Understanding Italian Administrative Texts: A Reader-Oriented Study for Readability Assessment and Text Simplification

Martina Miliani^{1,2}, Marco S. G. Senaldi³, Gianluca E. Lebani⁴ and Alessandro Lenci²

¹University for Foreigners of Siena

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

³Department of Psychology, McGill University

⁴Department of Linguistics and Comparative Cultural Studies, Ca' Foscari University of Venice

Abstract

The complexity of administrative texts can preclude citizens with language disparities from accessing relevant information. Recent deep-learning models of readability assessment and text simplification would greatly benefit from training materials that are annotated with the specific needs of the target readers. The aim of the present work is to investigate how differently second language learners of Italian and elderly Italian native speakers read and comprehend administrative texts of different readability levels in digital format, as compared to a control group of Italian native speakers. To this end, we conducted a study where 86 participants from the three groups were asked to perform a comprehension task via smartphone. Participants read administrative texts in their original and simplified form, where simplification was performed on the basis of linguistic features that previous literature considered typical of the administrative domain. Although the applied simplification did not seem to affect text comprehension, we observed differences across the three subject groups, especially in relation to participants' background.

Keywords

Natural Language Processing, Reading Comprehension, Public Administration, L2, Elderly, Automatic Readability Assessment, Automatic Text Simplification

1. Introduction

Even though public institutions communicate more and more through the web and innovative digital technologies and have been encouraged to use a plain language [1], Italian administrative texts appear still far from being easily readable [2]. Writing easy-to-read text is a non trivial task if we consider what text comprehension means. According to [3], text comprehension is determined by the interplay of three factors: the reader and their background, the reading

gianluca.lebani@unive.it (G. E. Lebani); alessandro.lenci@unipi.it (A. Lenci)

AIxPA 2022: 1st Workshop on AI for Public Administration, December 2nd, 2022, Udine, IT

m.miliani@studenti.unistrasi.it (M. Miliani); marco.senaldi@mcgill.ca (M. S. G. Senaldi);

thttps://colinglab.humnet.unipi.it/people/miliani/ (M. Miliani); www.unive.it/persone/gianluca.lebani (G. E. Lebani); https://people.unipi.it/alessandro_lenci (A. Lenci)

D 0000-0003-1124-9955 (M. Miliani); 0000-0003-2205-3843 (M. S. G. Senaldi); 0000-0002-3588-1077 (G. E. Lebani); 0000-0001-5790-4308 (A. Lenci)

^{© 02022} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

context (e.g., the medium used), and the text itself. For this reason, not only should algorithms for Automatic Readability Assessment (ARA) and Automatic Text Simplification (ATS) take into consideration the linguistic features of a text that pertain to its domain and genre [4], but should also be tuned to the specific needs of the target audience [5, 6].

These issues become particularly compelling when it comes to the administrative language. Its complexity can become a barrier to the accessibility of information related to citizens' rights [3]. This is especially true for citizens with language disparity, namely those who do not have an optimal level of language proficiency [7].

In this paper, we present a study that involved 86 participants belonging to two groups with language disparities, i.e., Italian second-language speakers and elderly Italian native speakers, and to a control group, i.e., Italian native speakers. Participants were asked to perform a comprehension task in a digital context (via smartphone) on original and simplified administrative texts. This simplification was carried out by only considering those features that previous literature considered typical of the linguistic complexity of the administrative domain. The goals of the present study are manifold:

- Assessing if there is a difference between the three groups in the comprehension of administrative texts with two different levels of readability;
- Exploring the effect of participants' background (e.g., education and digital literacy) on the comprehension of administrative texts;
- Detecting which linguistic features affect the comprehension performance across the three groups, over and above those strictly related to the administrative language.

Details on the experimental setting, including materials, selected participants, experimental design, and extracted linguistic features are described in Section 3. Section 4 shows the results of the three reading tasks and the linguistic feature analysis, whose implications are then discussed in Section 5¹.

2. Related Work

Classic readability formulae were designed in the early 20s to detect the complexity of a text in relation to educational stages [8, 9, 10]. Since these formulae took only raw linguistic features into consideration, such as word and sentence length, they were not fully reliable [11, 12]. Advancements in Machine Learning (ML) led to the implementation of more complex models that are informed by a wider and less superficial set of linguistic features [13, 14].

As for Italian, readability formulae started being implemented only in the late 80s, by [15] and [16], authors of the Flesch-Vacca formula and the Gulpease Index, respectively. The first ML-based index for Italian is Read-It [17]: The index measures the probability of a text to be labelled as complex by an SVM trained on newspaper articles on the basis of linguistic features ranging from the lexical to the syntactic level. Inspired by Coh-Metrix [18], [19] considered also discourse-level features, e.g., cohesion, to design Coease, an index for texts related to the educational domain. Finally, CTAP is a web-based readability tool available also for the Italian

¹Anonymized and aggregated data, and code are available at https://github.com/Unipisa/ita_admin_user_study

language [20], which extracts 253 different linguistic features. Such features are not related to any reference corpora, and it is up to the user to give an interpretation to the extracted values.

Some models for Italian text readability were also designed for targeted reader groups, such as Italian second language learners. [6] implemented MALT-IT2, a tool that automatically classifies texts by assigning one of the proficiency levels of the Common European Framework of Reference for Languages (CEFR). This tool is based on a SVM, trained on raw, lexical, morphosyntactic, syntactic, and discursive features.

For what concerns the administrative language, [21] automatically analyzed several linguistics features extracted from a parallel corpus composed of administrative texts and their simplified versions. The author aimed at distinguishing between features that are expression of the intrinsic complexity of the administrative language and those used in the so-called "bureaucratese", a term used to indicate the "artificial" and "obscure" style that sometimes characterizes the administrative writing [22]. [23] extracted complexity features from about 100 institutional texts for foreigners, showing the gap between the language used in these texts and those tailored for Italian second language speakers. This gap was also confirmed by a comprehension test carried out on specific target readers [24].

A way to detect which features best predict the readability of a certain text for a certain target is in fact to collect data from human participants. [25] built two models trained on several linguistic features to predict pairwise scores for text comprehension and reading time collected through online crowdsourcing. [26] showed that scrolling interactions are predictive of text readability also for specific target users, such as English second language speakers.

User studies were also conducted for the administrative domain. [27] collected judgments on readability from public administration staff and extracted linguistics features from the analyzed texts in French. [28] analyzed complexity features for administrative texts in German, and evaluated their model through the correlation of such features with non-experts' judgments on readability.

To the best of our knowledge, this is the first study on readability that focuses on the administrative Italian language and addresses multiple subject groups, such as second-language and elderly Italian readers in a digital context.

3. Experimental settings

In the comprehension task, three different groups of participants were involved: Italian secondlanguage (L2) speakers, Italian first-language speakers who were older than 60 (elderly), and Italian native speakers younger than 60 years old with a medium-high literacy level (control). All participants had to perform the test via smartphone.

3.1. Materials

We collected four portions of texts extracted from documents of different nature, covering various topics related to public administration, and published by official websites of Italian city halls from all over the country (see Table 1). Texts with a similar distribution of specific

Table 1

Details about the selected texts, i.e., the city hall that published the document, the document topic and type, and linguistic features, i.e., total number of tokens, Type/Token Ratio, the average length of tokens in characters, the average length of sentences in tokens, and the percentage of words belonging to the Base Vocabulary (BV).

Text	City	Торіс	Туре	#Tok	TTR	Tok len	Sent len	BV
A	Naples	Benefits	Web, public call	264	0.42	4.43	26.86	69.91
В	Rome	Civil registry	FAQ	259	0.48	4.45	27.3	68
С	Bari	Mobility	Act, regulation	219	0.48	4.55	29.18	64.49
D	Trento	Public housing	Act, public call	271	0.49	4.54	28.9	67.41

Table 2

The table on the left shows the operations applied on the administrative texts. The table on the right shows the motivations behind each simplification operation.

Operation	Count	Motivation	Count
Split	7	Uncommon and formal terms	41
Reordering	10	Parenthetic clauses and asides	20
Merging	0	Long and wordy sentences	15
Insert	16	Impersonal and passive sentences	10
Delete	26	Prepositional/conjunctive phrases	7
Transformation	75	Abbreviations and acronyms	4
– Lexical Subst. (word level)	20	Verb periphrasis	4
- Lexical Subst. (phrase level)	38	Pleonastic and stereotyped phrases	3
- Anaphoric replacement	2	Improper cohesion between sentences	3
- Noun to Verb	5	Abbreviations acronyms	3
– Verbal Voice	2	Fixed textual organization	2
– Verbal Features	8	Other	22
Total	134	Total	134

Table 3

A sentence of Text B before () and after (<a>) a simplification operation was applied.

linguistic features² were selected, such as the length of the whole text and the average length of sentences (in tokens), the token/type ratio, and the percentage of tokens belonging to the Base Vocabulary of the Italian language [30].

A simplified version of each text was then created based on the features of the Italian administrative language that were singled out by [21]. The presence of these features in the administrative texts is claimed not to be justified based on the complexity of the public bodies

²Texts were analyzed by using the Python NLP library Stanza [29].



Figure 1: On the left, the percentage of participants who choose the simplified (S) version of each administrative text (A, B, C, D) over its original (O) counterpart when asked "Which text is simpler?". On the right, the distribution of the degree of similarity between original and simplified versions of each text on a Likert scale given by subjects' answers to the question "How similar are the two texts?".

and the procedures they describe, nor on the performative nature of such language [31], but to rather lead to "bureaucratese" [22]. The adopted annotation schema was firstly presented by [32] and then used by [33] for the annotation of SIMPITIKI, where a single simplification operation was performed on each sentence. We adapted this schema by annotating all the simplification operations applied to each sentence and by indicating the motivation for the performed operation, i.e., the detected linguistic feature to be simplified (see statistics in Table 2 and Table 3, and see Table 3 for an example of the simplification operation). The simplification was validated through a test, which involved 43 Italian native speakers. For each originalsimplified pair, the participants were asked which text was the simpler one and how similar they were (Fig. 1), to assess if the information contained in the original text was preserved in the simplification process. Four multiple-choice questions for each text were formulated by analyzing their macro and micro informative structure. Drawing inspiration from [34], we split each text into sentences (microstructure) and, at a higher and more abstract level, we split the text according to the subject matter (macrostructure). Then, we checked that the obtained micro and macro structures were preserved after the simplification process and we selected the portion of texts on which to test the participants. This ensured that questions covered each element of the macrostructure. The same questions were asked to participants reading the original or the simplified version of each text. We choose to ask multiple-choice questions, since they are usually adopted in comprehension tasks [26], have already been used for an effective simplification of texts [3] and are widely adopted for assessing the proficiency of second language learners [35]. Item readability was then analyzed employing Read-It [17], which provides a readability score at the sentence level, and items were then simplified accordingly.

3.2. Participants

We recruited a total of 111 participants, 47 for the group L2, 29 for the elderly group and 35 for the control group. For the control and elderly group, we only included the participants who were born and were living in the Tuscany region, in order to limit the influence of regional varieties of Italian on the comprehension of texts. People without a high school diploma were excluded from the control group.³ For what concerns the L2 group, we eventually included only non-native speakers of Italian with A2 and B1 language certificates (according to the CEFR) and currently residing in Italy, as well as and non-native speakers of Italian without any proficiency certificate who had lived in Italy for at least 5 years. We assumed that people who had been living in Italy for at least 5 years had higher chances to be frequently exposed to public administration language in everyday life. By filtering participants based on these criteria, we were eventually left with 86 subjects: 26 for the group L2, 29 for elderly, and 31 for the control group. L2 group participants were aged 18 to 55 and 69.2% of them were female. They were born in Morocco (15.38%), Senegal (11.54%), Albania (11.54%), Georgia (7.69%), Indonesia (7.69%), Nigeria (7.69%), Russia (7.69%) and other countries (30.77%). For what concerns the education level, 61.54% of participants had at least a high school diploma. Elderly participants' age ranged from 60 to 82 and the 51.72% of them were female. In this case, only 55.17% of participants had at least the high school diploma.

We collected such information through a demographic questionnaire. We grouped the questions into different topics, starting with those regarding all the participants [36]. The questionnaire also included questions about digital literacy, familiarity with the administrative domain, education and reading habits.

3.3. Test implementation and design

We implemented the test on a multiple-step web page, using HTML, CSS and JavaScript. Choices about the test layout, such as line spacing, font type, and dimension were made following the Design Guidelines for Public Administration Web Sites and Services⁴ provided by AGID (Agency for Digital Italy).

We administered the test in a hybrid format, partly in person and partly remotely. Each participant read two texts, one presented in its original version and the other in its simplified version. The 8 texts (4 original and 4 simplified) that were obtained through the procedure described in Section 3.1 were rotated across participants so that no participant saw the same text in both conditions. In the four resulting lists, the order in which the original and simplified text appeared was counterbalanced. Firstly, participants answered to the demographic questionnaire and then completed the comprehension task for one text at a time.

We showed each single question on a different step page, right below the related text. For each multiple choice question, we provided a key, two distractors, and the "I don't know" option, to try to limit participants' guessing.

3.4. Feature Extraction

We extracted 13 linguistic features from each text to assess which ones mostly affected participants' reading speed and comprehension. Features were selected based on existing literature on the readability of administrative language [21, 2], Italian Second Language Learning [37, 6],

³In 2001, the average year of scholar education per person was about 11,7 years [7]. People without a high school diploma, which in Italy is obtained after 13 years of scholar education, were thus considered having a low literacy level.

⁴https://docs.italia.it/italia/design/lg-design-servizi-web/it/versione-corrente/index.html

and language processing in elderly people [38]. Such features are related to different linguistic levels:

- *Raw.* Average length of sentences in tokens;
- *Lexical.* Percentage of words belonging to the Fundamental Vocabulary⁵, average number of multiword units and entities per sentence, average number of collateral technicisms⁶ per sentence;
- *Psycholinguistic*. Percentage of abstract nouns ⁷;
- Morphosyntactic. Percentage of deverbal nouns, participles verbs, and indicatives verbs;
- *Syntactic.* Average depth of the parsing tree, ratio between subordinate and total number of clauses, average length of the prepositional chains;
- *Discourse and Style.* Average number of asides and parenthetical expressions per sentence, average number of common nouns among adjacent sentences.

3.5. Preprocessing

Participants' performance in comprehension questions was measured in terms of error rate. Sociolinguistic data from the demographic questionnaire were operationalized in three different indices. We measured the use of smartphone by averaging for each participant the Likert values for the questions "How many hours per day you spent on the phone last week?" and "Do you use the smartphone to work or study (i.e. reading books and articles, writings, making analysis, doing some research, etc.)". The second question was motivated by the fact that people used to quick interactions with their smartphone struggle in focusing on longer task [39]. This averaging procedure resulted in the *digital* index. Familiarity with the administrative domain was analyzed though the admin index, obtained by averaging Likert values for the questions "How often have you paid taxes, filled forms or asked for financial support in the last month?" and "How often did you read forms, notices, call for applications, regulations or similar in the last month?". Finally, we merged information about education and reading habits into the readedu index. This index was obtained by averaging responses to the questions: "How many books have you read in the last year?", "How often have you read newspapers and magazines (also online) in the last month?", and "Which is your highest degree?"⁸. Such indices were then centered and scaled.

4. Results

Analyses were run to assess if there was any difference between the three subject groups (L2, elderly, control) in the comprehension of administrative texts and of their easier-to-read versions, simplified according to the sole linguistic features that are specifically related to the

⁵A subset of the Italian Base vocabulary.

⁶Collateral technicisms are terms related to sectorial or special languages used to give the text a high linguistic registry but that lack specific communicative function.

⁷Given the lack of annotated data for the administrative domain, nouns' lemma were manually annotated by a linguist as abstract or concrete. We then computed the percentage of abstract nouns for each portion of text.

⁸For L2 and elderly, the index is only based on the answers to the first two questions for those participants who did not precised their educational level.



Figure 2: Interaction between text complexity and *digital* index for the three subject groups in terms of error rates. This interaction results from a linear mixed model where the error rate is a function of group, complexity, *admin, digital, readedu* index (as fixed effects), participants, and text (as random effects).



Figure 3: Main effect of the *admin* index for the three subject groups in terms of error rates. This effect results from a linear mixed model where the error rate is a function of group, complexity, *admin*, *digital*, *readedu* index (as fixed effects), participants, and text (as random effects).

administrative language. Furthermore we analyzed how familiarity with the administrative domain, digital literacy, reading habits, and education affected comprehension, by using the three indices described in Section 3.5.

4.1. Error rates

We were interested in detecting any significant difference among groups in relation to participants' error rate when answering the comprehension questions. A significant interaction



Figure 4: The significant trend in the interaction between text complexity and *digital* index on L2 participants' error rates. Participants with lower digital literacy where less accurate in answering questions on simplified texts. This interaction results from a linear mixed model where the error rate is a function of complexity, *admin* and *digital* index, educational level, reading habits, language certificate level, years lived in Italy, years spent studying Italian (as fixed effects), and participants (as random effects).

emerged between group, complexity, and the *digital* index (p = .012). While L2 speakers were overall less accurate in answering comprehension questions, they specifically made more errors on simpler texts when digital exposure was lower (see Fig. 2). We also observed a main effect of the *admin* index on participants' error rate (p = .040). L2 speakers' error rate was higher for both original and simplified texts when their familiarity with the administrative domain was lower (see Fig. 3). By contrast, the performance of participants seemed not to be impacted by reading habits and education level.

4.1.1. Focus on L2

In a subsequent step of our analysis, we zoomed in on L2 speakers only, to shed light on the role of L2-specific demographic variables in text comprehension. In particular, we analyzed the interaction between text complexity, the number of years spent in Italy, Italian proficiency (i.e., the language certificate), the years employed in studying Italian as a second language, and the use of Italian when communicating at home, at work, and with friends. In this case, texts were not included in the random effects. Finally, for this analysis we considered the education level separated from the information about reading habits.⁹ In line with the results obtained on the three groups, by analyzing L2 participants we observed a significant trend concerning the interaction between the *digital* index and text complexity on error rates (p = .085).¹⁰ Namely, the error rate was higher for those participants with lower digital literacy when answering questions on simplified texts (see Fig. 4).

⁹This analysis involved 24 participants out of 26: two participants where excluded since they did not indicated which was their education level.

¹⁰We considered only participants as random effect here.

4.2. Feature analysis

We conducted a preliminary analysis on the linguistic features we described in Section 3.4, to detect which ones are more predictive of participants' error percentage in the comprehension task. We performed a Principal Component Analysis (PCA) on the linguistic features to reduce the dimensionality of the data while preserving as much as possible of their original information. The first two principal components cumulatively accounted for 71% of the variance of the original variables. When inspecting the loadings matrix, PC1 seemed to be mostly influenced by morphological features, i.e., the number of participles and indicative verbs, and by features that affect the sentence length: the average number of multiwords units and entities per sentence, the average length of the prepositional chains, and the average length of sentences in tokens. By contrast, PC2 appeared to be influenced by the average number of common nouns in adjacent sentences, which is related to text cohesion, and morphosyntactic features, i.e., the average depth of the parsing tree per sentence and the number of deverbal nouns.

When predicting participants' error rates,¹¹ we observed an effect of PC2 on each group, and in particular an almost-significant effect on L2 participants (p = .060). As shown in Figure 5, when PC2 is higher, the error rate increases for the three groups, especially for L2.

A significant effect is observed in the interaction among group, complexity, and PC1 (p = .030). Figure 6 shows that L2 participants' error rates increases along with PC1 values for questions regarding simplified texts. An higher error rate is registered also for control participants, but such increment is not significant.

It is paramount to underscore the preliminary nature of this exploratory analysis. Future contributions will better clarify the role of specific linguistic features through finer-grained and targeted analyses.

4.3. Interactive task

The test originally included also an interactive task, where we asked participants to underline the portions of text they perceived as more difficult. By doing so, we wanted to compare response accuracy against a more subjective judgment of the text complexity. Participants were free not to underline any portion of texts. However, we tried to encourage the readers' participation by also providing a tutorial, to reduce the limitations posed by the low familiarity of some participants with digital devices. Unfortunately, only a few participants underlined a portion of text (10 L2, 6 elderly and 14 control participants) and thus, we did not carry out any statistical analysis on this data. We do believe that this happened because most participants did not perceive any part of the text as complex. However, when including data from participants that were left out in the initial filtering, we noticed that on average, L2 speakers underlined as many portions of text as control participants in the simpler condition, whereas they underlined more passages on original texts. However, when inspecting the underlined text more closely, we noticed that L2 speakers underlined fewer tokens, i.e., they tended to underline single words rather than entire phrases (see Fig. 7 for an example). We do not discuss data related to the elderly group, since only six people took part in the task.

¹¹We considered only the participants as random effect here.



Figure 5: Main effect of cohesion and morphosyntactic related features expressed by PC2 on groups' error rates. The higher is PC2 the higher is the error rate for the three groups, especially for L2 participants. This effect results from a linear mixed model where the error rate is a function of group, complexity, PC1, PC2 (as fixed effects), and participants (as random effects).



Figure 6: Interaction between morphological and sentence length features expressed by PC1, group and complexity in terms of error rates. This interaction results from a linear mixed model where the error rate is a function of group, complexity, PC1, PC2 (as fixed effects), and participants (as random effects).

5. Discussion

The higher error rate registered in L2 participants revealed a significant difference with respect to elderly and control participants in text comprehension. However, in light of our current results, we could not confidently conclude that text complexity affected participants' comprehension



Figure 7: The picture shows the portions of the simplified version of text "D" underlined by participants of the three subject groups.

across subject groups. The fact that participants' comprehension did not improve when dealing with simplified texts could highlight the need for a simplification strategy that focuses more on linguistic features specific to each target group.

Furthermore, we saw that participants' background affected text comprehension to some extent. For example, digital literacy seemed to specifically affect L2 learners' comprehension of simplified texts. Namely, participants who used their smartphone less frequently, in particular not for reading and writing, struggled more when answering the questions.

When focusing only on the L2 group, we also found a marginal effect of digital literacy on participants' error rates, whereas we did not register any effect of proficiency as assessed by a certificate or concerning years spent in Italy.

Furthermore, familiarity with the administrative domain seems to play a role in subjects' understanding. The lower the *admin* index, and therefore the exposure to administrative texts and public administration, the less accurate were L2 participants' in answering questions related to simplified texts.

The analysis of linguistic features showed that, regardless the text readability, each group and L2 in particular struggles in understanding sentences with a low number of common nouns among adjacent sentences and with a complex syntactic structure. In fact, by looking at the loading matrix, PC2 grows with sentences with a deeper parsing tree and when the percentage of deverbal nouns in the text decreases. Deverbal nouns, in fact, tend to condense information, even though this produces further complexity related to their abstractness and high information density (e.g., "la percezione dell'integrazione salariale" [*the receipt of wage subsidies*] instead of "i lavoratori che ricevono l'integrazione salariale" [*workers that receive wage subsidies*]).

For what concerns the analysis of the linguistic features expressed by PC1, we also observed that L2 participants also struggle when reading simple texts with long sentences. Furthermore, we could say that L2 participants find difficult understanding texts with compound tenses of verbs, since their error rate increase with a high number of participle verbs, and a lower number of indicative verbs. Moreover, L2 group's comprehension is also affected by the lexicon: their error rate increase with a higher number of multiwords and entities. According to what we observed in the interactive task, only this lexical aspect of complexity is pointed out by L2 participants', who seem to perceive lexical features as more complex than syntactic ones.

PC1's linguistic features do not seem to have an effect when participants deal with original texts. In particular, the presence or absence of such features does not help participants to answer questions correctly when dealing with original and - thus - more complex texts. On the contrary, with simplified texts, L2 participants' error rate is higher than for the other two groups: elderly participants seem to benefit from PC1's features, whereas for control, and even more so for L2, texts may require further simplification based on such features.

6. Conclusions

The goal of the present work was to investigate how differently second language learners of Italian and elderly Italian native speakers read and comprehend administrative texts in a digital context. We designed a comprehension task to be completed online which allowed us to collect the error rate in answering comprehension questions for each selected text. Furthermore, we wanted to assess if other linguistic features affected the comprehension of participants other than those strictly related to the administrative domain. Such features were used to simplify the selected texts, and these simplified versions were also shown within the two tasks.

We observed a difference in comprehension for L2 participants compared to the elderly and control group, and found out that text complexity did not affect text comprehension across the three groups. However, participants' background had some effect on the comprehension process, especially for what concerns speakers' digital literacy and their familiarity with the administrative domain. Finally, we detected the effect of specific linguistic features on participants' comprehension for all the three groups.

In future contributions, we would like to further investigate the registered effect of digital literacy on L2 participants' comprehension. For example, it would be interesting to analyze the comprehension of texts when read on paper and on digital devices.

We plan to increase our participant sample and the number of selected texts, in order to have a dataset that is more representative of the various types of documents that can be found in the administrative domain. Furthermore, we intend to include also other groups of target users, like low-literacy people. Then, we aim to propose a simplification procedure for administrative texts that takes into account a larger set of linguistic features focused on the need of specific target groups. The obtained data might be used to build a neural model for the simplification of administrative texts that takes into account the needs of such groups. For example, this would be possible by leveraging *Controllable simplification* [5, 40, 41], which aims at constraining the neural model output by using special tokens with information on sentence and word length, Levenshtein distance between input and generated sentence, and so on.

References

- [1] D. Fortis, Il dovere della chiarezza. quando farsi capire dal cittadino è prescritto da una norma, Rivista Italiana di Comunicazione Pubblica 25 (2005) 82–116.
- [2] M. A. Cortelazzo, Il linguaggio amministrativo: principi e pratiche di modernizzazione, Studi superiori, Carocci, 2021.
- [3] M. Vedovelli, T. De Mauro, Dante, il gendarme e la bolletta: la comunicazione pubblica in Italia e la nuova bolletta ENEL, Laterza, 1999.
- [4] F. Dell'Orletta, G. Venturi, S. Montemagni, Genre-oriented readability assessment: A case study, in: Proceedings of the Workshop on Speech and Language Processing Tools in Education, 2012, pp. 91–98.
- [5] L. Martin, B. Sagot, E. de la Clergerie, A. Bordes, Controllable sentence simplification, arXiv preprint arXiv:1910.02677 (2019).
- [6] L. Forti, G. Grego, S. Filippo, V. Santucci, S. Spina, Malt-it2: A new resource to measure text difficulty in light of cefr levels for italian l2 learning, in: 12th Language Resources and Evaluation Conference, The European Language Resources Association (ELRA), 2020, pp. 7206–7213.
- [7] T. De Mauro, L'educazione linguistica democratica, Gius. Laterza & Figli Spa, 2018.
- [8] B. A. Lively, S. L. Pressey, A method for measuring the vocabulary burden of textbooks, Educational administration and supervision 9 (1923) 389–398.
- [9] R. Flesch, A new readability yardstick, Journal of Applied Psychology 32 (1948) 221.
- [10] J. P. Kincaid, R. P. Fishburne, R. L. Rogers, B. S. Chissom, Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel, in: Institute for Simulation and Training, 1975.
- [11] L. Si, J. Callan, A statistical model for scientific readability, in: Proceedings of the tenth international conference on Information and knowledge management, 2001, pp. 574–576.
- [12] K. Collins-Thompson, Computational assessment of text readability: A survey of current and future research, ITL-International Journal of Applied Linguistics 165 (2014) 97–135.
- [13] S. E. Schwarm, M. Ostendorf, Reading level assessment using support vector machines and statistical language models, in: Proceedings of the 43rd annual meeting of the Association for Computational Linguistics (ACL'05), 2005, pp. 523–530.
- [14] S. Vajjala, D. Meurers, On improving the accuracy of readability classification using insights from second language acquisition, in: Proceedings of the seventh workshop on building educational applications using NLP, 2012, pp. 163–173.
- [15] V. Franchina, R. Vacca, Taratura dell'indice di flesch su testo bilingue italianoinglese di unico autore, in: Atti dell'incontro di studio su: Leggibilità e Comprensione, Linguaggi, a. III, 1986, pp. 47–49.
- [16] P. Lucisano, M. E. Piemontese, GULPEASE: una formula per la predizione della difficoltà dei testi in lingua italiana, La Nuova Italia (1988).
- [17] F. Dell'Orletta, S. Montemagni, G. Venturi, Read-it: Assessing readability of italian texts with a view to text simplification, in: Proceedings of the second workshop on speech and language processing for assistive technologies, 2011, pp. 73–83.
- [18] A. C. Graesser, D. S. McNamara, M. M. Louwerse, Z. Cai, Coh-metrix: Analysis of text on cohesion and language, Behavior research methods, instruments, & computers 36 (2004)

193-202.

- [19] S. Tonelli, K. M. Tran, E. Pianta, Making readability indices readable, in: Proceedings of the First Workshop on Predicting and Improving Text Readability for target reader populations, 2012, pp. 40–48.
- [20] N. Okinina, J.-C. Frey, Z. Weiss, CTAP for Italian: Integrating components for the analysis of Italian into a multilingual linguistic complexity analysis tool, in: Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), 2020, pp. 7123–7131.
- [21] D. Brunato, A study on linguistic complexity from a computational linguistics perspective. a corpus-based investigation of italian bureaucratic texts, Ph.D. thesis, University Of Siena, 2015.
- [22] S. Lubello, Il linguaggio burocratico, Le bussole, Carocci, 2014.
- [23] G. Lombardi, La leggibilità dei testi istituzionali italiani destinati agli stranieri, in: M. E. Favilla, S. Machetti (Eds.), Lingue in contatto e linguistica applicata: individui e società, AItLA Associazione Italiana di Linguistica Applicata, Bologna, 2021, pp. 199–214.
- [24] G. Lombardi, Capire i documenti in L2: dall'analisi della comprensibilità di un corpus di testi istituzionali per stranieri alla sperimentazione di approcci didattici e linguistici, Ph.D. thesis, Università degli Studi di Genova, 2020.
- [25] S. A. Crossley, S. Skalicky, M. Dascalu, Moving beyond classic readability formulas: New methods and new models, Journal of Research in Reading 42 (2019) 541–561.
- [26] S. Gooding, Y. Berzak, T. Mak, M. Sharifi, Predicting text readability from scrolling interactions, arXiv preprint arXiv:2105.06354 (2021).
- [27] T. François, L. Brouwers, H. Naets, C. Fairon, Amesure: a readability formula for administrative texts (amesure: une plateforme de lisibilité pour les textes administratifs)[in french], in: Proceedings of TALN 2014 (Volume 2: Short Papers), 2014, pp. 467–472.
- [28] T. vor der Brück, S. Hartrumpf, H. Helbig, A readability checker with supervised learning using deep indicators, Informatica 32 (2008) 429–435.
- [29] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, C. D. Manning, Stanza: A Python natural language processing toolkit for many human languages, arXiv preprint arXiv:2003.07082 (2020).
- [30] T. De Mauro, I. Chiari, Il nuovo vocabolario di base della lingua italiana, Internazionale, 28/11/2020. (2016). URL: https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/ 23/il-nuovo-vocabolario-di-base-della-lingua-italiana.
- [31] A. Fioritto, Manuale di stile. Strumenti per semplificare il linguaggio delle amministrazioni pubbliche, Il mulino, 1997.
- [32] D. Brunato, F. Dell'Orletta, G. Venturi, S. Montemagni, Design and annotation of the first italian corpus for text simplification, in: Proceedings of The 9th Linguistic Annotation Workshop, 2015, pp. 31–41.
- [33] S. Tonelli, A. P. Aprosio, F. Saltori, Simpitiki: a simplification corpus for italian., in: Proceedings of the Third Italian Conference on Computational Linguistics (CLiC-it), 2016.
- [34] W. Kintsch, T. A. Van Dijk, Toward a model of text comprehension and production, Psychological review 85 (1978) 363–394.
- [35] M. Barni, A. Villarini, La questione della lingua per gli immigrati stranieri: insegnare, valutare e certificare l'italiano L2, volume 39, FrancoAngeli, 2001.
- [36] D. A. Dillman, J. D. Smyth, L. M. Christian, Internet, phone, mail, and mixed-mode surveys: The tailored design method, John Wiley & Sons, 2014.

- [37] L. Forti, A. Milani, L. Piersanti, F. Santarelli, V. Santucci, S. Spina, Measuring text complexity for italian as a second language learning purposes, in: Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, 2019, pp. 360–368.
- [38] S. Norman, S. Kemper, D. Kynette, Adults' Reading Comprehension: Effects of Syntactic Complexity and Working Memory, Journal of Gerontology 47 (1992) 258–265.
- [39] L. E. Annisette, K. D. Lafreniere, Social media, texting, and personality: A test of the shallowing hypothesis, Personality and Individual Differences 115 (2017) 154–158.
- [40] J. Mallinson, M. Lapata, Controllable sentence simplification: Employing syntactic and lexical constraints, arXiv preprint arXiv:1910.04387 (2019).
- [41] M. Maddela, F. Alva-Manchego, W. Xu, Controllable text simplification with explicit paraphrasing, arXiv preprint arXiv:2010.11004 (2020).