TRABAJO FIN DE GRADO

GRADO EN ESTUDIOS INGLESES: LENGUA, LITERATURA Y CULTURA

Reading the Terms of Service

A longitudinal study addressing the readability of the Terms of Service of the top five most visited websites in Spain

AMALIA HAFNER TÁBOAS

AHAFNER1@ALUMNO.UNED.ES

TUTOR ACADÉMICO: Prof. Eva Samaniego Fernández

LÍNEA DE TFG: Línea 2 – Inglés Profesional y Académico

FACULTAD DE FILOLOGÍA

CURSO ACADÉMICO: 2020-2021- Convocatoria: junio 2021

2

Abstract

This research argues that public awareness on data privacy and automated decision-making processes, and the growing interest towards clarity and transparency in legal and technical writing, have influenced the presentation of Web applications' Terms of Service (ToS). This paper explores if their readability level has improved over the last decade, by applying readability tests to the ToS of the top five most visited websites in Spain. Additionally, the presence of Plain English resources and new concepts related to automated decision-making processes are also examined.

Keywords

Readability, Plain English, Terms of Service, Legal English

Table of Contents

1. Introduction

1.1.   Motivation

I have been following the public debates on Web-based services transparency since I started working as a technology teacher, almost 15 years ago. At that time, discussions were mostly centred on privacy policies, especially pointing to social networks, with Facebook as the usual suspect. Several public scandals have stained social networks' reputation ever since and the public debate warmed up somewhat. Students became increasingly aware of the risks of not reading (or not understanding) the terms and conditions of the services they used.

Later, over the last few years working for organisations related to technology policy and development, I became increasingly interested in transparency at an additional level: are users aware of the algorithms that make automated decisions affecting us? These algorithms can be harmless, even useful, as in the case of the recommended books that Amazon offers its customers. But others can be indeed harmful, negatively impacting citizens' access to health, credit, and justice, for example.

This area, where language, law, and technology meet, is of huge interest beyond the scope of the usual stakeholders, such as translators, technical writers, lawyers, policymakers, and tech-companies: citizens are increasingly concerned. The growing impact of online services is inescapable, affecting all areas of life. But do users have enough information on this when accepting the Terms of Service? If we do, do we read them? If we honestly say yes, do we understand them?

1.2.   Relevance of the topic

Every time users sign-up to Web-based services or applications, they need to agree to specific terms and conditions. These terms constitute the legal agreements that regulate the rights and obligations of the user and the company providing the service.

The Terms of Service (ToS) are legally binding. Both the service provider and the user can enforce the agreement by refusing service and filling a lawsuit, respectively.

Terms and conditions are subject to change. New versions of the terms can be issued following decisions made by the service providers (due to upgrades in the service they provide, mergers with other services, or downsizing, for example). In addition, the terms might be edited to comply with State regulations affecting the rights and obligations that the terms address.

The usage of online services and applications has extended worldwide, and a vast range of citizens interact with such services daily. The services have increasingly become a necessity for many users, companies, and institutions.

In this context, the way in which users read, understand, and interpret the terms and conditions has gained an increasing interest, raising the awareness of users, academics, and policymakers, especially regarding data protection and privacy.

These legal texts have been addressed as objects of study to evaluate their readability. According to the literature review included in section 1.4.2, the conclusions are not promising yet: the readability of the ToS is still quite low. However, since the growing concerns of users are quite recent, it is worth exploring if there are changes in the ToS that might suggest a potential improvement in readability, as well as the inclusion of new topics in them.

Several elements have probably affected the presentation of the ToS of Web-based applications. In the recent past, the plain English movement made an important impact on legal texts and technical writing, increasingly concerned with clarity and transparency. Moreover, the coming into force of the GDPR in Europe pushed the legal background of Web-centred companies into public debate: ToS, privacy settings and data policies gained users' attention. This debate nested on the growing public awareness about automated decision-making processes, reflected in many best-seller books and much-viewed documentaries.

The "right to explanation", specifically in the field of machine-learning, can be defined as the right users have to be given an explanation on the output of an algorithm. Algorithms used in this field do not directly depend on the designers' decisions in a

straightforward manner; they are modified by automated processes instead. Due to the inherent difficulties of explaining the processes behind automated decisions, and the reluctance of some companies to explicitly describe their trade secrets, the "right to explanation" is under strong debate.

This debate is reflected in the political and public media arena, especially so in Europe after the coming into force of the GDPR. However, it is also addressed by academia: several researchers from different fields -but especially Technology and Law- are focusing on the explanations that Web-based companies are offering consumers.

The number of academic sources addressing the "right to explanation" shows a steady growth. As an example, the Google Scholar database shows no results when searching for academic articles addressing the "right to explanation" for the years 1990-2000. One decade later, the period 2009-2010 throws five results. Now, when browsing the range 2010-2020, the database shows 1580 entries.

However, it is not easy to find straightforward information compliant with the public demand towards a "right to explanation" in the legal texts that Web-based services offer. Also, based on previous research, the availability of this kind of information would not necessarily guarantee that users will read it and understand it.

1.3. Hypothesis, research question and objectives

This research considers, as a point of departure, that recent social trends have an impact in the way that companies decide to present their Terms of Service. As a hypothesis, this paper argues that the increasing public awareness of the risks associated with data privacy and automated decision-making processes, and the growing interest towards clarity and transparency in legal and technical writing have influenced the presentation of Web applications' Terms of Service.

With this hypothesis in mind, the research question is: Are the Terms of Service of five popular Web-based applications improving their readability over time?

The specific objectives of this piece of research are the following:

- Analyse the readability of the Terms of Service of five popular Web-based applications in Spain, as of February 2021[1].

- Compare the readability level of the Terms of Service for two periods in time: 2010 and 2020. In case the Terms of Service for these years are not available, the comparison will consider the closest dated version.

- Identify the language resources recommended by the Plain English movement to improve the Terms of Service readability.

- Explore if new topics related to automated decision-making processes have been added to the scope of the Terms of Service over time for the sample under analysis.

These objectives will be addressed through an analysis of ten ToS texts (the applicable ToS of five Web-based services in two periods in time. The analysis will be conducted by applying well-established readability analysis techniques: the Flesch-Kincaid (F-K) test and the Dale-Chall test. Additionally, the analysis will consider approach, style, and register, and it will identify the presence of elements related to social trends such as the Plain English Campaign and the "right to explanation" debates.

1.4.   State of the art

1.4.1. Readability definition

According to the Cambridge Dictionary (2021), readability is a noun defined as "the quality of being easy and enjoyable to read"[2]. Readability, from a general-purpose language perspective, implies two elements: it depends on the text's *presentation*, related to typographic aspects and overall legibility, including the visual arrangement of the text- and on the text's *content* -lexical and syntactic complexity-.

---

[1] According to the periodic report published by the consultancy company Similarweb. Source: https://www.similarweb.com/top-websites/spain/ accessed 1 February 2021.
[2] Readability. 2021. In: Cambridge Dictionary. Retrieved December,1, 2021, from https://dictionary.cambridge.org/dictionary/english/readability

DuBay states that these elements should not be confused: "Readability is what makes some texts easier to read than others. It is often confused with legibility, which concerns typeface and layout" (DuBay, 2004, p. 3). This study will thus focus on the former.

This piece of research follows the definition of readability proposed by Edgar Dale and Jeanne Chall (1949): "The sum total (including all the interactions) of all those elements within a given piece of printed material that affect the success a group of readers have with it. The success is the extent to which they understand it, read it at an optimal speed, and find it interesting".

1.4.2. Readability tests

Following DuBay (2006; 2007), early works that aimed at addressing readability can be dated as early as the 19th century: In 1893, the English Literature scholar L.A. Sherman proposed a method to analyse and teach literature using a statistical approach. His work showed how sentences became shorter over time and hypothesized that the average length of sentences had an impact on the text's readability. A few years later, in 1889, in Russia, R.A. Rubakin published a comprehensive study of word frequency, with the aim to identify the vocabulary that most readers could understand.

The systematic interest of academics from different fields, ranging from Philology to Library Studies and from Marketing to Educational Psychology, developed into the first formulae to quantitatively assess readability (Dale and Tyler, 1934; Gray and Leary, 1935; Lorge, 1939). It is important to note that these early studies aimed at addressing several textual elements: content, form, organisation, and style. However, the quantitative approach was mostly focused on style. Additionally, vocabulary complexity was considered a relevant element to address in the analyses.

By the 1940s, readability analysis had become a well-established field of study. Rudolph Flesch, in his 1948 article, proposed a readability formula with two parts. The first one, called the Reading Ease formula, counted the number of syllables and the number of sentences per 100 words. The second part of the formula addressed the

text's potential interest for a hypothetical reader, counting the number of pronouns, personal names, quotes, and exclamations, among other elements.

Also, in 1948, Edgar Dale and Jeanne Chall published a different readability formula. Unlike Flesch's proposal, the Dale-Chall readability formula focused on vocabulary load to measure the difficulty of a text. This formula counted the number of "hard words" that appear in a given text. These are all those words not included in a list of frequent vocabulary ("the Dale list"), consisting of 3000 common terms in English. The test is further in section 2.2.1.

Flesch's Reading Ease Formula "became the most widely used formula and one of the most tested and reliable" (DuBay, 2007, p. 58). With the intention to further simplify this formula, the Reading Ease score was replaced with a grade-level scale. This modified formula, called the Flesch-Kincaid test, is the most extensively used technique to assess readability nowadays. The formula presented in Kincaid, Fishburne, Rogers and Chissom (1975) is described in section 2.2.1.

It is worth noting that these techniques were developed by English-speaking researchers addressing English-written texts. The syllable count measure is particularly relevant in English, since longer words tend to stem from Latin and Greek roots, which are usually identified with a higher register and thus, are more difficult to understand by the general public. The impact of word length in readability is not necessarily applicable to other languages.

Considering the current state of the art in the field of readability analysis, even though these tests are very popular, they are not exempt from criticism.

Meta-studies (Klare, 1974) compared different formulae applied to readability and suggested that the most appropriate approaches to address texts depend on the needs and kind of application -manual versus machine-, among other considerations. This means that the results obtained by applying readability tests to different kind of texts might not be comparable.

Other studies have centred on criticising the usefulness of the variables that popular readability tests try to measure. Davison and Kantor (1982) argue that, when adapting texts to make them compliant with the assumptions of readability tests (by shortening

sentences or choosing words with less syllables) experimental tests about reading comprehension suggest that the overall understanding does not necessarily increase.

More recent meta-studies (Zhou, Jeong and Green, 2017) focus on evaluating readability tests' consistency. These authors compared the scores resulting from the application of the most extensively used readability formula, the Flesch-Kincaid test, and other commonly used tests (the Gunning-Fog Index, the SMOG Index, the Coleman-Liau Index and the Automated Readability Index) to a sample of highly technical texts. The study concluded that there were substantial differences in the results, which affects the reliability of the tests results as indicators to measure readability.

Some studies have addressed the changes of readability over time. Longitudinal studies yield similar results for different fields: Pavlén-Sigray, Matheson, Schiffler and Thompson (2017) conclude that scientific texts' readability is decreasing; Jones (2012) arrives at a similar conclusion regarding corporate reports. As per legal texts - specifically centred around the digital realm- Milne, Culnan and Greene (2006) also found that readability had declined.

Within the field of readability analysis, several studies have applied mixed methods. These methods are related to very diverse fields. In some studies, quantitative and qualitative techniques are applied to a readability analysis. For example, Stahl (2003) explores how readability formulae can be complemented by an assessment of lexical complexity, beyond the syllable count. Paasche-Orlow, Taylor and Brancati (2003) pay attention to contextual factors that have a potential impact on readability, such as the local literacy rates and the level of research activity and funding in the area.

1.4.3. Terms of Service analyses

The ToS of Web-based services and applications are increasingly taken as an object of study. These analyses are conducted using a wide range of methods and techniques.

Several studies focused especially on the privacy policies of Web-based services. Graber, D'Alessandro, and Johnson-West (2002) analysed the privacy policies of 80 health-focused websites; Jensen and Potts (2004) examined 64 websites and Proctor, Ali and Vu (2008) investigated the policies published on 10 websites. All these studies concluded that the texts are not generally readable by the general public: the resulting scores produced by readability tests range between a 13th and 14.5th grade. Moreover, even in cases in which readability tests produced a 13th grade level score, subsequent experiments focused on reading comprehension showed that results were questionable. College students were only capable of answering correctly around 50% of questions the researchers asked on specific policies (Proctor, Ali & Vu, 2008).

Becher and Benoliel (2019) applied the Flesch Reading Ease test and the Flesch-Kincaid test to 500 sign-in contracts for Web-based services and applications. They concluded that the average readability level of these texts was equivalent to the average academic publication, which are not easily understood by the general public.

Peslak, Kovalchik and Conforti (2020) conducted a qualitative, linguistics and sentiment analysis of Google's Privacy Policy for each available year between 2000 and 2018. Despite encountering a slight decrease in readability (using the grade-level score, described in the previous section), the authors found that these policies have become "more positive in sentiment and more personal in approach".

Due to its increasing size and complexity, researchers are addressing the corpus consisting of Terms of Service texts through complex machine-based techniques, leveraging the fact that this corpus can be considered as a source for "big data", where machine-learning techniques can be successfully applied. Palka and Lippi (2019) studied the results of training a machine-learning tool to automatically detect unfair clauses in the Terms of Service.

From a Cognitive Psychology perspective, the way in which consumers usually read and accept these terms has been thoroughly studied from a vast array of perspectives. Steinfeld (2016) used eye-tracking technology to study if -and how- users read the privacy policy, comparing different ways to access the policy text (presented by default or accessible through clicking on a link) and different privacy attitudes of the participants. The author concluded that, regardless of the users' attitudes towards

privacy, users spend more time reading the privacy policy when it is presented by default.

Obar & Oeldorf-Hirsch (2018) designed an experimental study in which participants had to agree on the Privacy Policy and Terms of Service of a fictitious social network. The researchers included "gotcha clauses" as part of the legal texts. For example, by agreeing on the terms, users were committing to provide their first-born child as payment to access the social network service. Only 2% of participants read these clauses.

2. Analysis and results

This section is structured in three parts:

Firstly, it presents the **methodology** employed in this paper, including a description of the sample of texts that constitute the corpus, the methods and techniques, and the resources (repositories and software) used to analyse this corpus.

Secondly, through two complementary approaches, the **readability analysis** and its results are presented. This sub-section addresses objectives 1) and 2) of this piece of research. The objectives addressed in this section are marked in bold.

Last, this section addresses the exploration of elements linked to the two **social trends** identified in section 1.2.: The Plain English Campaign and the "right to explanation" debates. The outcome of this exploration feeds objectives 3) and 4). The objectives addressed in this section are marked in bold.

2.1.   Methodology

2.1.1. The sample

The sample of texts is based on the top 5 most visited websites in Spain as of February 2021[3], in the following order: Google, Youtube, Facebook, Twitter, and Amazon.

This study compares the currently applicable terms versus the terms applicable in 2010. The year 2010 was chosen as representative of older versions of the terms for all the companies under the scope of this study, regardless of their foundation date. The analysis will thus address the ToS of the five selected Web services in two distinct periods: the currently applicable ToS and the terms that were in force one decade earlier.

All of them publicly offer their currently applicable ToS online. Google, Twitter, Facebook and Amazon allow public access to an archive of previous terms. In the case of Google and Twitter, it is possible to access all versions of the ToS. Facebook and Amazon only display, as oldest available versions, 2018 and 2011, respectively. Youtube grants access to the immediately previous applicable terms, dated in 2019. The resources used to gather the ToS are described in section 2.1.3.

The corpus under analysis is composed of the ToS offered in English language and applicable in the European Union and UK.

As a summary, this research addresses ten texts, publicly accessible:

- Google: ToS applicable in 2020 and ToS applicable in 2010

- Youtube: ToS applicable in 2020 and ToS applicable in 2010

- Facebook: ToS applicable in 2020 and ToS applicable in 2010

- Twitter: ToS applicable in 2020 and ToS applicable in 2010

---

[3] According to the periodic report published by the consultancy company Similarweb. Retrieved 1 Febuary 2021 from https://www.similarweb.com/top-websites/spain/

- Amazon: ToS applicable in 2020 and ToS applicable in 2011[4]

A complete list of the sources from which the ToS were obtained is included as Annex 1.

2.1.2. Methods and techniques

Considering the limitations of the readability tests, mentioned in section 1.4.2, this research makes use of a mixed methodologies approach. The application of quantitative-focused tests is complemented with a qualitative assessment.

The sample of texts presenting Terms of Service will be analysed through:

1. A quantitative assessment, which consists of applying well-established readability analysis techniques: the Flesch-Kincaid (F-K) test and the Dale-Chall test. With these tests, this paper will assess the average length of sentences, the number of syllables per word and the vocabulary load as indicators of the difficulty implied in reading the texts.
2. A qualitative assessment that focuses on describing pragmatic elements, such as the approach (more or less personal), the style (more or less formal) and register (more or less high) employed in the ToS. Moreover, and from a qualitative perspective too, this study will aim to identify the presence of elements related to social trends such as the Plain English Campaign and the "right to explanation" debates.

2.1.3. Resources

To gather the texts that compose the corpus of this study, we used the official websites of each company. As previously mentioned, we were able to get most texts from the corresponding websites: all ToS for Google, Twitter, and Amazon. As per Facebook and YouTube, we extracted the currently applicable terms from the official website and the 2010 ToS using snapshots available through the Wayback Machine, offered by the

---

[4] The ToS of Amazon for the year 2010 (in English language and applicable to European countries) was not available. In this case, we analysed the ToS applicable in 2011.

Internet Archive[5]. There were no snapshots available for the Amazon ToS in 2010. For this reason, we used the text that was closer to the desired date, 2011.

A complete list of the sources from which the ToS were obtained is included as Annex 1.

The analyses conducted during this study made use of the following resources:

- The Python package Textstat was used to apply the Flesch-Kincaid test and the Dale-Chall test. "Textstat is an easy-to-use library to calculate statistics from text. It helps determine readability, complexity, and grade level"[6], including the scripts to process datasets applying the readability tests (see section 2.2.1).
- Additionally, a custom-developed Pyhton script was used to examine the frequency of occurrence of words considered difficult, based on the list of alternatives published by the Plain English Campaign[7] (see section 2.3.1) and the keywords related to the "right to explanation" (see section 2.3.2).
- To calculate the delta between different versions of the ToS per company, we used Microsoft Excel.

2.2.   Readability analysis

This section addresses the following objectives:

- Analyse the readability of the Terms of Service of the five most popular Web-based applications in Spain, as of February 2021.

- Compare the readability level of the Terms of Service for two periods in time: 2020 and 2010.

As discussed in section 2.1.2 on Methodology, this study presents the results of two distinct approaches in the analysis.

---

[5] The Wayback Machine allows access to a repository of saved snapshots from the Web. This tool is accessible through the Internet Archive, a non-profit "digital library of Internet sites and other cultural artifacts in digital form". Retrieved 31 March 2021 from https://archive.org/web/
[6] Readability (2020) Retrieved 31 March 2021 from https://pypi.org/project/textstat/
[7] The A to Z of alternative words. Retrieved 30 March 2021 from http://www.plainenglish.co.uk/files/alternative.pdf

Firstly, following a quantitative approach, the selected corpus is analysed through the application of the Flesch-Kincaid test and the Dale-Chall test. Readability scores for each text will be presented and compared.

Secondly, within a qualitative approach, the texts will be analysed and compared in terms of approach, style, and register.

2.2.1. Quantitative approach

As discussed in section 1.4.1, the concept of "readability" can be addressed from different viewpoints. Following the comprehensive definition of readability developed by Dale and Chall (1949), this analysis will focus on estimating the success a group of readers would have in understanding a given text.

It is important to stress the idea of *estimation* in this type of analysis. When measuring a complex concept through formulae, several peculiarities of the texts, its readers, and the context of the interaction between readers and texts, are not considered. Furthermore, readability is analyzed independently from legibility and, thus, it does not ponder the impact of the visual arrangement of the texts under analysis.

However, despite the criticism that readability tests have received in the past, they are useful to "provide an objective prediction of text difficulty" (DuBay, 2004, p. 3).

Aware of the readability formulae's limitations, this paper makes use of readability tests as a first step towards an analysis of the corpus, complemented by a qualitative approach addressed in sections 2.2.2. and 2.3.

- ▪ Flesch-Kincaid test

The Flesch-Kincaid formula (Kincaid, et al., 1975), as an adaptation of Flesch's Reading Ease formula (Flesch, 1948), is the most extensively used readability test in the field. This test considers two main factors: the average sentence length, or ASL (the number of words divided by the number of sentences) and the average number of syllables per word, or ASW (number of syllables divided by the number of words).

These values are taken from groups of 100 words within a text. Thus, the values are independent from the texts' length.

The output of the Flesch-Kincaid formula is the Grade Level Score, which is equivalent to school grades (or years of formal instruction) within the United States education system.

The Flesch-Kincaid formula is:

Grade Level Score = (0.39 x average sentence length) + (11.8 x average number of syllables per word) – 15.59

The following table shows the equivalence between the Reading Ease score, Grade level and a brief description of the estimated difficulty.

Score	Grade level	Difficulty
90-100	5th grade	Very easy to read. The average 10–11-year-old student can understand the content. Conversational consumer English should score at least 80.
80-90	6th grade	
70-80	7th grade	Fairly easy to read. An average 12-year-old students can understand the content.
60-70	8th and 9th grade	Understood by 13–15-year-old students. Plain English should score at least 60.
50-60	10th to 12th grade (high school)	Fairly difficult to understand.
30-50	College undergraduates	Difficult to understand.
0-30	College graduates	Very difficult to read. Understood by college graduates or specialized professionals.

Table 1. Reading Ease Score, Grade level and difficulty. Adapted from Flesch (1949) "How to write in Plain English"

The application of the Flesch-Kincaid test to the corpus of this study produced the results summarized in the table below.

Company	F-K test score for the 2020 ToS	F-K test score for the 2010 ToS	Delta	Readability
Google	19.7	28.4	8.7	↘
Youtube	15.7	24.4	8.7	↘
Facebook	21.8	45.0	23.2	↘
Twitter	20.2	19.9	-0.3	=
Amazon	16.7	15.0	-1.7	=

Table 2. Flesch-Kincaid test score per ToS under analysis and delta per company

These results show that all texts in our scope indicate a (very) high difficulty. Based on the Flesch Reading Ease score (1948), the lower the score, the higher the difficulty. Scores below 30 can be identified as suitable for college graduates; scores between 30-50 can be presumably understood by college students.

Except Facebook's ToS applicable in 2010, which would be, according to the Flesch-Kincaid test, understood by a college undergraduate, all other texts imply an even higher degree of difficulty. When comparing these results to the literature review (in section 1.4.3) the outcome is in the same line.

Moreover, the delta between both versions of the ToS per company shows that texts have become more difficult to read over the last decade. The only exceptions are shown by the slight decrease in Twitter and Amazon's ToS scores. However, with such low scores, this does not translate into an improvement in readability.

Peslak, Kovalchik and Conforti (2020) had arrived at a similar conclusion when comparing the results of the Flesch-Kincaid test to Google's privacy-related terms. The researchers found that the readability of the terms had decreased over time (between the years 2000 and 2018).

▪ Dale-Chall test

The Dale-Chall formula (1949) was developed to address one of the shortcomings its authors perceived in Flesch's Reading Ease test. Similarly to the latter (and thus to its modified version, the Flesch-Kincaid test), it assesses the average sentence length. But, instead of evaluating the average number of syllables per word, it considers vocabulary load as a variable. This is analyzed by checking the occurrence of words that are not included in a list of 3000 words commonly used in English, the Dale-Chall list.

The Dale-Chall formula reads as follows:

> Raw score = .1579 percentage of difficult words, or PDW (words not on the Dale-Chall list) + .0496 average sentence length, or ASL + 3.6365

The table below shows the equivalence between the raw score resulting from the Dale-Chall test and the corresponding Grade level, equivalent to the educational grade.

Raw score	Grade level
4.9	4 th and below
5.0 – 5.9	$5^{th} – 6^{th}$ grade
6.0 – 6.9	$7^{th} – 8^{th}$ grade
7.0 – 7.9	$9^{th} – 10^{th}$ grade
8.0 – 8.9	$11^{th} – 12^{th}$ grade

9.0 – 9.9	College undergraduates
10 and above	College graduates

Table 3. Dale-Chall score-correction chart (DuBay, 2004: 23)

As an output of the analysis of the 10 texts of this corpus, the Dale-Chall raw score achieved by each text is included in the table below.

Company	D-C test score for the 2020 ToS	D-C test score for the 2010 ToS	Delta	Readability
Google	7,68	8,54	0.86	⬆
Youtube	7,33	8,32	0.99	⬆
Facebook	7,86	10,58	2.72	⬆
Twitter	8,14	8,06	-0,08	=
Amazon	7,08	7,13	0.05	=

Table 4. Dale-Chall test score per ToS under analysis and delta per company

These results indicate that there are some improvements in terms of readability. The raw scores resulting from the application of the Dale-Chall test shows that the 2020 texts under analysis could presumably be understood by 10th graders (15 or 16-year-old high-school students).

Except for Twitter and Amazon's scores, which remained almost invariable, the scores show a slight improvement when compared to the older versions of the ToS. The 2010 ToS show scores equivalent to more advanced high-school students, and even college graduates in the case of Facebook.

By comparing the outcome of the Dale-Chall test application with the Flesch-Kincaid test, we can contribute to the conclusion held by Zhou, Jeong and Green (2017). Our analysis does not find consistent results when comparing the output of two different readability tests. While the Flesch-Kincaid test results indicate that readability has decreased over the last decade, the Dale-Chall formula application suggests the contrary.

In this context, we can hypothesize that despite the increased length of sentences, the choice of words has become more appropriate for the readers. We will get back to this idea in section 2.3.1

Preliminary conclusions

To sum up this section, and as a first finding of this study, the ToS of the five most popular Web-based applications in Spain have changed in terms of readability over the last decade.

When examining the average sentence length and the average number of syllables per word, the texts are very difficult, and they have become slightly more so over the years. This conclusion is in line with other pieces of research addressing the readability of ToS texts, mentioned in section 1.4.3 (Peslak, Kovalchik & Conforti, 2020; Graber, D'Alessandro & Johnson-West, 2002 and Jensen & Potts, 2004, Proctor, Ali & Viu, 2008).

However, if considering the vocabulary load, results indicate that the ToS have become more readable and, thus, presumably easier to understand.

2.2.2. Qualitative approach

This section focuses on other aspects that indicate the potential difficulty in reading a text. In this vein, approach, style, and register are explored, as a way to identify changes in each company's ToS. Furthermore, this section briefly addresses the information display, examining the length of the texts and the number of links offered throughout them.

Approach, style, and register

The corpus of this study is composed of legal texts. Legal English can be considered as a sub-field within English for specific purposes (ESP), sharing its tenets.

When analysing the ToS, we can identify some aspects of specialised English. However, there are some peculiarities worth mentioning.

In legal texts, the **approach** is generally impersonal, with a high occurrence of passive sentences. The use of passive constructions is one of the most salient syntactic features of ESP. When leaving the agent out of the sentence, the results of the actions are emphasised.

In addition, phrases tend to be written in third person singular, centred -in many legal texts- on the two parties to a private contract: the consumer and the company.

Even though the ToS under analysis were already showing a more personal approach, by using personal pronouns in the 2010 versions -referring to the user as "you" and the company as "we"-, it is interesting to note that some ToS show an increasing trend towards personalisation. Some examples of the more personal approach in the ToS are included below:

- Google
    - (2010) "In order to use the Services, you must firstly agree to the Terms. You may not use the Services if you do not accept the Terms"
    - (2020) "Understanding these terms is important because, to use our services, you must accept these terms".
- Facebook
    - (2010) "We do not guarantee that Platform will always be free".
    - (2020) "We don't charge you to use Facebook or the other products and services covered by these Terms. Instead, businesses and organizations pay us to show you ads for their products and services".

Regarding **style,** legal texts are normally rather formal, with a high **register**. It is possible to define register -similarly to style- as "a set of connected linguistic practices that are associated with a group of speech situations in a speech community, or with an institution in a speech community" (Kiesling, 2011, 94).

22

In a similar line to our analysis of approach, the ToS are generally closer to the traditional formal style of legal texts. However, there are some examples showing an increasingly informal style. This is achieved by creating a closer relationship with the reader through establishing complicity with the user and including examples with which they can identify (the bold is ours).

- Twitter (2010) "What you say on Twitter may be viewed all around the world instantly. **You are what you Tweet!"**
- Facebook (2020) "We allow advertisers to **tell us things** like their business goal, and the kind of audience they want to see their ads (**for example, people between the age of 18-35 who like cycling**)".
- Google (2020): **"We know it's tempting** to skip these Terms of Service, but it's important to establish what you can expect from us as you use Google services, and what we expect from you".

It is possible to identify the high register of a text through its use of words with Latin and Greek roots. The occurrence of these words is usually higher in legal English than in other specialized uses of the language. Words that stem from Latin and Greek are usually longer and have more syllables, impacting on a higher grade-level score in the Flesch-Kincaid test. In addition, these words are not generally included in the frequent word lists, like the one used by the Dale-Chall formula, thus potentially throwing higher scores.

However, as argued in section 2.2.1, we might identify an incipient trend towards improving readability through a less difficult vocabulary choice (the bold is ours):

- Facebook (2020) "Where we take such action, we'll **let you know** and explain any options you have to request a review […]"
- Twitter (2010) "You can **opt-out** of most communications from Twitter".
- Google (2020) "We want to maintain a respectful environment for everyone, **which means** you must follow these basic rules of conduct […]"

Information display

Even though this study does not analyse the visual arrangement of the texts within the corpus, it is useful to consider aspects of the information display that might impact the texts' difficulty.

Firstly, following Peslak, Kovalchik and Conforti (2020), it is useful to ponder the length of the texts under analysis. An increasing complexity in the texts can be estimated through an increasing **word count**. The following table shows the word count for each text in the corpus.

Company	Word count for the 2020 ToS	Word count for the 2010 ToS	Delta	Text length
Google	3586	3990	-404	↘
Youtube	3584	3731	-147	↘
Facebook	4134	4039	-439	↘
Twitter	3151	2537	175	↗
Amazon	6668	4547	2121	↗

Table 5. Word-count per ToS under analysis and delta per company

As displayed in the table above, the ToS of Google, Youtube and Facebook have become shorter over the last decade, with less words in use. For the same period, the ToS for Twitter and Amazon have grown longer.

However, these results should not be considered as a straightforward measure of complexity. A given text might have become shorter in terms of word count but, at the same time, it might have been enriched through other means.

It is very important to note that ToS are, by default, subject to change. As discussed in section 1.2 these changes might respond to company-internal motivations (new services, mergers, upgrades, etc.) or external obligations, such as new regulation. When the GDPR came into force in Europe, in 2019, the ToS had to be rewritten. This, however, did not necessarily result in longer texts, as we saw examining words count.

In parallel to the length of texts, it is useful to analyse how the information is structured. The texts under analysis are displayed online and, in this context, the information is

usually accessible through **links**. A reader might easily navigate the texts, clicking on the available links to read relevant further information, to move forward or come back to previous screens.

In this vein, it is interesting to note that, regardless of the texts' length, the number of links included in them has grown as well. The table below shows the number of links included in each ToS under analysis.

Company	Number of links in the 2020 ToS	Number of links the 2010 ToS	Delta	Links
Google	6	40	34	
Youtube	6	31	25	
Facebook	35	43	8	
Twitter	5	38	33	
Amazon	15	30	15	

Table 6. Number of links per ToS under analysis and delta per company. The number of links refers to the hyperlinks accessible through the body of the ToS and pointing to sites different from the ToS site. This does not include other links (navigation pane, contextual links, cross-references within the same site, or e-mail addresses, for example).

Preliminary conclusions

The results obtained from the application of readability tests, addressed in section 2.2.1 are complemented by the analysis of approach, style, and register. Even though the aspects usually identified with legal English as a genre are still central, we have identified slight changes in the ToS. These examples might indicate a trend towards a more personal approach and informal style. We will get back to this when exploring the Plain English Campaign in section 2.3.1. However, these preliminary conclusions need to be considered together with the potential difficulties posed by longer texts and a richer information display.

2.3.   Social trends

This section of the paper addresses two social trends that may have had an impact on the way the ToS are written: the Plain English Campaign and the "right to explanation"-related debates. The sub-sections below examine elements linked to both trends, which we identified in the texts under analysis.

2.3.1. The Plain English movement

Firstly, we will address the third research objective of this study:

- Identify the language resources recommended by the Plain English Campaign to improve the Terms of Service readability.

From a historical perspective, a social trend towards a clearer and plainer language in legal texts started a very long time ago. Looking at the history of the English language (Baugh & Cable, 2002), it is worth noting that legal texts were originally written in Latin or French, following the Battle of Hastings in 1066. Even though the vast majority of the people were English native-speakers, and the Statute of Pleading (1362) acknowledged that fact, it was not until then that English became the language used in Court.

Much closer to our time, Legal English still carries Greco Latin and French influence with it. Legal texts (Statutory Law, Case Law, but also contracts such as the Terms of Service users agree to when using online services) tend to be filled with a Latin and Greek-based vocabulary, in the form of borrowings, adaptations and archaisms. Regarding syntax, legal English tends to use longer and more complex sentences, with a high frequency of subordinated clauses. From a pragmatic perspective, legal texts tend to show an impersonal approach, a high register, and a very formal style.

All these elements contribute to increasing the difficulty in reading and understanding legal texts. The Plain English movement gained momentum during the 1960s and 1970s, to persuade members of the Parliament, the Judiciary, and bureaucrats to write clearer texts, in layman's terms. In addition to the publication of popular books like The

Language of Law (Mellinkoff, 1963) and Plain English for Lawyers (Wydick, 1978), the Plain English Campaign was launched in the United Kingdom in 1979[8].

The impact of the trend towards Plain English can be identified in examples such as, the Plain Writing Act, passed in 2010 in the United States and the Unfair Terms in Consumer Contracts Regulations (1999) in the United Kingdom.

The following list shows a summary of the Plain English Campaign guidelines[9]:

a) Keep sentences short
b) Avoid nominalizations
c) Use vocabulary that is appropriate for the target reader
d) Favour personal pronouns ("you", "yours", "we", "our", "us")
e) Choose active verbs over passive sentences
f) Use the imperative mood where appropriate

Some guidelines can be linked to the Flesch-Kincaid test and the Dale-Chall test, described in section 2.2.1. above. These tests addressed:

- The average sentence length, related to guideline (a)
- The average number of syllables per word, which might be a proxy for nominalizations, related to guideline (b), since these usually have more syllables.
- The vocabulary load related to guideline (c) can be inferred from the grade-level scores that both tests throw as results. However, it is worth stressing that we found inconsistent results when comparing the Flesch-Kincaid test and the Dale-Chall test.

Guidelines (d), (e) and (f) were addressed through section 2.2.2. These recommendations are closer to a personal approach and a neutral style as an alternative to the usual style of legal texts.

---

[8] Timeline of Plain English Campaign. Retrieved 30 March 2021 from
http://www.plainenglish.co.uk/about-us/history/timeline.html
[9] Adapted from How to write in Plain English. Retrieved 30 March 2021 from
http://www.plainenglish.co.uk/files/howto.pdf

In addition to these common guidelines, the Plain English Campaign recommends writers to replace difficult words by a simpler alternative[10]. These alternatives usually focus on choosing English-root words over Latin or Greek-based words and favouring a more personal approach and neutral style instead of a formal one.

We analysed the ToS in our corpus to explore if there were changes in the choice of vocabulary. Improvements in this line would contribute to explain why texts became more readable according to the Dale-Chall test results, despite having a low readability according to the Flesch-Kincaid test.

To conduct this analysis, we extracted the words in need of an alternative from the Plain English Campaign list mentioned above. We will call the words in need of an alternative "obscure words". Next, using a custom-developed Pyhton script, we counted the relative occurrence of such "obscure words" (the sum of occurrence of all words in the list, divided by the total of words per text), comparing both versions of the ToS for each company.

The table below shows the average frequency of occurrence of the "obscure words" for each text under analysis.

Company	Average frequency of occurrence of "obscure words" in the 2020 ToS	Average frequency of occurrence of "obscure words" the 2010 ToS	Delta	Lexical difficulty
Google	0,04	0,051	0,011	↘
Youtube	0,048	0,065	0,017	↘
Facebook	0,039	0,068	0,029	↘
Twitter	0,062	0,068	0,006	↘

[10] The A to Z of alternative words. Retrieved 30 March 2021 from
http://www.plainenglish.co.uk/files/alternative.pdf

Company	Average frequency of occurrence of "obscure words" in the 2020 ToS	Average frequency of occurrence of "obscure words" the 2010 ToS	Delta	Lexical difficulty
Amazon	0,062	0,062	0	=

Table 7. Average frequency of occurrence of "obscure words" per ToS under analysis and delta per company

Preliminary conclusions

These results indicate that, even though the texts under analysis cannot be considered as Plain English examples by default, it is possible to identify an incipient movement towards a plainer language. This trend might be inferred from the choice of vocabulary, indicated by the examination of words that could be replaced by a simpler alternative, according to the Plain English Campaign. In addition, these results are consistent with the slight improvements in readability that were observed through the Dale-Chall test results.

2.3.2. The "right to explanation"

In addition to examining aspects linked to the Plain English Campaign in the corpus, we investigated the potential influence of the "right to explanation" debates, following the fourth -and last- objective of this paper:

- Explore if new topics related to automated decision-making processes have been added to the scope of the Terms of Service scope over time for the sample under analysis.

As previously discussed, the public's concerns regarding data and privacy policies have increased during the last decade. The coming into force of the GDPR in the European Union, and several scandals involving data breaches and an unfair use or personal information, have had an impact in the field of technical and legal texts.

Compliant with the new data regulation applicable in Europe, companies offering their online services were obligated to modify their ToS. Additionally, in the context of the

consumers' growing concerns in terms of data privacy, some companies have tried to be clearer in their presentation of legal terms.

Following Araujo, Helberger, Kruikemeier and De Vreese (2020) users' perceptions regarding the "right to explanation" and algorithm transparency might impact the companies' decision to make such information more accessible.

Several academics (Wachter, Mittelstadt & Russell, 2017; Malgieri, 2019; Goodman & Flaxman, 2017; Edwards & Veale, 2018) have analysed the GDPR from different perspectives, mainly critical, in relation to automated decision-making.

We have examined the ToS in our corpus to identify terms related to automated decision-making. This section aims to explore if companies have included information compliant with the users' "right to explanation".

Based on the literature review on the topic, we identified four keywords that are related to the "right to explanation". Then, we investigated the texts under analysis to find out if this concern was incorporated in the ToS over the last years.

The keywords we browsed are the following:

- Algorithm (the search includes related words: algorithms; algorithmic)
- Accountability (the search includes related word: accountable)
- Transparency (the search includes related word: transparent)
- Automated decision-making (the search includes related words: automated; decision-making)

The table below displays the results for the ToS of 2020 in the corpus.

Company	Occurrence of algorithm*	Occurrence of accountability*	Occurrence of transparency*	Occurrence of automated decision-making*
Google	1	0	1	1
Youtube	0	0	0	0
Facebook	0	1	0	0
Twitter	0	0	0	0
Amazon	0	0	0	0

Table 8. Occurrence of "right to explanation" related keywords in the ToS applicable in 2020 per company. The * indicates that the analysis includes both the keyword and its related words for searching purposes.

There was no mention of any of these keywords within the ToS applicable in 2010. There are four occurrences in the currently applicable ToS, but only for Google and Facebook services.

Looking at the context in which these keywords occur in the texts, only Google has added relevant information for the topic we are examining. The keyword identified within Facebook's ToS does not relate to the automated processes that the service uses (and that might affect users). In this case, the word 'accountable' is found in the following phrase, under section "Your commitments to Facebook and our community" (the bold is ours):

Who can use Facebook

When people stand behind their opinions and actions, our community is safer and more **accountable**. For this reason, you must […]

Accountability, in this context, applies to the community and relies on users. There is no mention to the company accountability regarding their automated processes within their ToS.

As per Google, there is one explicit mention to the automated systems and algorithms that analyse users' data (the bold is ours):

> This license is for the limited purpose of:
>
> operating and improving the services, which means allowing the services to work as designed and creating new features and functionalities. This includes using **automated systems** and **algorithms** to analyze your content
>
> […]

Under the section titled "Handling requests for your data", Google's ToS include a direct link to their Transparency Report, which works as a menu to access several related sites (the bold is ours):

> For more information about the data disclosure requests that Google receives worldwide, and how we respond to such requests, see our **Transparency** Report […]

A thorough analysis of such content is beyond the scope of this paper. However, we can mention that some of Google's automated decision-making processes are addressed in these reports (linked in the ToS), with a varying depth of information.

In the case of Youtube, there is a specific link within the ToS pointing to a site dealing the Data Processing terms. However, the available information does not explain the automated decision-making processes in depth.

Preliminary conclusions

These results show that there is no significant incorporation of the terms related to algorithm transparency in the ToS under analysis. However, it is important to note that the number of links included in the texts has largely enriched the information they contain. For some of the companies in our scope (Google and Youtube, especially), it is possible for users to access information that might touch upon automated decision-making through links available at the ToS.

3. Conclusions

The research question that guided the development of this paper asked if the Terms of Service of five popular Web-based applications had improved their readability over the last decade. After completing our analyses, we anticipate that the answer is not straightforward.

In other words, the affirmative or negative answer to the question might depend on the tools used to address it. However, it is worth noting that it is possible to envision a trend towards a higher readability, at least when it comes to the Terms of Service of the most popular online services (in Spain, as of February 2021).

Below, we include the conclusions following each of the objectives of this study.

Objective 1: Analyse the readability of the Terms of Service of five popular Web-based applications in Spain.

Objective 2: Compare the readability level of the Terms of Service for two periods in time: 2020 and 2010.

- We applied two extensively used readability formulae: the Flesch-Kincaid test and the Dale-Chall test, with inconsistent results. The ToS of the five popular Web-based applications under analysis, in both versions (2010 and 2020), show that they have become slightly more difficult to read and understand, according to the Flesch-Kincaid test results. This conclusion is in line with other similar studies (Peslak, Kovalchik & Conforti, 2020; Graber, D'Alessandro & Johnson-West, 2002 and Jensen & Potts, 2004, Proctor, Ali & Viu, 2008).

- However, the results from the Dale-Chall test show slight improvements in the ToS readability. This conclusion might indicate that, despite using long sentences and many-syllable words, the choice of vocabulary has become easier to understand.

- Complementing the application of the Flesch-Kincaid test and the Dale-Chall test, we conducted a qualitative assessment, examining approach, style, register and information display. The texts that we examined are binding legal agreements and thus examples of legal texts. However, the features of legal English are not that central in our corpus.

- We have identified improvements in the ToS. This might indicate a trend towards a more personal approach and a more informal style, which is more appropriate for the target readers of these texts. In this vein, the "duty to read", in general considered as a one-sided responsibility, could start to push for its necessary counterpart: "the duty to draft readable contracts" (Benoliel & Becher, 2019).

- Texts are usually shorter, although the information they contain is richer, incorporating an increasing number of links. Even though we might conclude that the richer the texts, the higher their difficulty (following Peslak, Kovalchik and Conforti, 2020), readers have more freedom when navigating the information.

- The web of links leading to further information has the potential to facilitate the access to relevant data that a user is interested in finding, thus improving the texts' readability. We might arrive to such conclusion, especially when considering the classic definition of readability (Dale & Chall, 1949) that focuses on the success of readers: understanding the text, reading it at optimal speed and finding it interesting.

Objective 3: Identify the language resources recommended by the Plain English movement to improve the Terms of Service readability.

- Concerning the identification of elements linked to the Plain English movement, it is possible to see an incipient movement towards a plainer language through the vocabulary choice in the ToS under analysis.

- This conclusion is consistent with the results of the Dale-Chall test, which show slight improvements in readability. Furthermore, this analysis is in line with the qualitative assessment applied to exploring approach, style, and register.

Objective 4: Explore if new topics related to automated decision-making processes have been added to the scope of the Terms of Service scope over time for the sample under analysis.

- As regards to the identification of concepts related to the "right to explanation" concerns, we could not detect their incorporation into the ToS in a straightforward manner.

- However, when going through the ToS it is possible to access some information regarding automated decision-making processes when navigating the links (especially in the case of Google and Youtube). The availability of such information does not fully cover the requirements towards algorithm transparency that users are interested in, but it might indicate an incipient trend in the line of the conclusion from Araujo, Helberger, Kruikemeier and De Vreese (2020).

Bibliographic references

Araujo, T., Helberger, N., Kruikemeier, S. *et al.* (2002). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Soc* 35**,** 611–623. https://doi.org/10.1007/s00146-019-00931-w

Baugh, A.C. & Cable, T. (2002). *A History of the English Language*. Routledge.

Becher, S. I. & Benoliel, U. (2020). Law in Books and Law in Action: The Readability of Privacy Policies and the GDPR. *Consumer Law & Economics*, Klaus Mathis & Avishalom Tor eds., Springer, 179-204.

Benoliel, U., & Becher, S. I. (2019). The duty to read the unreadable. *BCL Rev.*, *60*, 2255.

Dale, E. & Chall, J. (1948). A Formula for Predicting Readability. *Educational Research Bulletin* Vol. XVII, 1.

Davison, A., & Kantor, R. N. (1982). On the failure of readability formulas to define readable texts: A case study from adaptations. *Reading research quarterly*, 187-209.

DuBay, W. H. (2004). *The Principles of Readability*. ERIC – Education Resources Information Center https://eric.ed.gov/?id=ED490073

DuBay, W.H. (Ed.) (2006). *The Classic readability studies*. ERIC – Education Resources Information Center https://files.eric.ed.gov/fulltext/ED506404.pdf

DuBay (2007). *Smart Language: Readers, Readability, and the Grading of Text*. ERIC – Education Resources Information Center https://files.eric.ed.gov/fulltext/ED506403.pdf

Edwards, L., & Veale, M. (2018). Enslaving the algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?. *IEEE Security & Privacy*, *16*(3), 46-54.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology* 32:221–33.

Flesch, R. (1949) *How to write in Plain English*. Archive.org https://web.archive.org/web/20160712094308/http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml)

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI magazine, 38*(3), 50-57.

Graber, M.A., D'Alessandro, D.M., & Johnson-West, J. (2002). Reading level of privacy policies on Internet health Web sites. *Journal of Family Practice*, 51, 642-645.

Jones, M. J. (1988). A longitudinal study of the readability of the chairman's narratives in the corporate reports of a UK company. *Accounting and Business Research*, *18*(72), 297-305.

Jensen, C., & Potts, C. (2004). Privacy policies as decision-making tools: An Evaluation of online privacy notices. *CHI 2004*, 6, 471-478.

Kiesling, S. F. (2011). *Linguistic Variation and Change.* Edinburgh University Press.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L. & Chissom. B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel.* CNTECHTRA Research Branch Report 8-75.

Klare, G. (1974). Assessing Readability. *Reading Research Quarterly*, 10(1), 62-102.

Lorge, I. (1939). Predicting reading difficulty of selections for children. *Elementary English Review* 16:229–233.

Malgieri, G. (2019). Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations. *Computer law & security review*, 35(5), 105327.

Obar, J. A. & Oeldorf-Hirsch, A., (2018) The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. *Information, Communication & Society*, 1-20.

Paasche-Orlow, M. K., Taylor, H. A., & Brancati, F. L. (2003). Readability standards for informed-consent forms as compared with actual readability. *New England journal of medicine*, *348*(8), 721-726.

Pałka, P. & Lippi, M. (2019). Big Data Analytics, Online Terms of Service and Privacy Policies. *Research Handbook on Big Data Law* (Forthcoming), https://ssrn.com/abstract=3347364

Peslak, A., Kovalchick, L., & Conforti, M. (2020). A Longitudinal Study of Google Privacy Policies. *JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH*. 13(2), 54-65.

Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *Elife*, *6*, e27725.

Proctor, R.W., Ali, M.A., & Vu, K.L. (2008). Examining usability of web privacy policies. *International Journal of Human-Computer Interaction*, 24(3), 307-328.

Stahl, S. A. (2003). Vocabulary and Readability: How Knowing Word Meanings Affects Comprehension. *Topics in Language Disorders*. 23(3), 241-247.

Steinfeld, N. (2016). "I agree to the terms and conditions": (How) do users read privacy policies online? An eye-tracking experiment. *Computers in human behavior*, *55*, 992-1000.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL & Tech.*, *31*, 841.

Zhou, S., Jeong H. & Green, P. A. (2017) How Consistent Are the Best-Known Readability Equations in Estimating the Readability of Design Standards?. *IEEE Transactions on Professional Communication*, 60(1), 97-111.

Annex 1

List of sources from which the Terms of Service under analysis were obtained

Google Terms of Service (2020). Retrieved 31 March 2021 from
https://policies.google.com/terms?hl=en-US

Google Terms of Service (2010). Retrieved 31 March 2021 from
https://www.google.com/intl/en_ES/policies/terms/archive/20070416/

Youtube Terms of Service (2020). Retrieved 31 March 2021 from
https://www.youtube.com/t/terms

Youtube Terms of Service (2010). Retrieved 31 March 2021 from
https://web.archive.org/web/20101026141237/https://www.youtube.com/t/terms

Facebook Terms of Service (2020). Retrieved 31 March 2021 from

https://www.facebook.com/legal/terms

Facebook Terms of Service (2010). Retrieved 31 March 2021 from

https://web.archive.org/web/20101209203534/http://www.facebook.com/terms.php

Twitter Terms of Service (2020). Retrieved 31 March 2021 from

https://twitter.com/en/tos#intlTerms

Twitter Terms of Service (2010). Retrieved 31 March 2021 from
https://twitter.com/en/tos/previous/version_4

Amazon Terms of Service (2020). Retrieved 31 March 2021 from
https://www.amazon.co.uk/gp/help/customer/display.html?ie=UTF8&nodeId=201909000&ref_=footer_cou

Amazon Terms of Service (2011). Retrieved 31 March 2021 from

https://m.media-amazon.com/images/G/02/legal/Previous_versions_Legal_policies/Conditions_of_Use_and_Sale/Amazon.co.uk_Conditions_of_Use_and_Sale_-_March_3_2011.pdf