CREATING A SILVER STANDARD FOR PATENT SIMPLIFICATION *

Silvia Casola
University of Padua
Fondazione Bruno Kessler
Italy
scasola@fbk.eu

Alberto Lavelli
Fondazione Bruno Kessler
Italy
lavelli@fbk.eu

Horacio Saggion
Universitat Pompeu Fabra
Spain
horacio.saggion@upf.edu

ABSTRACT

Patents are legal documents that aim at protecting inventions on the one hand and at making technical knowledge circulate on the other. Their complex style – a mix of legal, technical, and extremely vague language – makes their content hard to access for humans and machines and poses substantial challenges to the information retrieval community. This paper proposes an approach to automatically simplify patent text through rephrasing. Since no in-domain parallel simplification data exist, we propose a method to automatically generate a large-scale silver standard for patent sentences. To obtain candidates, we use a general-domain paraphrasing system; however, the process is error-prone and difficult to control. Thus, we pair it with proper filters and construct a cleaner corpus that can successfully be used to train a simplification system. Human evaluation of the synthetic silver corpus shows that it is considered grammatical, adequate, and contains simple sentences.

Keywords text simplification · natural language processing · patents

1 Introduction

To file a patent, an inventor must describe their invention in detail so that a person skilled in the art[2] can understand and reproduce it. The invention disclosure should benefit society by making innovation circulate. However, patent documents are extremely complex: they contain long sentences, novel multiterm entities, and complex syntax built out of noun phrases instead of clauses with recurring entities [1]. [2] performed a user study on patent text readability: Most participants strongly (35%) or somewhat (23%) agreed improving the readability of patents was important and considered documents extremely (29%) or somewhat (29%) difficult to read. 39% of the participants reported difficulties in finding information they were looking for. Even law scholars have advocated for simpler language in patent documents. [3] argues that "When the subject of the case is wrapped in complex and unfamiliar terms, it is tremendously difficult for legal actors to grapple with the theoretical content of the dispute. [...] Communication at the intersection of law and science will always be tremendously challenging. Nevertheless, there are elements of the current patent system that substantially exacerbate the problem". Moreover, patent text is difficult to process automatically and is significantly challenging for the information retrieval community.

Previous work on patent simplification has mainly targeted experts: systems search ways to help patent attorneys better understand the claims (which defines the scope of the legal protection) through ad hoc visualizations or compare documents in the same patent family[3]. Since the legal scope needs to remain unchanged, modifying the text presentation is preferred to rephrasing. We take a different stance and aim at creating a silver standard for simplifying the Description of the invention – the section that describes the invention embodiments in detail – through rephrasing. Simplifying the Description makes the technical knowledge more accessible to society, practitioners in the field, engineers, academics,

---

[2]A person skilled in the art has ordinary skills in a particular technical field.

[3]A patent family is a collection of patent applications covering the same or similar technical content.

and other laypeople; moreover, it can improve performance on further automatic downstream tasks, as shown in other domains [4].

The state of the art in sentence simplification through rephrasing relies on parallel corpora of complex and simple sentences. The problem is framed as a monolingual translation: using a sequence-to-sequence network, complex sentences get translated into their corresponding simple versions. However, parallel data for simplification are difficult to obtain, and many of the existing large-scale datasets derive from the automatic alignment of sentences in the English and Simple English Wikipedia [5, 6]. Creating parallel datasets in other domains requires substantial human effort and is slow and expensive. To the best of our knowledge, no simplified corpora exist for the patent domain.

This paper proposes a method to automatically create a parallel simplification corpus of patent sentences. In particular, we generate a huge noisy corpus of simplification candidate pairs (that we call "bronze") and clean it to obtain a higher-quality silver corpus. To generate the candidates, we adopt a paraphrasing system trained on general-domain text and show that using a zero-shot approach on patents results in simpler, shorter, and easier-to-read sentences. The process is, however, hard to control and might be error-prone. Thus, we discuss filters to select reliable candidates only and show how the silver corpus can be used for training controllable simplification systems for patent sentences. Our contributions are the following:

1. We propose the use of a paraphrasing system to obtain simplification candidates from complex patent sentences. The paraphraser is used in a zero-shot fashion and is only trained on general-domain sentences. We show that, while far from perfect, this zero-shot approach produces text which is more compressed and thus simpler to read. We call the simplifications obtained this way a "bronze corpus".

2. We discuss filtering to select candidates that are appropriate for simplification only. By doing so, we generate the first large-scale parallel silver standard for patent sentence simplification. We make the bronze and silver parallel datasets public for future research[4].

3. We show that the silver standard can be used to train a sequence-to-sequence state-of-the-art system for controllable patent simplification that we release.

4. Finally, we perform a human evaluation of the results and make the unaggregated data public. During the evaluation, we also collect non-expert-generated simplification, which can be used in future research.

2 Previous work

2.1 Sentence simplification

Automatic text simplification aims at making text easier to read and understand for a target group (e.g., people who are neurodivergent [7] or have a disability [8] or people with a low literacy level [9]). Most of the current research performs simplification at the sentence level; refer to [10] for a survey. To simplify complex sentences, the applied transformations can be lexical [11] (substituting unusual words or explaining hard concepts) or syntactic (with sentence splitting [12] and other syntactic structure simplification techniques). They can involve removing unimportant information and compressing the text to enhance clarity. The most popular paradigm for sentence simplification through rephrasing employs sequence-to-sequence architectures. The task is framed as a translation between complex and simple sentences [13]. Controlling the level of simplification (e.g., in terms of compression or extent of the rephrasing) is now deemed essential, and ad hoc control tokens are often used for this goal. [14], for example, proposed to enhance a transformer [15] with control tokens to guide the level of compression, similarity to the input, and lexical and syntactic complexity. [16] used a T5 model [17] and proposed a new control token to help the model replace long, complex words with shorter ones.

These architectures require a large number of parallel complex and simple sentences to train. A popular dataset for simplification is Wikilarge [6], obtained by automatically aligning sentences from corresponding documents in the English and Simple English Wikipedia. However, obtaining corpora in other domains and languages is more complex, as no natural alignment between documents usually exists. Only a few manually curated datasets have been created (e.g., the Newsela corpus [18]), as they imply a huge human effort, which is expensive and requires considerable time. Researchers have thus proposed ways to automatically generate silver standards for simplification. Starting from a Japanese corpus of mixed complexity, [19] first identified a complex and a simple subcorpus based on readability scores; then, they aligned sentences from the two corpora based on their word embeddings nearest neighbors; they experimented with several similarity metrics. Similarly, [20] used the LASER [21] sentence embedding and retrieved nearest neighbors with some filters to ensure quality. While interesting, these methods are hard to apply to the patent

---

[4]The corpora and models are available at `https://github.com/slvcsl/patentSilverStandard`. Data is shared under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License.

domain, whose textual content has largely the same (high) complexity, as strict rules and common patterns govern the style. In contrast, [22] proposed a method to turn a translation corpus into a simplification corpus. Given a pair of sentences in two languages, one is used as a bridge and translated into the other (target) language. Authors argue that the two sentences will likely have a different complexity level because machine translation models tend to output high-frequency tokens [23], and there is often a difference in complexity between languages in translation corpora [24]. If the translation is satisfactory, and if there is a difference in the complexity level, the sentence pair is added to the simplification silver standard. Authors show that using a large filtered silver corpus obtained this way outperforms a smaller, cleaner corpus. This approach shifts the burden from a parallel simplification corpus to a parallel translation corpus, which is typically easier to obtain. All described methods were proposed for general-domain simplification.

Our method is inspired by previous literature on generating silver data for simplification in that we propose to select relevant pairs from a larger corpus of possible candidates. However, we do not rely on in-domain simple data nor on external parallel data.

2.2 Patent simplification

Patent documents have attracted great interest from the Information Retrieval community, as demonstrated, e.g., by the CLEF-IP labs [25, 26, 27, 28, 29]. Refer to [30] and [31] for recent surveys in patent retrieval and to [1] for a review of previous work on patent simplification. As claims are hard to read even for patent professionals, most previous work has focused on improving the accessibility and readability of the Claim section. [32], for example, aimed at improving each claim presentation by segmenting it into fragments that are then formatted more clearly, e.g., by adding new lines. [33] used an Information Extraction engine that detects entity mentions, their type, and relations through distant supervision. They built an interface that shows the most salient elements in the Claim section to understand the patent structure and compare patents in the same family. These works target patent attorneys and other experts involved in the patent filing process. In contrast, [2] proposed ways to improve patent visualization to lay people, all of which were considered at least as good as the original patent text by users. Some previous research has also tried to improve the understanding of entities in the claims by linking them to the Description, where they are mentioned in the context of actual embodiments [34]. Finally, other previous work has visualized claims in a more structured way, e.g., through graphs [35] or trees [36]. Previous work on simplification through rephrasing is much more limited. A rewriting and rephrasing system was built as part of the PATExpert project [37]. Researchers considered two levels of simplification: one uses surface criteria to segment the input and reconstructs chunks into shorter, easier-to-read sentences [38]. The other [39] represented patents by their Deep Syntactic Structures. This representation is, in turn, used to rewrite a text that is simpler to process for the reader (possibly in another language). Both methods modify the patent text.

3 Method

3.1 Dataset

We use data in the Patent Translation Resource (PatTR) [40, 41] corpus. PatTR is a sentence-level parallel corpus extracted from the MAREC patent collection[5]. The data is available under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. It consists of sentence pairs in different languages (DE-EN, FR-EN, and DE-FR). Sentence pairs are indexed by language and patent sections, i.e., Title, Abstract, Description, and Claims. We chose the PaTR as it contains pre-tokenized sentences, and finding sentence delimiters is an error-prone task for patents. Moreover, the PaTR dataset allows us to compare our method with the general-domain approach by [22], which requires a parallel translation corpus.

In the following experiments, we will focus on the English sentences from the DE-EN pairs and specifically on sentences extracted from the Description only. The PatTR DE-EN Description dataset contains almost 12 million sentence pairs. For computational reasons, we sample 500 thousand sentences.

As a preprocessing step, we removed sentences shorter than 5 tokens or longer than 55 tokens. We also filter sentences where alphabetic characters account for less than 60% of the total. These sentences mainly contain long lists of references or complex chemical formulas. This leaves us with 425,148 sentences. Finally, sentences in the Description often contain references to the figures in the form of numbers in brackets. We use regular expressions to remove these references.

Table 1 contains some statistics on the English subset of the corpus. We use the following metrics as simplicity proxies:

---

[5]See: http://www.ifs.tuwien.ac.at/imp/marec.shtml [Last accessed January 2023]

Table 1: Statistics on the original English patent sentences (after preliminary filtering)

Metric	Mean $\pm$ std
Flesch Reading Ease [42]	32.5 ± 26.3
Flesch–Kincaid Grade Level [43]	61.1 ± 9.5
WordRank [14]	9.8 ± 1.1
Max dependency tree depth	6.7 ± 2.5
Length (chars)	170.3 ± 75.3

- Flesch Reading Ease score [42]: it computes sentence simplicity as a function of the mean word length (in syllables) and the mean sentence length (in words). A higher score indicates a text that is easier to read.

- Flesch Kincaid Grade Level [43]: it is similar in nature to the Flesch Reading Ease score, but it is weighted to correspond to U.S. education grade levels. We report it for completeness. A lower score indicates a text that is easier to read.

- WordRank score: this measure uses word frequency as a proxy of lexical simplicity (as text containing more common words is considered easier to understand). It is computed by taking the third quartile of log ranks (inverse frequency order) of all words in the sentence. We use the implementation proposed by [14].

- Maximum Dependency Tree: this score uses the height of the dependency tree as a proxy for syntactic complexity. We use the implementation proposed by [14].

All metrics are computed per sentence and then averaged.

3.2 Generating simplification candidates using a general-domain paraphrasing system

Previous work has explored using models trained on out-of-domain simple text to simplify complex text. [44], for example, propose an architecture with two decoders (one trained on complex text only and the other trained on simple text only) to control the level of simplification without supervision. Inspired by this line of work, we investigate using a sequence-to-sequence system trained on general-domain text only for simplifying complex patent text. Specifically, we employ a Pegasus-based [45] paraphrasing system for general text. The system was fine-tuned on a custom set of 60k examples from multiple datasets, including PAWS [46]. The trained model is available on Huggingface[6]. We use it as a black box and transform the preprocessed patent sentences.

Table 2 reports some random sentence pairs (no cherry-picking) generated using Pegasus. The simplification is mainly by sentence compression and lexical substitution; some other syntactical changes (e.g., use of the active voice instead of the passive voice) are also common.

While the generated candidates tend to be simpler than the original sentences, using the Pegasus model directly for simplification has two main limitations:

1. Some of the generated sentences contain errors and unknown tokens or are too similar or excessively compressed with respect to the original sentence.

2. The process is hard to control. While state-of-the-art simplification systems allow controlling the level of compression through appropriate tokens, the Pegasus model has a strong tendency to compression. However, this tendency can degenerate into text that does not contain the original core elements, lose too much content, or completely change the original meaning. The level of paraphrasing, similarity, and lexical simplification is also hard to control.

Due to these issues, in the following, we will call the candidates generated by our Pegasus model our "bronze corpus".

For comparison, we also report the same candidates obtained through translation (prior to any filtering) using the model proposed by [22] in Table 3. Note that the sentences are practically paraphrases, and there is no evident difference in complexity. We attribute these characteristics to the lack of complexity differences in sentences in different languages in the patent domain. Our preliminary experiments show that even after filtering these instances for simplification level, e.g., maintaining only pairs with a high relative difference in reading scores, as the original method proposes, the level of simplification is poor.

---

[6]`tuner007/pegasus_paraphrase` [Last accessed: January 2023]

Table 2: Simplification candidate pairs obtained through the Pegasus paraphrasing model (zero-shot) prior to any filtering. Pairs with a crossed-out number are later filtered.

	Original	Candidate
0̸	In interaction with the component secured on the cylinder head, the radial shoulder of the pressure medium distributor now prevents the camshaft from migrating axially further into the cylinder head.	The radial shoulder of the pressure medium distributor prevents the camshaft from moving further into the cylinder head when interacting with the component secured on the cylinder head.
1	Film 12 had holes, and these could be attributed to moisture in the blend material.	The blend material could be to blame for the holes in Film 12.
2̸	According to the invention, the formulations advantageously comprise one or more antioxidants.	The invention states that the formulations comprise one or more anti-oxidants.
3	In the pressure medium flow path between the consumers and the variable displacement pump a respective meter-in orifice is provided.	A meter-in orifice is provided in the pressure medium flow path between the consumers and the pump.
4̸	The yield of TMCH relative to the feed KIP was 27%.	The yield was 27%.
5	A throttle valve angle sensor for internal combustion engines which is described therein has a permanent magnet that is attached to one end of the throttle valve shaft and produces a constant magnetic flux having a direction which is turned as a function of the rotation of the throttle valve shaft.	A permanent magnet that is attached to one end of the throttle valve shaft produces a constant magnetic flux which is turned as a function of the rotation of the throttle valve shaft.
6	They have to be so thin in comparison with the wavelengths produced that they may be caused to vibrate by the ultrasonic waves.	They have to be so thin that they can't be seen by the naked eye
7̸	Mechanisms are provided which allow the guide rollers to yield under spring tension so that the force with which the rollers are pressed on the pane is held within a prescribed range regardless of installation and vehicle body tolerances.	The force with which the rollers are pressed on the pane is held within a prescribed range, regardless of installation and vehicle body tolerances, because of mechanisms provided which allow the guide rollers to yield under spring tension.
8	The line includes an outer conductor tube which is equipped on both ends with connection flanges .	The line has an outer conductor tube on both ends.
9	Thick pressure-sensitive adhesive coats may be present in unfilled, all-acrylate form or resin-blended form or filled with organic or inorganic fillers.	Thick pressure-sensitive adhesive coats can be filled with organic or inorganic fillers.
1̸0̸	The seamless tubular casing preferably has a thickness of from 40 to 200 \u03bcm, when it is made by a blown tube process, and a thickness of from 25 to 75 \u03bcm, when it was obtained by biaxial orientation (double bubble process).	When it is made using a blown tube process and a thickness of 25 to 75?m is obtained, the seamless tubular casing should have a thickness of 40 to 200?m
11	Furthermore, with this type of integration it is mainly the classic MC functionalities, as are relevant in particular for machine tools, that are supported.	The classic MC functionalities are relevant in particular for machine tools that are supported by this type of integration.
1̸2̸	Then, if a predetermined limiting pressure value is reached, the valve body member 19 is moved against the pretension force of the spring member 18, thus clearing a further conduit with respect to the liquid treatment apparatus 3.	If a limiting pressure value is reached, the valve body member 19 is moved against the pretension force of the spring member 18, thus clearing a further conduit with respect to the liquid treatment apparatus 3.
13	These process parameters for intermingling depend on the yarn titer and the lubricant applied to the yarn and must therefore be adapted accordingly.	The process parameters for intermingling are dependent on the yarn titer and the lubricant applied to the yarn.
14	The telephone network is formed as a digital telephone network (e.g. Integrated Services Digital Network, ISDN) or as an analogue telephone network (e.g. Public Switched Telephone Network, PSTN).	The telephone network can be formed as a digital telephone network or an analogue telephone network.
1̸5̸	This may be perceived as disadvantageous by the consumer.	This may be seen as disadvantageous by the consumer.
16	On the other end of the coupling rods 20, 21, these are connected by way of second ball-and-socket joints 22, 23 as the second connection joint of the parallelogram linkage with the second axis 24 which extends in parallel to the first axis 19.	The second connection joint of the parallelogram linkage with the second axis 24 extends in parallel to the first axis 19 on the other end of the coupling rods.

Table 3: Pair candidates obtained through translation, as proposed by [22].

	Original	Candidate
0	In interaction with the component secured on the cylinder head, the radial shoulder of the pressure medium distributor now prevents the camshaft from migrating axially further into the cylinder head.	In this way, the radial collar of the distributor forms the axial bearing of the camshaft in conjunction with the cylinder-head-resistant component.
1	Film 12 had holes, and these could be attributed to moisture in the blend material.	Foil 12 had holes that could be traced back to moisture in the glare material.
2	According to the invention, the formulations advantageously comprise one or more antioxidants.	According to the invention, the preparations contain advantageous one or more antioxidants.
3	In the pressure medium flow path between the consumers and the variable displacement pump a respective meter-in orifice is provided.	A diaphragm is provided in each case in the pressure medium flow path between the consumers and the variable displacement pump.
4	The yield of TMCH relative to the feed KIP was 27%.	The yield of THCM in relation to the KIP used was 27%.
5	A throttle valve angle sensor for internal combustion engines which is described therein has a permanent magnet that is attached to one end of the throttle valve shaft and produces a constant magnetic flux having a direction which is turned as a function of the rotation of the throttle valve shaft.	Here a throttle angle sensor for internal combustion engines is described with a permanent magnet attached to one end of the throttle shaft, which generates a constant magnetic flux, the direction of which is rotated depending on the rotation of the throttle shaft.
6	They have to be so thin in comparison with the wavelengths produced that they may be caused to vibrate by the ultrasonic waves.	These must be so thin in comparison to the wavelengths produced that they can be oscillated by the ultrasonic waves.
7	Mechanisms are provided which allow the guide rollers to yield under spring tension so that the force with which the rollers are pressed on the pane is held within a prescribed range regardless of installation and vehicle body tolerances.	In order to ensure that the force with which rollers 56 are pressed to disc 4 remains within a predetermined range, independent of installation and body tolerances, means 42.70 are provided to enable guide rollers 56 to be cushioned.
8	The line includes an outer conductor tube which is equipped on both ends with connection flanges .	The cable comprises an outer conductor tube 1, which is equipped with connecting flanges 2 at both ends.
9	Thick pressure-sensitive adhesive coats may be present in unfilled, all-acrylate form or resin-blended form or filled with organic or inorganic fillers.	Thick adhesive coatings can be available unfilled as pure acrylate or mixed with resin or filled with organic or inorganic fillers.
10	The seamless tubular casing preferably has a thickness of from 40 to 200 \u03bcm, when it is made by a blown tube process, and a thickness of from 25 to 75 \u03bcm, when it was obtained by biaxial orientation (double bubble process).	The seamless, tubular sheath of the invention preferably has a thickness of 40 to 200 \u00b5m if manufactured by a tube blowing process, and a thickness of 25 to 75 \u00b5m if obtained by biaxial stretch orientation.
11	Furthermore, with this type of integration it is mainly the classic MC functionalities, as are relevant in particular for machine tools, that are supported.	In addition, this type of integration mainly supports the classic MC functionalities, which are particularly relevant for machine tools.
12	Then, if a predetermined limiting pressure value is reached, the valve body member 19 is moved against the pretension force of the spring member 18, thus clearing a further conduit with respect to the liquid treatment apparatus 3.	If a certain limit value is exceeded, valve body 19 is shifted against the prestressing force of spring 18, thus releasing a further line with respect to liquid treatment device 3.
13	These process parameters for intermingling depend on the yarn titer and the lubricant applied to the yarn and must therefore be adapted accordingly.	These process conditions for fishing depend on the yarn titer and the preparation applied to the yarn and must therefore be adapted to these parameters.
14	The telephone network is formed as a digital telephone network (e.g. Integrated Services Digital Network, ISDN) or as an analogue telephone network (e.g. Public Switched Telephone Network, PSTN).	Telephone Network 2 is designed as a digital telephone network (e.g. Integrated Services Digital Network, ISDN) or an analog telephone network (e.g. Public Switched Telephone Network, PSTN).
15	This may be perceived as disadvantageous by the consumer.	This can be detrimental to the consumer.
16	On the other end of the coupling rods 20, 21, these are connected by way of second ball-and-socket joints 22, 23 as the second connection joint of the parallelogram linkage with the second axis 24 which extends in parallel to the first axis 19.	At the other end of the coupling rods 20, 21, these are connected via second ball joints 22, 23, as the second joint of the parallelogram guide, to the second axis 24, which runs parallel to the first axis 19.

Table 4: Number of instances removed by each filter from our original 426,963 bronze standard and examples of such instances. The filters were applied consecutively.

Filter	Removed	Examples
Bad tokens	8,602	**Simple:** *"It is1-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-65561-[...]-65561-65561-65561-6556"*
Non-alphabetical	695	**Simple:** *"R.sup.1, R.sup.2, R.sup.3, Z.sup.1, Z.sup.2 and Z.sup.3 are all related."*
Similarity	38,217	**Complex:** *"Example 23 (E,E)-3,3-[4-(4-Fluorophenyl)-6-isopropyl-1-methyl-pyrid-2one-3,5-diyl]-diprop-2-enal ##STR54## Analogously to Example 10, 2.70 g of crude product are obtained starting from 2.13 g (7.1 mmol) of the compound of Example 22."* **Simple:** *"There are 2.70 g of crude product."* **Complex:** *"Another method of solution attempts to achieve optimum operating cracks by means of temperature-dependent shifting of a conical rotor in the tapered case diameter with the help of an adjustment device."* **Simple:** *"Another method of solution attempts to achieve optimum operating cracks by means of temperature dependent shifting of a conical rotor in the tapered case with the help of an adjustment device."*
Partial similarity	5,937	**Complex:** *"In the treatment of parts of plants, the active compound concentrations in the use forms can be varied within a substantial range."* **Simple:** *"The active compound concentrations in the use forms can be varied within a substantial range."*
Sorted similarity	18,807	**Complex:** *"The solution polymerization without addition of other auxiliaries is the preferred process for the ethylene-vinyl acetate copolymers to be used according to the present invention."* **Simple:** *"According to the present invention, the solution polymerization without addition of other auxiliaries is the preferred method for the production of the ethylene-vinyl acetate copolymers."*
Compression	62,926	**Complex:** *"he bottom end of this rod passes through a hole 46 which leads into the fourth circular chamber 20."* **Simple:** *"The hole 46 leads into the fourth chamber 20."*
Simplicity	1,999	**Complex:** *"The preparation and further processing of the catalyst supports used according to the invention are well known to the person skilled in the art."* **Simple:** *"The person skilled in the art knows about the preparation and further processing of the catalyst supports used in the invention."*

3.3 Filtering bad candidates and generating a silver standard of patent sentences.

As the bronze corpus is noisy, we remove candidates according to the following criteria:

- Bad tokens: we remove any pair that contains an unknown token. We also noticed that the generation can degenerate into a sequence of numerical tokens (e.g., 65561). We remove any pair where the generated sentence contained any of these tokens. For example, this removes pair 10 in Table 2.

- Non-alphabetical characters: we remove pairs where the alphabetical characters in the generated sentence account for less than 60% of the total. These sentences are not necessarily the effect of a bad generation, but the pair is useless as a training example.

- Similarity: we remove sentences whose character-level Levenshteins similarity between the original sentences and its simplification (normalized by the sum of their length) is < 25% or > 90%. The low threshold removes simple sentences that are too different compared to the complex ones (generally very short); the high threshold removes sentences that are equal to the original one or only differ by a small number of characters. This removes, for example, pair 12 and 15. We also remove pairs in which the generated sentence has a similarity > 99% with a continuous substring of the original sentence. Finally, we removed pairs where the strings obtained by sorting the tokens have a similarity > 90%. This removes, for example, pairs 0, 2, and 7.

- Compression: we remove pairs where the ratio between the generated and the original sentence length was > 1.5 or < 0.5. This step avoids excessive compression, which generally corresponds to losing important content or having a modified meaning. This removes, for example, pair 4.

- Simplicity: we considered the Fresh Reading Ease score, the WordRank score, and the sentence dependency tree height as simplification proxies. We removed candidates where the generated sentence is not simpler according to any of these scores.

Table 4 shows the number of pairs removed by each filtering step, with examples. We chose the thresholds heuristically. Future work could investigate learning their most suitable values from a larger corpus of annotated silver sentences; moreover, it could study the best threshold between noise and size of the corpus. The remaining sentences compose our parallel silver standard. After filtering, our corpus consists of 287,965 samples.

Table 5: Results of the model trained using the silver standard

	SARI	BLEU
Validation (silver)	55.09	54.88
Test (silver)	55.22	54.99
Human simplification (with param. search)	39.25	56.86
Human simplification (without param. search)	36.99	57.78

Table 6: Statistics on sentence pairs and simplified sentences from the silver corpus

Metric	Complex	Simplified
Flesch Reading Ease [42]	33.7 ± 24.5	47.6 ± 24.2
Flesch–Kincaid Grade Level [43]	60.8 ± 9.1	56.6 ± 8.5
WordRank [14]	9.7 ± 1.1	9.5 ± 1.2
Max dependency tree depth	6.6 ± 2.5	5.6 ± 2.2
Length (chars)	157.5 ± 67.4	108.4 ± 46.7
Levenshtein similarity	67.7 ± 12.9	
BLEU [49]	38.6 ± 15.5	
BERTScore (avg) [50]	94.45 ± 1.82	

3.4 Using the corpus for training a controllable simplification system

A silver standard for simplification allows for experimentation with models and training processes. To demonstrate this point, we train ACCESS [14], a state-of-the-art system for automatic sentence simplification. We randomly split the silver standard into train (184,297), validation (46,075), and test (57,593) sets. We train the model with early stopping on the validation SARI [47] (patience: 20). We also evaluate the model on the human annotations we collected from the Prolific workers (Section 5.3).

Table 5 shows the results, as computed using the EASSE library [48]. As expected, the metrics on the silver standard are particularly high. This is likely because the corpus is synthetic and presents patterns that are easy to learn for the simplification model. Note, however, that when the model is evaluated on human-written sentences, its performance is close to those obtained on other general-domain datasets. Specifically, we have reported the results on the human evaluation set when using it both as a validation set (allowing the ACCESS model to optimize its preprocessing parameters) and as a test set (only using the optimal parameters computed on the silver standard validation set). In future work, we plan to compare our results with those obtained by training sequence-to-sequence models on different data sets.

Having a silver parallel corpus also allows experimenting with transfer learning. We plan to explore the impact of consecutive fine-tunings (for example, further fine-tuning a model already fine-tuned on Wikipedia-style data). On the other hand, the performance of a model first fine-tuned on a large-scale silver dataset can usually be improved with a small human-curated dataset; this approach would diminish the human effort and, thus, the time and money required for obtaining high-quality simplifications. We plan to investigate these aspects in future work.

4 Corpus quality estimation

4.1 Automatic metrics

To study the dataset characteristics, we compute several automatic metrics. In addition to the simplicity proxies discussed in Section 3.1, we also compute similarity scores among the complex and the simple sentences:

- Normalized character-level Levenshtein similarity: it is the number of insertions, deletions, and substitutions between the complex and the simple sentences normalized by the sum of their length.

- BLEU [49] and BERTScore [50]: these metrics compute the token-based and semantic similarity of each pair.

All metrics are computed per sentence or sentence pairs and then averaged. When required, the original complex sentence is used as the reference. Figure 1 reports the metric distribution; means and standard deviations are summarized in Table 6.
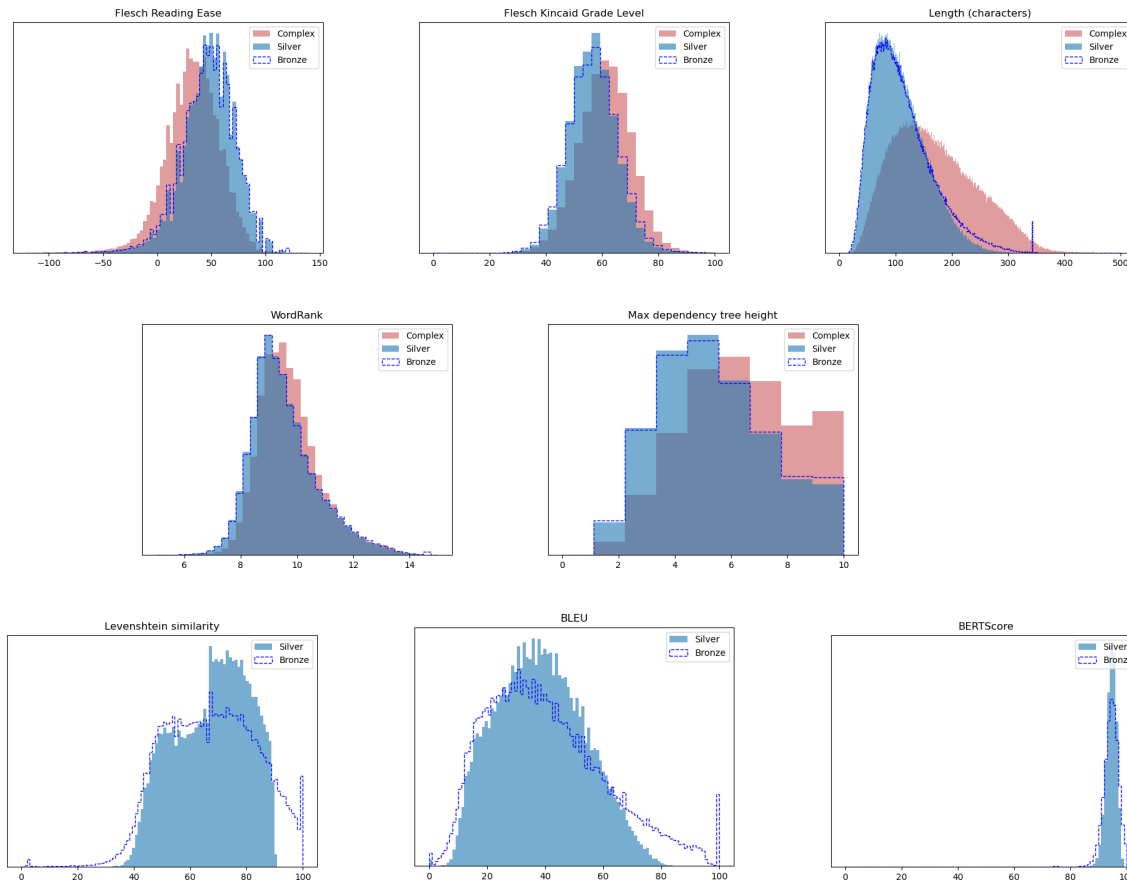
Figure 1: Automatic metrics as computed on the complex and simple sentences of the silver corpus. All histograms are normalized to have unit area.

One can notice that sentences in the silver corpus are simpler than the original ones as measured by the reading scores. The effect on the lexicon is smaller as measured by WordRank. Future work might investigate how to combine our approach with [22], which appears to perform some lexical simplification. The effect of simplification on the syntax complexity is clear as measured by the maximum dependency tree height, and simple sentences are generally significantly shorter than the original. Note how the filtering process removes several sentences practically identical to the original one.

4.2 Qualitative error analysis

We manually analyzed a subset of 100 sentences to identify the remaining errors. Generated sentences are grammatical and fluent. Errors in meaning preservation are mainly of two types: excessive or wrong compression or problems with compositionality. With respect to incorrect sentence compression, coordinate elements in a sentence can be removed. For example, the original sentence *"Both the solid and the corrugated sheets preferably exhibit on one or both outer sides a layer consisting of the compositions according to the invention."* is paired with the sentence *"The corrugated sheets have a layer consisting of compositions on one or both sides."* In this case, the solid sheets, which appear in the original sentence, disappear in the simplified one. Another possible but rare error is the removal of important adjectives. For example, the sentence *"In some cases, it has proved advantageous to use emulsion polymers exhibiting reactive groups at the surface."* has a simplified pair *"In some cases, it has proved to be beneficial to use emulsion polymers with groups at the surface."* Another class of errors derives from the high compositionality of some sentences. In these cases, the relations among elements might not be fully grasped and can be reversed. An example is the following sentence *"The contact lugs, projecting from the bearing plate of the motor, for the electrical feed of the motor are easily accessible there for the connecting cabling in the housing installation orifice."* which is incorrectly paired with

9

"There is an easy way to connect the electrical feed of the motor to the housing installation orifice with the contact lugs projecting from the bearing plate." We also noticed a few hallucinated sentences.

5 Human evaluation

5.1 Evaluation details

Human evaluation is considered the gold standard when evaluating open-ended tasks [51, 52]. We recruited workers for human evaluation through Prolific[7] [53] to quantify the quality of the silver standard as judged by non-experts of the patent domain. Prolific has been used for human evaluation and annotation, and previous work has found results are more reliable in the analyzed cases than those obtained with Amazon Mechanical Turk [54]; it is also often regarded as having higher ethical standards. The survey interface was built using Qualtrics[8].

The survey consisted of a statement describing the goal of the study and the data collection process and asking for consent. Subjects were required to be at least 18 years old and competent to give consent. The workers were informed that the data were collected for research purposes only but that they would be made available to other researchers and could be used for published work in scientific venues. Workers were required to be native English speakers.

After a brief demographic section, workers were introduced to the task. Following previous work, we asked workers to judge the grammaticality and core meaning preservation (adequacy) on a 0-5 Likert scale [55]. Simplicity was measured in a -2 (the Simplified sentence is much more complex than the original one) to +2 (the Simplified sentence is much simpler than the original one) scale. Subjects were also asked to provide an overall 0-5 score and to write a simplification of the original sentence that they considered adequate. Numerical values had to be chosen through sliders. Instructions included a description of the meaning of Grammaticality, Core Meaning Preservation, and Simplicity, together with clarifications (e.g., regarding judging grammaticality by looking at the Simplified sentence only). We excluded subjects who did not conclude the survey or who did not report adequate simplifications. After a small pilot study and after filtering inadequate responses, we collected evaluations for 96 sentence pairs. 78 workers participated in the study, so each sentence pair was evaluated, on average, by 3.93 participants. We chose to evaluate the silver standard with laypeople as they are the target of our simplification. In future work, we plan to further evaluate the dataset with experts of the patent domain and the target technical domain, which might be more reliable, particularly when judging the meaning preservation.

5.2 Numerical scores

Table 7 reports the results of the scores as assigned by the Prolific workers. Figure 3 reports the answer distribution for each dimension. Results show that sentences are considered rather grammatical. Grammatical sentences were described in the instructions as "free of typos, grammatical or syntactical errors" and workers were asked not to consider style when evaluating this dimension. However, by manually validating the answers, we noticed that complex sentences and sentences containing infrequent grammatical constructs (which are, however, frequent in patent documents) might have a low grammatical score despite no obvious errors. Most Simplified sentences were considered to retain the original core meaning. Regarding simplicity, the vast majority of the sentences were considered somewhat or much simpler than the original.

As a first preliminary experiment, we investigated how each dimension correlated with the overall quality score. A simple linear model ($R^2 = 0.66$) scores the dimensions as:

$$\text{overall} = 0.51 + 0.20 \times \text{G} + 0.45 \times \text{M} + 0.40 \times \text{S}$$

where G is the grammaticality score, M is the core meaning score and S is the simplicity score.

5.3 Human-written simplifications

During the survey, we also asked participants to provide a simplification, in their own words, of the original sentence. Table 9 shows some examples generated by the workers. Table 9 reports some metrics computed on the collected data. Unsurprisingly, the human-written sentences are more similar to the simplified sentence; note, however, that their similarity is not as high as to consider them derivative. Moreover, they are shorter than synthetic simplified sentences. Considering the simplification scores, they are in line with that generated by the model. The only exception is that of the dependency tree, which is sensibly less deep in the case of sentences generated by humans. Human-written

---

[7]https://www.prolific.co/
[8]https://www.qualtrics.com/

Figure 2: Human evaluation interface. For each sentence, we also included instructions at the top of the page.

Table 7: Human evaluation results

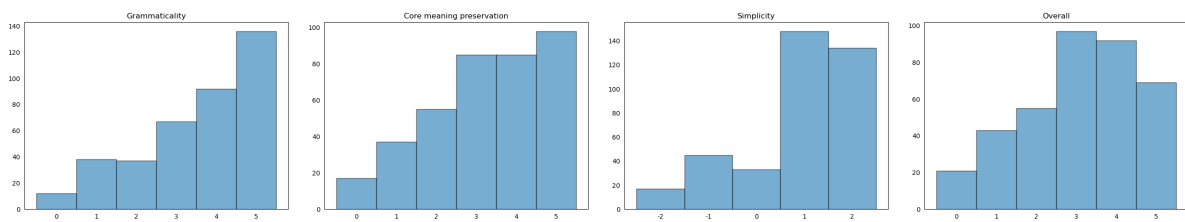| Dimension	Mean $\pm$ std
Grammaticality [0,5]	3.59 ± 1.45
Core meaning preservation [0,5]	3.27 ± 1.46
Simplicity [-2,2]	0.89 ± 1.15
Overall [0,5]	3.07 ± 1.44

Figure 3: Distribution of votes for each dimension. The y-axis shows the absolute number of workers who assigned a given score (x-axis) to sentence pairs in the corpus.

Table 8: Examples of human simplifications

Original	Simplified (automatic)	Simplified (human)
In yet another embodiment, the outer wall can exhibit at least one guide element by which the heating device is guided in an insertion direction during installation or during removal.	The outer wall can have at least one guide element that can be used to guide the heating device during installation or removal.	The outer wall will have at least 1 guide to help guide the heating device during installation or removal. During installation or removal the outer wall has 1 or more guide element.
A lamp holder or fitting carrier for at least one lamp is arranged in a position in which it lies outwardly opposite a recess in the reflector and is releasably connectable with the base frame by second mounting means.	A lamp holder or fitting carrier for at least one lamp is arranged in a way that it is releasably connectable with the base frame by second mounting means.	A lamp holder or fitting carrier for at least one lamp is arranged in a way that it is connectable with the base frame by second mounting means. A lamp holder or fitting carrier for at least one lamp is arranged in a position releasably connectable with the base frame by second mounting means. A lamp holder is arranged in a way that lets the base frame be accessible for a second mount. this product is easily changed from a one lamp to a 2 lamp item via the changeable base
In an emergency, the safety valve can also be opened by hand, when the pressure in the feed line is too high for unknown reasons.	When the feed line is too high for unknown reasons, the safety valve can be opened by hand.	In an emergency, the safety valve can be opened by hand, when the pressure is too high When the feed line pressure is too high, the safety valve can be opened by hand. In an emergency, the feed line is too high for unknown reasons, the safety valve can be opened by hand. when feed line is too high open safety valve if the line feed is too high, hand open the valve

Table 9: Statistics on simplified sentences produced by humans

Metric	Complex	Simplified
Levenshtein similarity	58.87 ± 19.47	65.98 ± 20.13
BLEU [49]	23.48 ± 19.39	35.56 ± 26.47
BERTScore (avg) [50]	92.15 ± 4.03	93.58 ± 4.11
Flesch Reading Ease [42]	45.92 ± 29.91	
Flesch–Kincaid Grade Level [43]	56.40 ± 11.83	
WordRank [14]	9.47 ± 1.46	
Max dependency tree depth	4.85 ± 2.02	
Length (chars)	88.67 ± 41.96	

simplifications allow evaluating models trained on our silver standard with data that are not synthetic. However, users should be aware of their limitations, as they were produced by workers lacking expertise in the patent or technical domain.

6 Conclusions

In this paper, we have discussed a method to generate a parallel silver standard for simplifying patent sentences through rephrasing. To the best of our knowledge, this is the first parallel simplification corpus for patents. We have analyzed the corpus quantitatively and qualitatively, showing that it can be used to train a sequence-to-sequence simplification model. We have also conducted a large-scale human evaluation of the corpus and collected human-written simplifications. While not exempt from the shortcoming of its automatic origin, we have shown that filtering out faulty candidate allows

us to obtain a corpus that has been considered grammatical, adequate, and with a significant simplification. In future work, we plan to explore alternative models to generate the simplification candidates. We will also further explore whether transfer-learning techniques can further improve results. Finally, we plan to investigate whether the method exposed in this paper can be successfully adopted for other types of technical text.

Acknowledgments

References

[1] Silvia Casola and Alberto Lavelli. Summarization, simplification, and generation: The case of patents. *Expert Systems with Applications*, 205:117627, 2022.

[2] Hanna Suominen, Gabriela Ferraro, Jaume Nualart Vilaplana, and Leif Hanlen. User Study for Measuring Linguistic Complexity and Its Reduction by Technology on a Patent Website. In *Conference: 34 International Conference on Machine Learning*, ICML'17, 02 2018.

[3] R. Feldman. Plain Language Patents. volume 17, page 289, 2008.

[4] Richard J. Evans. Comparing methods for the syntactic simplification of sentences in information extraction. *Literary and Linguistic Computing*, 26(4):371–388, 08 2011.

[5] Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. A monolingual tree-based translation model for sentence simplification. volume 2, pages 1353–1361, 08 2010.

[6] Xingxing Zhang and Mirella Lapata. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[7] Eduard Barbu, Maria Martín-Valdivia, Eugenio Martínez-Cámara, and L. López. Language technologies applied to document simplification for helping autistic people. *Expert Systems with Applications*, 42:5076–5086, 07 2015.

[8] Victoria Yaneva, Irina Temnikova, and Ruslan Mitkov. Evaluating the readability of text simplification output for readers with cognitive disabilities. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 293–299, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA).

[9] Gustavo Paetzold and Lucia Specia. Unsupervised lexical simplification for non-native speakers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[10] Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics*, 46(1):135–187, 2020.

[11] Sanja Štajner, Daniel Ferrés, Matthew Shardlow, Kai North, Marcos Zampieri, and Horacio Saggion. Lexical simplification benchmarks for english, portuguese, and spanish. *Frontiers in Artificial Intelligence*, 5, 2022.

[12] Christina Niklaus, André Freitas, and Siegfried Handschuh. Minwikisplit: A sentence splitting corpus with minimal propositions. *CoRR*, abs/1909.12131, 2019.

[13] Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P. Dinu. Exploring neural text simplification models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 85–91, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[14] Louis Martin, Éric de la Clergerie, Benoît Sagot, and Antoine Bordes. Controllable sentence simplification. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4689–4698, Marseille, France, May 2020. European Language Resources Association.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[16] Kim Cheng Sheang and Horacio Saggion. Controllable sentence simplification with a unified text-to-text transfer transformer. In *Proceedings of the 14th International Conference on Natural Language Generation*, pages 341–352, Aberdeen, Scotland, UK, August 2021. Association for Computational Linguistics.

[17] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

[18] Wei Xu, Chris Callison-Burch, and Courtney Napoles. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297, 2015.

[19] Tomoyuki Kajiwara and Mamoru Komachi. Text simplification without simplified corpora. *Journal of Natural Language Processing*, 25:223–249, 03 2018.

[20] Louis Martin, Angela Fan, Éric de la Clergerie, Antoine Bordes, and Benoît Sagot. MUSS: Multilingual unsupervised sentence simplification by mining paraphrases. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1651–1664, Marseille, France, June 2022. European Language Resources Association.

[21] Mikel Artetxe and Holger Schwenk. Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610, 2019.

[22] Xinyu Lu, Jipeng Qiang, Yun Li, Yunhao Yuan, and Yi Zhu. An unsupervised method for building sentence simplification corpora in multiple languages. In *EMNLP*, 2021.

[23] Shuhao Gu, Jinchao Zhang, Fandong Meng, Yang Feng, Wanying Xie, Jie Zhou, and Dong Yu. Token-level adaptive training for neural machine translation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1035–1046, Online, November 2020. Association for Computational Linguistics.

[24] Christian Bentz, Tatyana Ruzsics, Alexander Koplenig, and Tanja Samardžić. A comparison between morphological complexity measures: Typological data vs. language corpora. In *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity (CL4LC)*, pages 142–153, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.

[25] Giovanna Roda, John Tait, Florina Piroi, and Veronika Zenz. Clef-ip 2009: Retrieval experiments in the intellectual property domain. In Carol Peters, Giorgio Maria Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peñas, and Giovanna Roda, editors, *Multilingual Information Access Evaluation I. Text Retrieval Experiments*, pages 385–409, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[26] Florina Piroi. CLEF-IP 2010: Retrieval experiments in the intellectual property domain. In Martin Braschler, Donna Harman, and Emanuele Pianta, editors, *CLEF 2010 LABs and Workshops, Notebook Papers, 22-23 September 2010, Padua, Italy*, volume 1176 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2010.

[27] Florina Piroi, Mihai Lupu, Allan Hanbury, and Veronika Zenz. Clef-ip 2011: Retrieval in the intellectual property domain. 01 2011.

[28] Florina Piroi, Mihai Lupu, A. Hanbury, Walid Magdy, Alan Sexton, and I. Filippov. Clef-ip 2012: Retrieval experiments in the intellectual property domain. *CEUR Workshop Proceedings*, 1178, 01 2012.

[29] Florina Piroi, Mihai Lupu, and Allan Hanbury. Overview of clef-ip 2013 lab. In Pamela Forner, Henning Müller, Roberto Paredes, Paolo Rosso, and Benno Stein, editors, *Information Access Evaluation. Multilinguality, Multimodality, and Visualization*, pages 232–249, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[30] Walid Shalaby and Wlodek Zadrozny. Patent retrieval: a literature review. *Knowledge and Information Systems*, 61:631–660, 2019.

[31] Mihai Lupu, Allan Hanbury, et al. Patent retrieval. *Foundations and Trends® in Information Retrieval*, 7(1):1–97, 2013.

[32] Gabriela Ferraro, Hanna Suominen, and Jaume Nualart. Segmentation of patent claims for improving their readability. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 66–73, Gothenburg, Sweden, April 2014. Association for Computational Linguistics.

[33] Masayuki Okamoto, Zifei Shan, and Ryohei Orihara. Applying Information Extraction for Patent Structure Analysis. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, page 989–992. Association for Computing Machinery, 2017.

[34] Akihiro Shinmori, Manabu Okumura, Yuzo Marukawa, and Makoto Iwayama. In *Proceedings of the ACL-2003 Workshop on Patent Corpus Processing - Volume 20*, PATENT '03, page 56–65, USA, 2003. Association for Computational Linguistics.

[35] Linda Andersson, Mihai Lupu, and Allan Hanbury. Domain Adaptation of General Natural Language Processing Tools for a Patent Claim Visualization System. In Mihai Lupu, Evangelos Kanoulas, and Fernando Loizides, editors, *Multidisciplinary Information Retrieval*, pages 70–82, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[36] Svetlana Sheremetyeva. Automatic Text Simplification For Handling Intellectual Property (The Case of Multiple Patent Claims). In *Proceedings of the Workshop on Automatic Text Simplification - Methods and Applications in the Multilingual Society (ATS-MA 2014)*, pages 41–52, Dublin, Ireland, August 2014. Association for Computational Linguistics and Dublin City University.

[37] Leo Wanner, Ricardo Baeza-Yates, Sören Brügmann, Joan Codina, Barrou Diallo, Enric Escorsa, Mark Giereth, Ioannis Kompatsiaris, Symeon Papadopoulos, Emanuele Pianta, Gemma Piella, Ingo Puhlmann, Gautam Rao, Martin Rotard, Pia Schoester, Luciano Serafini, and Vasiliki Zervaki. Towards content-oriented patent document processing. *World Patent Information*, 30:21–33, 03 2008.

[38] N. Bouayad-Agha, G. Casamayor, G. Ferraro, L. Wanner, H.W. Guesgen, and H.C. Lane. Simplification of Patent Claim Sentences for their Paraphrasing and Summarization. In *FLAIRS - PROCEEDINGS, International Florida Artificial Intelligence Research Society Conference, 22nd, International Florida Artificial Intelligence Research Society Conference*, pages 302–303. Aaai Press;, 2009.

[39] Simon Mille and Leo Wanner. Making Text Resources Accessible to the Reader: the Case of Patent Claims. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, May 2008. European Language Resources Association (ELRA).

[40] Katharina Wäschle and Stefan Riezler. PatTR: Patent Translation Resource, 2014.

[41] Katharina Wäschle and Stefan Riezler. Analyzing parallelism and domain similarities in the marec patent corpus. In Michail Salampasis and Birger Larsen, editors, *Multidisciplinary Information Retrieval*, pages 12–27, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[42] R. Flesch. *How to Write Plain English: A Book for Lawyers and Consumers*. Harper & Row, 1979.

[43] J. Peter Kincaid, Robert P. Fishburne, R L Rogers, and Brad S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. 1975.

[44] Sai Surya, Abhijit Mishra, Anirban Laha, Parag Jain, and Karthik Sankaranarayanan. Unsupervised neural text simplification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2058–2068, Florence, Italy, July 2019. Association for Computational Linguistics.

[45] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11328–11339. PMLR, 13–18 Jul 2020.

[46] Yuan Zhang, Jason Baldridge, and Luheng He. PAWS: Paraphrase adversaries from word scrambling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1298–1308, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[47] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415, 2016.

[48] Fernando Alva-Manchego, Louis Martin, Carolina Scarton, and Lucia Specia. EASSE: Easier automatic sentence simplification evaluation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 49–54, Hong Kong, China, November 2019. Association for Computational Linguistics.

[49] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.

[50] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. BERTScore: Evaluating Text Generation with BERT. In *International Conference on Learning Representations*, 2020.

[51] Asli Celikyilmaz, Elizabeth Clark, and Jianfeng Gao. Evaluation of text generation: A survey. *ArXiv*, abs/2006.14799, 2020.

[52] Elena Lloret, Laura Plaza, and Ahmet Aker. The Challenging Task of Summary Evaluation: An Overview. *Lang. Resour. Eval.*, 52(1):101–148, March 2018.

[53] Akihiro Shinmori and M. Okumura. Aligning Patent Claims with Detailed Descriptions for Readability. In *NII Testbeds and Community for Information Access Research*, 2004.

[54] Eyal Peer, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer. Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 54(4):1643–1662, Aug 2022.

[55] John Robinson. *Likert Scale*, pages 3620–3621. Springer Netherlands, Dordrecht, 2014.