Article

Public communication of technoscience in the news: A cross-linguistic Multidimensional analysis of English and Italian newspapers

Virginia Zorzi iD
University of Bergamo, Italy

Federico Neresini
University of Padova, Italy

Alberto Cammozzo
Politecnico di Milano, Italy

Abstract

Technoscience has historically been endowed by modern societies with cultural authority, which has undergone significant questioning since the end of WWII. In this study, we consider technoscience communication in online newspapers to test whether its linguistic and communicative features differentiate it from other types of news, possibly reflecting greater cultural authority. Focusing on news in English versus Italian, we apply Multidimensional analysis to a comparable English-Italian online news corpus comparing technoscience-related articles with all remaining articles. Results indicate overall linguistic and communicative homogeneity between science and other domains. Nevertheless, small differences seem to suggest an effort to make technoscience accessible and engaging in English, whereas Italian technoscience-related news tends to be more formal and richer in specialised information.

Corresponding author:
Virginia Zorzi, University of Bergamo, 2, Piazza Rosate, Bergamo 24129, Italy.
Email: virginia.zorzi@unibg.it

Introduction

Science and technology, pervasive in modern societies, play a dominant role in the production of knowledge and decision processes (Beck, 1986; Stehr, 1994). Historically, modern societies have endowed science with cultural authority (Bauer et al., 2018), which has however undergone significant changes, especially after WWII, when technoscience started to be problematised as not necessarily completely beneficial to humans and the environment (Beck, 1986). At the same time, technoscience – an expression introduced by Latour within the field of Science and Technology Studies (Neresini, 2023) – acquired a strategic function in the eyes of powerful political and economic actors, which started to allocate funds in various research sectors to enhance economic wealth, military power and social progress (Bucchi and Trench, 2016: 153).[1]

Since then, the intertwining among politics, economics and technoscience has increased, as has the pressure on scientists to seek public attention and support (Bauer et al., 2018: 7). Technoscience has become more and more visible in the public sphere. Public attention to any social issue from any domain – intended as any area of knowledge and/or activity – tends to be a scarce resource (Hilgartner and Bosk, 1988). Therefore, technoscience-related concerns need to compete and interact with many other domains, including policy, economics and social concerns, in the media as well as in other arenas of public discourse. As a result of this entanglement, the cultural authority of technoscience – although still overall permeating modern societies – has been somewhat weakened by the need for accountability and by the competing authorities of other domains. Yet more recently, it has been further threatened by post-truth views (Lynch, 2020; Sismondo, 2017), which openly subject scientific explanations to personal or group opinions and interests.

Verbal language, essential in technoscience communication, can have an influence on the features and authority attributed to technoscience. Therefore, it may be asked whether the type of language used to cover technoscience-related content in non-specialised media environments represents and constructs technoscience as an area of knowledge with a unique status, in terms of prestige and cultural authority, with respect to the other domains which populate those environments. Our study takes this inquiry as a starting point, with a focus on English versus Italian-speaking contexts, assuming that a comparison between them may highlight similarities and differences in their approach to technoscience.

Online news can be regarded as one of the most important media for the public communication of technoscience (Neresini and Lorenzet, 2018: 155), and provides large amounts of potentially analysable linguistic data. We therefore chose online newspapers as the object of this study. In our analysis, we made an effort to tackle linguistic variation beyond that determined by the use of more or less technoscience-related vocabulary, to find out whether other features, such as verbal tenses, syntactic structures or pronoun use, operated differently when technoscience was covered. We formulated the following research questions:

RQ1: Is the language used to cover technoscience in English and Italian online news different in terms of grammatical, syntactic and lexical features,[2] from that used to cover other domains, intended as areas of knowledge and/or activity?

RQ2: How are English and Italian data similar or different in their linguistic characterisation of technoscience?

Answering these questions might be useful in assessing whether any of the abovementioned linguistic aspects may frame technoscience as different from other areas of knowledge and/or activity by conveying somehow distinctive communicative functions. This may allow us to verify whether any observed differences suggest some form of epistemological dominance and/or cultural authority of technoscience with respect to other domains appearing in the news. No previous assumption is made about which linguistic features and communicative functions can represent a domain as more or less authoritative; rather, the distribution of several linguistic features is explored in order to uncover its variation and the consequences it has in terms of communicative functions.

Background

The nature and communicative purposes of the public communication of technoscience have been researched in many different areas, including sociology (Bucchi and Trench, 2014; Shinn and Whitley, 1985), history (Cooter and Pumphrey, 1994), communication studies (Nisbet et al., 2002; Scheufele, 2013), rhetoric (Fahnestock, 1986; Gross, 1994) and linguistics, where it has been considered a type of Language for Specific Purposes (LSP) (Gea-Valor et al., 2010: 9; Gotti, 2003: 203–216) or through the lens of Discourse Studies (Calsamiglia, 2003; Myers, 2003). Literature has highlighted that public discourses concerning technoscience most prototypically flow from specialised settings (e.g. labs, research journals) to increasingly non-expert communicative environments (e.g. news). In this process, technoscience is progressively represented in terms of facts, losing the nuances of tentativeness and uncertainty it had within the research community (Cloître and Shinn, 1985; Fleck, 1979; Latour, 1987). However, this should not be simplistically understood as a one-way process in which pristine and objective knowledge is simplified and transferred to a passive audience.

Firstly, technoscientific knowledge can follow multiple, non-linear paths among scientists and, for example, political and economic stakeholders, advocacy groups and the general public (Bucchi, 2004: 114–122). Secondly, the production and communication of technoscience cannot be considered completely neutral nor objective, in specialised as well as in lay contexts (Bucchi, 2004: 65; Kamenova, 2017; Latour, 1987). Moreover, audiences' beliefs and experiences have a role in technoscience reception and interpretation, which has prompted new communication models that feature aspects of public engagement, dialogue and collaboration (Bucchi, 2004; Callon, 1999; Winter, 2004). Thirdly, when knowledge flows from experts to lay people, it does not undergo a mere reduction from its original version to a simplified (possibly corrupted or incomplete) one. Public communication has been conceptualised as a form of translation (Jakobson, 1959: 233) in its broad sense of adaptation to different contexts (Garzone, 2006). Scholars have also described it in terms of reformulation and rewriting (Gotti, 2012), re-elaboration and re-contextualisation (Incelli, 2018; Motta-Roth and Scherer, 2016). 'Translation' is also a key-concept in Science and Technology Studies, particularly Actor-Network

Theory (ANT), which aims to analyse how networks of actors – or actants – are assembled in order to shape and to maintain scientific knowledge and technological innovation (Callon, 1984; Latour, 1987; Law, 1991). In ANT, translation can tie together in networks heterogeneous actors/actants with different interests. In general, a change of context affects language (e.g. Hyland, 2005: 94–100), content and communicative goals. Studies have thus identified technoscience in the news as a kind of hybridisation between science and journalism (Motta-Roth and Scherer, 2016), or of pedagogy mixed with infotainment (Elorza, 2014).

Most research on the language of the public communication of technoscience has focused on English, but other languages have also been considered (e.g. Calsamiglia and van Dijk, 2004 on Spanish; Maeseele et al., 2015 on Belgian; Morfakis, 2017 on Greek; Motta-Roth and Dos Santos Lovato, 2011 on Portuguese), with some cross-linguistic analyses (e.g. Almansa, 2021 on German vs Spanish; Elorza, 2014 on English vs Spanish). Different outlets have been examined; examples include the press (Garzone, 2006), popular science books (Hunston, 2013), blogs (Mahrt and Puschmann, 2014) and podcasts (Ye, 2021). Methodologies have ranged from more qualitative to predominantly quantitative, and most studies have focused on one particular topic or issue within the broader domain of technoscience, whereas the present study offers a more comprehensive view of technoscience in the news.

Corpus and method

Multidimensional analysis

To carry out our analysis, we applied a corpus-based method called Multidimensional Analysis (henceforth MDA). MDA draws upon multivariate statistics – that is, procedures designed to handle data featuring many variables – to uncover patterns in the use of a set of selected linguistic features in a given corpus. These patterns, and their variation within the corpus, can then be interpreted and used to describe linguistic and communicative variation in that corpus, on the basis of the assumption that linguistic features that tend to be used together (here meaning that they appear in the same text) potentially contribute to realising shared communicative functions.

MDA was first devised by Biber (1988) to describe linguistic and communicative variation in a general English corpus, providing measurements of the degree of informational focus, narrativity, and overt persuasion, among others, of the analysed texts. It has since then been applied to several languages and different corpora (see Berber Sardinha and Pinto, 2014); despite its complexity (Stamatatos et al., 2000; Xiao and McEnery, 2005) and sensitivity even to small methodological changes, demanding caution in result interpretation (Lee, 2000) – it remains a powerful, statistically grounded method which accounts for multiple linguistic elements.

We used MDA to highlight any linguistic or communicative aspect which may point to salient ways of covering technoscience with respect to all other domains. We carried out two separate MDAs – in English and Italian – to compare results across languages. The patterns of variation detected through MDA are based on language-specific sets of linguistic features and are therefore language-specific themselves, which means that they

cannot be directly compared across different languages. Rather, cross-linguistic comparisons based on MDA allow analysts to notice differences and similarities between patterns emerging from different languages and their communicative implications (see Biber, 1995). In addition, they may enable cross-linguistic comparisons concerning the same genre or text type – that is, where it is situated within the linguistic and communicative patterns of variation detected in the different languages, which is the focus of the present study. Ours is, to the best of our knowledge, the first Italian version of MDA to have been created.

Corpus *collection*

We collected a comparable English-Italian online news corpus. As a technical requirement, MDA needs to be performed on a corpus where the sub-corpora among which variation is to be assessed are, on the whole, quantitatively balanced. In order to address our research questions, we therefore needed to perform MDA on a comparable corpus whose English and Italian components were representative of the online newspaper genre in the two languages. Moreover, each of the two components needed to include a balanced sample of the various types of news and of the different domains which normally appear in online newspapers (see research questions in the Introduction). These included technoscience-related news, which we wanted to situate within the linguistic and communicative variation potentially described through MDA. At the level of corpus collection, this meant achieving – both in the Italian and the English sub-corpora – balance among different news sub-genres, which in turn represented different thematic sections and different types of online newspaper articles (including technoscience-related ones).

We chose to treat online newspaper publication sections (e.g. politics, economics, sports, etc.) as a proxy for news sub-genres, and since sections vary across newspapers, we identified seven section categories which we used to classify and collect articles: business, culture and leisure, homepage news, news (general news, especially concerning politics), op-ed (comment and opinion articles), science and technology, sport.

Our source for corpus collection was the media monitoring platform developed within the TIPS (Technoscientific Issues in the Public Sphere) research initiative,[3] which collects articles from different online newspapers daily and makes them available for textual analysis (Cammozzo et al., 2020; Di Buccio et al., 2022). TIPS also features a machine learning procedure called 'classifier', which automatically estimates how likely it is that an individual text is relevant to technoscience,[4] regardless of its publication section. This meant that we had two distinct systems at our disposal to identify technoscience-relevant news articles: one based on publication sections, which was also the basis to balance our samples among news sub-genres; the other based on the actual content of each article, evaluated by the TIPS classifier as more or less likely to be technoscience-relevant (see Supplemental Appendix 3 for examples). We were able to exploit both criteria in the final part of MDA.

Our English-Italian comparable corpus consisted of articles published between 2013 and 2020.[5] The sources comprise, in both languages, national daily newspapers representative of diverse editorial stances among those available in the TIPS database: *The Guardian, The New York Times, Financial Times, The Telegraph* in English; *Avvenire,*

Corriere della Sera, Il Giornale, La Stampa, Il Mattino, Il Messaggero, Repubblica, Il Sole 24 Ore in Italian.[6]

We extracted the balanced random samples of articles from the English and Italian sections of the TIPS database, where each article was assigned to one of the seven above-mentioned section categories. We established a minimum length of 100 words for articles to be collected, to avoid including incomplete articles or items only containing a headline or a single sentence, which can sometimes be expected in automatic news text collection. Our sample extraction plan for both languages required ideally the same number of articles in each section category, for each newspaper, and for each year. We gave priority to balance among section categories, for each of which we aimed at collecting a minimum of 3800 articles. Whenever our ideal sample extraction plan could not be followed (i.e. there were not enough articles of at least 100 words available within the TIPS database for a given section category), some articles were drawn from other newspapers, and subsequently from other years to compensate the gap in that section category.

We managed to collect an average of 3,766.43 articles (standard deviation 194.05) and 2,599,440.29 word tokens (standard deviation 316,016.26) for each section category in English; and an average of 3,891.71 articles (standard deviation 194.72) and 1,867,973.43 word tokens (standard deviation 298,194.33) for each section category in Italian. The English sub-corpus comprised 26,365 articles and 18,196,082 word tokens, while the Italian one comprised 27,242 articles and 13,075,814 word tokens (see Tables 1 and 2 and Figures 1 and 2 for an overview of word token distributions within the sub-corpora).

Multidimensional analysis on a comparable English-Italian news corpus

The first stage of MDA consists in selecting a set of linguistic features (hence LFs) whose use may vary, following variation in communicative purposes, within the corpus. The English LF set was based on that used in Biber (1988, see Supplemental Appendix 1), while the Italian one was derived and adapted from that used by Biber et al. (2006) for Spanish, given the closeness between the two languages (see Supplemental Appendix 2). Subsequently, an ad-hoc software was devised to produce frequency counts of each LF in each article, within each sub-corpus.[7]

The second stage of MDA involves Exploratory Factor Analysis (EFA), which uses LF frequency counts to identify (1) LFs which tend to appear in the same texts, potentially sharing communicative purposes, and (2) LFs which tend not to appear in the same texts, potentially contributing to realising diverging communicative purposes. EFA arranges LFs following (1) and (2) in a series of 'factors', which can be regarded as hidden variables underlying LF variation.[8] These factors are interpreted as 'dimensions of variation' (Biber, 1988: 91), whose description includes the communicative purposes which most likely characterise them.

The presence and importance of each dimension varies across texts in the corpus; thus, the third stage of MDA exploits EFA-related techniques to make dimensions measurable in the texts through 'dimension scores' (Biber, 1988: 93–97, from the concept of 'factor scores', see DiStefano et al., 2009). Dimension scores enable analysts to compare single texts or groups of texts along each of the dimensions identified.

Table 1. Distribution of word tokens in the English sub-corpus across section categories and newspapers.

	Financial Times		The Guardian		The New York Times		The Telegraph		Tot word tokens per section category	% Grand Total
	Word tokens	% Grand Total	Word tokens	% Grand Total	Word tokens	% Grand Total	Word tokens	% Grand Total		
Business	645,336	3.55%	752,655	4.14%	671,730	3.69%	347,145	1.91%	2,416,866	13.28%
Culture	198,212	1.09%	1,049,321	5.77%	696,814	3.83%	564,821	3.10%	2,509,168	13.79%
Home	0	0.00%	1,395,818	7.67%	1,301,214	7.15%	466,698	2.56%	3,163,730	17.39%
News	845,230	4.65%	714,012	3.92%	646,705	3.55%	380,069	2.09%	2,586,016	14.21%
Op-ed	523,887	2.88%	819,582	4.50%	588,369	3.23%	755,178	4.15%	2,687,016	14.77%
Sci-Tech	0	0.00%	1,420,137	7.80%	962,077	5.29%	321,320	1.77%	2,703,534	14.86%
Sport	282,454	1.55%	0	0.00%	1,342,489	7.38%	504,809	2.77%	2,129,752	11.70%
Total word tokens per newspaper	2,495,119	13.71%	6,151,525	33.81%	6,209,398	34.12%	3,340,040	18.36%	18,196,082	100.00%

Table 2. Distribution of word tokens in the Italian sub-corpus across section categories and newspapers.

	Avenire		Corriere della Sera		Il Giornale		La Stampa	
	Word tokens	% of Grand Total	Word tokens	% of Grand Total	Word tokens	% of Grand Total	Word tokens	% of Grand Total
Business	281,289	2.15%	244,719	1.87%	0	0.00%	247,143	1.89%
Culture	403,777	3.09%	287,323	2.20%	0	0.00%	268,772	2.06%
Home	237,924	1.82%	189,714	1.45%	403,027	3.08%	230,847	1.77%
News	284,664	2.18%	266,619	2.04%	0	0.00%	251,908	1.93%
Op-ed	201,794	1.54%	47,945	0.37%	0	0.00%	77,639	0.59%
Sci-Tech	33,778	0.26%	347,124	2.65%	0	0.00%	373,680	2.86%
Sport	204,932	1.57%	240,855	1.84%	0	0.00%	258,911	1.98%
Total word tokens per newspaper	1,648,158	12.60%	1,624,299	12.42%	403,027	3.08%	1,708,900	13.07%

	Il Mattino		Il Messaggero		Repubblica		Il Sole 24 Ore	
	Word tokens	% of Grand Total	Word tokens	% of Grand Total	Word tokens	% of Grand Total	Word tokens	% of Grand Total
Business	231,732	1.77%	173,220	1.32%	233,512	1.79%	256,986	1.97%
Culture	162,803	1.25%	206,076	1.58%	274,630	2.10%	306,062	2.34%
Home	153,379	1.17%	176,924	1.35%	279,452	2.14%	234,675	1.79%
News	193,396	1.48%	204,060	1.56%	274,374	2.10%	254,962	1.95%
Op-ed	435	0.00%	0	0.00%	1,120,608	8.57%	1,049,805	8.03%
Sci-Tech	209,284	1.60%	246,901	1.89%	274,232	2.10%	246,984	1.89%
Sport	189,046	1.45%	220,260	1.68%	257,867	1.97%	259,765	1.99%
Total word tokens per newspaper	1,140,075	8.72%	1,227,441	9.39%	2,714,675	20.76%	2,609,239	19.95%

Table 2. (Continued)

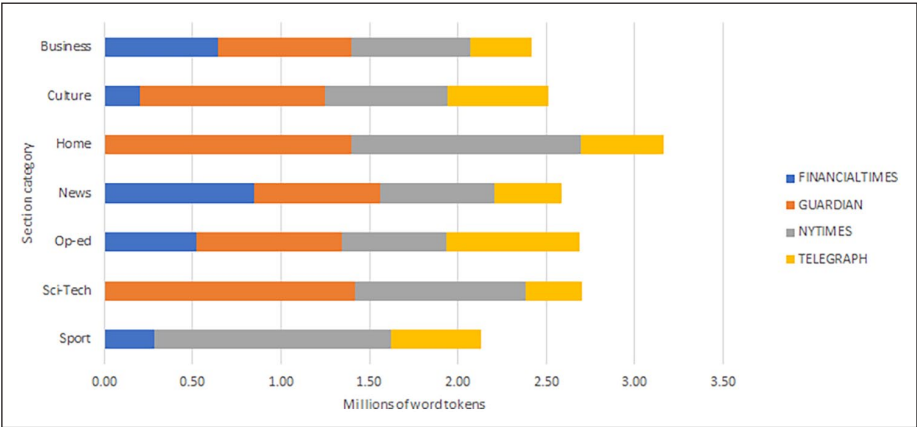|	Tot. word tokens per section category	% Grand Total
Business	1,668,601	12.76%
Culture	1,909,443	14.60%
Home	1,905,942	14.58%
News	1,729,983	13.23%
Op-ed	2,498,226	19.11%
Sci-Tech	1,731,983	13.25%
Sport	1,631,636	12.48%
Total word tokens	13,075,814	100.00%

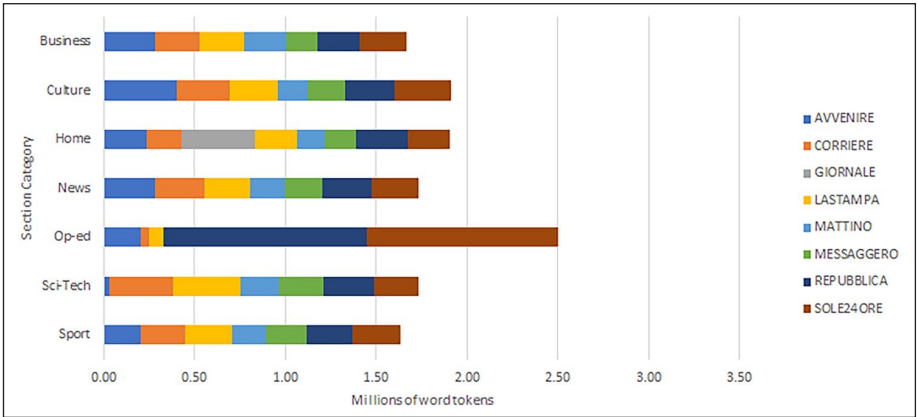**Figure 1.** Structure of the English sub-corpus.

Figure 2. Structure of the Italian sub-corpus.

Therefore, after performing EFAs in English and Italian using the R software (R Core Team, 2017), we used dimension scores to compare different groups of texts in both languages, namely:

a)   articles published in 'Science and Technology' (henceforth TS, standing for technoscience) sections vs articles published in all other sections (i.e. all remaining articles);

b)   articles categorised as relevant to technoscience by the above-mentioned TIPS classifier – that is whose probability of being technoscientifically relevant was higher than 50% according to the classifier, regardless of their publication section – vs all remaining articles, which we considered non-relevant.

c) maximally relevant articles, that is the top 25% most relevant articles among those already classified as relevant by the classifier – probability thresholds vary depending on the distribution of probability scores in English and Italian texts – vs all remaining articles.

We used the JASP software (JASP Team, 2021) to perform the comparisons and apply the Mann-Whitney *U*-test (Kabacoff, 2011: 166) – which we deemed the most suitable to our data – for statistical significance, to make sure that any difference observed between article groups was worth reporting and not due to chance. The maximum *p*-value threshold to regard a difference as statistically significant was 0.05.

Results

Following the EFA procedure, we tested several factor/dimension configurations (Biber, 1988: 81–91; for a more detailed account, see Bartholomew et al., 2008: 184–192), in order to identify the ones which most closely and plausibly represented the latent variables underlying linguistic and communicative variation within each sub-corpus. Each dimension can be thought of as a continuum between two stylistic poles, respectively and randomly assigned a positive and negative value by the EFA software. Similarly, the scores used to measure the presence and importance of dimensions in single texts can be positive or negative, placing the text more or less towards either end of the continuum, on the basis of their absolute value.

English MDA

Identification and interpretation of the dimensions. For the English sub-corpus, four dimensions were identified. Table 3 shows the LFs which the software selected as most representative of each factor/dimension – an aspect measured by factor loadings (Biber, 1988: 85–86); LFs with positive and negative loadings characterise, respectively, texts with positive and negative dimension scores on each dimension. This should in turn reflect the communicative and stylistic profile of those texts. If one dimension only has LFs with positive loadings, it means that texts with a negative score on that dimension are characterised by a low frequency or absence of the positively-loaded LFs, and therefore that they might be characterised by communicative functions that diverge from those potentially expressed by the features found in the factor.

Dimension 1 (D1), similar to the first dimension identified by Biber (1988: 104–108), consists in a continuum between articles with a relatively informal, personal and conversational style – whose structure tends to be fragmented rather than compact – and articles with a more formal and informational focus. The former are characterised by the co-occurrence of LFs such as contractions, first and second-person pronouns and determiners, vague/unspecific language – most notably the *it* pronoun and indefinite pronouns – and adverbs, the combination of which points to conversational interaction, and LFs such as *be* used as a main verb, which produces a more fragmented sentence structure with respect, for instance, to adjectival pre-modification (Example 1).[9] Articles at the opposite side of the continuum are characterised by LFs which create high information

Table 3. Factorial structure chosen for the English MDA.

Factor/Dimension 1		Factor/Dimension 2		Factor/Dimension 3		Factor/Dimension 4	
LFs	Factor loading	LFs	Factor loading	LFs	Factor loading	LFs	Factor loading
Contracted forms	0.69	Public verbs	0.79	Present tenses	0.73	Total nouns	0.47
Be as a main verb	0.66	Past tenses	0.74	3rd-pers. pron./det.	−0.42	Proper nouns	−1.03
1st-pers. pron./det.	0.64	Subordinator *that* deletion	0.58	Past tenses	−1.21		
Do as a non-main verb	0.61	*That* as a verb complement	0.49				
Total adverbs	0.60	Suasive verbs	0.38				
Present tenses	0.53	Perfect aspect (verbs)	0.37				
pronoun *it*	0.53	Agentless passive	0.37				
Analytic negation	0.52						
Clause coordination	0.49						
Indefinite pron.	0.48						
Predicative adjectives	0.47						
Demonstrative pronouns	0.45						
2nd-pers. pron./det.	0.42						
Private verbs	0.39						
Nominalisation	−0.35						
Preposit. Phrases	−0.38						
Avg. sentence length	−0.40						
Proper nouns	−0.57						
Avg. word length	−0.69						

density and specificity, due to the high frequency of nominalisations; they also have longer words and sentences on average, as well as many prepositional phrases, all of which reflect potentially complex, integrated sentence and phrase structures (Example 2). This dimension was called 'Interactional vs Informational focus'.

1)   Cook for **Others.** Hello again! **Still** here, **still** trying to fill Sam's sneakers. (Or tennis shoes? This map of regional terms resurfaced last week, **and I'm still** processing.) (2019, Op-ed, *Financial Times*)

2)   The French competi<u>tion</u> author<u>ity</u> imposed fines <u>of</u> €951m (£748m) <u>on</u> personal hygiene and cleaning products companies <u>on</u> Thursday <u>over</u> price fixing <u>in</u> supermarkets. (2014, Business, *The Telegraph*)

At the positive end of Dimension 2 (D2) are articles which rely heavily on speech attribution (signalled by the use of public verbs, which include reporting verbs like *say*) and on passive verbs with no explicit agents, mostly in past tenses. Speech attribution allows authors to attribute content to external sources, which may provide it with external support and/or reduce the author's commitment and responsibility towards that content. Moreover, passive verbs contribute to a detached, depersonalised tone, which may help the articles sound more objective (Example 3). At the opposite side of the D2 continuum are texts with relatively few of the above-mentioned features (Example 4), where authors tend not to resort to speech attribution, fully committing to their propositions or simply presenting content as factual and/or objective. We therefore called D2 'Agentless reported communication'.

3)   Three people **have been injured** after a suspected chemical substance **was delivered** to a London restaurant in an envelope. The unidentified substance, believed to have been corrosive, is understood to **have been sent** to Feng Sushi restaurant in Borough Market. The London Fire Brigade **said** it **had been called** to 'a chemical incident' shortly before 2pm. (2017, News, *The Telegraph*)

4)   Catalogued by Charles Messier in the late 18th century, M41 is a collection of about one hundred stars in a volume about 25 light years across. It lies at a distance of around 2300 light years from Earth. (2020, Science and Technology, *The Guardian*)

Dimension 3 (D3) can be represented as a continuum between articles that mainly focus on present events or situations, which is reflected by the predominance of present tenses (Example 5), and articles concerning past events, with a more narrative style, reflected by the predominance of past tenses in combination with third-person pronouns and determiners (Example 6). Therefore, we called D3 'Present vs Past-narrative focus'.

5)   [. . .] songbirds, which **form** lifelong mating pairs, **have** brain systems perfectly tuned to fit together. While you **sort** through the messages of admirers, deciding whom to make your Valentine, **consider** finches. Young males in this family of feathered crooners **learn** the song of <u>their</u> father [. . .]. (2018, Homepage, *The New York Times*)

Table 4. Mean dimension scores for articles related to technoscience (TS) as compared to the rest of the English sub-corpus.

		D1 Mean score	D2 Mean score	D3 Mean score
Section-based comparison	TS sections	0.37	0.26	−0.02
	Non-TS sections	−0.07	−0.05	0.004
Classifier-based comparison (I)	Relevant articles	−0.18	−0.22	0.13
	Non-relevant articles	0.03	0.04	−0.02
Classifier-based comparison (II)	Most relevant articles	0.4	0.02	0.17
	All other articles	−0.02	−0.0007	−0.007

Three comparisons are shown: the first one is based on publication section, the remaining ones on the TIPS technoscientific relevance classifier. A grey background is used to indicate cases in which the difference in dimension scores between the compared groups is statistically significant (p-value $\leqslant 0.05$). Here, no cases were found where the magnitude of the difference, measured through the rank biserial correlation,[11] was above 0.1, indicating a large difference as compared to the overall differences analysed.

   6)   A plane carrying 189 people over Indonesia <u>crashed</u> into the sea on Monday, officials <u>said</u>. Lion Air Flight JT-610 <u>had been flying</u> north from the capital, Jakarta, to the city of Pangkal Pinang on the island of Bangka when <u>it went</u> missing. (2018, Homepage, *The New York Times*)

Finally, we found that Dimension 4 (D4) had resulted from the presence of articles almost entirely consisting of noun or proper noun lists (e.g. products, sports teams or stock market data). Since we found that D4 was not relevant enough with respect to the communicative purposes characterising the corpus, and since it was due to texts which, in our perspective, do not fully qualify as news stories, we excluded it from our subsequent analyses.

Comparing news on technoscience with news on other domains. All observed differences were extremely small, suggesting that any linguistic and communicative aspect peculiar to technoscience in the news appears to be quite limited. Yet, some differences were statistically significant (see Table 4) and were therefore considered in more detail.

   Along D1, the TS-section group resulted as being slightly more informal and conversational, and consequently less informational, than the non-TS one. As for D2, TS-section articles scored higher than articles in other sections considered together, which points to a tendency for the former to adopt a more objective and impersonal style where speech attribution is relatively frequent.

   When we used the classifier to distinguish between technoscientifically relevant and non-relevant articles, we observed significant differences in D2 and D3. More specifically, relevant articles appeared to contain less speech attribution and agentless passives than non-relevant articles, contrary to what we had observed in the section-based comparison. This may be because the classifier detects technoscience regardless of publication section; therefore, higher D2 scores may be more related to publication section than the actual presence of technoscience content. Along D3, relevant articles appeared to deal with present events significantly more than non-relevant ones.

Significant differences between the maximally relevant articles (whose minimum relevance probability was 88%) and all remaining articles were observed in D1 and D3. Along D1, maximally relevant articles appeared to be slightly more informal, interactional and conversational, in line with the section-based comparison; along D3, maximally relevant articles appeared to be more focused on the present than the past.

Italian MDA

Identification and interpretation of the dimensions. For the Italian sub-corpus, we opted for a five-dimension solution, shown in Table 5.

We interpreted D1 as representing a continuum between articles that combine description and argumentation or explanation (Example 7), and articles that lack any of these communicative purposes (Example 8), where authors report on events or situations without elaborating on them nor explicitly expressing specific views on them. Description here was determined by the presence of manner and quantity adverbs; argumentation or explanation were mainly determined by the presence of emphasis and downtoning expressions – which, together with evaluation adverbs, comparative/superlative adjectives, and adversative conjunctions, can be used for explanation and/or supporting one's viewpoint (i.e. argumentation). Therefore, we called D1 'Descriptive and argumentative/ explanatory style'.

7) Dunque, **come allora** ci si attendeva un Bush **ancora più determinato** sul fronte estero, per esempio nei confronti dell'Iran, **ora** ci si attende un Obama **più determinato** nelle politiche liberal sui diritti e contro gli eccessi di Wall Street[. . .]. **Poi** i fatti, **probabilmente**, non seguiranno (**troppo**). (2013, Op-ed, *Il Sole 24 Ore*)

Therefore, exactly as Bush was then expected to be firmer on foreign policy, for instance towards Iran, today Obama is expected to be firmer concerning liberal policies on civil rights and against the excesses of Wall Street. [. . .] However, actions are not likely to follow (too closely).

8) Nicoletta Passet, moglie di Diego Bianciotto, difesa dall'avvocato Andrea Cianci, questa mattina, martedì 20 ottobre, ha dimostrato di essere estranea al reato di bancarotta: è stata assolta. (2020, Homepage, *La Stampa*)

Nicoletta Passet, Diego Bianciotto's wife, represented by lawyer Andrea Cianci, proved this morning that she was not involved in the bankruptcy case, and was acquitted.

On the positive side of the D2 continuum we found articles that tend to deal with known entities and present processes or situations, combining determinate articles with verbs in the third-person and/or in the indicativemood and or in present-tenses (Example 9). On the opposite side of the continuum are texts with a relatively less verbal – and therefore potentially more nominal – style, as in Example 10. We called D2 'Present-focused processes or states'.

Table 5. Factorial structure chosen for the Italian MDA.

Factor/Dimension 1		Factor/Dimension 2		Factor/Dimension 3	
LFs	Factor loading	LFs	Factor loading	LFs	Factor loading
Adverbs	1.08	3rd person verbs	1.31	1st person verbs	1.12
Quantity adverbs	0.73	Indicative mood	0.90	1st person pron./determiners	0.98
Evaluation adverbs	0.61	Finite present verbs	0.71	possessive pron./determiners	0.47
Adverbs of manner	0.50	Determinate articles	0.36	2nd person pron./determiners	0.36
Time adverbials	0.49			3rd person pron./determiners	−0.52
Emphatics	0.43				
adversative conjunctions	0.42				
Downtoners	0.42				
Compar./superl. adjectives	0.42				

Factor/Dimension 4		Factor/Dimension 5	
LFs	Factor loading	LFs	Factor loading
Post-modifying attr. adjectives	0.69	Finite past verbs	0.96
Relational adjectives	0.68	Finite present verbs	−0.56
Mean word length	0.67		
Nominalisation	0.64		
Common nouns	0.58		
Proper nouns	−0.54		

9) 'Da parte mia c'**è** grande preoccupazione, **l'**Empoli **ha cambiato** modulo tattico, **abbina** qualità tecniche ad una fase difensiva molto interessanti, partite facili in questo campionato non **esistono** [. . .]'. (2019, Sport, *Il Messaggero*)

There's great concern on my part, Empoli [soccer team] has changed its formation, and combines very interesting technical qualities and defence strategies; easy matches don't exist [. . .].

10) **È indetto** un concorso pubblico, per titoli ed esami, a 50 posti per **l'**accesso alla qualifica iniziale del**la** carriera prefettizia. (2017, Economy, *Avvenire*)

A public call has been announced for the selection, by means of qualifications and examinations, of 50 positions with access to prefect tenure.

D3 was understood as representing a distinction between articles characterised by a relatively dialogic and personally involved tone (Example 11), given by first and second-person verbs, pronouns and determiners, and much less personal and involved articles (Example 12), where third-person pronouns and determiners are more frequent. We called D3 'Personal, involved interaction'.

11) Cara [. . .], [. . .] Da allora **ho continuato** a «seguir**ti**», neanche **fossi** uno stalker. **Digitavo** il **tuo** nome su internet e **mi apparivi**. Nel corso degli anni **sei cambiata**, ma l'amore per il **tuo** papà è sempre rimasto lo stesso. (2013, Homepage, *Il Giornale*)

Dear [. . .], [. . .] Since then, I have been "following you", as if I were some sort of stalker. I typed your name on the internet and there you were. You changed through the years, but your affection towards your dad remained unchanged.

12) Va dal dentista, presumibilmente dopo molti anni, e quello che il dottore vede <u>lo</u> lascia pietrificato. In rete è stato diffuso il video di un uomo che ha in bocca una strana macchia nera. (2016, Culture and Leisure, *Il Mattino*)

He goes to the dentist – most likely for the first time in years – and the doctor sees something that leaves them speechless. A video was published online of a man with a strange dark stain in his mouth.

D4, somewhat similar to D1 in English, consists in a continuum between primarily informational articles, rich in specific and technical content (Example 13), and less information-dense articles where proper nouns seem to be particularly frequent (Example 14). Information density is here conveyed by frequent nouns and nominalisations, which also indicate information specificity, further supported by the use of longer words on average, along with adjectival post-modification and relational adjectives.[10] We called D4 'Informational, technical and specific focus'.

13)   **Lauree magistrali**, il 13 settembre la **prova** d'**accesso** per **Scienze** della **Formazione primaria**, il 25 ottobre **Professioni sanitarie** – Sono stati pubblicati sul **sito** del **ministero** dell'**Istruzione date** e **modalità** delle **prove** d'**accesso** per il **corso** di **Laurea magistrale** a **ciclo unico** in **Scienze** della **Formazione primaria** [. . .]. (2019, Culture and Leisure, *La Repubblica*)

Master's degrees: entry tests to Primary Education Sciences will be on 13th September, Health Professions test will be on 25th October – the dates and procedures for the entry tests to the five-year Primary Education Sciences Master's degree have been specified on the website of the Ministry of Education.

14)   Nella seconda **giornata** di **campionato** di **serie** A, **vittoria** 'esagerata' dell'Inter che supera, come lo scorso **anno**, in **casa** il Sassuolo 7-0. Stecca il Napoli al San Paolo sconfitto dal Chievo 0-1 [. . .]. (2014, Homepage, *La Repubblica*)

In the second day of the major league championship, "exaggerated" triumph by Inter [soccer team], which passes Sassuolo [soccer team] playing at home, as it did last year. Napoli [soccer team] hits the wrong note at San Paolo [stadium], being defeated by Chievo [soccer team] by 0-1 [. . .].

D5 mirrors our English D3, as it represents a continuum between narrative articles focused on the past – where past tense verbs are particularly frequent, as in Example 15 – and present-focused articles, rich in present tense verbs (Example 16). We called D5 'Past vs present focus'.

15)   '[. . .] Mi spiace che la vostra pizza **sia arrivata** tardi, ma **dovevo** salvare un cane'. Si **è giustificato** così Steven Donovan ai suoi clienti. Il ragazzo **stava** facendo una consegna a domicilio quando improvvisamente **ha visto** un cagnolino saltare nel mezzo di una strada molto trafficata. (2017, Culture and Leisure, *La Stampa*)

"I'm sorry I was late in delivering your pizza, but I had to rescue a dog". This is how Steven Donovan justified himself to his clients. He was delivering food when he suddenly noticed a small dog jumping into a congested road.

16)   'Da due anni seguo una dieta sana e non perdo peso: com'è possibile?' Ma come fanno molte persone a perdere anche 8–10 kg in un mese? Io mi **sono messa** a dieta (ho 42 anni, una buona salute), ma non riesco a vedere questi risultati mirabolanti. (2018, Science and Technology, *Il Corriere della Sera*)

I've been following a healthy diet for two years but I'm not losing weight: how is this possible? How on Earth can many people lose up to 8-10 Kg in one month? I've started a diet (I'm 42 and healthy), but I cannot achieve such amazing results.

Comparing news on technoscience with news on other domains.   All observed differences were very small; yet some of them were statistically significant, as shown in Table 6. More specifically, articles in TS sections feature more description and explanation/argumentation

Table 6. Mean dimension scores for articles related to technoscience as compared to the rest of the Italian sub-corpus.

		D1 Mean score	D2 Mean score	D3 Mean score	D4 Mean score	D5 Mean score
Section-based comparison	TS sections	0.292	−0.043	−0.481	**1.85**	−0.48
	Non-TS sections	−0.045	0.007	**0.074**	−0.28	**0.07**
Classifier-based comparison (I)	Relevant articles	0.46	0.007	−0.5	**2.86**	−0.62
	Non-relevant articles	−0.06	−0.0008	0.061	−0.35	**0.07**
Classifier-based comparison (II)	Most relevant articles	0.37	−0.08	−0.69	**4.11**	−0.56
	All other articles	−0.01	0.002	**0.10**	−0.11	**0.02**

Three comparisons are shown: the first one is based on publication section, the remaining ones on the TIPS technoscientific relevance classifier. A grey background is used to indicate cases in which the difference in dimension scores between the compared groups is statistically significant (p-value ⩽ 0.05). Bold is used to indicate cases where the magnitude of the difference, measured through the rank biserial correlation10, was above 0.1, indicating a large difference as compared to the overall differences analysed.

(D1), less personal involvement and dialogic interaction (D3), a more specific informational focus (D4) and more present events (D5) than articles published in non-TS sections. The classifier-based comparisons essentially confirmed these findings.

Discussion and conclusions

The overall small size of the differences observed, even if statistically significant, allows us to maintain that the language used to report on technoscience in English and Italian online newspapers is not generally very different, in terms of grammar, syntax, general lexis and related communicative purposes, from that used to report on other domains, which points to a certain degree of homogeneity in news language, at least regarding the use of the LFs here considered. This lack of specificity calls for a re-consideration of technoscience communication and its features, often considered in isolation, sometimes as a specific genre relying on a relatively specific set of features. On the other hand, in this study the communication of technoscience was re-contextualised within one of its possible wider media contexts – online newspapers. From a linguistic and communicative perspective, technoscience communication appeared to be shaped more, as suggested by MacLuhan, by its own medium than by the actual presence of technoscientific content in the texts.

Nevertheless, some differences do emerge, reflecting some circumscribed elements of uniqueness for technoscience news. In English, both articles published in technoscience-related sections and articles recognised by the TIPS classifier as strongly related to technoscience were slightly more conversational and less information-dense than other articles. This might reflect an effort to make technoscientific knowledge more accessible to non-experts by conveying specialised content in an engaging and relatively informal tone. This may as well point to the effort, on the part of actors in technoscientific communities, to compete for public attention in news media (see the Introduction). Moreover, articles whose content is related to technoscience are significantly more focused on

present than past events. This is unsurprising, since technoscience reporting is likely to deal with recent developments of research and innovation, as well as with the description of reality in the present.

Italian results were consistent among all the comparisons we carried out: both the one examining articles in technoscience-related publication sections versus articles in other sections, and those examining articles with technoscience-related content versus articles without it, regardless of publication section (i.e. as established by the TIPS classifier). Firstly, results regarding our first Italian dimension of variation suggest that technoscience-related articles are more descriptive and argumentative/explanatory than other articles; qualitative checks revealed that in articles with a high score along this dimension, descriptions were often combined with elaborate explanations – hence the relatively high frequency of adverbs, adversative conjunctions and comparatives/superlatives – as a strategy to convey specialised knowledge to non-experts. Italian Dimensions 3 – 'Personal, involved interaction' – and 4 – 'Informational, technical and specific focus' – both somehow concern the level of formality and interaction: we noticed a tendency, in technoscience-related articles, to adopt a slightly more formal, information-dense and less interactional style. We interpreted this finding as reflecting comprehensiveness and detachment, as an expression of cultural and epistemological authority attributed to technoscience. Although small in size and restricted to some texts only, this would suggest that the Anglophone and Italian cultural contexts frame technoscience in the public sphere differently: the former tends to bring the communication of technoscience close to non-specialised conversational language, possibly to achieve greater involvement and support, while the latter might maintain, in some cases, an idea of cultural authority, if not superiority, of technoscience with respect to other domains.

This finding, based on wide-ranging and quantitatively relevant evidence, is partly consistent with the awareness, developed especially in translation and cross-cultural studies, that the author-reader relationship does not only change across genres and registers, but is influenced by culture. On that account, scholars observed that certain linguistic and cultural contexts, including the Italian one, prefer writer-oriented styles high in informational load and details, whereas other contexts, including the most influential English-speaking ones, favour more reader-oriented styles lower in informational load and thus more synthetic and clearer (Scarpa, 2020: 138). Further research might be conducted on whether such stylistic cross-cultural differences mostly (if by a small extent) pertain to technoscience news communication, among all the news domains, as our study seems to suggest.

Lastly, technoscience coverage in Italian also tends to focus on the present rather than the past, showing how both linguistic and cultural contexts associate technoscience to state-of-the art research and innovation.

In conclusion, the present analysis approached the public communication of technoscience in English and Italian online newspapers replicating Biber's Multidimensional method and creating an Italian version of it. Our analysis revealed some linguistic and communicative homogeneity characterising news language, which suggests that the cultural specificity of technoscience may not be clearly foregrounded in its public image, although it still emerges in some news pieces, suggesting a difference in said public image between English and Italian-speaking societies.

Author's Note

Alberto Cammozzo is now affiliated with University of Padova, Italy.

ORCID iD

Virginia Zorzi (iD) https://orcid.org/0000-0002-5168-2465

Supplemental material

Supplemental material for this article is available online.

Notes

1. The authors have jointly discussed and conceived this paper. Nevertheless, individual contributions are identified as follows. Virginia Zorzi: conceptualisation, data curation, formal analysis, methodology, writing (original draft, review & editing); Federico Neresini: conceptualisation, supervision, writing (review & editing); Alberto Cammozzo: data curation, writing (review & editing).
2. The focus here is on linguistic features and their functions in discourse, adopting a context, genre and topic-independent approach as a starting point (Biber, 1988: 72). Reference to grammar, syntax and lexis is here made to convey that various levels of the language are being considered in selecting features; these features are selected because they potentially fulfil certain communicative functions in discourse, with the aim of uncovering how communicative functions may vary across different categories of texts, such as, in the present case, news on technoscience and news unrelated to it. Moreover, grammar and lexis are here understood as poles of a 'lexico-grammatical continuum', with syntax embedded in grammar itself (Halliday and Matthiessen, 2004: 24; Morley, 2000). Our understanding of the term 'linguistic features' is therefore in line with that found in Biber's work, and is different from the meaning currently attributed to the term in machine learning approaches, including sentiment analysis, fake news detection, document similarity detection, recommendation systems etc., which were not available at the time of Biber's analysis.
3. See http://www.tipsproject.eu/tips/#/public/home
4. Relevance to technoscience is defined by the creators of the classifier as reference, in the text, to at least two of the following: 'a scientist; a scientific journal; a research center/laboratory; a scientific discipline (excluding humanities and the social sciences); a generic reference to research processes and/or technological innovations; a discovery, an innovation, a scientific instrument or a medical apparatus'. (Crabu et al., 2021: 7/19). For more details see Di Buccio et al. (2022).
5. This time span was chosen on the basis of availability in the TIPS platform.
6. Dealing with two different linguistic contexts required arranging different configurations with regards to the number of newspapers to be included. For instance, we had to include the

Vatican's newspaper *Avvenire* and the markedly conservative *Il Giornale* in order to more accurately reflect the Italian context.

7. To produce frequency counts, we annotated the texts with a part-of-speech tagger (TextPro – (Pianta et al., 2008)) and then applied a connected series of regular expressions – sequences of characters which can be used to search for specific strings within a text – to the part-of-speech-tagged texts, in order to identify and count each linguistic feature. We automated the repetitive part of the work, making it possible to analyse multiple texts.

8. These are generally called latent variables. For an account on factor analysis and other similar multivariate methods in the social sciences, see Bartholomew et al. (2008).

9. In all examples, when present, positive LF of the dimension being discussed are shown in bold, while negative LFs on the same dimension are underlined.

10. Relational adjectives like *muscolare/muscular*, *industriale/industrial* or *presidenziale/presidential*, derive from nouns and indicate a relationship between the noun they modify and the noun they derive from (Bisetto, 2010: 65).

11. For more on rank biserial correlation, see https://www.scalestatistics.com/rank-biserial.html (accessed 03/08/2022).

References

Almansa IR (2021) German-Spanish contrastive analysis for the translation of science journalism: Science popularization in news on coronavirus. *Mutatis Mutandis* 14(1): 240–265.

Bartholomew DJ, Steele F, Galbraith J, et al. (2008) *Analysis of Multivariate Social Science Data*. London, New York: CRC Press.

Bauer MW, Pansegrau P and Shukla R (2018) Image, perception and cultural authority of science–by way of introduction. In:Bauer MW, Pansegrau P and Shukla R (eds) *The Cultural Authority of Science*. Milton Park: Routledge, pp.3–21.

Beck U (1986) *Risikogesellschaft*. Frankfurt A.M: Suhrkamp Verlag.

Berber Sardinha T and Pinto MV (eds) (2014) *Multi-Dimensional Analysis, 25 Years on: A Tribute to Douglas Biber*. Amsterdam: John Benjamins Publishing Company.

Biber D (1988) *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber D (1995) *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

Biber D, Davies M, Jones JK, et al. (2006) Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora* 1(1): 1–37.

Bisetto A (2010) Relational adjectives crosslinguistically. *Lingue e linguaggio* 9(1): 65–85.

Bucchi M (2004) *Science in Society: An Introduction to Social Studies of Science*. Milton Park: Bucchi.

Bucchi M and Trench B (eds) (2014) *Routledge Handbook of Public Communication of Science and Technology*. Milton Park: Routledge.

Bucchi M and Trench B (2016) Science communication and science in society: A conceptual review in ten keywords. *Tecnoscienza (Italian Journal of Science & Technology Studies)* 7(2): 151–168.

Callon M (1984) Some elements of a sociology of translation. Domestication of the scallops and the fishermen of St Brieuc Bay. *Sociological Review* 32(1): 196–233.

Callon M (1999) The role of lay people in the production and dissemination of scientific knowledge. *Science Technology & Society* 4(1): 81–94.

Calsamiglia H (2003) Popularization discourse. *Discourse Studies* 5(2): 139–146.

Calsamiglia H and van Dijk TA (2004) Popularization discourse and knowledge about the genome. *Discourse & Society* 15(4): 369–389.

Cammozzo A, Di Buccio E, Neresini F, et al. (2020) Monitoring technoscientific issues in the News. In:Koprinska I (ed.) *ECML PKDD 2020 Workshops*. Berlin: Springer, pp.536–553.

Cloître M and Shinn T (1985) Expository practice: Social, cognitive and epistemological linkages. In:Shinn T and Whitley R (eds) *Expository Science. Forms and functions of popularization*. Dordrecht: Reidel, pp.31–60.

Cooter R and Pumphrey S (1994) Separate spheres and public places: Reflections on the history of science popularization and science in popular culture. *History of Science* 32(3): 237–267.

Crabu S, Giardullo P, Sciandra A and Neresini F (2021) Politics overwhelms science in the Covid-19 pandemic: Evidence from the whole coverage of the Italian quality newspapers. *PloS one* 16(5): e0252034.

Di Buccio E, Cammozzo A, Neresini F, et al. (2022) TIPS: search and analytics for social science research. In: *Proceedings of the 2nd Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2022)*, Samatan, Gers, France, 4–7 July 2022, vol. 3178

DiStefano C, Zhu M and Mîndrilă D (2009) Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research and Evaluation* 14(20): 1–11.

Elorza I (2014) Newsworthiness, attribution and lexicogrammatical strategies in two types of news articles in English and Spanish. *Topics in Linguistics* 14(1): 16–33.

Fahnestock J (1986) Accommodating science: The rhetorical life of scientific facts. *Written Communication* 3(3): 275–296.

Fleck L (1979) *Genesis and Development of a Scientific Fact*. Chicago, London: University of Chicago Press.

Garzone G (2006) *Perspectives on ESP and Popularization*. Milano: CUEM.

Gea-Valor ML, García-Izquierdo I and Esteve MJ (eds) (2010) *Linguistic and Translation Studies in Scientific Communication*. Bern: Peter Lang.

Gotti M (2003) *Specialized Discourse: Linguistic Features and Changing Conventions*. Bern: Peter Lang.

Gotti M (2012) La riscrittura del testo da specialistico a divulgativo. *Altre Modernità* 11: 145–159.

Gross AG (1994) The roles of rhetoric in the public understanding of science. *Public Understanding of Science* 3(1): 3–24.

Halliday MAK and Matthiessen CMIM (2004) *An Introduction to Functional Grammar*. London: Hodder Arnold.

Hilgartner S and Bosk CL (1988) The rise and fall of social problems: A public arenas model. *American Journal of Sociology* 94(1): 53–78.

Hunston S (2013) Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text & Talk* 33(4–5): 617–640.

Hyland K (2005) *Metadiscourse: Exploring interaction in writing. London*: Continuum.

Incelli E (2018) Popularising the Higgs boson: A corpus-assisted approach to reporting scientific discovery in online media. *Corpora* 13(2): 169–203.

Jakobson R (1959) On linguistic aspects of translation. In:Brower RA (ed.) *On Translation*. Cambridge: Harvard University Press, pp.232–239.

JASP Team (2021) *JASP (Version 0.14.0.0)*. [Computer software]. https://jasp-stats.org/ (accessed 03 July 2023).

Kabacoff R (2011) *R in Action: Data Analysis and Graphics With R*. Shelter Island; NY: Manning Publications.

Kamenova K (2017) Media portrayal of stem cell research: Towards a normative model for science communication. *Asian Bioethics Review* 9(3): 199–209.

Latour B (1987) *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, MA: Harvard University Press.

Law J (ed) (1991) *A Sociology of Monsters. Essays on Power, Technology and Domination*. Milton Park: Routledge.

Lee DYW (2000) *Modelling Variation in spoken and written English: The multi-dimensional approach revisited*. PhD dissertation, Lancaster University, UK.

Lynch M (2020) We have never been anti-science: Reflections on science wars and post-truth. *Engaging Science, Technology, and Society* 6: 49–57.

Maeseele PA, Deneckere K, Panis K, et al. (2015) The energy question in the Belgian daily press during 2010: The role of region, newspaper type and newspaper section. *Journal of Communication Science* 14(1): 1–17.

Mahrt M and Puschmann C (2014) Science blogging: An exploratory study of motives, styles, and audience reactions. *Journal of Communication Science* 13(03): 1A05–1A17.

Morfakis C (2017) Human Gene Mapping: The Mass Media Iconography of the Human Genome Project in the Most Popular Greek Newspapers. In:Petermann HI, Harper PS and Doetz S (eds) *History of Human Genetics*. Berlin: Springer, pp.285–315.

Morley GD (2000) *Syntax in Functional Grammar: An Introduction to Lexicogrammar in Systemic Linguistics*. London, New York: Continuum.

Motta-Roth D and Dos Santos Lovato C (2011) The hegemonic power of science in the scientific popularization discourse. *Calidoscópio* 9(3): 251–268.

Motta-Roth D and Scherer AS (2016) Science popularization: Interdiscursivity among science, pedagogy, and journalism. *Bakhtiniana: Revista de Estudos do Discurso* 11(2): 164–189.

Myers G (2003) Discourse studies of scientific popularization: Questioning the boundaries. *Discourse Studies* 5(2): 265–279.

Neresini F (2023) Technoscience. In:Colombo E and Rebughini P (eds) *Framing Social Theory*. Milton Park, New York: Routledge, pp.123–138.

Neresini F and Lorenzet A (2018) The great narrative: Analysing the cultural authority of science through media attention in Italy. In:Bauer MW, Pansegrau P and Shukla R (eds) *The Cultural Authority of Science*. Milton Park: Routledge, pp.155–170.

Nisbet MC, Scheufele DA, Shanahan J, et al. (2002) Knowledge, reservations, or Promise? A media effects model for public perceptions of science and Technology. *Communication Research* 29(5): 584–608.

Pianta E, Girardi C and Zanoli R (2008) The TextPro tool suite. In: *Proceedings of the 6th edition Language Resources and Evaluation Conference (LREC)*, Marrakech, Morocco, 28–30 May 2008.

R Core Team (2017) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (accessed 10 October 2022).

Scarpa F (2020) *Research and Professional Practice in Specialised Translation*. Hoboken, NJ: Palgrave Macmillan.

Scheufele DA (2013) Communicating science in social settings. *Proceedings of the National Academy of Sciences* 110(Supplement 3): 14040–14047.

Shinn T and Whitley RP (eds) (1985) *Expository Science: Forms and Functions of Popularisation*. Berlin: Springer.

Sismondo S (2017) Post-truth? *Social Studies of Science* 47(1): 3–6.

Stamatatos E, Fakotakis N and Kokkinakis G (2000) Text genre detection using common word frequencies. In: *Proceedings of the 18th conference on computational linguistics*, vol. 2, pp.808–814. Association for Computational Linguistics.

Stehr N (1994) *Knowledge Societies*. London: Sage.

Winter E (2004) Public communication of science and technology: German and European perspectives. *Science Communication* 25(3): 288–293.

Xiao Z and McEnery A (2005) Two approaches to genre analysis: Three genres in modern American English. *Journal of English Linguistics* 33(1): 62–82.

Ye Y (2021) From abstracts to "60-second science" podcasts: Reformulation of scientific discourse. *Journal of English for Academic Purposes* 53: 101025.

Author biographies

Virginia Zorzi is a research fellow in English language and linguistics at the University of Bergamo. She received a Ph.D. in Linguistic, Philological and Literary Sciences from the University of Padova. Her research can be situated within the fields of discourse analysis and contrastive studies, and her work chiefly draws on methods related to corpus linguistics. She has carried out research on the public communication of technoscience, conspiracy theory discourses and migration discourses.

Federico Neresini (Ph.D. Sociology and Social Research) teaches Sociology of Innovation and Digital Sociology at the University of Padova (Italy). His research interests are focused on Science & Technology Studies, in particular on the construction processes of scientific knowledge and the analysis of technoscience in the public sphere. Recently, he has become interested in the relationships between big-data and scientific research activities, as well as the implications of the availability of large amounts of digital data for the social sciences. Since 2008, he has coordinated the Padua Science, Technology and Innovation Studies research unit (PaSTIS - www.pastis-research. eu); in particular, he is leading the TIPS project (Technoscientific Issues in the Public Sphere), aimed at analysing the societal relevance and evolution of technoscience using the media as a source through a big data approach.

Alberto Cammozzo is an ICT consultant and independent researcher affiliated with the University of Padova, Italy, and involved in multidisciplinary work.