statuS Quaestionis
language text culture

Laura Di Ferrante
Università degli studi di Milano

How passive voice gets affected by popularization. A quantitative analysis of scientific research articles and university press releases

Abstract

The passive voice has been explored for its communicative potential offered by the possibility to omit the agent, which makes it suitable for both ideological and academic discourse. Research has shown that the passive voice is a characterizing element of academic discourse, but there is a gap in the research concerning the incidence of the passive voice in popularized science discourse. In this paper, the presence and form of the passive voice are analyzed and compared in two bijective English corpora: one comprising published research papers (n = 30) and the other their respective popularized versions published as press releases on university websites (n = 30). The results of the quantitative analysis show statistically significant differences with regard to the amount of passive voice constructions between the two corpora; the difference in the use of the be-passives is also significant; the differences of get-passives, conjoined passives, and bare passives are not statistically significant between the two corpora. Additionally, some novel tendencies seem to arise: the scientific texts present instances of get-passives, which might point toward a contamination between academic language and the objects of analysis, such as text corpora, interview transcripts, or other verbal materials analyzed in the articles.

Keywords: passive voice, academic language, press releases, English for Academic Purposes (EAP), comparable corpora

1. *Introduction*

In the outcome of proceedings on the "high-quality, transparent, open, trustworthy and equitable scholarly publishing" published by the Council of the European Union, it is emphasized "the importance of supporting the development of such models ["that do not charge fees to authors or readers and

where authors can publish their work without funding/institutional eligibility criteria"] led by public research organisations" (Council 2023, 4;5). Consistently with this, the Council recommends that the member states support and reward open science culture (Council 2021). It turns out that free accessibility of science is a goal to be pursued and supported by governments and institutions and one of the reasons for this to happen is to narrow the divide between scientific research and society.

Although some observers believe that open access (OA) is the future of research, many still oppose it, resulting in a significant amount of research being locked behind paywalls. While achieving global, comprehensive, and enduring access to scientific knowledge on a worldwide scale remains an ongoing endeavor that still needs time, one way through which research findings are being disseminated to wider society is science communication. The popularization of scientific discourse (Fahnestock 1998; Garzone 2014; Garzone 2020; Luzón 2013) consists in disseminating science by adapting a format and language appealing and accessible to audiences of non-experts, thus contributing to the promotion of scientific literacy and to the engagement of the public with scientific issues. Universities worldwide, as privileged research environments, contribute to science communication by posting on their websites press releases reporting on findings of published scientific research (Autzen 2014; Di Ferrante et al. 2021; Lindh 2020; Zhang 2018). Research has demonstrated that the language that is used to present scientific findings to the general public has distinct characteristics that differentiate it from academic discourse.

Most of the time, press releases are derived directly from newly published scientific articles. While they are adapted in terms of length, format, and language to suit the general public, they still retain linguistic traits of English for Academic Purposes (EAP) (Petrocelli et al. 2022). In this context, it is important to understand if and to what extent specific features of academic English are manipulated and adapted to science communication genres. This understanding helps shed light on the linguistic mechanisms used to foster science-society relationships.

Several functions have been attributed to the use of passive voice in academic discourse, including the preservation of objectivity, the concealment of the author, the focus on process, and the emphasis on formality. These aspects have been shown to be particularly appropriate for the presentation of scientific information in English. Consistently with this, passive voice is listed as a

marked feature (positive loadings) in the fifth dimension (Abstract - Non-abstract) of the multi-dimensional linguistic analysis of genre variation (Biber 1992). However, it has been demonstrated that the use of the passive voice is decreasing and many writing manuals are suggesting to avoid its use in writing. Moreover, grammar checker tools like Grammarly often advise to replace passive voice with active voice – sometimes inaccurately (O'Neill and Russell 2019; Zinkevich and Ledeneva 2021). Some associate such decline with a tendency to use language that is more informal and colloquial.

In this paper, a corpus-assisted analysis is conducted with the objective of observing and measuring the variation in the use of the passive voice. Specifically, such variation is observed when transforming the content of scientific papers to make it more accessible to a wider audience through press releases. To achieve this goal, the following research questions are addressed:

RQ1. How does the incidence of passive voice in university press releases compare to that in the source scientific articles?
RQ2. What is the distribution of different types of passives in press releases and scientific articles? Is there any significant difference in distribution between the two corpora?
RQ3. Is the get-passive, deemed as the most informal passive, present in one or both corpora? And if so, how is it used?

In the following sections, quantitative studies focusing on the use of passive voice in scientific discourse are reviewed to identify methods and criteria employed to explore this specific research subject and to gather information on data that could facilitate comparisons and/or contrastive reflections. Subsequently, a quantitative analysis is carried out. This analysis aims to contribute quantitative data and understanding of the popularization processes involved in scientific discourse.

2. *Literature review*

The analysis of passive voice has a long and solid tradition. Researchers have analyzed passive voice both diachronically and synchronically with a particular interest in scientific discourse. These studies were motivated by interest in the passive voice as a characterizing feature of registers, particularly in written discourse. Biber's fifth dimension – abstract vs. non-abstract style (Biber

1988; Biber 1992) – serves as a prototypical example: genres that exhibit a high frequency of be-passives are typically characterized by abstract and technical content, as well as a formal style.

Research on the use of passive voice in written texts also provides insights into the role of academics in the relationship with each other and their readers. Several diachronic investigations, however, found that the use of passive voice is generally declining (Biber 1999), including in scientific texts and particularly in American English compared to British English (Seoane and Loureiro-Porto 2005). Hundt, Schneider, and Seoane (2016) summarize the motivations adopted to justify this tendency in two general trends: one that concerns the "democratisation of discourse in Western society" (33) and the other is related to the competitiveness in academia. The democratization trend would result in keeping the distance from the detached and formal style associated with a conception of academia as confined within the ivory tower. Competitiveness, instead, worked as a force toward more efficient communication. It turns out that "the tendency to make academic writing more accessible and engaging goes together with the trend towards informalisation" (Ibid.: 33). As Hyland and Jang (2017) suggest, authors are no longer constrained by a stylistic system that imposes strict norms on objectivity. This increased freedom allows authors to generate texts that are less formal and to adopt a more inclusive approach in establishing relationships with their readership.

Over the years, research on passive voice has been motivated by the need to contribute research for the understanding of several aspects: from impersonality (Biber et al. 1999; Fairclough 1995) to written and spoken discourse (Xiao, McEnery, and Qian 2006), from authorial presence in texts (Seoane and Hundt 2018) to development of disciplinary language across time and disciplines (Leong 2020), as well as emphasizing stance (Baratta 2009). Research has also shown differences in the use of passive voice across English varieties and across disciplines (Hundt, Schneider, and Seoane 2016; Mair and Leech 2006; Seoane and Hundt 2018).

Some information exists about the use of passive voice across different registers, thanks to the significant amount of studies carried out by Biber and colleagues with multidimensional analysis (Biber 1988; Biber 1999; Conrad and Biber 2001). Academic prose and official documents are heavily characterized by the presence of passive voice, followed closely by press reportage, which provides valuable insight into the use of passive voice in press releases.

To the best of my knowledge, quantitative research on passive voice has not looked at press releases yet. With reference to the passive voice, multidimensional analysis has been applied to some genres like press reviews and press reportages that are similar to press releases insofar as they are related to journalistic genres. However, the language of press releases cannot be identified with that of other journalistic genres as press releases are shown to be an inherently hybrid genre (Catenaccio 2007; Catenaccio 2008; McLaren and Gurău 2005) in that they convey content which is both informational and promotional (Di Ferrante et al. 2021; Lindh 2020; Petrocelli et al. 2022).

The presence of passive voice in corporate press releases has been recorded in a number of qualitative studies, like in McLaren and Gurău (2005), who noticed that passive voice in press releases tends to be present in the first part of the text, which is concerned with the Announcement and its Elaboration, a part of the text that appears "more informative than persuasive" (Ibid.: 19); Jacobs (1999) also noticed the strategic use of passives to convey authoritativeness in some excerpts of press releases. Hyland (2010) compared research papers and popular science articles, specifically focusing on proximity, which he defined as the manner in which writers utilize rhetorical features to convey both their authoritative expertise and their attitude toward the subjects being discussed.

2.1 *Passive voice in written discourse*

Research on the use of passive voice in academic writing is rather scattered and hardly allows for cross comparisons. Differences among studies concern a wide range of aspects regarding both corpus selections and methodological choices and procedures. In order to give a sense of the variability across these studies, it is interesting to notice that some studies focused on specific journals and counted the passive voices used in their published articles (Amdur, Kirwan, and Morris 2010; Leong 2014; Lu 2013). In addition, several scholars analyzed articles from a specific discipline, mainly scientific ones with a few exceptions – for example, Leong (2021) focused on History articles and Lu (2013) on Theoretical and Applied Linguistics – but the bulk of the studies took into account articles from scientific disciplines: Banks focused on Oceanography in 1994 and on Physical and biological sciences in 2017, Tarone et. al (1981; 1998) on Astrophysics, Seoane (2006) on Natural science and Medicine, and Riley

(1991) on Speech-language pathology. Other studies focused on papers simultaneously from the science and the humanities fields, for example, on Biochemistry and Literature (Iddings 2007) or on Biology and History (Rachul 2008). Biber (2006) focused on multiple disciplines but looked mainly at university textbooks. Moreover, in different studies, the passive voice is categorized with different criteria and methodological choices. For instance, Bakalar et al. (2014) and Millar et al. (2013) did not consider bare passives when counting passives; Banks (2017) eliminated single-author papers or other instances; Seoane and Loureiro-Porto (2005) did not take into account -ing passives.

Additionally, in some studies, the percentage of passives was calculated over the total number of sentences (Lu 2013), and some of these provided the mean value (Leong 2014) while some the median (Amdur, Kirwan and Morris 2010); other investigations counted the number of passive forms over all verb phrases (Biber 2006); some instead counted the passives over the total number of finite verbs (Banks 1994; Banks 2017; Biber et al. 1999); other scholars counted the passives over the total number of clauses (Leong 2020), or over the total number of non-modal verb forms (Barber 1962), or over the total number of words in the selected corpus (Chen and Ye 2014). This brief picture of the different approaches to the count of passive voice constructions should suffice to present the heterogeneity of the landscape, where passives are examined through very diverse timelines and disciplinary fields, diverse operational definitions, and diverse computational bases. It turns out that on the one hand the research on passives is very rich and provides information on their use qualitatively, quantitatively, and diachronically, but on the other hand the different approaches and methodologies make it difficult to attempt comparisons among the results obtained in the various studies.

Furthermore, another aspect that makes the study of passives quite hard to carry out is that counting passive voices is not as straightforward as with other linguistic categories for several reasons. First, passive voice takes different forms because of the variety of tenses and modals it can be formed by, because of the presence of the auxiliary *be*, which can also be omitted (bare passive) or substituted by *get* (see section 3 for a classification of the passive forms); this means that passives take multiple forms, which are expressed with one or more words (e.g., the different length of the two passive forms in the following sentence: "the lesion could have been caused by the trauma described"). Second, in the case of bare passives, the past participles are sometimes similar to adjectives in terms of

form. Both the first and the second reason contribute to make it difficult for most corpus software to detect all the different variants. This resulted in the choice, operated by many scholars, to manually count the passives in the articles, which, to date, seems to be the only way to detect all forms of passives. Manually counting determines that in some studies, the number of analyzed articles is low enough – sometimes less than five articles were analyzed – to not allow for generalizations.

In the context of investigations focusing on the passive voice, some studies combined the methodology of corpus linguistics with contrastive analysis, comparing the presence of passives in two different corpora, as it is the case of the present study.

Xiao, McEnery, and Qian (2006) conducted a comparative analysis of data between two different languages, British English and Mandarin Chinese. They utilized four corpora, two for each language, consisting respectively of written and spoken language samples. They compared the occurrences of get- and be-passives in the English corpora looking at their use in positive, negative, or neutral contexts and also looked at other features of the passive voice, such as variation across genre, long vs. short passives (with or without the by-phrase), and adverbials in be- and get-passives.

However, most contrastive analysis of passive voice focused on written texts. Chen and Ye (2014) analyzed the presence of the passive voice in two corpora in the field of translation studies: the first corpus consisted of 160 original English abstracts, the other consisted of 160 English abstracts translated from Chinese. In order to identify passive voices, they used the tagging tool CLAWS (the Constituent Likelihood Automatic Word-tagging System) and manually tagged the instances of bare passives. They found that the passive voice accounted for 1,28% of the total number of words in the first corpus and 1,13% in the second. They also found that be-passives are 1,22% and 1,08% respectively of the total number of words of each corpus. They also analyzed the occurrence and frequency of tenses and types of pronouns.

Hiltunen (2016) also compared two corpora of academic writing: one of 256 research articles and the other of 388 university student essays. The corpora were tagged with CLAWS. The analysis revealed cross-disciplinary differences in the presence of passives in research articles, characterizing particularly the distinction between hard and soft sciences.

Seoane and Loureiro-Porto (2005) analyzed texts from scientific articles written in American English and British English. The two corpora consist of

110,000 words each distributed across three time periods: 1905/1925, 1960/1975, and 1985/1990. The scholars demonstrated that the use of passive has been actually declining over time and more so in American English than in British English, however they show that the reason of such decline cannot be ascribed to the integration of traits of spoken language into written language; instead, the authors propose that the underlying reasons for this decline might be rooted in events and dynamics that are sociolinguistic in nature.

2.2 *Use of be-passive and get-passive*

Quantitative analysis has enabled the monitoring of language change over time which has revealed how written language has been significantly impacted by the increase of colloquialisms and traits of spoken language (Mair 1998). This influence can be attributed, in large part, to the rise of digital platforms which fostered the development of informal writing. As a result, these new forms have also had a partial impact on academic style.

Consistently with this tendency, between the end of the 20th century and the first two decades of the 21st century, research has recorded an increase of the use of the get-passive construction (Biber and Gray 2016; Givón and Yang 1994; Mair and Leech 2006). This is a construction where get is used as an auxiliary verb accompanied by the past participle of the main verb in the passive voice: typically, it is less common than the be-passive and has traditionally been documented as a trait of spoken language (Biber et al. 1999). Given its different nature from the be-passive, it does not substitute it and in this regard, it has been found that it is mostly used in colloquial and informal texts (Xiao, McEnery, and Qian 2006). Additionally, Leech et al. (2009) note that: "the get-passive and the mediopassive are less typical as passive constructions because they are both grammatically and semantically different from the be-passive." (Leech et al. 2009, 145; on the difference between get-passive and be-passive, see also Chappell 1980; Collins 1996; Matthews 1993). The reasons for the increase of the use of the get passive are attributed to the process of grammaticalization of this construction, facilitated by the evolution of get into an auxiliary verb in passive constructions and the increasing presence of spoken traits in written English. Additionally, it should be noted that a number of studies demonstrated that the use of the get-passive is not a perfect alternative to the be-passive,

hence the choice between them is not arbitrary. In particular, it has been found that the get-passive constructions are usually related to specific meanings, implying, for example, a transition or change of state, but also the speaker/writer's intention to convey an adversative context or unpleasant outcome for the subject (Chappel 1980; Carter and McCarthy 1999; Xiao, McEnery, and Qian 2006), namely "a state of affairs that is signalled contextually by the conversational participants as unfortunate, undesirable, or at least problematic" (Carter and McCarthy 1999, 49). Xiao, McEnery, and Qian (2006) provided separate frequencies for get-passives and be-passives and, consistently with the literature, found that get-passives are mostly present in the spoken corpus (BNCdemo) and more frequently (46.5% of the cases) associated with negative meanings in the written corpus (FLOB). It is important to point out that most studies on the get-passive are based on spoken corpora (or written-to-be-spoken – see, for example, Schwartz 2015) and not all of them agree in connecting the get-passive to mostly adversative contexts. For example, Xiao, McEnery, and Qian (2006) found that most of the get-passives were neutral, however, it should be specified that data from multiple genres of spoken and written English were aggregated. Coto Villalibre (2015), who only focused on spoken data, found that the majority of get-passives in his study had "semantically neutral and non-adversative implications for the subject" (24).

3. *Data and Method*

The study presented here is based on two bijective corpora of articles, where each paper in one corpus corresponds to a paper in the other corpus. The first corpus, named University Research Press Releases (UNREP), consists of university press releases reporting on findings from scientific articles; all the press releases were collected from the university websites where they were first issued. The second corpus, Scientific Articles (SCAR), includes the original scientific papers on which the articles in the UNREP corpus are based. The two corpora are very different in number of words as the popular-science articles are inherently shorter than their respective scientific articles.

While this study is part of a much larger project on the discursive practices emerging from the comparison of scientific texts and their popularization, for the purposes of the present work, 30 pairs of articles and press releases were

analyzed. The study consists in a blended approach that combines Corpus Linguistics and Discourse Analysis (i.e., the so-called CADS, Corpus Assisted Discourse Studies) (Partington, Duguid, and Taylor 2013; Gillings, Mautner, and Baker 2023). Both the academic articles and the press releases were manually tagged. The tagging was compared with the results obtained by corpus analysis tools and, when discrepancies would arise, inconsistencies were checked again and addressed.

The process of identification and tagging of the passive voices was time-consuming, and this impacted considerably on the final number of articles analyzed. Manual tagging was selected as the most accurate process to make sure that all the passive voices in the articles were identified and marked, because corpus analysis tools are not always able to identify all the different types of passives, particularly in specific cases like bare passives. In particular, piloting tests were run with some tools like LancsBox (Brezina, Weill-Tessier, and McEnery 2020) and UAM Corpus Tool, and it was found that the bare passives and some other passive constructions like modal + be + past participle or the conjoined passives (see below) were not detected. The tagging included the following four different labels for four types of passive voice.

1) *Be passives*, which are passive voices formed with the auxiliary verb "to be", and following Leong (2014), include the following subtypes:

   a. Basic (be + Ven)

   The study also found that a positive chatbot experience <BE PASSIVE> *was associated* with customer loyalty.[1]

   b. Progressive (be + being + Ven)

   Lee and her colleagues set out to understand how they <BE PASSIVE> *were being deployed* throughout the social media universe.

   c. Perfective (have + been + Ven)

   The report says that since 1970, large areas of land <BE PASSIVE> *have been urbanized* [...].

   d. Modal (modal + be + Ven)

   It <BE PASSIVE> *can be used* to clean car headlights.

1 The examples are taken from the two corpora in this study.

e. Modal perfective (modal + have + been + Ven)

Future studies could also investigate how materialistic values <BE PASSIVE> *may have been affected* [...].

f. To-infinitive (to + be + Ven)

Temperature anomalies <BE PASSIVE> *have also been found* <BE PASSIVE> *to be associated* with moderate stunting in Ethiopia.

g. Non-finite –ing (being + Ven)

Christopher Bail et al. [8] have also shown how <BE PASSIVE> *being exposed* to more opposing views can actually increase political polarization.

2) *Bare passives or passives with no auxiliary verb* (Ven)[2]

The other attributes <BARE PASSIVE> *associated* with customer satisfaction were [...].

[...] through the same procedure <BARE PASSIVE> as *used* in experiment 3.

3) *Conjoined passives*, which are "passives where the repeated 'be' auxiliary in a conjoined clause is omitted to avoid repetition" (Leong 2014, 3). In the present work, the auxiliary could also be "get".

When consumers <BE PASSIVE> *are* either *primed* with a general sense of nontraditional product functionality (experiment 5) or explicitly <CONJOINED PASSIVE> *fixated* on the traditional functionality of a product (experiment 6).

4) *Get passives*, which are passive voices formed with *get* used as an auxiliary verb

Uninsured children are still less likely to <GET PASSIVE> *get vaccinated*.

Since the present study is heavily based on counting frequencies of passive voices, it is important to clarify some methodological choices. The first methodological remark concerns some particular cases which although might be possibly interpreted as instances of passives, either implicitly or in terms of early grammatical genesis, were not tagged as passives in this study:

- Past participles in a left-branching position (Garzone 2020, 100), e.g., *based on*, *given*, *compared*, *granted*, *provided*, *taken*, *set*:

2 Most of the instances under this umbrella label are non-finite relative clauses, or "reduced" relative clauses (Biber, 2006, p. 74).

> Taken together, the regression analysis finds a positive and meaningful treatment effect for the marketing intervention. (SCAR)
>
> Compared with less descriptive logos, more descriptive logos elicit stronger impressions of authenticity. (SCAR)
>
> Given that the consumption of authentic brands provides important identity benefits to consumers, consumers typically appreciate and value authenticity in brands. (SCAR)
>
> Set against the rapid pace of technological innovation, this simple question has grown into a pressing concern for scientists, care givers, and policymakers. (SCAR)

- Idioms like *be bound to*, *be supposed to*, *get rid of*, *as opposed to*.
- Two instances of the so-called causative passive or have-pseudo-passives (but with get instead of have) were found in the corpora and they were not counted as passives. One case was "Newswise — Investing in product safety, employee diversity and carbon footprint reduction are all examples of corporate social responsibility (CSR) that can result in high praise for a chief executive — or get them fired— according to new research from the University of Notre Dame." (UNREP)

The second methodological remark concerns the denominator. As it was mentioned in the previous section, multiple studies have been conducted on passive voice, however the baseline, or denominator, in relation to which passives have been counted has not been consistent across different studies which counted passives over units as vary as total number of clauses, of verbs, of finite verbs, of verb phrases etc. Moreover, while corpus linguistics commonly uses per million words (pmw) as a baseline for normalization (see, for example, Brezina 2018, 43), other studies are based on different choices, such as the percentage over the total number of words (Chen and Ye 2014), over 100K words (Sulaiman 2022), over 1,000 words and the ratio of passive voices to active transitive clauses (Seoane and Loureiro-Porto 2005) or to active voices (Seoane 2009). These choices were probably related to the sizes of corpora, their comparability (or difficulty thereof); other issues may be related to passive voice being often manually tagged rather than automatically tagged through corpus linguistics software, which is also one of the issues mentioned in some critiques to using pmw as a baseline for some linguistic features; other concerns are related to the opportunity of selection of a given feature (for an extensive discussion on this, see Wallis and Mehl 2022). In this paper, relative frequencies of the passives

are presented together with percentages calculated over the total number of words with normalization. The total number of words in a corpus allows for easily replicating the counting and therefore more easily comparing data across different studies. As a matter of fact, other measures used in other studies, like the number of active voices or the number of clauses or sentences rely on operational definitions that may differ across software and research designs.

The criterion for the choice of the articles has been that of looking for popular articles first. The search has been performed by making sure that each popularized article was overtly based on a scientific paper. The topics range from Biology to Psychology, from Economy and Marketing to Information Technology and Education.

4. *Results*

As mentioned before, 30 scientific papers and 30 press releases were manually tagged for four types of passives. Table 1 illustrates differences between the scientific articles in the SCAR corpus and the press releases in the UNREP corpus in terms of total number of words, average article length, total number of passive voices and average number of passive voices in each article.

Corpora	Number of articles	Total number of words	Average article length (number of words)	Total number of passive voices	Average number of passives per article	Average passives*
SCAR corpus	30	216,874	7,229	3,512	117	1.66%
UNREP corpus	30	24,084	802	308	10,2	1.28%

Table 1. Number of words and passive voices in the two corpora

Note: References were not included in the word count.
Percentages were calculated on the total number of words.
*The average percentage of passives was determined by computing the average of the percentages of passive constructions present in each article.

The count shows that press releases have approximately nine times less words than scientific articles (24,084 vs. 216,874). The average length of the scientific articles is 7,229 words, while on average, the number of words in press releases

is 802, approximately 10% of the number of words in scientific articles. In order to answer the first research question and thus compare the incidence of passive voice in the two corpora, the passive constructions were counted in each article. Throughout the 30 scientific articles, a total of 3,512 instances of the passive voice were detected, while 308 were found in the press releases: 11 times fewer passive voices than in the scientific articles, with an average percentage of the passive occurrence of 1.66% in the scientific articles of the SCAR corpus and 1.28% in the press releases of the UNREP corpus. The observed differences are compatible with the two distinct genres of scientific articles and press releases and with the substantial variation in article length. Interestingly, 1.28% is also the exact percentage of passive voice reported in Chen and Ye's (2014) study on scientific abstracts.

A Wilcoxon Paired Test was performed to determine whether the difference in the number of passives between the two corpora is statistically significant, and therefore whether a relationship exists between the use of passives and the textual genre of scientific articles and press releases. The Wilcoxon Paired Test was selected after a Shapiro-Wilk Test was used to test the null hypothesis of normal distribution of the data at a significance level of 0.05. While it did not reject the null hypothesis in the UNREP corpus ($W = 0.9636$, p-value = 0.3816), the distribution of data in the SCAR corpus was significantly different from a normal distribution: ($W = 0.92497$, p-value = 0.03616). It turns out that a t-test could not be used.

The Wilcoxon Paired Test was computed with the data expressed as percentages of passives in each article. The test produced a significant result with p-value < 0.05 (p-value = 0.001341), which shows a statistically significant difference in the number of occurrences of passive voices in the scientific articles of the SCAR corpus and in the press releases of the UNREP corpus. It turns out that our data confirm that passive voice is a discriminating feature of the two types of texts, based on these two corpora. This is particularly interesting because despite the fact that the press releases analyzed in this study are based on their corresponding scientific articles, they don't seem to be completely adherent to the scientific articles and their language use, at least in terms of passive voice incidence.

The second research question pertained to the distribution/occurrence of the different types of passives in the two corpora and whether any potential difference might be determined by the genre specificity of press releases and scientific articles.

Table 2 shows the distribution of the different types of passive voice in each corpus. Each type of passive construction (be-passive, conjoined passive, get-passive, bare passive) is represented in terms of its frequency and percentage (ratio) within both corpora. The percentages in the table are calculated on the total number of passives.

	bare passive		be passive		conjoined passive		get passive		total number of words	
	frequency	ratio	frequency	ratio	frequency	ratio	frequency	ratio	frequency	ratio
SCAR corpus	1124	32.0	2296	65.37	88	2.50	4	1.59	3512	100
UNREP corpus	101	32.68	198	64.07	8	2.58	2	0.64	309	100

Table 2. Distribution of the types of passives in each corpus on the total of passives

Looking at the distribution of each type of passive, it is apparent the percentages on the total number of passives are very similar across the two corpora. Consistently with the literature (see, for example, Leong 2014), the be-passive is the dominant type in both corpora constituting over 64% of all passives in both corpora. This proportion is partially comparable to Chen and Ye's (2014) findings, where the be-passives accounted for 95.71% of all passives, which is over 30% more compared to our corpus; however, Chen and Ye's (2014) study focused solely on paper abstracts, where the most typical form of passive might be more prevalent compared to the main body of the paper. This is merely a hypothesis, and further investigation is required to confirm its validity. Additionally, the bare passive is also very similar counting around 32% in both corpora. This particular finding is partly comparable with the ones in Hiltunen's (2016) study which found that bare passives ranged between 24.7% and 32.8% on the total number of passives in research articles, depending on the discipline. The bare passives in the SCAR corpus are 32% of the total number of passives. This percentage is within the range, and hence compatible with Hiltunen's findings.

While the data presented in Table 1 are calculated on the total number of passives to show how the four types of passives are apportioned, in Table 3 and Figure 1 below, the distribution of the passives is instead computed on the total number of words in each corpus.

	bare passive		be passive		conjoined passive		get passive		total number of words	
	frequency	ratio	frequency	ratio	frequency	ratio	frequency	ratio	frequency	ratio
SCAR corpus	1124	0,518	2296	1,059	88	0,041	4	0,002	216.874	100
UNREP corpus	101	0,419	198	0,822	8	0,033	2	0,008	24.084	100

Table 3 Distribution of the types of passives in each corpus on the total number of words

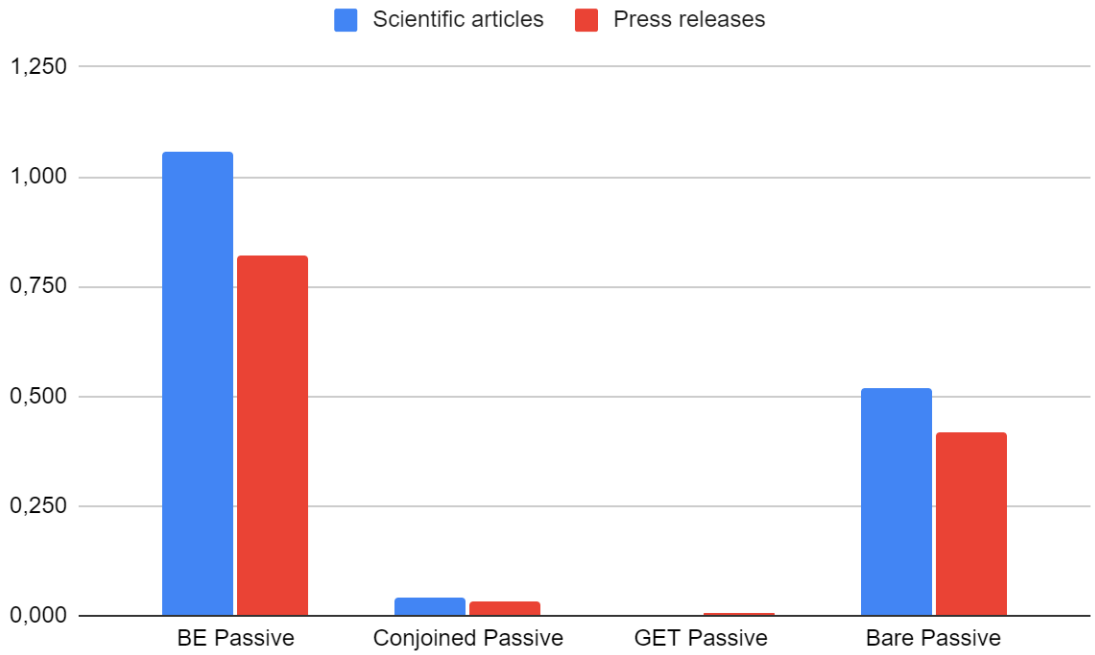

Figure 1 Distribution of types of passives in the two corpora

As shown in Table 3 and Figure 1, the distribution of the types of passives in the two corpora is similar in terms of proportions: in both the corpus of scientific articles and the one of press releases, the be-passive is the most used type of passive, followed by the bare passive. The conjoined passive has relatively few occurrences in the two corpora and the get-passive is barely used and is the only type of passive voice that occurs comparatively more frequently in press releases than in scientific articles.

A Wilcoxon Paired Test was used to assess the differences between the two corpora in the distribution of each type of passive. Since multiple statistical tests were performed simultaneously, to account for the type-I error inflation, the Bonferroni correction (Dunn 1961) was applied to the 5 tests. The procedure consists of rescaling the targeted significance level of 0.05 to account for the multiple tests, which is 0.05/5 = 0.01.

Table 4 shows the mean and median number of each type of passive construction in each of the two corpora; it also presents the results obtained for the p-values and their significance. A visual summary of the variability of passive voice in the two corpora is reported in the box plot in Figure 2.

Passives	Mean (SCAR)	Mean (UNREP)	Median (SCAR)	Median (UNREP)	p-value	significance
Bare Passive	0.536	0.433	0.490	0.425	0.038418418	ns
Be Passive	1.083	0.816	1.010	0.770	0.003222989	**
Conjoined Passive	0.042	0.027	0.023	0.000	0.111650797	ns
Get Passive	0.001	0.011	0.000	0.000	0.583882421	ns
Total Passives	1.663	1.287	1.671	1.311	0.001340603	**

Table 4 Mean and median number of passive constructions and their statistical significance

Note: statistical significance is indicated through an asterisk (*); ns: P ≥ 0.01; ** P ≤ 0.01.

Because for the Wilcoxon Paired Test the Bonferroni correction was used on 5 simultaneous comparisons (total passives, be passives, bare passives, conjoined passives and get passives), only those comparisons with p <0.01 are considered statistically significant.

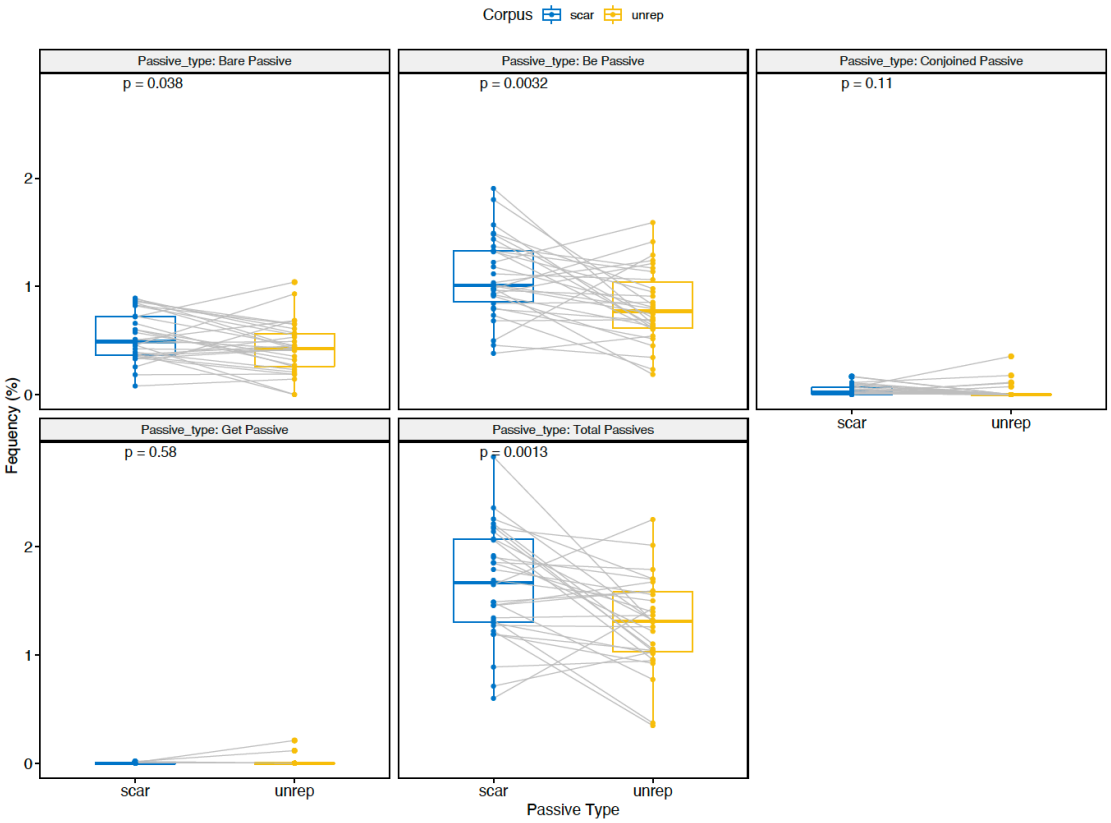

Figure 2. Box plots of the distribution of the different passives in the two corpora and statistical p-value based on a Wilcoxon Paired Test

Table 4 and Figure 2 show the results obtained for each comparison of passive types. In particular, the box plots in Figure 2 present a visualization of the distribution of the different passives in the SCAR and UNREP corpora, along with the Wilcoxon Paired Test results. The horizontal lines within the blue and the yellow rectangles are the median values of each type of passive in the two corpora. The lines coming out of each box span the range between the highest and lowest values within each set. Each blue whisker is paired with a yellow whisker through grey lines: these lines allow us to see how the use of passive constructions changed from the research article to its respective press release. The type of passive construction visualized is indicated in the row at the top of each of the five plots. At the top center of Figure 2, the box plot shows the comparison on the be-passive, which turned out to be statistically different (p-value = 0.00322). The tests on the other types of passives did not show sta-

tistical significance with p-values always greater than 0.01 (bare passive: p-value = 0.038; conjoined passive: p-value = 0.11; get-passive: p-value = 0.58). These results are interesting for several reasons: first, they suggest that the be-passive is the only passive form whose use varies between scientific articles and university press releases. Hence, it seems reasonable to state that it is the be-passive that tows the statistical difference in the general use of passives in the scientific articles compared to the press releases. Second, these results also show that the bare passive, which is the second most used type of passive, does not seem to heavily characterize either text type and this suggest that the bare passive might be perceived as less formal or less academic. As mentioned before, many studies on passive voice disregarded bare passives and many taggers fail to detect them; for these reasons, the different incidence of the two types of passives in the texts was probably overlooked.

It should be pointed out though, that although these data are based on two comparable, bijective corpora, the sample of analyzed articles is still limited, so it seems appropriate to suggest that such significance might account for a probable tendency which should be tested in future studies, with larger samples.

As far as the get-passive is concerned, its presence is very low, which makes it impossible to make generalizations or even observe tendencies, yet the presence itself of a total of six get-passives in the two corpora deserves to be noticed. In the literature, the get-passive is usually regarded as a trait of spoken discourse whose use is being recorded as increasing, but still in the realm of spoken and informal discourse. Moreover, it is regarded as not typical of written discourse, and most studies on passive voice in academic discourse reported zero or very low occurrences of get-passive. Its presence in the present corpora of scientific discourse, and particularly in the one of academic articles, is therefore unexpected and it prompted the analysis to delve into the individual cases. As a matter of fact, academic articles often report chunks of discourse from other contexts for research-related needs: for example, interviews, spontaneous conversations, etc. Moreover, the get-passive used in scientific articles might have been repeated or reported in the matching press release. For all these reasons, it seemed relevant to observe the individual cases qualitatively other than quantitatively.

Interestingly, the analysis of the individual cases revealed that in two cases, the get-passives are found in both the scientific article and the press release belonging to the same pair, hence four of the matching instances belong to two matching pairs. However, surprisingly, it is not the same passive being used:

the passives in each article are used with different verbs and in different contexts. Still, it could be assumed that the source text, namely the scientific article, may influence textual choices of its derivative text, namely, press releases (Biber 1992; Biber et al. 1999; on the influence of textual choice, see also below, the comments on Example 3). Examples 1 and 2 below report two instances of get-passives in a scientific article and its corresponding press release in the field of Mathematical Modeling:

Example 1. Scientific article:
A recovered node cannot infect or <GET PASSIVE> *get infected*. The ratio R0 = rI /rR is the basic reproduction number and is a single adjustable parameter for the infectiousness of an SIR process. (SCAR)

Example 2. Press release:
In the last decade or so, much handwringing has transpired over "fake news" – stories that are not factually accurate that seem <GET PASSIVE> *to get wedged* in the popular psyche and stick. Many now accept the conventional wisdom that fake news spreads "farther, faster, deeper and more broadly than the truth." (UNREP)

This presence of get-passives, even in such low frequency, is important with regard to the reasons why it is used. We could assume that the source texts, namely the scientific articles, may influence textual choices of its derivative texts – namely, the press releases (see Biber 1992; Biber et al. 1999).

As far as the particular get-passives presented in Example 1 and 2 are concerned, consistently with the literature (see Carter and McCarthy 1999), it seems that they both portray an adversative context. As a matter of fact, the verb *infect* intrinsically conveys negative implications as it implies the transmission and spread of harmful agents, often resulting in adverse effects on individuals or systems. Similarly, *wedge*, in Example 2, is lexically connoted as evaluative: it refers to the insertion of an external object in a way that creates disruption, obstruction, or difficulty in removing it. Both these instances of get-passive thus align with the literature that reports such construction as not semantically neutral. Similarly, three more of the six instances present get-passives which are not neutral nor unproblematic, as in Examples 3, 4, and 5 below:

Example 3. Press release:
From facts to fake news: How information <GET PASSIVE> *gets distorted*. (UNREP)

Example 4. Scientific article:
Harris <BE PASSIVE> *has been accused* of forcing businesses to donate to his foundations. Mayor Ludovic <GET PASSIVE> *got entangled* because he supposedly instructed his employee[s] to collect money for the cult-like foundations. (SCAR)

Example 5. Scientific article:
As one user lamented, "I wish more of the public would do some research into them and see how much of a risk they are but sadly most wont [sic] — because once you do and you see the truth on them, you <GET PASSIVE> *get labeled* as an 'anti vaxxer' which equates to fool. In the next few years, the vaccine industry <BE PASSIVE> *is set* to be a nearly 105 billion dollar industry. People should really consider who profits off of our ignorance". (SCAR)

The get-passive in Example 3 has prominence as it is located in the title of the press release. The adversative meaning is here attached to the passive *gets distorted* by the context, as it refers to information which is manipulated, possibly with the intent of deceiving by presenting a biased or misleading perspective. In Example 4, the passive is used in the past and it should be stressed that it is part of a written passage that is reported and analyzed in the study, but it has nothing to do with scientific language or the language of the authors of the scientific article. As for the previous examples, this get-passive has adversative meaning: *got entangled* is here used figuratively, as it is referred to a person who was caught in a situation which generates conflict and adversity. Similarly, the get-passive used in Example 5, *get labeled*, clearly implies adversative implications for the subject, as it suggests the attribution and/or impositions of an undesired or unrequested association between the subject and the label. Therefore, it is plausible that also in this case, the choice of a get-passive is connected with the adversative context of the text. Additionally, this particular passive is included in a *verbatim* quotation from a Facebook post, which forms part of the material analyzed in a scientific article in the field of Data Visualization. As mentioned above, some scientific papers are based on verbal data, which are often cited and quoted within the articles: it turns out that chunks from other genres (from Twitter posts to transcriptions of spoken interactions) are juxtaposed with the academic language used in academic articles. As a result, the academic language gets contaminated when quantitatively analyzed with corpus linguistic tools that do not discriminate between the two.

Moreover, one hypothesis is that there is an additional type of contamination which consists of the authors of the scientific articles being influenced by

the stylistic and linguistic strategies used in the materials they analyze. Such a possibility can be observed in Example 6 below.

Example 6. Scientific article:
Informants used selfies on social networks to <GET PASSIVE> *get noticed* (e.g. Darla: "I want people to see me. I want people to *notice* me") and get responses from others (e.g. Eden: "To *get* Likes or any kind of reaction out of it or comments out of it. That's usually my intention). (my emphasis, SCAR)

Example 6 shows that the get-passive used in this particular scientific article is *get noticed*. Looking at the subsequent quoted statement, the informant in the first passage uses *notice* and the one in the second passage uses *get* even though not with the same function: 'get' is used as a main verb – not in a passive construction – in the phrase *to get likes*. This might suggest that the passive *get noticed* was used by the authors of the scientific article because their writing style was being – probably unconsciously – influenced by the linguistic style and lexicon of materials (often spoken English or written social media interactions) they were analyzing. Clearly, these data are not enough to demonstrate this, but it might be an interesting hypothesis to take into consideration for future research.

5. *Conclusions*

This study is rooted in the field of research aimed at understanding the communicative and linguistic strategies underlying the transformation of technical-scientific language into science-communication language, a more accessible variety that enables larger audiences of non-experts to read and understand such content. In particular, the analysis conducted here focused on passive voice which is one characterizing feature of written academic texts. Specifically, the data consist of two bijective corpora where every scientific article in the SCAR corpus is matched with its popularized version, namely a university press release in the UNREP corpus. This methodological design provided comparable data in terms of topics, disciplines, and individual linguistic phenomena.

The data used in this analysis have a distinctive characteristic that is uncommon (but not novel: see Garzone 2014; Sumner et al. 2016); the data at hand here are exceptionally well-suited to reveal if and how discursive transformations occur

when comparing expert-oriented scientific texts with non-expert oriented ones. In terms of results, it should be first noted that although the press releases in this study are strictly related to their matching scientific articles, the difference in the use of passives between the two corpora is still statistically significant. When considering the total number of passives in the two corpora, it is expectedly higher in the corpus of scientific articles and the difference is statistically significant. The lower presence of passive voices in press releases compared to scientific articles is probably mostly due to the difference in technicality and abstraction of the two genres, but it could also be partly attributed to the difference in focus. Typically, the methodological sections of academic articles are the densest in terms of passive voice compared to other sections (see Leong 2014; Riley 1991). Conversely, the specifics of the methods are usually overlooked in the press releases, which instead prioritize presenting the research findings and their implications (Fahnestock 1998, 335). As a result, the necessity for passive voice decreases, leading to a lower presence of it in the press releases.

Furthermore, it has been found that there is a predominance of be-passives in both corpora and this difference is statistically significant. More precisely the be-passive is the only type of passive whose presence is statistically significant in the two corpora. This could be related to the fact that it is the most prototypical of the four types of passives and also the most easily recognized and avoided by writers of press releases who try to informalize their texts. This being said, also the difference in the presence of bare passive is still noticeable and contributes to draw the distance between the two types of texts.

Finally, the get-passive is – minimally – present in both corpora, even if this is the only one of the four types of passives whose occurrence is higher in the corpus of press releases. The presence of the get-passive appears to be simultaneously surprising and reasonable. It is surprising because the use of get-passives has been mostly recorded in colloquial and informal registers and it is interesting to identify a number of occurrences in the two corpora, limited to a small number of articles and referring to scientific content. It is also reasonable, though, because on the one hand, the press releases accommodate the language to the needs of a wider, nonexpert audience and it is to be expected that some less prototypically formal linguistic means are used. On the other hand, scientific articles in the field of communication, linguistics, sociology, but also statistics and mathematical modeling often rely on authentic language material from spoken interactions, interviews, social media posts, written summaries.

This material is often quoted in the scientific articles, contaminating/affecting their language and probably also the academic writing style of the researchers themselves. In this paper, scientific language was conceptualized as a register with overarching macro-characteristics shared across all disciplines. However, future research in this field may carry out analogous studies, while distinguishing between various subjects to ascertain whether the frequency of passive voice usage varies across disciplinary fields. Moreover, such research could investigate potential disparities in the extent to which both the subjects under investigation and the methodological approaches impact scientific discourse.

Acknowledgments

The author wishes to express her gratitude to Dr. Salvatore Attardo and Dr. Nina Deliu for their time and the invaluable comments on a previous version of this paper.

Bibliography

Amdur, Robert J., Kirwan, Jessica and Morris, Christopher G. 2010. "Use of the passive voice in medical journal articles." *AMWA Journal: American Medical Writers Association Journal 25*, no. 3: 98-110.

Autzen, Charlotte. 2014. "Press releases—the new trend in science communication." *Journal of Science Communication* 13, no.3: 1-8. https://doi.org/10.22323/2.13030302.

Bakalar, Tomas, Magdalena Razusova, Henrieta Pavolova, and Barbara Hlavnova. 2014. "Use of Active and Passive Verb Forms in Papers from the Field of Chemical Engineering." *Communications-Scientific letters of the University of Zilina* 16, no. 3: 102-108.

Banks, David. 1994. *Writ in water, aspects of the scientific journal article*. Brest Cédex: E.R.L.A. Université de Bretagne Occidentale.

Banks, David. 2017. "The extent to which the passive voice is used in the scientific journal article, 1985-2015." *Functional Linguistics* 4, no. 1: 1-17. https://doi.org/10.1186/s40554-017-0045-5.

Baratta, Alexander. 2009. "Revealing stance through passive voice." *Journal of Pragmatics* 41: 1406-1421. https://doi.org/10.1016/j.pragma.2008.09.010.

Barber, Charles L. 1962. "Some measurable characteristics of modern scientific prose." *Contributions to English syntax and philology*, 21-43.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 1992. "The multi-dimensional approach to linguistic analyses of genre variation: an overview of methodology and findings." *Computers and the Humanities* 26, no. 5/6: 331-345. https://doi.org/10.1007/BF00136979.

Biber, Douglas. 2006. *University Language: A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan and Finegan, Edward. 1999. *The Longman Grammar of Spoken and Written English*. New York: Pearson Education.

Biber, Douglas and Gray, Bethany. 2016. *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.

Brezina, Vaclav. 2018. *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.

Brezina, Vaclav, Weill-Tessier, Pierre, and McEnery, Anthony. 2020. #LancsBox v. 5.x. [software]. Available at: http://corpora.lancs.ac.uk/lancsbox.

Carter, Ronald and McCarthy, Michael. 1999. "The English get-passive in spoken discourse: description and implications for an interpersonal grammar." *English Language & Linguistics* 3, no. 1: 41-58. https://doi.org/10.1017/S136067439900012X.

Catenaccio, Paola. 2007. "New(s) Genre and Discursive Identity. The Changing Face of the Press Release in the Age of the Internet." In *Multimodality in corporate communication. Web genres and discursive identity*, edited by Giuliana Garzone, Gina Poncini and Paola Catenaccio, 55-72. Milano: Franco Angeli.

Catenaccio, Paola. 2008. "Press Releases as a Hybrid Genre: Addressing the Informative/Promotional Conundrum." *Pragmatics* 18, no.1: 9-31. https://doi.org/10.1075/prag.18.1.02cat.

Chappell, Hilary. 1980. "Is the get-passive adversative?" *Papers in Linguistics* 13: 411-452. http://dx.doi.org/10.1080/08351818009370504.

Chen, Mingyao and Ye, Qiongxia. 2014. "A Corpus-Based Study on Original English Abstracts and Translated English Abstracts: A Case Study of Passive Voice and Pronouns." *International Journal of English Linguistics* 4, no. 6: 52. http://dx.doi.org/10.5539/ijel.v4n6p52.

Collins, Peter. 1996. "Get-passives in English." *World Englishes* 15: 43-56. https://doi.org/10.1111/j.1467-971X.1996.tb00091.x.

Conrad, Susan, and Biber, Douglas. 2001. *Variation in English: Multi-Dimensional Studies*. Routledge.

Coto Villalibre, Eduardo. 2015. "Is the get-passive really that adversative?." *Miscelánea: A Journal of English and American Studies* 51: 13-26.

Council (Council of the European Union). 2021. "Council Recommendation (EU) 2021/2122 of 26 November 2021 on a Pact for Research and Innovation in Europe." *Official Journal of the European Union* L 431/1, Brussels, 26 November 2021. Accessed November 8, 2023. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021H2122.

Council (Council of the European Union). 2023. "Council conclusions on high-quality, transparent, open, trustworthy and equitable scholarly publishing, 3949th meeting." 9616/23, RECH 190, EDUC 169, PI 77, DIGIT 96, Brussels, 23 May 2023. Accessed November 7, 2023. https://data.consilium.europa.eu/doc/document/ST-9616-2023-INIT/en/pdf.

Di Ferrante, Laura, Lattanzi, Giulia, and Petrocelli Emilia. 2021. "Those who get lost in communication science. Promotion and downplay in university press releases." In *Discursive perspectives on knowledge dissemination in corporate and professional communication: focus on ethical and ideological aspects*, edited by Giuliana Garzone, Walter Giordano, and Sergio Pizziconi. 31-57, Timișoara: Editura Politehnica.

Dunn, Olive Jean. 1961. "Multiple comparisons among means." *Journal of the American statistical association* 56, no. 293: 52-64. https://doi.org/10.1080/01621459.1961.10482090.

Fairclough, Norman. 1995. *Critical discourse analysis*. London: Longman.

Fahnestock, Jeanne. 1998. "Accommodating Science: The Rhetorical Life of Scientific Facts." *Written communication* 3, no. 3: 275-296. https://doi.org/10.1177/0741088398015003006.

Garzone, Giuliana. 2014. "News production and scientific knowledge: Exploring popularization as a process." In *Die Sprache der Popularisierung [The Language of Popularization]*, edited by Giancarmine Bongo and Giuditta Caliendo, 73-107. Bern: Peter Lang.

Garzone, Giuliana. 2020. *Specialized Communication and Popularization in English*. Roma: Carocci Editore.

Gillings, Mathew, Gerlinde Mautner, and Paul Baker. 2023. *Corpus-Assisted Discourse Studies*. Cambridge: Cambridge University Press.

Givón, Talmy and Yang, Lynne. 1994. "The Rise of the English Get-Passive." In *Voice: Form and Function*, edited by Barbara Fox and Paul J. Hopper, 119-149. Amsterdam: Benjamins.

Hiltunen, Turo. 2016. "7 Passives in Academic Writing: Comparing Research Articles and Student Essays Across Four Disciplines." In *Corpus Linguistics on the Move*, edited by María José López-Couso, Belén Méndez-Naya, Paloma Núñez-Pertejo, and Ignacio M. Palacios-Martínez, 132-157. Leiden: Brill. https://doi.org/10.1163/9789004321342_008.

Hundt, Marianne, Schneider, Gerold and Seoane, Elena. 2016. "The Use of the Be-Passive in Academic Englishes: Local Versus Global Usage in an International Language." *Corpora* 11, no. 1: 29-61. https://doi.org/10.3366/cor.2016.0084.

Hyland, Ken. 2010. "Constructing proximity: Relating to readers in popular and professional science." *Journal of English for academic purposes* 9, no. 2: 116-127.

Hyland, Ken, and Jiang, Feng Kevin. 2017. "Is academic writing becoming more informal?" *English for Specific Purposes* 45, 40-51. https://doi.org/10.1016/j.esp.2016.09.001.

Iddings, John. 2007. "A functional analysis of English humanities and biochemistry writing with respect to teaching university composition." *Novitas-ROYAL (Research on Youth and Language)* 2, no. 1: 60-87.

Jacobs, Geert. 1999. *Preformulating the News: An Analysis of the Metapragmatics of Press Releases*. Vol. 60. Amsterdam: John Benjamins Publishing.

Leech, Geoffrey, Marianne Hundt, Christian Mair, and Nicholas Smith. 2009. *Change in Contemporary English: A Grammatical Study*. Studies in English Language. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642210.008.

Leong, Ping Alvin. 2014. "The Passive Voice in Scientific Writing. The Current Norm in Science Journals." *Journal of Science Communication* 13, no. 1: 1-16. https://doi.org/10.22323/2.13010203.

Leong, Ping Alvin. 2020. "The passive voice in scientific writing through the ages: A diachronic study." *Text & Talk* 40, no. 4: 467-489. https://doi.org/10.1515/text-2020-2066.

Leong, Ping Alvin. 2021. "The passive voice in scholarly writing: A diachronic look at science and history." *Finnish Journal of Linguistics* 34: 77-102. https://journal.fi/finjol/article/view/103168.

Lindh, Karolina. 2020. "Press releases as medical knowledge." In *Movement of Knowledge: Medical Humanities Perspectives on Medicine, Science, and Experience*, edited by Kristofer Hansson and Rachel Irwin, 111-132. Förlagssystem in Falun: Nordic Academic Press.

Lu, Lu. 2013. "A Contrastive Study of the Passive Voice in Journal Articles in Theoretical and Applied Linguistics." *Chinese Journal of Applied Linguistics* 36, no. 4: 465-478. https://doi.org/10.1515/cjal-2013-0031.

Luzón, María José. 2013. "Public Communication of Science in Blogs: Recontextualizing Scientific Discourse for a Diversified Audience." *Written Communication* 30, no. 4: 428-457. https://doi.org/10.1177/0741088313493610.

Mair, Christian, and Leech, Geoffrey N. 2006. "Current changes in English syntax." In *The Handbook of English Linguistics*, edited by Bas Aarts and April M. McMahon, 318-342. Oxford: Blackwell.

Mair, Christian. 1998. "Corpora and the study of the major varieties of English: issues and results." In *The Major Varieties of English*, edited by Hans Lindquist, Staffan Klintborg, Magnus Levin and Maria Estling. Papers from MAVEN 97, Växjö 20-22 November 1997, 139-157. Växjö: Edman and WesterlundsTryckeri.

Matthews, Richard. 1993. *Papers on Semantics and Grammar*. Frankfurt am Main: Peter Lang.

McLaren, Yvonne, and Gurău, Călin. 2005. "Characterising the Genre of the Corporate Press Release." *LSP and Professional Communication* 5, no.1: 10-29.

Millar, Neil, Budgell, Brian, and Fuller, Keith. 2013. "'Use the active voice whenever possible': The Impact of Style Guidelines in Medical Journals." *Applied Linguistics* 34, no. 4: 393-414. https://doi.org/10.1093/applin/ams059.

O'Neill, Ruth and Russell, Alex. M. 2019. "Grammarly: Help or Hindrance? Academic Learning Advisors' Perceptions of an Online Grammar Checker." *Journal of Academic Language and Learning* 13, no. 1: A88-A107. https://journal.aall.org.au/index.php/jall/article/view/591.

Partington, Alan, Alison Duguid, and Charlotte, Taylor. 2013. *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Vol. 55. John Benjamins Publishing.

Petrocelli, Emilia, Di Ferrante, Laura, and Ghia, Elisa. 2022. "Multi-Word Combinations in Science Communication: A Corpus-Driven Study of University Press Releases." *Online Journal of Communication and Media Technologies* 12, no. 4: 1-19. https://doi.org/10.30935/ojcmt/12247.

Rachul, Christen M. 2008. *Comparing apples and oranges: a socio-rhetorical analysis of writing in biology and history*. PhD diss., Carleton University.

Riley, Kathryn. 1991. "Passive voice and rhetorical role in scientific writing." *Journal of technical writing and communication* 21, no. 3: 239-257. https://doi.org/10.2190/Y51Y-P6QF-3LCC-4AUH.

Schwarz, Sarah. 2015. "Passive Voice in American Soap Opera Dialogue." *Studia Neophilologica*, 87, no. 2: 152-70, https://doi.org/10.1080/00393274.2015.1049831.

Seoane, Elena. 2006. "Changing styles: On the recent evolution of scientific British and American English". In *Syntax, Style and Grammatical Norms: English from 1500-2000*, edited by Christiane Dalton-Puffer, Dieter Kastovsky and Nikolaus Ritt, 191-211. Bern: Peter Lang.

Seoane, Elena. 2009. "Syntactic complexity, discourse status and animacy as determinants of grammatical variation in Modern English." *English Language & Linguistics* 13, no. 3: 365-384. https://doi.org/10.1017/S1360674309990153.

Seoane, Elena, and Hundt, Marianne. 2018. "Voice alternation and authorial presence: Variation across disciplinary areas in academic English." *Journal of English Linguistics* 46, no. 1: 3-22. https://doi.org/10.1177/0075424217740 9.

Seoane, Elena, and Loureiro-Porto, Lucia. 2005. "On the Colloquialization of Scientific British and American English." *ESP Across Cultures* 2: 106-118.

Sulaiman, Eman. 2022. *A corpus-based study of the passive voice in texts produced by Norwegian learners and native speakers of English*. Master's thesis. University of Oslo.

Sumner, Petroc, Solveiga Vivian-Griffiths, Jacky Boivin, Andrew Williams, Lewis Bott, Rachel Adams, Christos A. Venetis, Leanne Whelan, Bethan

Hughes, and Christopher D. Chambers. 2016. "Exaggerations and caveats in press releases and health-related science news." *PloS one* 11, no. 12: e0168217.

Tarone, Elaine, Dwyer, Sharon, Gillette, Susan. and Icke, Vincent. 1981. "On the use of the passive in two astrophysics journal papers." *ESP Journal* 1, no. 2: 121-139. https://doi.org/10.1016/0272-2380(81)90004-4.

Tarone, Elaine, Dwyer, Sharon, Gillette, Susan. and Icke, Vincent. 1998. "On the use of Passive and Active Voice in Astrophysics Journal Papers: With Extensions to Other Languages and Other Fields." *English for Specific Purposes* 17, no. 1: 113-132. https://doi.org/10.1016/S0889-4906(97)00032-X.

Xiao, Richard, McEnery, Tony and Qian, Yufang. 2006. "Passive Constructions in English and Chinese." *Languages in Contrast* 6, no. 1: 109-149. Web. https://doi.org/10.1075/lic.6.1.05xia.

Wallis, Sean, and Mehl, Seth. 2022. "4 Comparing Baselines for Corpus Analysis. Research into the get-passive in speech and writing." In *Data and Methods in Corpus Linguistics: Comparative Approaches*, edited by Ole Schützler and Julia Schlüter, 101-126. Cambridge: Cambridge University Press.

Zhang, Yiqiong. 2018. "Retailing science: genre hybridization in online science news stories." *Text & Talk* 38, no. 2: 243-265. https://doi.org/10.1515/text-201.

Zinkevich, Nina A., and Ledeneva, Tatiana V. 2021. "Using grammarly to enhance students' academic writing skills." *Professional Discourse & Communication* 3, no. 4: 51-63. https://doi.org/10.24833/2687-0126-2021-3-4-51-63.

Laura Di Ferrante is Assistant Professor of English Language and Linguistics at the University of Milan, in the Department of International, Legal, Historical and Political Studies. Co-founding editor and editor in chief of E-JournALL, EuroAmerican Journal of Applied Linguistics and Languages, her education background and work experience have been carried out both in Italy and the United States. Her main research interests focus on workplace and media discourse, intercultural pragmatics, and applied linguistics.