



# A Short Story about my First Year in the OEG

**Isam Diab**

Ontology Engineering Group  
Universidad Politécnica de Madrid, Spain

## Who am I?

Isam Diab Lozano (Valladolid, 1996)



## Background (Academics and Work)



**Universidad de Valladolid**

Spanish Language and Literature

2014-2018



**UNIVERSITAT DE  
BARCELONA**

Master in Cognitive Science  
and Language

2018-2019

*emagister:*



Computational Linguist

2018-2020

Since March 2021...



Cognitive Accessibility research line thanks to a working contract from the Community of Madrid, supervised by  
**Mari Carmen Suárez de Figueroa Baonza**

- **Objective** of the proposal: to explore and apply Artificial Intelligence (AI) technologies to solve cognitive accessibility problems in documents.
  - **Multidisciplinary** line:



- **Target**: persons with disabilities or with different situations.

# Need of the Research Line



Jaime Gómez-Obregón ✅  
@JaimeObregon

¡Hola, Gobierno de España! 🙌

¡Hola, amigos de las Administraciones Públicas! 🙌

Millones de ciudadanos nos topamos con avisos administrativos como este que habéis claveteado en el frontispicio del nuevo buzón de notificaciones electrónicas.

¿Puedo proponer una mejora? 🖕

The screenshot shows a web interface for 'dehú' notifications. At the top, there's a header with the Spanish flag, 'GOBIERNO DE ESPAÑA', and the 'dehú' logo. Below the header, there are several buttons: 'Notificaciones Pendientes', 'Notificaciones Realizadas', 'Comunicaciones', 'Mis datos de contacto', and 'Gran Destinatario'. A dropdown menu indicates the recipient is 'JOHN RAMBO' and 'FUERZAS ESPECIALES EN VIETNAM'. There's also a 'Cerrar sesión' button. The main content area has a green banner at the top with the word 'AVISO'. The text in the banner is in Spanish and discusses article 44.5 of the Royal Decree 203/2021 regarding electronic communication. Below the banner, there's a section titled 'Comunicaciones' with a sub-section 'Avisos importantes'. A red box highlights a specific message from 'JOHN RAMBO'.

## Comunicaciones

Desde aquí tiene acceso a las comunicaciones que hayan emitido los diferentes Organismos de la Administración a JOHN RAMBO con NIF 0000000000.

Puede consultar la lista completa de administraciones adheridas en el enlace de [Organismos Emisores](#).

Por defecto se mostrarán las comunicaciones de los últimos 30 días. Podrá modificar los filtros para realizar otras consultas.

Buscador de Comunicaciones

Criterio de búsqueda: Fecha de puesta a disposición: dd/mm/aaaa Desde: dd/mm/aaaa Hasta: dd/mm/aaaa Límpiar Buscar

Buscar avanzado Q+



Jaime Gómez-Obregón ✅ @JaimeObregon · 17 abr.  
Y es que esto nos estás haciendo leer:

- Una sola frase... ¡de 5 líneas y 113 palabras!
- ¡Tiene diez comas! Y un par de paréntesis.
- ¡Hacéis 11 incisos!
- ¡Con incisos dentro de incisos! Incisos anidados.

¿Por qué nos hacéis esto? 😳

¿Por qué os dirigís a nosotros así? 😰



Jaime Gómez-Obregón ✅ @JaimeObregon · 17 abr.  
Eso no es español. Es marciano.

Pensad que la ciudadanía es diversa. Hay personas de muy variadas capacidades, culturas y habilidades cognitivas. Y vuestra responsabilidad es administrarnos a todos. 🎉



Jaime Gómez-Obregón ✅ @JaimeObregon · 17 abr.

Esta arcana jerga administrativa erige una barrera artificial entre vosotros, la Administración, y nosotros, los administrados. Y eso es malo para el país, porque provoca desafección de la ciudadanía hacia las instituciones.

Pero es que, además,

- Nos distancia
- Nos separa

1:34 p. m. · 17 abr. 2022 · Typefully



Find [here](#) the original tweet link!



# PhD Thesis Plan: A Three-Part Story

Proposition

Heart

Denouement

P

h

D



March – October 2021

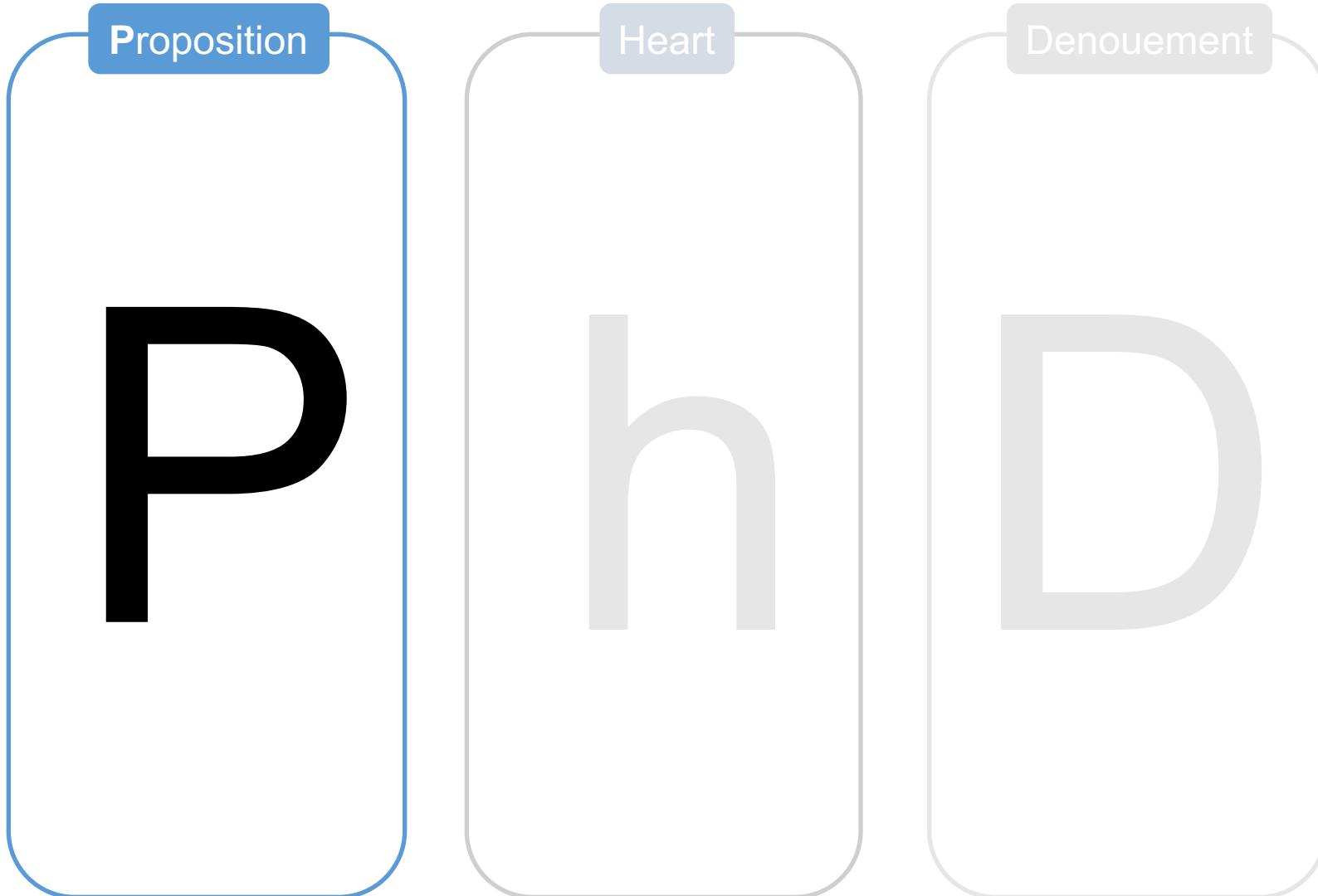
## Research Assistant

Analysis of AI  
techniques for  
adapting texts

Creation of  
some user  
studies

Data Analysis  
and Results  
Extraction

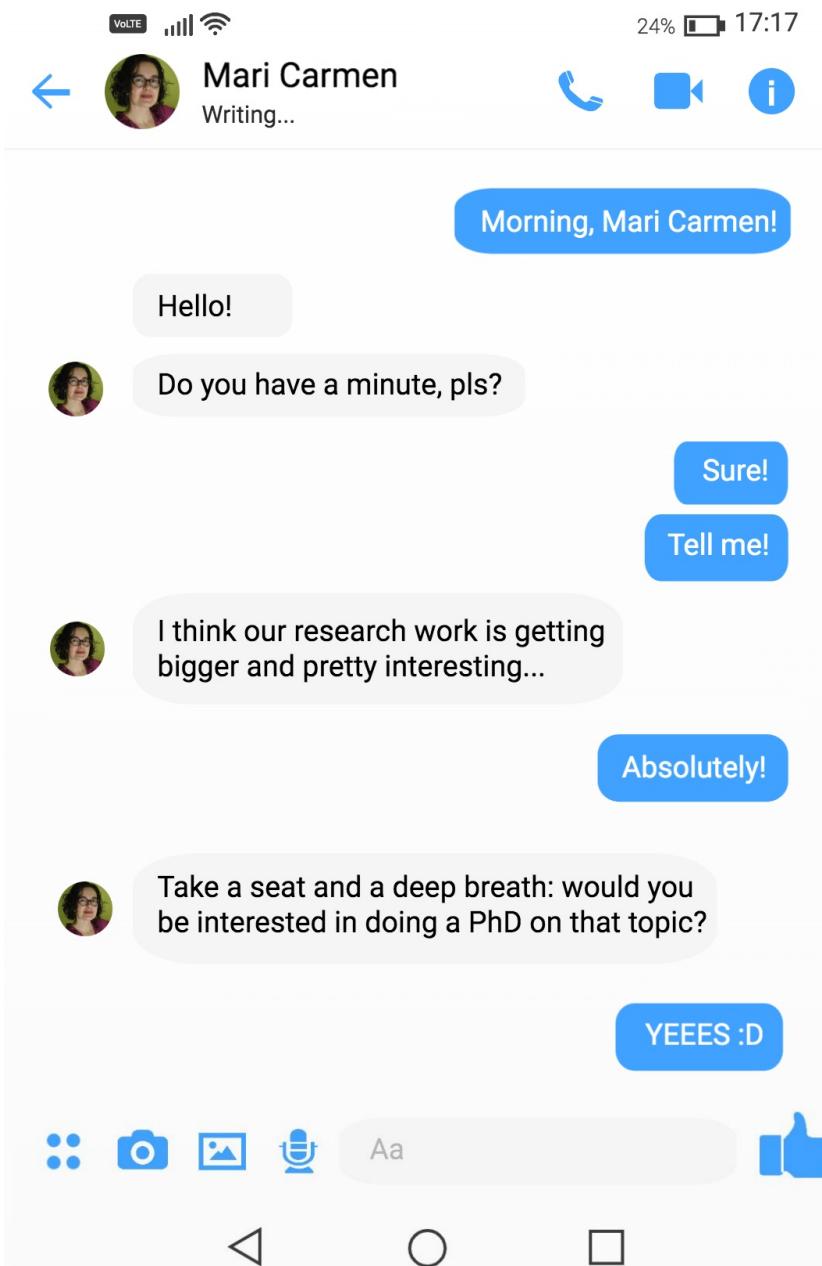
Proposals for  
implementation  
methods



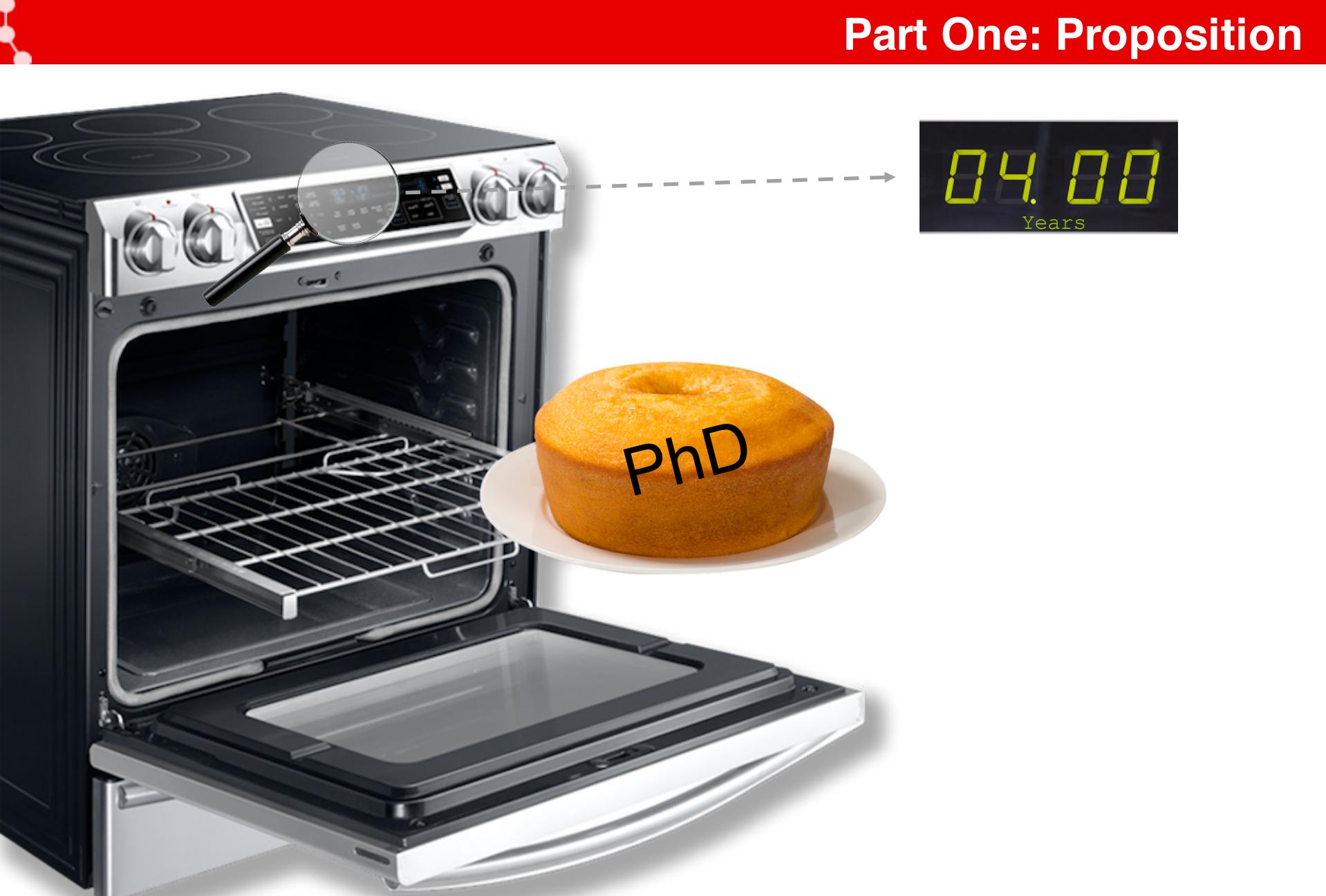
# Part One: Proposition

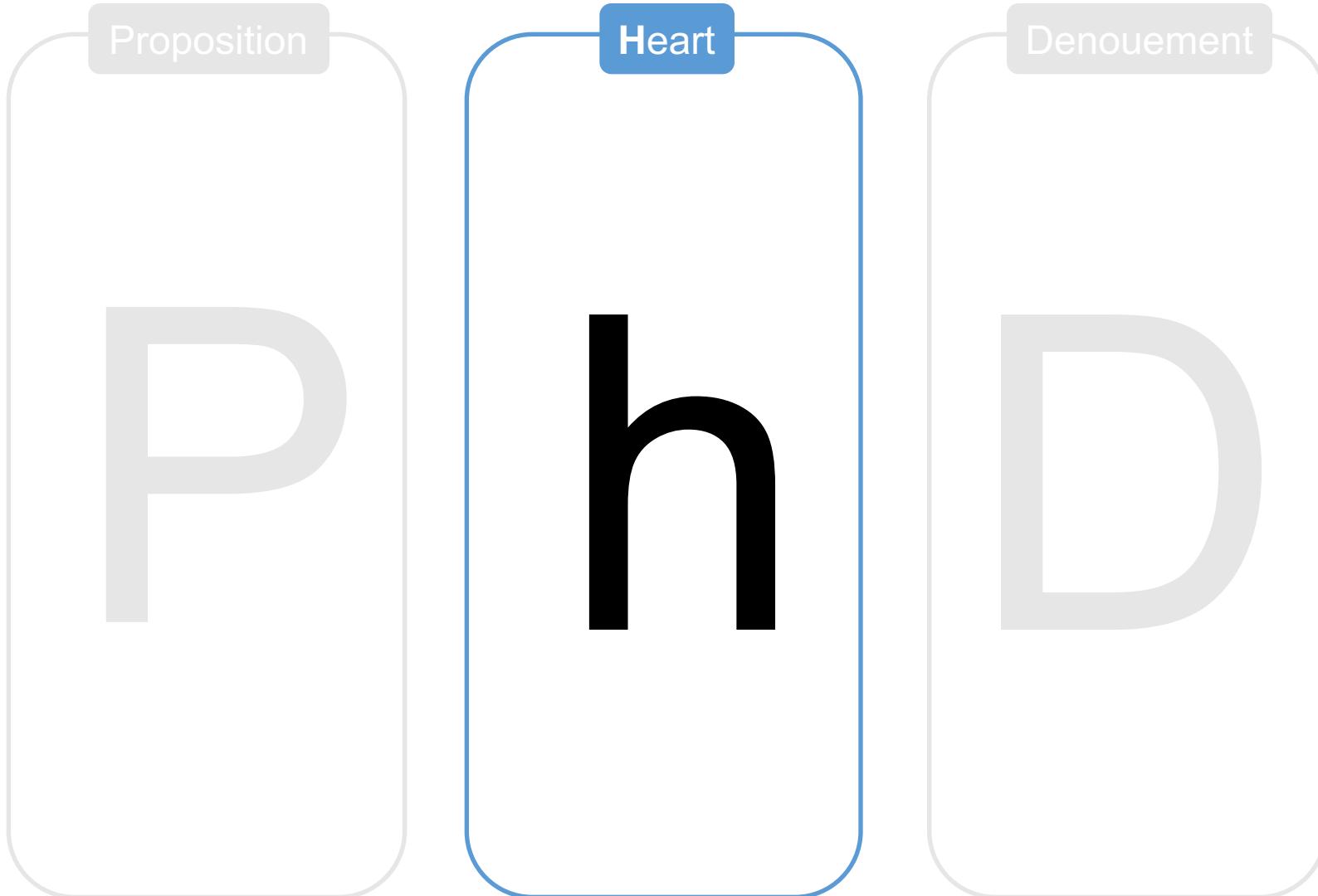


# Part One: Proposition



# Part One: Proposition







October 2021 – Now

## PhD Candidate

**Towards An Artificial Intelligence (AI)-based  
Framework to Facilitate the **Reading Comprehension**  
of **Short Stories in Spanish****

## Introduction and Background

- Socio-cultural **problem**: there are people who for different reasons present difficulties in reading comprehension, and among them are groups of people with disabilities.
  - Cognitive Disabilities: some data (2019):
    - **Spain**: 270.00 people
      - Same as Granada population
    - **UK**: 1.5 million people
    - **Earth Planet**: 2,6% of the world's population

- How to **solve** the problem?

## E2R METHODOLOGY



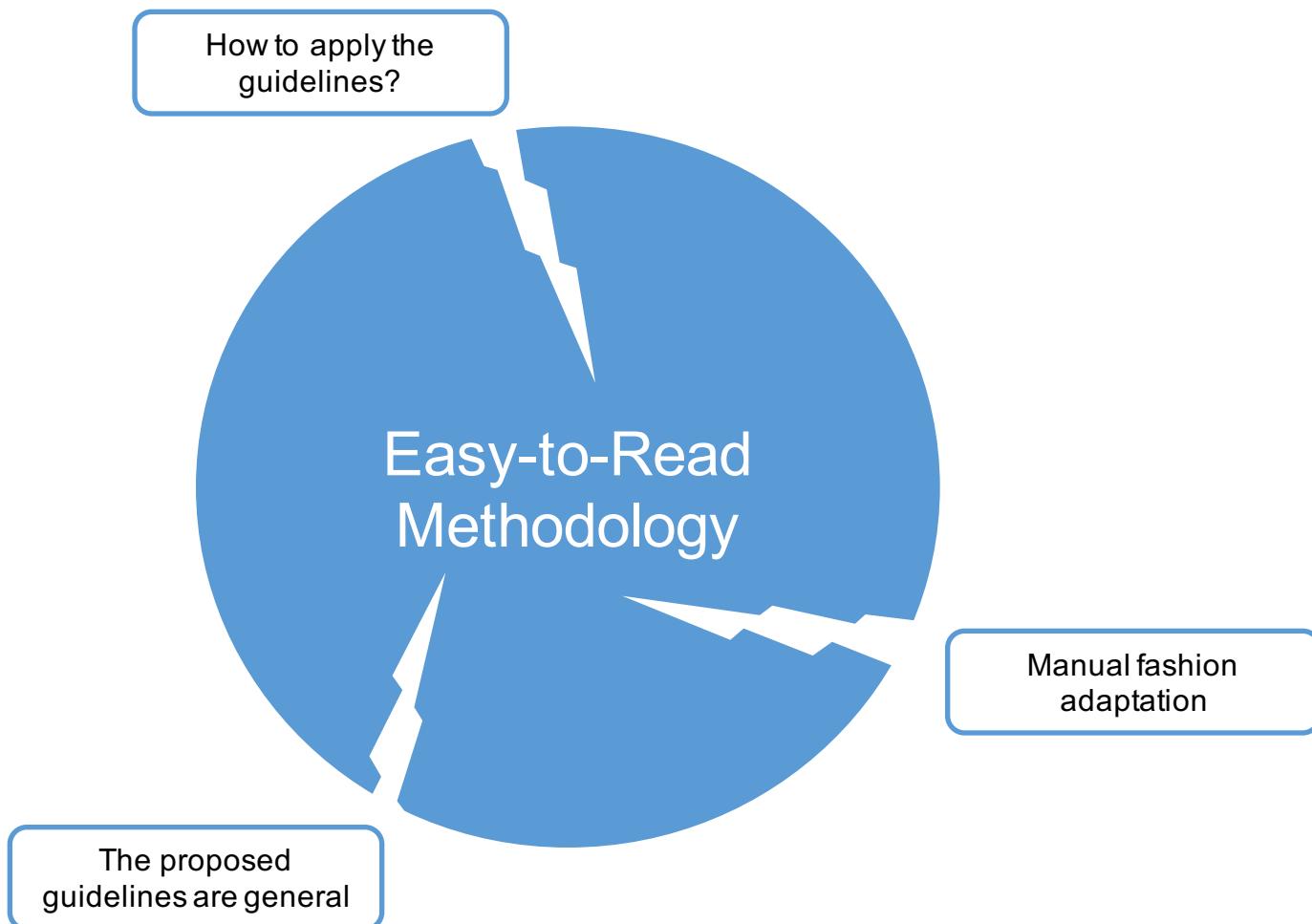
Nomura, M., Nielsen, G.S., International Federation of Library Associations and Institutions, Library Services to People with Special Needs Section: **Guidelines for easy-to-read materials**. IFLA Headquarters, (2010).



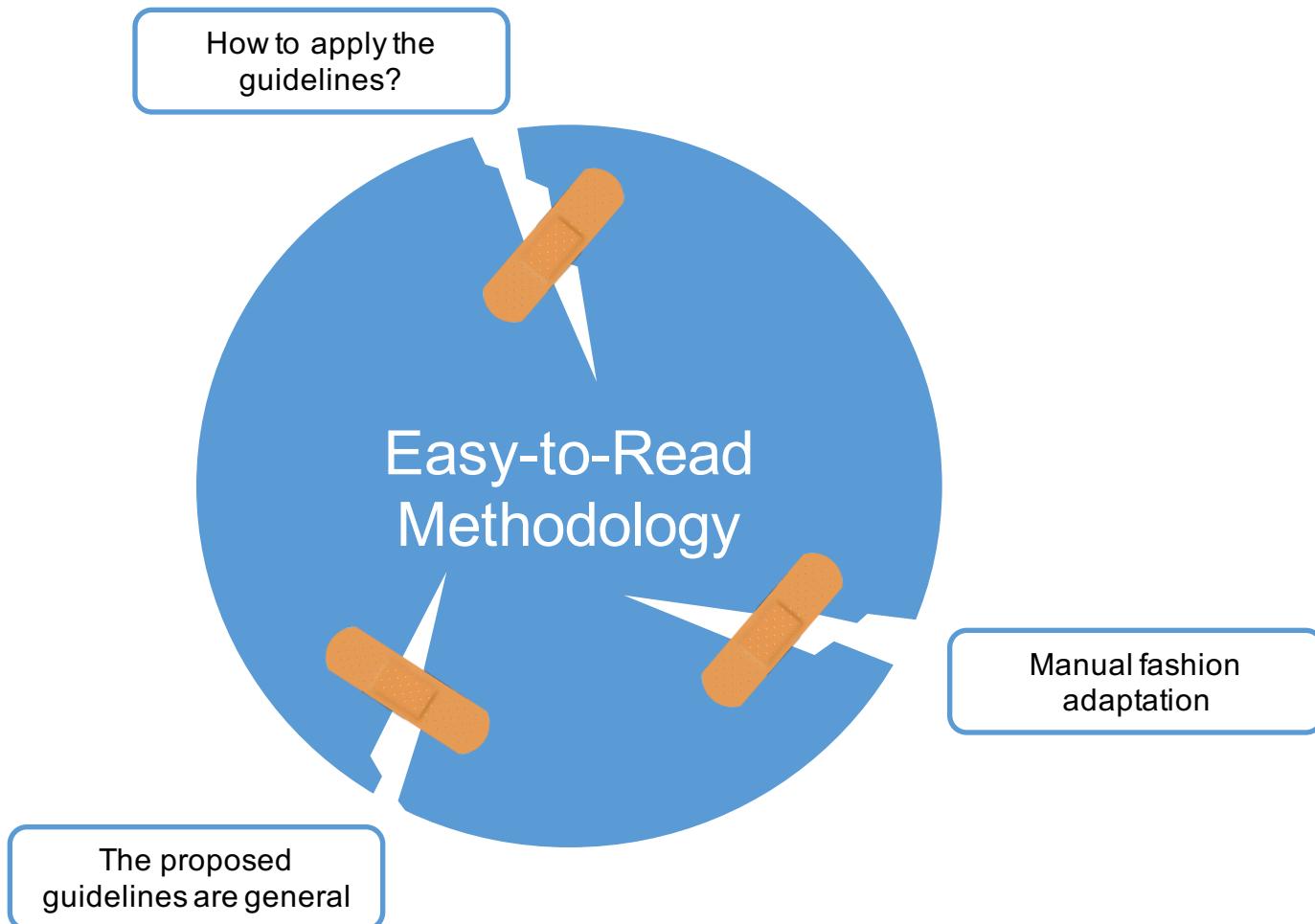
AENOR: **Lectura Fácil. Pautas y recomendaciones para la elaboración de documentos (UNE 153101:2018 EX)**. Asociación Española de Normalización (2018)



But not all that glitters is gold... It has **limitations**:



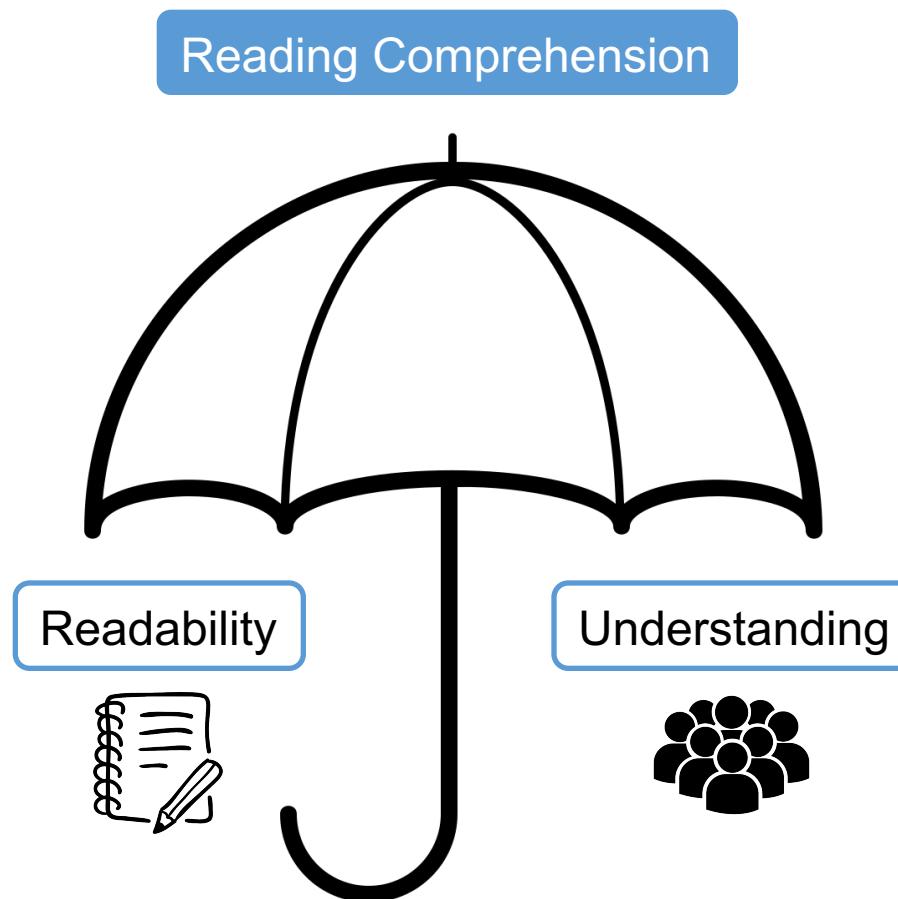
- Thus, our aim:



- 1. What is Reading Comprehension?**
- 2. Why do we focus the research on Short Stories?**

- 1. What is Reading Comprehension?**
- 2. Why do we focus the research on Short Stories?**

# What is Reading Comprehension?



- 1. What is Reading Comprehension?**
- 2. Why do we focus the research on Short Stories?**

## What is a short story?

A short story is a piece of prose fiction that typically can be read in one sitting and focuses on a self-contained incident or series of linked incidents, with the intent of evoking a single effect or mood. The short story is one of the oldest types of literature and has existed in the form of legends, mythic tales, folk tales, fairy tales, tall tales, fables and anecdotes in various ancient communities around the world.



## 1. Linguistic Issues.

Pragmatics

Semantics

Syntax

Layout

Vocabulary

Un auténtico flechazo, como en tantos otros casos, fue el comienzo de aquella historia de amor. Ella salía de su despacho de ejecutiva de la agencia de publicidad en la que trabajaba, y al volver la cabeza lo vio junto a un compañero. Siempre es difícil de explicar esa mezcla de admiración y deseo, de sorpresa ante lo inesperado, de emocionado deseo de posesión, esa extraña química que envuelve un flechazo, y que nada tiene que ver con los fríos análisis racionales. Algo, dentro de ella, suscitó la pasión que le llevaría a buscar su compañía, cosa que logró al día siguiente, porque si no se trataba de una mujer de gran belleza, sí que poseía una enorme capacidad de resolución, una admirable celeridad para decidir.

## 2. Structure.

- Normally, short stories have a fixed structure and content organization.
- Short stories are of moderate length.

Proposition

Heart

Denouement

## 3. Social Impact.

“Lee y conducirás, no leas y serás conducido”.

Santa Teresa de Jesús



**OPR1** The number of adapted short stories in Spanish into easy-to-read versions is quite scarce. Thus, persons with reading comprehension difficulties do not have access to literature.

**H1** The (semi)-automatic adaptation of short stories written in Spanish into E2R versions is **possible**.

**H2** The result of the (semi)-automatic adaptation of short stories written in Spanish into E2R versions will **improve the comprehension** of such texts by people with reading comprehension difficulties, in terms of (a) **readability** and (b) **understanding**.

## Restrictions

**R1** No specific comprehension difficulties will be addressed.

# Objective and Sub-Objectives

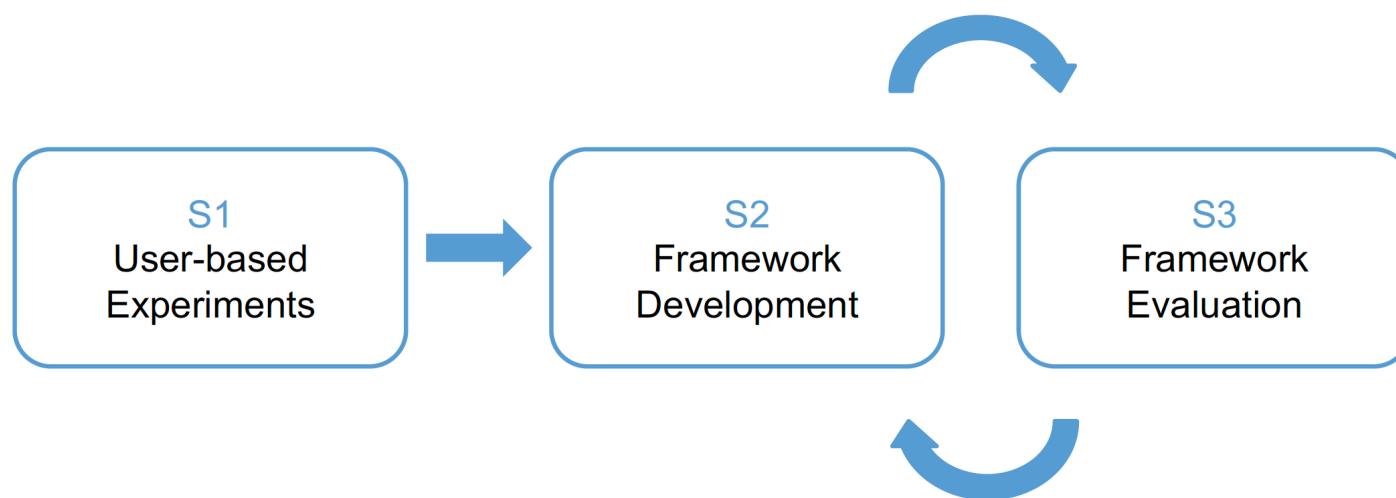
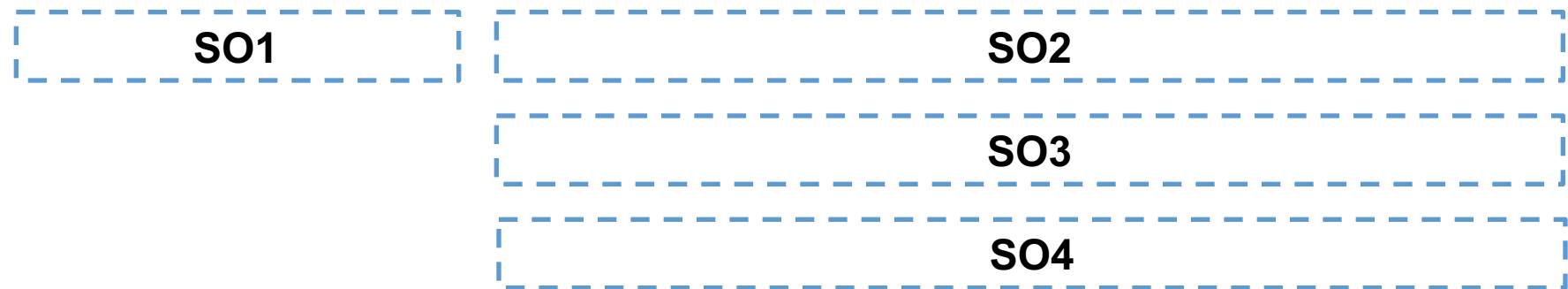
**Research Objective (RO)** To develop (a) a technological framework for adapting (semi)-automatically short stories in Spanish into E2R versions, and (b) a framework to evaluate the understanding of short stories by persons with reading comprehension difficulties.

**SO1** To identify the main **cognitive problems** in short stories regarding reading comprehension by people with comprehension difficulties.

**SO2** To determine and classify the different **linguistic criteria** concerning the identified problems, as well as to relate them to E2R guidelines.

**SO3** To analyze which **AI-based methods and techniques** are the most **appropriate** to be used in the technological framework.

**SO4** To **design** and **develop** specific **methods** for adapting (semi)-automatically short stories in Spanish into E2R versions.



## Stage 1. User-based Studies



### Type/Nature

Supervised

### Sample

Persons with reading comprehension difficulties

- E2R Validators
- No validators

### Techniques

- Free recall
- Induced recall methods
- Metacognition

### Metrics

- Time
- Nonverbal language

- Currently...
  - ACCEDES (Madrid)
  - Fundación Roncalli (Madrid)
  - Others?
- Example of a questionnaire we launched:  
<https://forms.gle/sW5eTCk5xDsQjFkm9>

## Stage 2. Framework Development

- To classify and relate the problems we found in the studies to the E2R guidelines in order to:
  - Complete them
  - Improve them
  - Find new ones
- Once related: to automate them by means of AI methods and techniques:

Natural Language Processing

Pattern Matching

Machine Learning

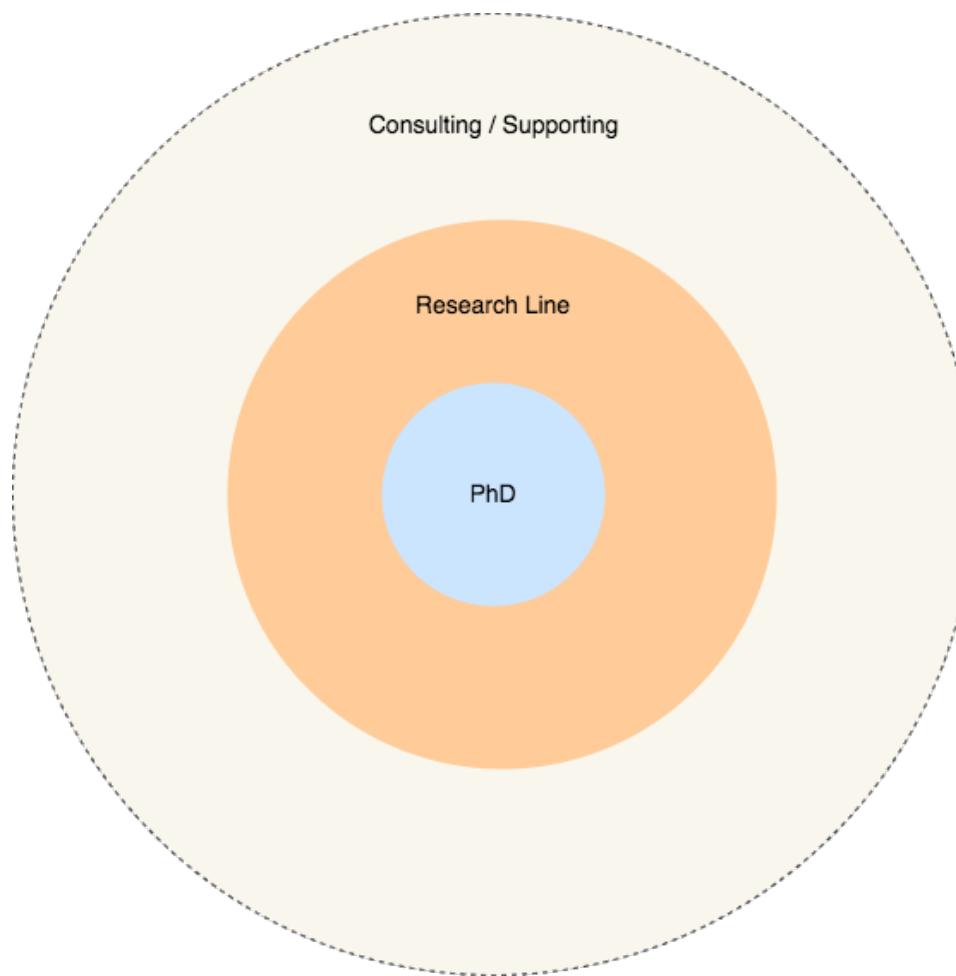
## Stage 3. Framework Evaluation



### User-based Studies

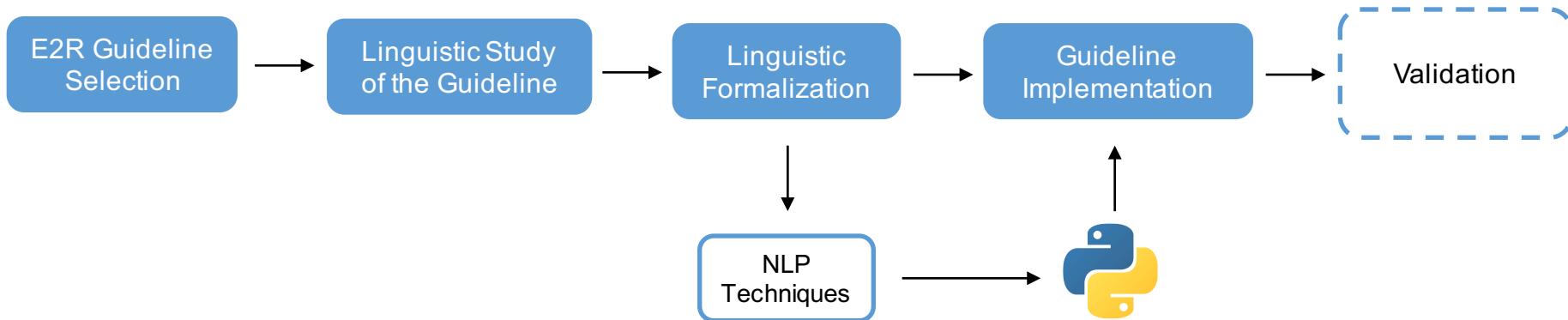
- E2R Professionals
- Persons with RC difficulties

Implementation of **readability and understanding indices** to compare the original short story and its E2R version provided by our software



# Consulting or Supporting

- Early implementation of E2R guidelines (for identification and/or transformation):
  - Avoid the use of **superlatives** (*riquísmo* → *muy rico*).
  - Avoid the use of adverbs ending in **-mente** (-ly) (*automáticamente* → *de manera automática*).
  - Avoid the use of **complex words** (complex word → easy synonym/glose).
  - Avoid the use of the **gerund** form (*estaba comiendo* → *comía*)
  - Avoid the use of verbal **periphrases** (*volver a llamar* → *llamar de nuevo*).
  - Avoid the use of **metaphorical** language.
  - Avoid the use of **repetitions** (*te quiero tanto tanto* → *te quiero mucho*).
  - Avoid the use of **sexist language** (*los alumnos* → *el alumnado*).
  - Avoid the use of **abbreviations and acronyms** (*Sra.* → *Señora*).
  - Avoid the use of **passive voice** (*fueron creados* → *crearon*).



## Some Results (so far)



Your APP for E2R Adaptation



## Example of E2R Guidelines Identification (screenshot)

**FACILE** Your APP for E2R Adaptation

CONGRESO.pdf ▾

Página Anterior Siguiente Página Página 1 de 1

The screenshot shows a document page titled "CONGRESO.pdf". At the top left is a back arrow icon. Below the title bar are navigation links: "Página Anterior", "Siguiente Página", and "Página 1 de 1". The main content area features a large image of the Palacio de las Cortes in Madrid, with the caption "Congreso de los Diputados de España". To the right of the image is a text block: "El Congreso de los Diputados es la cámara baja de las Cortes Generales, el órgano constitucional que representa al pueblo español. Se reúnen para sesiones en el Palacio de las Cortes ubicado en la Plaza de las Cortes en la Plaza de las Cortes de Madrid." Below this is another text block: "El Congreso de los Diputados está compuesto por un mínimo de 300 y un máximo de 400 diputados. Los diputados son elegidos por sufragio universal, libre, igual, directo y secreto (art. 68 CE). Las elecciones se celebran ordinariamente cada cuatro años, o antes en caso de elecciones tras una disolución anticipada de las Cámaras." At the bottom of the page is a footer with the question "¿Qué Pautas de Lectura Fácil cumple el documento?" followed by a list of readability guidelines with green checkmarks and red X's.

MAQUETACIÓN

- La fuente del texto pertenece a los estilos aceptados
- El tamaño del texto tiene que ser como mínimo 12 y como máximo 16
- No existe texto en cursiva
- No existen más de un % de palabras en negrita
- No existen más de un % de palabras subrayadas
- No existen textos con sombreado
- No existen más de un % de palabras en mayúsculas
- Color de fuente negro
- Color de fondo blanco sólido
- Cantidad de palabras en la diapositiva no supera el límite establecido

Descargar

## Example of Text Transformation (screenshot)



Concepto      Analizar Texto      Subir Ficheros      Contacto

### Analizar Texto

Eficacia: Capacidad de lograr el resultado que deseamos después de realizar una acción.

Ejemplo de uso: El mecánico está comprobando la eficacia de los frenos del coche.

Una persona vacunada **está protegida** frente a la enfermedad pero no frente al virus. Las vacunas tienen un **altísimo** porcentaje de efectividad y eficacia

Las vacunas **protegen** a las personas frente a la enfermedad pero no frente al virus. Las vacunas tienen un porcentaje de efectividad y **eficacia muy alto**

Analizar otro texto >

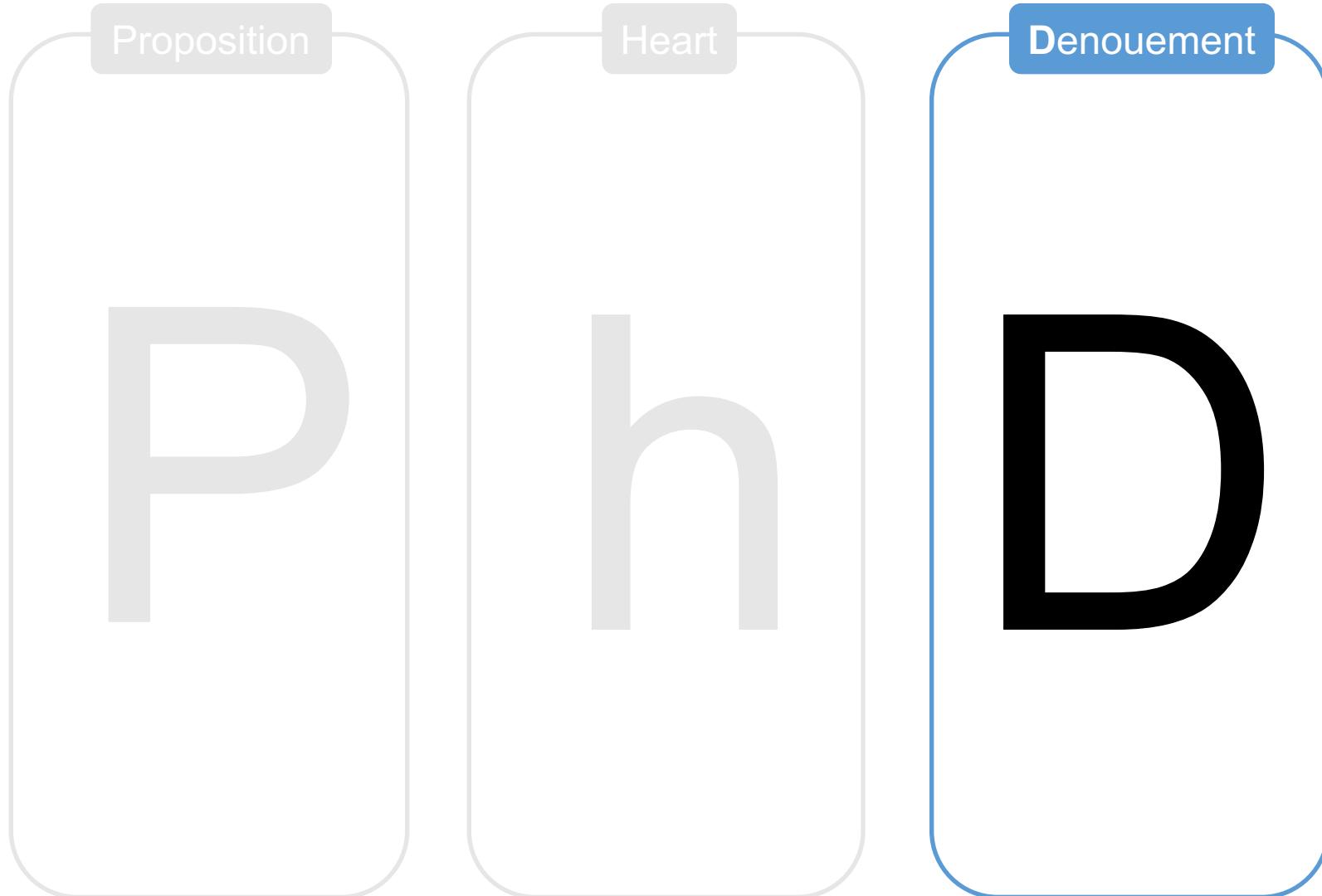


-  Mari Carmen Suárez-Figueroa, [Isam Diab](#), Edna Ruckhaus, Álvaro González. First steps in the development of a support application for Easy-to-Read adaptation. *Universal Access in the Information Society*, (2021)
-  Mari Carmen Suárez-Figueroa, [Isam Diab](#), Álvaro González, Jesica Rivero-Espinosa. First Attempt to an Easy-to-Read Adaptation of Repetitions in Captions. *ICCHP-AAATE Conference*, (2022)
-  [Isam Diab](#), Mari Carmen. Shall the Easy-to-Read Adaptation of Micropoems affect Emotions?. *ICCHP-AAATE Conference*, (2022)
-  Mari Carmen Suárez-Figueroa, [Isam Diab](#), Álvaro González, Jesica Rivero-Espinosa. Towards an Automatic Easy-to-Read Adaptation of Superlatives in Spanish. *ACM ASSETS Conference*, (2022)

[Ongoing] 2-3 more for the resto of the year.



# PhD Thesis Plan: A Three-Part Story



Hopefully, in two or three years...



I would like to thank their support and work:

- Supervisor: Mari Carmen Suárez de Figueroa Baonza
- Institutions such as ACCEDES and Fundación Roncalli (special mention to Isa Cano)
- Bachelor's degree and Master students
- Daily partners in the OEG

