

"DUE TO HEIGHTENED RISK" Qualifying Risk in the Debate over Covid-19

MARINA BONDI¹, SILVIA CAVALIERI²

¹UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

²UNIVERSITÀ DI VERONA

Abstract – Our paper presents an analysis of the conceptualisation and framing of ‘risk’ in the Covid-19 pandemic. We investigate all occurrences of risk-based vocabulary in two corpora dealing with the spread and the political/social/economic consequences of Covid19 from February 2020 to February 2021 (i.e. the InterDiplo Covid-19 corpus and the CORD-19 corpus). We apply a mixed-method approach to show how the concept of ‘risk’ is lexically qualified in the public communication of a hot issue such as the pandemic. Therefore, after a first recognition of all related expressions of the nominal realization of “risk” in the worldlists of the two corpora, we investigate their concordances to find typical collocations and phraseology used to qualify risk in Covid-19 epidemic. Our results show a clear tendency towards a negative thematisation of “risk” which is characterized by patterns highlighting uncertainty and fear in the fight against a disease spreading on an unprecedented scale and posing total threats that not only involve individuals but affect humanity globally. The focus on the phraseology around the lexical elements helps reconstruct communicative functions and approaches that characterize the two discursive contexts.

Keywords: risk communication; corpus linguistics; phraseology; collocations; Covid-19.

1. Introduction¹

The Covid-19 pandemic has changed the world as we knew it. Our lives have been completely overturned as for what concerns our habits and social relations; we now live in a constantly evolving scenario and the way we live, relate, and communicate with others has been transformed permanently. In this context, risk communication has been a key element worldwide during the pandemic, with a view to maintaining an accessible and transparent relationship in the different flows and transmissions of information and

¹ The article has been jointly planned by the two authors: Marina Bondi has dealt with sections 1, 2, 4.2.2 and the parts of the other sections dealing with the COVID-19 Open Research Dataset (CORD-19), while Silvia Cavalieri with sections 3, 5, 4.2.1 and the parts of the other sections dealing with the InterDiplo Covid-19 corpus.

communication, which is the premise of collective action (de Las Heras-Pedrosa et al. 2020: 2). In such unstable and precarious times, events are changing rapidly, and information is in constant evolution. As a consequence, risk communication has played a crucial role when conveying information in society. Covid-19 has determined a real need for risk communication management in terms of impact, preparedness, response, and mitigation by governments, health organizations, non-governmental organizations (NGOs), mass media, and stakeholders (Zhang et al. 2020: 1). At the same time, with the urgency of the need for effective diagnostic and therapeutic solutions, scientific publishing has soared dramatically (Hyland and Jiang 2021) and the discourse of research has gained great prominence, especially in relation to open access issues.

Risk communication, however, needs to adapt according to the purposes, the receivers and the channels used and may involve different linguistic strategies to convey the message. In the field of health communication, a number of studies have dealt with the role of risk communication in the media, social media and governmental releases (see among others De Paula et al. 2022; Bernard et al. 2021; Dyer and Kolic 2020) but, to the best of our knowledge, little has been written specifically on the linguistic realizations of risk communication during Covid-19 pandemic. Müller et al. (2021) offer an interesting overview of expressions of uncertainty in the discourse of the press, but no studies have considered the phraseological profile of risk-related notions comparing scientific language and the discourse of international relations about Covid-19.

With this objective in mind, we take a discourse-analytical perspective on Covid-19 discourse focusing on the following research questions: 1) what role has risk-related vocabulary played in public and scientific communication at the outbreak of the most dangerous pandemic of the last centuries? 2) more specifically: how has the vocabulary of risk thematization been contextualised and used in international relations and scientific language? 3) Does the local grammar of risk-related terms change according to the different purposes of the two discourses analyzed?

The paper is structured as follows: Section 2 is devoted to the review of the literature of discourse analytical studies on the notion of risk and risk communication. Section 3 introduces the corpora for the analysis and the methodological framework adopted in this work. In Section 4 results from the two corpora are described starting with a quantitative overview of the risk-related terms and then moving on to a more qualitative perspective on their local grammar and phraseology. The conclusions summarize the comparative results.

2. 'Risk': a view from discourse analysis

A focus on the language of risk communication should be set more generally within the framework of a growing body of research on the discourse of risk communication. This is central to engaging with the public and promoting acceptance, compliance and policy support in ways that are shaped by institutional and sociocultural conditions (Rothstein et al. 2022), especially in matters of health communication (WHO 2017). Discourse analytic approaches have rapidly consolidated within this framework by looking at both general issues and specific language tools. Candlin et al. (2016: 5), for example, following Giddens (1998), remind us that communicating risk is not only associated with hazards, but also with positive projections and highlight the close link between risk communication and issues of power, categorization, distribution, regulation, negotiation, and mediation.

When focusing on the language of risk, on the other hand, an important starting point is provided by Fillmore and Atkins' (1992) seminal study on *risk* (verb and noun) as an illustration of frame semantics. The authors analysed the semantics of *risk* in terms of a "valence description", combining semantics and syntax into a representation of the conceptual structures associated to words (ibid.: 78), mapping formulation categories onto conceptual categories. Within the "risk frame", *risk* is considered as a polysemous word, taking into account both its positive meaning of 'chance' and its negative one related to 'harm'. Based on their analysis of a general corpus, the authors created different categories which were functional for the valence description of *risk*. These include Chance, which refers to the uncertainty of the future, Harm, which is potential unwelcome behaviour, Victim, the individual that suffers if Harm occurs, Valued Object, which is a valued possession of the victim seen as endangered, Risky Situation, a state of affairs where someone is at risk, Deed, the act that brings about a risky situation, and the Actor, which is the person that performs the Deed (ibid.: 81-83). They also included subcategories for the description of the Actors' intentions, such as Intended Gain (the Actors' hope-for gain in taking risk), Purpose (what an actor intends in performing the Deed), Beneficiary (the person for whose benefit something is done), and Motivation (the psychological source for someone's behaviour) (ibid.: 84-84).

In a corpus-based framework, work on risk-related lexis has often concentrated on collocations and positive or negative semantic prosody. Hamilton et al. (2007) have carried out a corpus-based study of *risk* as noun and verb across different domains using the Collins Wordbanks Online and the conversational Nottingham Corpus of Discourse in English (CANCODE). Results have confirmed the negative semantic prosody; they have also shown that in both corpora *risk* is used more often in the context of health and illness

than in that of finance (2007: 169). Moreover, the connection to the medical semantic field is highly dependent on collocates associated with health and illness (*ibid.*, 178. Hamilton et al.'s (2007) work on how the semantics of risk is dependent on the specific context of use was further developed by Hardy and Colombini (2011) focusing on the collocations, semantic associations, and semantic prosody of the noun *risk* in the COCA corpus, a balanced sample of texts in terms of domains (Hamilton et al. 2007: 170-171). Outcomes from Hardy and Colombini's analysis confirmed the negative connotations of *risk* which are usually associated with medical discourse, mainly found in magazine and academic genres, even in apparently positive contexts such as *good risk*, *risk worth taking* and *x be worth risk*.

Others have looked at the semantics of risk and related concepts, such as causality (Boholm 2009), danger or threat (Battistelli and Galantino 2019, Boholm 2012), security and safety (Boholm et al. 2016) or uncertainty (Müller et al 2021). Boholm et al. (2016: 4) have also argued that risk associations are established by linguistic practices in a “layered model of risk association”, revolving around the word *risk* (2016, 4). Boholm's (2016) “onion model” is presented as a framework to methodologically analyse the construction of *risk* in discourse (2016: 8) across four layers: the core and central layer including the noun *risk* and its compounds, the first layer including the morpheme *risk* and its derived words (e.g. verb *risk* and adjective *risky*), the second one gathering close synonyms of *risk* (e.g. *danger*, *hazard* and *peril*) and their derived forms (e.g. *dangerous*), the third one collecting *risk*'s antonyms (e.g. *safety* and *security*) with their derived forms and backformations (e.g. *safe*), while the fourth layer comprising other related words.

In a diachronic perspective, Zinn (2010) and Zinn and MacDonald (2018) analyze the diachronic change of the semantics of *risk* and related items over the news coverage of the *New York Times*, while claiming the need for corpus-assisted discourse analyses in research in order to carry out more empirical analyses. They notice for example that health-related risks actually became dominant only after the second world war. Müller and Mell (2022) also explore diachronically quantitative and semantic changes of the concept of ‘risk’ in the parliamentary discourse of the German Federal Republic from 1949 to 2017, underlining new emerging lexical paradigms.

Lexico-semantic choices in the area of the notion of ‘risk’ vary according to the specific contexts. If the discourse of *risk* seems to be key in the domain of health and illness (Zinn and MacDonald 2018), the discourse of *threat* often becomes dominant in the domain of war, terrorism and migration crises (Galantino 2020). The notion of *danger* also becomes central in the discursive representation of possible harm (Battistelli and Galantino 2019). Starting from Luhman's (1993: 23) distinction between the attribution

of negative outcomes to one’s decision (as with *risk*) or to external sources (as with *danger*), the authors propose a representation of the semantic space of the three expressions based on the notions of agency and intentionality. The assumption is that the uncertainty about future harm ranges from the possibility of negative outcomes that can still be attributable to *positive* intentionality (*risk*) to the harm being produced by ill-intentioned actors (*threat*) or by negative circumstances (*danger*) (Battistelli and Galantino 2019, 70), as represented in Figure 1.

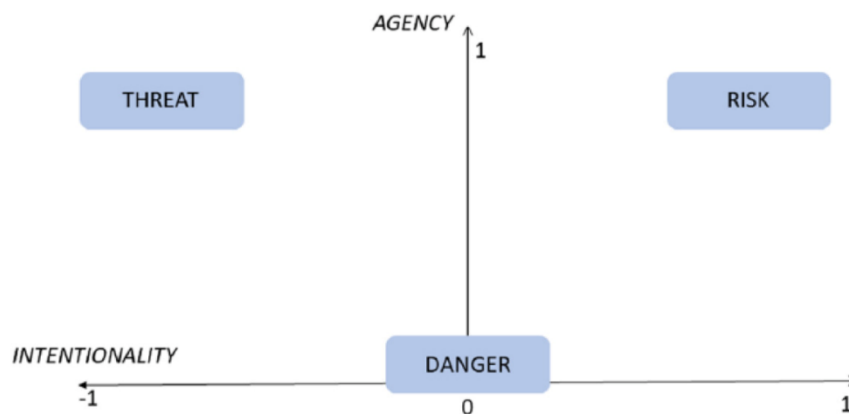


Figure 1

A typology of danger, risk and threat. (Battistelli and Galantino 2019: 70)

A growing body of corpus-based studies have contributed to developing risk research with specific reference to the pandemic. Semino (2021: 54), for example, has investigated different metaphors adopted in the communication of the pandemic, focusing on the Fire metaphor, and emphasizing how it conveys danger and urgency, the risk of contagion, the role of health workers or the connection with health inequalities. Wicke and Bolognesi (2021) have looked at the construction of the WAR frame in Covid-19 tweets, including the word *threat* in the building of their WAR frame. In a corpus-based perspective, Müller et al (2021) focus on an annotation scheme examining elements of personal and situational uncertainty in relation to ‘risk’ in British and German newspapers. Müller et al (2021) map different markers of uncertainty in a corpus of British and German newspaper articles and highlight different trends in the two contexts, showing for example that while markers of possibility and probability are similarly frequent, uncertainty is more often realized through reference to anxiety in the British context and to disagreement in the German context. Lexical markers like *risk*, *danger*, *threat* are actually classified as constructions of situational uncertainty (Danger) in their model.

Our own analysis aims at exploring the contextual meanings of different lexical markers of risk in different forms of public discourse: the discourse of international relations and the discourse of research.

3. Methods and materials

3.1. *Materials*

The analysis was carried out on two corpora of texts dealing with Covid 19 pandemic.

The first one consists of texts that were released as part of the COVID-19 Open Research Dataset (CORD-19). The CORD-19 is a free resource of more than 280,000 scholarly articles about the novel coronavirus for use by the global research community, produced by The Semantic Scholar team at the Allen Institute for AI, in partnership with leading research groups. The dataset is updated regularly and contains all COVID-19 and coronavirus-related research (e.g. SARS, MERS, etc.) from a range of sources: PubMed, the WHO COVID-19 database, bioRxiv and medRxiv pre-prints. It is thus taken to be representative of the discourse of scientific research in the most typical form of the research article.

The second is a corpus of interviews in which diplomats and international operators are interviewed on the spread and the political/social/economic consequences of Covid19, i.e. the InterDiplo-Covid 19 corpus². This corpus is part of a larger corpus, i.e. the InterDiplo corpus, currently under construction at the Dep. of Foreign Literatures and Languages - University of Verona³. The interviews were collected from the most famous international broadcasting companies (e.g. BBC, CNN, CGNT, ARIRANG, SKY NEWS UK, FRANCE 24 ENGLISH) or, due to VPN issues, on their YouTube channel where they often publish complete interviews.

Diplomats and international operators are interviewed in English by journalists who do not share the same lingua-cultural background as they can be both native and non-native speakers of English. Videos were saved and then transcribed through Happy Scribe, a transcription software which uses speech recognition technology. It has an accuracy of up to 85% thus txt output files were then manually proofread. After transcription the txt files

² Cavalieri, S., S. Corrizzato & R. Facchinetti (2021). The InterDiplo-Covid 19 Corpus.

³ Project of excellence: “Le Digital Humanities applicate alle lingue e letterature straniere” – Dep. of Foreign Languages and Literatures – University of Verona

were transposed in xml and tagged for metadata, parts of speech, discursive features, questions and answers.

The InterDiplo-Covid 19 corpus includes 80 interviews, and it was collected within a year timespan from February 2020 February 2021. It consists of 236,000 tokens and, to have a balanced corpus, the 80 interviews were grouped into 4 sub-corpora: 1) 20 interviews in which the interviewer and the interviewee are both native speakers of English; 2) 20 interviews in which the interviewer is a native speaker of English whereas the interviewee is a non-native speaker of English; 3) 20 interviews in which the interviewer is a non-native speaker of English whereas the interviewee is a native speaker of English; 4) 20 interviews in which the interviewer and the interviewee are both non-native speakers of English.

3.2. Methods

We adopted a mixed-method approach, i.e., a data-driven quantitative and qualitative analysis (Baker 2006; Müller 2017; Müller and Stegmeier 2019; Müller 2022; Taylor and Marchi 2018) in which corpus linguistics tools are combined with a discourse analytical perspective to investigate the development of *risk* thematization in the two corpora. As for the corpus tools used, we relied on the ones offered by *Sketch Engine*, i.e., the wordlist and concordance analysis.

First of all, we generated a wordlist in the two corpora to identify the most frequent vocabulary nominal realizations associated to the notion of 'risk' and we then extracted its possible related expressions with a normalized frequency higher than 10 pmw. This allowed us to compare the frequency patterns of *risk* and semantically similar nouns in order to map its lexical field in the pandemic. Indeed, as shown by the literature, 'risk' concepts can be expressed not only with the word *risk*, but also by synonyms to emphasize certain aspects of the risk frame.

Risk-related vocabulary was selected according to the definition provided by the Merriam-Webster dictionary online⁴ (e.g., *danger*, *challenge*, *threat*) and also taking into consideration the list of near-synonyms in the thematization of *risk* provided in Müller and Mell (2022).

Secondly, for all the selected items, concordances were analyzed looking for collocational and phrasal patterns (Sinclair 1996, 2004), starting from an analysis of their collocation, colligation and semantic preference. We identified the semantic sequences in which the vocabulary of *risk* is involved, and then we studied elements of local grammar (e.g., patterns of pre- and post-modification) of 'risk' and its related expressions. In this way, we

⁴ <https://www.merriam-webster.com/thesaurus/risk> (Last accessed on December 10, 2021).

determined the phraseological contexts of risk thematisation and could thus identify the most frequent communicative functions of risk in Covid-19 pandemic in a discourse perspective.

4. Results

4.1. Frequency list: risk and semantically related words

The first step of our analysis consisted in the selection of nominal realisations of the notion of risk in the wordlists of the two corpora. To do so, we relied on the definition of *risk* provided by the Merriam-Webster Dictionary online and on the listed risk-related terms:

Risk (noun)

1. *something that may cause injury or harm*
danger, hazard, challenge, imminence, menace, peril, pitfall, threat, trouble
2. *the state of not being protected from injury, harm, or evil*
danger, distress, endangerment, harm's way, imperilment, jeopardy, peril, trouble

A wordlist was then generated, and the focus was placed in particular on the frequency list of nouns to detect the nominal realization of the notion of 'risk'. An overview of the frequency wordlist of nouns in the two corpora offers an interesting picture of the thematization of 'risk'. Table 1 reports rank, frequency and normalized frequency (per million words, pmw) of the top lexical items within the first 10,000 positions.

CORD-19				INTERDIPLO-COVID19			
Rank	Word	Frequency	Pmw	Rank	Word	Frequency	Pmw
182	<i>risk</i>	157,958	562.60	334	<i>risk</i>	77	314.24
398	<i>challenge</i>	76,586	272.78	413	<i>challenge</i>	64	261.18
1,246	<i>threat</i>	23,036	82.05	489	<i>threat</i>	52	212.21
2,397	<i>distress</i>	10,003	35.63	1,018	<i>danger</i>	22	89.78
2,889	<i>hazard</i>	7,593	27.04	-----	-----	-----	-----
5,358	<i>danger</i>	2,937	10.46	-----	-----	-----	-----

Table 1
Wordlist of "risk"-related nouns in the corpora.

As Table 1 shows, the nouns related to the notion of ‘risk’ are semantically connoted by the “reference to the possibility of an unwelcome outcome” (Fillmore and Atkins 1992: 79). All the nominal expressions found suggest that *risk* and semantically related words convey the idea of uncertainty for the future which is defined between the two semantic sub-frames identified by Fillmore and Atkins (1992: 81-84) of *chance* and *harm*.

Interestingly, we can see that in the InterDiplo-Covid 19 corpus the notion of risk seems to develop from a positive valence description to negative ones. On the one hand, the most frequent semantic realization of ‘risk’ is *challenge* (98 occurrences), which can imply a positive experience that usually involves the accomplishment of a difficult task. On the other hand, the other three “risk”-related nouns carry a negative value (a possibility of being exposed to danger or loss) that goes from one that can still be attributable to *positive* intentionality (*risk*, 84 occurrences) to the harm being produced by ill-intentioned actors (*threat*, 52 occurrences) or by negative circumstances (*danger*, 22 occurrences) (Battistelli & Galantino 2019: 70). Quantitatively speaking, this thematic development of the notion of “risk” in the corpora concerning agency and intentionality seems to perfectly fit the typology proposed in Battistelli and Galantino (2019: 70)

In the CORP-19 corpus, on the other hand, the emphasis of scientific discourse lies very much on *risk*, which is by far more frequent than the more positive (or neutral) *challenge*. In terms of intentionality and agency, we notice the same cline, from positive intentionality to negative intentionality to negative agency, but the range of values is much wider: the use of *risk* is much more frequent in scientific discourse (the frequency is almost double), whereas reference to *threat* and especially *danger* is much more limited.

The presence of *distress*, on the other hand, points in a totally different direction, which is not associated to any probabilistic notion of risk. The word is clearly marked in the CORP-19 by its terminological use in classifying pathological states (e.g. *respiratory distress*) and in identifying diagnostic and therapeutic processes: *distress experienced by, characterized by respiratory distress, as respiratory distress progresses, respiratory distress requiring intubation, alleviate distress*, etc. It is thus mostly used to refer to pathological states, rather than to a state of danger or risk. We will therefore exclude it from the qualitative analysis of risk-related terms.

To explore the use of the word *risk* and the other semantically similar nouns identified with reference to this framework, we need to study their contextualization. Thus, in the next section, we will pay close attention to the phraseology involved in the qualification of risk in the corpora.

4.2. Local grammar of “risk”

4.2.1. The InterDiplo-Covid19 corpus

Taking into consideration the phraseology around the notion of risk in the two corpora, we discuss elements of the local grammar (Hunston & Su 2017) of the most frequent semantically related words identified in the wordlist. The focus is on pre- and post-modification of each noun, as well as on its verbal collocates.

Starting with the InterDiplo-Covid19 corpus, the noun *risk* is qualified by pre-modifying adjectives indicating the degree and we find collocates such as *potent*, *main*, *great/greatest*, *high/highest*, *increased*, and *significant risk*. All these patterns highlight the perception of Covid 19 as a disease of unprecedented dimension that has changed societies and are often used to describe the risks associated to chronic conditions or advanced age. Examples 1 to 2 provide instances of these patterns:

- (1) [...] if someone has *heart disease* and they're collecting fluid in their lungs they're going to be **at greater risk** well if *someone has poorly managed diabetes mellitus* they will be **at greater risk**.
- (2) it seems that the **main risk** at the moment is *older people* and *those with preexisting medical conditions*

Moreover, the noun *risk* is also qualified with post-modification, mainly realized through prepositional phrases (*risk of ...*, *risk for...*) used to describe the types of risks brought about by Covid19 (*zoonotic disease*, *death*, *dying*, *serious conditions*, *infection*, *exposure*, *severe complications*, *developing serious illness*) or by its quick spreading (*risk of the disease spreading like wildfire*, *risk of spread is high*, *risk of global catastrophe*, *risk of a pandemic worldwide*, *risk of the transmission throughout the world*) as shown in example 3:

- (3) We can't say it's not acceptable it must stop but we can say that all the evidence suggests that wet markets and the eating of bushmeat and similar practices are contributing to **an increased risk of zoonotic disease**

However, the representatives of world diplomacy also focus on the Coronavirus as a risk that needs to be accepted and mitigated/minimized and reduced through specific measures such as social distancing. This is shown by the frequency of verbs collocating with the noun *risk* as object, such as: *manage/minimize/mitigate/accept/reduce*. Examples are provided in 4 and 5:

- (4) You're going to reduce your risk of exposure to the virus. So the outside of 10 Feet you're going to **reduce the risk** and the outdoor environment is going to have much more potential to dilute
- (5) Our approach has been from the very beginning by default respecting the people's right to freedom of movement and then to implement measures where necessary and proportionate to the needs to **manage the risk** and and now because the risk factor comes with the people traveling coming from overseas into the countries we have placed all travelers coming from outside the country to 14 days of self isolation either in their homes or in government sponsored facilities

The local grammar of *risk* highlights how, in the case of Covid19, this notion seems to be affected by human decisions and by deliberate choices between alternative courses of action (Battistelli & Galantino 2019, 70), i.e. protecting people with pre-existing medical conditions, adopting social distancing.

Moving on to the noun *challenge*, which is the most frequent risk-related noun in the InterDiplo-Covid19 corpus, Covid19 is perceived as a “global challenge” and is often connoted by pre-modifying adjectives of size (*big/biggest, enormous, unprecedented, huge*). The local grammar of “challenge” shows that the pandemic experience involves the accomplishment of a difficult task that can only be faced by nations all together. Thus, ‘cooperation’ becomes a key notion in the context around the noun *challenge*, as in examples 6 and 7:

- (6) This pandemic is a huge wake up call for all of us in terms of giving ourselves the means globally in a **cooperative** manner to fight a **global challenge** that knows no border that knows no passport called coronavirus
- (7) This is a a **challenge** that we have to **overcome together** as a **global community**

When we consider post-modification, the noun *challenge* is also often characterized by the globality of its nature. We often find that challenge is followed by prepositional phrases introduced by *for* plus expressions involving the people facing the challenge - *the world, all of us, most countries, the entire world, the entire humanity, the democratic governments*. These patterns highlight again the need for a collective action to face Coronavirus as shown in (7) and (8):

- (8) Yes covid-19 I think him as a as a big crisis **big challenge for the entire humanity** so Pakistan was no exception
- (9) So it was a painful process of discovery more learning more about this virus and knowing how to deal with it I think **this is a challenge for all of us**

Another interesting pattern in the local grammar of *challenge* is its post-modification through *that-clause*. Once more, Covid19 is described as a global challenge that could only be faced by the world as a community (e.g. *a global/big challenge that knows no border/we have to overcome together/we cannot win alone*)

The cooperative action of nations is also reflected in the verbal patterns around the noun *challenge* when it is used as direct object. These include verbs of action that express the need of dealing with the pandemic (*face the challenge, meet the challenge, address the challenge*) sometimes also recurring to stronger verbs such as *fight* as in example (9):

- (10) This pandemic is a huge wake up call for all of us in terms of giving ourselves the means globally in a cooperative manner **to fight a global challenge** that knows no border

Similarly to what has been observed for *risk*, the noun *challenge* seems to be used in the language of diplomacy in patterns where the action of governments and people is perceived as fundamental to overcome the pandemic

Moving on to *threat*, we find pre-modification patterns that are similar to those of *risk*. It is possible to observe frequent elements that qualify the intensity and the globality of the threat posed by Coronavirus (e.g. *global, common, transnational*), as well as its possible reiteration (e.g. *continuing, constant*). A further instance of pre-modification of *threat* deals with the nature of the threat described (e.g. *health, Covid19, public health, pandemic*). See the examples below:

- (11) The really important point is to stress that it's not going to suddenly go away it's going to stay with us for the foreseeable future so this capacity to defend against it and to learn how to live with **the constant threat** of the virus is going to be the key for the future of humanity.
- (12) [...] we are fighting together a common threat and there is a **common threat** that which **transnational threat** which doesn't respect borders.
- (13) [...] last week the United Nations chief described the pandemic as the most challenging crisis we have faced since the Second World War the secretary general called for a coordinated global response to the **health threat** with an emphasis on developing nations reminding countries that we are only as strong as the weakest health system in our interconnected world

Post-modification of *threat* is extensively realized by prepositional phrases introduced by *to* and identifying the entities concerned by the threat. As it was for *challenge*, these consist of people perceived as a global community:

e.g. *a common threat to the whole world, to every country, to the wellbeing and the life of people*. The noun *threat* is also post-modified by prepositional phrases with *for* that qualify the kind of threat (e.g. *threat of Covid-19, of the virus, of Coronavirus*).

The most frequent verbs collocating with *threat* are *pose, face, fight* and they are used with a twofold purpose: on the one hand, to establish the existence of the threat when the noun is used as subject (*the threats posed by Coronavirus*); on the other hand, to describe the need for a common action of people when *threat* is the object of verbs such as *face* or *fight* as in the following example:

- (14) [...] every country needs to do both at the same time **fight covid-19 threat** but also address the other threats and challenges which are faced [...]

The last risk-related term for frequency in the InterDiplo-Covid 19 corpus is *danger* which is far less used by diplomats and international operators. As for *threat*, *danger* is qualified through pre-modification to highlight its intensity (*huge, great, grave*) and it is often post-modified by prepositional phrases indicating the type of *danger* (*of viruses, of pandemics*).

If we compare the core notion *risk* to the other risk-related nouns found in the InterDiplo-Covid 19 corpus using the the Wordsketch difference tool in Sketchengine, the most interesting variation we observe is when we compare the collocational patterns of *risk* and *challenge* (Figure 2). They only include verbal patterns, but the difference concerns the kind of actions needed to deal with those two issues: on the one hand, *risk* can only be identified (*be*), differentiated or reduced, so with a limited agency and intentionality; on the other hand, *challenge* requires more interventions to be addressed, met or even fought.

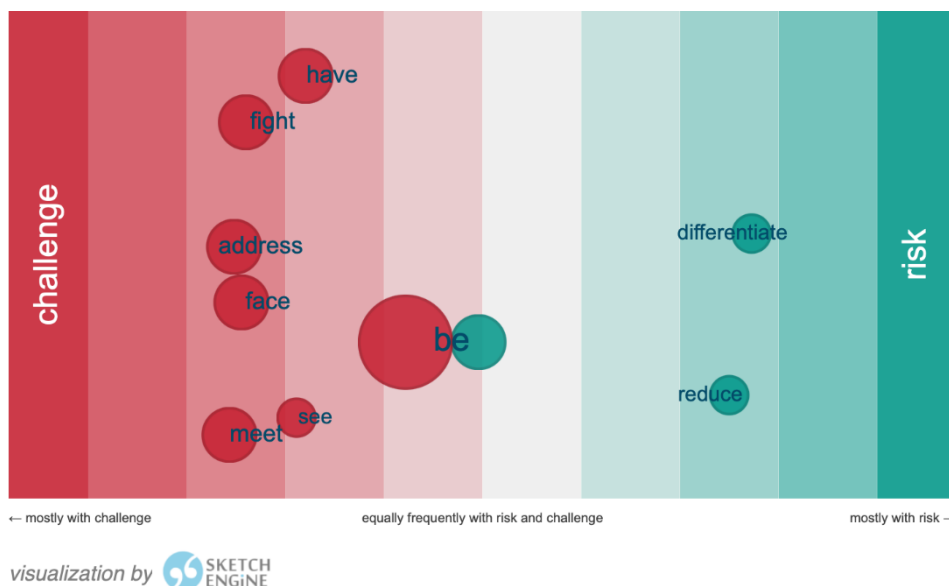


Figure 2

Risk and Challenge compared in terms of their associations.

In the next section, the phraseology around the notion of risk-related terms will be discussed in relation to the CORON-19 corpus considering their local grammar.

4.2.2. The CORON-19 corpus

The key role played by *risk* in the CORON-19 corpus and the nature of the corpus may concur in determining a wider range of pre-modification forms. If the most frequent elements qualify the intensity and the probability of the risk (*high, greater, relative, potential, low, elevated, significant, serious risk*), another interesting set of pre-modifiers qualifies the nature of the risk in question (*infection, health, transmission, disease, exposure, mortality, zoonotic, public, occupational, fatality, personal, cancer risk*). Pre-modification often combines with post-modification in introducing types of risks with the prepositions *of* and *for*: *risk of severe complications, the relative risks of cardiac failure, the potential risks of infectious agents; the potential risk for public health, additional risks for infectious complications, risks for errors; the risk of covid-19 for hospital staff and patients, etc.*

A study of verbs collocating with *risk* as object or subject provides an overview of the contexts in which risk is discussed. The verbs that have *risk* as object can be roughly divided into three main semantic categories:

- Verbs of change (increase/decrease): *increase, reduce, minimize, decrease, assess, mitigate a risk;*
- Verbs that establish a relational connection between risk and a potential agent of risk: *pose, carry a risk;*
- Verbs of (risk) assessment (with a potential “senser” of risk): *assess, perceive, associate, estimate, predict, model a risk.*

The examples below illustrate typical contexts in which hypotheses are formulated as to what helps contain risk (example 15), what poses a risk (and why) (example 16), and how risk can be assessed (example 17):

- (15) Conservation of wildlife and limits of the exposures of humans to wildlife will be important **to minimize the spillover risks** of coronaviruses from wild animals to humans.
- (16) This phenomenon is well documented in many viral zoonoses, yet Coronaviruses **pose exceptional risk to human populations due to** the variety of potential reservoirs and the virulence of the emergent pandemics they cause
- (17) By combining three different modelling approaches, we created a tool to **assess the risk of** 2019-nCoV outbreaks in countries outside of China.

Similarly, verbs with *risk* as subject often refer to processes establishing the risk as such (as in example 18), providing an explanation of risk in terms of its causes (as in example 19) and providing its relative position and assessment (20):

- (18) Nevertheless, **a finite risk exists** of droplet and airborne disease transmission while traveling in a commercial airplane.
- (19) **This risk depends on three key parameters:** the cumulative number of cases in areas of China which are not closed, the connectivity between China and the destination country, and the local transmission potential of the virus.
- (20) Given the **risk of death conditional on** disease of 0.6% for the group of 50-60 yrs old, **the daily mortality risk outweighs** the wages lost due to absenteeism (around \$170 per day in the US) for most workers assuming the value of statistical life of \$3mln.

A distinctive feature that characterizes the representation of *risk* in our academic corpus is also the higher frequency of modality markers hedging the degree of (un)certainty in the calculation of risk. Epistemic modal markers like *may, appear* and *seem* are particularly frequent around the notion, highlighting risk assessment as a key issue in science:

- (21) The risk of recombination involving other attenuated veterinary herpesvirus vaccines has not been evaluated comprehensively, but **the risk appears to be higher** in some viruses (such as EHV-4) and lower in others (such as EHV-1). The results from this study suggest that the risk for FHV-1 vaccine recombination **may also be low**.

As shown in the preliminary overview, *challenge* is far less used in scientific discourse. It is preferably pre-modified by adjectives expressing the relevance of the challenge (*major, great, big, unique, significant, key, serious, main*) and more often associated with the challenges of scientific research than with the notion of specific health risks, as in the following examples:

- (22) **One of the major challenges** of the current pandemic to date is disease detection and diagnosis.
- (23) **To tackle this major challenge, we developed** a geo-stratified debiasing estimation framework based on the following observation.

Similarly, post-modification shows a preference for prepositional elements introduced by *for*, which typically identify either the nature of the challenge itself or the people facing the challenge: *for doctors, for the globe, for clinicians throughout the world* etc.

The verbs collocating with *challenge* as subject or object are widely dominated by relational verbs, establishing the existence (or persistence) of a challenge: *pose, face, address, present, meet, arise, exist, encounter* etc. These contribute to identifying the problem, monitoring the situation and highlighting the relevance of the issues dealt with, rather than specifying elements of interpretation and assessing what is actually taking place. Examples are provided below:

- (24) Cities outside Hubei are also **facing great challenge** and require implementing of effective and feasible strategy in precision diagnosing novel coronavirus pneumonia (NCP).
- (25) So far, there is no specific drug for SARS-CoV-2, and **challenges remain**.

Elements of interpretation become really prominent when a *threat* is identified. Not only is *threat* qualified in a wider range of ways; it is also more often subjected to interpretative processes.

Pre-modification shows that the word is qualified in terms of intensity (*serious, major, great, significant*), but also of dimension (*global, worldwide*), epistemic status (*potential, constant, real, ongoing*) and in terms of type of threat (*health, public, pandemic, disease, security, infectious, biological, bioterrorism, microbial*). Post-modification is widely dominated

by prepositional phrases introduced by *to* and identifying the entities affected by the threat: e.g. *a serious threat to public health, to human health, to the unprotected population of southern Brazil, to the poultry industry, to the patient, to FMD-free countries, to the environment, to tourism, to front-line health workers.*

Verbs collocating with *threat* include words like *pose* and *face*, used to introduce the notion and establish the existence of a threat (*pose a threat, facing the threat, constitute a threat*), but also more interpretative verbs of projection, perhaps required by the negative semantic meaning of *threat* in terms of agency and intentionality. Verbs like *perceive, represent, consider, recognize* emphasize the interpretative task required by using the word *threat*, while verbs of change like *emerge, remain, become, arise, continue* emphasize the need to be on the alert in a constant re-interpretation of events as threats:

- (26) The continuing pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, previously known as 2019-nCoV) **has now become an international public health threat**, causing inconceivable loss of lives and economic instability

Hazard and *danger* are also clearly related to *risk* and *threat*, though typically referred to contextual circumstances. As shown by the Wordsketch difference tool in Sketchengine (Figure 2), *danger* is more clearly associated with *difficulty* and *fear*, whereas *hazard* is associated both to issues of *vulnerability* and *communicability* and to its potential nature and consequences (*chemical, disaster*). Modifiers of the two words also mark a slight difference between the two nouns: *danger* is qualified by its epistemic status and intensity (*imminent, impending, grave, mortal*), while *hazard* is more often classified into types (*occupational, moral, smog-related, biological, environmental*) and generally recognized as a specific object of study in occupational medicine (as a condition arising from exposure to agents potentially affecting normal physiological mechanisms). Verbs collocating with the two words also show a higher presence of mental interpretative processes such as *perceive, realize, sense* in collocation with *danger*, thus perhaps confirming the view that *hazard* refers to a situation where danger is always only potential, whereas reference to *danger* rather indicates a probability of the negative outcome.

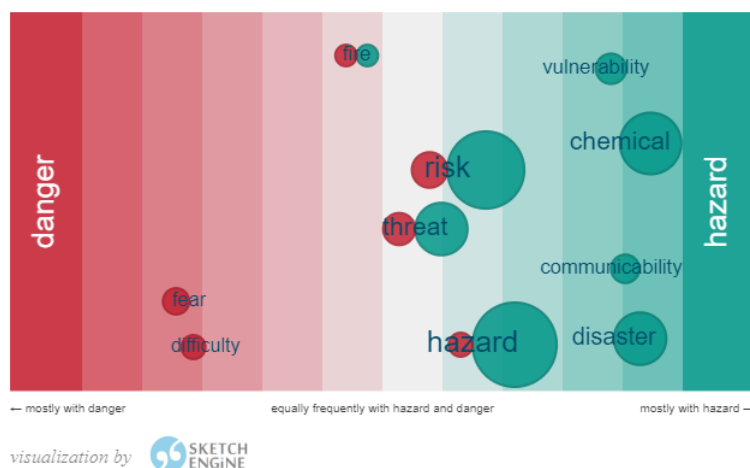


Figure 3

Danger and hazard compared in terms of their associations.

The various lexical realizations of the notion of ‘risk’, however, are often closely related and not always clearly distinguished, as we can see from the contextual use of *risk*, *threat*, *danger* and *hazard* in the next example:

- (27) [...] at least 2 broad methods exist for analyzing the role of **risk** in social science, and the choice of method is important. [...] The most common approach is the realist approach, in which **risk is seen as an objective threat or danger** that can be measured independent of the social context within which it occurs [...]. The alternative is the social constructionist approach, which describes **risk as being based on objective facts about danger and hazard**, amenable to rationalistic calculations, which are then mediated, perceived, and responded to in particular ways through social, cultural, and political processes

5. Conclusions

Our paper has delved into the conceptualisation and framing of ‘risk’ in the Covid-19 pandemic. We explored occurrences of risk-based vocabulary on two corpora dealing with the spread and the political/social/economic consequences of Covid19 from February 2020 to February 2021 (i.e. the InterDiplo Covid-19 corpus and the CORD-19 corpus). Our aims were to study the role of risk-related vocabulary at the outbreak of the most dangerous pandemic of the last centuries focusing in particular on the thematization of risk in two widely differing contexts: international relations and scientific language. To reach these purposes, we adopted a mixed-method approach, i.e., a data-driven quantitative and qualitative analysis (Baker 2006, Müller 2017, Taylor and Marchi 2018) in which corpus linguistics tools were combined with a discourse analytical perspective to investigate the development of *risk* thematization in the two corpora. Therefore, after a first recognition of all related expressions of the nominal realization of “risk” in the worldlists of the two corpora, we analyzed their concordances to find typical collocations and phraseology used to qualify risk in Covid-19 epidemic.

Starting with quantitative results, risk vocabulary is more varied in the CORD-19 corpus where we find 6 risk related lexical items (*risk, challenge, threat, distress, hazard* and *danger*) with normalized frequency higher than 10 pmw. In the InterDiplo-Covid 19 corpus, we observe only 4 risk related lexical realizations (*risk, challenge, threat, danger*) and the notion of risk swings from a positive valence description to negative ones. This may of course depend on the huge size difference between the two corpora, but what is most interesting is the different relative importance of the various lexicalizations. Considering frequency, the most frequent element in the InterDiplo Covid-19 corpus is *challenge*, which is often found in relation to a positive experience usually involving the accomplishment of a difficult task. Whereas the other three risk-related nouns carry a negative value (a possibility of being exposed to danger or threat). On the other hand, in the CORD-19 corpus, the situation is reversed, and the emphasis of scientific discourse seems to lie more on *risk*, which is by far more frequent than the more positive (or neutral) *challenge*. In terms of intentionality and agency, we noticed the same cline in the two corpora, from positive intentionality to negative intentionality to negative agency, but the range of values is much wider: the use of *risk* is much more frequent in scientific discourse, whereas reference to *threat* and *danger* is much more limited.

Moving on to the phraseology involved in the qualification of *risk* in the corpora, and to the communicative functions realized through this phraseology, we observed some differences due to the different aims that

communication tries to fulfil according to the discourses. Indeed, in the academic corpus the main purpose seems to be finding an understanding within the scientific community on the on how to assess the risks of the pandemic in terms of identifying risks as such, their nature, the subjects affected and the likely causes. On the other hand, diplomats and international relation experts are more inclined to evaluate the pandemic risk as for its dimension and global relevance in order to shift the focus on its consequences on world communities and the necessity for a joint effort of all countries to fight Coronavirus.

A similar trend is actually found with *threat*: diplomats are on the whole more interested in threats (with their implied negative intentionality) than scientists, but this also involves giving people a more active role in facing the threat, while scientific discourse is more interested in the more neutral notion of risk, but phraseology and the communicative functions involved – whether with risk or threat - have more to do with finding tools for assessing risks and threats.

Another interesting difference between the two corpora concerns the use of modality markers in the representation of risk. In the CORDIS-19 academic corpus, the frequency of modality markers is high, and they are used to hedge the degree of (un)certainty in the calculation of risk. Thus, risk assessment can be considered as a key issue in science when dealing with a pandemic. On the other hand, in the InterDiplo Covid-19 corpus, modality is almost absent, and the key focus is placed on action against the risk in its globality without any epistemic evaluation, pointing rather at issues of risk management.

Moreover, the two corpora also show a different perception of risk when we consider the phraseology around the noun *challenge*: in the CORDIS-19, *challenge* is more often associated to scientific research than with the notion of specific health risks, hence expressing a more neutral and objective interpretation that includes the identification of the problem and the monitoring of the pandemic situation. In the InterDiplo-Covid19, conversely, *challenge* acquires a far more positive connotation and frequently collocates with verbs indicating the possibility of overcoming the pandemic thanks to a strong cooperative effort of people.

This study is a preliminary investigation, and the findings should be seen in light of some limitations: first of all, it would be interesting to address the diachronic development of risk-related items following the development of the pandemic and study its evolution also after the vaccination diffusion process. A closer study of the lexico-grammatical patterns around the key lexical elements could also lead to a more precise study of the grammatical patterns of the clauses in which nominal elements are found. Comparison of nominal and verbal elements expressing the notion of ‘risk’ would also add to

the picture. Finally, the data are limited to specific contexts and specific lexical elements, which could probably be further illuminated by a wider perspective on the whole range of expressions related to the notion of uncertainty (cf. Müller et al. 2021) or by further comparison with the language of the media.

On the other hand, we hope to have illustrated how an analysis of the collocation, colligation and semantic preference of risk-related vocabulary can contribute to identifying the common (and divergent) communicative functions of different genres and discourses. We also hope to have shown how studying the representation of risk and its phraseological lexico-grammatical patterns can actually contribute to identifying not only different communicative functions, but ultimately different thematizations across discourses, e.g. highlighting the centrality of risk assessment in scientific discourse or risk management in diplomatic discourse.

Bionotes: Marina Bondi is a full professor of English Linguistics at the University of Modena and Reggio Emilia (Italy), where she is the coordinator of the PhD programme in Human Sciences. She has published extensively in the field of genre analysis, EAP and corpus linguistics, knowledge dissemination and the impact of digital media on specialized discourse.

Silvia Cavalieri holds a Ph.D in Comparative Languages and Cultures from the University of Modena and Reggio Emilia (UNIMORE). She is currently a temporary assistant professor (RTD-B) at the Department of Foreign Languages and Literatures of the University of Verona. Her areas of research and publication include ESP, legal language, courtroom discourse, academic discourse, and medical discourse with a focus on corpus linguistics and digital humanities. She is currently working on the InterDiplo project, which focuses on the language used in spoken interactions by diplomats, ambassadors, and international relations.

Authors' addresses: marina.bondi@unimore.it; silvia.cavalieri@univr.it

References

- Baker, P. 2006, *Using Corpora in Discourse Analysis*. Continuum, London / New York
- Battistelli, F. and M.G. Galantino 2019. *Dangers, Risks and Threats: An Alternative Conceptualization to the Catch-All Concept of Risk*, in “Current Sociology” 67 (1), pp. 64–78. doi:<https://doi.org/10.1177/0011392118793675>.
- Bernard, N. R., Basit, A., Sofija, E., Phung, H., Lee, J., Rutherford, S. and N. Wiseman 2021, *Analysis of Crisis Communication by the Prime Minister of Australia during the COVID-19 Pandemic*, in “International Journal of Disaster Risk Reduction” 62, 102375.
- Boholm, M. 2009, *Risk and Causality in Newspaper Reporting*, in “Risk Analysis: An Official Publication of the Society for Risk Analysis” 29 (11), pp. 1566–1577.
- Boholm, M. 2012, *The Semantic Distinction between “Risk” and “Danger”: A Linguistic Analysis*, in “Risk Analysis: An Official Publication of the Society for Risk Analysis” 32 (2), pp. 281–293.
- Boholm, M. 2016, *Risk Association: Towards a Linguistically Informed Framework for Analysing Risk in Discourse*, in “Journal of Risk Research” 21 (4), pp. 480–501. <http://dx.doi.org/10.1080/13669877.2016.1223158>
- Boholm, M., N. Möller and S. O. Hansson 2016, *The Concepts of Risk, Safety, and Security: Applications in Everyday Language*, in “Risk Analysis: An Official Publication of the Society for Risk Analysis” 36 (2), pp. 320–338.
- Candlin, C. N., J. Crichton and A.S. Firkins 2016, *Crucial Sites and Research Orientations: Exploring the Communication of Risk*, in Crichton J., C.N. Candlin and A.S. Firkins (eds.), *Communicating Risk*, Palgrave Macmillan, Basingstoke, pp. 1-14.
- COVID-19 Open Research Dataset (CORD-19). 2020. Version 2020-05-02. Retrieved from <https://pages.semanticscholar.org/coronavirus-research>. Accessed 2020-05-02. doi:10.5281/zenodo.3715505
- DePaula, N., Hagen, L., Roytman, S., Dyson, D., Alnahass, D., Patel, M. and A. Hill 2022, *A Framework of Social Media Messages for Crisis and Risk Communication: A Study of the Covid-19 Pandemic*. In *Proceedings of the 55th Hawaii International Conference on System Sciences*.
- Fillmore, C.J. and B. T. Atkins 1992, *Toward a Frame-Based Lexicon: The Semantics of RISK and its Neighbors*, in Lehrer A. and E. F. Kittay (eds.), *Frames, Fields, and Contrasts: New Essays in Semantics and Lexical Organisation*, Lawrence Erlbaum Associates Publishers, Hillsdale, pp. 75-102.
- de Las Heras-Pedrosa, C., P. Sánchez-Núñez, and J.I.Peláez 2020, *Sentiment Analysis and Emotion Understanding during the COVID-19 Pandemic in Spain and its impact on Digital Ecosystems*, in “International Journal of Environmental Research and Public Health” 17(15), pp. 5542.
- Dyer, J., and B. Kolic 2020, *Public Risk Perception and Emotion on Twitter during the Covid-19 Pandemic*, in “Applied Network Science” 5 (1), pp. 1-32.
- Galantino, M.G. 2020, *The Migration-Terrorism Nexus: An Analysis of German and Italian Press Coverage of the ‘Refugee Crisis’*, in “European Journal of Criminology” 10, pp. 1-23. Doi: <http://dx.doi.org/10.1177/02F1477370819896213>
- Giddens, A. 1998, *Risk Society: The Context of British Politics*, in Franklin, J. (ed.), *Politics of Risk Society*, Polity Press, Cambridge, pp. 23-24.
- Hamilton, C., S. Adolphs and B. Nerlich 2007, *The Meanings of ‘Risk’: A View from*

- Corpus Linguistics*, in “Discourse & Society” 18 (2), pp. 163–181. doi:<https://doi.org/10.1177/0957926507073374>.
- Hardy, D.E. and C.B. Colombini 2011, *A Genre, Collocational, and Constructional Analysis of RISK*, in “International Journal of Corpus Linguistics” 16 (4), pp. 462—485. doi 10.1075/ijcl.16.4.02
- Hunston, S. and H. Su 2019, *Patterns, Constructions, and Local Grammar: A Case Study of ‘Evaluation’*, in “Applied Linguistics” 40 (4), pp. 567–593, <https://doi.org/10.1093/applin/amx046>
- Hyland, K. and K. Jiang 2021, *Competition and the Hying of Virus Research*, in “International Journal of Corpus Linguistics” 26 (4), pp. 444-468.
- Luhmann, N. 1993, *Risk: A Sociological Theory*, de Gruyter, Berlin/New York.
- Müller, M. 2017, *Digital Discourse Analysis*. LitLab Pamphlete #5. Online publication: https://www.digitalhumanitiescooperation.de/wp-content/uploads/2019/08/p05_mueller_eng.pdf (Last accessed on March 13 2022)
- Müller, M. and R.M. Mell 2022, ‘Risk’ in Political Discourse. A Corpus Approach to Semantic Change in German Bundestag Debates, in “Journal of Risk Research” 25 (3), pp. 1-16.
- Müller, M. and J. Stegmeier 2019, *Investigating Risk, Uncertainty and Normativity within the Framework of Digital Discourse Analysis. The Example of Future Technologies in Climate Change Discourse*, in Olofsson A. and J.O. Zinn (eds.), *Researching Risk and Uncertainty – Methodologies, Methods and Research Strategies*, Palgrave, Basingstoke, pp. 309-355.
- Müller, M., S. Bartsch and J. O. Zinn 2021, *Communicating the Unknown. An Interdisciplinary Annotation Study of Uncertainty in the Coronavirus Pandemic*, in “International Journal of Corpus Linguistics” 26 (4), pp. 498–531. doi:<https://doi.org/10.1075/ijcl.21096.mul>.
- Rothstein, H., D. Demeritt, R. Paul and L. Wang 2022, *True to Type? How Governance Traditions Shaped Responses to Covid-19 in China, Germany, UK and USA*, in Brown P. and J. Zinn (eds.), *Covid-19 across 6 Continents – Social Challenges, Responses and Consequences*, Palgrave Macmillan, Cham Switzerland.
- Semino, E. 2021, “Not Soldiers but Fire-fighters” – Metaphors and Covid-19, in “Health Communication” 36 (1), pp. 50—58. DOI: 10.1080/10410236.2020.1844989
- Sinclair, J. 2004, *Trust the Text: Language, Corpus and Discourse*, Routledge, London.
- Taylor, C. and A. Marchi 2018, *Corpus Approaches to Discourse: A Critical Review*, Routledge, London
- WHO (World Health Organisation). 2017. *Communicating Risk in Public Health Emergencies. A WHO Guideline for Emergency Risk Communication (ERC) Policy and Practice*. Geneva: World Health Organization.
- Wicke P. and M. M. Bolognesi 2021, *Covid-19 Discourse on Twitter: How the Topics, Sentiments, Subjectivity, and Figurative Frames Changed Over Time*, in “Frontiers in Communication” 6, Article 651997 DOI=10.3389/fcomm.2021.651997
- Zhang, L., H. Li and K. Chen 2020, *Effective Risk Communication for Public Health Emergency: Reflections on the COVID-19 (2019-nCoV) Outbreak in Wuhan, China*, in “Healthcare” 8 (1), p. 64.
- Zinn, J.O 2010, *Risk as Discourse: Interdisciplinary Perspectives*, in “Critical Approaches to Discourse Analysis across Disciplines” 4 (2), pp. 106-124.
- Zinn, J.O. and D. McDonald 2018, *Risk in the New York Times (1987–2014). A Corpus-Based Exploration of Sociological Theories*, Palgrave, London.