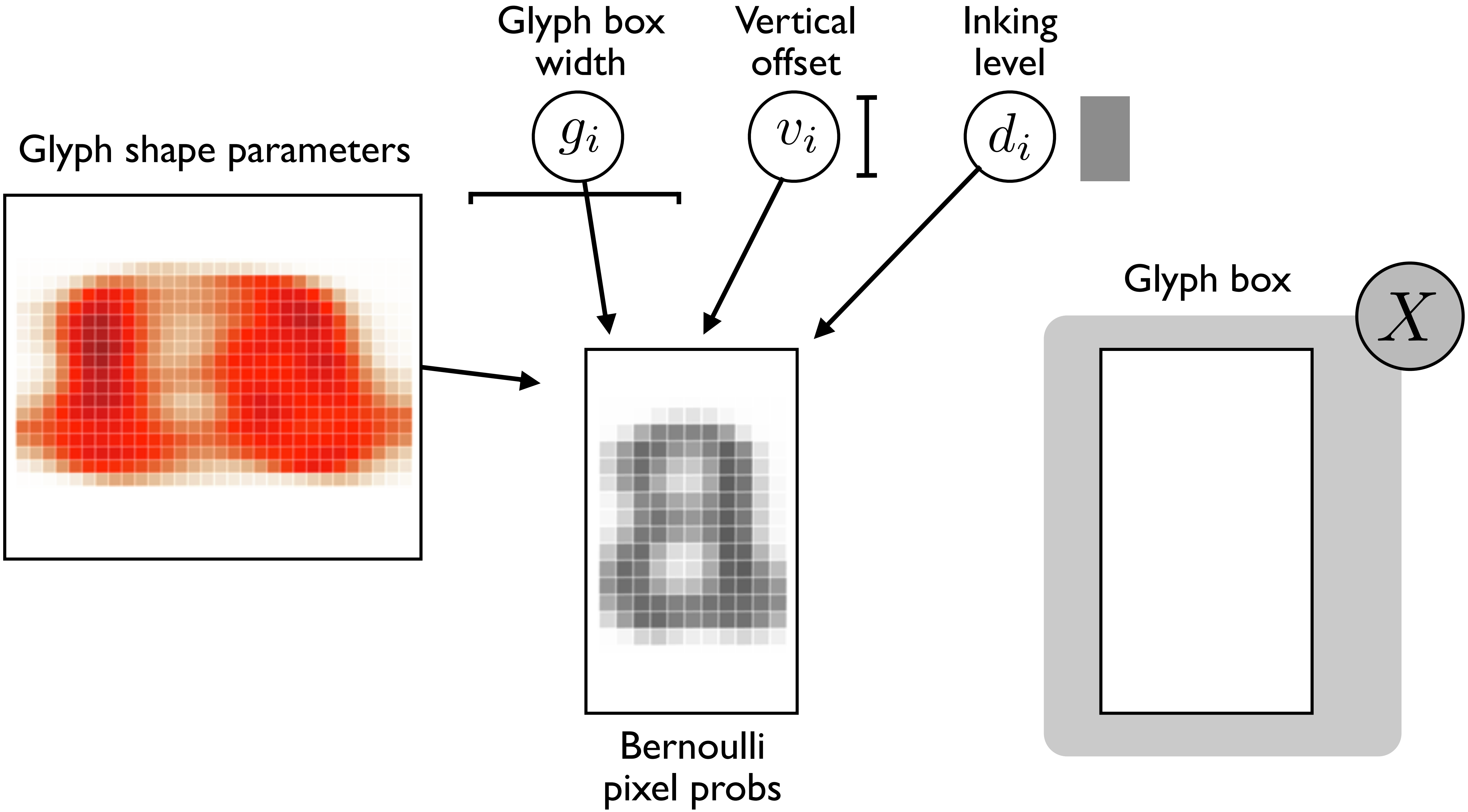
d_i

Glyph shape parameters

Glyph box

X

Rendering Model



Glyph box width

Vertical offset

Inking level

g_i

v_i

d_i

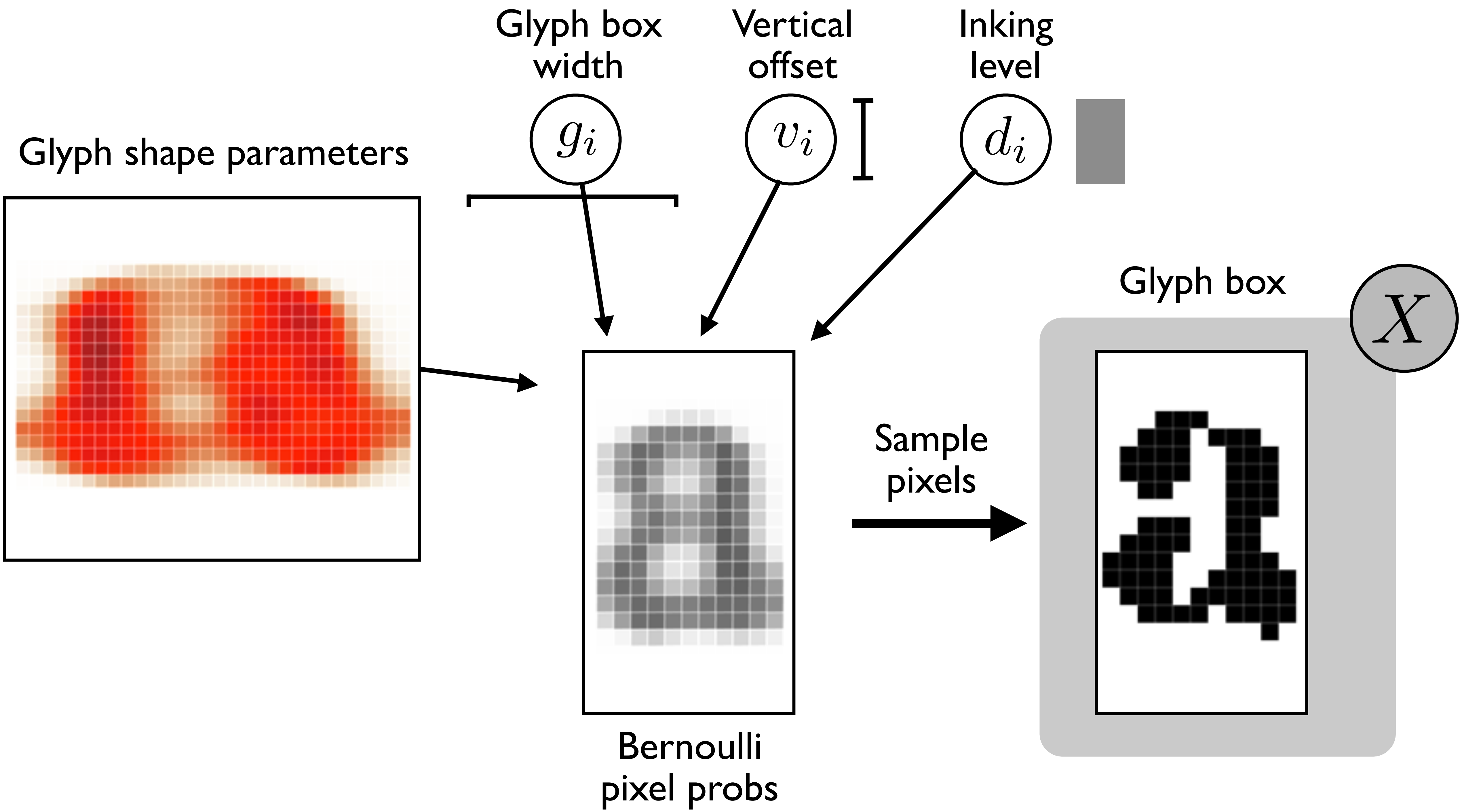
Glyph shape parameters

Glyph box

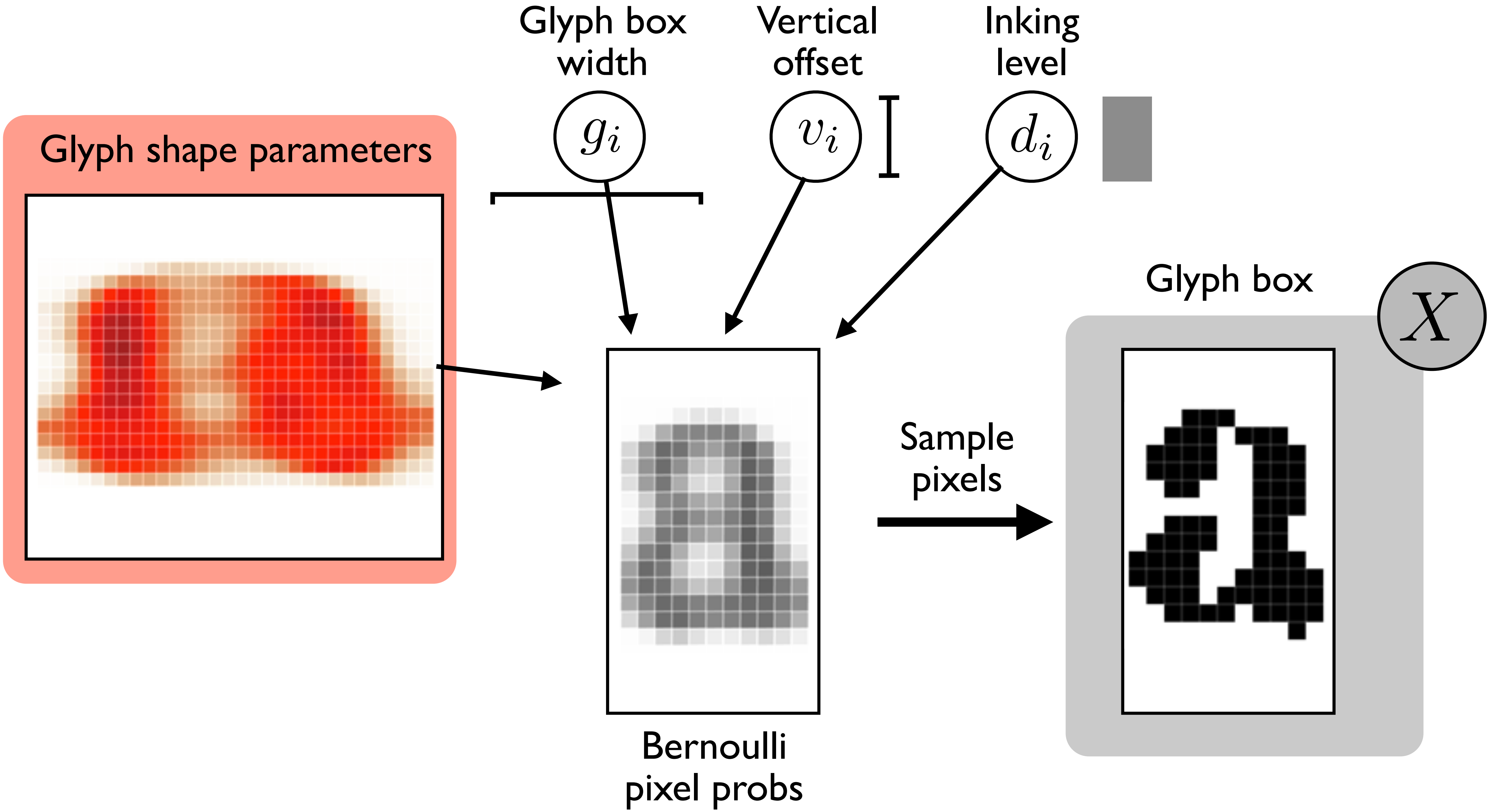
X

Bernoulli pixel probs

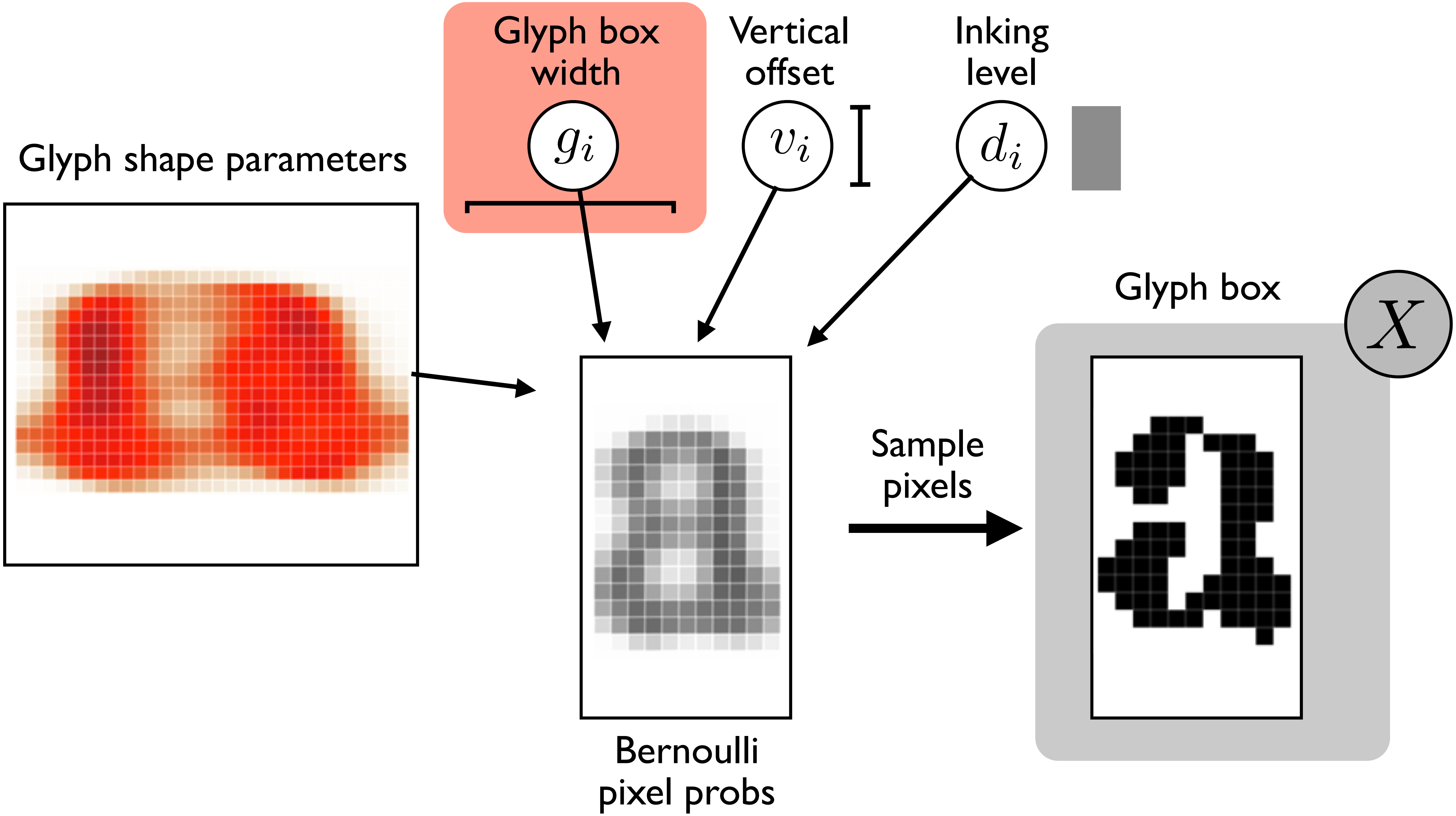
Rendering Model



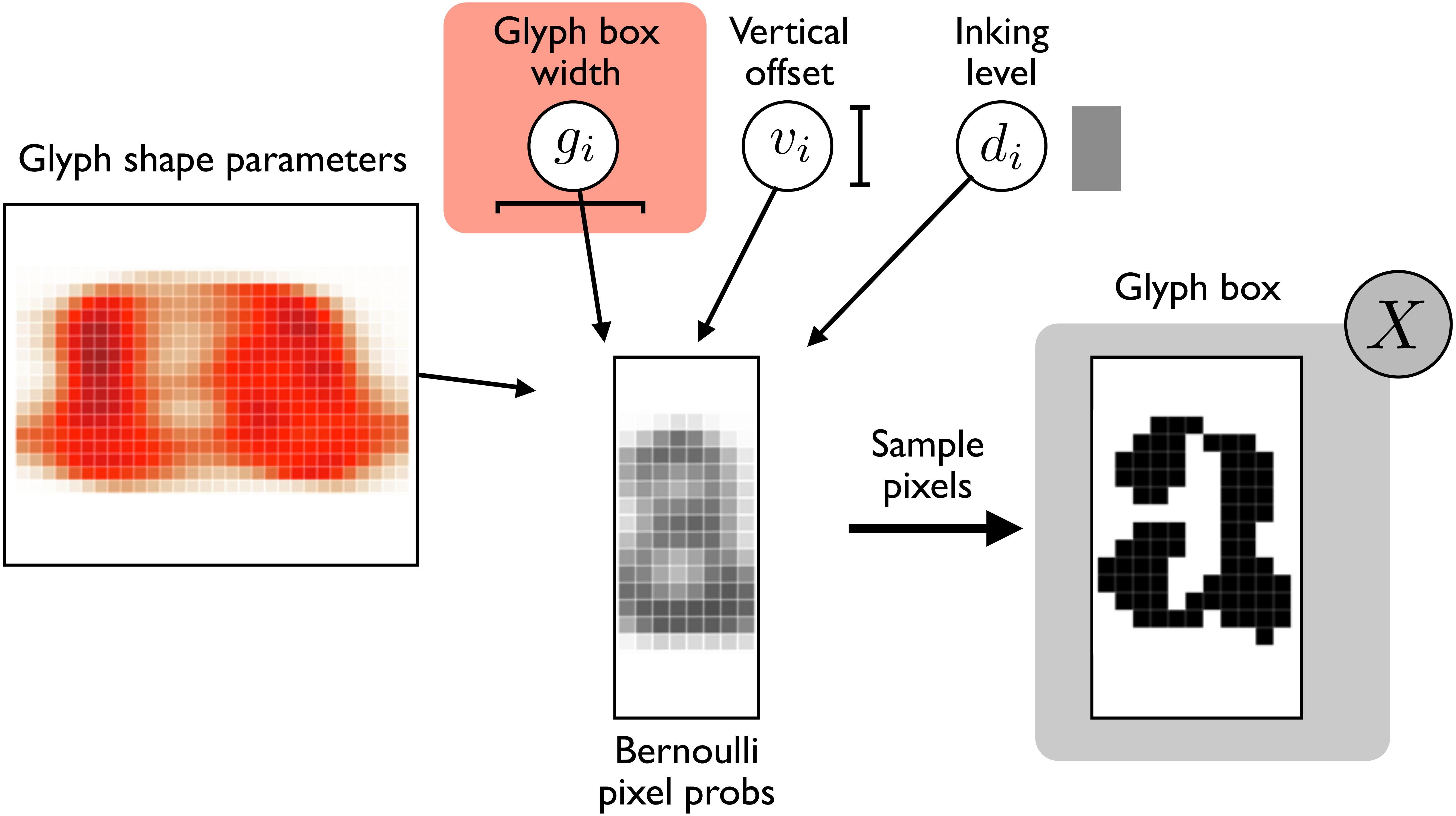
Rendering Model



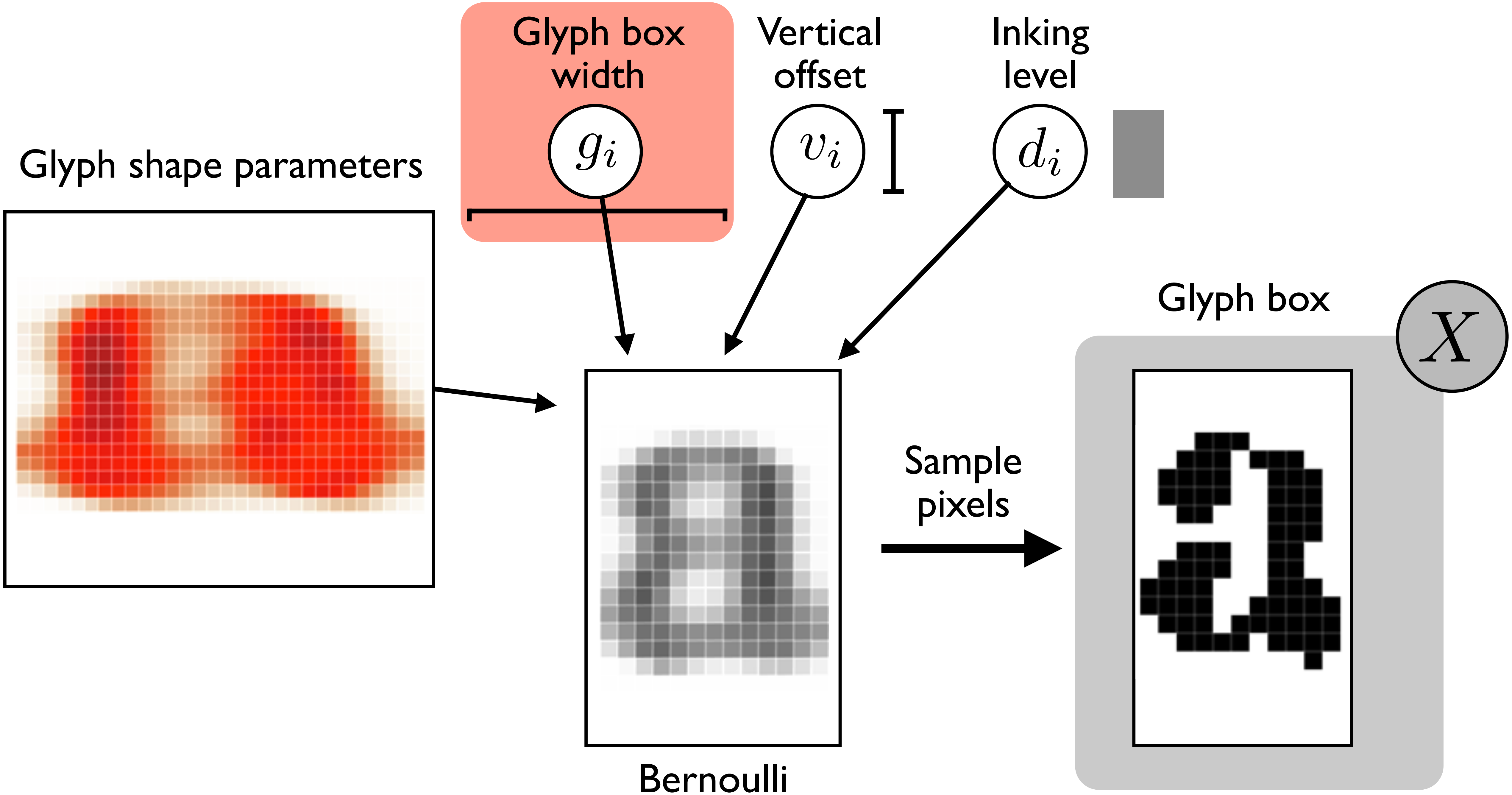
Rendering Model



Rendering Model

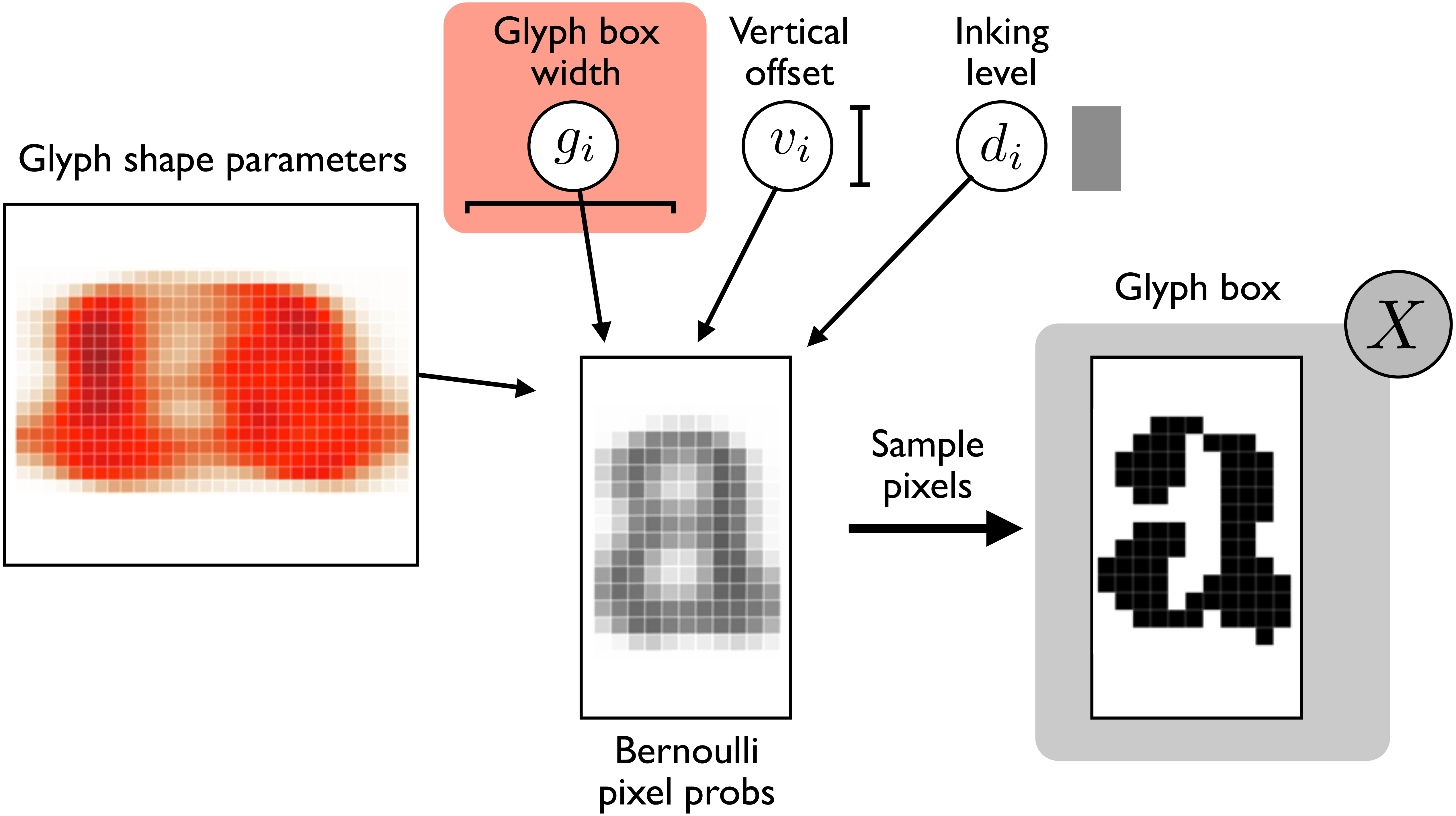


Rendering Model

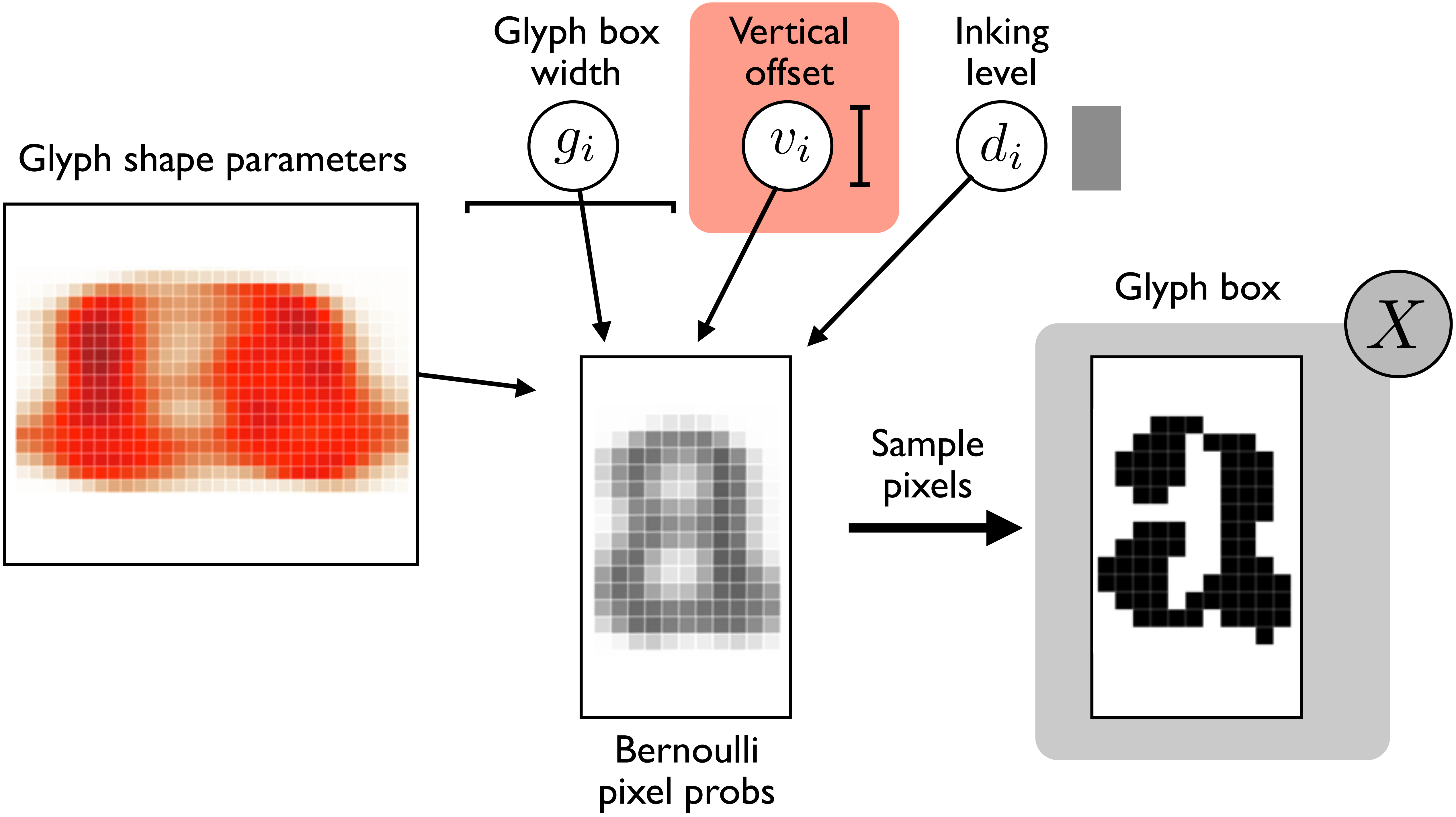


Bernoulli pixel probs

Rendering Model



Rendering Model



Glyph box width

Vertical offset

Inking level

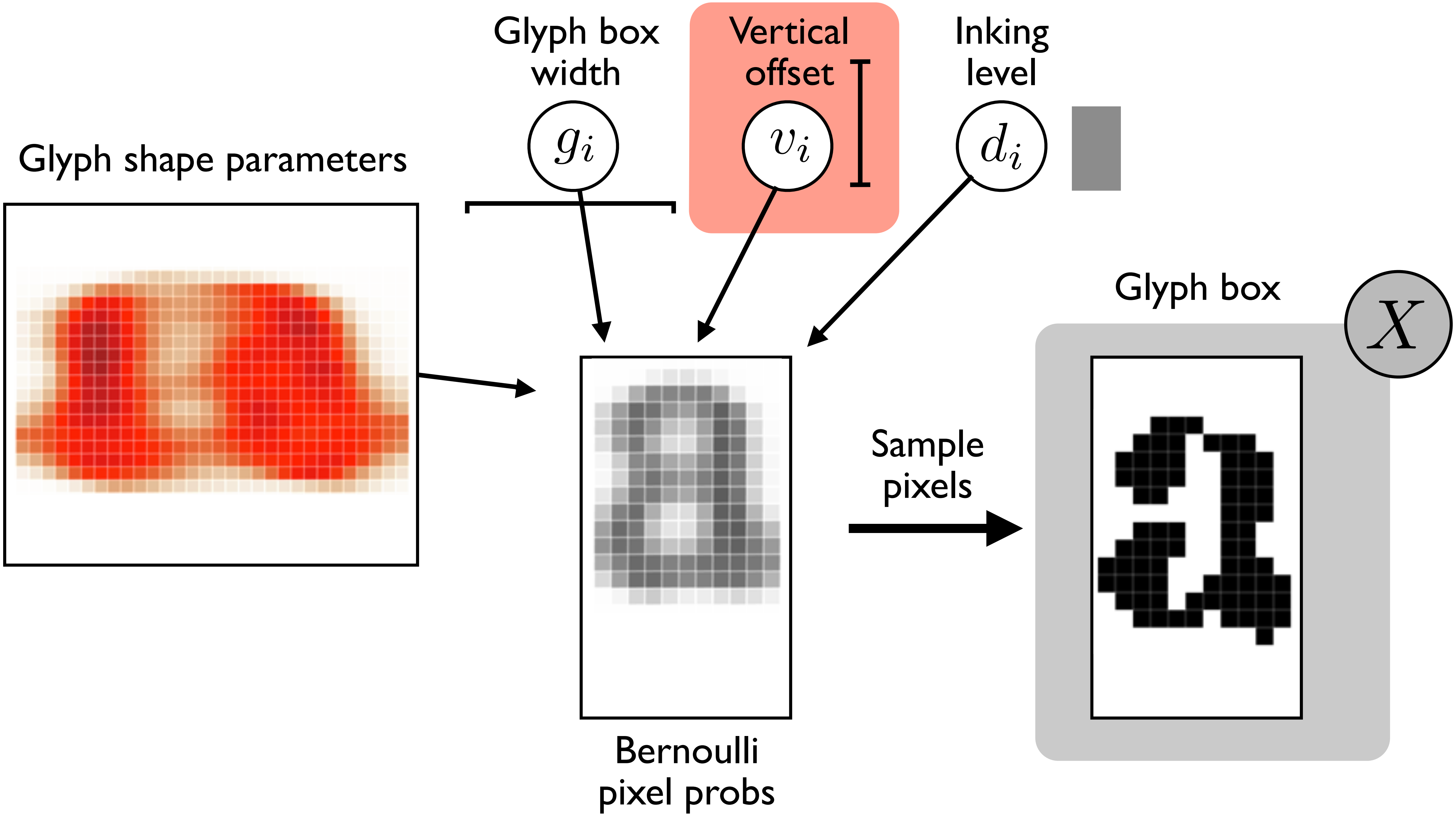
Glyph shape parameters

Glyph box

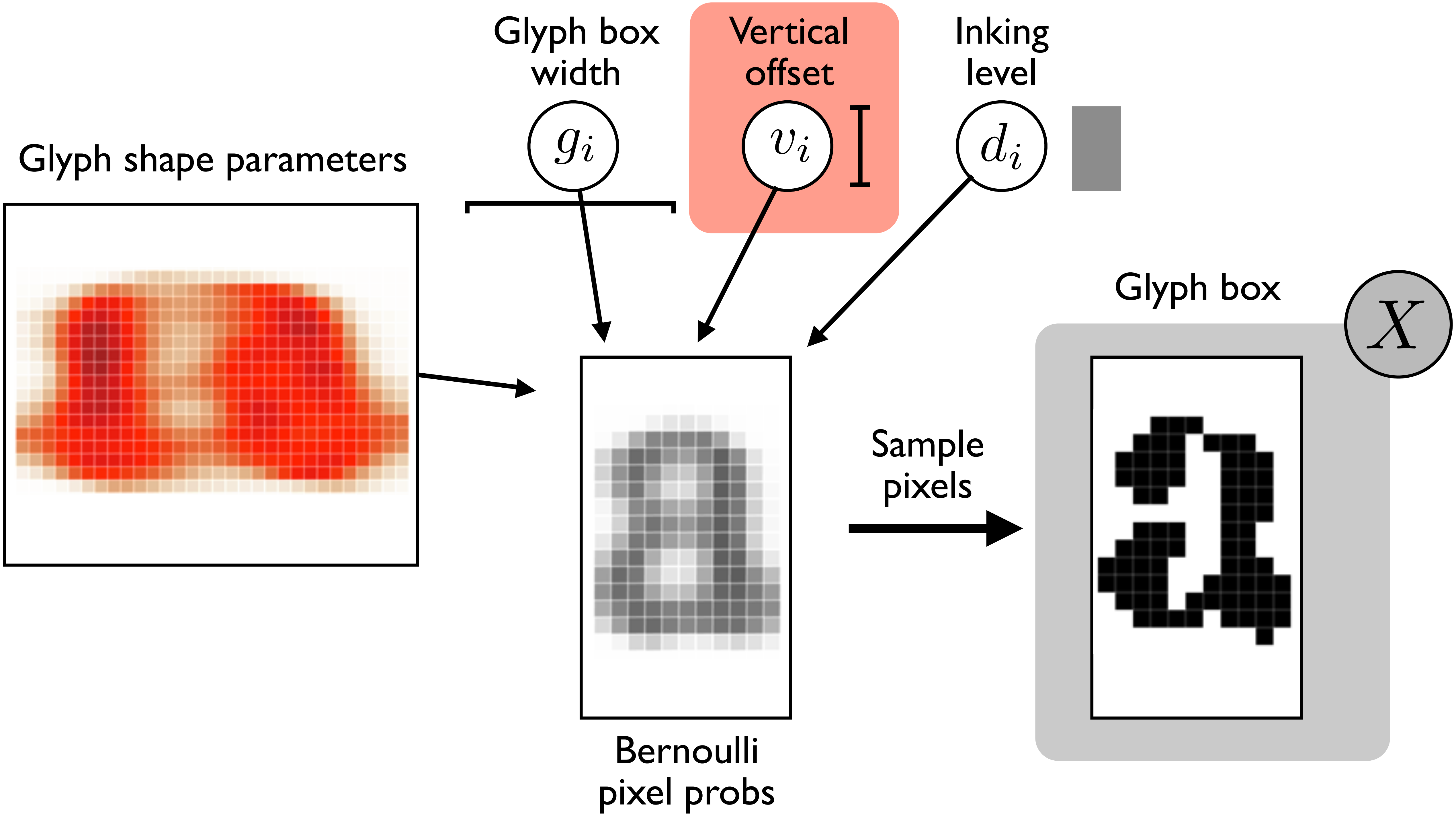
Sample pixels

Bernoulli pixel probs

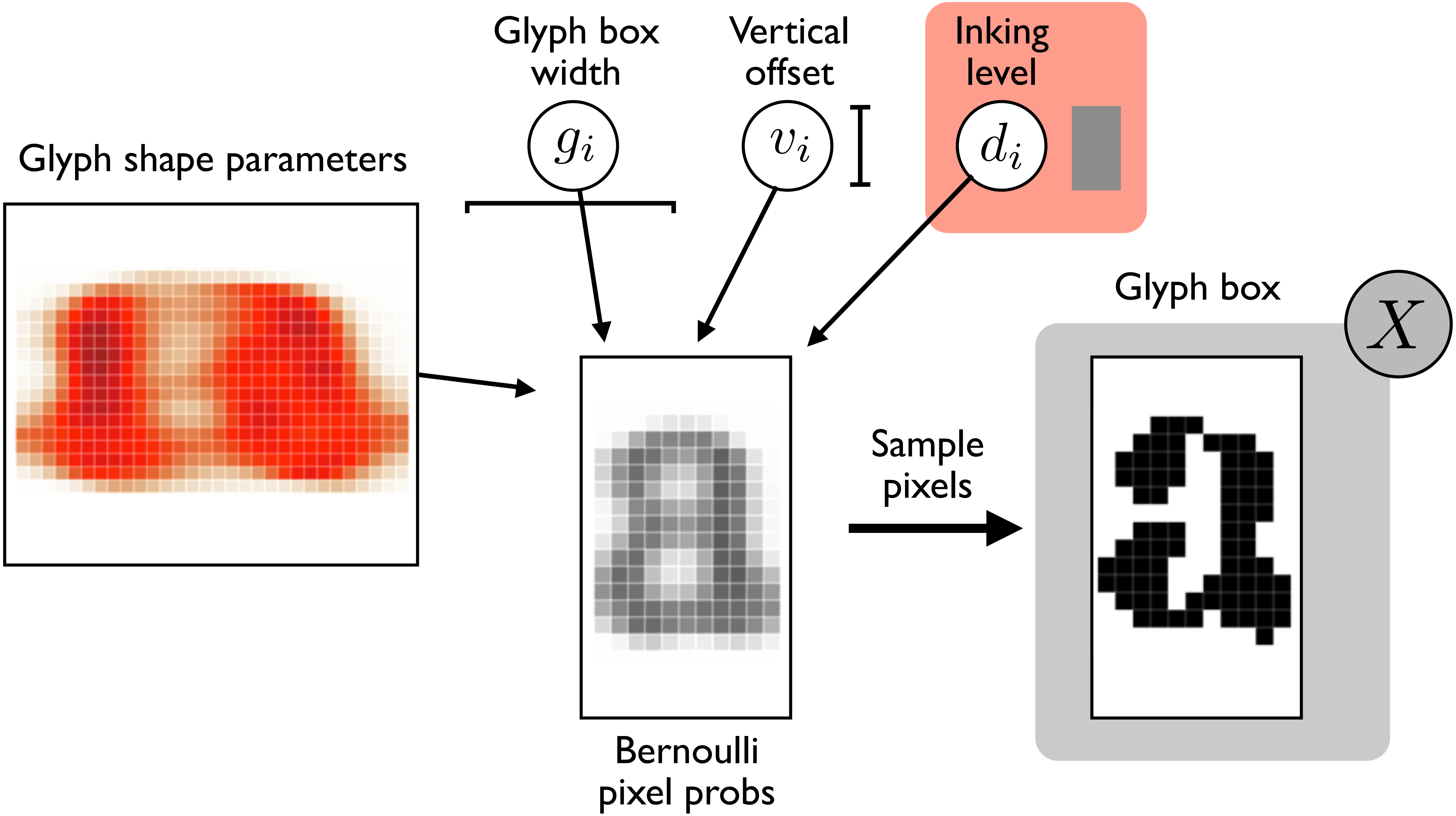
Rendering Model



Rendering Model



Rendering Model



Glyph shape parameters

Glyph box width

g_i

Vertical offset

v_i

Inking level

d_i

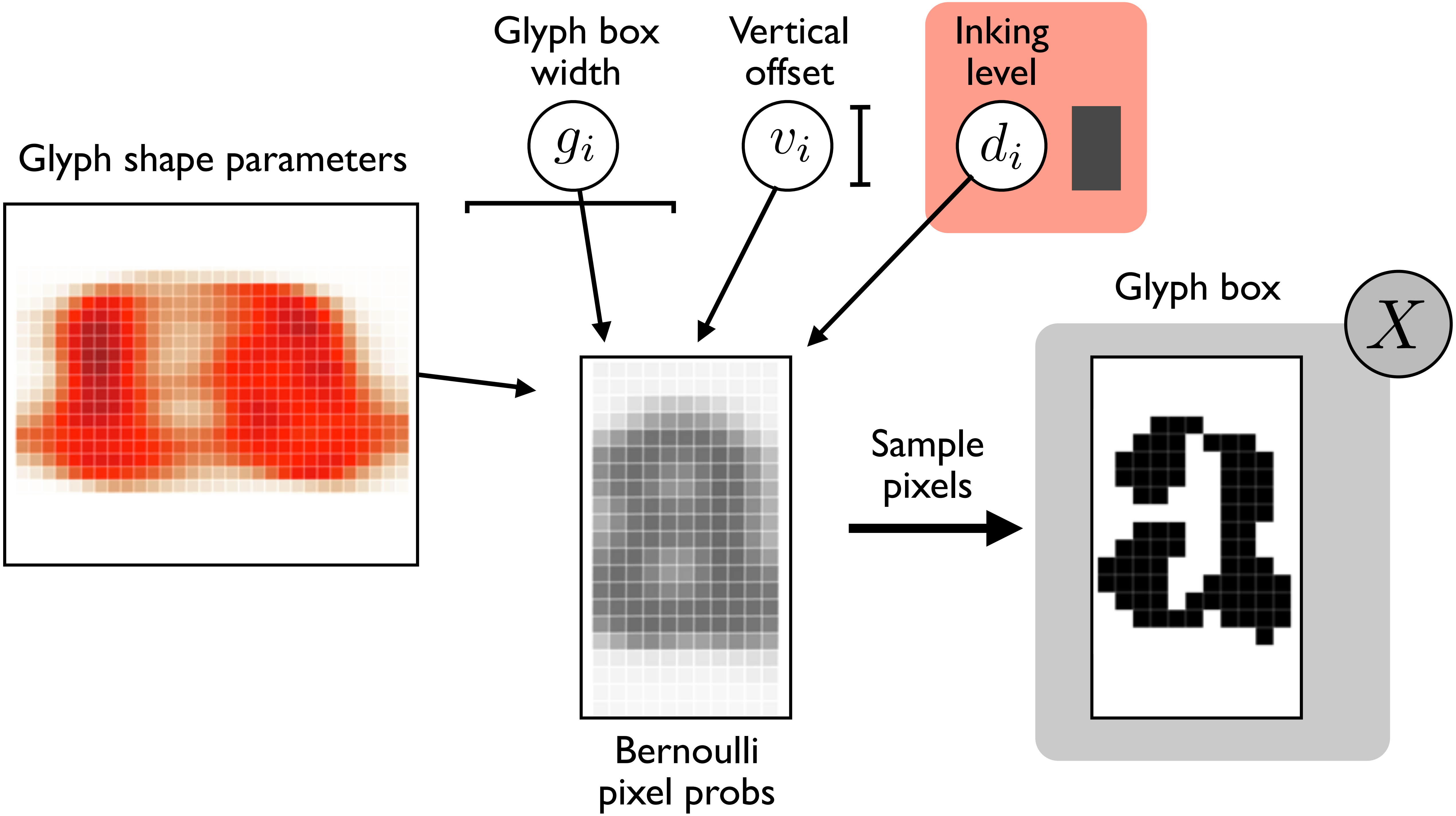
Glyph box

X

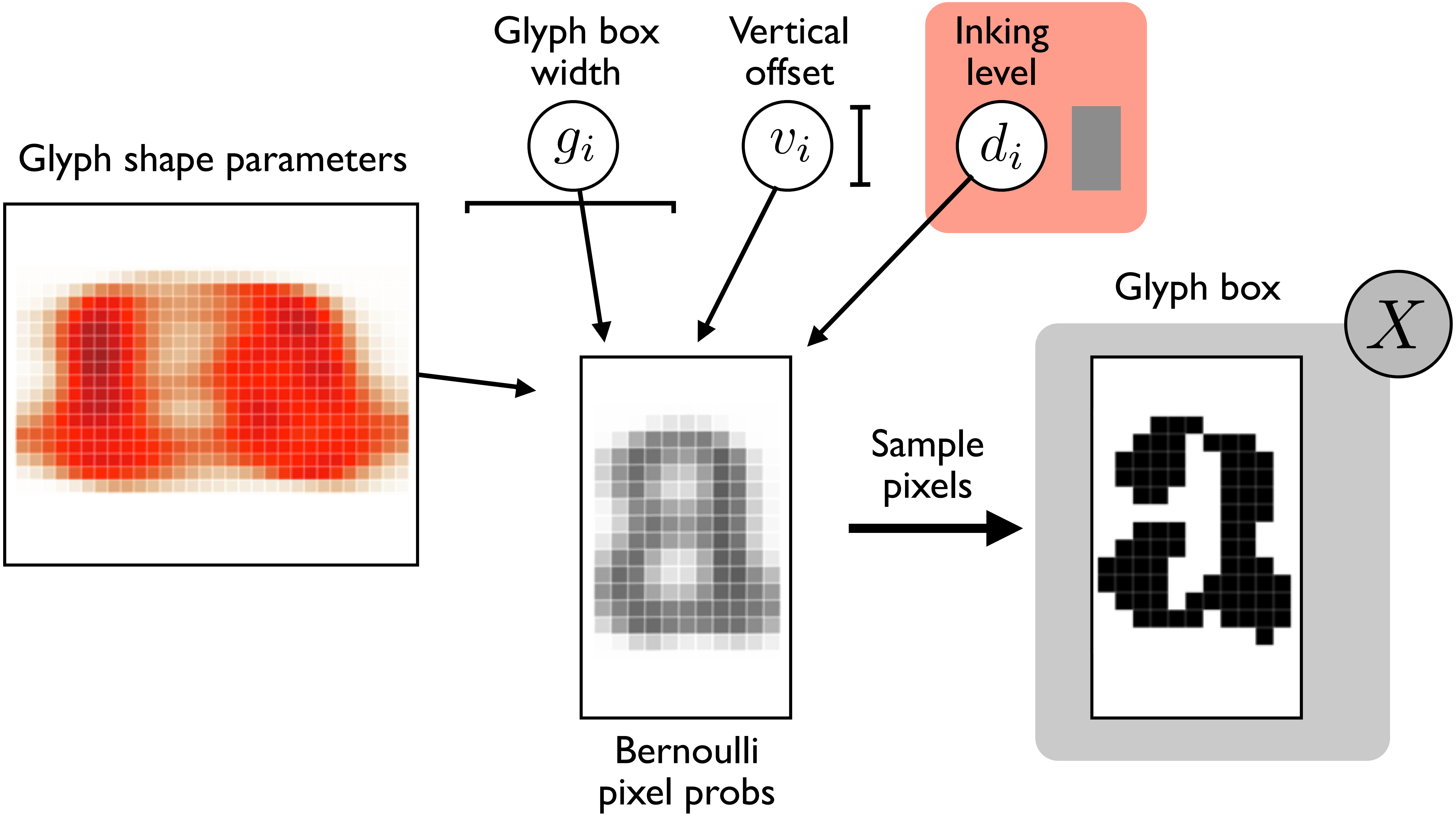
Sample pixels

Bernoulli pixel probs

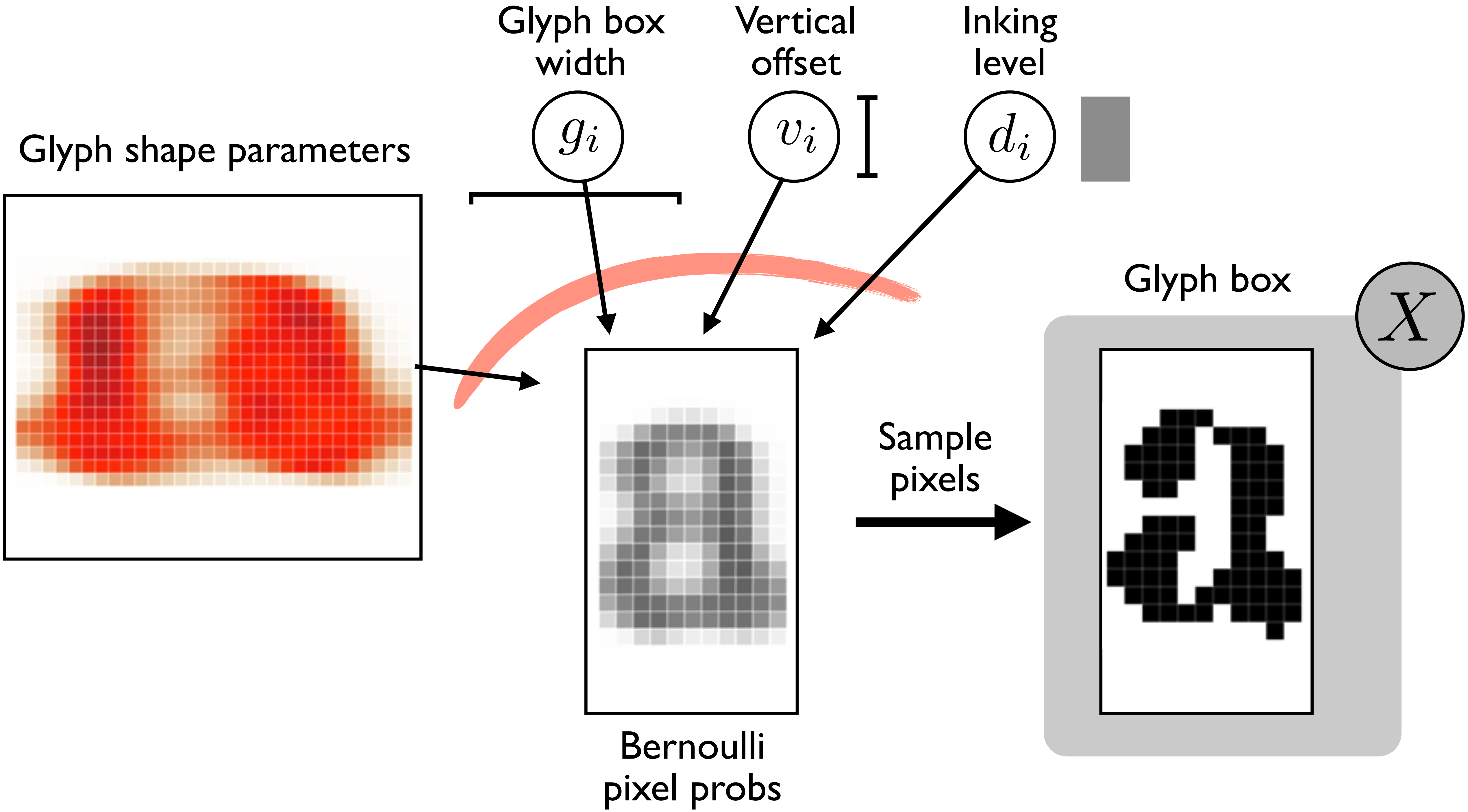
Rendering Model



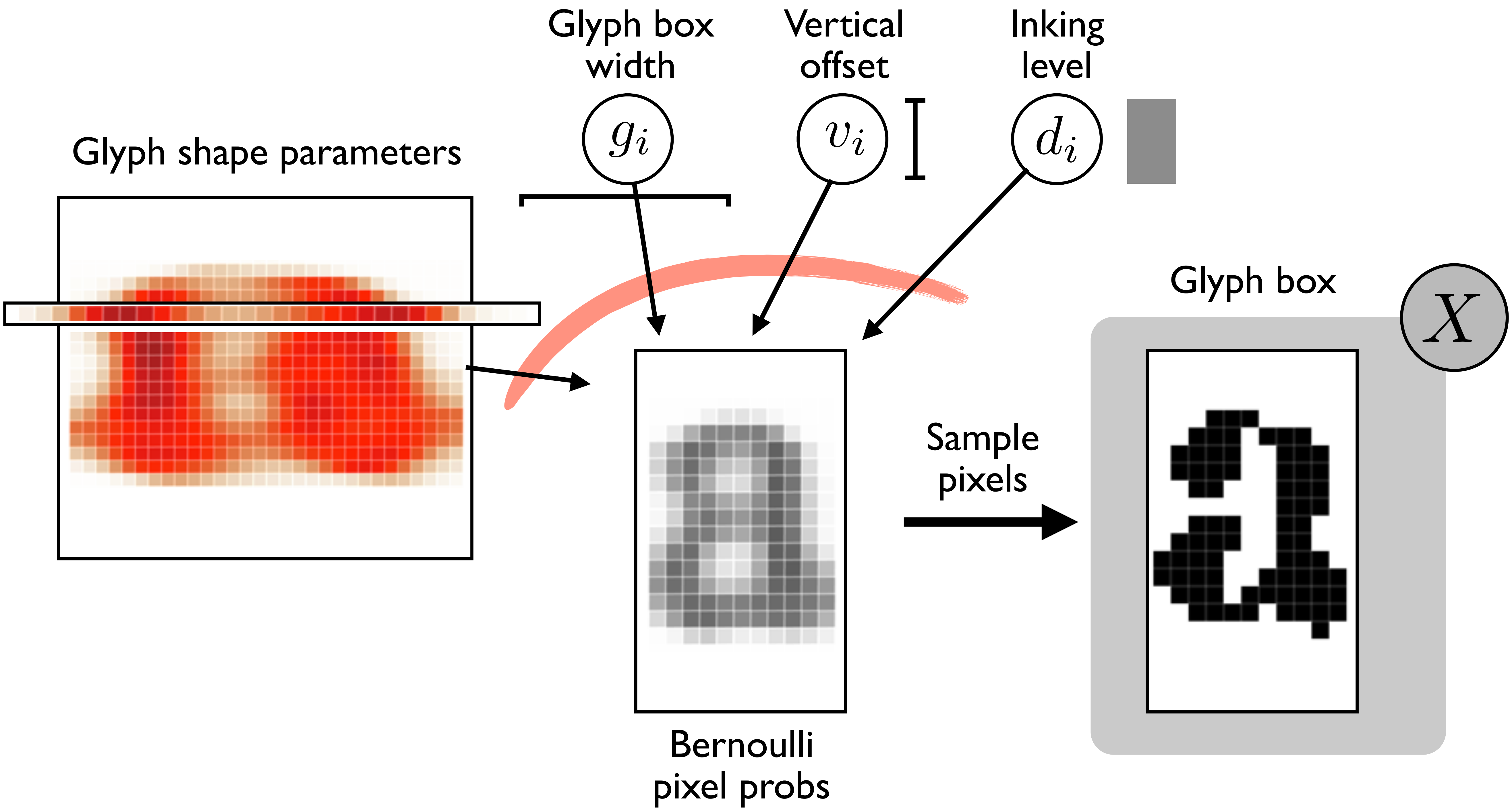
Rendering Model



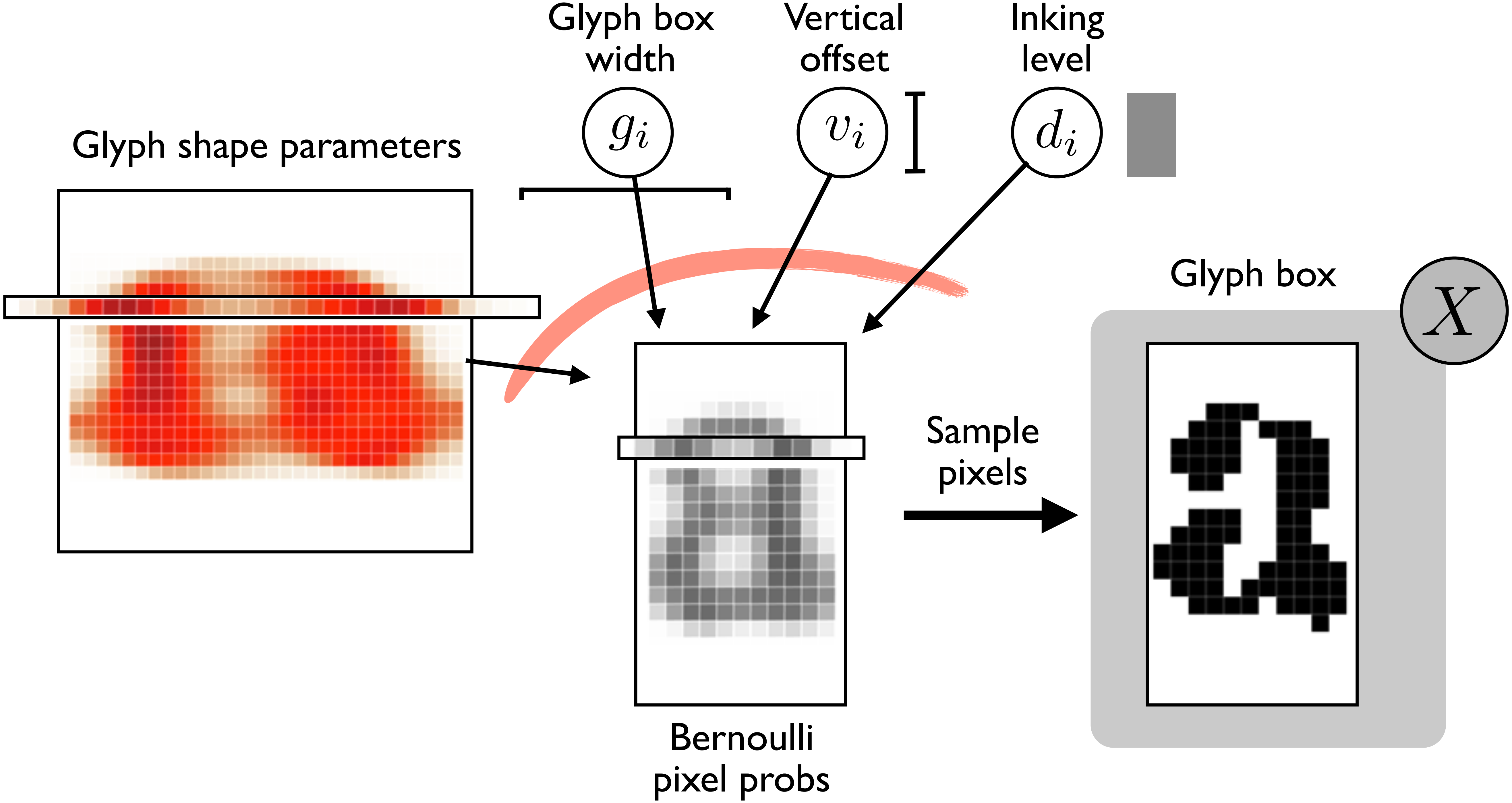
Rendering Model



Rendering Model



Rendering Model



Glyph box width

Vertical offset

Inking level

Glyph shape parameters

Glyph box

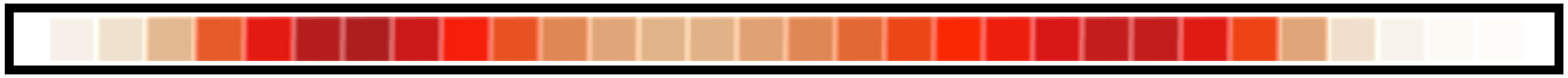
Sample pixels

Bernoulli pixel probs

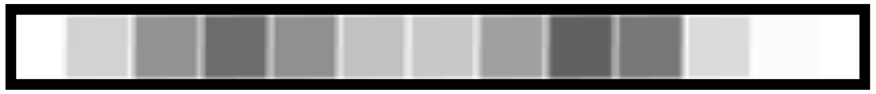
Log-linear Interpolation

Log-linear Interpolation

Glyph shape parameters ϕ

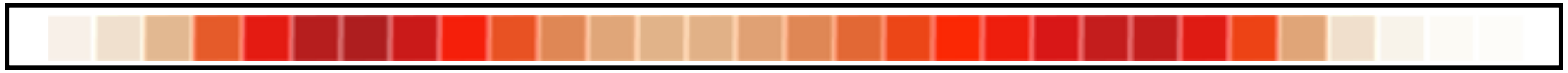


Bernoulli pixel probs θ

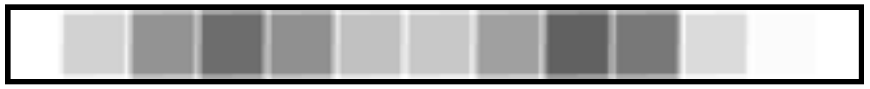


Log-linear Interpolation

Glyph shape parameters ϕ



Bernoulli pixel probs θ



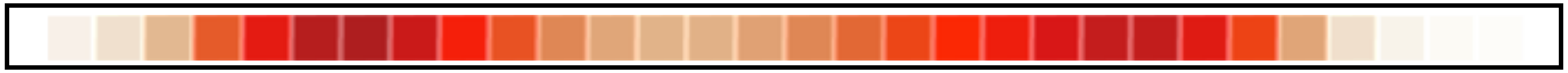
Log-linear Interpolation

Glyph shape parameters ϕ

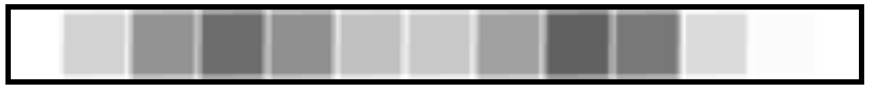
Bernoulli pixel probs θ

Log-linear Interpolation

Glyph shape parameters ϕ



Bernoulli pixel probs θ

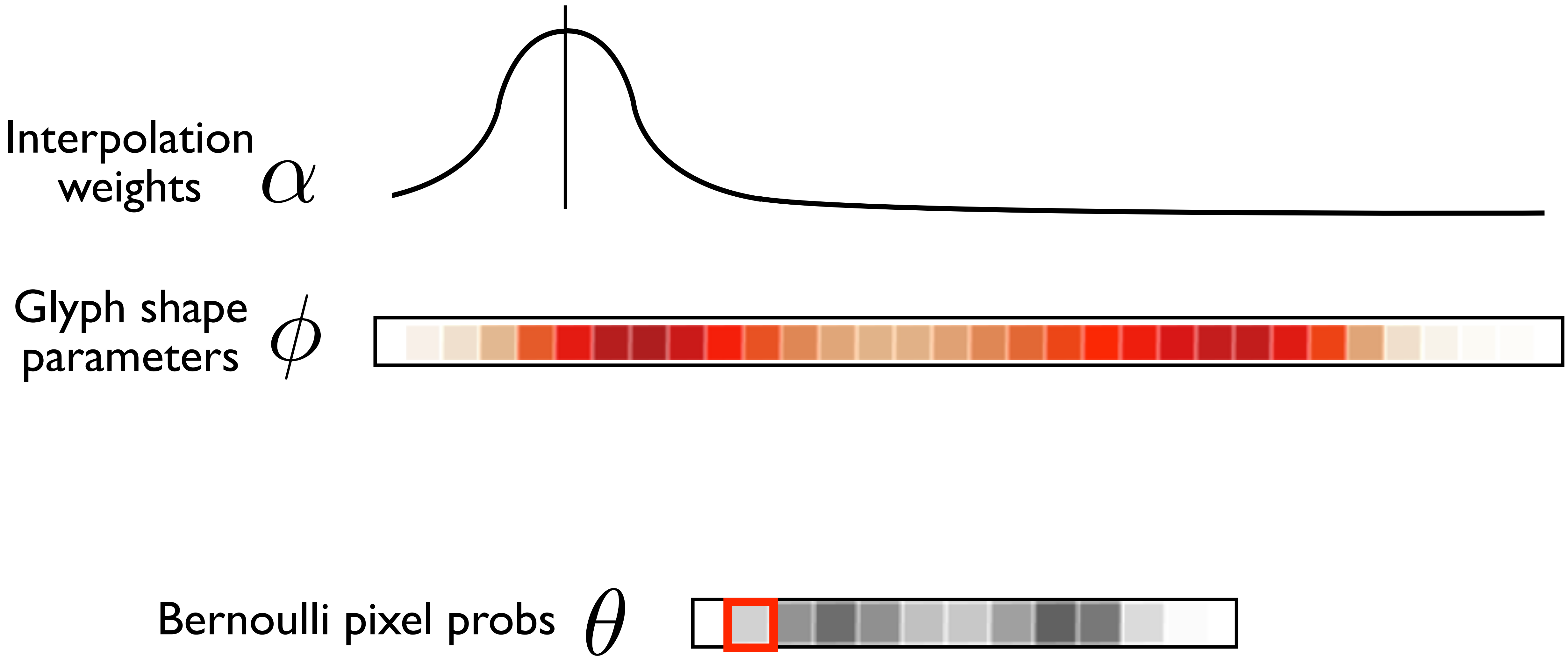


Log-linear Interpolation

Glyph shape parameters ϕ

Bernoulli pixel probs θ

Log-linear Interpolation

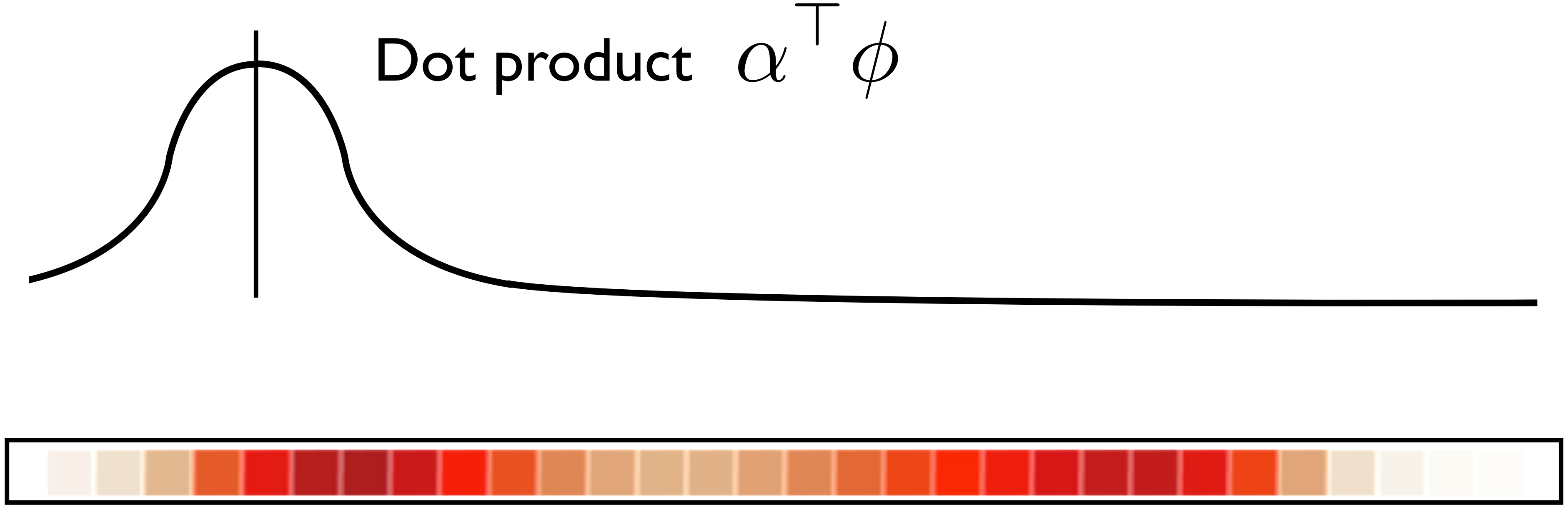


Interpolation weights α

Glyph shape parameters ϕ

Bernoulli pixel probs θ

Log-linear Interpolation



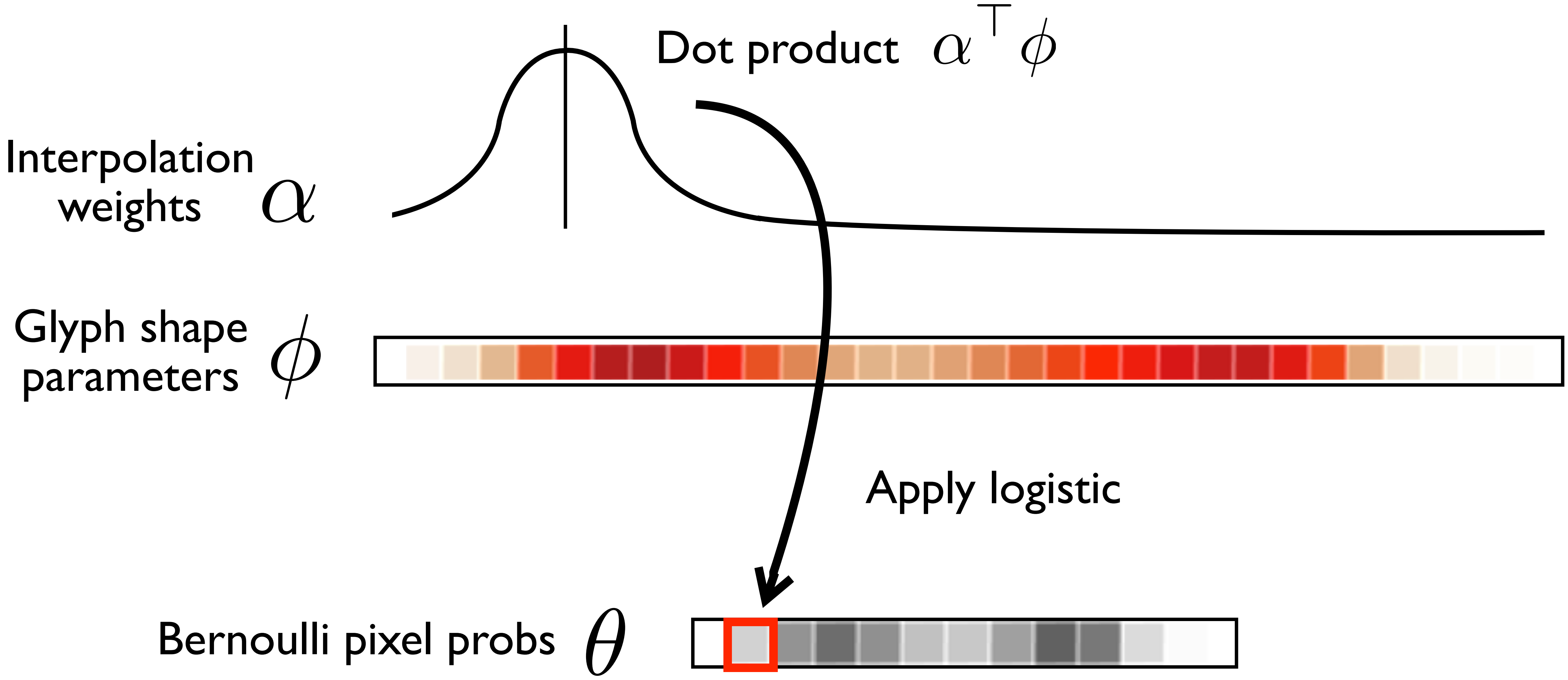
Dot product $\alpha^\top \phi$

Interpolation weights α

Glyph shape parameters ϕ

Bernoulli pixel probs θ

Log-linear Interpolation



Dot product $\alpha^\top \phi$

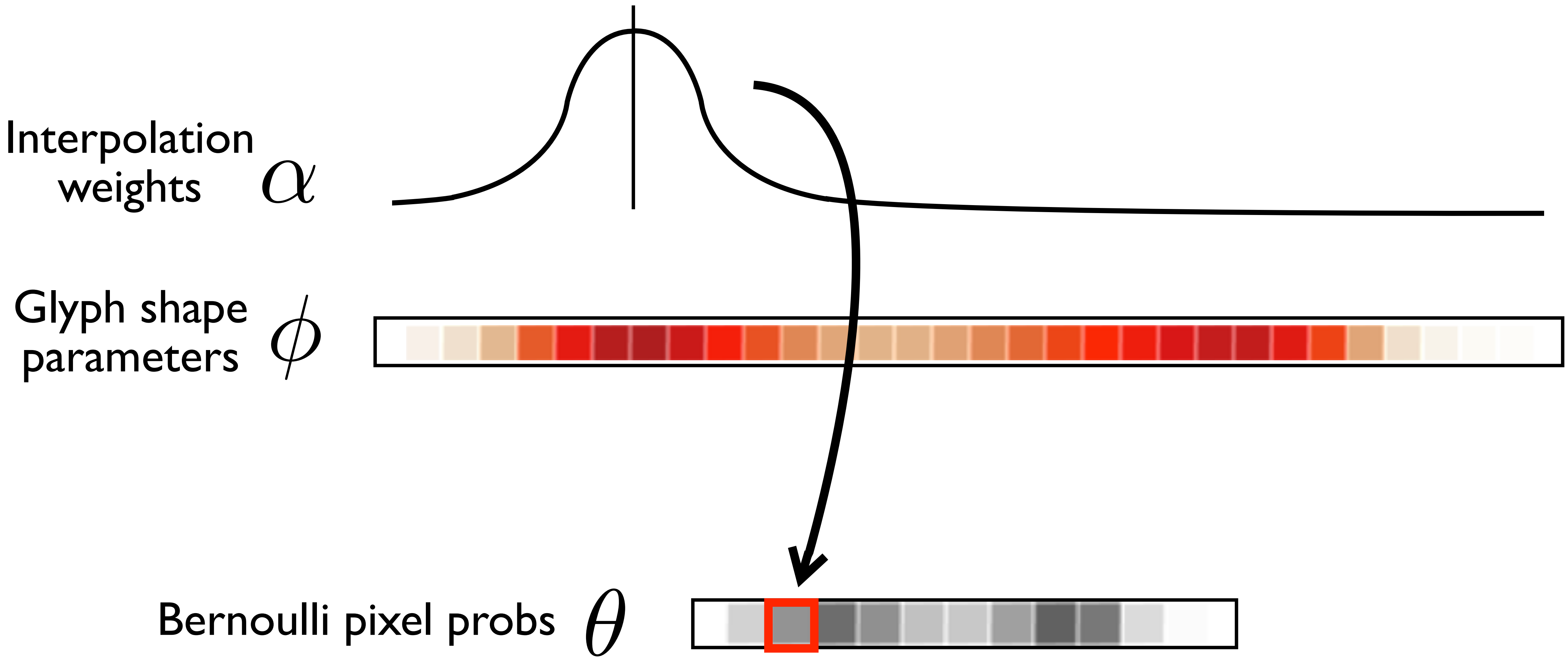
Interpolation weights α

Glyph shape parameters ϕ

Apply logistic

Bernoulli pixel probs θ

Log-linear Interpolation

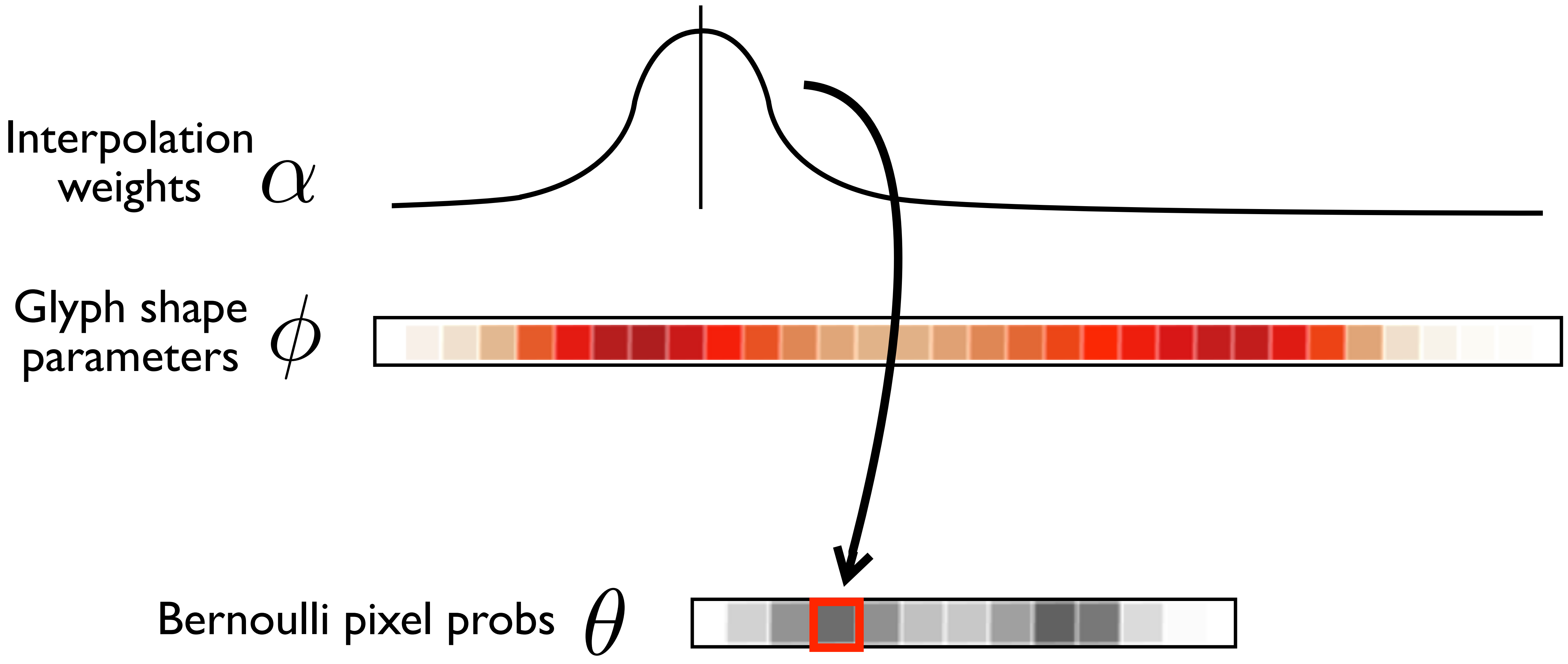


Interpolation weights α

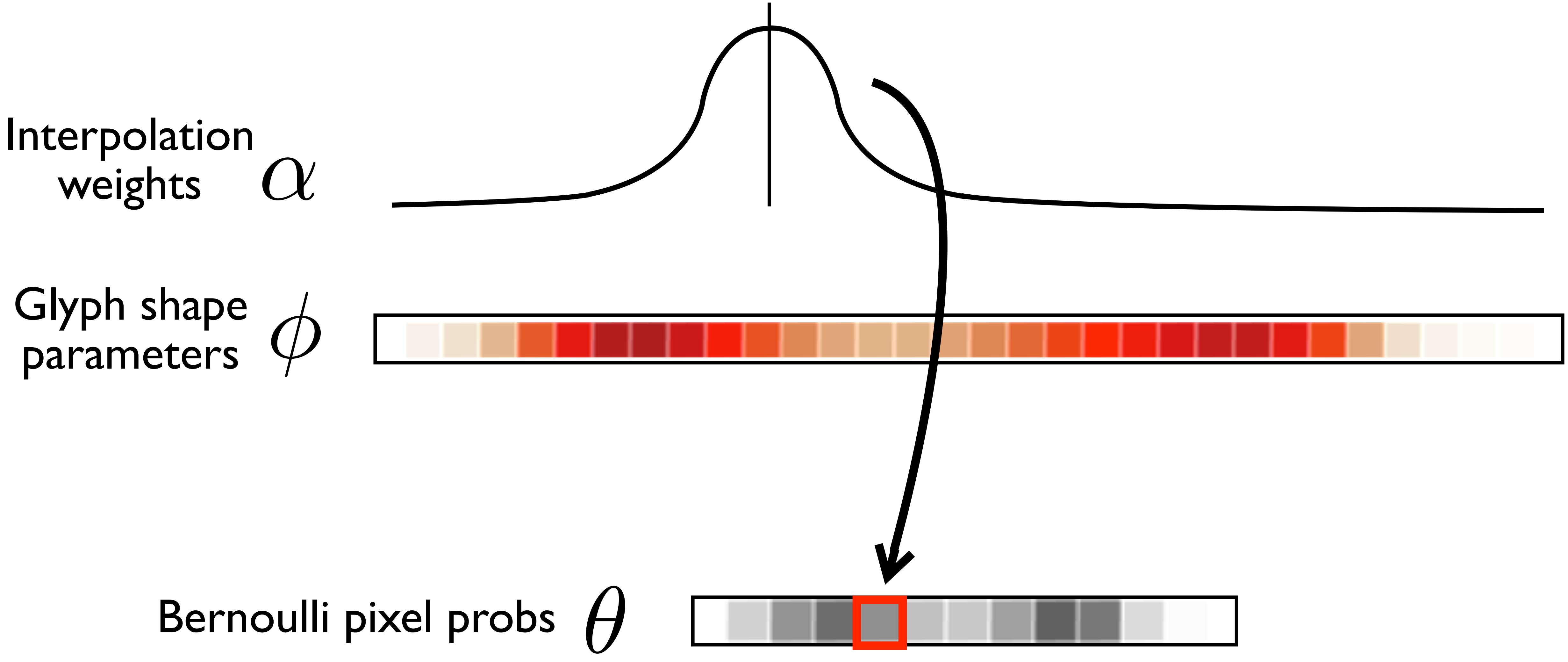
Glyph shape parameters ϕ

Bernoulli pixel probs θ

Log-linear Interpolation


Interpolation weights α
Glyph shape parameters ϕ
Bernoulli pixel probs θ

Log-linear Interpolation

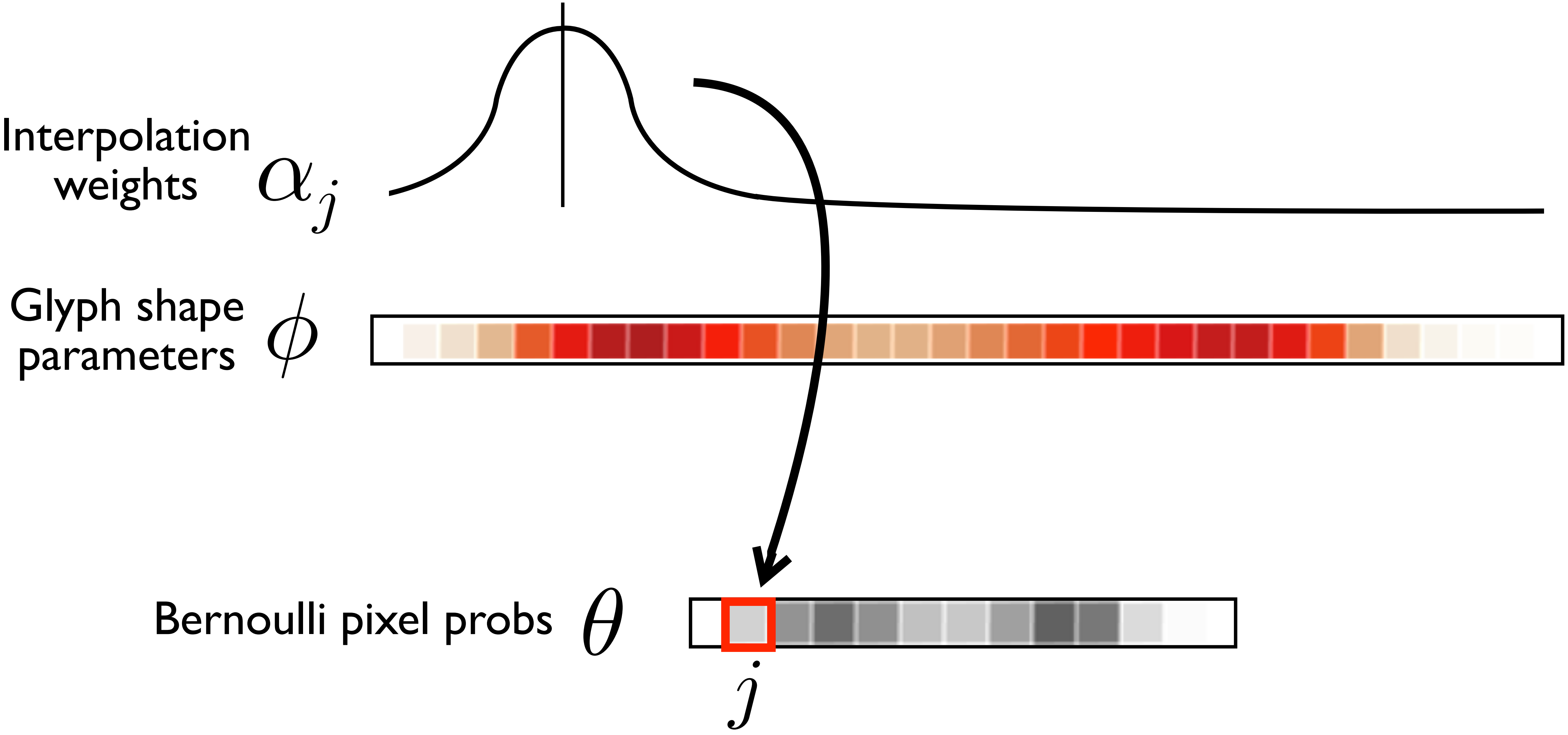


Interpolation weights α

Glyph shape parameters ϕ

Bernoulli pixel probs θ

Log-linear Interpolation



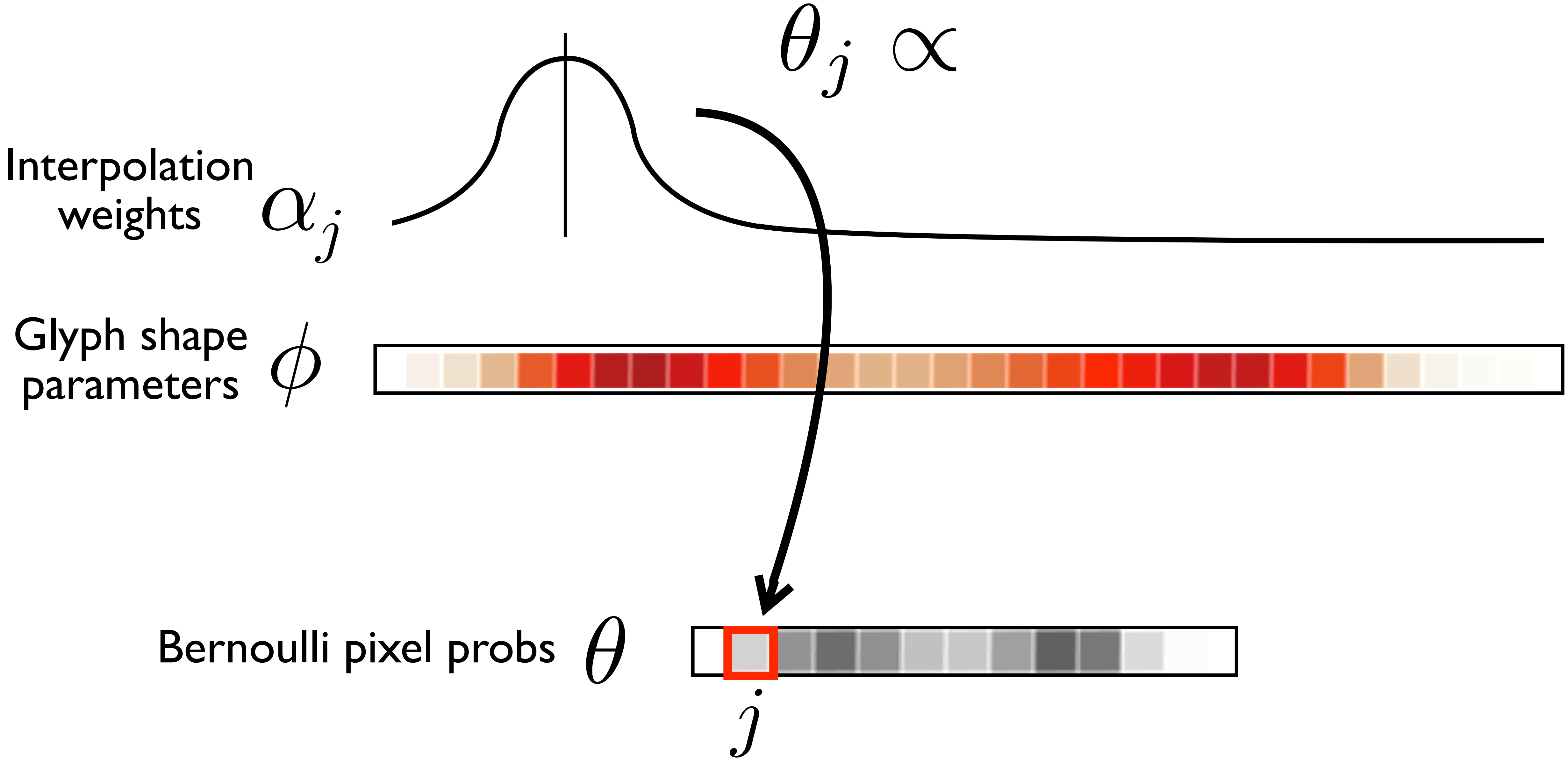
Interpolation weights α_j

Glyph shape parameters ϕ

Bernoulli pixel probs θ

j

Log-linear Interpolation



$\theta_j \propto$

Interpolation weights α_j

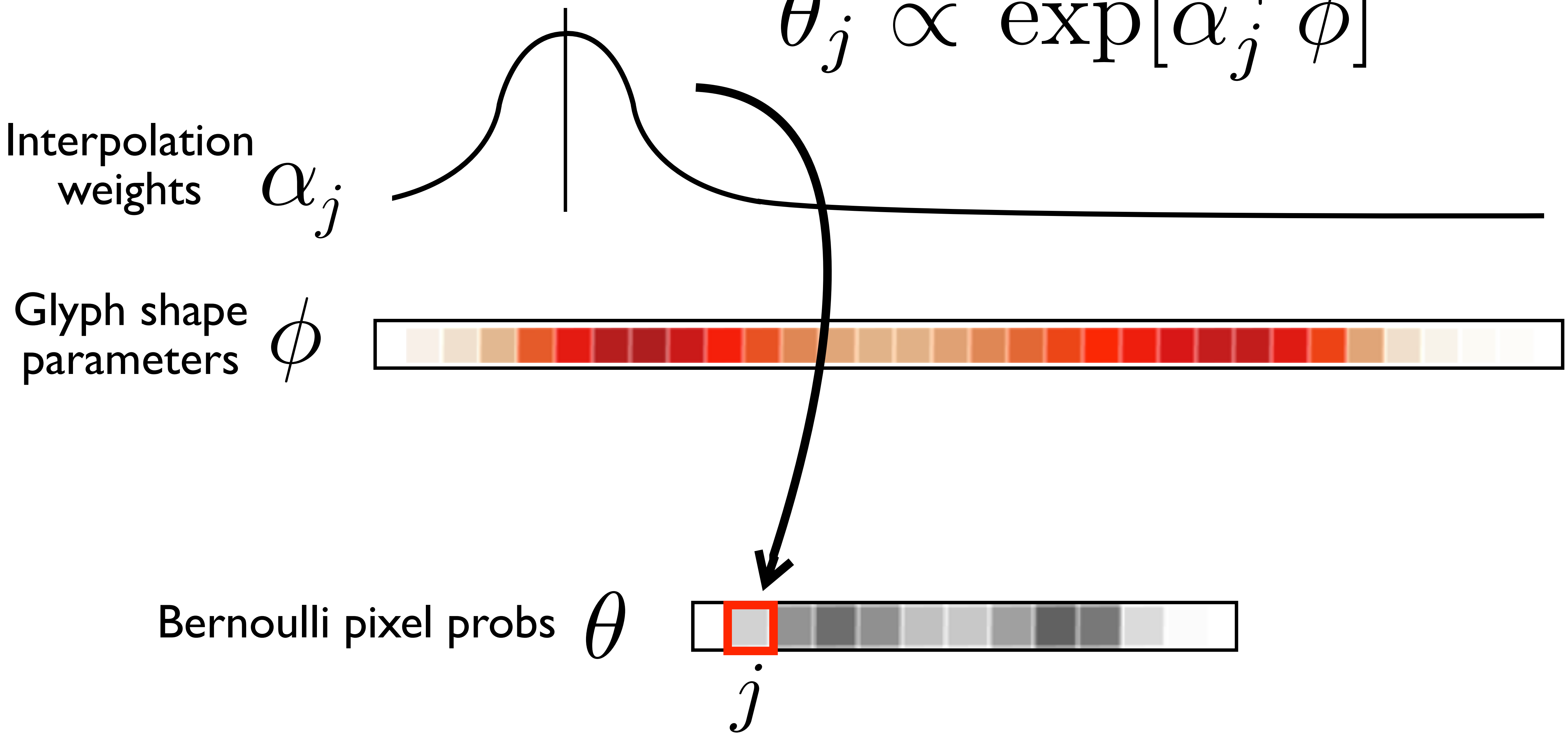
Glyph shape parameters ϕ

Bernoulli pixel probs θ

j

Log-linear Interpolation

$$\theta_j \propto \exp[\alpha_j^\top \phi]$$



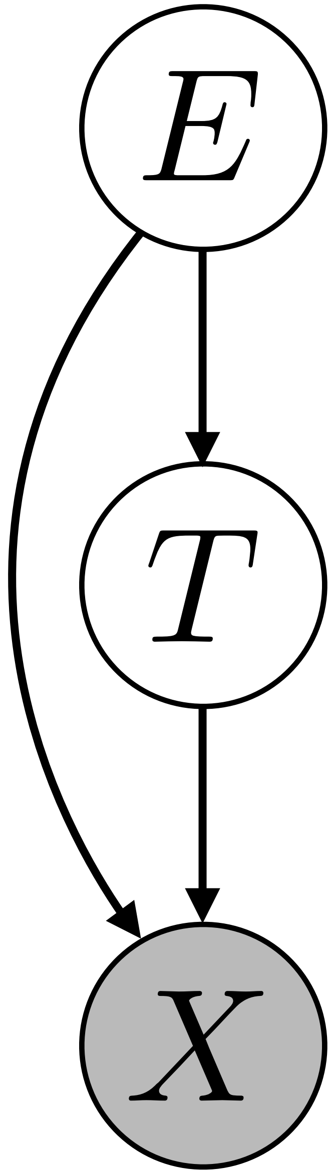
Interpolation weights α_j

Glyph shape parameters ϕ

Bernoulli pixel probs θ

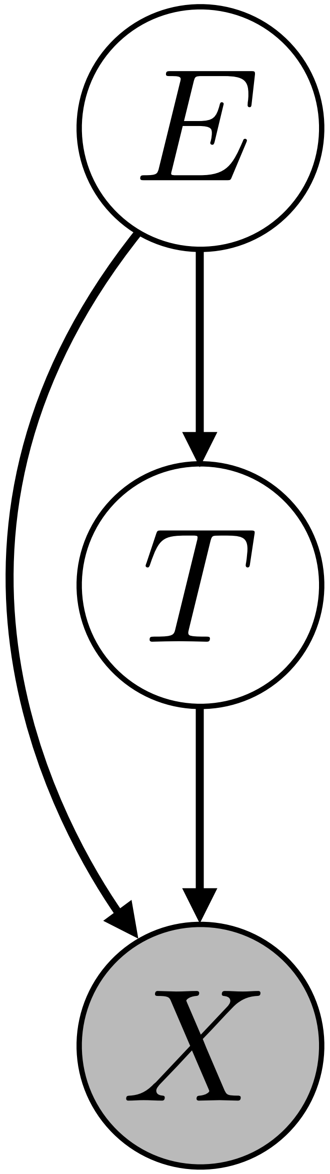
j

Learning and Inference



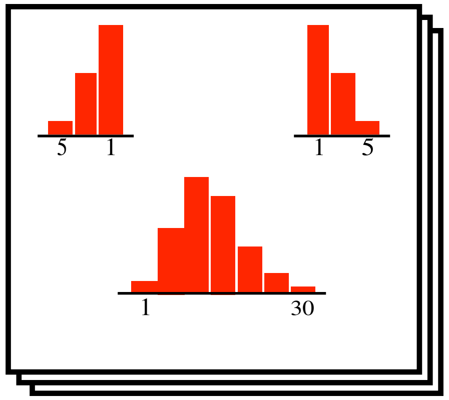
Learning and Inference

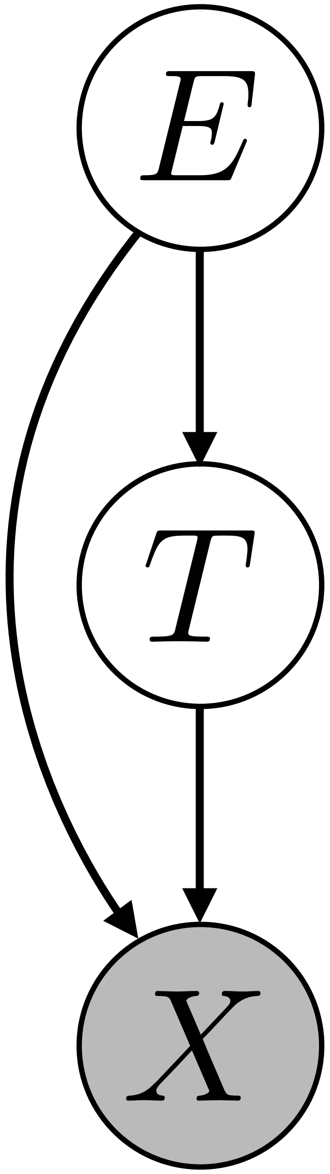
- Learn font parameters using EM



Learning and Inference

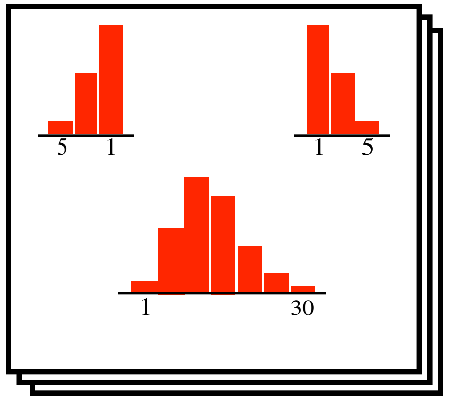
- Learn font parameters using EM

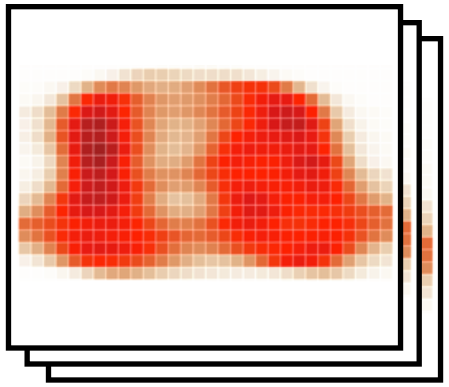


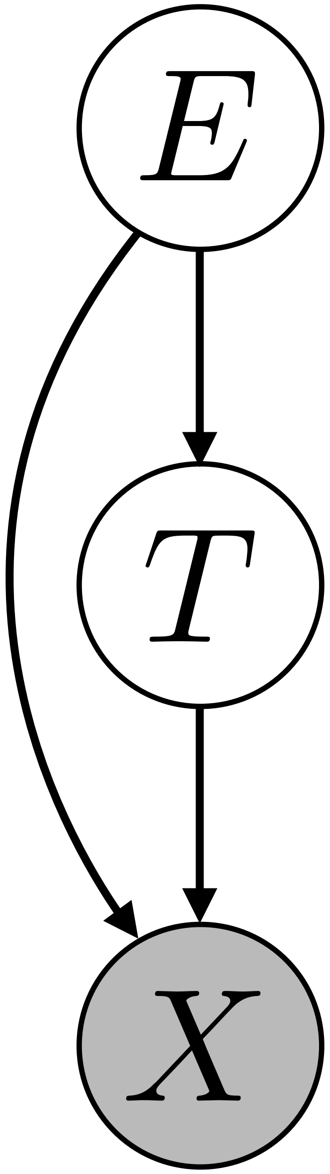


Learning and Inference

- Learn font parameters using EM

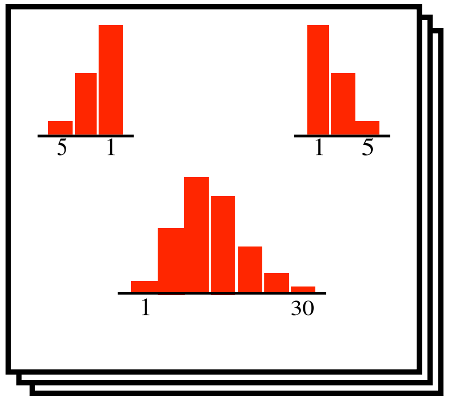


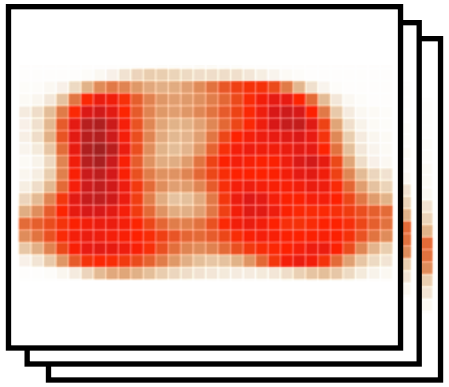




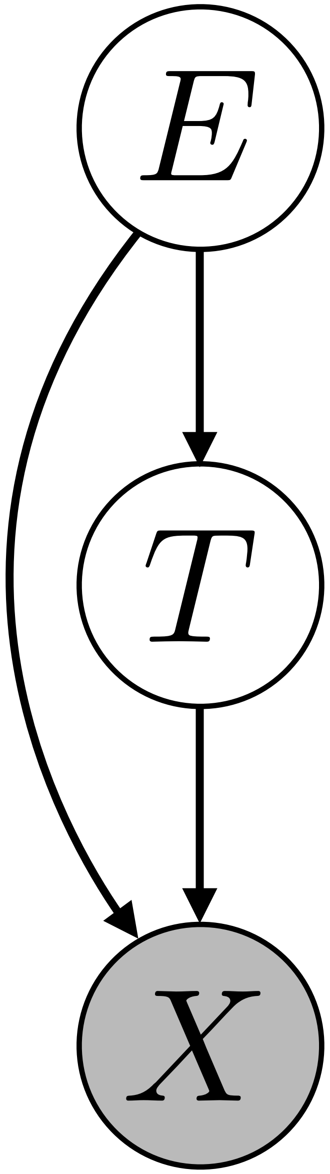
Learning and Inference

- Learn font parameters using EM



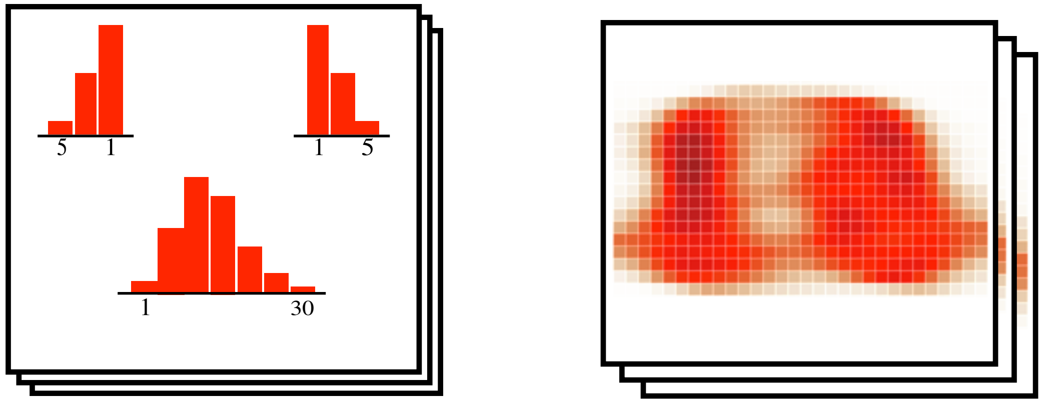


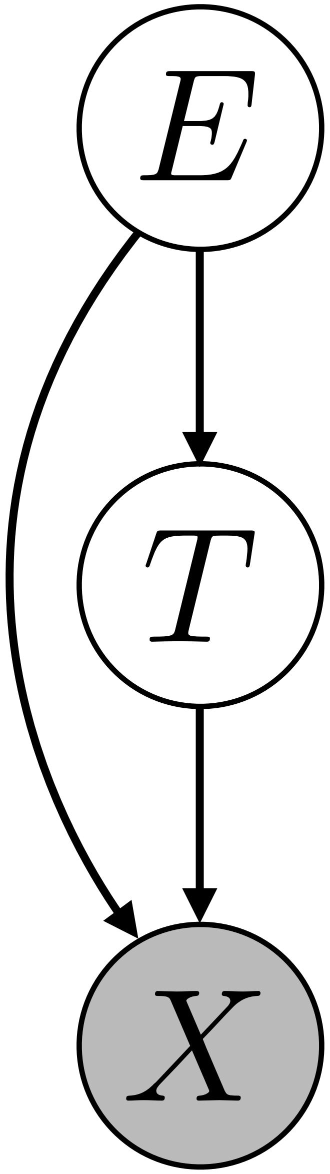
- Initialize font parameters with mixtures of modern fonts



Learning and Inference

- Learn font parameters using EM



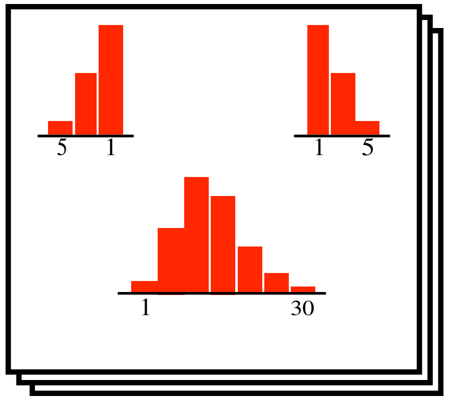


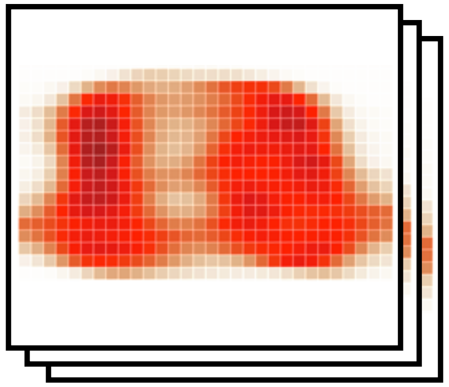
- Initialize font parameters with mixtures of modern fonts

- Semi-Markov DP to compute expectations

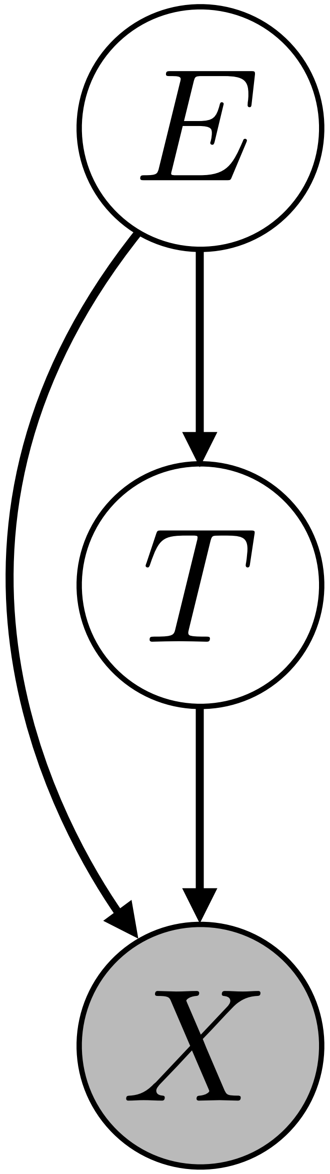
Learning and Inference

- Learn font parameters using EM

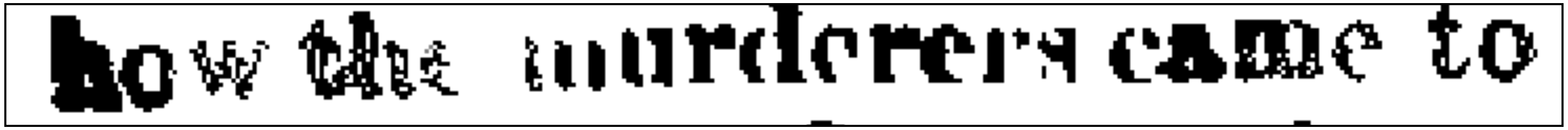




- Initialize font parameters with mixtures of modern fonts
- Semi-Markov DP to compute expectations
- Efficient inference using a coarse-to-fine approach

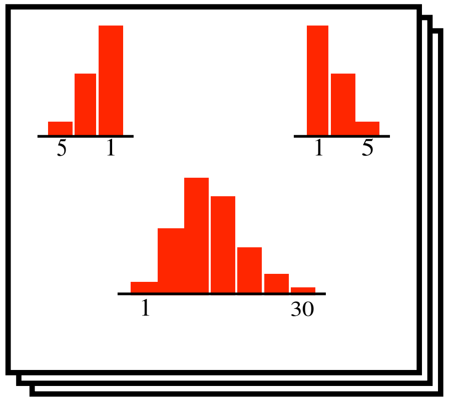


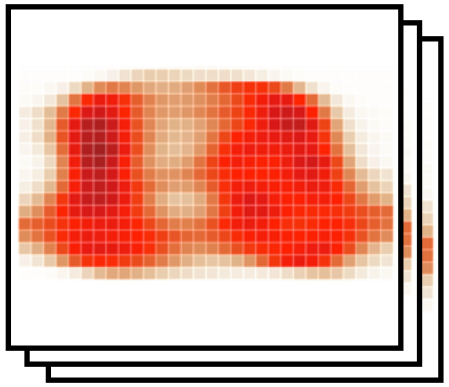
System Output Example

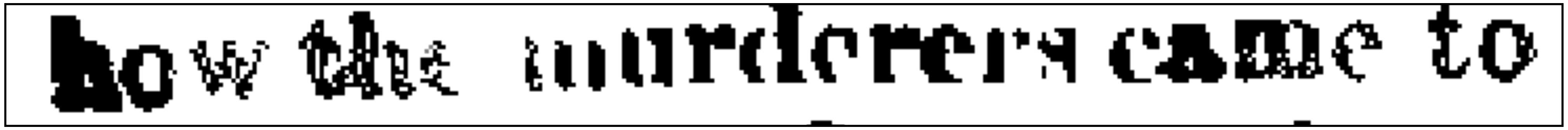


how the murderers came to

System Output Example

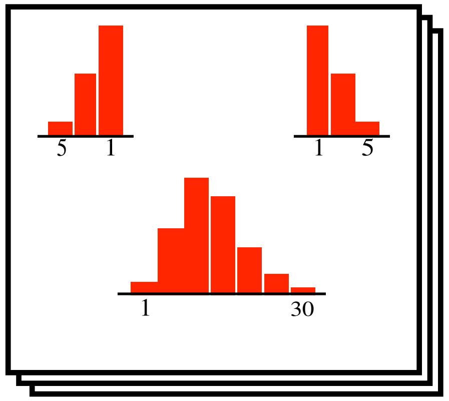


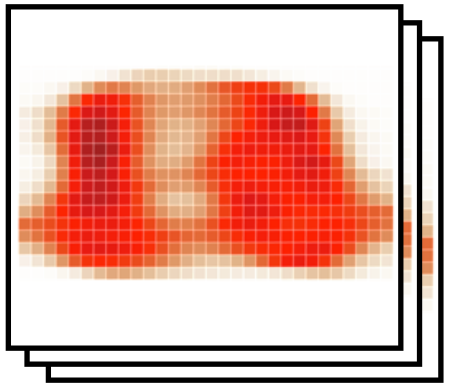




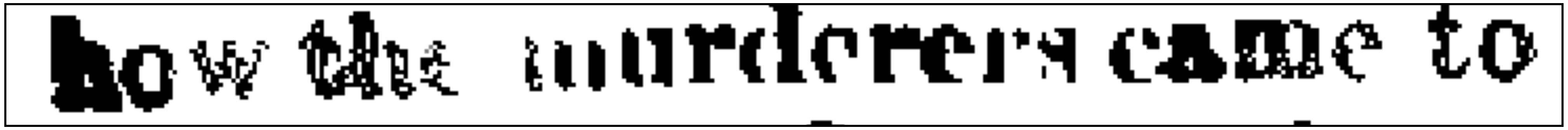
how the murderers came to

System Output Example

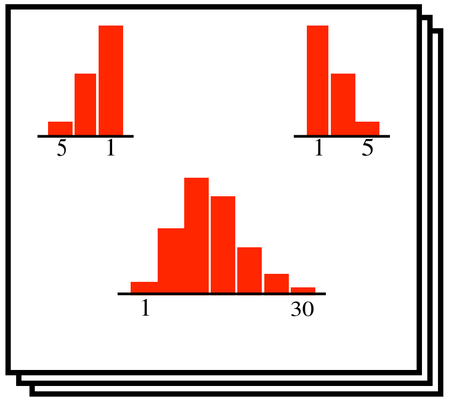


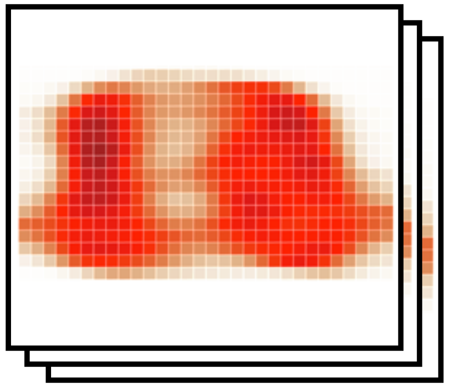


```
how the murderers came to
```

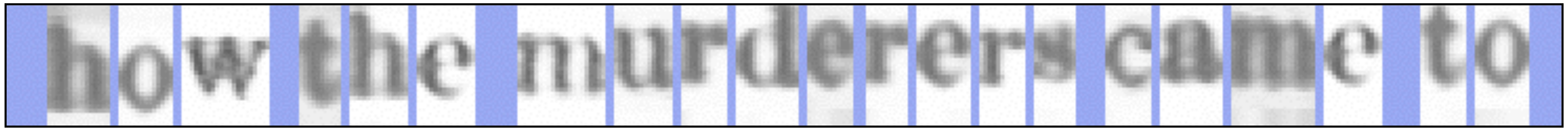


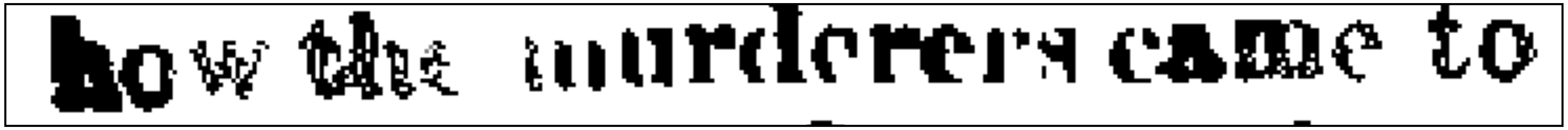
System Output Example



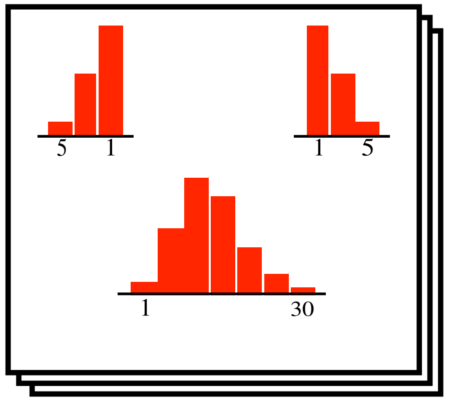


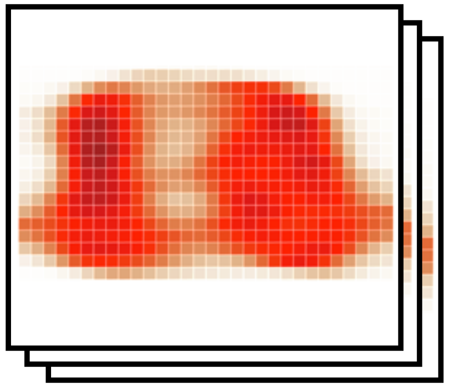
how the murderers came to



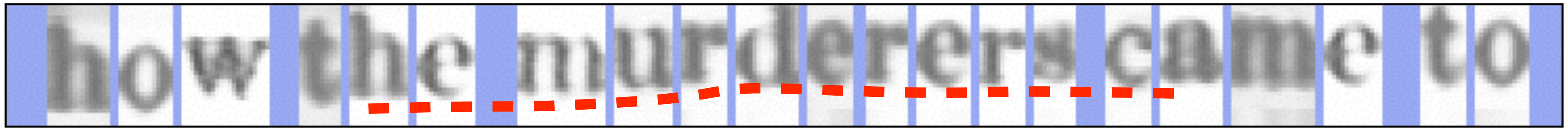


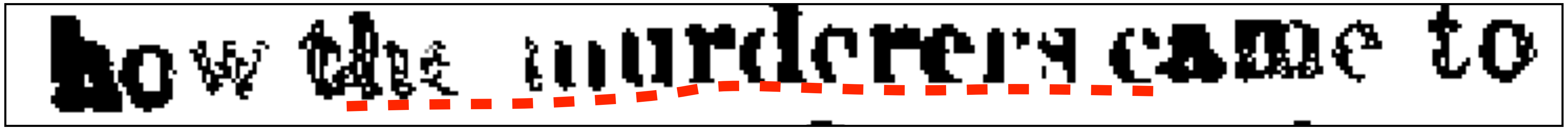
System Output Example



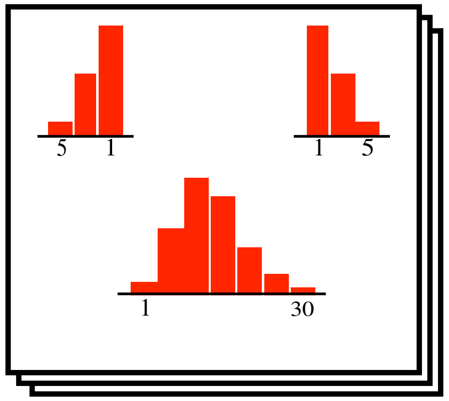


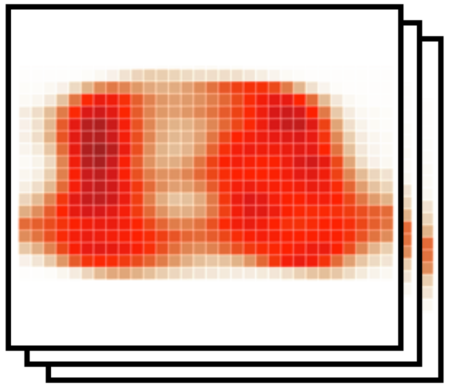
```
how the murderers came to
```



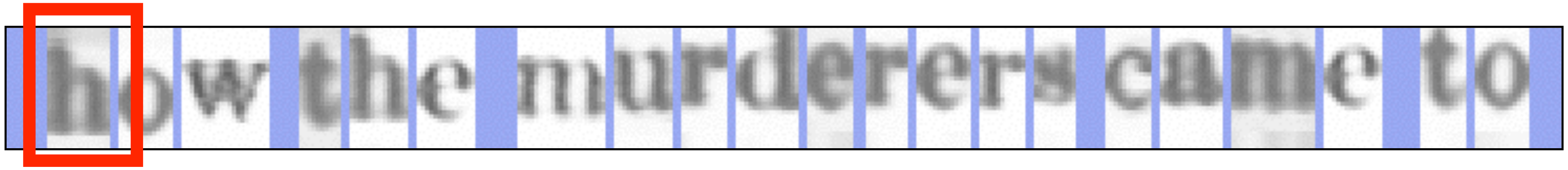


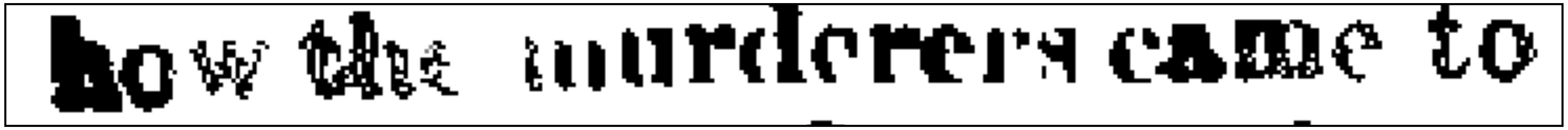
System Output Example



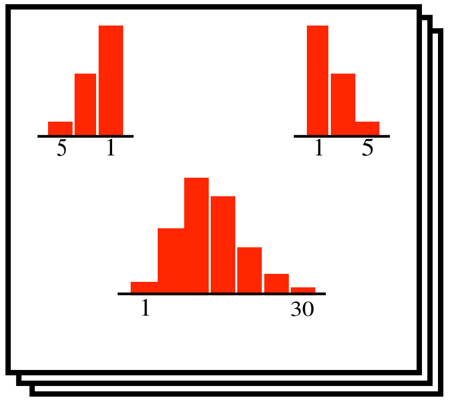


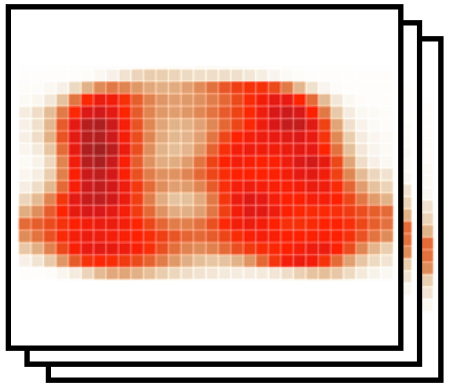
```
how the murderers came to
```


how the murderers came to

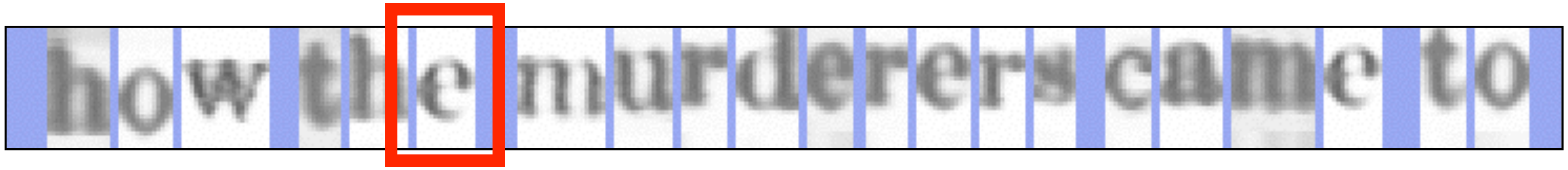

how the murderers came to

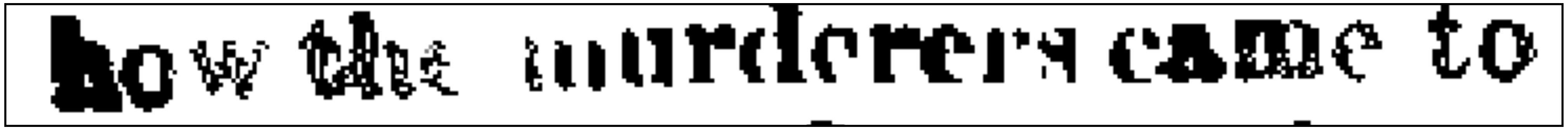
System Output Example



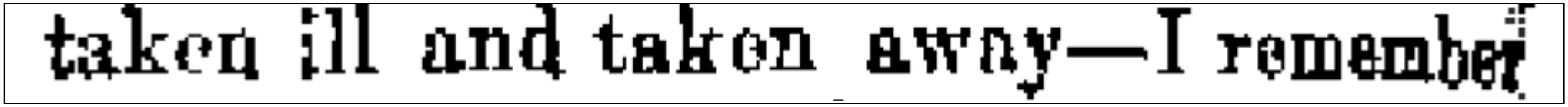


how the murderers came to



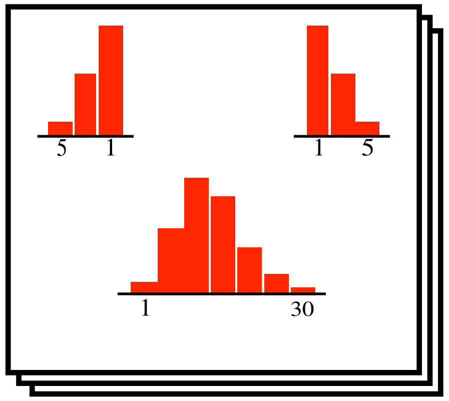


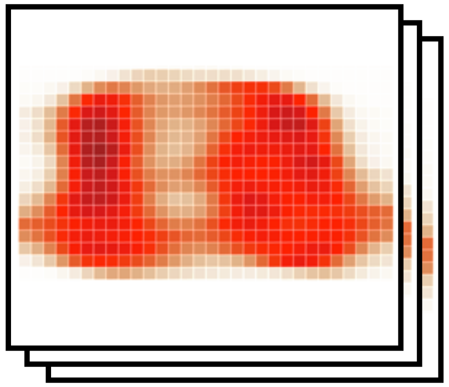
System Output Example

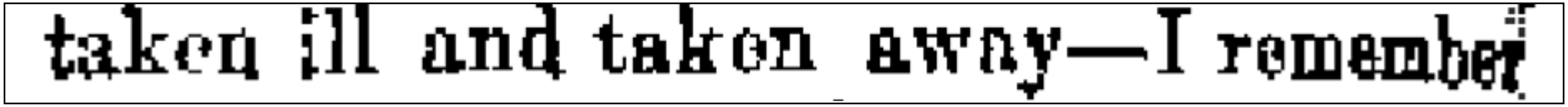


taken ill and taken away—I remember

System Output Example

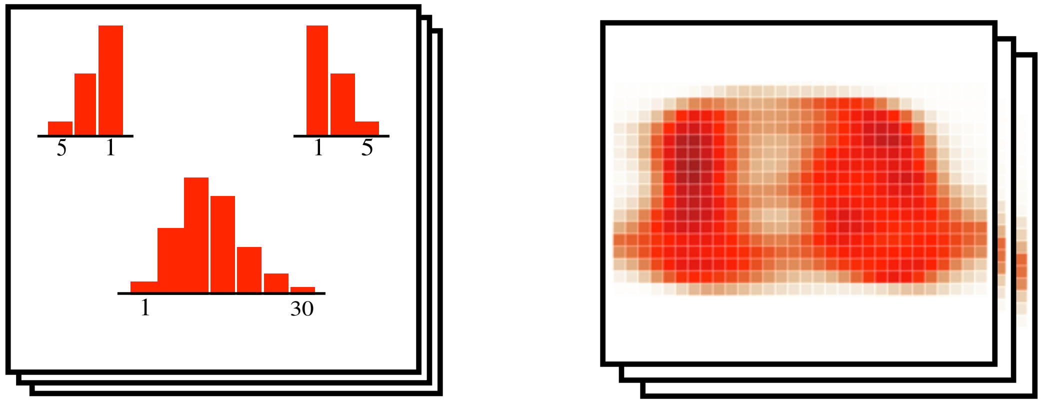




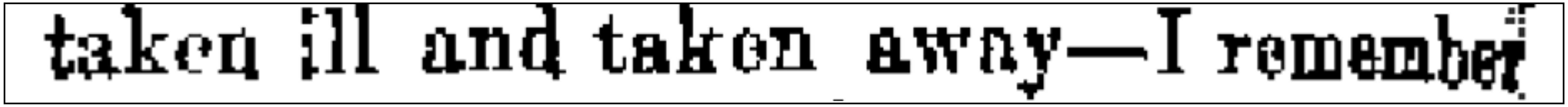


taken ill and taken away—I remember

System Output Example

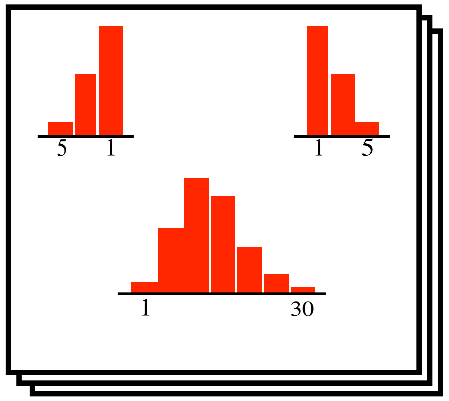


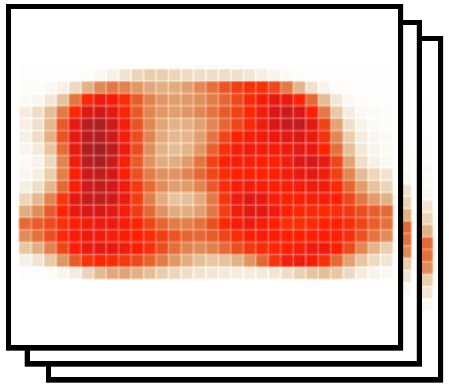
```
taken ill and taken  away -- I remember
```



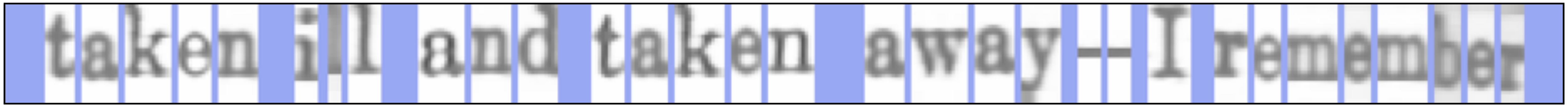
taken ill and taken away—I remember

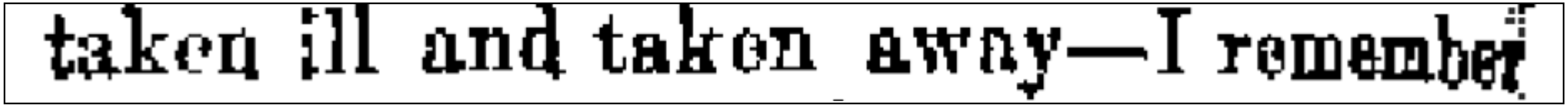
System Output Example



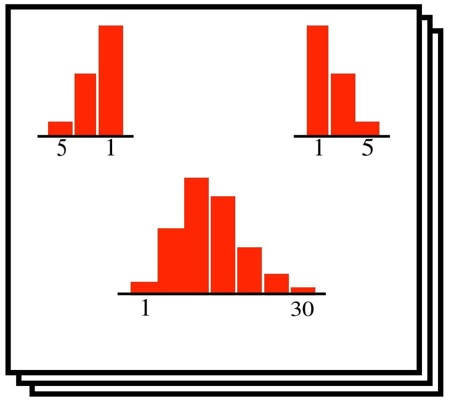


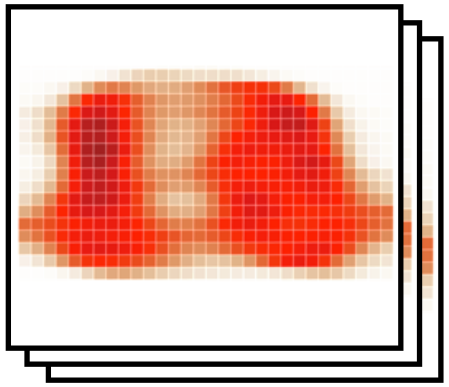
taken ill and taken away -- I remember



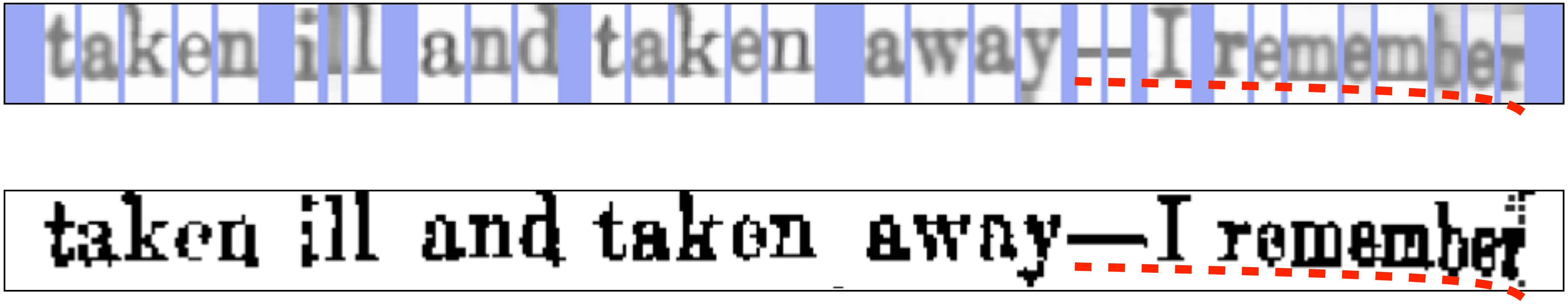


System Output Example

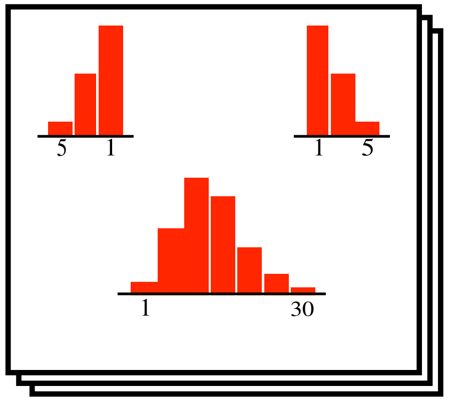


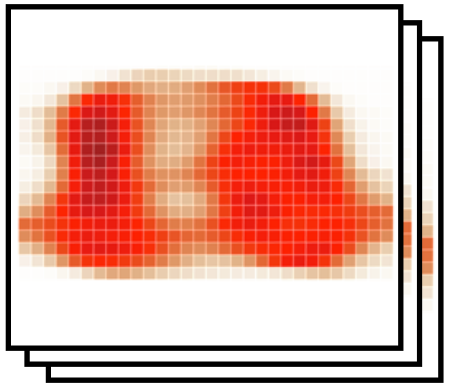


taken ill and taken away -- I remember

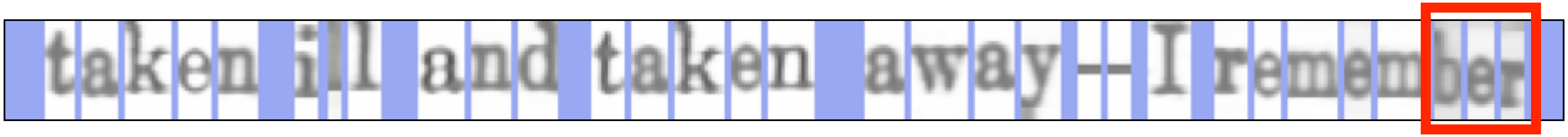

taken ill and taken away -- I remember
taken ill and taken away—I remember

System Output Example





taken ill and taken away -- I remember



Experiments

Experiments

Test data

Experiments

Test data

- Old Bailey (1715-1905)

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Baselines

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Baselines

- Google Tesseract

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Baselines

- Google Tesseract
- ABBYY FineReader 11

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Baselines

- Google Tesseract
- ABBYY FineReader 11

Language models

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Baselines

- Google Tesseract
- ABBYY FineReader 11

Language models

- New York Times

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Language models

- New York Times

  34M words NYT Gigaword

Baselines

- Google Tesseract
- ABBYY FineReader 11

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Baselines

- Google Tesseract
- ABBYY FineReader 11

Language models

- New York Times

  34M words NYT Gigaword

- Old Bailey

Experiments

Test data

- Old Bailey (1715-1905)

  20 images, 30 lines each

- Trove (1803-1893)

  10 images, 30 lines each

Language models

- New York Times

  34M words NYT Gigaword

- Old Bailey

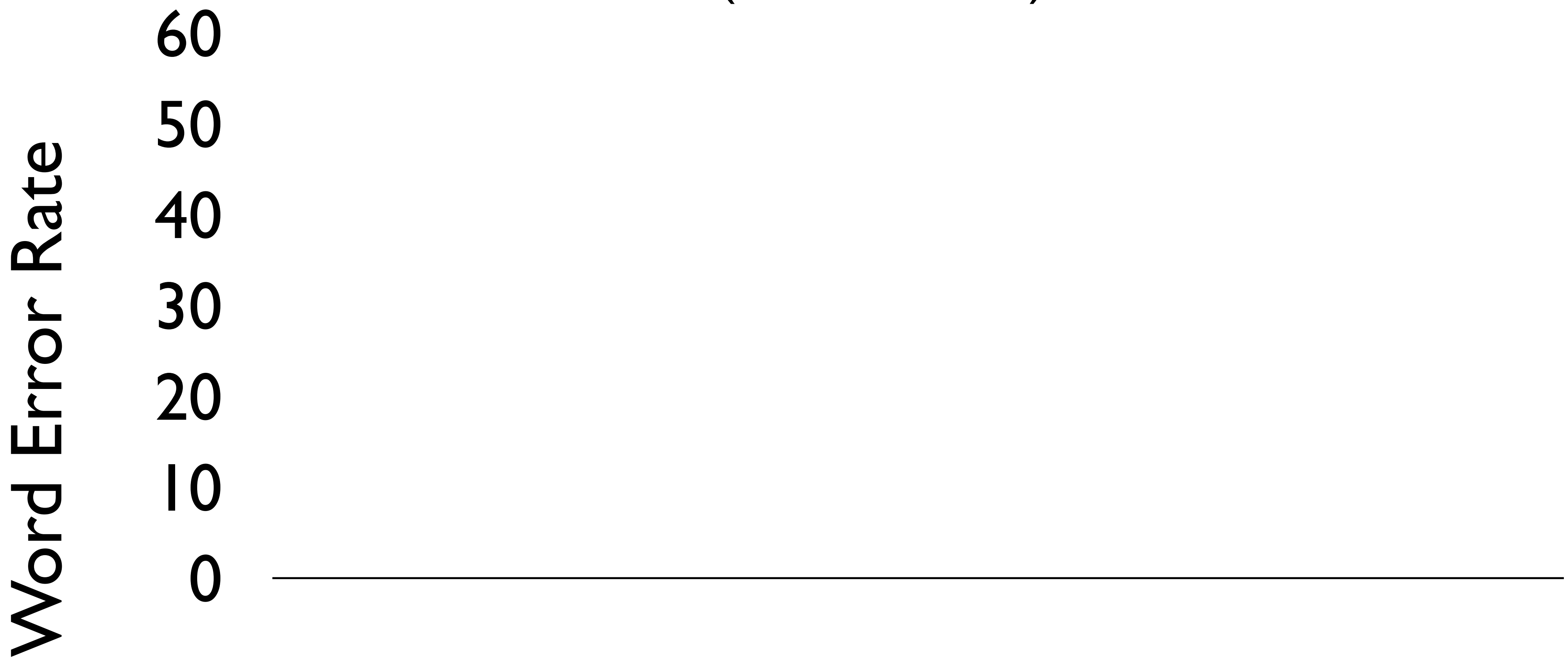
  32M words manually transcribed

Baselines

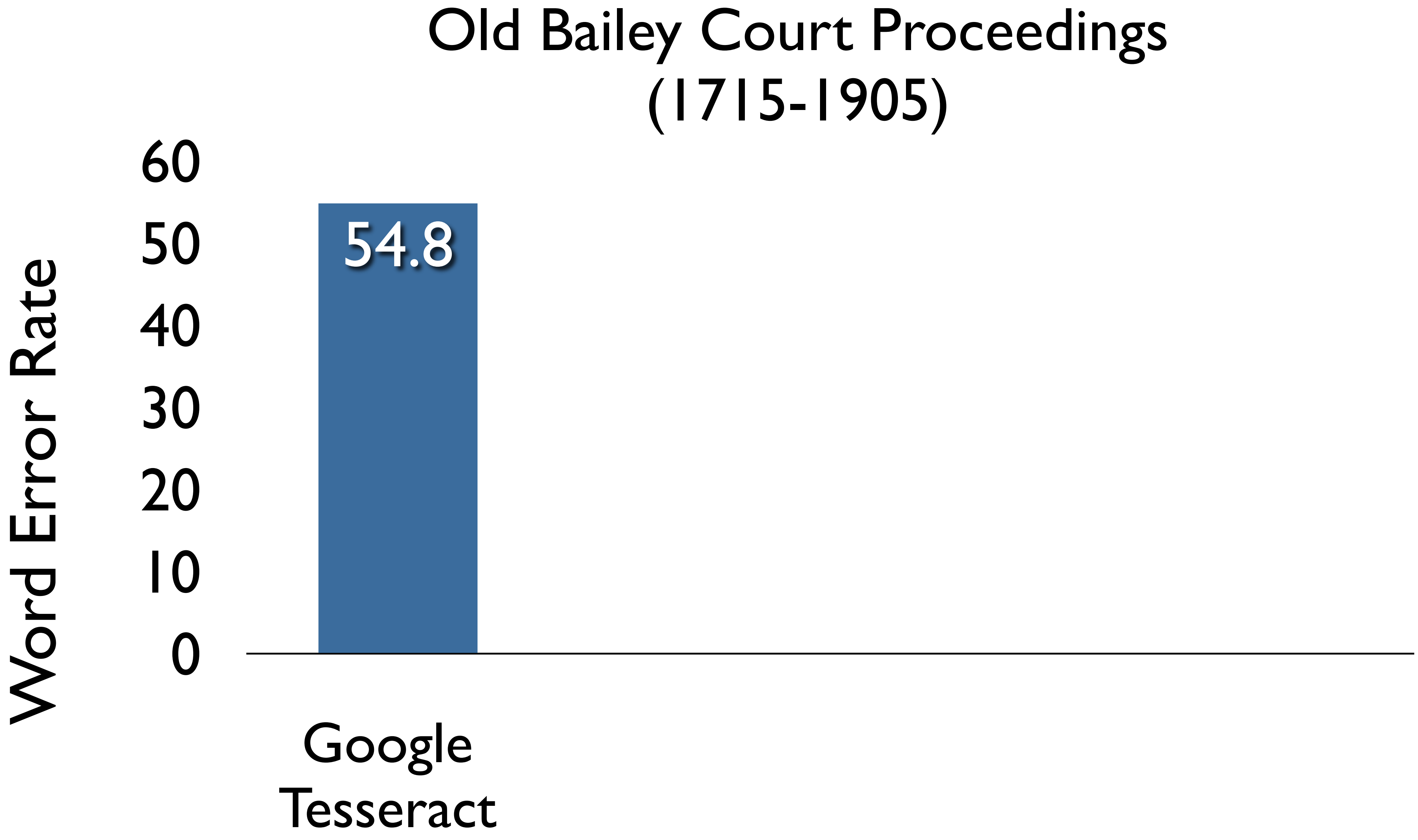
- Google Tesseract
- ABBYY FineReader 11

Results

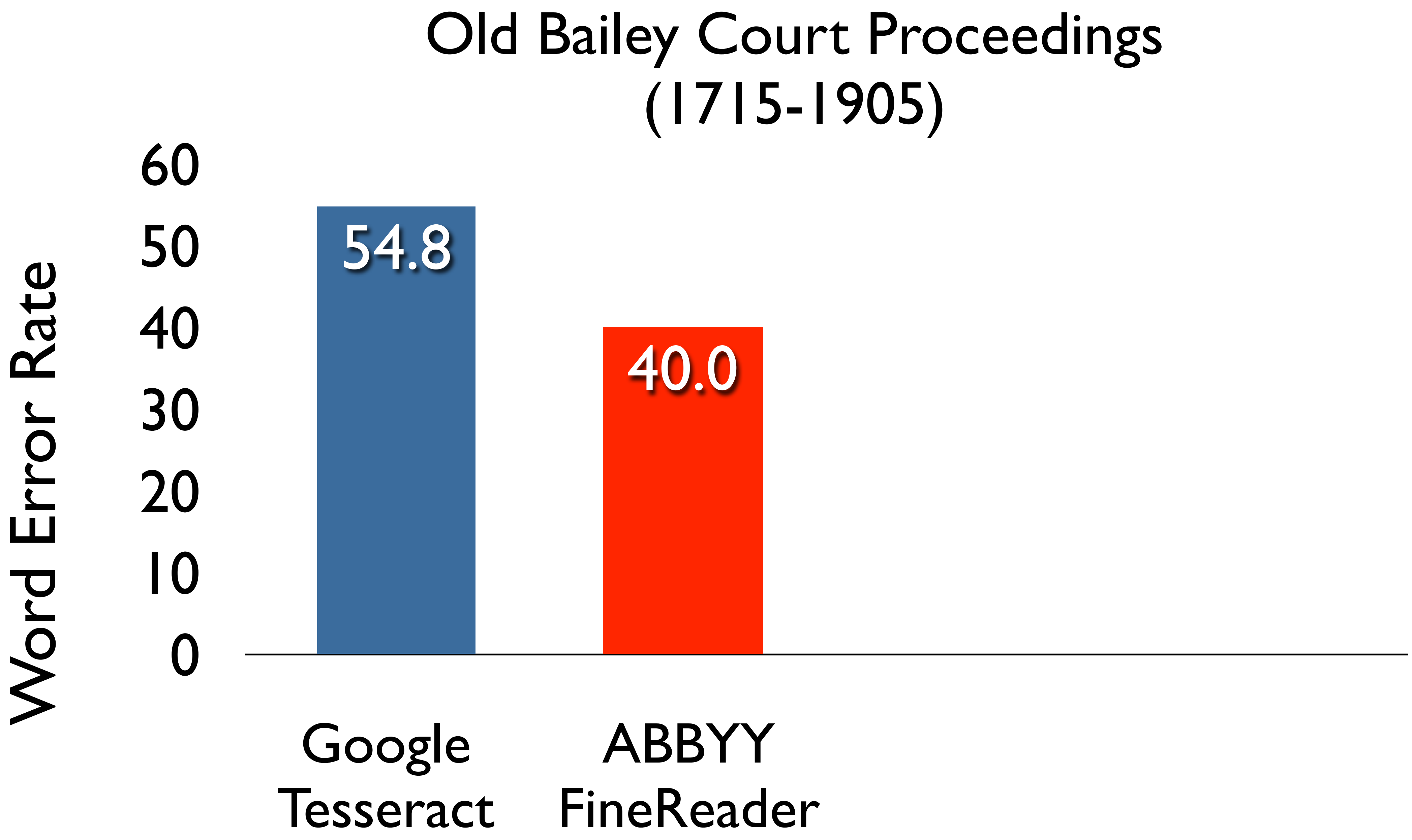
Old Bailey Court Proceedings
(1715-1905)



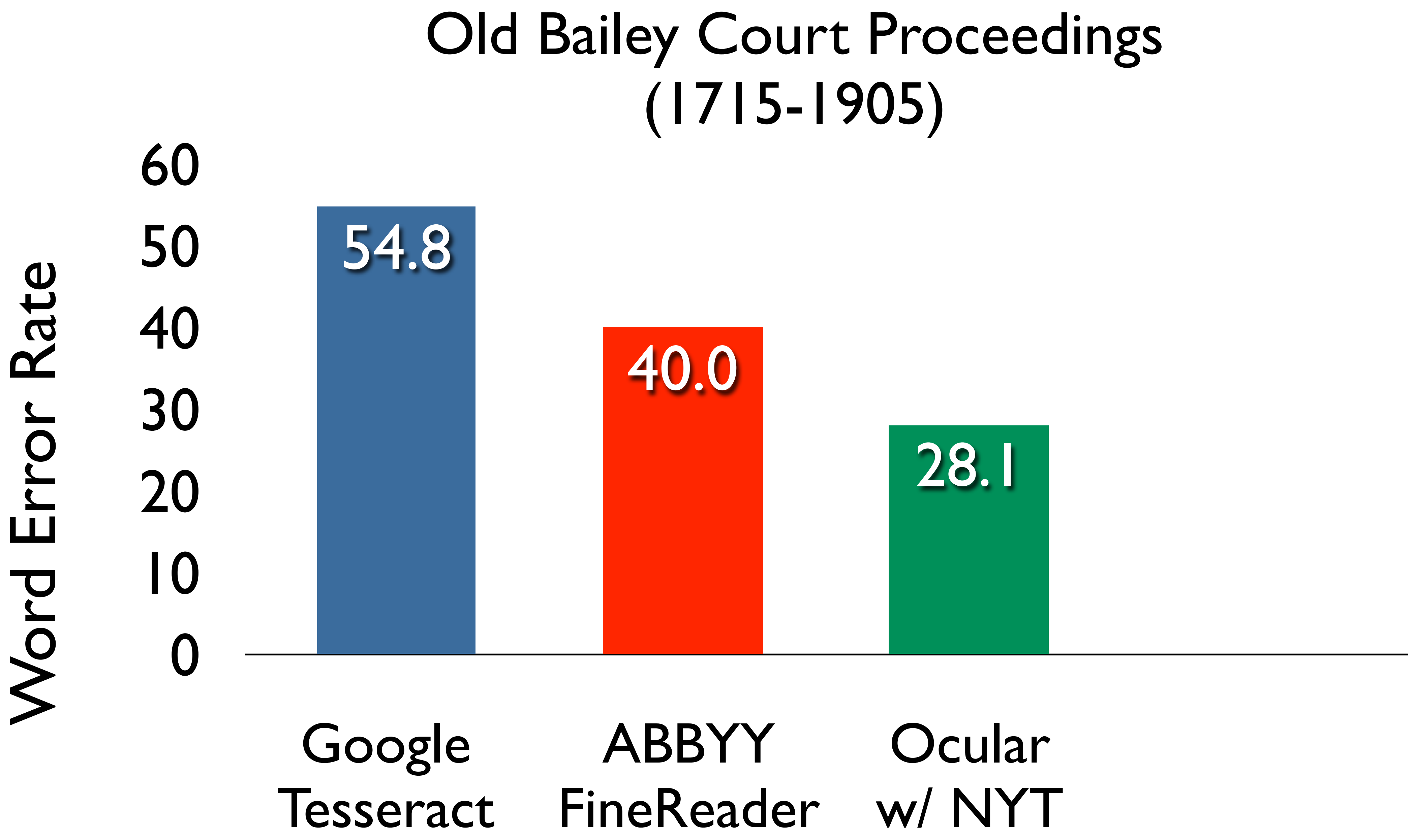
Results


Old Bailey Court Proceedings
(1715-1905)
Word Error Rate
60
50
40
30
20
10
0
54.8
Google
Tesseract

Results

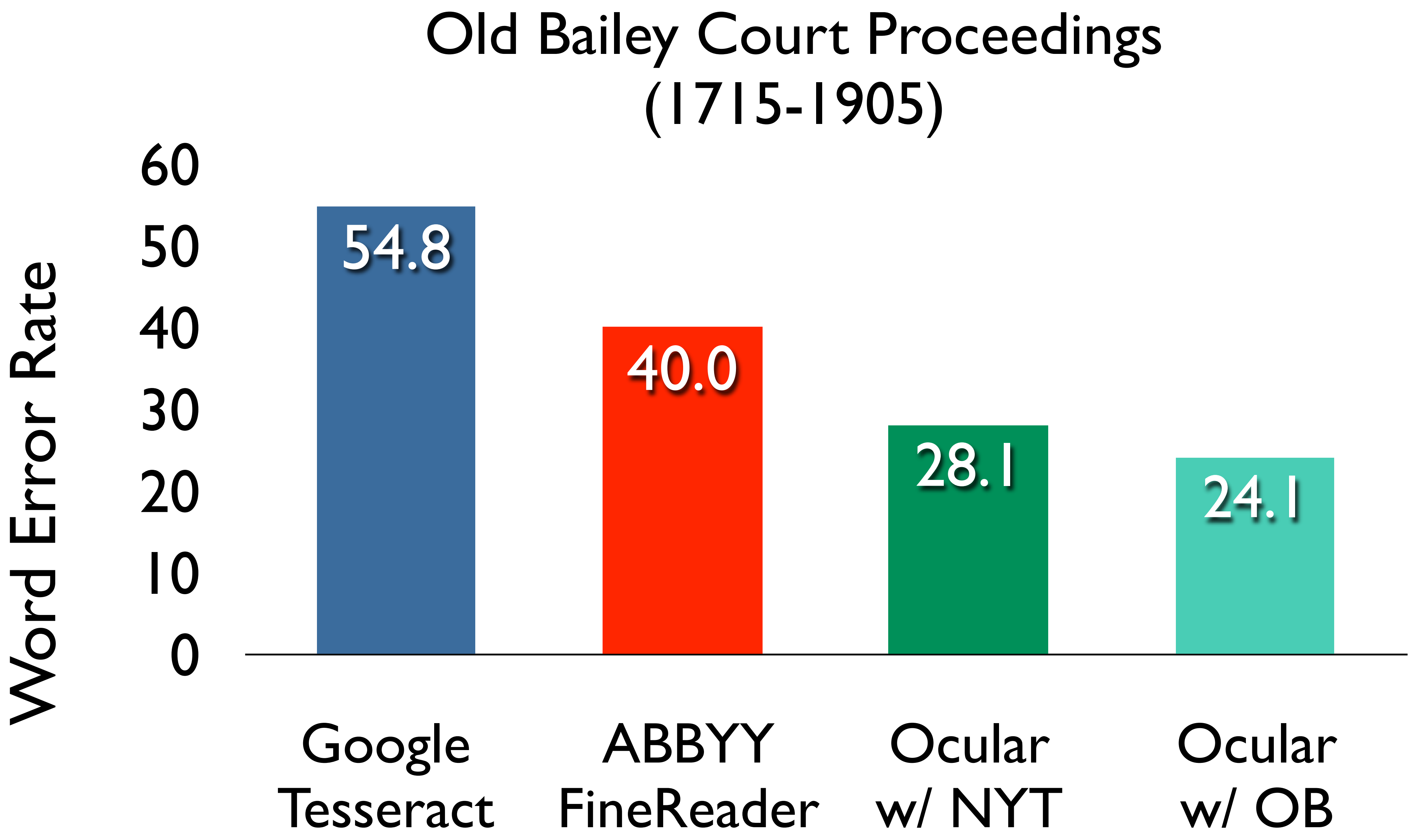

Old Bailey Court Proceedings (1715-1905)

Results



[Berg-Kirkpatrick et al. 2013]

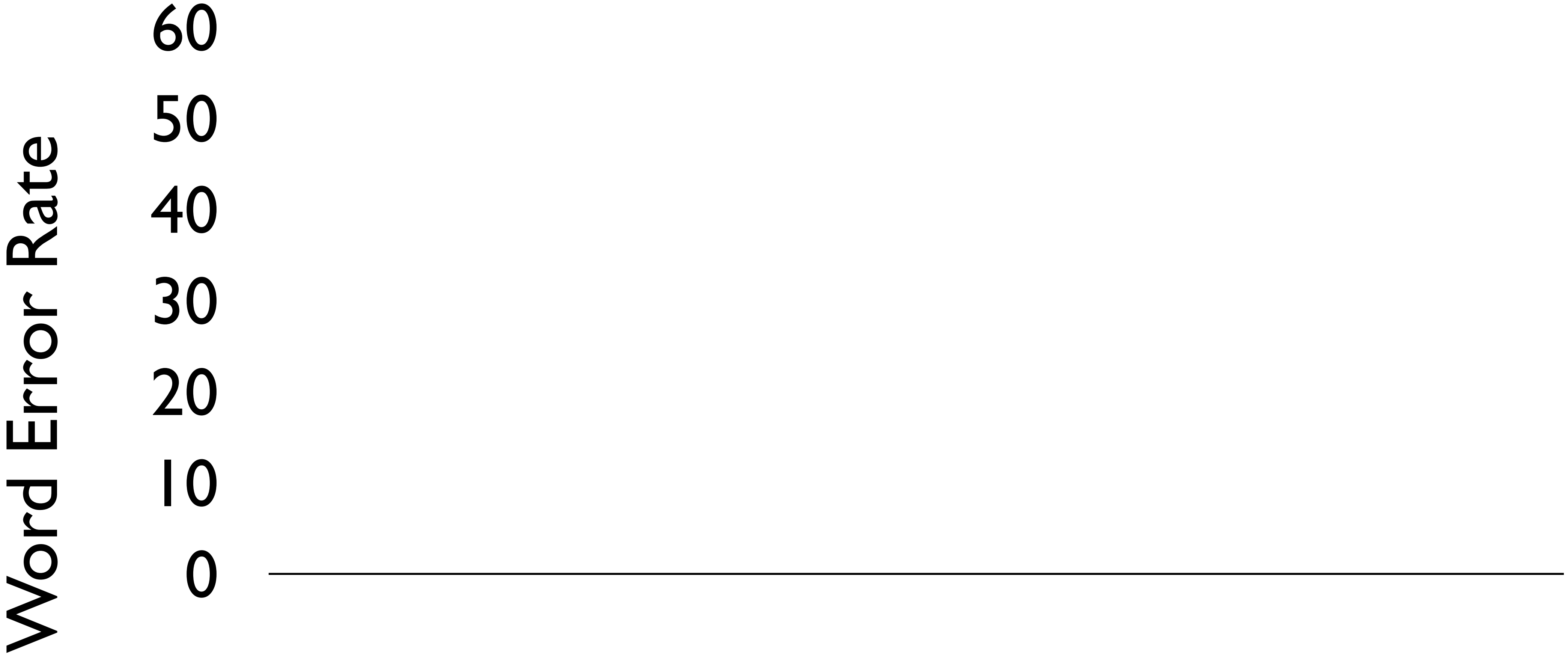
Results



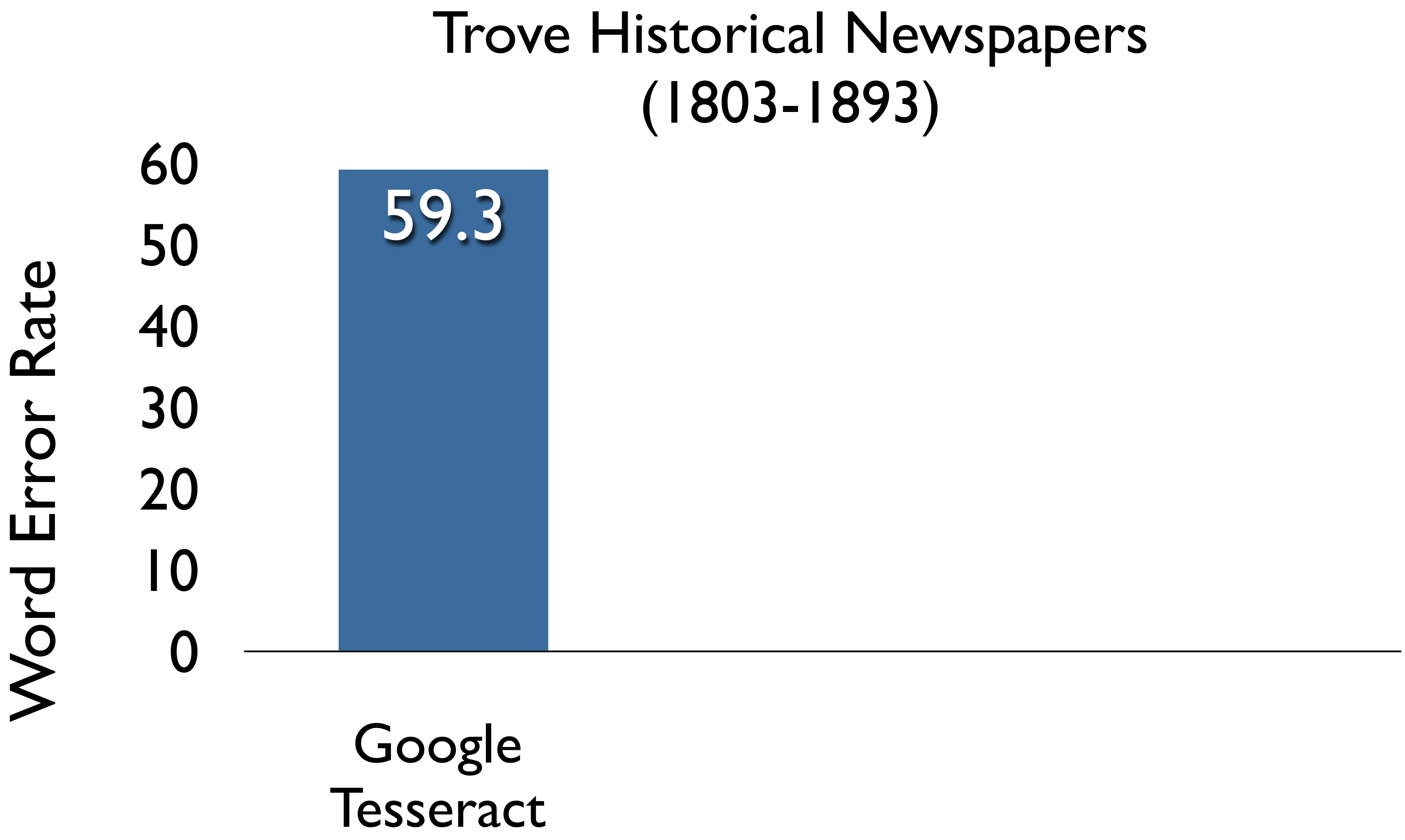
[Berg-Kirkpatrick et al. 2013]

Results

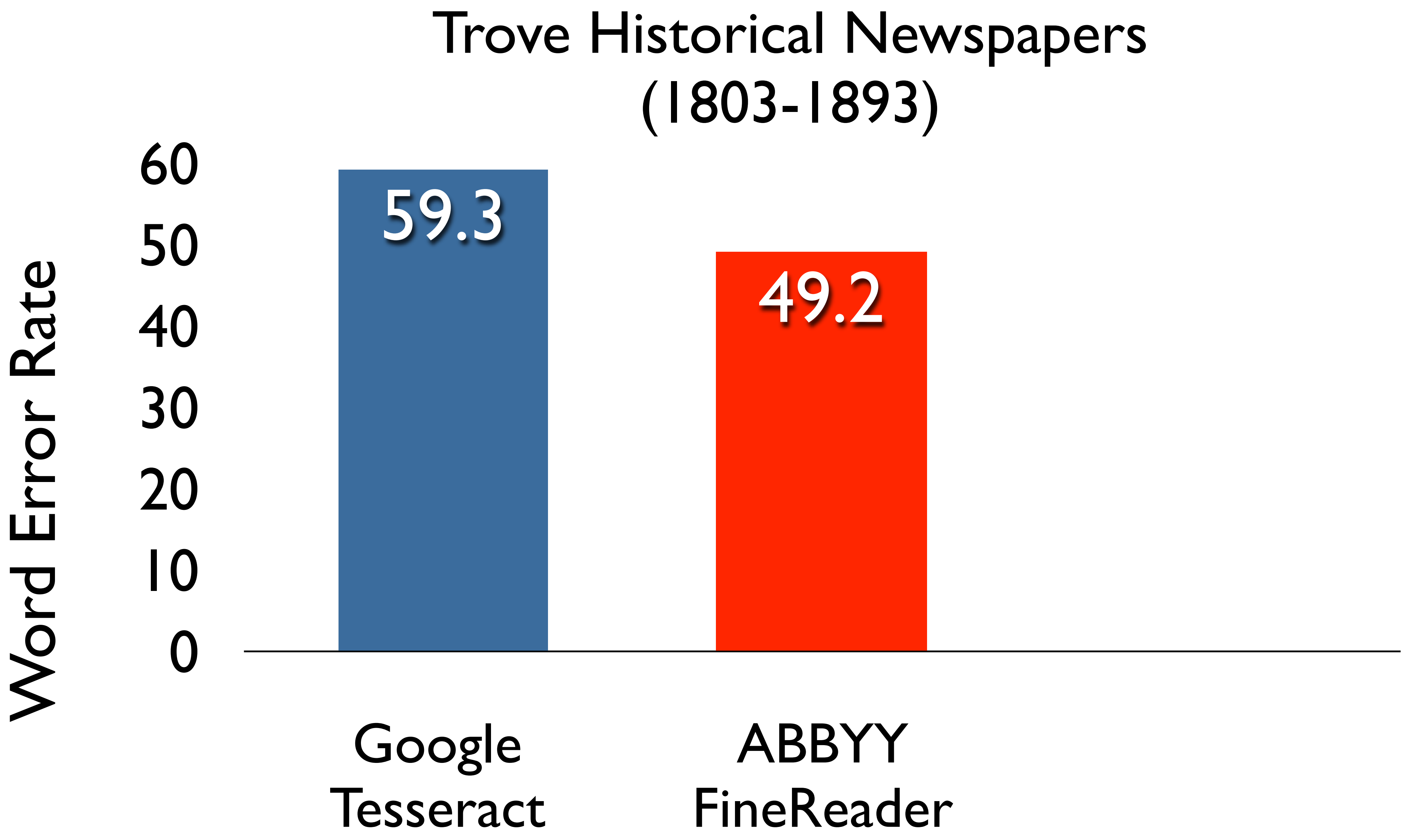
Trove Historical Newspapers
(1803-1893)



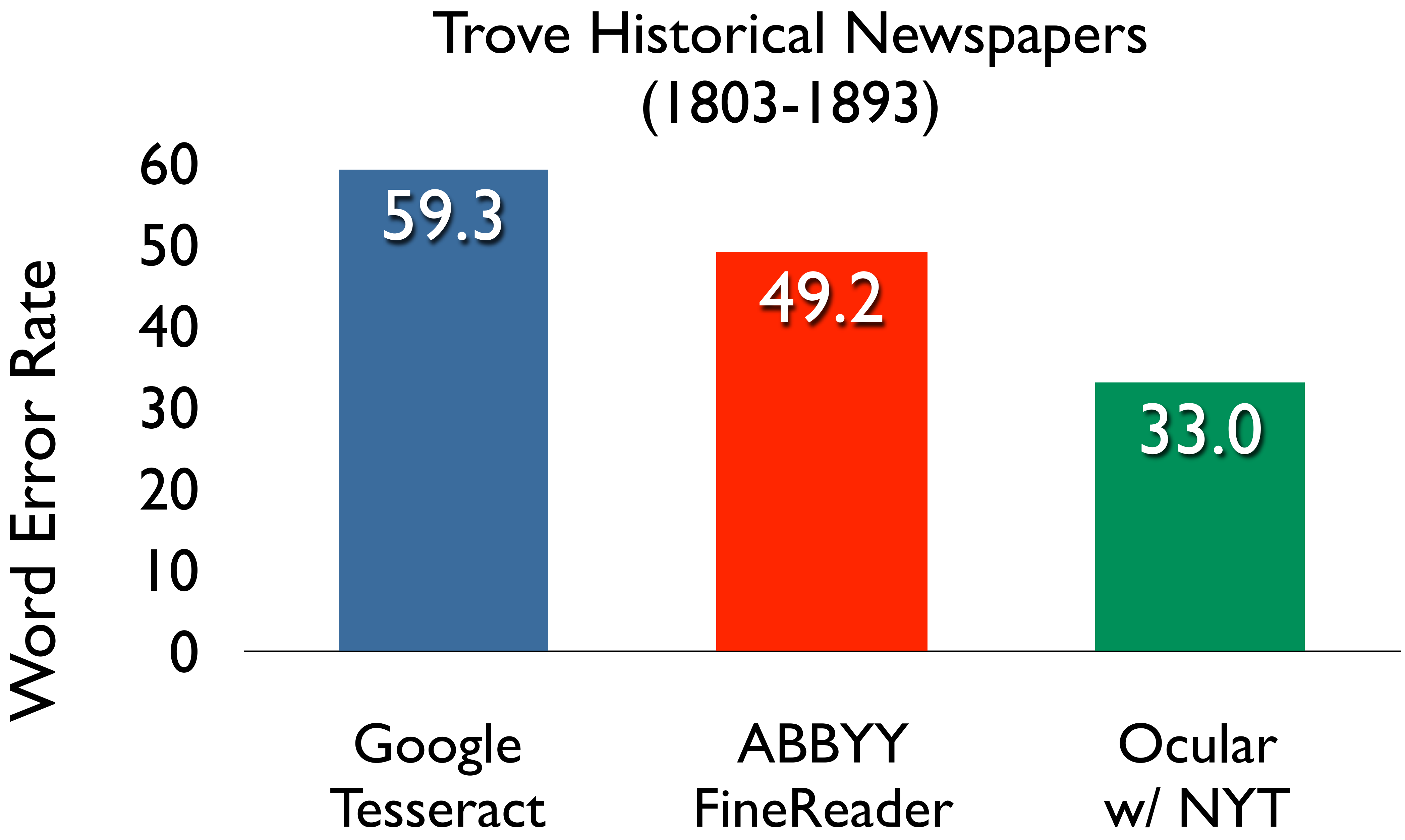
Results


Trove Historical Newspapers
(1803-1893)
Word Error Rate
60
50
40
30
20
10
0
59.3
Google
Tesseract

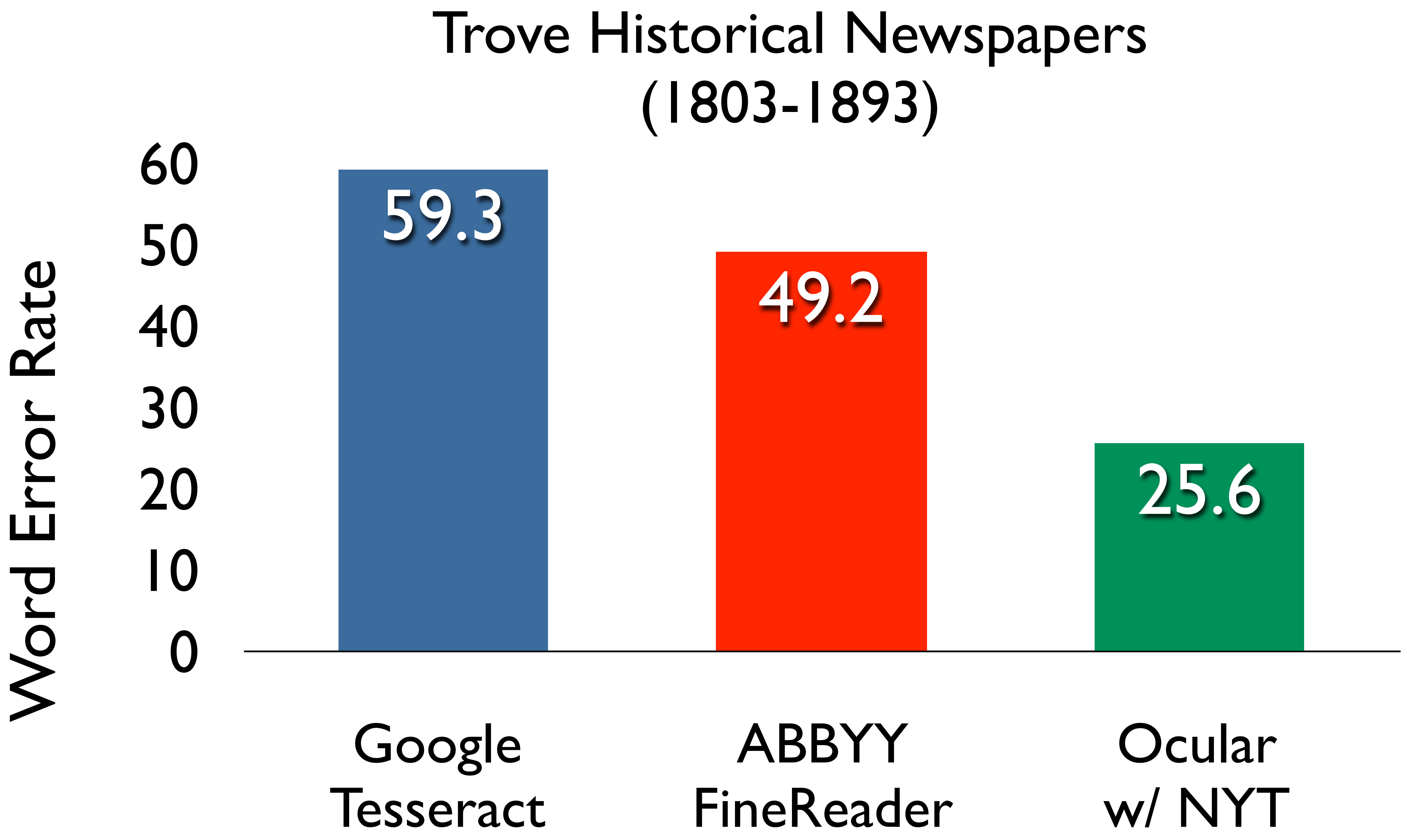
Results


Trove Historical Newspapers
(1803-1893)
Word Error Rate
60
50
40
30
20
10
0
59.3
49.2
Google
Tesseract
ABBYY
FineReader

Results


Trove Historical Newspapers
(1803-1893)
Word Error Rate
60
50
40
30
20
10
0
59.3
49.2
33.0
Google
Tesseract
ABBYY
FineReader
Ocular
w/ NYT

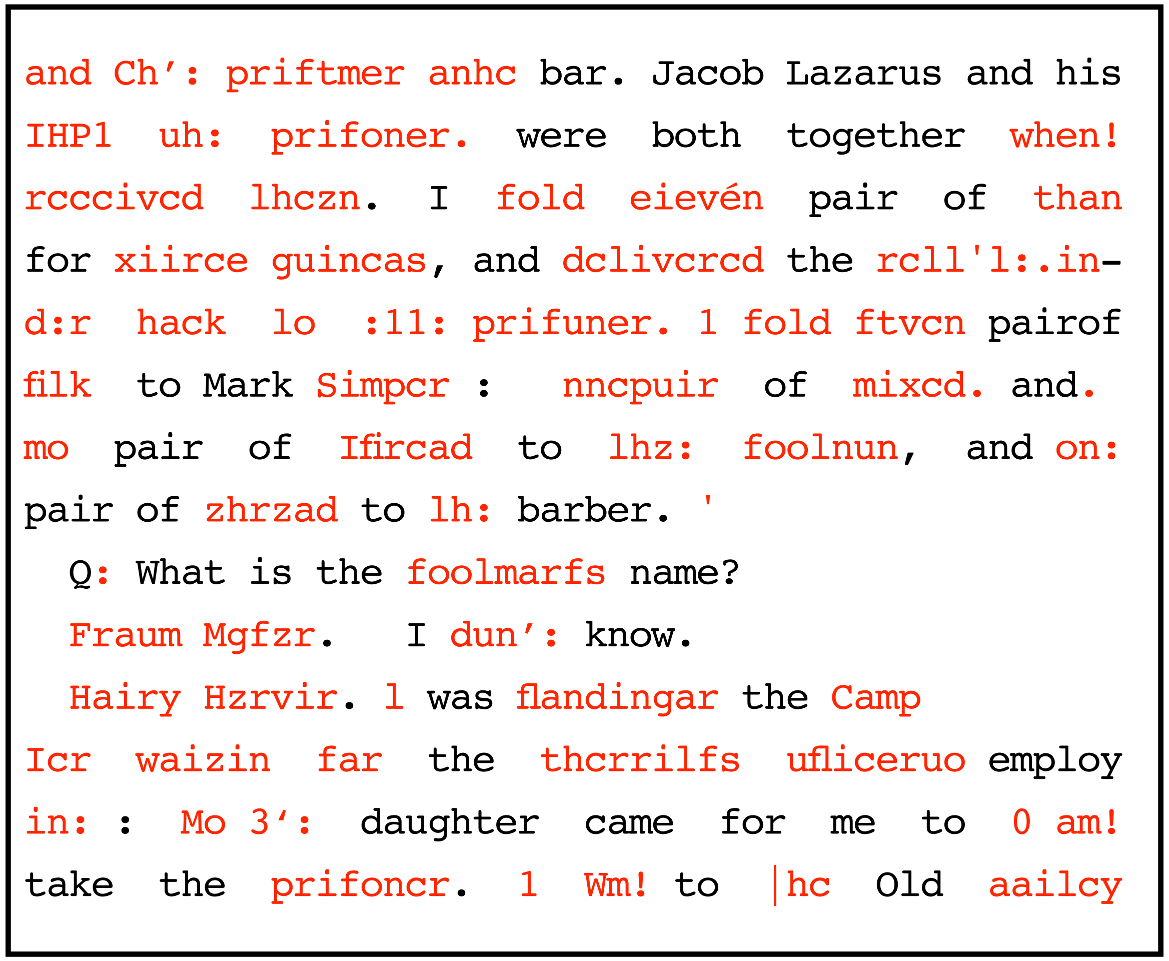
Results



[Berg-Kirkpatrick et al. 2014]

Transcription

Google Tesseract



```
and Ch’: priftmer anhc bar. Jacob Lazarus and his
IHP1  uh:  prifoner.  were  both  together  when!
rcccivcd  lhczn.  I  fold  eievén  pair  of  than
for xiirce guincas, and dclivcrcd the rcll'l:.in-
d:r  hack  lo  :11: prifuner. 1 fold ftvcn pairof
filk  to Mark Simpcr :   nncpuir  of  mixcd. and.
mo  pair  of  Ifircad  to  lhz:  foolnun,  and on:
pair of zhrzad to lh: barber. '
  Q: What is the foolmarfs name?
  Fraum Mgfzr.   I dun’: know.
  Hairy Hzrvir. l was flandingar the Camp
Icr  waizin  far  the  thcrrilfs  ufliceruo employ
in: :  Mo 3‘:  daughter  came  for  me  to  0 am!
take  the  prifoncr.  1  Wm!  to  |hc  Old  aailcy
```

Transcription

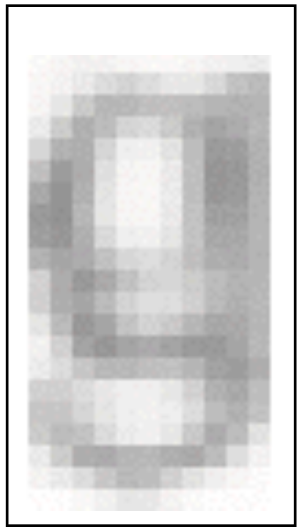
Google Tesseract

```
and Ch’: priftmer anhc bar. Jacob Lazarus and his
IHP1 uh: prifoner. were both together when!
rcccivcd lhczn. I fold eievén pair of than
for xiirce guincas, and dclivcrcd the rcll'l:.in-
d:r hack lo :11: prifuner. 1 fold ftvcn pairof
filk to Mark Simpcr : nncpuir of mixcd. and.
mo pair of Ifircad to lhz: foolnun, and on:
pair of zhrzad to lh: barber. '
  Q: What is the foolmarfs name?
  Fraum Mgfzr. I dun’: know.
  Hairy Hzrvir. l was flandingar the Camp
Icr waizin far the thcrrilfs ufliceruo employ
in: : Mo 3‘: daughter came for me to 0 am!
take the prifoncr. 1 Wm! to |hc Old aailcy
```

Ocular

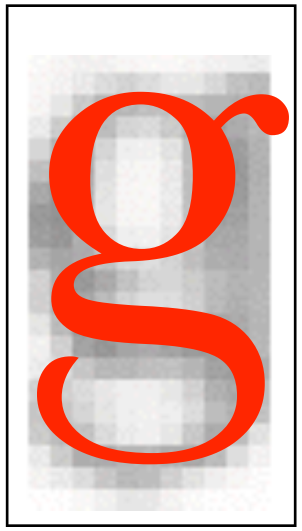
```
the prisoner at the bar. Jacob Lazarus and his
wife, the prisoners were both together when I
received them. I sold eleven pair of them
for three guineas, and delivered the remain-
der back to the prisoner. I sold, seven pair of
silk to Mark Simpert one pair of mixed, and
two pair of thread to the footman, and one
pair of thread to the barber,
  Ms. What in the footman's name?
  Franco Asyut, I don't know-
  Nearly Norris. I was standing at the Comp-
ter waiting for the sherrill's officers to employ
me a Moses's daughter came for me to go and
take the prisoner. I went to the Old Bailey
```

Learned Fonts



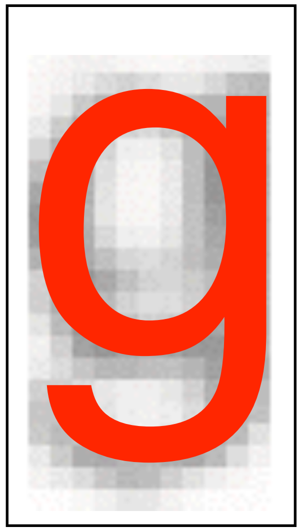
Initializer

Learned Fonts



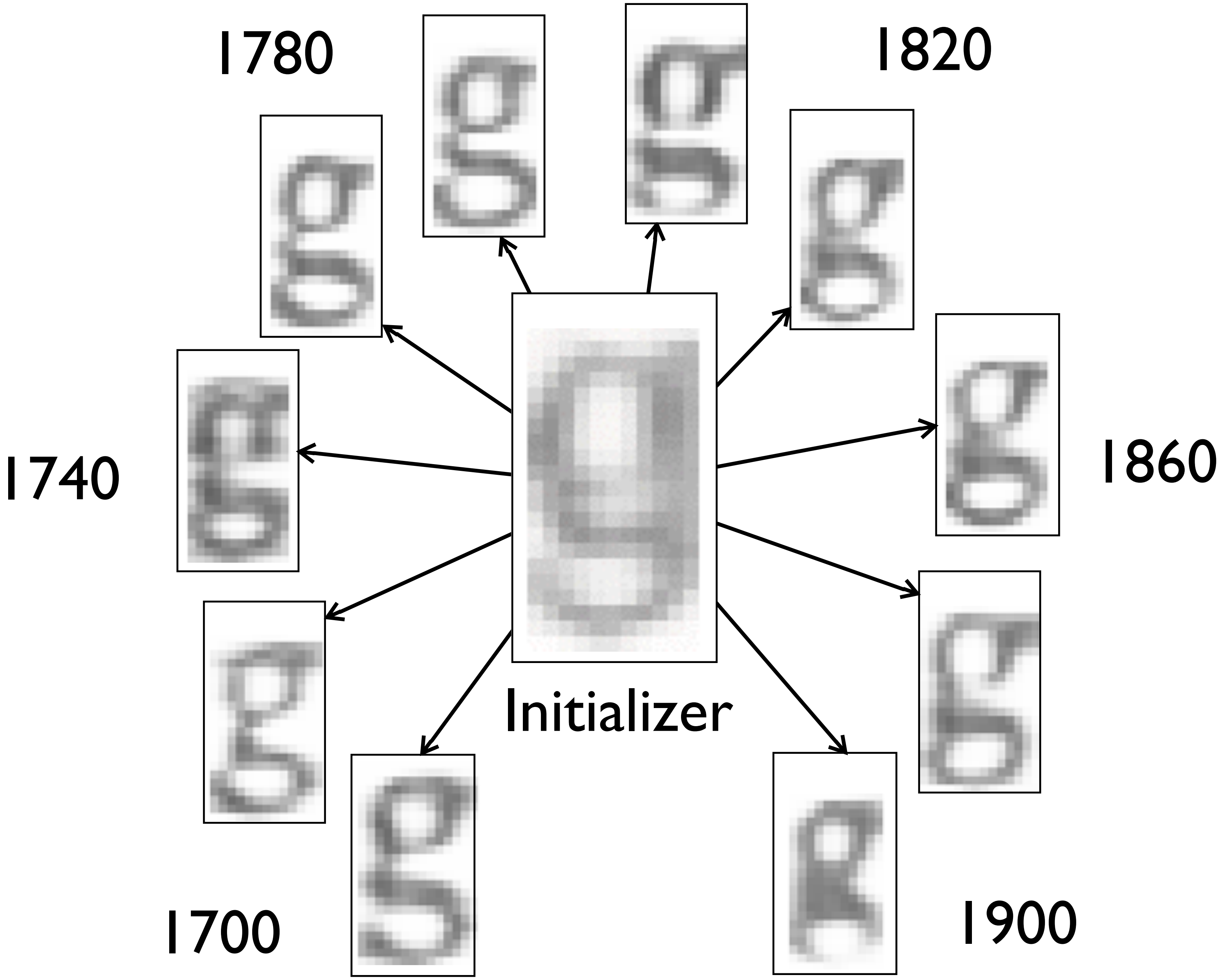
Initializer

Learned Fonts

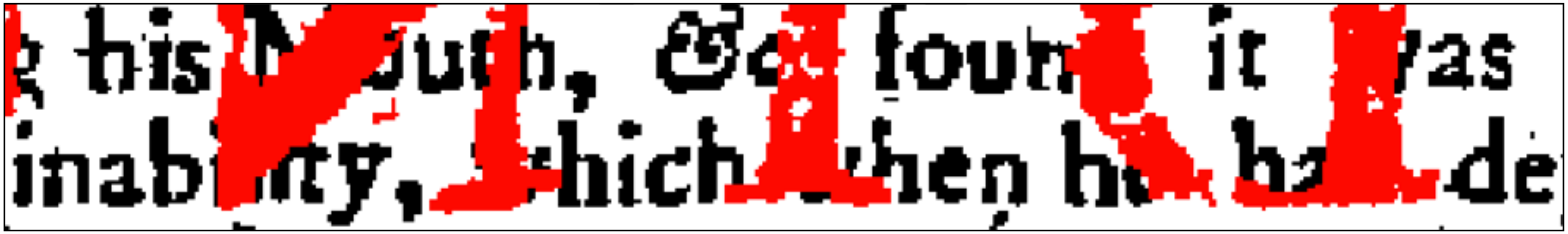


Initializer

Learned Fonts

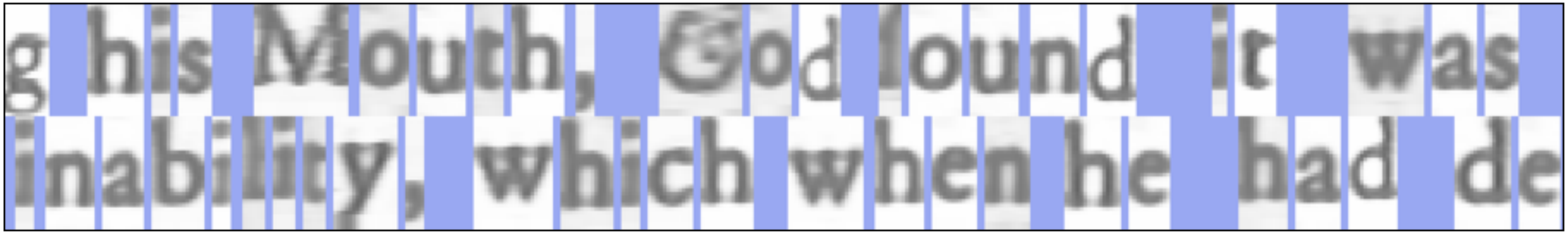


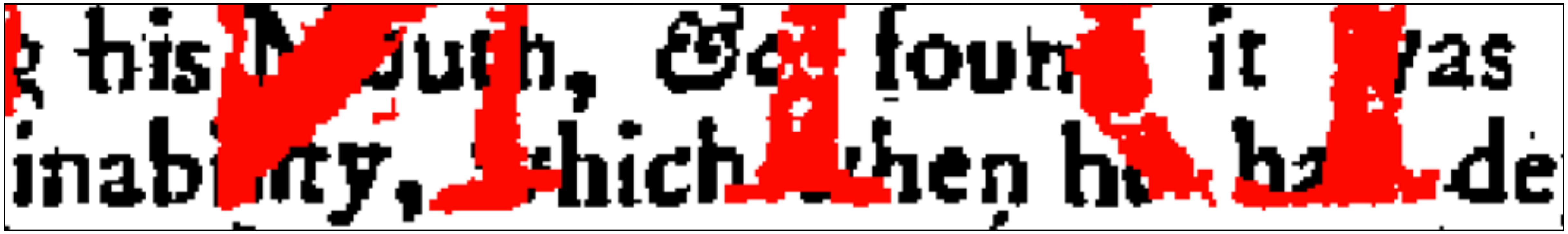
Unobserved Pixels



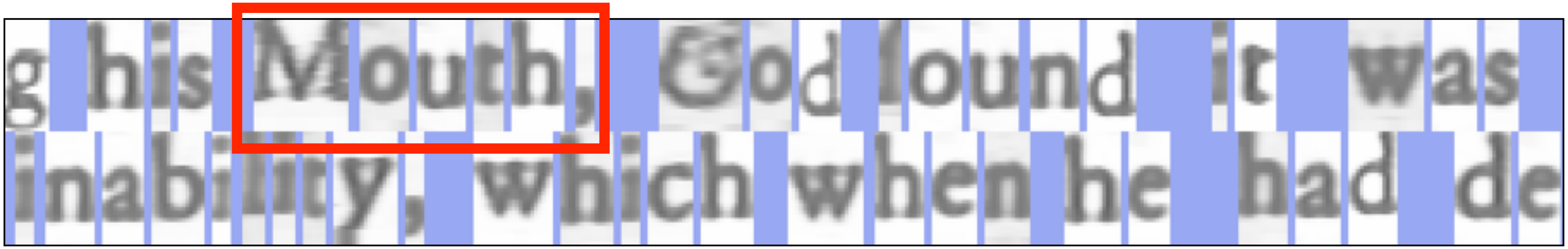
g his M uth, &c foun it as
inab ty, hich hen h h de

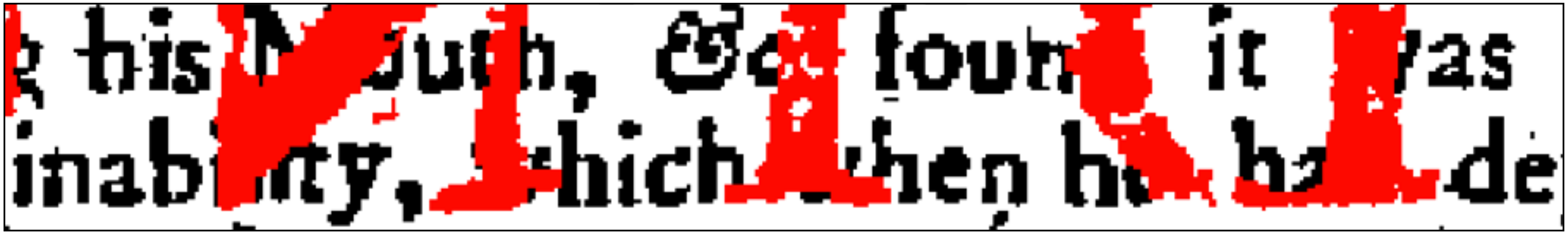
Unobserved Pixels


g his Mouth, God found it was
inability, which when he had de

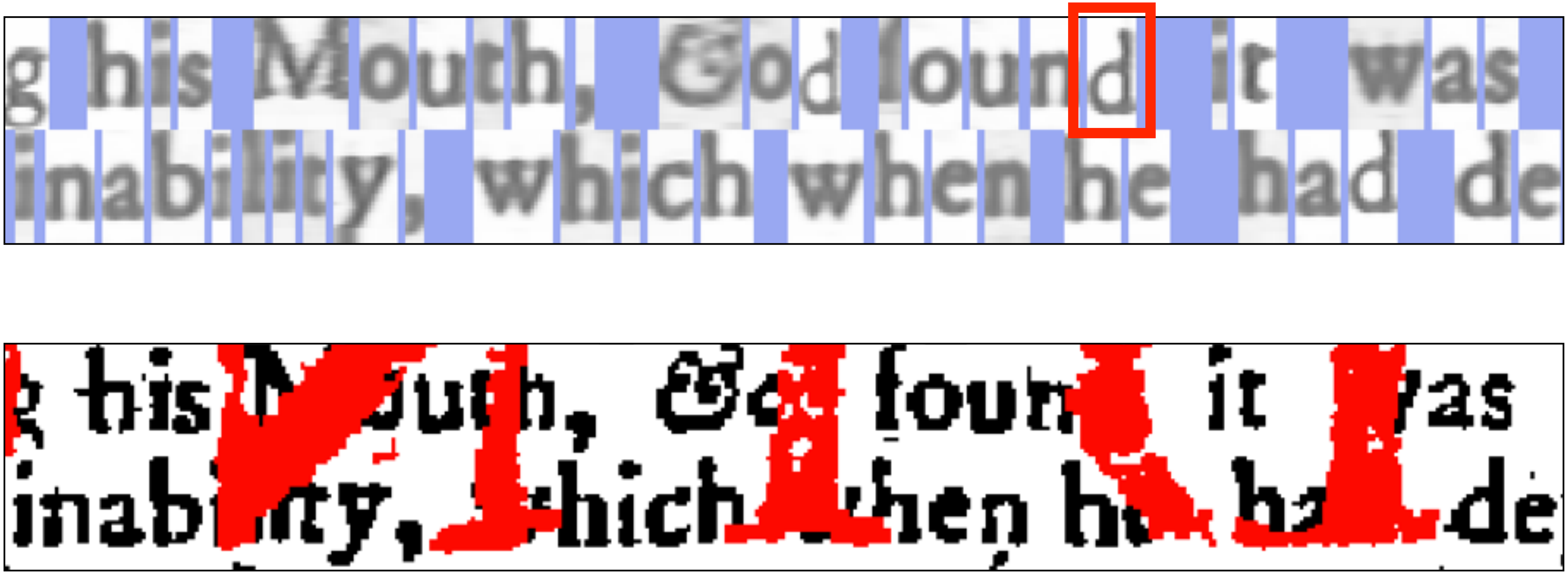


Unobserved Pixels


g his Mouth, God found it was
inability, which when he had de



Unobserved Pixels



g his Mouth, God found it was
inability, which when he had de

Conclusion

Conclusion

- Unsupervised font learning yields state-of-the-art results on documents where font is unknown

Conclusion

- Unsupervised font learning yields state-of-the-art results on documents where font is unknown
- Generatively modeling sources of noise specific to printing-press era documents is effective

Conclusion

- Unsupervised font learning yields state-of-the-art results on documents where font is unknown
- Generatively modeling sources of noise specific to printing-press era documents is effective
- Ocular available as a downloadable tool: nlp.cs.berkeley.edu/ocular.shtml

Conclusion

Thanks!