

Simplifying Administrative Texts for Italian L2 Readers with Controllable Transformers Models: A Data-driven Approach

Martina Miliani¹, Fernando Alva-Manchego² and Alessandro Lenci¹

¹Dipartimento di Filologia, Letteratura e Linguistica, Università di Pisa, Italy

²School of Computer Science and Informatics, Cardiff University, UK

Abstract

This paper presents a data-driven study focused on the automatic simplification of in-domain texts for specific target readers, which is “controlled” through data collected from behavioral analysis. We used these data to create Admin-It-L2, a parallel corpus of original-simplified sentences in the Italian administrative language, in which simplifications are aimed at Italian L2 speakers. Then, we used this corpus to test controllable models for text simplification based on Transformers.

Although we obtained a high SARI score of 39.24, we show that this datum alone is not fully reliable in evaluating text simplification.

Keywords

Automatic Text Simplification, Transformers, Italian L2, Italian Bureaucratic Language

1. Introduction

Reading a text in a language that is easy to understand becomes necessary if the information it conveys is crucial to people’s daily lives. This is the case for acts and communications of Public Administrations (PA). In Italy, despite institutions repeatedly encouraging clear writing [1], PA texts have not abandoned the stylistic figures of “bureaucratese” [2]. Suffering the most from the effects of these linguistic choices are speakers with language disparity. In particular, Italian L2 speakers deal with the Italian bureaucracy, for example, to obtain a visa for which A2-level language certification is required. Yet, comprehension of administrative texts is attested only with the C2 level, according to the CEFR.

This paper presents a data-driven approach to simplification, which is “controlled” through data collected from behavioral analysis. We use these data to create Admin-It-L2, a parallel corpus of original-simplified sentences in the administrative language, in which simplifications are aimed at Italian L2 speakers. Then, we employ this corpus to test computational models. The adopted approach exploits the potential of Transformer-based multilingual models, which permit the use of training data also in languages other than Italian [3]. Leveraging corpora in different languages allows us to overcome the limited availability of parallel corpora in Italian and the adminis-

trative domain in particular.

To sum up the main contributions of this paper:

- We release Admin-It-L2,¹ a parallel corpus in the Italian administrative language with simplifications aimed at Italian L2 speakers (Sec. 3).
- To the best of our knowledge, this is the first attempt to create a controlled simplification model specific to the Italian administrative language (Sec. 4). In particular, we train controllable models (Sec. 5) with multilingual data, to simplify texts aimed at Italian L2 speakers, thanks to data collected with behavioral analysis.
- We use Admin-It-L2 to test such models and we evaluate them with the available metrics for Automatic Text Simplification (ATS). We also manually analyze the produced simplification to assess the validity of such metrics for the Italian bureaucratic language (Sec. 6).

Finally, we also show the results of a preliminary and exploratory experiment with ChatGPT.² Based on GPT-3.5, this Large Language Model (LLM) gained popularity also thanks to the impressive performance it reaches in several NLP tasks [4]. Thus, we decided to give a brief overview of its potential and limitations when applied to the simplification of Italian administrative texts for Italian L2 speakers (Sec. 7).

2. Related Work

ATS aims at simplifying a text while maintaining its meaning [5]. Since the spread of Neural Networks, models’ architectures are mostly taken from “Neural Machine

CLiC-it 2023: 9th Italian Conference on Computational Linguistics,

Nov 30 – Dec 02, 2023, Venice, Italy

✉ martina.miliani@fileli.unipi.it (M. Miliani);

alvamanchego@cardiff.ac.uk (F. Alva-Manchego);

alessandro.lenci@unipi.it (A. Lenci)

ORCID 0000-0003-1124-9955 (M. Miliani); 0000-0001-6218-8377

(F. Alva-Manchego); 0000-0001-5790-4308 (A. Lenci)

© 2023 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)

¹<https://github.com/Unipisa/admin-it-l2>

²<https://openai.com/blog/chatgpt>

Translation” (NMT) [6]. For example, to build their simplification model, [7] started from OpenNMT [8], consisting of an encoder-decoder with two layers of LSTM. [9] followed a similar approach and trained a model for the Italian language thanks to data augmentation techniques.

In some of the most recent models, simplification can be “controlled” to generate texts for specific groups of readers. CROSS [10] proposed a Transformer-based model able to control the level of simplicity and the type of applied simplification. To this aim, at a lexical level, they marked the token to replace, whereas, at a syntactic level, they leveraged templates. ACCESS [11] introduced the use of control tokens to bind the simplification with specific attributes, such as the amount of paraphrased content and lexical and syntactic complexity. [12] used ACCESS to simplify Italian texts and trained it on the automatic translation of NewsEla [13] (see Sec. 5.1), obtaining promising results. MUSS [14] presented an unsupervised method to collect parallel data to train a model based on ACCESS adopting BART and its multilingual version, mBART [15]. The authors’ goal was to deal with the paucity of available parallel corpora in languages different from English. To this aim, we leverage cross-language and fully translated data, by exploiting the ability of multilingual models to use linguistic knowledge from different languages. [16] also used ACCESS but replaced BART with T5 [17]. With this controllable model, [18] achieved SOTA performance for SARI on Spanish texts.

Differently from the described work, we focus our simplification on in-domain texts written in the Italian bureaucratic language. Furthermore, we leverage data on Italian L2 speakers to create a test set to control the simplifications generated by Transformer-models. Such a dataset is described in the next section.

3. Admin-It-L2

In this section, we present Admin-It-L2, a parallel corpus of complex-simplified sentences in the Italian administrative language for ATS aimed at Italian L2 speakers.

We manually simplified 134 sentences by focusing on the linguistic traits that emerged from a comprehension test conducted over 86 participants [19], involving also Italian L2 speakers (30,2%), and elderly Italian native speakers (33,72%). The participants were asked to answer questions about an original text and a simplified version of another text, both from the administrative domain. Such simplification focused only on the typical traits of the bureaucratic language [20].

By analyzing the participants’ answers to the comprehension test, the authors observed that L2 struggled more when reading simple texts with long sentences, long prepositional chains, a high number of participle

verbs, and a lower number of indicative verbs. Their answer error rate also increased with a higher number of multi-words and entities.

Admin-It-L2 counts 134 pairs of sentences extracted from the texts used in this study (34 sentences) and from **Admin-It_{RS}** (100 sentences), a subsection of Admin-It, a parallel corpus of Italian administrative texts [3]. The original sentences of this subset were selected from websites of Italian municipalities, and from the longest sentences of the PaWaC Corpus [21]. [3] manually rewrote the sentences simplifying them both at lexical and syntactic levels. This simplification as well was only based on the typical traits of the administrative language [2, 20].

The similar nature of the simplification applied, allowed us to further simplify the simple sentences of Admin-It_{RS} and the simple sentences employed in the comprehension test (together referred to as *Admin-It_{simp}* in Table 1) by considering the results of the study related to Italian L2 speakers. The annotation was conducted by a single annotator among the authors, and validated through a quantitative analysis shown in Table 1.³

4. The Controllable Simplification Model

Our models are based on the implementation released by [16].⁴ The two authors used ACCESS and replaced BART with T5, a model with an encoder-decoder architecture pre-trained on various tasks through supervised and unsupervised approaches [17].

We adopted the multilingual version of T5, mT5 [22], for all our experiments. [23] observed that multilingual models proved to be better for tasks like text summarization or switching from a formal to a more informal language. The authors suggest that this is probably due to the different distributions of linguistic data used for training the model. In particular, the multilingual model would perform better with automatically translated texts.

Unlike [16], due to hardware constraints, we employed the basic mT5 model and fine-tuned it on a batch of size 16. We retained the values of all other parameters, including the number of epochs (5). Fine-tuning was performed on an NVIDIA A100 Tensor Core GPU, with 40 GB of RAM.

5. Experimental Settings

Given the limited availability of parallel corpora in Italian, we created training sets for fine-tuning by adopting two

³Details about the simplification process are presented in Appendix A.

⁴https://github.com/KimChengSHEANG/TS_T5

Table 1

Some statistics of Admin-It-L2 in the original version of the sentences (*Admin-It-L2-comp*), the simplified version based on administrative language traits only (*Administ-semp*) and the simplified version of the sentences in Admin-It-L2 (*Admin-It-L2-semp*).

	<i>Admin-It-L2-comp</i>	<i>Administ-semp</i>	<i>Admin-It-L2-semp</i>
<i>Avg token per sentence</i>	52.11	29.86	24.83
<i>Avg character per token</i>	4.65	4.52	4.52
<i>Number of sentences</i>	179.0	276.0	270.0
<i>Participial verbs (%)</i>	22.56	7.39	6.83
<i>Infinitive verbs (%)</i>	17.69	21.72	22.04
<i>Finite-mode verbs (%)</i>	38.85	44.68	49.87
<i>Indicative verbs (%)</i>	31.92	42.55	49.48
<i>Avg dept of the syntactic tree</i>	5.17	3.87	3.07
<i>Avg length of prepositional chains</i>	1.65	1.38	1.34
<i>Levenshtein Distance</i>	-	170.44	202.73

main strategies. The first involves the full translation into Italian of parallel English corpora with Google Translate.⁵ We translated only from English since translation systems for this language generally perform better, thanks to the amount of data available for model training. The second strategy involves the creation of cross-language datasets, in which the original sentences are in English, whereas their simplified counterparts were translated into Italian. On the one hand, translating simple texts is an easier task for machine translation systems [24, 25]. On the other hand, we intended to take advantage of the capabilities of multilingual generative models to succeed in applying linguistic knowledge from different languages. This potential has previously been tested for summarizing documents in different languages [26].

5.1. Datasets

We used Admin-It-L2 as the test set, to assess whether the models succeed in producing simplifications close to the needs of Italian L2 speakers. To train the controllable models we combined the following datasets:⁶

NewsEla (NewsEng) consists of newspaper articles manually simplified by experts according to different degrees of complexity. NewsEla was then automatically sentence-aligned by [27].

OneStopEnglish (OSE) contains articles from the British newspaper The Guardian rewritten by teachers in three levels of readability for English L2 learners [28].

SimPA is a parallel corpus of sentences in English administrative language [29]. The authors applied two steps of simplification: lexical and syntactic.

Spanish NewsEla (NewsEs) sentence-aligned version was created by [9]. Documents were translated from English, and then simplified manually.

Terence contains short stories in Italian for children rewritten by a group of experts [20].

Teacher contains teaching materials manually simplified by a teacher for Italian L2 learners [20].

PaCCSS-IT was presented by [30], who collected sentences in Italian from the web through a semi-supervised method.

Simpitiki_W is the portion of Simpitiki composed of Italian Wikipedia articles selected from the edits labeled as simplifications. The simplified sentences are obtained by applying one simplification operation at a time [31].

Admin-It_{OP} and **Admin-It_{RD}** are subsections of Admin-It. The former corresponds to another section of the Simpitiki corpus and is composed of sentences from administrative texts manually simplified by applying a single operation, whereas in the latter, the simplification is applied at a document level, then the sentences were manually aligned [3].

As for the automatic translations, firstly we used Google Translate only on the simple sentences of English corpora. The resulting corpora are then referred to as *NewsEn-Ita*, *OSE-EnIta*, and *SimPA-EnIta*. Then, we automatically translated datasets including both simple and complex sentences. In this case, the resulting corpora are referred to as *NewsIta*, *OSE-Ita*, and *SimPA-Ita*.⁷

We extracted two random samples from NewsEn,⁸ instead of fine-tuning models on the whole dataset, since [18] achieved SOTA performance for SARI on Spanish on a relatively small amount of data.⁹

⁵<https://translate.google.com/>

⁶From these datasets, pairs containing the same sentence in their original and simplified versions were filtered out. Additional processing was applied to Terence and Teacher, which were provided in XML format. Datasets statistics are reported in Appendix B.

⁷The simplified sentences in these corpora show variations in the two translation versions, as the statistics in Appendix B show.

⁸One for each kind of translation.

⁹They fine-tuned their model on a sample of Spanish NewsEla that counted about 7k sentences.

5.2. Evaluation metrics

Three metrics were used to evaluate the models: BLEU [32], SARI [33] and BERTScore_p [34].

The first, BLEU, measures the n-gram overlap between gold and generated sentences. It is inherited from Machine Translation, and [35] observed that BLEU is less reliable when sentence splitting is applied in the simplified sentence. SARI is specific to evaluating simplification models and measures the effectiveness of copy (KEEP), insertion (ADD), and deletion (DEL) operations applied [36]. This metric, therefore, is widely used to evaluate simplification at the lexical level. By design, SARI takes as input several gold references. Since Admin-It-L2 contains only one reference per sentence, this metric might not be fully reliable. The third evaluates texts generally created by generative models. BERTScore sums up the cosine similarity of token pairs with the highest similarity. Specifically, BERTScore_p measures the similarity of tokens in the predicted sentence with respect to the tokens in the gold sentence. [35] observed that BERTScore_p correlates with the simplicity values given by human annotators when such values are low.

We applied these metrics by using the implementation of the EASSE evaluation tool [37] provided by [16]. As for BERTScore, since there is no variant of BERT available for the Italian language for this metric, we employed a multilingual model, `xlm-roberta-large`,¹⁰ which performs better than mBERT on the Italian administrative language [38].

5.3. Control tokens

For these experiments, we used the same control tokens as [16]:

- *NbChars*. Synthesis: the ratio of the length in characters between source and simplified sentences;
- *LevSim*. Paraphrasing: the normalized *Levenshtein distance* between source and simplified sentences;
- *WordRank*. Lexical complexity: the ratio between word frequencies in the two sentences, original and simplified;
- *DepTreeDept*. Syntactic complexity: the ratio between the maximum depth of the syntactic tree of source and target sentences;
- *NbWords*. Lexical complexity: the number of words in the simplified sentence divided by that of the complex sentence.

Although *WordRank* leverages English language data, we included this control token since [12] drew some advantage from it on Italian text simplification. To extract

¹⁰<https://huggingface.co/xlm-roberta-large>

these linguistic features, we added a specific function to the implementation of [16] to detect the language of the examined sentence.¹¹ Given the results obtained through behavioral analysis [19], we employed all available control tokens for these experiments. A more suitable simplification for Italian L2 speakers can be obtained by controlling word frequency through *WordRank* and *NbWords*. This affects morphological aspects because verbs in the indicative present tense are generally more frequent than those in other verb tenses and modes. *LevSim* and *NbChars* operate more on sentence length whereas the depth of the syntactic tree can be reduced with *DepTreeDept*, also affecting propositional chains average length. Regarding the number of complex entities and terms, this trait was traced back to aspects related to sentence length, as emerged from the behavioral analysis.

Differently from [11] and [16], who conducted their experiments by extracting training, validation, and test sets from the same corpus, the sentences used for our experiments come from different domains and languages. The goal here is to modify the values of the control tokens to optimize the performance of the model on the test set. For this purpose, we extracted the values of the controlled variables from each pair of sentences in Admin-It-L2, and then we computed the average.¹² We fed the models with such values to condition the simplification generated during evaluation.

5.4. Baselines

We selected two baseline models. Admin-It_{OP} is obtained by fine-tuning mT5 using the subsection of Admin-It. This subset corresponds to a subsection of Simpitiiki, the largest Italian dataset of parallel sentences in the administrative language. We also evaluated this model by setting the values of the control tokens with the average values on Admin-It-L2. The second baseline is a hypothetical model that generates a copy of the complex sentences of the test set. This baseline is reported in Table 2 as Admin-It-L2_C. We assume that models obtaining results close to this baseline are very conservative in their simplification process.

6. Results and discussion

The results of our experiments are shown in Table 2. The baseline obtained by fine-tuning mT5 only on Admin-It_{OP} reached results close to zero for both BLEU and BERTScore_p indicating that this dataset might be too small for the model to learn any simplification rule. Also,

¹¹We used the *language identification* module provided by fastText [39, 40]. In this case, as to compute the corpora statistics we used: `it_corenews_sm` and `en_coreweb_sm`.

¹²Table 11 in Appendix C contains the values given to the controllable tokens.

Table 2

Results achieved by each model. The generated simplifications are evaluated with BLEU, SARI, and BERTScore_p. The highest results obtained for each metric are written in bold.

Model	BLEU	SARI	SARI _{ADD}	SARI _{KEEP}	SARI _{DEL}	BERTScore _p
Admin-It_{OP}	0.02	28.89	0.21	3.28	83.19	0.00
Admin-It-L2_C	25.16	14.04	0.00	42.11	0.00	0.50
All-Ita	24.73	21.63	0.21	42.09	22.60	0.50
All-Ita+SimPA-EnIta	25.34	27.51	0.80	41.91	39.83	0.52
All-Ita+All-EnIta	22.39	24.97	0.93	40.57	33.42	0.48
OSE-EnIta+SimPA-EnIta	12.02	39.24	3.19	36.72	77.82	0.51
OSE-Ita+SimPA-Ita	24.86	27.57	1.00	42.74	38.98	0.51
All-TransIta	15.96	35.85	2.98	41.94	62.64	0.44
All-Ita+SimPa-Ita	25.0	20.16	0.36	42.28	17.85	0.51

the low BERTScore_p is indicative of incorrect simplification [35]. Admin-It-L2_C, obtained a lower SARI score and higher BLEU and BERTScore_p than the first baseline.

The first model we tested, All-Ita, was obtained by fine-tuning mT5 on all available Italian corpora. However, this model obtained a lower SARI score than Admin-It_{OP} (-6.66), and even though it equals Admin-It-L2_C for BERTScore_p, it reached a lower BLEU score (-0.43).

We built a second model by adding SimPA-EnIta to All-Ita. This model scores slightly better than Admin-It-L2_C on both BLEU (+0.61) and BERTScore_p (+0.02). We also included OSE-EnIta and NewsEn-Ita in the fine-tuning of a third model, All-Ita+All-EnIta. However, adding all the cross-language data worsened the performance. Therefore, we ran further experiments focusing on the characteristics of each corpus used.

We leverage cross-language corpora employing only OSE-EnIta and SimPA-En-Ita, and although this model achieves the worst results for BLEU (12.02), its BERTScore_p is only slightly lower than All-Ita+SimPA-EnIta (-0.01), and above all, it surpasses this model for 11.73 points for SARI, obtaining the best score for this metric: 39.24%. This model also achieved the highest value for SARI_{ADD} (3.99) - although it remains low - and SARI_{DEL} (77.82), and the lowest for SARI_{KEEP} (36.72). Such scores may indicate how much less conservative this model is than the others in generating simplifications. Our next experiment involved the full translation of these two corpora, OSE-Ita and SimPA-Ita. Although the model achieves the same BERTScore_p (0.50) as OSE-EnIta+SimPA-En-Ita, the performance is significantly worse for SARI (-11.67). Finally, for BLEU, although it outperforms All-Ita (+0.13), the best model for this score remains All-Ita+SimPa-EnIta. Next, we decided to fine-tune mT5 using all the corpora fully translated from English. The resulting model is reported in Table 2 as All-TransIta and improves over the previous one only for SARI, but without reaching the score obtained by OSE-EnIta+SimPA-En-Ita. We conducted a final experi-

ment with All-Ita+SimPA-Ita. This model scores similarly to All-Ita+SimPA-EnIta for BLEU (-0.34) and BERTScore_p (-0.01), but the SARI score is significantly worse (-7.35), generating rather conservative simplifications, with a SARI_{KEEP} even higher than Admin-It-L2_C (+0.17).

Albeit considering the variations in the translations, the contribution of the cross-language data for specific corpora would seem to make a difference compared to the texts fully translated into Italian (e.g., OSE-EnIta+SimPA-En-Ita vs. OSE-Ita and SimPA-Ita). Even though they remain low, the best results were obtained by All-Ita+SimPA-EnIta for BERTScore_p (0.52) and BLEU (25.34), while the highest score for SARI is reached by OSE-EnIta+SimPA-En-Ita (39.24), not too far from the SOTA for English.¹³

6.1. Manual analysis

Then, we manually analyzed a random sample of 50 simplified sentences. We observed that All-Ita+SimPA-EnIta tends to copy the content of the original sentences. The model merely eliminates some portion of text, but in several cases, this happens with relevant information. In OSE-EnIta+SimPA-EnIta simplifications, the deletion operations are even more massive, and much of the source text is usually removed, producing inconsistent sentences. In other cases, OSE-EnIta+SimPA-EnIta produces “artificial hallucinations”. However, OSE-EnIta+SimPA-EnIta succeeds in producing better simplifications when the complex sentences are shorter, even when All-Ita+SimPA-EnIta still reproduces them entirely (see Table 3).¹⁴

The results confirmed what was pointed out by [35]: BERTScore_p is a reliable metric in detecting mostly low-quality simplifications, and metrics like BLEU and SARI cannot be analyzed in isolation. Besides the low quality of the generated simplification, the low values obtained by BERTScore_p may also find an explanation both in the

¹³Martin et al. [14] obtained a SARI score of about 42 on NewsEn.

¹⁴Other examples of models output can be seen in Appendix D.

Table 3

The Table shows an example of the generated simplifications. All-Ita+SimPA-EnIta exactly reproduces the text of the original sentence in Admin-It-L2, while OSE-EnIta+SimPA-EnIta eliminates the more complex portion of text contained within an aside (in italic in the original sentence).

Original Admin-It-L2: È, inoltre, possibile richiedere, per non più di 3 giorni consecutivi, un pasto “in bianco”, *in presenza di disturbi gastroenterici, senza la presentazione di certificazione o prescrizione medica.*

Simplified Admin-It-L2: Se il bambino ha disturbi allo stomaco o all’intestino, il genitore può richiedere un pasto “in bianco” senza presentare una certificazione o prescrizione medica per massimo 3 giorni di fila.

All-Ita+SimPA-EnIta: È possibile richiedere, per non più di 3 giorni consecutivi, un pasto “in bianco”, in presenza di disturbi gastroenterici, senza la presentazione di certificazione o prescrizione medica.

OSE-EnIta+SimPA-EnIta: È possibile richiedere un pasto “in bianco per non più di 3 giorni consecutivi, senza la presentazione di certificato o prescrizione medica.

BERT model employed for the extraction of the vector representations, which is not specific to Italian (Sec. 5.2), and in the composition of the test set, Admin-It-L2. Its sentences are more complex and longer than the other corpora and underwent a deeper simplification process.

7. A quick chat with ChatGPT

We conducted a preliminary and exploratory experiment to get a brief overview of the capabilities of ChatGPT when asked to simplify Italian administrative texts for Italian L2 speakers in a zero-shot setting.

In the conversation shown in Table 4,¹⁵ when asked to simplify a sentence contained in Admin-It-L2, ChatGPT provides a short and syntactically accurate text, even though the long period in the original sentence has not been broken down into shorter sentences. The model shows that it can provide the meaning of “procura alle liti” (power of attorney) and can insert a short definition into the simplified sentence. However, when asked to simplify the sentence further by phrasing it as if it were addressing a person who is not fluent in Italian, the model adopts a very informal register: the sentence sounds almost ungrammatical and less coherent in some parts.¹⁶ When prompted to use a more polished language, ChatGPT raised the register, although apparently, the model

¹⁵We tested ChatGPT between December 2022 and January 2023.

¹⁶The absence of the pronoun “lo” referring to “documento” makes less explicit what the expression “e non doveva essere” is referring to.

loses sight of the identity of the complainants, namely citizens, and instead refers to a generic “person”. This way, ChatGPT shows to be able to simplify administrative text in Italian at both the syntactic and the lexical level and to tune the simplification according to the prospected scenario, i.e., when addressing Italian L2 speakers. Nevertheless, the model is not entirely accurate when inferring that “the uncertainty as to the antecedence of the power of attorney” led to the “confusion about who is representing the citizen in the process”. Indeed, this latter case could juridically give rise to defects of a different nature than the one depicted by the Province of Pistoia.¹⁷ ChatGPT seems not to grasp here the intrinsic and necessary complexity of the content, which is closely linked to the complexity of the subject matter described [41].

It is also important to remark that proprietary LLMs like ChatGPT are not open-source and are available only via web APIs. This leads to a lack of transparency in the data used for training, limiting the reproducibility of experiments and quality control of the results [42].

8. Conclusions and future work

In this paper, we presented the first benchmark for the automatic simplification of Italian administrative texts aimed at Italian L2 speakers. We employed data collected from behavioral analysis to build Admin-It-L2, a parallel corpus of original-simplified sentences. Then, we used this corpus to test controllable models for text simplification based on Transformers. Although we obtained high results for SARI, we observed that SARI and BLEU alone cannot be used to evaluate text simplification and that a fully reliable automatic metric for the Italian language is still missing. Finally, we asked ChatGPT to simplify one of the sentences in Admin-It-2, obtaining a higher quality simplification than mT5, although the model presents some limitations, especially in reproducibility, and in some cases, in answer accuracy.

For future work, we aim at validating Admin-It-L2 involving human annotation, and, more specifically, domain experts and Italian L2 speakers, to properly evaluate the employed models and the quality of their generations. As for the models, we aim to conduct an ablation study to assess the contribution of each control token. We also would like to adopt a control token to encode the language used in the sentences and employ the corpora in their original versions. Moreover, we aim to create a specific automatic metric to evaluate Italian text simplification. Finally, we would like to test open-source LLMs on administrative text simplification focused on specific target readers, with different prompts, in zero and few-shot settings.

¹⁷We consulted a legal expert to verify the correctness of the answers of ChatGPT.

Table 4

The excerpt of a conversation with ChatGPT in Italian. In bold are the instructions given to the chatbot and in italics are the translations in English of both questions and answers.

Semplifica questa frase: La Provincia di Pistoia eccepisce anche l'inammissibilità del ricorso per nullità della procura alle liti rilasciata dai cittadini ricorrenti in quanto apposta, a differenza di quella dell'associazione attrice, su appositi fogli aggiunti e spillati all'atto introduttivo del giudizio, in difformità della previsione contenuta nell'art. 83 c.p.c. e ciò in quanto, in tal caso non vi sarebbe certezza in ordine alla anteriorità del conferimento del mandato rispetto alla notifica dell'atto di resistenza o di impugnazione (cfr. Cass. civ., sez. iii, 14 marzo 2006, n. 5443).

Simplify this sentence: The Province of Pistoia also objects to the inadmissibility of the appeal due to the nullity of the power of attorney issued by the citizen plaintiffs in that it was affixed, unlike that of the plaintiff association, on special sheets added and stapled to the writ of summons, contrary to the provision contained in Art. 83 c.p.c. and this because, in such a case, there would be no certainty as to the anteriority of the conferment of the mandate concerning the notification of the act of resistance or appeal (see Cass. civ, sec. iii, March 14, 2006, no. 5443).

La Provincia di Pistoia contesta l'ammissibilità del ricorso per nullità della procura alle liti in quanto è stata apposta su fogli aggiunti e non nell'atto introduttivo del giudizio, violando l'articolo 83 c.p.c. e causando incertezza sulla anteriorità del mandato rispetto alla notifica dell'atto di resistenza o di impugnazione.

The Province of Pistoia contests the admissibility of the appeal on the grounds that the power of attorney was null and void because it was affixed on supplementary sheets and not in the writ of summons, violating Article 83 of the Code of Civil Procedure and causing uncertainty as to the antecedence of the grant of power of attorney with respect to the service of the notice of resistance or appeal.

cos'è la procura alle liti?

what is the power of attorney?

La procura alle liti è un documento formale con cui una persona (il mandante) conferisce a un'altra (il mandatario) il potere di rappresentarla e agire in suo nome in una causa legale. La procura alle liti deve essere redatta in forma scritta, sottoscritta dal mandante e autenticata da un notaio o da un avvocato. In essa vengono specificati i poteri che il mandatario ha in relazione alla causa, come ad esempio il potere di depositare documenti, di comparire in giudizio e di transigere.

A power of attorney is a formal document by which one person (the principal) grants another (the agent) the power to represent and act on his behalf in a legal case. The power of attorney must be in writing, signed by the principal, and notarized by a notary public or lawyer. It specifies the powers that the agent has in relation to the lawsuit, such as the power to file documents, to appear in court, and to settle.

Puoi inserire questa definizione nella semplificazione che hai fornito poco fa in modo che il testo sia chiaro anche alle persone che non sanno il significato di procura alle liti. La frase non deve risultare quindi troppo lunga

You can include this definition in the simplification you provided earlier so that the text is clear even to people who do not know the meaning of power of attorney in litigation. The sentence should therefore not be too long

La Provincia di Pistoia contesta l'ammissibilità del ricorso per nullità della procura alle liti, un documento in cui una persona conferisce a un'altra il potere di rappresentarla in una causa legale, in quanto è stata apposta su fogli aggiunti e non nell'atto introduttivo del giudizio, violando l'articolo 83 c.p.c. e causando incertezza sulla anteriorità del mandato rispetto alla notifica dell'atto di resistenza o di impugnazione.

The Province of Pistoia challenges the admissibility of the appeal on the grounds that the power of attorney, a document in which one person confers on another the power to represent him or her in a legal case, is null and void because it was affixed on additional sheets of paper and not in the writ of summons, violating Article 83 of the Code of Civil Procedure and causing uncertainty as to the antecedence of the power of attorney from the service of the notice of resistance or appeal.

Adesso semplifica ancora di più la frase come se stessi parlando a qualcuno che conosce poco l'italiano

Now simplify the sentence even more as if you were speaking to someone who is not fluent in Italian

La Provincia di Pistoia dice che il documento che permette a qualcuno di rappresentare un'altra persona in un processo legale è sbagliato, perché è stato scritto su fogli aggiunti e non doveva essere e questo causa confusione su chi rappresenta la persona nel processo.

The Province of Pistoia says that the document that allows someone to represent another person in a legal process is wrong because it was written on added sheets and should not have been, and this causes confusion about who is representing the person in the process.

Potresti migliorare un po' lo stile? il tuo interlocutore conosce poco l'italiano ma il testo deve essere un po' più curato

Could you improve the style a little? your interlocutor is not fluent in Italian, but the text needs to be a little more polished

La Provincia di Pistoia sostiene che il documento chiamato "procura alle liti" che consente a una persona di rappresentare un'altra persona in un processo legale è stato scritto in modo errato. Infatti, è stato scritto su fogli aggiunti anziché nell'atto ufficiale del processo, e questo può causare confusione su chi rappresenta la persona nel processo.

The Province of Pistoia claims that the document called "power of attorney" that allows a person to represent another person in a legal process was written incorrectly. In fact, it was written on added sheets instead of in the official document of the trial, and this can cause confusion about who represents the person in the trial.

Acknowledgments

Research partially supported by PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGeneration EU programme.

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. URL: http://www.francoangeli.it/Riviste/Scheda_rivista.aspx?idArticolo=25382.
- [2] M. A. Cortelazzo, Il linguaggio amministrativo: principi e pratiche di modernizzazione, *Studi superiori*, Carocci, 2021.
- [3] M. Miliani, S. Aurienemma, F. Alva-Manchego, A. Lenci, Neural readability pairwise ranking for sentences in italian administrative language, in: *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, 2022*, pp. 849–866.
- [4] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, et al., Summary of chatgpt-related research and perspective towards the future of large language models, *Meta-Radiology (2023)* 100017.
- [5] F. Alva-Manchego, C. Scarton, L. Specia, Data-driven sentence simplification: Survey and benchmark, *Computational Linguistics* 46 (2020) 135–187.
- [6] K. Cho, B. van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using rnn encoder-decoder for statistical machine translation, in: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1724–1734.
- [7] S. Nisioi, S. Štajner, S. P. Ponzetto, L. P. Dinu, Exploring neural text simplification models, in: *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers)*, 2017, pp. 85–91.
- [8] G. Klein, Y. Kim, Y. Deng, J. Senellart, A. M. Rush, Opennmt: Open-source toolkit for neural machine translation, in: *Proceedings of ACL 2017, System Demonstrations*, 2017, pp. 67–72.
- [9] A. Palmero Aprosio, S. Tonelli, M. Turchi, M. Negri, A. Di Gangi Mattia, Neural text simplification in low-resource conditions using weak supervision, in: *Workshop on Methods for Optimizing and Evaluating Neural Language Generation (NeuralGen)*, Association for Computational Linguistics (ACL), 2019, pp. 37–44.
- [10] J. Mallinson, M. Lapata, Controllable sentence simplification: Employing syntactic and lexical constraints, *ArXiv abs/1910.04387* (2019). URL: <https://api.semanticscholar.org/CorpusID:204008990>.
- [11] L. Martin, È. V. De La Clergerie, B. Sagot, A. Bordes, Controllable sentence simplification, in: *Proceedings of the 12th Language Resources and Evaluation Conference*, 2020, pp. 4689–4698.
- [12] A. L. Megna, D. Schicchi, G. Lo Bosco, G. Pilato, A controllable text simplification system for the italian language, in: *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, 2021, pp. 191–194. doi:10.1109/ICSC50631.2021.00040.
- [13] W. Xu, C. Callison-Burch, C. Napoles, Problems in current text simplification research: New data can help, *Transactions of the Association for Computational Linguistics* 3 (2015) 283–297.
- [14] L. Martin, A. Fan, È. de la Clergerie, A. Bordes, B. Sagot, MUSS: Multilingual unsupervised sentence simplification by mining paraphrases, in: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, 2022, pp. 1651–1664. URL: <https://aclanthology.org/2022.lrec-1.176>.
- [15] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7871–7880.
- [16] K. C. Sheang, H. Saggion, Controllable sentence simplification with a unified text-to-text transfer transformer, in: *Proceedings of the 14th International Conference on Natural Language Generation*, Association for Computational Linguistics, Aberdeen, Scotland, UK, 2021, pp. 341–352. URL: <https://aclanthology.org/2021.inlg-1.38>.
- [17] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, et al., Exploring the limits of transfer learning with a unified text-to-text transformer., *J. Mach. Learn. Res.* 21 (2020) 1–67.
- [18] S. Štajner, K. C. Sheang, H. Saggion, Sentence simplification capabilities of transfer-based models, *Proceedings of the AAAI Conference on Artificial Intelligence* 36 (2022) 12172–12180. URL: <https://ojs.aaai.org/index.php/AAAI/article/view/21477>. doi:10.1609/aaai.v36i11.21477.
- [19] M. Miliani, M. S. Senaldi, G. E. Leboni, A. Lenci, Un-

- derstanding italian administrative texts: A reader-oriented study for readability assessment and text simplification, in: Proceedings of the 1st Workshop on AI for Public Administration (AIXPA), 2022, pp. 71–87.
- [20] 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.
- [21] L. C. Passaro, A. Lenci, Extracting terms with extra, Computerised and corpus-based approaches to phraseology: Monolingual and multilingual perspectives (2016) 188–196.
- [22] L. Xue, N. Constant, A. Roberts, M. Kale, R. AlRfou, A. Siddhant, A. Barua, C. Raffel, mT5: A massively multilingual pre-trained text-to-text transformer, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, Online, 2021, pp. 483–498. URL: <https://aclanthology.org/2021.naacl-main.41>. doi:10.18653/v1/2021.naacl-main.41.
- [23] G. Sarti, M. Nissim, It5: Large-scale text-to-text pretraining for italian language understanding and generation, arXiv preprint arXiv:2203.03759 (2022).
- [24] S. Štajner, M. Popović, Can text simplification help machine translation?, in: Proceedings of the 19th Annual Conference of the European Association for Machine Translation, 2016, pp. 230–242.
- [25] K. Mishra, A. Soni, R. Sharma, D. M. Sharma, Exploring the effects of sentence simplification on hindi to english machine translation system, in: proceedings of the workshop on automatic text simplification-methods and applications in the multilingual society (ATS-MA 2014), 2014, pp. 21–29.
- [26] L. Perez-Beltrachini, M. Lapata, Models and datasets for cross-lingual summarisation, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, pp. 9408–9423.
- [27] C. Jiang, M. Maddela, W. Lan, Y. Zhong, W. Xu, Neural CRF model for sentence alignment in text simplification, CoRR abs/2005.02324 (2020). URL: <https://arxiv.org/abs/2005.02324>. arXiv:2005.02324.
- [28] S. Vajjala, I. Lučić, Onestopenglish corpus: A new corpus for automatic readability assessment and text simplification, in: Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications, 2018, pp. 297–304.
- [29] C. Scarton, G. Paetzold, L. Specia, SimPA: A sentence-level simplification corpus for the public administration domain, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, 2018, pp. 4333–4338. URL: <https://aclanthology.org/L18-1685>.
- [30] D. Brunato, A. Cimino, F. Dell’Orletta, G. Venturi, Paccss-it: A parallel corpus of complex-simple sentences for automatic text simplification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 351–361.
- [31] S. Tonelli, A. P. Aprosio, F. Saltori, Simpitiiki: a simplification corpus for italian, in: Proceedings of CLiC-it, 2016.
- [32] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, Bleu: a method for automatic evaluation of machine translation, in: Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.
- [33] W. Xu, C. Napoles, E. Pavlick, Q. Chen, C. Callison-Burch, Optimizing statistical machine translation for text simplification, Transactions of the Association for Computational Linguistics 4 (2016) 401–415.
- [34] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, Y. Artzi, Bertscore: Evaluating text generation with bert, in: International Conference on Learning Representations, 2019.
- [35] F. Alva-Manchego, C. Scarton, L. Specia, The (un)suitability of automatic evaluation metrics for text simplification, Computational Linguistics 47 (2021) 861–889.
- [36] E. Sulem, O. Abend, A. Rappoport, Bleu is not suitable for the evaluation of text simplification, arXiv preprint arXiv:1810.05995 (2018).
- [37] F. Alva-Manchego, L. Martin, C. Scarton, L. Specia, Easse: Easier automatic sentence simplification evaluation (2019) 49–54.
- [38] S. Auriemma, M. Miliani, A. Bondielli, L. C. Passaro, A. Lenci, Evaluating pre-trained transformers on italian administrative texts, in: Proceedings of 1st Workshop on AI for Public Administration, 2022, pp. 54–70.
- [39] A. Joulin, É. Grave, P. Bojanowski, T. Mikolov, Bag of tricks for efficient text classification, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, 2017, pp. 427–431.
- [40] A. Joulin, E. Grave, P. Bojanowski, M. Douze, H. Jégou, T. Mikolov, Fasttext.zip: Compressing text classification models, ArXiv abs/1612.03651 (2016). URL: <https://api.semanticscholar.org/CorpusID:16196524>.
- [41] A. Fioritto, Manuale di stile. strumenti per semplificare il linguaggio delle amministrazioni pubbliche (1997).
- [42] A. Liesenfeld, A. Lopez, M. Dingemans, Opening

up chatgpt: Tracking openness, transparency, and accountability in instruction-tuned text generators, in: Proceedings of the 5th International Conference on Conversational User Interfaces, 2023, pp. 1–6.

A. Admin-It-L2 Annotation

In this section, we present further details about the realization of Admin-It-L2. As described in Section 3, the behavioral analysis shows that Italian L2 speakers struggle in reading simplified texts with long sentences, long prepositional chains, a high number of participle verbs, and a lower number of indicative verbs. Their answer error rate also increased with a higher number of multi-words and entities, but also these lexical features were associated in the behavioral analysis with sentence length.

Regarding morphological aspects, we applied transformations from the subjunctive (6) or the infinitive (5) to the indicative, and verbal periphrases were replaced with single verb forms of similar meaning, as in the example 8. Concerning sentence length reduction, references to laws whose citation in the text was not crucial to sentence comprehension were removed (6, 9). The idea is that the automatic simplification operated by the neural models is always flanked by the original text, which complements the absence of some information that might lower text readability [2]. The depth of the syntactic tree was also reduced (-0.80) by limiting redundant expressions, from an average sentence length in tokens of 29.86 to 24.83 (see Table 1), without affecting the cohesion of the text (6). A finding also emerges from Levensthein’s distance from the original sentences to the simplified sentences (170.44) and the simplified sentences in Admin-It-L2 (202.73). We simplified the sentences syntactically by intervening on prepositional chains, which in some cases were eliminated (7) or reduced (9). Finally, 27 sentences were not further simplified.

B. Dataset statistics

Table 10 shows the details of the corpora used in the simplification experiments. The upper part of the table lists the Italian corpora, then the English-Italian corpora, and at the bottom, the English corpora fully translated into Italian. Finally, the test set, Admin-It-L2. Since the complex-simple pairs constitute in several cases portions of text consisting of several sentences (especially in the simplified version), Table 10 shows also the number of the sentences recognized by spaCy within each portion of text.

C. Values of controllable variables

Table 11 shows the values associated with each controllable token in the prediction phase of the fine-tuned models. We computed the controllable variables from each sentence pair in Admin-It-L2 and then we calculated their average.

D. Examples of simplification

In this section, we reported some examples of the sentences generated by two of the created models: All-Ita+SimPA-EnIta and OSE-EnIta+SimPA-EnIta. All-Ita+SimPA-EnIta tends to copy most of the content from the original sentence (Table 13) or delete portions of the texts regardless of their relevance (Table 12). In the other two reported tables, there are interesting examples of the sentences returned by OSE-EnIta+SimPA-EnIta. In Table 13, the model tends to delete most of the content and generates inconsistent text. Such inconsistencies can also lead to artificial hallucinations, as in the example in Table 14.

Table 5

The simplified sentence in Admin-It_{RS} has been further manipulated morphosyntactically by replacing a verb in the infinitive with one in the present indicative, and syntactically by preferring parataxis to hypotaxis and eliminating information that is not strictly necessary.

Original: La gestione della pesa sarà interamente informatizzata, collegando l'apparato di pesatura ad un sistema informativo; tutte le operazioni di pesatura verranno così memorizzate all'interno di un server, *per poter essere storicizzate, consultate ed elaborate* con la massima comodità e sicurezza e la gestione principale dei dati verrà svolta attraverso l'impiego di un software commerciale leader nel settore, *in grado di soprintendere* opportunamente all'intero iter di trattamento del rifiuto, dal momento del conferimento fino allo stoccaggio o smaltimento finale.

Simplified: La gestione della pesatura sarà informatizzata, collegando l'apparato di pesatura ad un sistema informativo. I dati sulle operazioni di pesatura verranno così conservati all'interno di un server, *per poter essere consultati ed elaborati* con la massima comodità e sicurezza. La gestione dei dati verrà svolta attraverso l'impiego di un software *commerciale leader nel settore, in grado di monitorare* l'intero processo di trattamento dei rifiuti: dal conferimento fino allo stoccaggio o smaltimento.

Admin-It-L2: La gestione della pesatura sarà informatizzata, collegando l'apparato di pesatura ad un sistema informatico. I dati sulle operazioni di pesatura verranno così conservati all'interno di un server, *e potranno essere consultati ed elaborati* con la massima comodità e sicurezza. La gestione dei dati verrà svolta attraverso l'impiego di un software *che monitora* l'intero processo di trattamento dei rifiuti.

Table 6

The simplified sentence in Admin-It_{RS} was further manipulated by intervening both morphologically on verbal modes and length-wise, removing references to norms and redundant information.

Original: La domanda di abitabilità o agibilità di quanto edificato, deve essere proposta all'Ufficio Tecnico Comunale, allegando a questa il certificato di collaudo delle opere edili, ove necessario, la dichiarazione presentata per l'iscrizione al catasto dell'immobile, *restituata dagli uffici catastali con l'attestazione dell'avvenuta presentazione e la dichiarazione del direttore lavori che certifichi*, sotto la propria responsabilità, *la conformità* di quanto realizzato rispetto al progetto approvato, *l'avvenuta prosciugatura* dei muri e la salubrità degli ambienti, *ai sensi del d.p.r. 22 aprile 1994, n. 425*.

Simplified: La domanda di abitabilità o agibilità dell'immobile deve essere presentata all'Ufficio Tecnico Comunale. Alla domanda va allegato il certificato di collaudo delle opere edili. Se necessario, va allegata anche la *dichiarazione dell'iscrizione dell'immobile al catasto (certificata dall'Ufficio del Catasto)* e la dichiarazione del direttore dei lavori *che attesti* sotto la propria responsabilità: che quanto realizzato *sia conforme* al progetto approvato; che i muri *siano asciugati*; che l'ambiente *sia salutare* (d.p.r. 22 aprile 1994, n. 425).

Admin-It-L2: La domanda di abitabilità o agibilità dell'immobile va presentata all'Ufficio Tecnico Comunale. Alla domanda va allegato il certificato di collaudo della costruzione. Se necessario, va allegata anche *la dichiarazione dell'iscrizione dell'immobile al catasto* e la dichiarazione del direttore dei lavori. Il direttore *dichiara* sotto la propria responsabilità che l'immobile *è conforme* al progetto approvato, che i muri *sono asciugati* e che l'ambiente *è sano*.

Table 7

The simplified sentence in Admin-It_{RS} was further manipulated to reduce prepositional chains and superfluous expressions.

Original: *Per l'espletamento dei Servizi oggetto del presente contratto*, nient'altro è dovuto dalla Provincia alla Società oltre a quanto previsto nel presente Contratto, salvo il reintegro delle somme relative ad agevolazioni tariffarie e a riduzioni imposte unilateralmente rispetto a quanto previsto nell'allegato 6 "sistema tariffario", salvo eventuali richieste giudicate ammissibili *da parte del Comitato Tecnico* di cui al successivo art. 49 e salvo provvedimenti di Autorità comunitarie, nazionali, regionali e locali destinati direttamente o indirettamente ai servizi oggetto del presente contratto, senza alcuna decurtazione.

Simplified: *Per l'esecuzione dei Servizi oggetto di questo contratto*, la Provincia deve alla Società solo quanto previsto dal contratto stesso, e l'eventuale rimborso di: somme relative a riduzioni di tariffa e a sconti non concordati nell'allegato 6 "sistema tariffario"; *eventuali* richieste giudicate ammissibili *da parte del Comitato Tecnico* (elencate al successivo art. 49); provvedimenti di Autorità comunitarie, nazionali, regionali e locali destinati direttamente o indirettamente ai servizi oggetto di questo contratto.

Admin-It-L2: La Provincia deve alla Società che esegue i servizi solo quanto previsto dal contratto, e l'eventuale rimborso di somme legate a: riduzioni non presenti nel "sistema tariffario" (allegato 6); richieste approvate *dal Comitato Tecnico* (elencate all'art. 49); provvedimenti di enti comunitari, nazionali, regionali e locali destinati ai servizi oggetto del contratto.

Table 8

The simplified sentence in Admin-It_{RS} was further manipulated by operating on morphological aspects and reducing redundant and superfluous information.

Original: I coniugi che *intendono procedere* alla separazione personale consensuale, allo scioglimento o cessazione degli effetti civili del matrimonio e modifica delle condizioni di separazione o di divorzio *dinanzi all'Ufficiale di Stato Civile* devono compilare una richiesta (ENTRAMBI I CONIUGI, vedi modello allegato) ed inviarla all'Ufficio di Stato Civile via mail all'indirizzo ufficio.anagrafe@comune.it, alla PEC comune@postacert.it oppure via fax al numero 000/0000000 insieme alle copie dei documenti di identità.

Simplified: I coniugi che *intendono procedere presso l'Ufficiale di Stato Civile a:* separazione personale consensuale, scioglimento o cessazione degli effetti civili del matrimonio, modifica delle condizioni di separazione o di divorzio, devono compilare il modulo allegato e inviarlo all'Ufficio di Stato Civile. Il modulo e la copia dei documenti di identità possono essere inviati per mail all'indirizzo ufficio.anagrafe@comune.it, alla PEC comune@postacert.it oppure via fax al numero 000/0000000.

Admin-It-L2: I coniugi che *chiedono* la separazione consensuale, lo scioglimento o la cessazione degli effetti civili del matrimonio, o la modifica delle condizioni di separazione o di divorzio, devono compilare il modulo allegato e inviarlo all'Ufficio di Stato Civile. Il modulo e la copia dei documenti di identità possono essere inviati per mail all'indirizzo ufficio.anagrafe@comune.it, alla PEC comune@postacert.it oppure via fax al numero 000/0000000.

Table 9

The simplified sentence in Admin-It_{RS} has been further manipulated to reduce prepositional chains and references to acts.

Original: In attuazione dell'articolo 6 delle norme tecniche di attuazione del piano del parco di cui alla deliberazione del Consiglio Regionale 12.12.1989 n. 515, per le zone costituenti riferimento ambientale dell'assetto del parco, il piano di gestione è finalizzato al mantenimento dell'esistente, al ripristino di morfologie preesistenti dei luoghi ed alla costruzione di elementi di connessione tra le varie connotazioni naturalistiche del territorio; il piano di gestione definisce anche le funzioni ricreative e produttive compatibili con il mantenimento, il restauro ed il ripristino dell'assetto paesaggistico e lo disciplinano nel tempo.

Simplified: Secondo l'articolo 6 della delibera del Consiglio Regionale 12.12.1989 n. 515 relativo alle norme tecniche di attuazione del piano del parco, il piano di gestione delle zone ambientali ha l'obiettivo di: tutelare il paesaggio esistente, di ristabilire l'aspetto originario di questi luoghi e di costruire connessioni tra gli elementi naturalistici del territorio. Il piano di gestione definisce e disciplina nel tempo anche le funzioni ricreative e produttive compatibili con gli obiettivi del piano stesso.

Admin-It-L2: Il piano del parco tutela il paesaggio esistente, ristabilisce l'aspetto originario di questi luoghi e costruisce connessioni tra gli elementi naturalistici del territorio. Il piano definisce e disciplina nel tempo anche le funzioni ricreative e produttive compatibili con gli obiettivi del piano stesso.

Table 10

The Table shows statistics computed over the datasets used for fine-tuning and testing the controllable models.

Dataset	Genre	Pairs	Sentences		Avg tok. per sent.		Avg char. per tok.		Avg Lev.
			C	S	C	S	C	S	
Admin-It _{OP}	PA	588	597	612	32.86	32.05	6.12	6.06	13.64
Admin-It _{RD}	PA	48	104	107	24.79	18.65	5.67	5.68	171.92
Simpitiki _W	Edu	568	1334	1330	30.27	29.98	5.60	5.59	14.01
Terence	Narr	1012	1125	1146	17.07	16.17	5.19	5.10	23.76
Teacher	Edu	171	271	242	15.40	13.67	4.96	4.79	60.76
PaCCSS-IT	Misc	21958	26217	24422	8.45	8.14	4.79	4.80	17.56
OSE-EnIta	News	5994	6084	6665	26.35	22.89	5.38	5.03	-
NewsEn-Ita	News	192001	201056	198490	24.27	13.99	5.27	5.34	-
SimPA-EnIta	PA	4637	4654	5930	30.66	24.32	5.42	5.80	-
OSE-Ita	News	5994	6682	6664	24.93	22.89	5.45	5.39	32.93
NewsIta	News	250000	274222	260245	26.22	15.99	5.56	5.39	96.46
SimPA-Ita	PA	4637	4924	5921	29.98	24.36	5.86	5.80	44.35
Admin-It-L2	PA	134	197	285	43.25	21.92	5.70	5.96	202.94

Table 11

The table shows the values associated to the control tokens to constrain the simplification of models on Admin-It-L2.

<i>Control Token</i>	<i>Value</i>
NbChars	0.76
LevSim	0.63
WordRank	1.02
DepTreeDept	0.82
NbWords	0.78

Table 12

The Table shows an example of simplification given by the top-scoring models in the presented experiments. In this example, the deletion of a portion of the text by All-Ita+SimPA-Enlta generates a sentence in which relevant information is missing.

Original Admin-It-L2: Per quanto riguarda i permessi temporanei è possibile richiederli attraverso il servizio online per il rilascio dei permessi oppure al comando di Polizia Municipale - solo permessi temporanei giornalieri con denominazione "T".

Simplified Admin-It-L2: L'interessato può richiedere i permessi temporanei giornalieri di tipo "T" attraverso il servizio online apposito oppure alla sede della Polizia Municipale.

All-Ita+SimPA-Enlta: Per quanto riguarda i permessi temporanei è possibile richiederli attraverso il servizio online per il rilascio dei permessi oppure al comando di Polizia Municipale.

OSE-Enlta+SimPA-Enlta: Per quanto riguarda i permessi temporanei, è possibile richiedere i permessi temporanei tramite il servizio online per il rilascio dei permessi temporanei giornalieri con denominazione "****".

Table 13

The Table shows an example of simplification given by the top-scoring models in the presented experiments. The example reports the most usual behavior observed for the two models: All-Ita+SimPA-Enlta tends to reproduce the original sentence contained in Admin-It-L2, while OSE-Enlta+SimPA-Enlta formulates inconsistent text.

Original Admin-It-L2: Il Comune può imporre limitazioni all'apertura di nuovi esercizi limitatamente ai casi in cui ragioni, non altrimenti risolvibili, di sostenibilità ambientale, sociale e di viabilità, rendano impossibile consentire ulteriori flussi di pubblico nella zona senza incidere in modo gravemente negativo sui meccanismi di controllo, in particolare per il consumo di alcolici, e senza ledere il diritto dei residenti alla vivibilità del territorio e alla normale mobilità.

Simplified Admin-It-L2: Il Comune può ostacolare l'apertura di nuovi negozi solo se ci sono problemi di sostenibilità ambientale, sociale e di viabilità, che non consentono ulteriori flussi di pubblico nella zona. Questi limiti servono a non ostacolare i controlli, in particolare sul consumo di alcolici, e a rispettare il diritto dei residenti a vivere tranquillamente e a muoversi normalmente.

All-Ita+SimPA-Enlta: Il Comune può imporre limitazioni all'apertura di nuovi esercizi limitatamente ai casi in cui ragioni, non altrimenti risolvibili, di sostenibilità ambientale, sociale e di viabilità, rendano impossibile consentire ulteriori flussi di pubblico nella zona senza incidere in modo gravemente negativo sui meccanismi di controllo, in particolare per il consumo di alcolici e senza ledere il diritto dei residenti alla vivibilità del territorio e alla normale mobilità.

OSE-Enlta+SimPA-Enlta: Il Comune può imporre limiti di apertura di nuovi esercizi entro i casi in cui le ragioni della sostenibilità ambientale e della viabilità non siano risolvibili, in particolare per il consumo di alcolici e alla mobilità normale.

Table 14

The Table shows an example of simplification given by the top-scoring models in the presented experiments. Here the simplification of the title of a document used for behavioral analysis is examined. On the one hand, All-Ita+SimPA-Enlta tends to complete the sentence, albeit with a noun-article agreement error, by inserting a predicate. On the other hand, OSE-Enlta+SimPA-Enlta generates text that is almost completely unrelated to that contained in Admin-It-L2.

Original Admin-It-L2: Requisiti di residenza e soggiorno

Simplified Admin-It-L2: Caratteristiche di residenza e soggiorno

All-Ita+SimPA-Enlta: Il Requisiti di residenza e soggiorno sono questi.

OSE-Enlta+SimPA-Enlta: Il requisito di locazione è l'affitto della città.
