

COMPUTATIONAL LINGUISTICS MEETS THE PORTUGUESE SIGN LANGUAGE

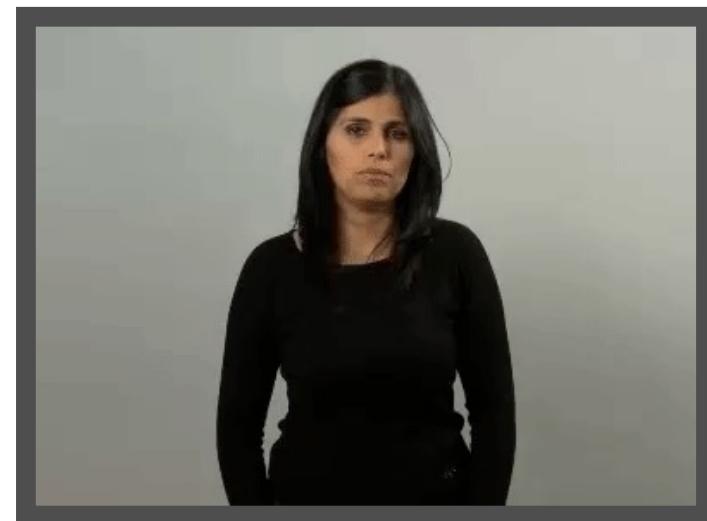
Luísa Coheur

OUTLINE

- Sign languages and the Portuguese Sign Language
- The dream
- The pet project
- The project
- What we have now
- Next?

SIGN LANGUAGES

- Are natural human languages
- Merge manual communication and body language:
 - different hand shapes, orientation and movement of the hands
 - body movements (upper torso)
 - facial expressions



SIGN LANGUAGES



Weekend



Excited



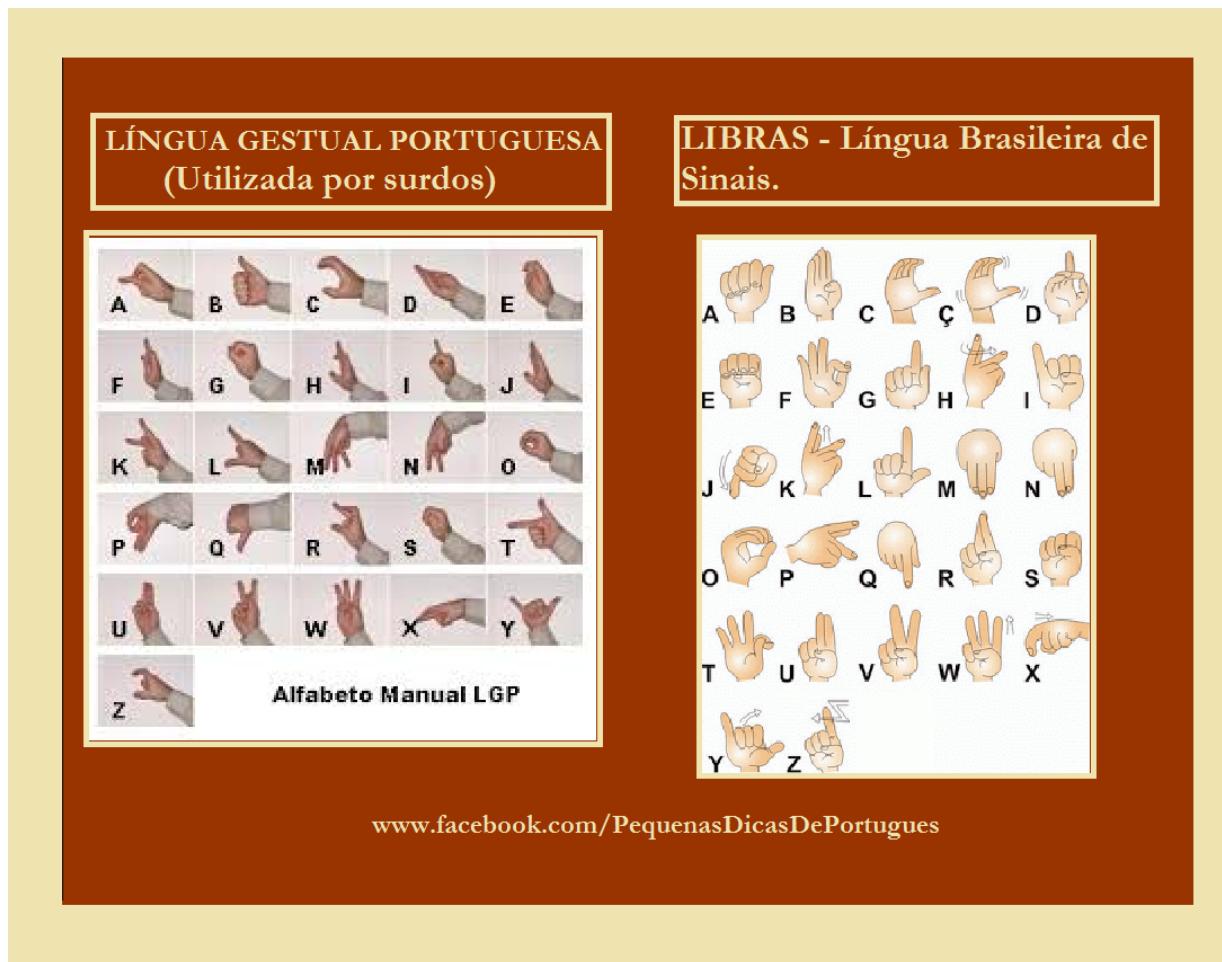
I don't know.

SIGN LANGUAGES

- Have their own vocabulary and grammatical rules
 - American Sign Language and British Sign Language are different and not mutually understandable
 - They are different languages with distinct grammars and vocabularies

SIGN LANGUAGES

- The same for Língua Gestual Portuguesa (LGP) and Libras (BR)

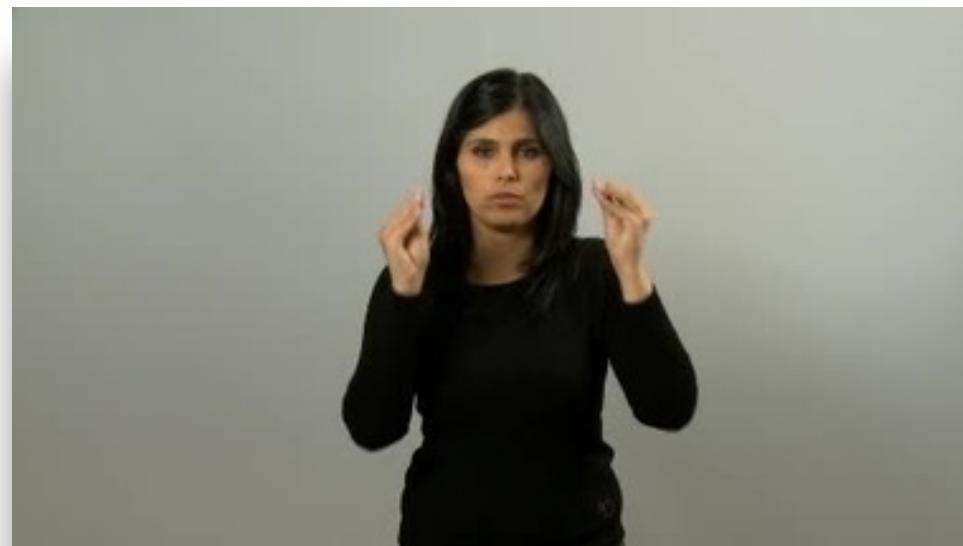
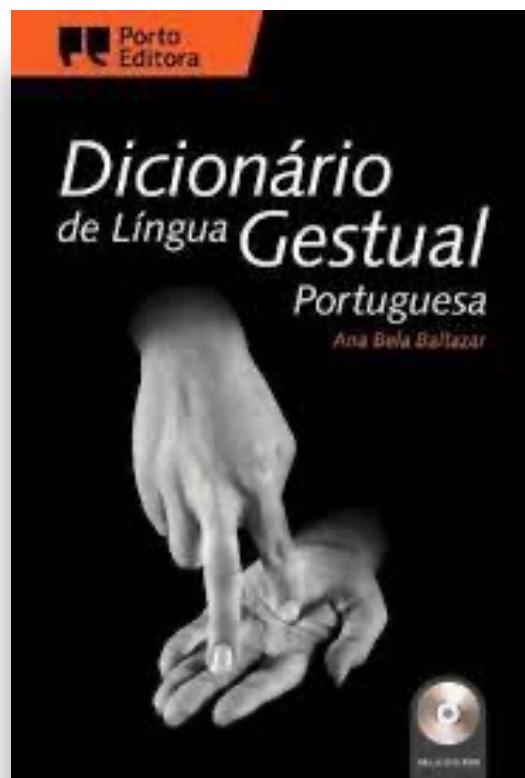


PORTUGUESE SIGN LANGUAGE

- Língua Gestual Portuguesa (LGP)
 - Official recognized since 1997
 - LGP resources:
 - Some books:
 - A “grammar” from 1994
 - Practical guide to LGP
 - ...
 - (Online) dictionary from Porto Editora (infopédia)
 - Online courses: Escola Virtual LGP
 - The Spreadthesign project
 - Master/PhD Thesis

PORTUGUESE SIGN LANGUAGE

- Dicionário da Porto Editora



PORTUGUESE SIGN LANGUAGE

- The Spreadthesign project



table x Search

Words Sentences Locations

table Noun

table Noun

work table Noun

night table Noun

vaulting table Noun

table tools Noun

pedestal table Noun

card table Noun

table 1

mesa 1

Select a language

English (United States) English (United States)

English (United Kingdom) English (United Kingdom)

Spanish (Spain) Spanish (Spain)

French (France) French (France)

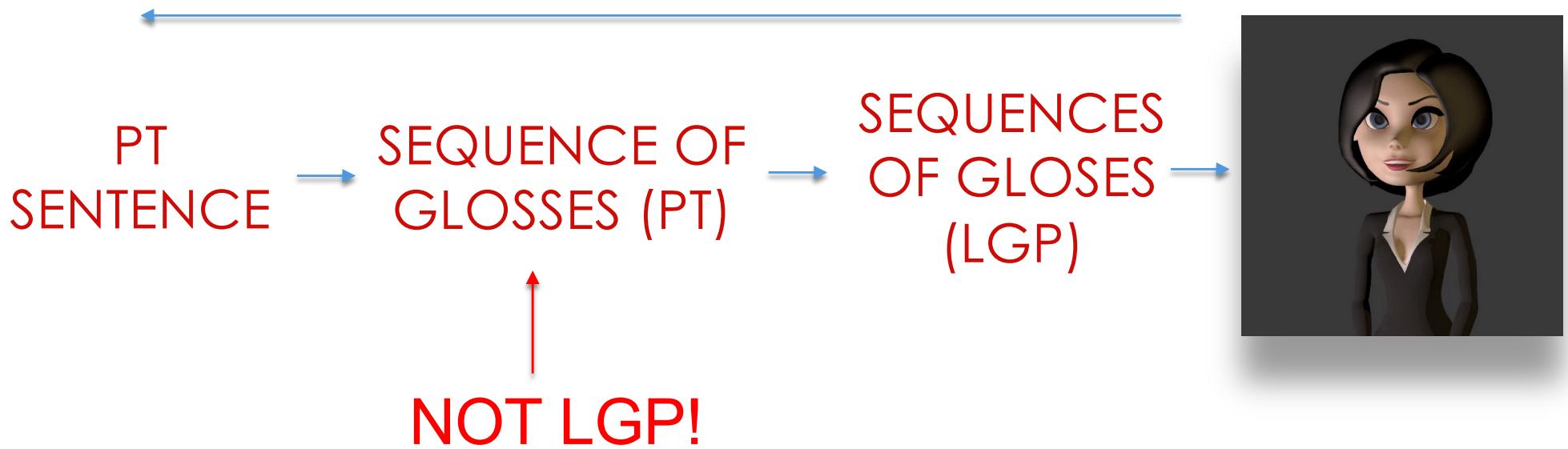
German (Germany) German (Germany)

OUTLINE

- Sign languages and the Portuguese Sign Language
- The dream
- The pet project
- The project
- What we have now
- Next?

THE DREAM

- Develop a free, open-source platform, that translates, in real time, European Portuguese (EP) into LGP (and vice-versa)



THE DREAM

EP	Glosses in LGP
A Ana gosta de massa? (Ana likes pasta?)	{DT(A-N-A) GOSTAR }(q) ({(DT(A-N-A) LIKE) }(q))
A rainha foi à praia. (The queen went to the beach.)	MULHER REI PRAIA IR (WOMAN KING BEACH GO)

We have seen this...

OUTLINE

- Sign languages and the Portuguese Sign Language
- The dream
- [The pet project](#)
- The project
- What we have now
- Next?

2014

THE PET PROJECT

- A EP2LGP translator includes a module for Natural Language Processing (NLP)
- A module for (adding signs to the) avatar



Inês Almeida

THE NLP MODULE – V0

A coelhinha Ana come a cenoura (*little female rabbit Ana eats a carrot*)

Tokenizer + Lemmatizer + PoS tagger

[(a, art), (coelhinha, cn), (ana, pn), (come, v), (a, art),
(cenoura, cn)]

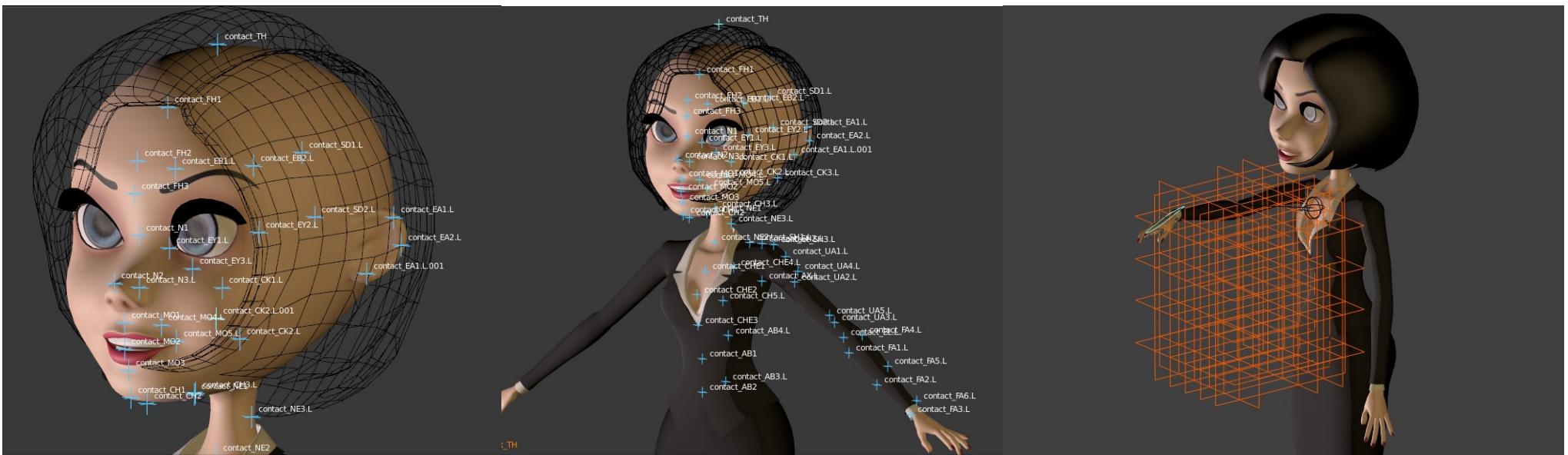
Named Entity Recognition + Lexical Transfer

[(G, coelho + fêmea + pequeno, cn), (FS, ana, pn), (G,
comer, v), (G, cenoura, cn)]

Syntactic Rules

[(G, cenoura, cn), (G, coelho + fêmea + pequeno, cn), (FS,
ana, pp), (G, comer, v)]

A MODULE FOR ADDING SIGNS – V0



blender

2016

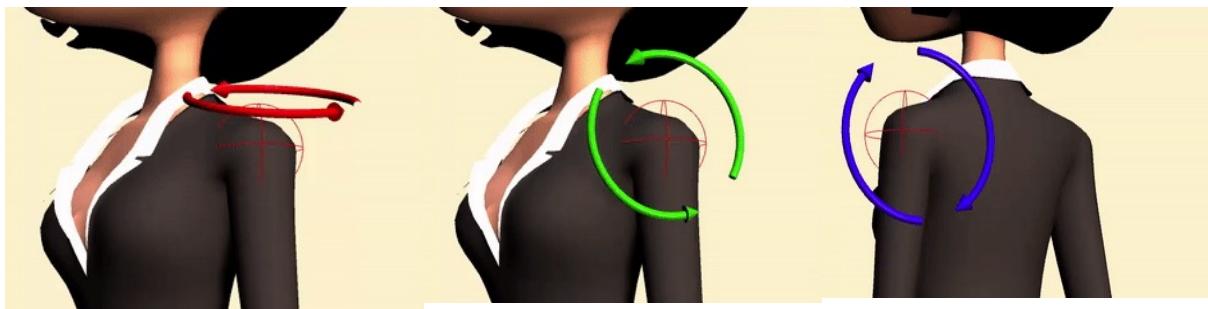
A MODULE FOR ADDING SIGNS – V1



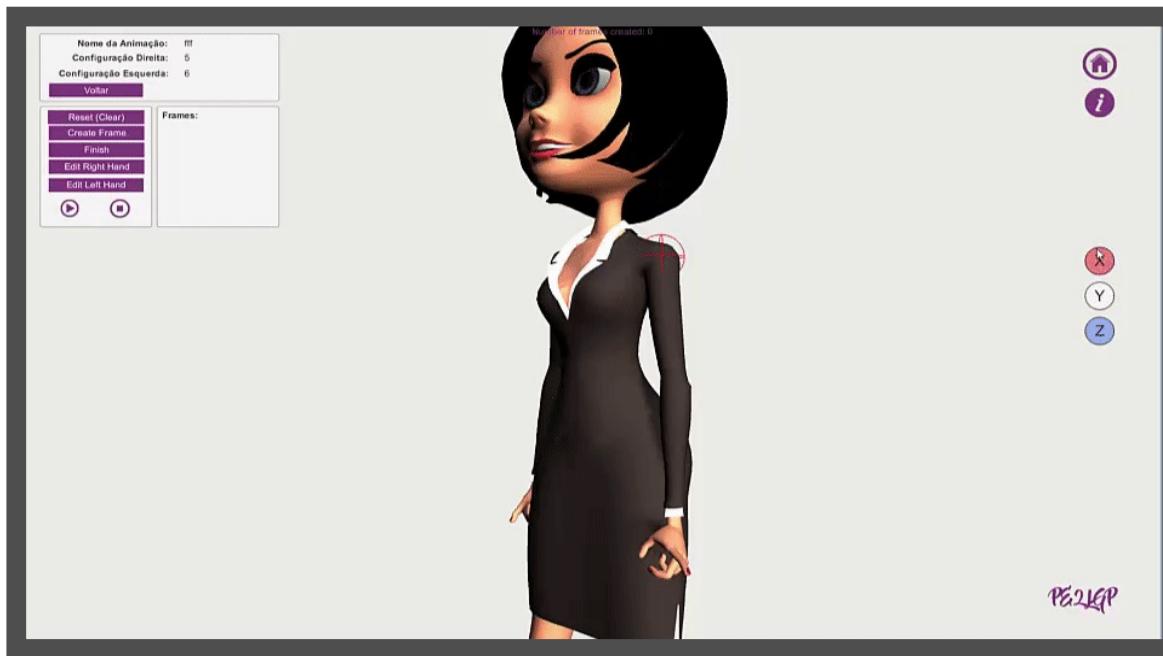
Ruben Santos

TRANSLATING PORTUGUESE TO LGP

- A module for adding signs – v1



A MODULE FOR ADDING SIGNS – V1

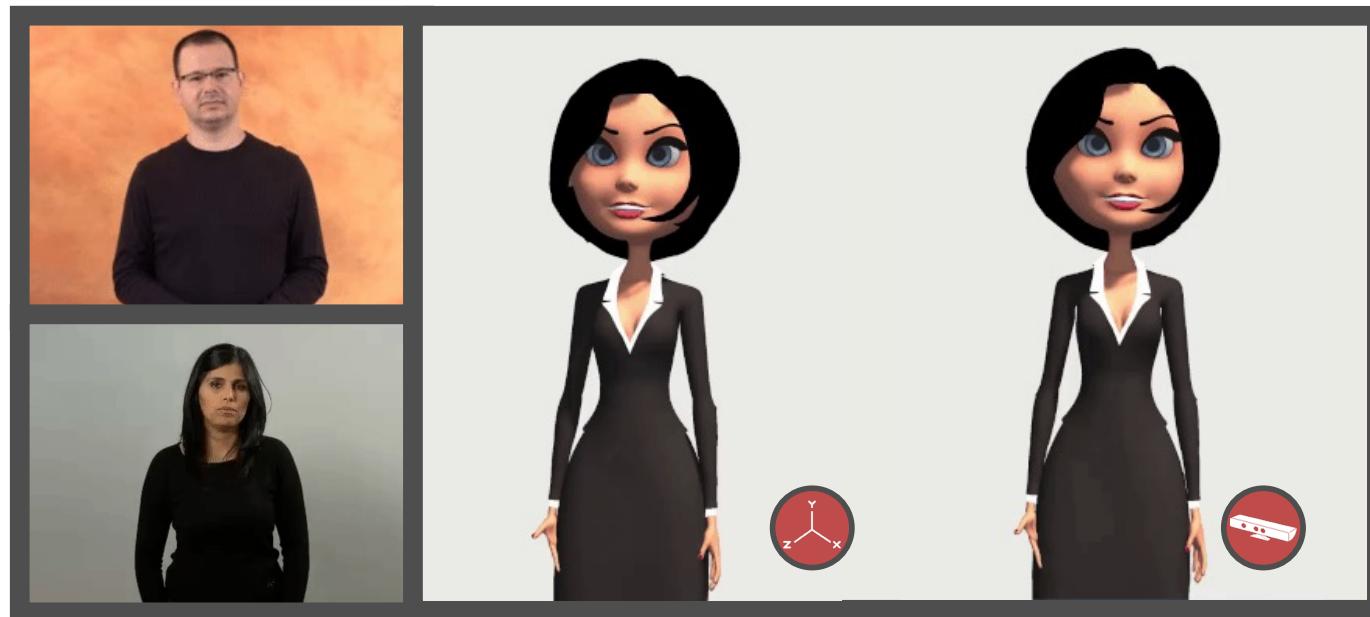


A MODULE FOR ADDING SIGNS – V1

Ruben Santos



EXAMPLE: WORD “BAIXO” (SHORT)



2018

A MODULE FOR ADDING SIGNS – V2

- We decided to explore **phonetic transcription systems**
- Phonetic transcription systems are method of visually representing the sounds of speech or **gestures** using a set of symbols
 - It allows to accurately document and analyze the pronunciation of words or the **specific movements and expressions in sign languages**

A MODULE FOR ADDING SIGNS – V2

HamNoSys

$\overline{\text{ə}} \text{r} \text{o} \square \cdot ([\rightarrow \rightarrow \overline{\text{o}}])$
 $[\overline{\text{o}} \text{i} \text{n} \text{o}, \text{o} \text{g} \text{o}] \sim \chi \xrightarrow{\circ} +$
 $\sim [\overline{\text{o}} \text{i} \text{n} \text{o}, \text{o} \text{h} \text{o}] \sim (\leftarrow \chi)$
 $[\text{y} \text{i} \text{n} \text{o}, \overline{\text{e}} \text{g} \text{o}] [\nwarrow \rightarrow \overline{\text{o}} \text{n} \text{o}] \sim \chi$
 $\overline{\text{d}} \text{i} \text{n} \text{o}, \overline{\text{q}} \text{h} \text{o}, \overline{\text{g}} \text{h} \text{o}$



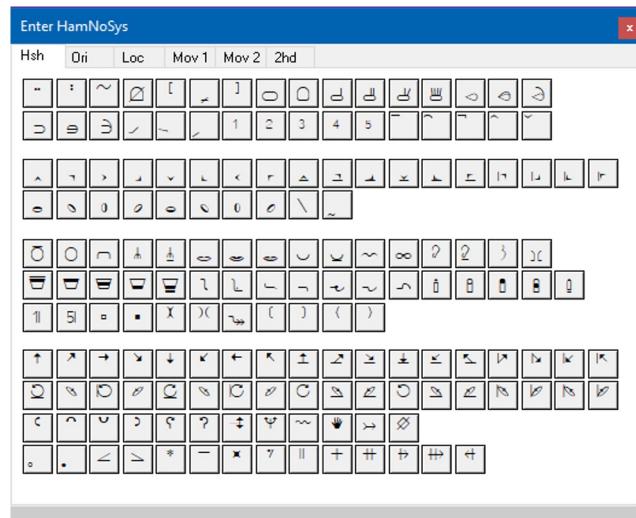
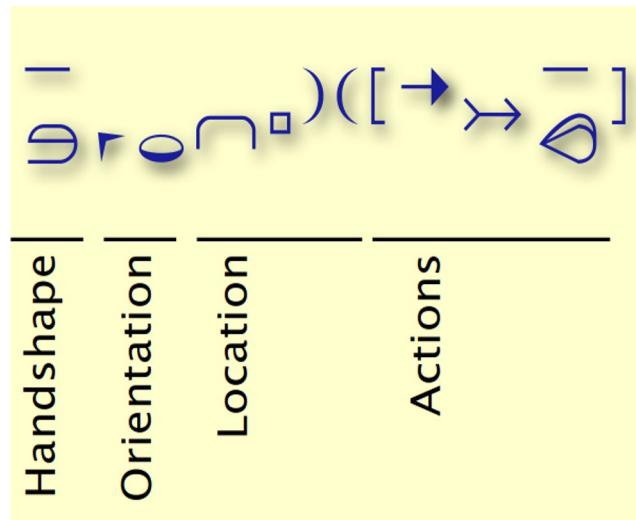
Stokoe (Wikipedia)

B_a B_a z~ N N a· 3¹ [] √C[†] √C[×] 3 Y⁰ √G_A < v <
 B_a √B_A w G¹ B_A † B_A v D A^{⊗x} B_v B_v 1
 G[>] ^ 5 x [] √C[†] √C[×] X₁ X₁ a B_T V_D v ·
 B_a L #· X₁ X₁ a

From Wikipedia

A MODULE FOR ADDING SIGNS – V2

Rui Ferreira



Avatar: marc

Speed (log₂ scale): +0.0

SiGML URL: scotland-H.sigml

SiGML Text:

```
<?xml version="1.0" encoding="utf-8"?>
<sigml>

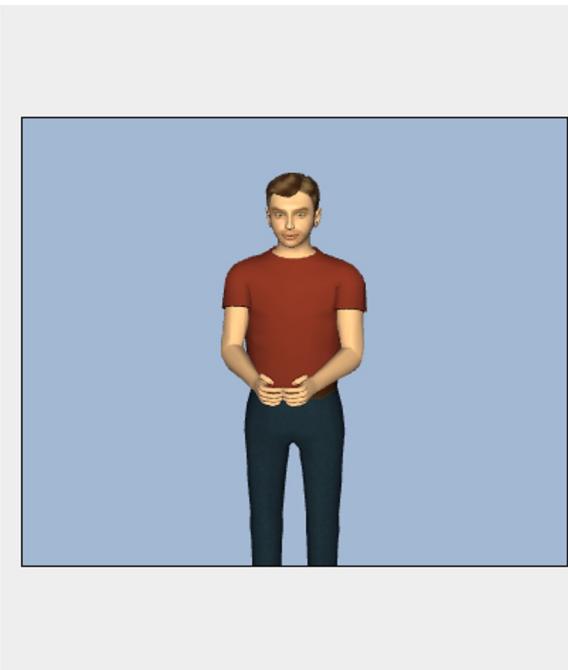
<hng_sign gloss="mug">
  <ham_nomanual>
    <ham_mouthpicture picture="mVg"/>
  </ham_nomanual>
  <ham_manual>
    <hamfist/> <hamthumbacrossmod/>
    <hamextfingerul/> <hampalm/>
  </hamManuals/>
  <hamparbegin/> <hammoveu/> <hamarcu/>
  <hamreplace/> <hamextfingerul/> <hampalmrl/>
  <hamparend/>
</hng_sign>
```

Play SiGML URL Play SiGML Text Stop

Suspend Resume Frames: -1 +1

Sign/Frame: 0/0 Gloss: [none] FPS: 25.00

Status: Client: Animgen by Client



<http://vh.cmp.uea.ac.uk/index.php/JASigning>

OUTLINE

- Sign languages and the Portuguese Sign Language
- The dream
- The pet project
- The project
- What we have now
- Next?

2018-2021

PROJECT: CORPUS & AVATAR DA LGP



UNIVERSIDADE
CATÓLICA
PORTUGUESA



Fundação
para a Ciência
e a Tecnologia

PROJECT CORPUS & AVATAR LGP

- Goal:
 - Create the first LGP corpus
 - Develop a EP2LGP, being the Natural Language Processing module based on the corpus

PROJECT CORPUS & AVATAR LGP



Matilde Gonçalves

Carolina Neves

Inês Lacerda



Pedro Cabral Prof. Hugo Nicolau



Profª Ana Mineiro + amazing team from Católica

PROJECT CORPUS & AVATAR LGP

Arquivo Editar Anotação Trilha Tipo Buscar Visualizar Opções Janela Ajuda

Grade Texto Legenda Lexicon Comments Reconhecedores Metadados Controles

The screenshot shows the Project Corpus & Avatar LGP software interface. At the top, there's a menu bar with options like Arquivo, Editar, Anotação, Trilha, Tipo, Buscar, Visualizar, Opções, Janela, and Ajuda. Below the menu is a toolbar with buttons for Grade, Texto, Legenda, Lexicon, Comments, Reconhecedores, Metadados, and Controles. A video player window on the left shows a woman in a blue t-shirt with a heart graphic, standing in front of a chalkboard. To the right of the video are volume and speed controls. The volume is set to 100, and the speed is set to 70. Below these controls is a timeline with a selection range from 00:00:21.331 to 00:00:22.804 and a frame count of 1473. The timeline displays several transcription tracks. From left to right, the tracks are: LP_P1 transcrição livre [37], Come_P1Literal [29], LGP_P1Trans_Literal [101], GLOSAS_P1 [456], GLOSA_P1-M1 [451], GLOSA_P1-M2 [153], M2_ClassGram [144], and M1_ClassGram. The right side of the timeline shows the corresponding text for each track: Cultura, Arte, Teatro., CULTURA ARTE TEATRO, CULTURA, and CULTURA.

We have seen this...

PROJECT CORPUS & AVATAR LGP



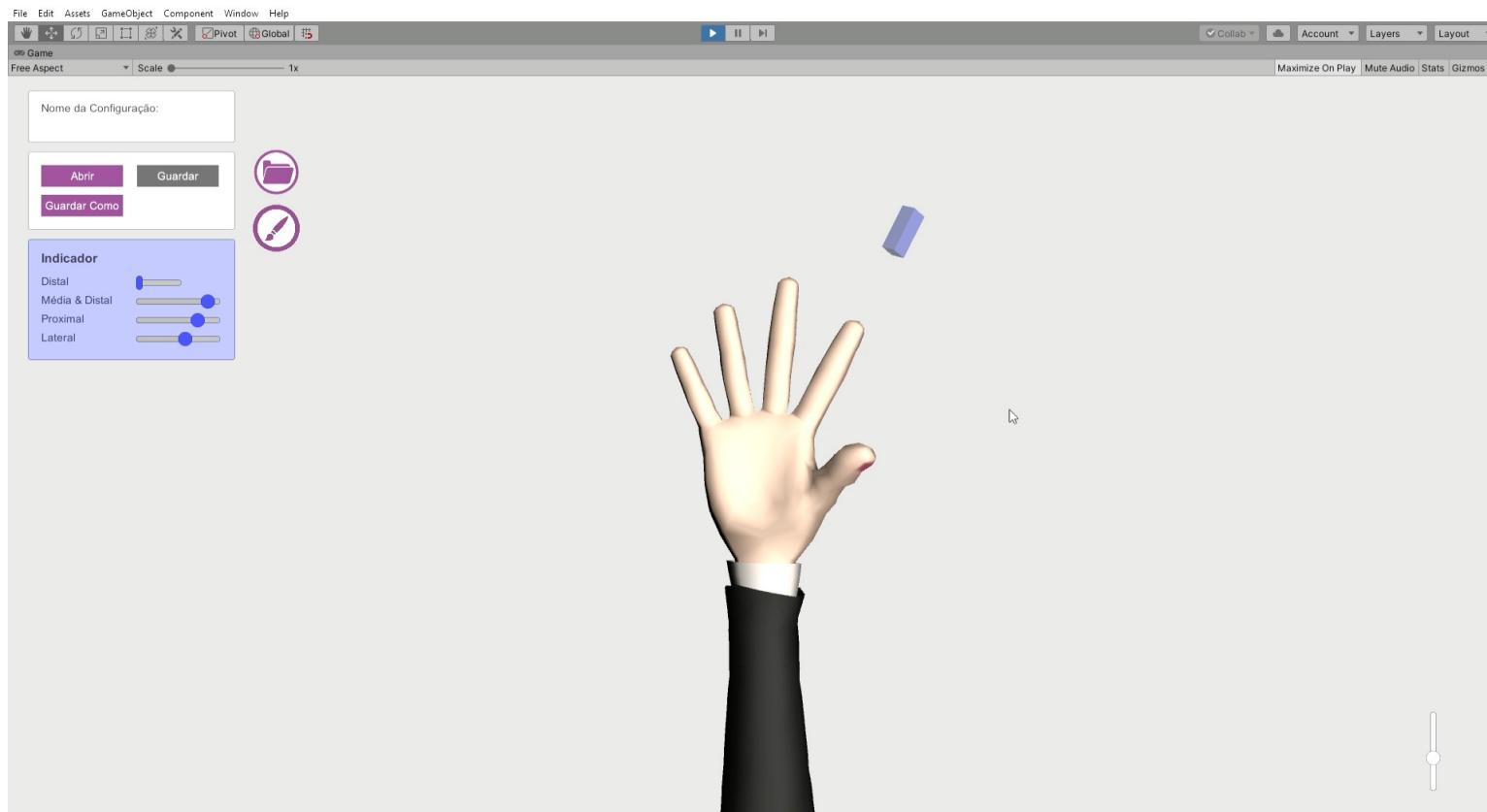
OUTLINE

- Sign languages and the Portuguese Sign Language
- The dream
- The pet project
- The project
- What we have now
- Next?

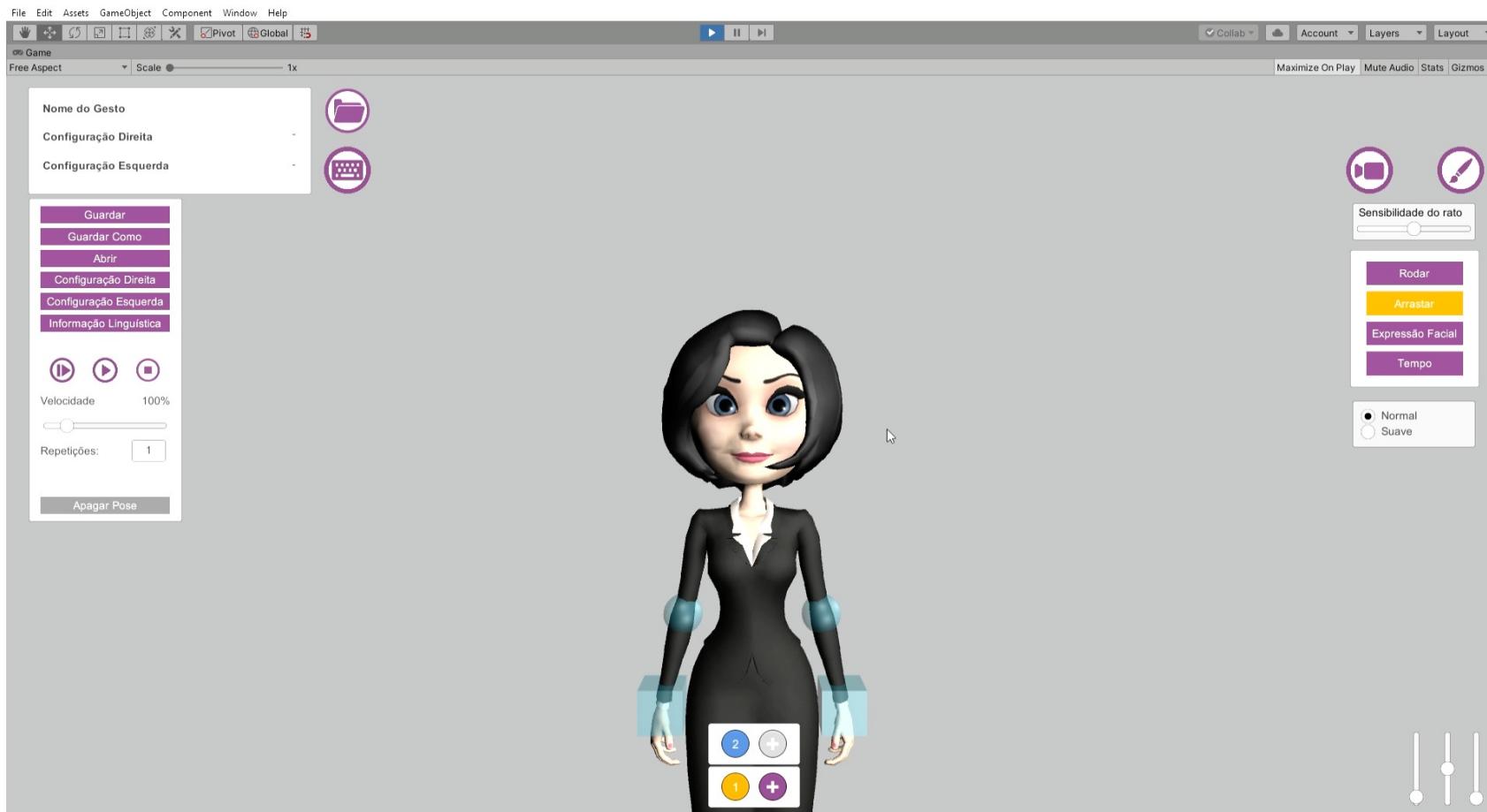
WHAT WE HAVE NOW

- First LGP corpus
 - 110 hours of video, 5 hours annotated at several levels (lexical, syntactic, ...)
 - Part was used to feed the new grammar
- 1000 signs

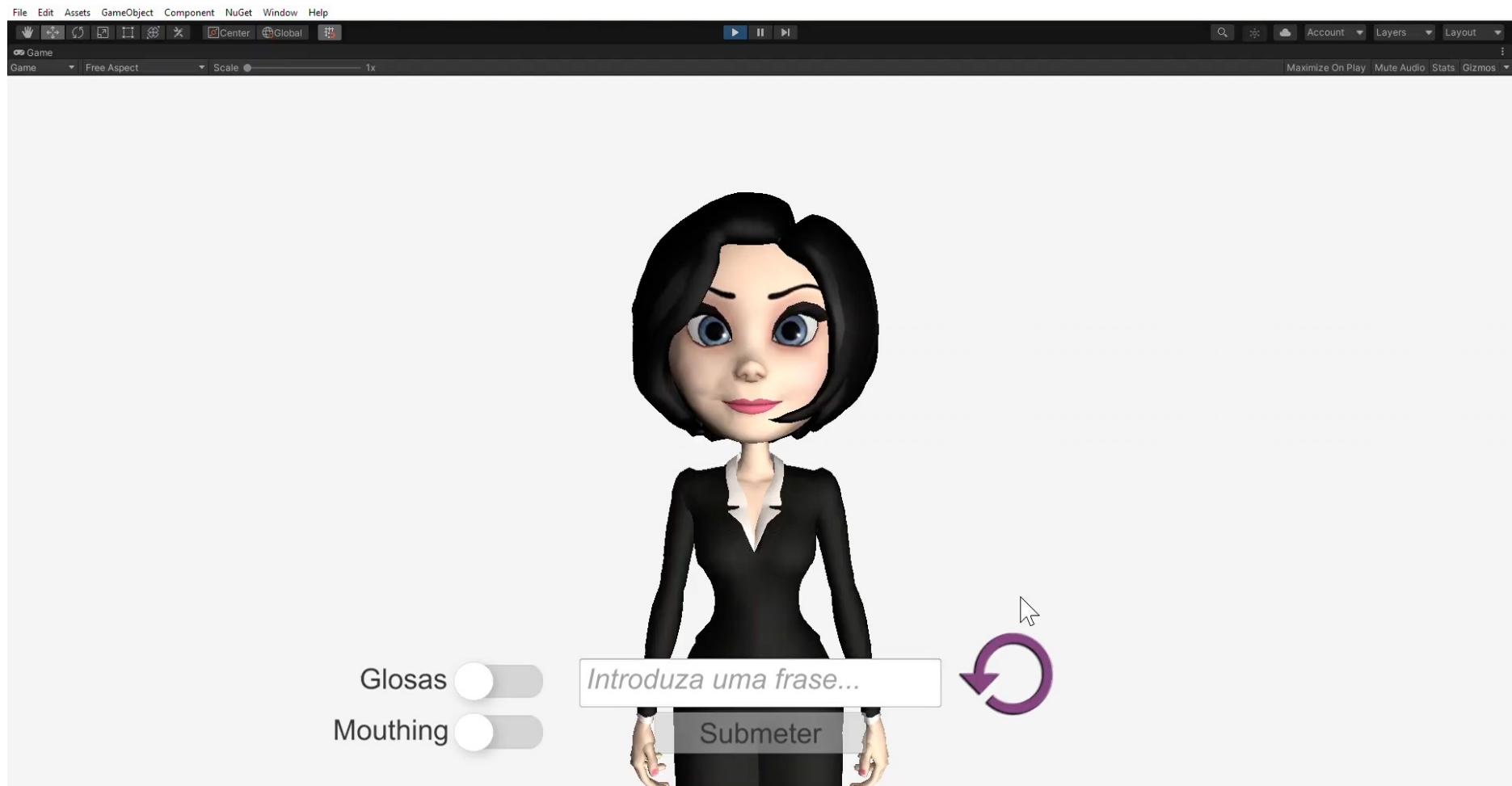
WHAT WE HAVE NOW



WHAT WE HAVE NOW



WHAT WE HAVE NOW



2023

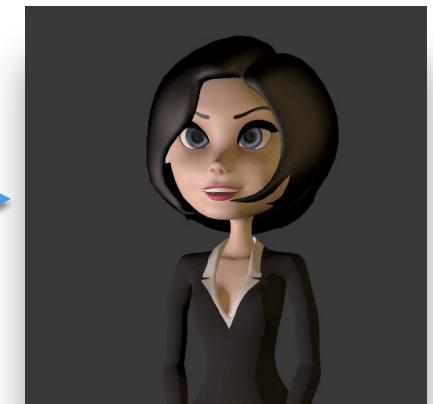
WHAT WE HAVE NOW

Catarina Sousa's Master thesis (2023)

- Co-supervised by Mara Moita (Católica)
- We finally moved to deep learning



Catarina Sousa



- RULE-BASED TRANSLATIONS WITH RULES EXTRACTED FROM CATOLICA'S CORPUS
- RULES USED TO TRANSLATE CORPORA
- DEEP LEARNING

We have seen this...

WHAT WE HAVE NOW

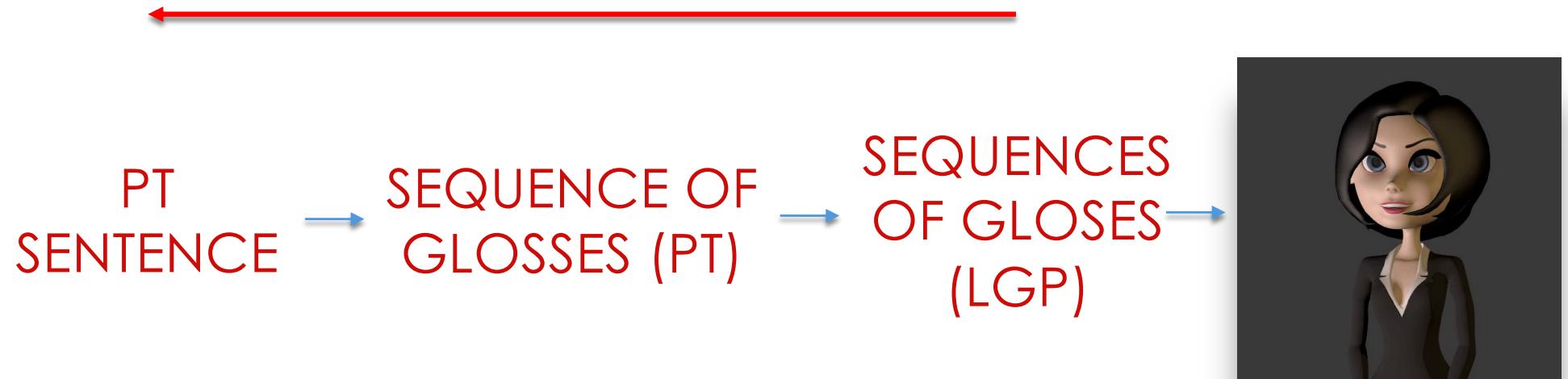
RB	M2M	Reference
NÓS COMER TUDO ISSO ACREDI-TAR CONSEGUIR (NÃO) (WE EAT EVERYTHING THAT BE-LIEVE CAN (NO))	COMER TUDO ISSO ACREDITAR EU CONSEGUIR (NÃO) (EAT EVERYTHING THAT BE-LIEVE I CAN (NO))	NÓS COMER TUDO ISSO EU ACREDITAR CONSEGUIR (NÃO) (WE EAT EVERYTHING THAT I BE-LIEVE CAN (NO))
TAL DEUS MEU PERMITIR (NÃO) (SUCH GOD MY ALLOW (NO))	TAL DEUS MEU TU PERMITIR (NÃO) (SUCH GOD MY YOU ALLOW (NO))	MEU DEUS TAL PERMITIR (NÃO) (MY GOD SUCH ALLOW (NO))
ONDE DT(T-O-M) EU MORAR LEMBRAR CONSEGUIR (NÃO) (WHERE DT(T-O-M) I LIVE RE-MEMBER CAN (NO))	ONDE DT(T-O-M) EU LEMBRAR MORAR CONSEGUIR (NÃO) (WHERE DT(T-O-M) I REMEMBER LIVE CAN (NO))	DT(T-O-M) ONDE MORAR EU LEMBRAR CONSEGUIR (NÃO) (DT(T-O-M) WHERE LIVE I RE-MEMBER CAN (NO))

Table 5: Example sentences where our rule-based model performs better than the fine-tuned M2M model when compared with the reference translation.

2024

WHAT WE HAVE NOW

Inês Pereirinha

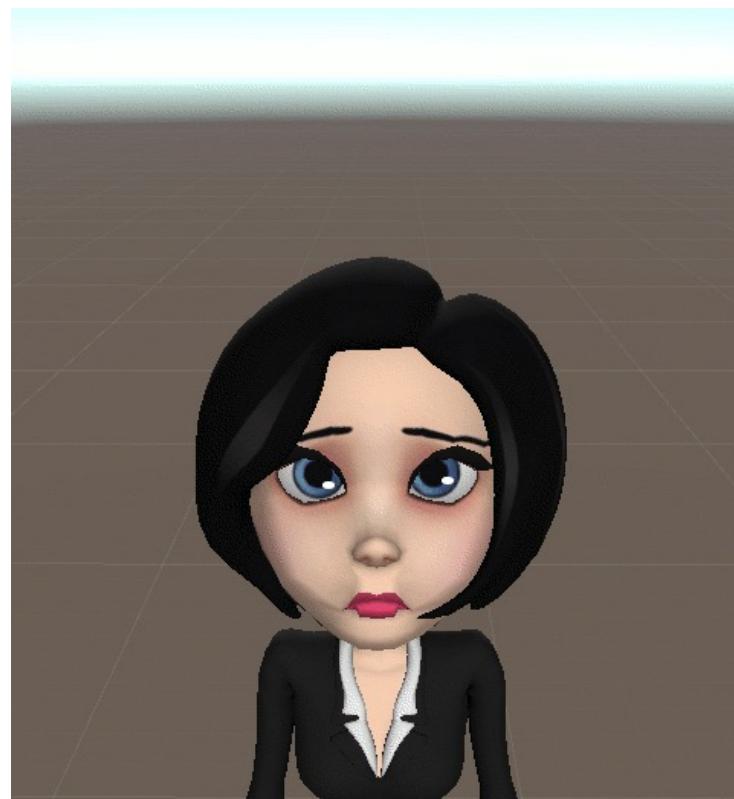


OUTLINE

- Sign languages and the Portuguese Sign Language
- The dream
- The pet project
- The project
- What we have now
- Next?

NEXT

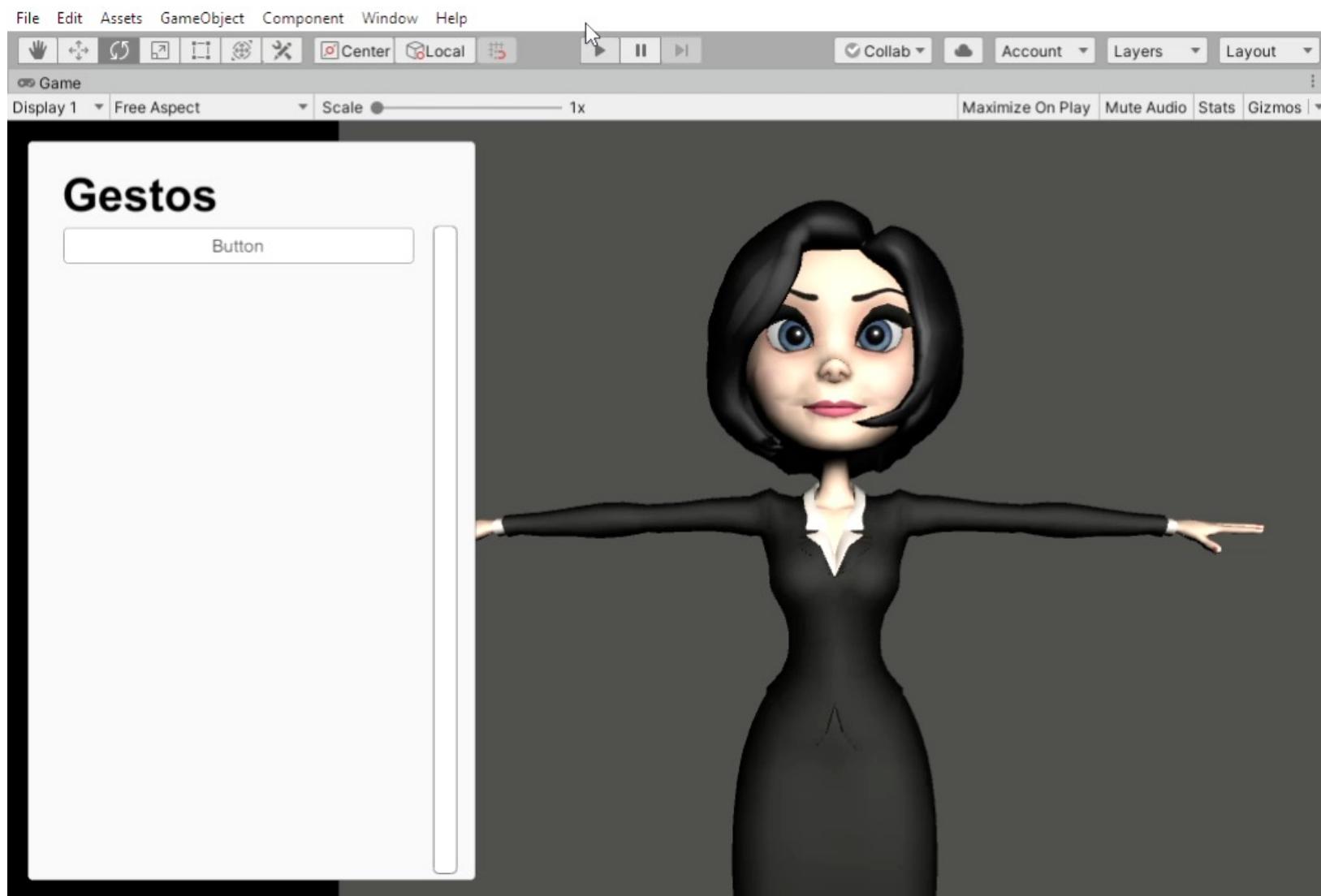
Another project? (already two projects rejected)



NEXT

- LGP grammar and the avatar still need to be improved (obviously)
- The crowd is probably needed to insert more signs in the database
 - how can we validate the signs?
- New ***free*** avatar
- Signs search engine (next, v0)

BLOOPERS



PUBLICATIONS

- Towards Realistic Sign Language Animations. Inês Lacerda, Hugo Nicolau and Luísa Coheur. 23rd ACM International Conference on Intelligent Virtual Agents. 2023
- Enhancing Accessible Communication: from European Portuguese to Portuguese Sign Language. Catarina Sousa, Luísa Coheur, Mara Moita. The 2023 Conference on Empirical Methods in Natural Language Processing (NLP -- Findings). 2023.
- Avaliação de recursos computacionais para o Português. Matilde Gonçalves, Luísa Coheur, Jorge Baptista e Ana Mineiro. *Linguamática*, 12(2), 51-68, 2021
- PE2LGP: tradutor de português europeu para língua gestual portuguesa em glosas. Matilde Gonçalves, Luísa Coheur, Hugo Nicolau, e Ana Mineiro. *Linguamática*, 13(1), 3-21. 2021
- HamNoSys2SiGML: Translating HamNoSys Into SiGML. Carolina Neves, Luísa Coheur e Hugo Nicolau. Proceedings of The 12th Language Resources and Evaluation Conference, 6035–6039, 2020
- PE2LGP Animator: A Tool to Animate a Portuguese Sign Language Avatar. Pedro Bertrand Cabral, Matilde Gonçalves, Ruben dos Santos, Hugo Nicolau e Luísa Coheur. Proceedings of the 9th Workshop on the Representation and Processing of Sign Languages, 33–38, 2020

KEY TAKEAWAYS

KEY TAKEWAYS

- Sign languages are also natural languages
- Sign languages have their own idiosyncrasies, including specific grammars
- There are not much datasets for training sign languages