



AIUCD 2021

Language Disparity in the Interaction with Chatbots for the Administrative Domain

M. Miliani^{1,2}, M. Benedetti¹, A. Lenci²

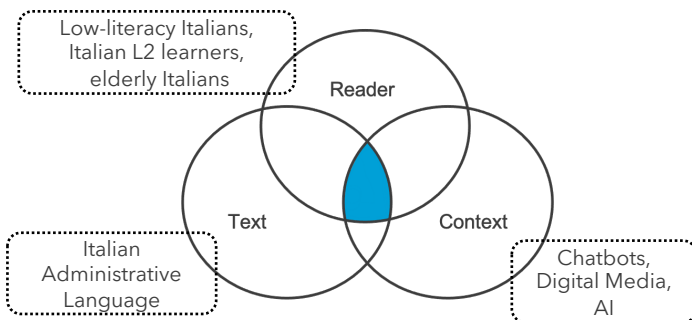
¹University for Foreigners of Siena

²CoLing Lab, Department of Philology, Literature, and Linguistics, University of Pisa

Introduction

- The writing style of **administrative texts** is artificial and obscure¹.
- Since citizen broadly access to the Internet and use extensively **mobile devices**, PAs are increasingly choosing to communicate through digital technologies, like **chatbots**.
- However, **language disparity** could impede full access to administrative texts.

Text Comprehension²



The **main goal** of this work is to detect which factors affect the **comprehension process** of citizens with limited reading skills when they access administrative texts, in the interaction with conversational AI systems, such as chatbots.

Is a simplification based on linguistic traits enough?

Reading comprehension test (10 questions per text)

4 texts in their original or **simplified versions**

- Passages from administrative acts and city halls' web pages:
 - A:** Web page about public subsidies (Naples)
 - B:** FAQ on eIDs (Rome)
 - C:** Municipality act about disability services (Bari)
 - D:** Call for bids for social housing (Trento)
- Text simplification adopting PACCSS-IT annotation schema³
- Simplification based on linguistic traits of the administrative Italian language

Evaluating Text simplification

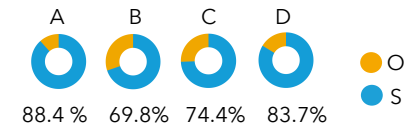
1) **Readability assessment tools** based on NLP:

- GULPEASE (% of simplicity)
- READ-IT⁴ (% of complexity)

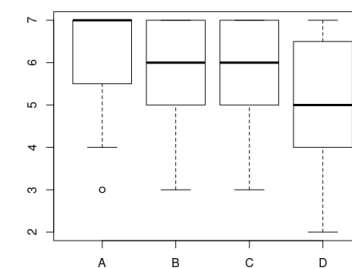
	A		B		C		D	
	O	S	O	S	O	S	O	S
GULPEASE	48.2	56.7	45.5	51.7	41	48	42.4	56.1
R-IT Basic	74.6	20.3	90.9	56.7	89.7	40	97.1	24.3
R-IT Lexical	17.7	0.4	4.2	0.7	92.7	12.6	98.2	30.9
R-IT Syntactic	99.6	90.7	91.2	88.5	95.9	65	100	88.4

2) **Preliminary test:** 43 Italian native speakers read the original (O) and simplified (S) text to decide:

- Which text is easier to understand?



- How similar are they? (Likert scale, 1-7)



- Since B was **simplified less** than other texts (G. +6.2), more participants chose the original version
- Since D was simplified more than other texts (G. +13.7), original and simplified text were judged **less similar**
- Participants' **familiarity** with administrative acts may affect the perception of complexity

Conclusion and future work

- The reading comprehension test can be used to predict the **actual level of informants' comprehension**.
- If simplified texts are still hard to read, this test will help to **detect features affecting readers comprehension process**, other than complexity linguistic traits

[1] Franceschini, F., & Gigli, S.. Manuale di scrittura amministrativa. Agenzia delle entrate, 2003.

[2] De Mauro, T., & Vedovelli, M.. (eds). Dante, il gendarme e la bolletta: la comunicazione pubblica in Italia e la nuova bolletta Enel. Laterza, 1999.

[3] Brunato, D., A. Cimino, F. Dell'Orletta, and G. Venturi. "Paccss-it: A parallel corpus of complex-simple sentences for automatic text simplification." EMNLP 2016.

[4] Dell'Orletta, F., S. Montemagni, and G. Venturi. "READ-IT: Assessing readability of Italian texts with a view to text simplification." SLPAT 2011.