

Looking into international research groups' digital discursive practices: Criteria and methodological steps in the compilation of the *EUROPRO* digital corpus

Daniel Pascual – Pilar Mur-Dueñas – Rosa Lorés
University of Zaragoza / Spain

Abstract –The *EUROPRO* digital corpus was designed by the *InterGedi* research group, based at the University of Zaragoza (Spain). The main focus of *InterGedi* is the analysis of the textual resources used by international research groups as part of their dissemination and visibility strategies. The corpus comprises a collection of 30 international research project websites funded by the *European Horizon2020 Programme (EUROPROwebs corpus)*. By looking into their websites, 20 projects were observed to maintain a *Twitter* account and the tweets from these accounts were the basis for the compilation of the *EUROPROtweets* corpus. This paper delves into the criteria used for the selection of the research project websites and the methodological steps taken to classify, label and tag the verbal component in these websites and tweets. The paper discusses the challenges in the compilation of the corpus because of the dynamic, hypermodal, and hypermedial nature of the digital texts it contains. The paper closes by underlining the potential uses and applications of *EUROPRO* in order to gain insights into the digital discursive and professional practices used by international research groups to foster their visibility online.

Keywords – corpus design; digital discourse; research project websites; *Twitter*; e-visibility; Computer-Mediated Communication

1. INTRODUCTION¹

Professional practices are increasingly influenced by digital communication. This is no exception for scholars who need to deploy digital discursive practices, especially when it comes to disseminating the results of their research. It is not only necessary for

¹ The compilation of the corpus has been carried out within the framework of the project *Towards Greater Visibility and Dissemination of Scientific Research: A Linguistic, Rhetorical and Pragmatic Study of Digital Genres in English as a Language of International Communication* (grant number FFI2017-84205-P). We would like to thank Dr Enrique Lafuente-Millán for coordinating different stages during the process of compilation. We would also like to express our gratitude to our research colleagues in the *InterGedi (Interpersonality in Digital Genres)* research group for their collaboration in this endeavour. Finally, we are also grateful to Dr Rosana Villares Maldonado, a research fellow from our research group *CIRES* (Gobierno de Aragón – H16_17R), for providing technical support in the compilation of the corpus.



academics to produce primary output, which certifies and legitimises new knowledge (Puschmann 2015: 31), but also to disseminate it broadly, which is frequently done online and in English. With the aim of undertaking lexico-grammatical, pragmatic, discursive and genre analyses of digital texts in the international research project websites, the *InterGedi* research group² compiled a database of 100 research project websites funded by the *European Horizon2020 Programme*, henceforth H2020. Out of these websites, a corpus of 30 webs fulfilling the criteria described below was compiled (see Section 2.1), and the texts were downloaded and tagged. This digital corpus was named *EUOPRO* and consists of two collections: *EUOPROwebs* which includes the texts downloaded from the 30 research project websites that were selected, and *EUOPROtweets* which includes the tweets from the 20 projects which had a *Twitter* account.

In order to compile the *EUOPRO* digital corpus, the World Wide Web was used ‘for’ a corpus rather than ‘as’ a corpus (Fletcher 2013), as a careful selection of webs and their texts was made to compile our own corpus. This is an effective alternative to foster fine-grained analyses but requires solving the challenges of the digital environment before the texts can be processed and worked on, as will be discussed presently (see Section 2.2). *EUOPRO* can also be described as a specialised and static corpus (Gries and Newman 2013: 259), since it delves into the context of international research communication by compiling, at a given point, digital instances of researchers’ discursive practices for the dissemination of their projects. The analysis of this type of discourse through specialised and ready-made corpora, as is the case in the *EUOPRO* digital corpus, is advantageous because such corpora tend to be of a manageable size. Compiling a specialised ready-made corpus allows for more qualitative analyses and may help overcome the de-contextualisation of texts and of the particular discursive and linguistic features that are analysed. Such a de-contextualisation of texts and their features is a frequently criticised methodological aspect in corpus-based studies. In this sense, the present paper intends to provide metadata which may contribute to contextualising the digital material included in the corpus, thus facilitating its use and its application.

² For information on the research group, see <http://intergedi.unizar.es/>

The *EUROPRO* digital corpus is also static in that the texts were compiled at a specific moment and not modified or extended, regardless of their evolution in the digital sources where they are hosted. As a result, and to deal with the organic nature of websites and *Twitter* accounts, our corpus captures texts as published final products which constitute the basis for our analyses. The compilation of the corpus over time would have been complex and so would have been its use, since multiple, on-going versions of the texts would have been available. Instead, we decided *EUROPRO* to remain static and to retrieve enough contextual information, as to cater for the process of text crafting and publication.

While it is true that texts online can be easily accessed and saved, important challenges and decisions need to be made when compiling corpora emerging from ‘Computer-Mediated Communication’, since these are distinct from those conformed by off-line texts or speech genres (Collins 2019). Some of these challenges are related to the selection of the specific sites and the texts in the corpora to ensure representativeness and to the coding of contextual information, which is of great relevance, especially in the website. The *EUROPRO* digital corpus contains textual documents in a reduced form consisting of character strings (Beißwenger and Storrer 2008: 297), but including a prominent layout and structure as well as multimodal elements (see Section 2), in the understanding that the combination of different modes—verbal, visual and audiovisual— makes meaning as a multimodal ensemble (Kress and van Leeuwen 2001; Jewitt 2016).

The aim of this paper is to describe the selection and nature of the texts compiled in the *EUROPRO* digital corpus and to justify the criteria followed in its compilation as regards size, balance, representativeness and topic (Sinclair 2005). Section 2 provides the description of the corpus. Section 2.1 discusses the criteria followed for its compilation while Section 2.2 outlines the methodological aspects considered when compiling, downloading and storing such texts in the belief that texts need to be gathered according to explicit design criteria (Tognini-Bonelli 2001: 2). Section 2.3 offers information on the process of labelling and tagging the *EUROPRO* digital corpus and Section 2.4 describes some important contextual factors of the corpus. Finally, Section 3 deals with the uses and applications of the corpus and Section 4 provides some concluding remarks.

2. DESCRIPTION OF THE CORPUS

The websites and *Twitter* accounts selected for the *EUROPRO* digital corpus emerge as part of the communication and dissemination plans included in the *European Horizon2020 Framework Programme for Research and Innovation*. Thus, they are taken to be instances of current digital scientific writing practices which, among others, serve the purpose of accounting for the adequate investment of public expenditure.

From a technical viewpoint, research project websites, as all websites, contain fluid texts which are featured by hypermediality and divided into webpages. These webpages play the role of dynamic nodes and host verbal and audiovisual content that users can navigate through hyperlinks (Djonov 2007: 145). In the case of the research project websites, they render visible a hybrid nature interweaving traditional and digital sources in their design and content organisation. Consequently, the overall function of the genres and texts housed in the research project websites sits “uneasily somewhere between a commercial, technical description of the product and a more formal report on facts” (Stein 2006: 5). These websites serve as repositories of the activities and productivity of the project, as transmitters of the current values of scientific research and as venues to strategically engage with interested users and make the research available to a broad audience (Lorés 2020: 1).

The tweets gathered in the *EUROPROtweets* collection are intended to illustrate the use that research groups make of social networks for the dissemination of their projects, assuming that “interactions in social media contexts may enable self-promotion strategies that result in social or economic gain” (Page 2012: 182). The dynamicity, immediacy and addressivity of social media such as *Twitter* enable research groups to develop a distinct kind of communication about the projects. In *Twitter*, they report on scientific progress and also devote space to daily issues, related topics and social and professional bonds. Hence, tweets are maximised to mediate everyday routines of professional research work and connect users, collaborators and beneficiaries at scholarly events (Kuteeva 2016: 440). In all, *Twitter* users may enact an ‘ambient identity’ to address the mass online audience and to construe an experience of semiotic belonging to different groups (Zappavigna 2014: 2–3). Such an identity originates in users’ discourse choices, in the values conveyed through them and in their exploitation of the affordances of this medium.

2.1. Criteria for the compilation of the EUROPRO digital corpus

The *EUROPRO* digital corpus emerges from a database of 100 websites of H2020 projects. It comprises two collections of texts: *EUROPROwebs* and *EUROPROtweets*. The former contains 30 research project websites funded under the H2020 programme with a word count of 394,072 and an average of 13,136 words per website. We here followed Biber *et al.* (1998: 243), who point out that “representativeness refers to the extent to which a sample includes the full range of variability in a population.” Therefore, we attempted to collect a specialised corpus which included samples of websites, within the context of the H2020 programme, and would allow to draw reliable insights on aspects of genre theory, metadiscourse, pragmatic strategies and multimodality.

Several criteria were established on the selection of the 30 websites for the corpus to be “a good sampling” (Koester 2010: 69). First, the research projects whose websites were selected had to aim at knowledge creation and dissemination in their respective disciplinary fields and not at training PhD students or professionals. Second, we followed a convenience sampling method which entailed choosing research projects with at least one member from the University of Zaragoza (Spain) or a research institution based in Zaragoza. This allowed to complement our text-based analysis with valuable contextual evidence from potential informants, as discussed in Section 3. Third, at some point, the date of the projects should coincide with the development of our own research project (2018–2021), so that the most recent digital academic practices could be studied.

Given the importance of social media in general —and *Twitter* in particular— for dissemination purposes, *EUROPROtweets* was compiled as an extension of *EUROPROwebs*. This collection of *Twitter* accounts of 20 research projects consists of 4,219 tweets containing 88,970 words, with an average of 211 tweets and 4,449 words per account. The use of social media is indeed highlighted by the communication plans endorsed by Horizon2020 and is generally adopted by research groups. The choice of *Twitter* as the object of study was also based on an observational analysis of the range of social networks maintained by the H2020 research projects within the representative sample, as described in Table 1 below. Out of the 30 research projects, only eight did not make use of any social networks to disseminate their research results. 20 research projects made use of *Twitter*, which was the most frequent social network, 13 made use

of *LinkedIn*, ten made use of *Facebook* and five made use of *YouTube*. Interestingly, it was also observed that over 50% of the webs were linked to, at least, two social networks, as shown in Table 1.

NUMBER OF SOCIAL NETWORKS	NUMBER OF H2020 PROJECTS	PERCENTAGE OF H2020 PROJECTS
0	8	26.6%
1	6	20%
2	8	26.6%
3	6	20%
4	2	6.6 %

Table 1: Range of social networks maintained in *EUROPROTweets* by the research projects of H2020

2.2. Methodological decisions in the compilation of the EUROPRO digital corpus

Once the selection of websites was determined, several methodological decisions were made for downloading and storing the texts. All texts from the websites were downloaded and labelled using different codes which referred to the pages or sections of the website where such texts were housed, as illustrated in Table 2. We excluded those texts housed in the websites as external downloadable documents, mainly deliverables in PDF format, as they were not considered to share the same digital nature and purpose as those texts generated for the website.

DIFFERENT LABELS USED FOR THE SECTIONS	CODE FOR MENU SECTIONS
Home; Homepage	HOME
About; Objectives; Project; Summary	ABOUT
Partners; Researchers; Consortium; Related projects	PARTNERS
Work packages; Actions; Demos	WORK
News; Events; News and events; Blog	NEWS
Outreach; Publications; Reports; Deliverables; Repository	OUTPUT

Table 2: Codification for salient menu sections and range of headings of sections in *EUOPROwebs*

After that, information related to the extent to which the text could be directly accessed from the website menu was also recorded. In cases where the menu sections were included in the codes provided above, no additional information was required. However, if the section showed up in an unfolding menu of options, the code SUB (subordinate) was noted. Likewise, if the section was included in the website but had no label in the menu, the code EMB (embedded) was added. In this way, a representation of the options of the menu section could give us insights into the preferred position and relevance of some sections or pages throughout the websites.

All texts were downloaded in May 2019 because of the dynamic nature of the research project websites. At that time, the projects in the sample had been developed to different extents just like their websites. For this reason, it was key to record information about the start and end date of the projects, as well as information about their degree of development when the texts were compiled. This information may be of great importance when discussing and interpreting the data retrieved from the analysis. Compiling a ‘Monitor Corpus’ (McEnery and Wilson 2001), which would need to be updated regularly, was disregarded as a feasible objective in our own research project. Such a compilation would entail constantly comparing updated versions of the websites and tracing them at different points in time throughout the duration of the project which, adding to being extremely time-consuming, would not have been relevant for the purposes of our research.

The ‘hypermodal’ and ‘hypermedial’ nature (Petroni 2014) of the websites led to the tagging of (external, internal and peripheral) hyperlinks,³ of visuals, such as tables, figures, pictures or logos, and of videos and audios. Thus, although the focus of the analyses, as well as the corpus, is mainly grounded on the verbal component, these multimodal and multimedial elements were not overlooked in websites, since they are affordances that combine with the verbal component as meaning-making devices (Kress and van Leeuwen 2001; Jewitt 2016). Similarly, because of the importance of layout and web design aspects, we stored screenshots for every page in the website as part of

³ Internal hyperlinks specifically link to other sites within the project website while external hyperlinks lead to external sources of information. Peripheral hyperlinks refer to project-related pages and downloadable documents which are located outside the project website.

the corpus, since this would allow us to go back to them when analysing verbal features in the research project websites.

The methodological decisions to compile *EUROPROTweets* were similar to those taken in the compilation of *EUOPROwebs*. Here, given the dynamic nature of *Twitter* accounts, all tweets and retweets were downloaded in June 2019. Since these platforms are also hypermodal and hypermedial (Petroni 2014), tweets were coded and tagged for hyperlinks and multimodal elements such as pictures, videos or GIFs. One key feature of *Twitter* accounts is their potential interactivity. As a result, we retrieved and saved information about (1) the number of likes in each tweet at the specific date, (2) the number of retweets by other users and (3) the number and types of hashtags (#) used in the tweets by the research group and their mentions (@) to other *Twitter* accounts. This information should be taken into account when carrying out textual analyses based on this corpus, as there may be a correlation between discursive choices and their likely dynamic nature.

2.3. Labelling and tagging of the EUROPRO digital corpus

The verbal component of the 30 websites of H2020 which conform *EUOPROwebs* was downloaded and saved into TXT format documents, both as a document corresponding to the whole content of the website (labelled 'NAMEOFTHEPROJECT') and as documents corresponding to common web sections or pages which were labelled with the codes pointed out in Table 2. The tagging of *EUOPROwebs* was performed manually. First, a number of general tags was determined in the light of the texts downloaded and saved, namely <hyperlink>, <image>, <video>, <table>, <graph>, <map>, <presentation>, <questionnaire> and <language>. The whole corpus was then annotated using these tags, as illustrated in Figure 1 below. Specific codes — metadiscursive, pragmatic, ethnographic or multimodal— were further applied to the different analyses undertaken by the members of the *InterGedi* research group.

Projects

[LEARN MORE ABOUT OUR PROJECTS](#)



<image>

The MIGRATE Members

The Transmission System Operator (TSO) TenneT is the project coordinator of MIGRATE. Together with 23 other partners (10 TSOs, 12 Universities/Labs and 1 manufacturer) of the consortium the research and investigation will be done.

[LEARN MORE ABOUT THE MIGRATE MEMBERS](#)



<image>

News and Newsletter

Here you will find recent news about MIGRATE and the possibility to register for our newsletter.

[VIEW MORE](#)

Figure 1: Example of manual tagging of *EUROPROWebs*

Manual annotation was leveraged over XML language in the belief that interpretation and evaluation play a significant role in this process at a syntactic, semantic, discursive and pragmatic level (Collins 2019) in the specialised texts that make up the *EUROPRO* digital corpus. Since ‘big data’ tends to be advocated for in the use of XML—and since the corpus does not contain large collections of dissimilar texts—a consistent coding and tagging carried out by the *InterGedi* members was preferred. This system facilitated keeping in mind the design and layout of the sites, pages and texts under analysis at all times and in a clear visual way.

A similar procedure was followed in the compilation of *EUROPROTweets*. The verbal component of the tweets was downloaded and saved into a TXT document corresponding to each account and labelled ‘NAMEOFTHEPROJECT_T’. The tagging of *EUROPROTweets* was also carried out manually. As Figure 2 illustrates, several tags were determined in the light of the texts downloaded and saved: <link>, <hashtag>, <image>, <video> and <language>.

Hammerschmid, <mention>@kaiwegrich1</mention> <mention>@EU_H2020 project</mention> “TROPICO” to research <hashtag>#egovernance</hashtag>, <hashtag>#digitalisation</hashtag> in public sector. <link>https://t.co/fjx6F94B68</link><ext> <link>https://t.co/Yy5kEilsbL</link><ext>

Do formal rules limit or foster <hashtag>#collaboration</hashtag> in and by governments? Insights on the alleged “maze of rules” and <hashtag>#DigitalTransformation</hashtag> in public sectors from 10 European countries available now: <link>https://t.co/6XR4oRGcn3</link><int>

Figure 2: Example of manual tagging of *EUROPROTweets*

2.4. Contextual factors of the *EUROPRO* digital corpus

The *EUROPRO* digital corpus was designed to undertake generic, discursive, pragmatic and multimodal analyses of digital texts. In this endeavour, contextual information about the texts to be analysed was essential. For this reason, special efforts were made to obtain and code specific details about the H2020 programme in general, and about each of the research projects selected to compile the *EUROPROWebs* corpus in particular. Table 3 provides an example of the contextual information recorded in the first three websites of our data set. This information mainly includes: (1) the name of the research project, (2) the link to the *CORDIS*⁴ web where the details of the project can be accessed, (3) the start and end date of the project and (4) information about the researcher affiliated to the University of Zaragoza or to the research institution based in Zaragoza.

RESEARCH PROJECT	LINK TO H2020 CORDIS	LINK TO WEBSITE	PROJECT START AND END DATE	DATE OF WEBSITE DOWNLOAD	CONTACT AND E-MAIL
ADREM	cordis.europa.eu/project/id/680777	spire2030.eu/adrem	01/10/2015-30/09/2019	07/05/2019	xxx
AGROinLOG	cordis.europa.eu/project/id/727961	agroinlog-h2020.eu/en/home/	01/11/2016-30/04/2020	13/05/2019	xxx
AIDA-2020	cordis.europa.eu/project/id/654168	aida2020.web.cern.ch/	01/05/2015-30/04/2019	26/05/2019	xxx

Table 3: Example of contextual information recorded about some of the projects in the *EUROPRO* corpus

⁴ *CORDIS* is a platform that gathers information on EU-funded Projects of Research and Development activities. It is the primary source for the consultation of updates about the results and publications of projects participating in different European programs, among with *HorizonH2020* is included. Access is available at <https://cordis.europa.eu/>.

3. CORPUS APPLICATIONS FOR THE ANALYSIS OF DIGITAL ACADEMIC DISCOURSE

The *EUROPRO* digital corpus is a corpus “that may serve to support empirical research on linguistic aspects of Computer-Mediated Communication discourse” (Beißwenger and Storrer 2008: 293), more specifically on digital academic communication within the context of international research projects. Consequently, its analysis can be undertaken from a range of linguistic theoretical frameworks and analytical perspectives, such as Computer-Mediated Communication, corpus linguistics, discourse analysis, ethnography studies, pragmatics, metadiscourse and multimodality. All of these can help to understand how research groups communicate their results online and allow to delve into other associated concepts such as knowledge dissemination, project accountability and e-visibility.

Studies of a contrastive nature can be carried out using the *EUROPRO* digital corpus for the purpose of looking for features that characterise digital research communication. Such features would be necessarily determined by both the technological level, which deals with the type and degree of exploitation of web-mediated affordances, and the linguistic or discursive level, which focuses on the ways language is employed. Two main directions are reckoned to be particularly rewarding when undertaking contrastive analyses around *EUROPRO*. First, a comparative study of *EUROPROwebs* and *EUROPROtweets* involves contrasting the discourse attested in the websites of the projects with that of *Twitter* accounts held by research groups. Such a study would cast light on the similarities and differences in the use of discursive resources made in both digital platforms when it comes to disseminating and promoting the project and the investigation that is being carried out. Second, findings from the analysis of *EUROPRO* can be contrasted with those from studies on other digital modes and media for research dissemination purposes, such as research reports and research group blogs.

A closer look at the *EUROPROwebs* collection can also be taken at a more rhetorical level by carrying out analyses that focus on the exploration and comparison of sections and/or pages. Here, studies of move analysis (Swales 1990, 2004), which have proved to be insightful for offline academic texts, may help explain potential structural, textual and discursive choices in the different webpages within the research project website, such as ‘About’, ‘Partners’, ‘Work Packages’, ‘Output’ or ‘News and Events’. In turn, move analysis may also allow the identification of prototypical patterns that

stand out throughout the corpus and help to generalise researchers' choices in these discursive practices. Furthermore, the relationship between tweets and webs can also be explored from the perspective of genre studies by identifying the genre relations that can be established among texts in both platforms in the light of concepts such as 'generic integrity', 'genre colonies' or 'genre constellations' (Bhatia 2004).

Pragmatic analyses of both *EUROPROWebs* and *EUROPROTweets* may foreground the identification and reasoning of researchers' intents when communicating their projects digitally. The study of the pragmatic mechanisms and resources exploited to disseminate the research that is undertaken can be approached from various pragmatic theories such as Speech Act Theory (Austin 1965; Searle 1969), Relevance Theory (Sperber and Wilson 1995) or Politeness Theory (Brown and Levinson 1987). Pragmatically speaking, the comparison of the aforementioned aspects in the research project websites, as opposed to social networks, such as *Twitter*, may lead to discover researchers' dissimilar intents and strategies that depend on the digital environment employed.

Moreover, the discursive analysis of the *EUROPRO* digital corpus may contribute to exploring the actual use of the language in such a digital scenario. The spectrum of linguistic items deployed in the communication of the project could be accounted for through different types of studies, for instance, at the lexico-grammatical level or at the level of metadiscourse. Additionally, corpus-assisted analyses (e.g. frequency, keywords, collocation and cluster analyses) could make a significant contribution at unveiling meaningful patterns by offering quantitative data. The findings at the discursive level would surely help establish connections with analyses at the rhetorical and pragmatic levels, ranging from a rather abstract and implicit level to the linguistic components that are used.

These analyses of the *EUROPRO* digital corpus —at the rhetorical, pragmatic and discourse levels— should ideally be combined with ethnographically-informed qualitative data. In other words, contextual information gathered from informants in ethnographic analyses can complement and expand the textual results. Thanks to the sampling method followed in the compilation of *EUROPRO* (see Section 2.1), this sort of evidence is at hand. Hence, the role of researchers and their attitude towards the digital communication used in their research projects can be unravelled. This would

allow for a better explanation on how international funded research projects develop their research and make their results visible online.

Finally, multimodal analyses are also necessary because there is a need to understand how the combination of languages, modes and media works in those texts hosted in the research project websites and *Twitter* accounts. This may contribute to unveiling the discursive and pragmatic functions that elements such as images, videos, interactive visuals, hyperlinks and other technical affordances perform in digital communication in general, and in Computer-Mediated scientific Communication in particular.

4. SUMMARY AND CONCLUSIONS

As has been pointed out, the *EUROPRO* digital corpus has been compiled to cater for the need to analyse scholarly discourse and scientific communication, as adapted to the digital environment. To narrow down the scope of the various possibilities offered by the Internet to develop such discursive practices, two digital platforms were specifically chosen, namely research project websites and *Twitter* accounts. Accordingly, two collections of texts make up the *EUROPRO* digital corpus: *EUROPROwebs* (texts downloaded from research project websites) and *EUROPROtweets* (texts retrieved from *Twitter* accounts of research projects). Thus, *EUROPRO* will allow to explore different web-mediated affordances and user-dependent linguistic decisions when disseminating research online.

In this paper, the emphasis has been placed on the criteria used to compile the *EUROPRO* digital corpus, which ensure the identification of current digital practices through the exploration of the texts, and the necessary contextual information around them for the analyses to be carried out. Moreover, methodological explanations have been offered to determine how digital texts of an inherent dynamic, fluid, hypermodal and hypermedial nature have been dealt with when compiling a static corpus, reflecting on potential hurdles posed by technical and structural features of the sites where the texts are hosted. Given the scarce number of specialised digital corpora, we believe that the decisions we have made in the compilation of the corpus can help others in the compilation of future corpora.

Furthermore, the fact that the specialised corpus we have compiled may serve a wide range of applications encourages us to make it publicly available in the near future, once the objectives in our current national research project have been achieved. We believe that this corpus may be of use to scholars interested in websites as a digital environment for exploration in genre studies. It will also be of interest to academics who conceive websites and social networks as spaces of engagement for scientific communication and interaction.

Finally, potential uses of the *EUROPRO* digital corpus have also been outlined in the paper. These uses comprise discourse and pragmatic studies, contrastive analyses between texts on websites and on *Twitter* or across website sections, as well as analyses that would complement the textual evidence, either from a multimodal perspective or from ethnographically-collected data. The compilation of *EUROPRO* is the first step to carry out analyses that will allow to gain insights into new, changing, digital discursive and professional practices of researchers nowadays.

REFERENCES

- Austin, John L. 1965. *How to Do Things with Words*. Oxford: Oxford University Press.
- Beißwenger, Michael and Angelika Storrer. 2008. Corpora of computer-mediated communication. In Anke Lüdeling and Merja Kytö eds. *Corpus Linguistics: An International Handbook*. Berlin: Mouton de Gruyter, 292–308.
- Bhatia, Vijay K. 2004. *Worlds of Written Discourse: A Genre-based View*. London: Continuum.
- Biber, Douglas, Susan Conrad and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Brown, Penelope and Stephen Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Collins, Luke Curtis. 2019. *Corpus Linguistics for Online Communication: A Guide for Research*. London: Routledge.
- Djonov, Emilia. 2007. Website hierarchy and the interaction between content organization, webpage and navigation design. *Information Design Journal* 15/2: 144–162.
- Fletcher, William H. 2013. Corpus analysis of the World Wide Web. In Carol Chapelle ed. *Encyclopedia of Applied Linguistics: Volume 3*. New Jersey: Wiley-Blackwell, 1339–1347.
- Gries, Stefan T. and John Newman. 2013. Creating and using corpora. In Robert J. Podesva and Devyani Sharma eds. *Research Methods in Linguistics*. Cambridge: Cambridge University Press, 257–287.
- Jewitt, Carey. 2016. Multimodal analysis. In Alexandra Georgakopoulou and Tereza Spilioti eds. *Routledge Handbook of Language and Digital Communication*. London: Routledge, 69–84.

- Koester, Almut. 2010. Building small specialised corpora. In Anee O’Keeffe and Michael McCarthy eds. *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 66–79.
- Kress, Gunther and Theo van Leeuwen. 2001. *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London: Arnold.
- Kuteeva, Maria. 2016. Research blogs, wikis and tweets. In Ken Hyland and Philip Shaw eds. *The Routledge Handbook of English for Academic Purposes*. London: Routledge, 431–444.
- Lorés, Rosa. 2020. Science on the web: The exploration of research websites of energy-related projects as digital genres for the promotion of values. *Discourse, Context and Media* 35: 1–10.
- McEnery, Tony and Andrew Wilson. 2001. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Page, Ruth. 2012. The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. *Discourse & Communication* 6/2: 181–201.
- Petroni, Sandra. 2014. Collaborative writing and linking: When technology interacts with genres in meaning construction. In Paola E. Allori, John Bateman and Vijay K. Bhatia eds. *Evolution in Genre: Emergence, Variation, Multimodality*. Bern: Peter Lang, 289–306.
- Puschmann, Cornelius. 2015. A digital mob in the ivory tower? Context collapse in scholarly communication online. In Marina Bondi, Silvia Cacchiani and Davide Mazzi eds. *Discourse in and through the Media: Recontextualizing and Reconceptualizing Expert Discourse*. Newcastle upon Tyne: Cambridge Scholars Publishing, 22–45.
- Searle, John Rogers. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Sinclair, John. 2005. Corpus and text – Basic principles. In Martin Wynne ed. *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 1–16.
- Sperber, Dan and Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Stein, Dieter. 2006. The web as a domain-specific genre. *Language@Internet* 3: <https://www.languageatinternet.org/articles/2006/374> (10 May, 2020.)
- Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Swales, John. 2004. *Research Genres: Explorations and Applications*. Cambridge: Cambridge University Press.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Zappavigna, Michele. 2014. Enacting identity in microblogging through ambient affiliation. *Discourse & Communication* 8/2: 209–228

Corresponding author

Daniel Pascual

University of Zaragoza

Department of English and German Studies

Calle San Juan Bosco, 7

50009. Zaragoza

Spain

e-mail: dpascual@unizar.es

received: January 2020

accepted: July 2020