```
a', ' mixture', ' of']
1:3
Patching hidden states…

100%|      | 36/36 [00:15<00:00,  2.27it/s]

1000000
['Les', 'lie', ' Moon', 'ves', ' is', ' employed', ' by']
['Les*', 'lie*', ' Moon*', 'ves*', ' is', ' employed', ' by']
0:4
Patching hidden states…

100%|      | 36/36 [00:10<00:00,  3.54it/s]

1000000
['The', ' original', ' language', ' of', ' De', ' fin', 'ibus', ' bon', 'orum',
' et', ' mal', 'orum', ' is', ' the', ' same', ' as', ' the']
['The', ' original', ' language', ' of', ' De*', ' fin*', 'ibus*', ' bon*',
'orum*', ' et*', ' mal*', 'orum*', ' is', ' the', ' same', ' as', ' the']
4:12
Patching hidden states…

100%|      | 36/36 [00:25<00:00,  1.43it/s]
```


Gaussian Noise (mean = 0, std = 1e-06)
```
p*,h(Antarctica)-p*(Antarctica)
```

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(Apple)-p*(Apple)

201

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 1e-06)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.001)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.01)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.1)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 0.5)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 1)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 1)

Gaussian Noise (mean = 0, std = 1)

Gaussian Noise (mean = 0, std = 1)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 1)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 1)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 1)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 1)

Gaussian Noise (mean = 0, std = 1)

Gaussian Noise (mean = 0, std = 1)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 1.5)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 2.5)

p*,h(CBS)-p*(CBS)

![Gaussian Noise (mean = 0, std = 2.5) heatmap with y-axis labels: The, original, language, of, De*, fin*, ibus*, bon*, orum*, et*, mal*, orum*, is, the, same, as, the. Colorbar labeled p*,h(Latin)-p*(Latin) with scale 1e-5.](image)

![Gaussian Noise (mean = 0, std = 5) heatmap with y-axis labels: V*, inson*, Mass*, if*, is, located, in, the, continent, of. Colorbar labeled p*,h(Antarctica)-p*(Antarctica).](image)

239

Gaussian Noise (mean = 0, std = 5)

Gaussian Noise (mean = 0, std = 5)

Gaussian Noise (mean = 0, std = 5)

Gaussian Noise (mean = 0, std = 5)

241

Gaussian Noise (mean = 0, std = 5)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 5)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 5)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 5)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 5)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 10)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 10)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 10)

p*,h(Amazon)-p*(Amazon)

Gaussian Noise (mean = 0, std = 10)

Gaussian Noise (mean = 0, std = 10)

Gaussian Noise (mean = 0, std = 10)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 10)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 10)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 10)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 10)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 100)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 100)

Gaussian Noise (mean = 0, std = 100)

Gaussian Noise (mean = 0, std = 100)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 100)

p*,h(Apple)-p*(Apple)

251

Gaussian Noise (mean = 0, std = 100)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 100)

p*,h(Europe)-p*(Europe)

Gaussian Noise (mean = 0, std = 100)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 100)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 100)

p*,h(Latin)-p*(Latin)

Gaussian Noise (mean = 0, std = 1000000)

p*,h(Antarctica)-p*(Antarctica)

Gaussian Noise (mean = 0, std = 1000000)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 1000000)

p*,h(Amazon)-p*(Amazon)

255

Gaussian Noise (mean = 0, std = 1000000)

p*,h(CBS)-p*(CBS)

Gaussian Noise (mean = 0, std = 1000000)

p*,h(Apple)-p*(Apple)

Gaussian Noise (mean = 0, std = 1000000)

p*,h(Italy)-p*(Italy)

Gaussian Noise (mean = 0, std = 1000000)

p*,h(Europe)-p*(Europe)

257

Gaussian Noise (mean = 0, std = 1000000)

p*,h(English)-p*(English)

Gaussian Noise (mean = 0, std = 1000000)

p*,h(CBS)-p*(CBS)