



A Persona-Based Corpus in the Diabetes Self-Care Domain - Applying a Human-Centered Approach to a Low-Resource Context

U F *m* G



Rossana Cunha, PhD

Thiago Castro Ferreira, PhD

Adriana Pagano, PhD

Fabio Alves, PhD

{rossanacunha, thiagocf05, apagano, fabio-alves}@ufmg.br

Outline

I	Background & Motivation	4
II	Hypotheses Development	7
III	Methodology	12
IV	Development	17
V	Conclusion and Discussions	25



**ROSSANA CUNHA,
PhD**

Software Engineer and
NLP Researcher at
CIASC/Brazil

Previously a PhD
Researcher at **LETRA**
(UFMG)



**THIAGO CASTRO
FERREIRA, PhD**

NLP Specialist at
aiXplain

Previously a Postdoc
Researcher at **LETRA**
(UFMG)



**ADRIANA
PAGANO, PhD**

Full Professor in
Translation Studies
(UFMG/Brazil)

Principal Researcher at
LETRA (UFMG)



**FABIO ALVES,
PhD**

Full Professor in
Translation Studies
(UFMG/Brazil)

Principal Researcher at
LETRA (UFMG)

Background and Motivation

Background and Motivation

While **Natural Language Processing** (NLP) has gained substantial attention with the most recent LLMs¹ (ChatGPT), only in recent years has research opened new paths for tackling **Human-Computer Design** (HCD) from the perspective of natural language.



LLMs¹ - Large Language Models

Background and Motivation

Limited research on **Human-Computer Design (HCD)** and **Natural Language Generation (NLG)** with output

- more fluent to human readers;
- more relevant to people's interests;
- part of a human-centered approach.



Hypotheses Development

Aim



Provide a resource that could help the development of more human-centered NLP systems and could overcome the lack of in-domain data in a **low-resource context**.

Assumptions: Human-centered approach

(Norman, 2023, p.2)

1. is people-centered (i.e., focuses on people within their context);
2. understands and solves source of problems;
3. assumes everything is an interconnected system;
4. performs small and straightforward interventions to refine solutions in an iterative form.



A human-centered corpus,
more specifically,
a persona-based corpus

Assumptions: Human-centered NLP

We explore human-computer interaction from a **human-centered NLP** perspective, which according to Yang (2023a, 10) “seeks to create NLP systems that are more user-friendly, accessible, and inclusive”.



Methodology

NLG Output Limitations to user proficiency levels

Example: **Healthcare domain.**

Boa tarde gente hoje acordei com as pernas e os pés inchados e adormecidos está doendo muito não sei o que fazer.

[Good afternoon people today I woke up with swollen and numb legs and feet, it hurts a lot I don't know what to do.]

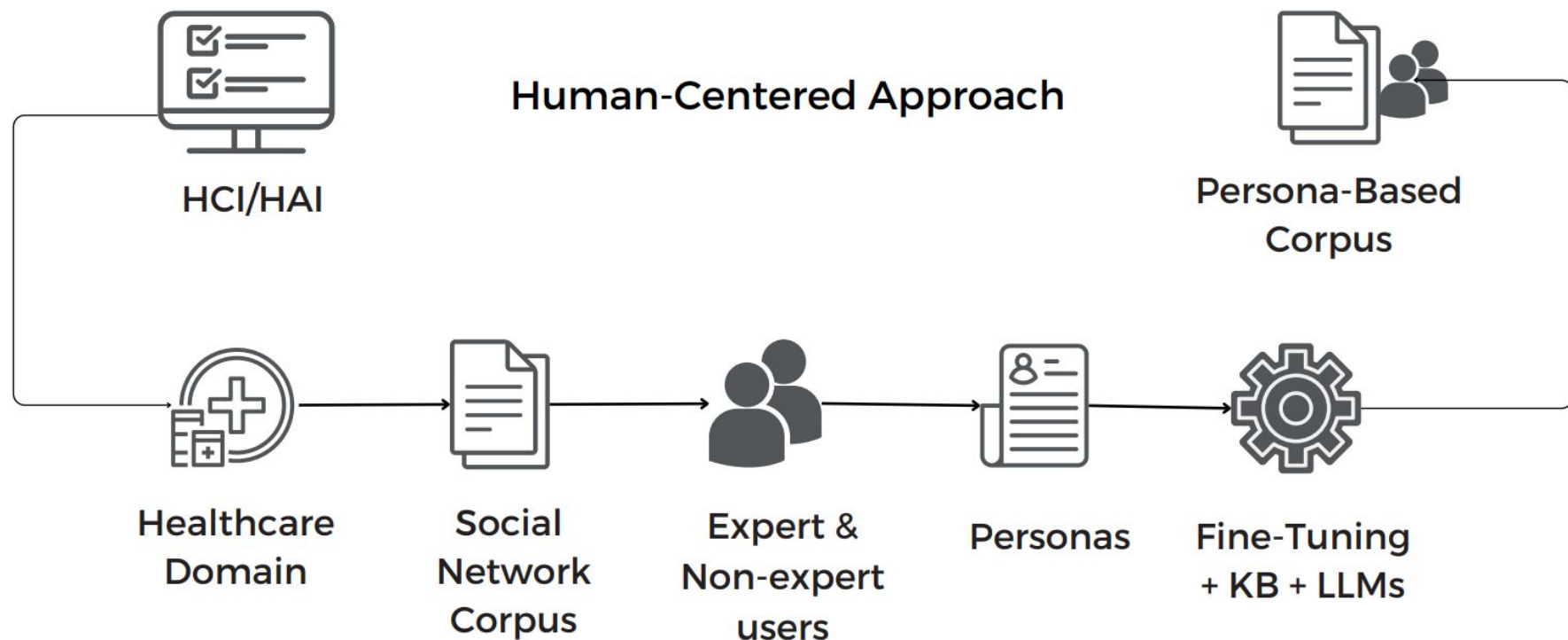


Adaptation to users with **different proficiency levels**

Target population: Brazilian Healthcare Service users

Language: **Brazilian Portuguese**





01

Human-Centered Approach

Contributes to a relevant comprehension within the domain: expert to lay users, complex to simplified texts, by revisiting Text Simplification applications guided by a Human-Centered approach.

03

Empoder@ Project

Corpus targets users by providing answers on two levels of complexity. That is why each answer drafted in response to a post or question has two versions, one of them being more simplified than the other (Pagano, 2015).

02

Personas

We addressed this need for a more user-driven generation by applying a group of personas (Nielsen, 2019) that are more frequently encountered within the healthcare domain (Pagano et al., 2023) ranging from more expert to non-expert users.



Developing an NLG artifact from a Human-Centered Perspective

HCD approach: Personas



ERIS CARVALHO

BIO

Eris has always worked as a housekeeper and retired by age. Eris lives alone near the oldest daughter's house and takes care of a small grocery shop. Eris has a family history of diabetes and is currently suffering from memory lapses and weight gain. Eris doesn't like technology because of poor eyesight and can't read without glasses anymore. Eris is caring and loves cooking for family and friends. Eris' social life is limited to grandchildren and ballroom dance classes. She is looking for information that could help to better understand her new condition.

GOALS

- Age healthily.
- Be an essential family figure.
- Reduce the effects of weight and memory lapses.
- Learn more about how to improve her diet.

FRUSTRATIONS

- Limited use of technology.
- Reading is difficult.
- Loves to eat sweets and does not like people regulating her consumption.
- Having difficulty finding out about a possible diabetes diagnosis.

"I love being with my grandchildren, but I take time for my recipes and taking care of my garden"

Age: 60-65

Education: High school

Sex: Female.

Pronouns: She/Her

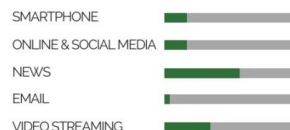
PERSONALITY



HEALTH



MEDIA / INFORMATION



LITERACY LEVEL



ALEXIS ROCHA

BIO

Alexis has worked with diabetes type 1 and 2 patients for the past five years and has become project coordinator at a municipal hospital in Northeastern Brazil. She also works at a private clinic for people with nutrition deficiency. Alexis uses a smartphone and computer to communicate with her patients and colleagues. Besides using a computer and smartphone at work, she also uses her laptop to share information through social media and read news about her work, her Yoga classes, and occasionally to make vacation bookings. Alexis is 20 weeks pregnant and recently discovered gestational diabetes. She noticed the lack of published information about the subject, especially regarding diet, and has decided to investigate how to disseminate more information online.

GOALS

- Seek information to help her patients.
- Give community guidelines.
- Reduce the effects of diabetes type 1 and 2 on teenage patients.
- Learn more about how to improve her skills.

FRUSTRATIONS

- Instruct her patients in a more effective way.
- Help her patients to elaborate on their emotional vocabulary.
- Spare time is difficult.
- Find nutritional information about gestational diabetes.

"The challenge is to teach patients how to develop new habits."

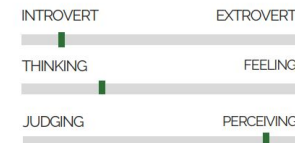
Age: 30-35

Education: Postdoctoral degree

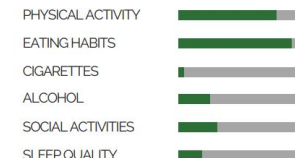
Sex: Female.

Pronouns: She/Her

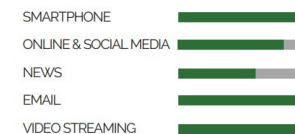
PERSONALITY



HEALTH



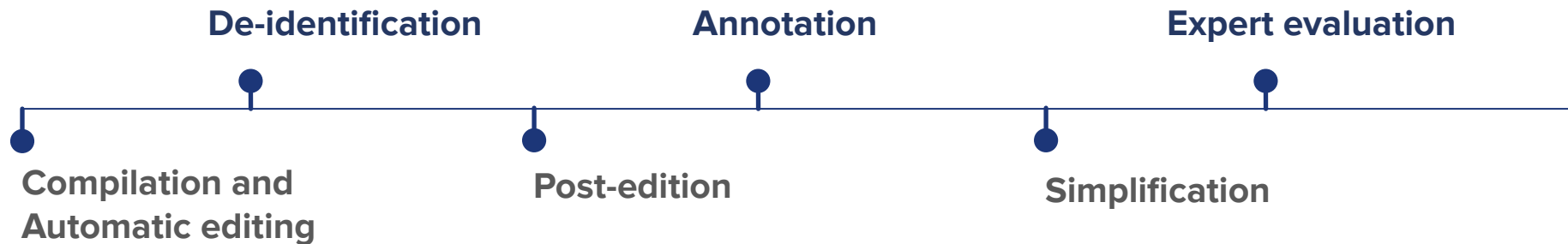
MEDIA / INFORMATION



LITERACY LEVEL



Persona-based corpus compilation timeline



Persona-based corpus compilation

QUESTION	Omitted due to ethical considerations.
AUTOMATIC EDITING	<i>Olá! Como vou saber se o aparelho de medir a glicemia está com defeito!</i>
POST-EDITING	<i>Olá! Como vou saber se o aparelho de medir a glicemia está com defeito?</i>
GLOSS	Hello! How will I know if the device for measuring blood glucose is failing?
ANSWER	<i>É recomendado que se calibre o glicosímetro (aparelho de medir a glicemia) sempre que for abrir um novo pacote de fitas de testagem. Essa calibragem pode ser realizada em farmácias.</i>
SIMPLIFIED VERSION	<i>É recomendado que se calibre o glicosímetro (aparelho que mede a quantidade de açúcar no sangue) sempre que for abrir um novo pacote de fitas de testagem. Essa calibragem pode ser realizada em farmácias.</i>
GLOSS	<i>We recommend you calibrate the glucose meter (device for measuring blood glucose) whenever you open a new package of testing strips. You can calibrate the glucose meter at a pharmacy.</i>
GLOSS FOR NON-EXPERT ANSWER	<i>We recommend you calibrate the glucose meter (a device that measures the amount of sugar in your blood) whenever you open a new package of testing strips. You can calibrate the glucose meter at a pharmacy.</i>
PROFICIENCY LEVEL: 4 DISEASE ACCEPTANCE LEVEL: 5	DOMAIN KNOWLEDGE: 3 CATEGORY: 6.3 - Glucose

Persona-based corpus

Features:

- Expert and Lay personas
- Proficiency level
- Domain knowledge
- Level of response complexity
- Acceptance level:
denial, anger, bargaining,
depression, and acceptance
(Kübler-Ross et al., 1972).

Annotation:

- Domain annotation:
date, time, food, glucose,
symptom, complication,
medication
- Sentimental analysis:
very positive, positive,
neutral, negative, and very
negative

HCD evaluation

LREC-COLING 2024

HCD

- Persona-based exploratory test (Nielsen, 2019);
- Usability test (Cunha, 2016; Cunha and Fernandes, 2016; Suojanen et al., 2015).



Persona-based Corpus Statistics

Persona-based corpus				
Origin	Train	Dev	Test	All
Original	881	109	109	1099
Nutrition	655	81	82	818
Total	1536	190	191	1917

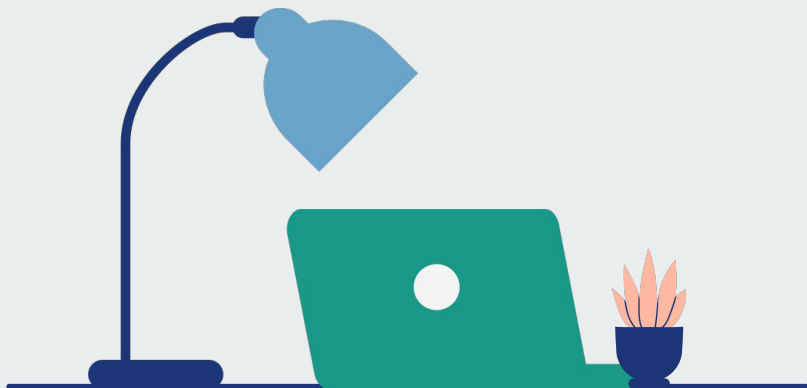
Literacy	Persona	Literacy
1	0	simple
2	0	simple
3	1	technical
4	1	technical
5	1	technical

Available at <https://github.com/Dia-Bete/PersonaBasedCorpus>

Outcome

Persona-based corpus

- A unique corpus with the combination of persona + healthcare domain + HCD approach
- A HCD evaluation at each utterance of the corpus compilation.



Main Contributions to Applied Linguistics and Natural Language Generation

01

Human-centered Approach

Benefits for humans and NLP;

Promoting literacy in low-resource contexts - domain, language;

→ Extra: Balancing expertise level (research group).

02

Text Simplification

Improvement of user comprehension, health literacy, and user experience in a low-resource context.



01

Persona-based English Corpora

Augmenting training data from
persona-based English corpora.

02

Clinical Data

Extending the persona-based
corpus with clinical data.

Future
Work



References

- Rossana Cunha. 2016. Avaliação de um sistema com base em corpus para a pesquisa, ensino e prática da tradução através de critérios de usabilidade e ergonomia cognitiva [Evaluation of a corpus-based system for the research, teaching and practice of translation under the perspective of ergonomics and usability]. Master's thesis, Federal University of Santa Catarina, Florianópolis/SC - Brazil.
- Rossana Cunha and Lincoln P. Fernandes. 2016. Avaliação comparativa de ferramentas web de apoio à tradução através de critérios de usabilidade e ergonomia cognitiva. Tradução em Revista, 21:67–90.
- Adriana Pagano. 2015. A linguagem na construção das práticas educativas nas ciências da saúde. Heloísa de Carvalho Torres, Ilka Afonso Reis, and Adriana Pagano. Empoderamento do pesquisador nas ciências da saúde. Belo Horizonte: FALE/UFMG, pages 19–36.
- Adriana Pagano, Claudia Moro, Elisa Terumi Rubel Schneider, Lilian Mie Mukai Cintho, and Yohan Gumiel. 2023. PLN na Saúde, book chapter 21. BPLN.
- Elisabeth Kübler-Ross, Stanford Wessler, and Louis V Avioli. 1972. On death and dying. Jama, 221(2):174–179
- Lene Nielsen. 2019. Personas-user focused design. 2nd edition. Springer.
- Don Norman. 2023. What is Human-Centered Design (HCD)?, chapter What is Human-Centered Design (HCD)? Interaction Design Foundation - IxDF.
- Tytti Suojanen, Kaisa Koskinen, and Tiina Tuominen. 2015. User-centered translation. Routledge.
- Diyl Yang. 2023. Cs329x: Human-centered nlp – human-centered nlp. Accessed: 13 July 2023.

Grazie mille!
Thank you!
Muito obrigada!



*Rossana Cunha
Fabio Alves
Adriana Pagano
Thiago Castro Ferreira*

UF *m* G

FALE
FACULDADE
DE LETRAS

POSLIN-UFMG

LETRA

CAPES

CNPq

aiXplain

CIASC
CENTRO DE INFORMÁTICA
E AUTOMAÇÃO DO ESTADO DE SANTA CATARINA S.A