

Dialbot: muturretik muturrerako dialogo sistema

Julen Etxaniz eta Aitor Zubillaga

Sarrera

Dialogo sistema sare errekorrenteak eta pelikulatako azpitutuluak erabiliz. Dialogoa itzulpen ataza bezala landu.

Itzulpen automatikoa :

- Sarrera -> esaldia jatorri hizkuntzan: “Egun on guztioi.”
- Irteera -> esaldia helburu hizkuntzan: “Buenos días a todos.”

Dialogoa:

- Sarrera -> dialogoko partaide baten esaldia: “Egun on guztioi.”
- Irteera -> sarrerako esaldiari erantzuna: “Baita zuri ere.”

Helburuak

- Z1: Ingeleserako entrenatutako sistema aztertu eta probatu.
- Z2: Sistema berri bat entrenatu euskararako.
- Z3.1: Sistema moldatu telegrameko bot bezala funtziona dezan.
- Z3.2: Testuingurua kontuan hartzen duen sistema berri bat garatu.

Datuak eskuratu

- Opensubtitles filmetako azpitituluak
- Lerro bakoitzean dialogoko partaide baten esaldia
- Euskararako 1.000.000 lerro inguru
- 500.000 lerro erabili ditugu

```
11 -Karswell doktorea?  
12 -Nor zaitugu?  
13 Harrington irakaslea.  
14 Esaiozu oso kontu larria dela.  
15 Karswell jauna ez dago etxean.  
16 Badakit etxean dagoela.  
17 Ez naiz hemendik joango.  
18 -Sentitzen dut, jauna.  
19 -Zoaz...  
20 -Karswell.  
21 -Ondo da, Bates.  
22 Sartu.  
23 Hemendik.  
24 -Arratsalde on.  
25 -Baita zuri ere.
```

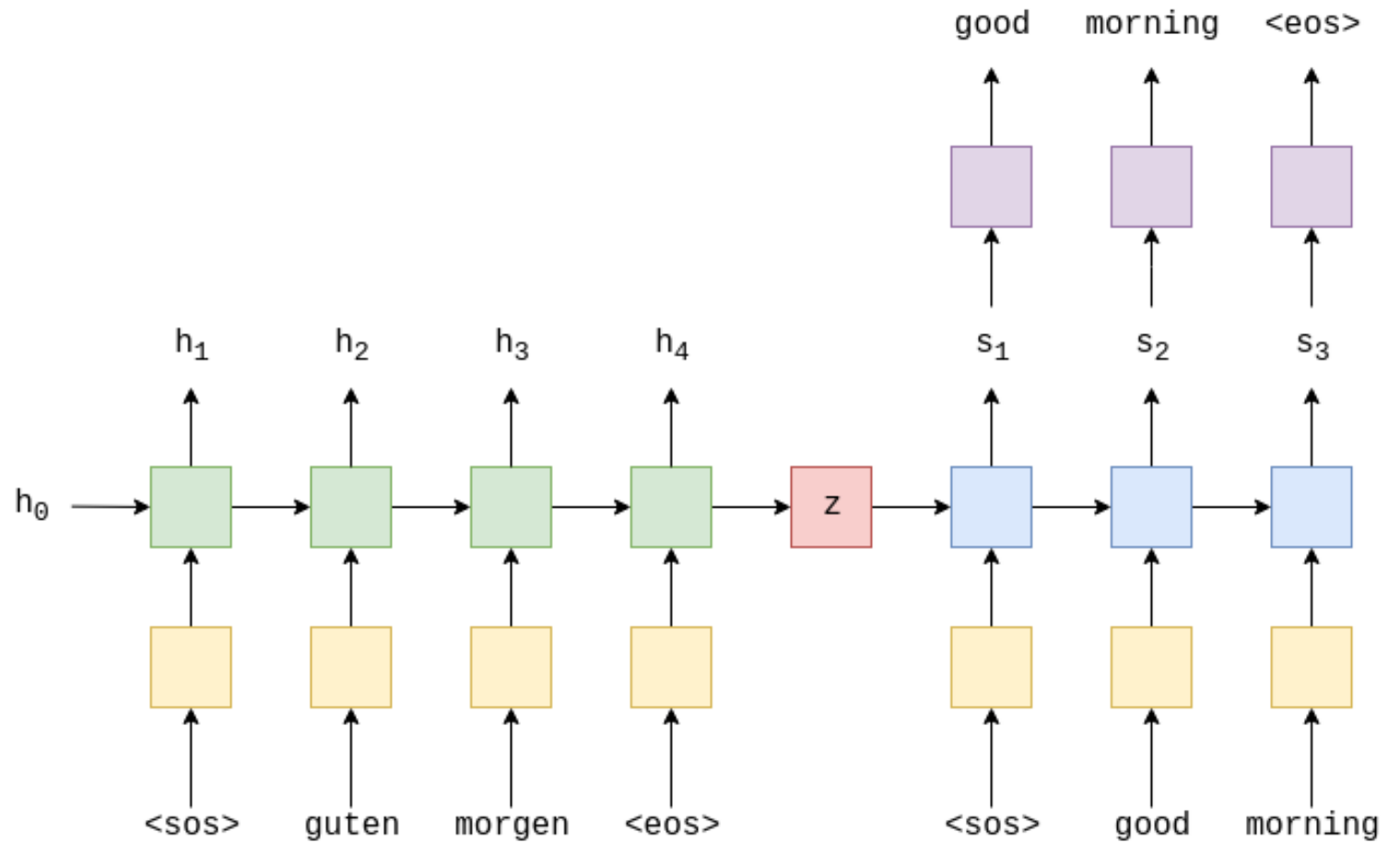
Datuak prozesatu

- Testua garbitu:
 - Minuskulatara bihurtu
 - Letrak eta .?! ikurrak bakarrik utzi
- Testua tokenizatu
- Dialogo formatuan jarri
 - Lerro bakoitzean dialogoan jarraian dauden bi esaldi
- Datuak banatu: train %80, valid %10, test %10

```
11  karswell doktorea ? nor zaitugu ?
12  nor zaitugu ?  harrington irakaslea .
13  harrington irakaslea .  esaiozu oso kontu larria dela .
14  esaiozu oso kontu larria dela .  karswell jauna ez dago etxean .
15  karswell jauna ez dago etxean .  badakit etxean dagoela .
16  badakit etxean dagoela .  ez naiz hemendik joango .
17  ez naiz hemendik joango .  sentitzen dut jauna .
18  sentitzen dut jauna .  zoaz .
19  zoaz .  karswell .
20  karswell .  ondo da bates .
21  ondo da bates .  sartu .
22  sartu .  hemendik .
23  hemendik .  arratsalde on .
24  arratsalde on .  baita zuri ere .
```

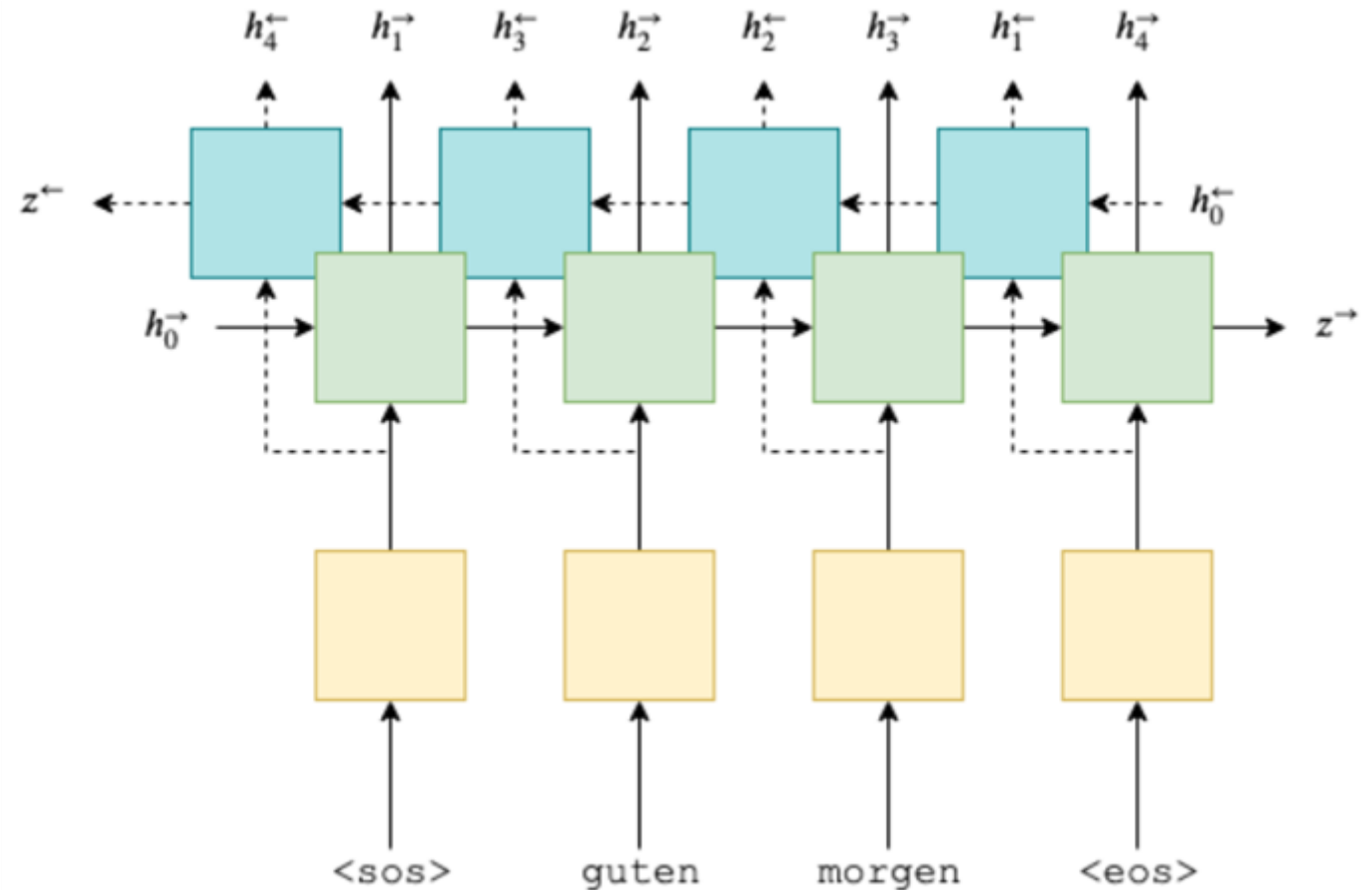
Eredua

- Seq2Seq eredua
- Itzulpenerako eredua
- Bi RNN
 - Kodetzailea
 - Deskodetzailea



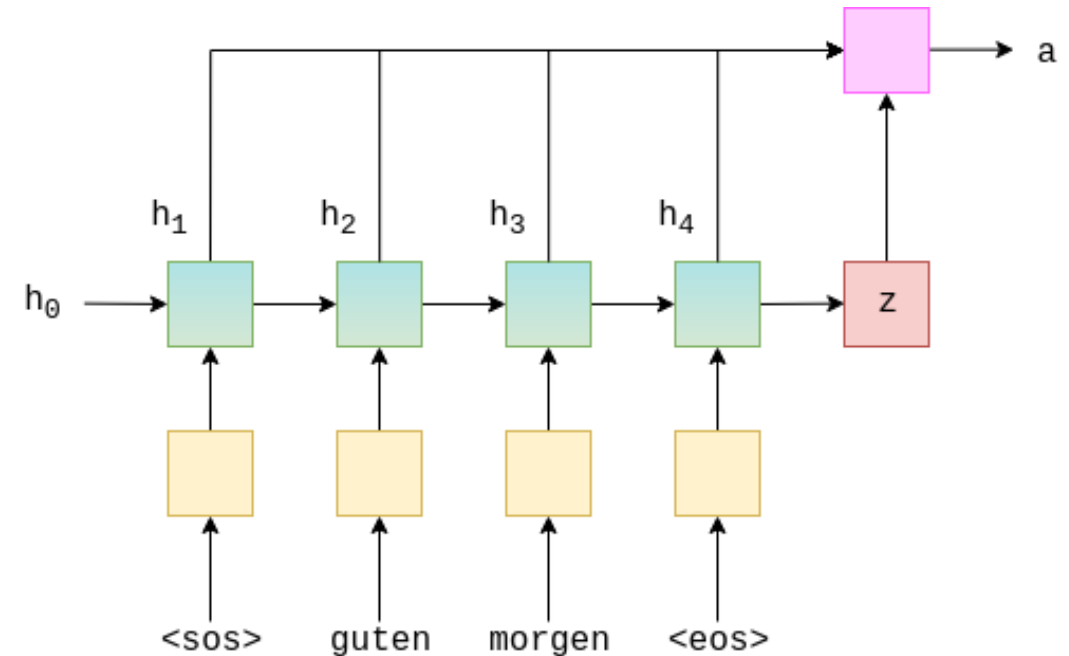
Kodetzailea

- Bidirectional RNN
- GRU
- Horiak embedding
- Berdeak forward
- Urdinak backward
- h egoera ezkutuak
- z kontextua



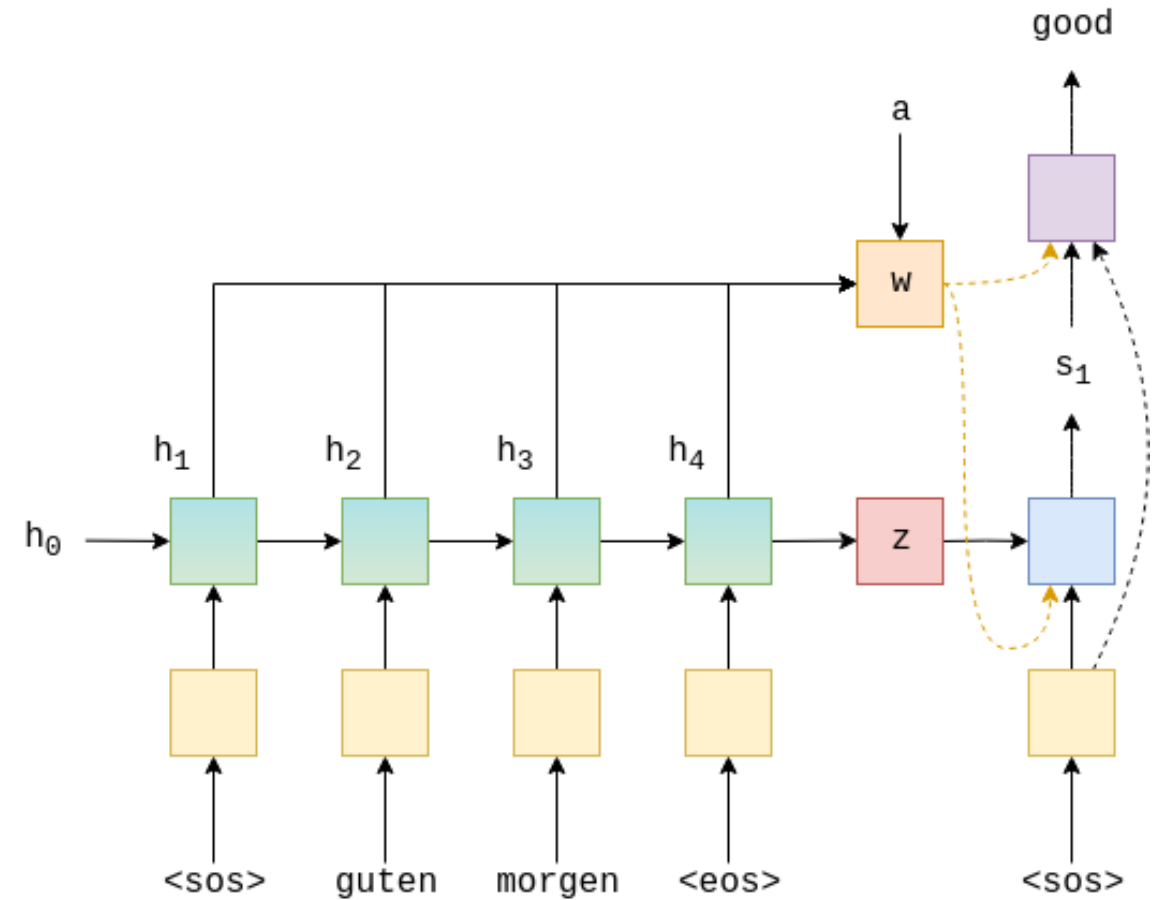
Atentzioa

- Sarrerako hitzetan atentzioa jarri
- Bi sarrera
 - Kodetzailearen irteera
 - Deskodetutako zatia
- Irteera 0 eta 1 arteko probabilitatea
- Irudian lehenengo atentzio bektorea



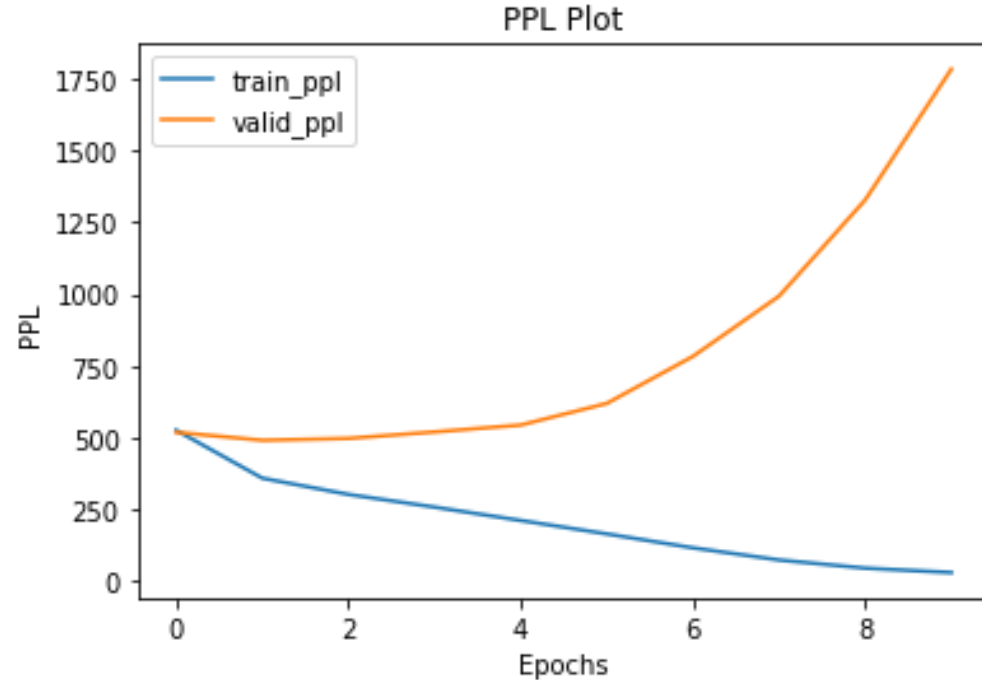
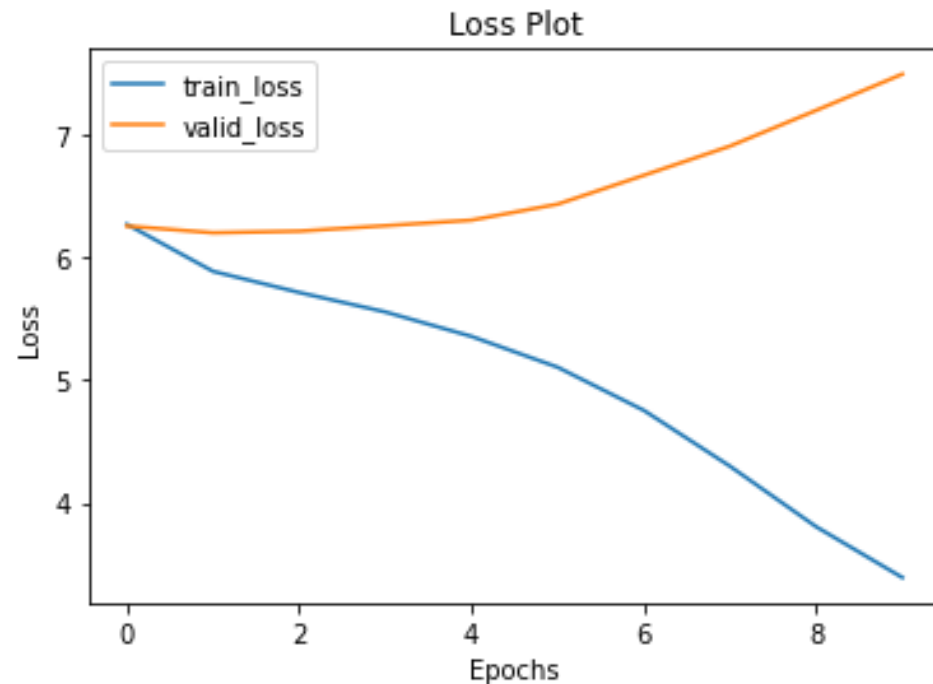
Deskodetzailea

- Hiru sarrera
 - Kodetzailearen kontextua
 - Atentzioa
 - Aurreko hitza
- Morea geruza lineala
- Irudian lehenengo hitzarena



Entrenamendua

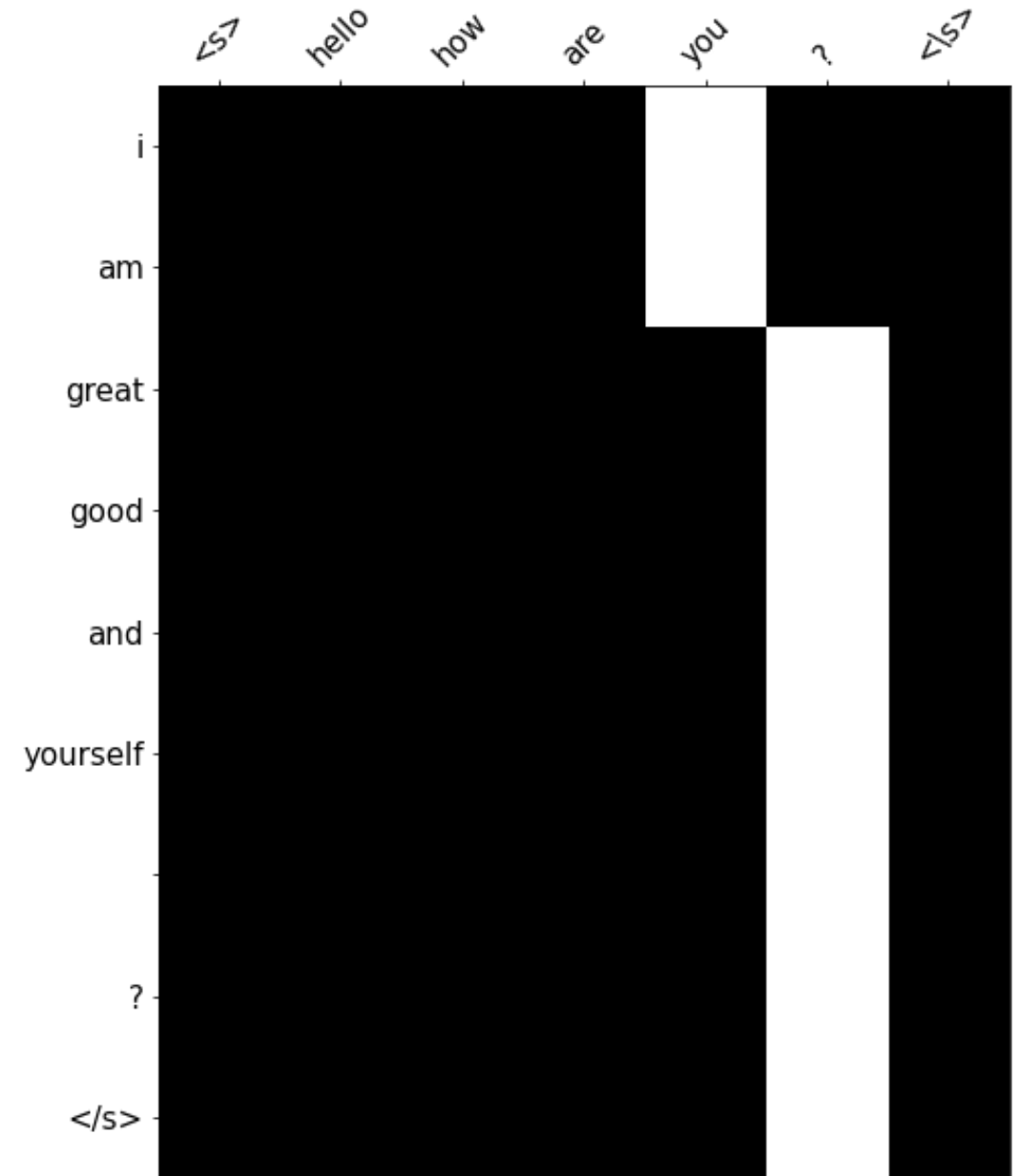
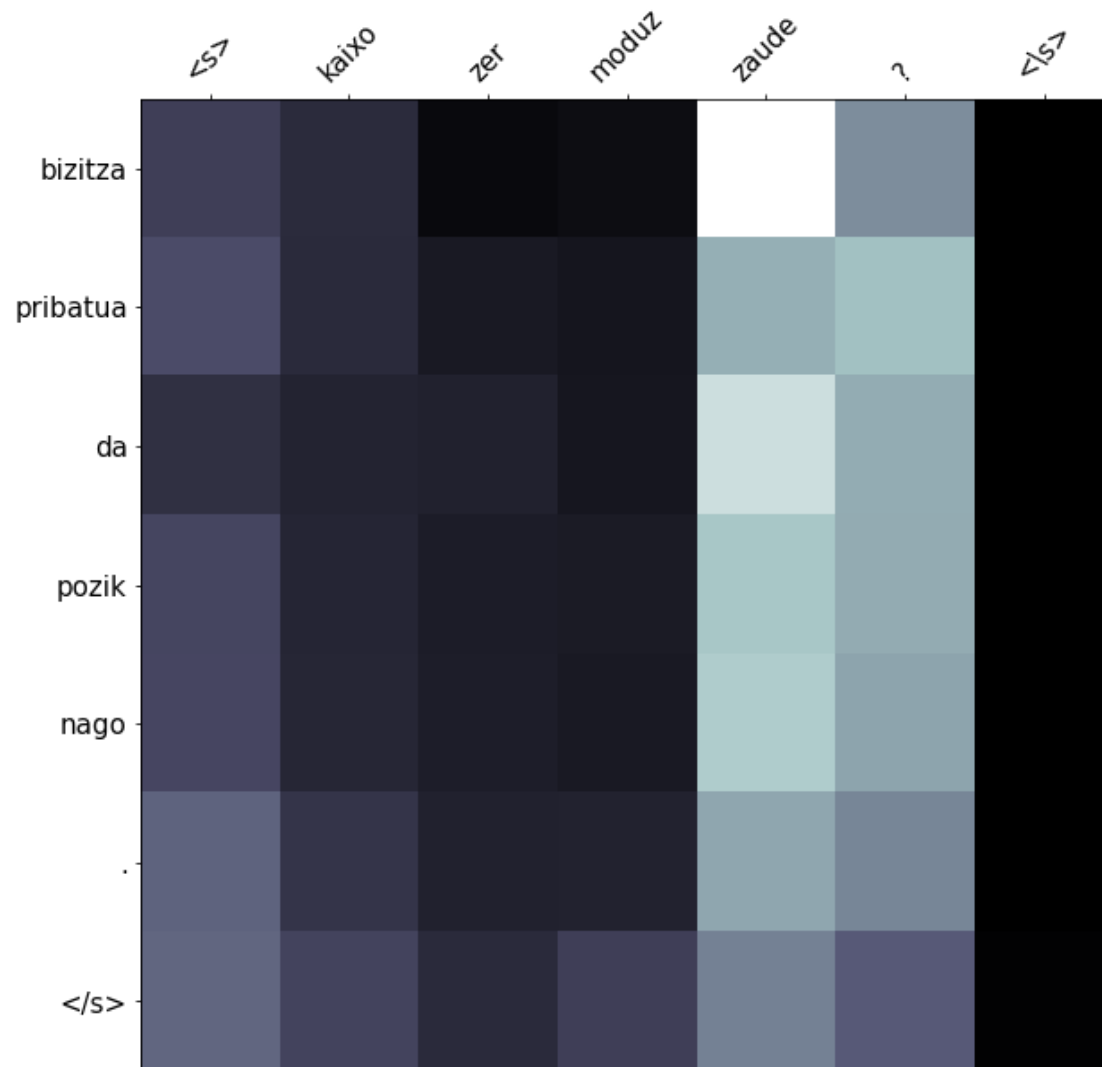
- 50 epoch
- Epoch bakoitza 20 min
- Eredua txikitu
- Overfitting



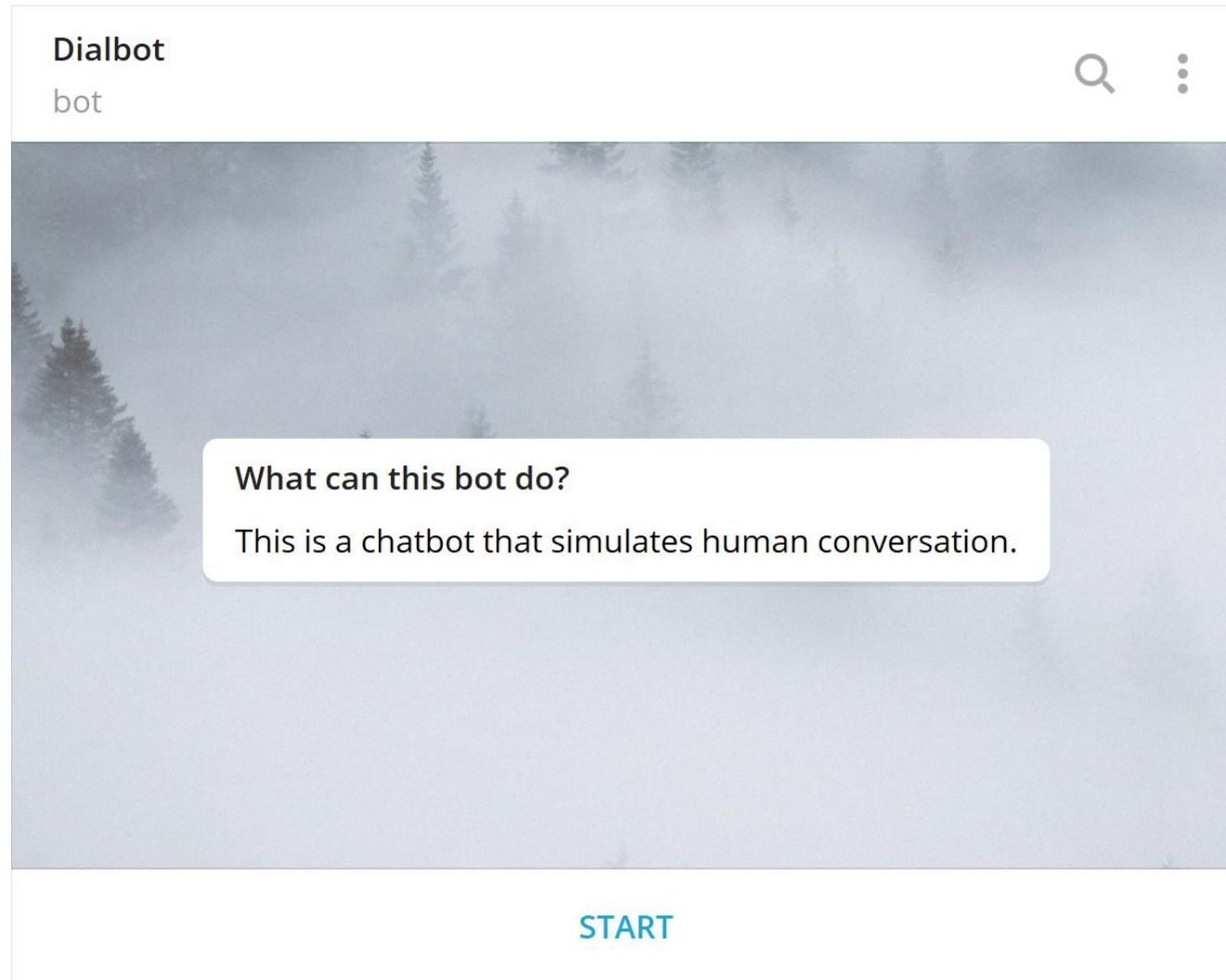
Emaizak

	Loss	PPL	BLEU
Train EN	1.686	5.400	15.01
Train EU	4.678	107.558	0.93
Valid EU	4.675	107.263	1.05
Test EU	4.689	108.763	1.19

Inferentzia



Telegrameko bota



Dialbot

bot



What can this bot do?

This is a chatbot that simulates human conversation.

May 6

/start 13:50 ✓✓

Hi! I'm dialbot. Use [/help](#) to see a list of all comands 🤖. 13:51

Hello 13:54 ✓✓

Hi there ! how you ? 13:54

Good 13:54 ✓✓

What do you are your plans this this this evening ? 13:54

I'm doing some work 13:55 ✓✓

Nice . do you like or or fun ? 13:55



Write a message...



Dialbot

bot



Hi! I'm dialbot. Use [/help](#) to see a list of all comands 🤖. 22:30

/eu 22:30 ✓✓

Hizkuntza aldatu da. 22:30

Kaixo 22:30 ✓✓

Kaixo . 22:30

Zer moduz? 22:30 ✓✓

Ondo . 22:30

Zer duzu gustuko? 22:33 ✓✓

Ez dakit . 22:33



Write a message...



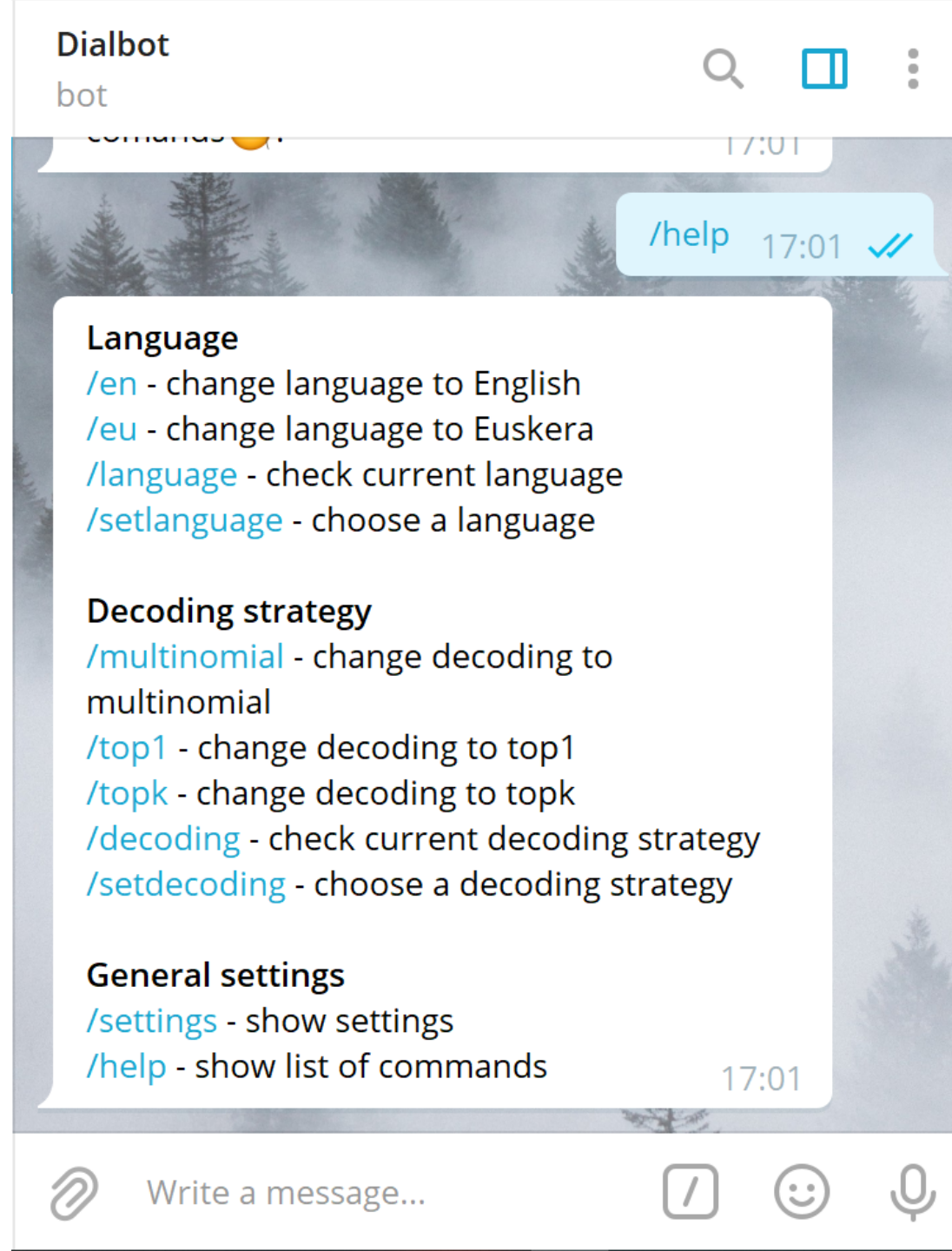
14



Komandoak

Aukera desberdinak:

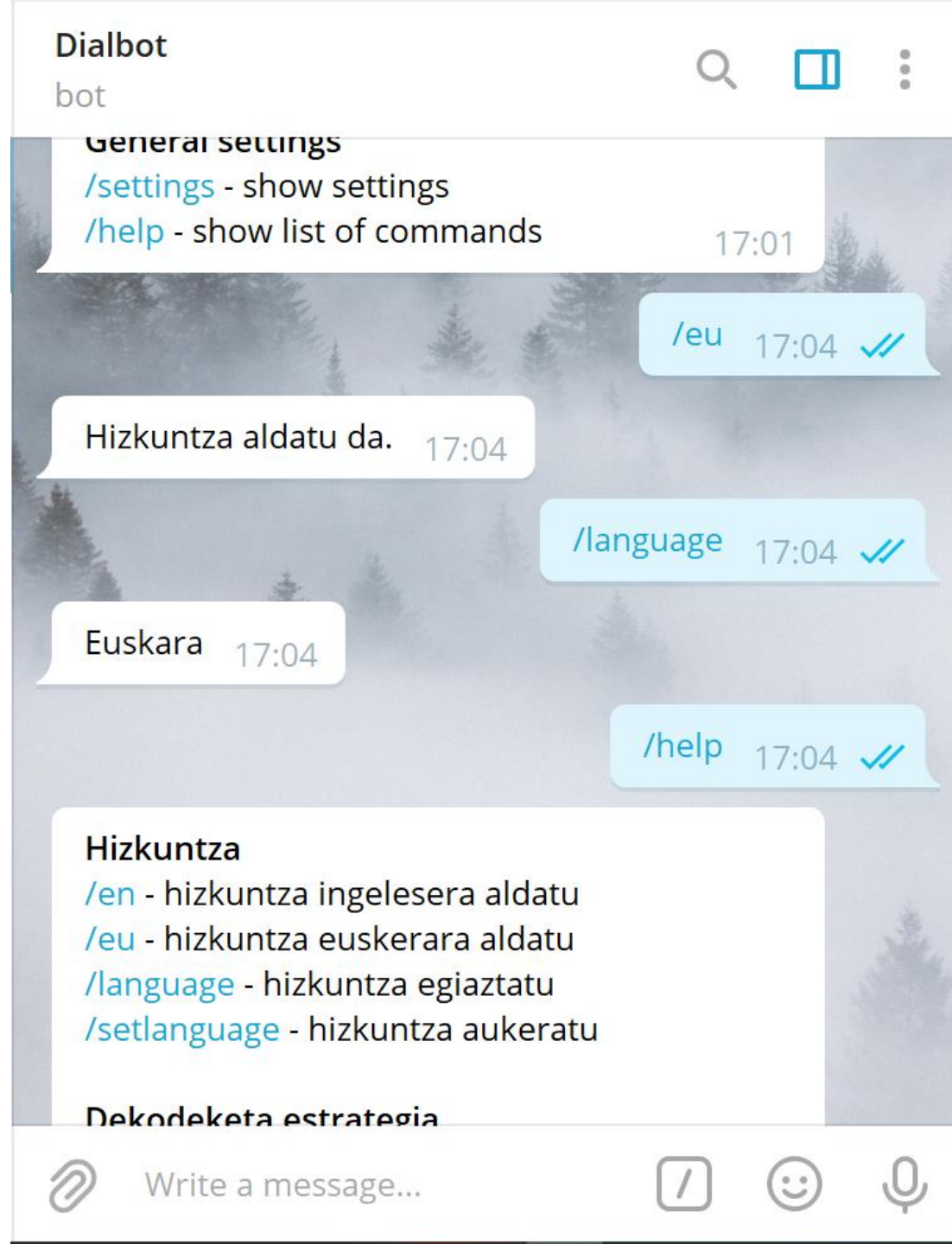
- Hizkuntza
- Dekodeketa estrategia



Komandoak

Hizkuntza:

- **/eu**
- **/en**
- **/setlanguage**
- **/language**



Komandoak

Hizkuntza aldatzeko:

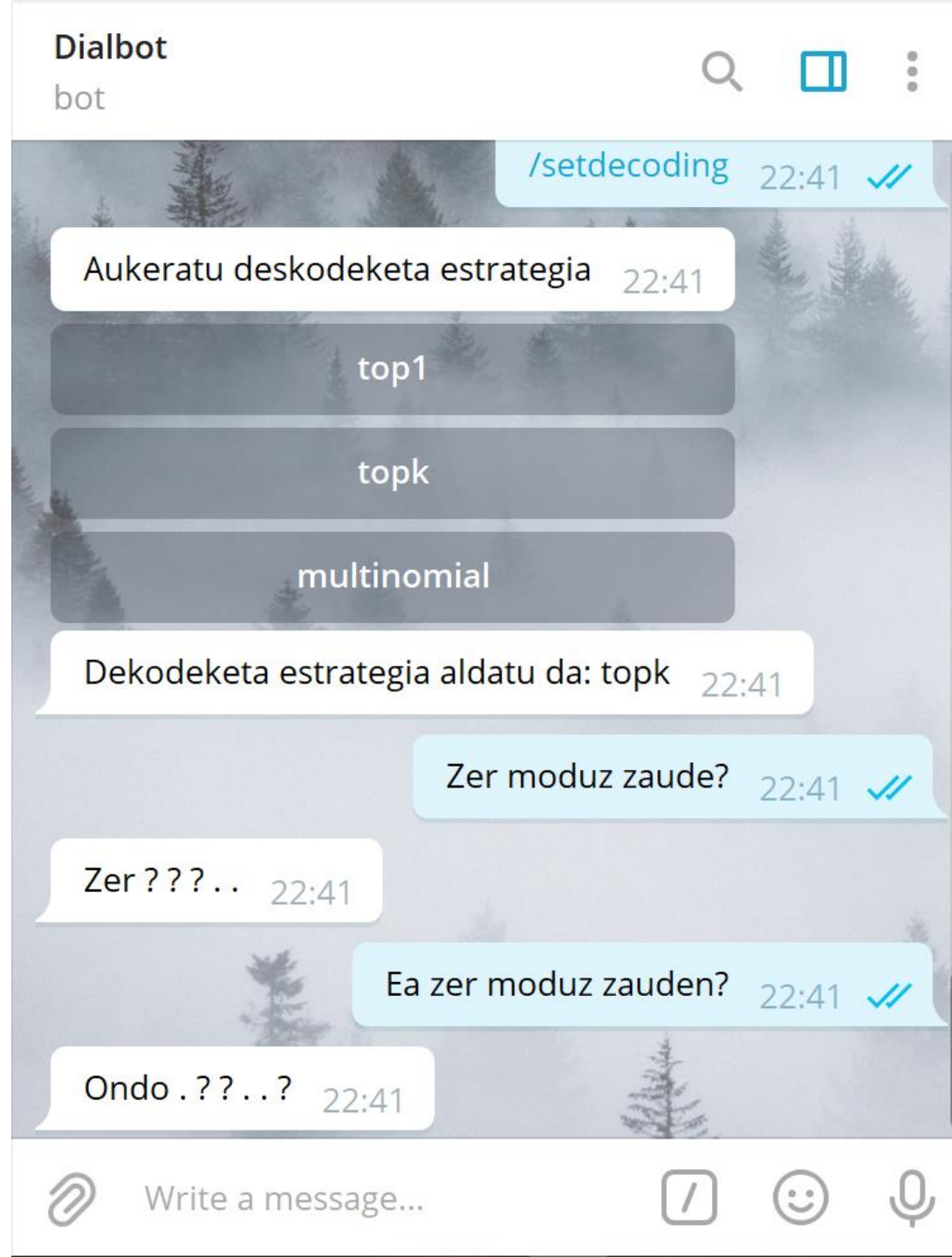
- /eu
- /en
- **/setlanguage**
- **/language**

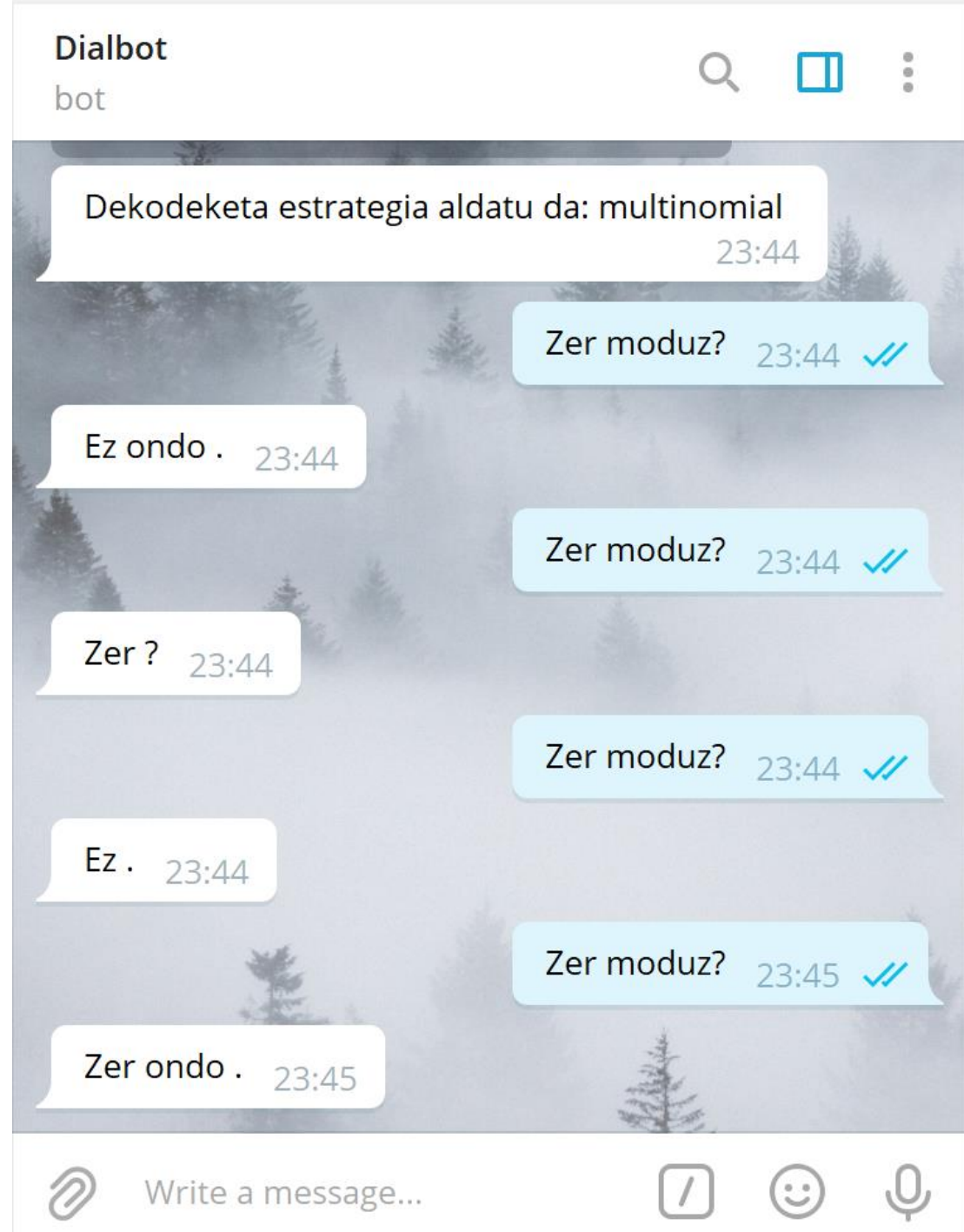
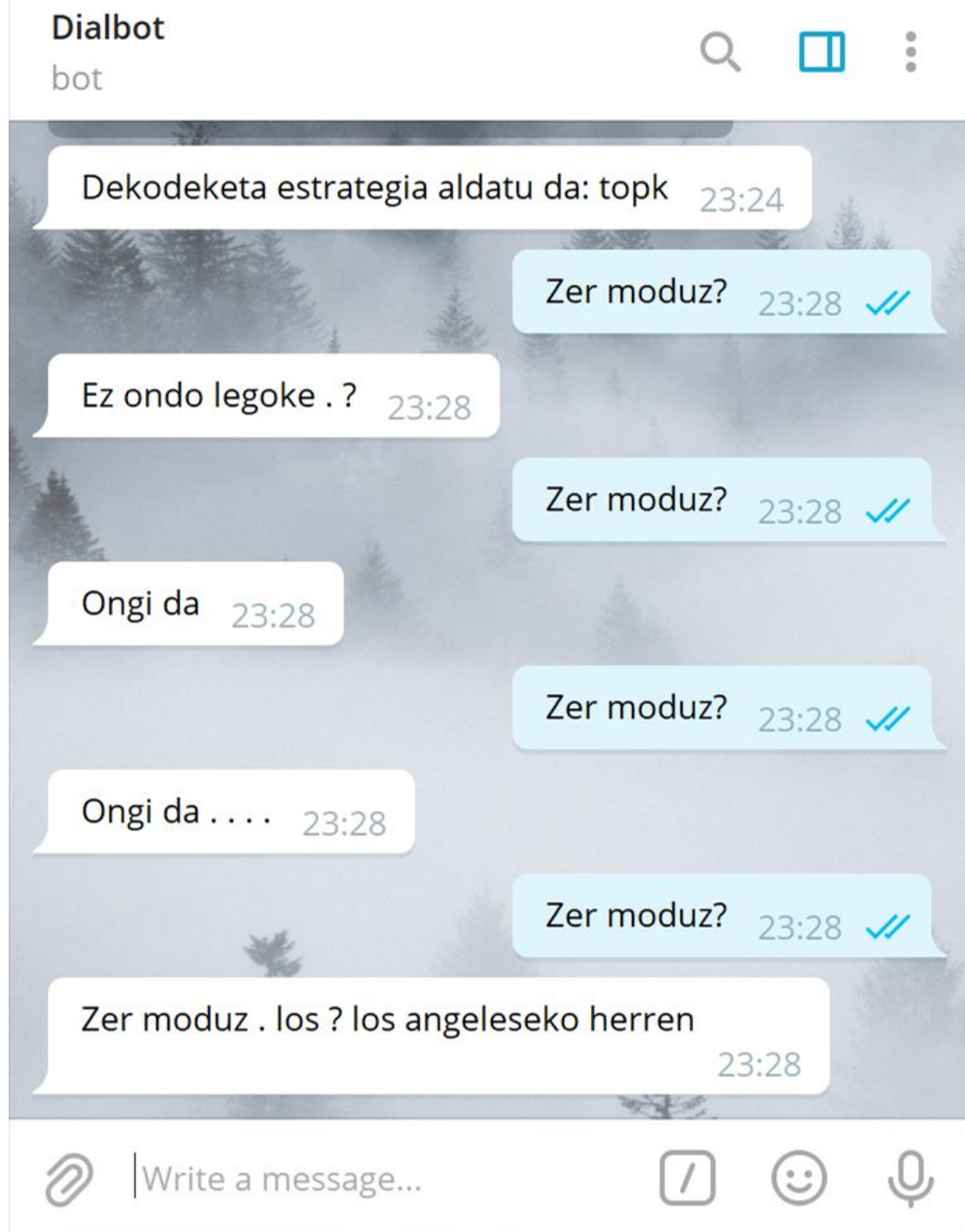


Dekodeketa estrategia

Dekodeketa aldatzeko:

- /top1
- /topk
- /multinomial
- **/setdecoding**
- **/decoding**





Testuingurua

1. Txanda:

Erabiltzailea: Kaixo.

Sistema:

Input --> Kaixo

Output --> Kaixo

2. Txanda:

Erabiltzailea: Zer moduz?

Sistema:

Input --> Kaixo Kaixo Zer
moduz?

Output --> Ondo eta zu?

3. Txanda

Erabiltzailea: Ni ere

Sistema:

Input --> Zer moduz?
Ondo eta zu? Ni ere

Output --> Pozten
naiz.

Eskerrik asko!