Laura Giacomini

Specialised language and conceptual knowledge in lexicographic portals

Abstract: This contribution focuses on the current state of specialised language and conceptual knowledge coverage in general-language or generalist lexicographic portals. The final goal is to establish whether their terminological and conceptual treatment is extensive enough to be compared with that found in specialised dictionaries used by terminologists, translators, and technical writers. Terminology of two domains, i. e. the medical field of virology related to the COVID-19 pandemic and the technical field of photovoltaics, is analysed in selected lexicographic portals. Results show that hybrid lexicographic and lexicological portals display the highest level of terminological coverage and detailed conceptual systems.

Keywords: lexicographic portal, terminology, specialised language, wordnet, conceptual knowledge

1 Introduction

Contemporary electronic lexicography is characterised, on the one hand, by single, stand-alone lexicographic resources and, on the other hand, by lexicographic portals which, through modalities such as external links or embedded networks, allow for the consultation of multiple lexicographic and lexicological resources. To the best of our knowledge, the vast majority of lexicographic portals focus on the general lexicon of one or more languages. Data about specialised languages are usually covered if they are also relevant to common language. This leads to considerations about missing specialised language portals as a crucial gap in lexicography.

In this contribution, the current state of specialised language and conceptual knowledge coverage in lexicographic portals will be ascertained. The aim is to establish to what extent special languages (LSP) are treated in general-language (LGP) or gener-

Laura Giacomini, Universität Innsbruck, Institut für Translationswissenschaft, Herzog-Siegmund-Ufer 15, 6020 Innsbruck, Austria, E-Mail: laura.giacomini@uibk.ac.at

https://doi.org/10.1515/lex-2023-0008

alist portals and whether their terminological and conceptual treatment is extensive enough to be compared with that found, for example, in specialised dictionaries used by terminologists, translators, and technical writers.

After introducing a typology of lexicographic portals, analysis will be carried out on selected types and on terminology of two domains, namely the medical field of virology related to the COVID-19 pandemic (Field I) and the technical field of photovoltaics (Field II).

The analysis of terminology coverage in Field I, a specialised field that has recently come to the forefront in the media, also aims at stating how quickly lexicographic portals are updated in relation to the introduction of new terms or to terminological risemantisation.

2 Typology and structural properties of lexicographic portals

The apparent heterogeneity of online lexicographic systems makes it complicated to establish a generally valid typology. This is primarily true of single online resources, for which the use of the generic label *lexicographic information system* (Leroyer 2011) is useful, precisely because it helps avoiding distinct categories such as *dictionary*, *glossary* or *database*. However, this is equally true of online portals that collect several lexicographic resources or are a combination of lexicographic resources. The term *dictionary* is therefore to be understood in a broad sense as indicating an online lexicographic resource aimed at the linguistic description of the lexicon and characterised by a specific, typically hierarchical, entry structure (Wiegand 1989).

In order to highlight the central features of lexicographic portals, this contribution takes inspiration from the typology proposed by Engelberg/Storrer (2016), to which useful elements will be added with regard to the assessment of conceptual knowledge coverage.

We will not go into the details of the characteristics of single digital lexicographic resources, which can be related to the medium itself (e. g. digitised or digital native dictionaries) or be independent of it (e. g. type of user, function or number of languages). For such classification, we refer again to the typology proposed in Engelberg/Storrer (2016), as well as to models valid for paper dictionaries (cf. Wang 2001). A necessary premise to be made is that the portals described in this contribution are all classified a priori as lemma-related, i. e. as resources that are primarily semasiological. An onomasiological classification, although potentially interesting, is not relevant at present, since the conceptual component in the existing resources is subordinate to the lexical one. Based on the degree of independence and integration of the resources included in the portal, lexicographic portals can be divided into lexicographic collections, lexicographic search engines and lexicographic networks.

1. **Lexicographic collections** offer outer, disjoint access to linked dictionaries, which are independent of each other. These are typically lists of dictionary links that a user can access according to specific needs. They enable quick reference to a number of resources that may be very different in nature. This category of lexicographic portals, for instance, may grant access to general language and/or LSP resources, monolingual and/or multilingual resources. Examples of lexicographic collections are The European Dictionary Portal (with outer access to dictionaries and portals provided by the European Network of e-Lexicography), the TermCoord Glossary links, and the Glossaries in the CLARIN infrastructure. From a structural point of view, these collections are equally heterogeneous and elude a unitary description. A common feature are their sources, usually academic institutions and language services, e. g. addressed towards translators or technical writers.

2. **Lexicographic search engines** offer indirect, joint access to the entries of a number of independent external dictionaries, each with a unique layout. Examples of lexicographic search engines are The Free Dictionary (aggregating general language, special, and LSP dictionaries, as well as encyclopaedic resources for many languages), and OneLook (with dictionaries and glossaries for several languages). The advantage of this kind of lexicographic portal lies in the possibility of accessing a number of entries from different dictionaries on the same webpage, thus enabling a direct comparison between entries.

The macrostructure of a lexicographic search engine typically involves a start page with a search mask for external semasiological access. Aggregated resources are immediately visible or accessible via links. There may be more or less complex outer features (cf. Klosa/Gouws 2015): the Free Dictionary includes, for instance, sections such as “Word of the Day”, “The Free Dictionary Blog”, “Daily Grammar Lesson”, “Flashcards” and many others. The microstructure of the lexicographic search engine portal is a set of specific microstructures of the dictionaries simultaneously accessed when performing a query.

Some lexicographic search engines also provide access to LSP resources. An example of this kind is The Free Dictionary portal, through which it is possible to view, for a certain lemma, the corresponding entries in various medical, legal and financial dictionaries. Conceptual coverage is conditional on the possibility of accessing resources, such as the thesaurus in The Free Dictionary, which add the possibility of onomasiological access to data.

This contribution focuses on lexicographic networks as well as hybrid networks. The latter bring together lexicographical and lexicological resources. The interesting fact about portals of this kind is that they offer direct access to embedded resources and the connection between these resources is the result of targeted planning. In such portals it will therefore be possible to check the intentional coverage of specialised languages and knowledge data.

3. **Lexicographic networks** are portals in which individual lexicographic resources are directly accessed and interconnected in terms of content and appearance (layout)

of the portal. Examples of this type of portal include Lexico.com (featuring OUP English and Spanish dictionaries), OWID (an online lexical information system for German by the Institute for the German Language), Treccani.it (with a general dictionary, a dictionary of synonyms, a dictionary of neologisms, and an encyclopaedia for Italian), Svenska Akademiens Ordböcker (providing dictionaries of the Swedish language by the Swedish Academy), and Diccionarios.com (containing a large number of dictionaries mainly for the Spanish, English, and French languages featured by Larousse and Vox).

The macrostructure of a lexicographic network is usually made up of a relatively small number of lexicographic resources as well as outer features. Dictionaries can be queried independently or jointly through a search mask, as is the case in lexicographic search engines. The only or main access structure is semasiological. The way in which results of a query are visualised varies. Also, microstructural features depend on the specific lexicographic resources but tend to be more homogeneous than in lexicographic search engines. This is due to the fact that lexicographic networks typically collect dictionaries from the same source, e. g. the same publisher.

In this category we also include portals that bring together, instead of dictionaries, a number of interconnected multilingual wordnets, such as MultiWordNet and the Open Multilingual Wordnet. The structural element that wordnet-like lexicographic networks typically have in common is a semasiological access based on a search mask, for which one or more languages can be set. The semasiological search for a word results in the identification of the synsets to which the word belongs, each connected to a distinct concept. Semasiological organisation and onomasiological organisation are therefore naturally linked in this type of portal. The microstructure of the entries is variable but reflects the usual architecture of wordnets, with the indication of synsets, lexical and conceptual relations, as well as synsets in further languages. The extended version of the Open Multilingual Wordnet, with its integrated Wiktionary data and the external links to SUMO, TempoWordNet and SentiWordNet, shares some common features with hybrid networks such as BabelNet but is much simpler in its structure and shallow in its content. For this reason, we decided to treat the Open Multilingual Wordnet among simple lexicographic networks.

Engelberg/Storrer (2016) point out that lexicographic networks are the prototypical form of a lexicographic portal, since they combine the highest level of integration with the highest level of autonomy of the individual resources.

4. The structure of **hybrid lexicographic-lexicological networks** is not uniform and largely depends on the focus of each portal, which may be more linguistic or more encyclopaedic. Purely encyclopaedic portals such as Encyclopedia.com, a collection of online encyclopaedias by renowned publishers, have not been taken into account.

Examples of hybrid portals are BabelNet (a multilingual encyclopaedic dictionary), Wortschatz Leipzig (a collection of corpus-based multilingual dictionaries), DWDS (a corpus-based information system for German), and ConceptNet (a multilingual knowledge base). Wortschatz Leipzig, for instance, has a more linguistic orientation, BabelNet a more encyclopaedic one. Common aspects are the macrostructural complexity and,

from a mediostructural point of view, the possibility of starting a query from any portal component, while access to all data connected to it remains ensured. We will illustrate this by considering some structural aspects of BabelNet, in which very heterogeneous resources are linked together.

In terms of macrostructure, BabelNet is very straightforward and clear. External access to the data is via a search mask in which a term is entered. When performing a query, the language of the term must be set, possibly together with a second language. Various outer features can be accessed from the main page. The microstructural properties of BabelNet are mentioned in Section 3.1.2.

Table 1 summarises the main features of lexicographic portals and highlights the types dealt with in this contribution.

Table 1: Main features of lexicographic portals. Types 3 and 4 are the core topic of this contribution.w

1. LEXICOGRAPHIC COLLECTIONS	2. LEXICOGRAPHIC SEARCH ENGINES	3. LEXICOGRAPHIC NETWORKS	4. HYBRID NETWORKS
e.g. Glossaries in the CLARIN infrastructure	e.g. OneLook	e.g. Lexico.com	e.g. BabelNet
Outer, disjoint access to a set of external dictionaries.	Indirect, joint access to a set of external dictionaries (and encyclopaedias).	Direct access to a set of interconnected dictionaries or wordnets.	Direct access to a set of interconnected lexicographic, lexicological, and sometimes encyclopaedic resources.

Mention should be made at this point of the numerous Linked Data resources, even though they will not be dealt with because of their extreme heterogeneity and because they go beyond the topic of portals in the narrow sense. Some of them are lexicographical and lexicological in nature, e. g. DBPedia, Geonames, and ConceptNet (the latter as part of the LLOD, a linguistic (sub-)cloud of Linked Open Data).

3 Specialised language and knowledge representation in LGP lexicographic portals

In order to explore the coverage of specialised language and conceptual knowledge in more detail, examples of each category have been chosen and their description will be accompanied by general considerations of the corresponding dictionary portals.

3.1 The coverage of specialised language in LGP lexicographic portals

In order to test the coverage of specialised languages in common language lexicographic portals, simple and complex terms from two specialised fields have been chosen:
– Field I: the medical field of virology related to the COVID-19 pandemic.
– Field II: the technical field of photovoltaics.

Terms belonging to Field I are listed in Table 2 together with the corresponding part of speech.

Table 2: Term set used for LSP coverage testing in Field I. Terms are followed by the part of speech under which they are searched for in the online portal: noun (N), adjective (ADJ), or verb (V).

FIELD I: COVID-19 PANDEMIC		
English	**Italian**	**German**
COVID-19 (N)	COVID-19 (N)	COVID-19 (N)
SARS (N)	contagio (N)	Lockdown (N)
asymptomatic (ADJ)	letalità (N)	Impfausweis (N)
sanitization (N)	quarantena (N)	Maskenpflicht (N)
disinfectant (N)	linfodema (N)	Quarantäne (N)
quarantining (ADJ)	droplet (N)	Letalität (N)
nonessential (ADJ)	tampone (N)	R-Zahl (N)
furlough (N)	sierologico (ADJ)	durchimpfen (V)
immune system (N)	paziente zero (N)	Herdenimmunität (N)
social distance (N)	immunità di gregge (N)	Social distancing (N)

Three languages, English, Italian, and German, were tested depending on the portal. Term sets in the different languages were employed independently of each other and were not necessarily equivalents. We chose terms with a homogeneous communicative role from the point of view of the vertical dimension, i. e. terms suitable for communication between experts and non-experts (or semi-experts). This level of specialised communication can reasonably be expected in common language lexicographic resources, since it reflects everyday consumption of specialised language. In order to obtain a representative set of terms for this communicative level, terms related to the COVID-19 pandemic were chosen for each language as follows:
– Ten of the most searched terms from March-May 2020 on Dictionary.com, an online lexicographic resource that uses the Random House Unabridged Dictionary as its primary source, as well as other sources such as American Heritage and Harper Collins.

– Ten of the most searched terms on Treccani.it, a lexicographic network that brings together interlinked lexicographic and encyclopaedic resources for the Italian language.
– Ten terms from the DWDS thematic glossary about the COVID-19 pandemic. As no similar list of the most searched terms for German was found as for the other two languages, the choice of the ten terms was based on criteria of equivalence to English and Italian.

The lists of terms were deliberately heterogeneous with regards to part of speech, term complexity, neologism status, or abbreviated form status. In addition to specific medical terms (e. g. *asymptomatic*, *droplet*, *R-Zahl*), terms referring to the social consequences of the pandemic were also selected (e. g. *furlough*, *quarantena*, *Lockdown*).

Terms belonging to Field II are listed in Table 3 together with their part of speech.

Table 3: Term set used for LSP coverage testing in Field II. Terms are followed by the part of speech under which they are searched for in the online portal: noun (N) or adjective (ADJ).

FIELD II: PHOTOVOLTAICS		
English	**Italian**	**German**
solar film (N)	pellicola solare (N)	Solarfolie (N)
substructure (N)	sottostruttura (N)	Unterbau (N)
highly efficient (ADJ)	ad alta efficienza (ADJ)	hocheffizient (ADJ)
photovoltaic (ADJ)	fotovoltaico (ADJ)	photovoltaisch (ADJ)
voltage (N)	voltaggio (N)	Spannung (N)
photovoltaic system (N)	impianto fotovoltaico (N)	Photovoltaikanlage (N)
in-roof (ADJ)	sopra tetto (ADJ)	Indach- (ADJ)
solar cell (N)	cella solare (N)	Solarzelle (N)
solar inverter (N)	invertitore fotovoltaico (N)	Solarwechselrichter (N)
renewable energy (N)	energia rinnovabile (N)	erneuerbare Energie (N)

Terms from the field of photovoltaics, which are less widely used in the media than those related to the new pandemic, were selected from a translation memory created on the basis of online texts published by manufacturers of photovoltaic systems and aimed at semi-expert users.

Although quantitative data on the occurrence of individual terms in the resources under analysis will be reported, such data must be considered in relation to the scope of the study and do not allow for generalisations. A primarily qualitative analysis will be carried out, which aims to answer the following question: is the term recorded as a lemma on the lexicographic portal and, if this is the case, in which specific resource(s)?

3.1.1 Specialised language in lexicographic networks

The presence of the terms selected for the two specialised fields was tested in OWID, Lexico.com, Treccani.it, and Open Multilingual Wordnet. The reference languages depend on the specific portal. It should be noted that, among the various lexicographic networks, OWID offers a specific dictionary dedicated to COVID-19 vocabulary ("Neuer Wortschatz rund um die Coronapandemie").

The dictionaries in which the selected terms are recorded as lemmas are listed below:
LEX1: Lexico.com, UK Dictionary
LEX2: Lexico.com, US Dictionary
LEX3: Lexico.com, Synonyms
TREC1: Treccani, Vocabolario on line (General dictionary)
TREC2: Treccani, Sinonimi e Contrari (Dictionary of synonyms and antonyms)
TREC3: Treccani, Enciclopedia on line (Encyclopaedia)
TREC4: Treccani, Neologismi (Dictionary of neologisms)
OWID1: Neuer Wortschatz rund um die Coronapandemie (COVID-19 vocabulary)
OWID2: Fremdwörterbuch (Dictionary of foreign words)
OWID3: Neologismenwörterbuch (Dictionary of neologisms)
OMW: Open Multilingual Wordnet, wordnets for English and Italian

Coverage of the term sample is indicated in Table 4:

Table 4: Coverage of selected COVID-19 and photovoltaics terminology in lexicographic networks: OWID, Lexico.com (LEX), Treccani.it (TREC), Open Multilingual Wordnet (OMW).

	TERM (English)	COVERAGE	TERM (Italian)	COVERAGE	TERM (German)	COVERAGE
FIELD I: COVID-19	COVID-19	LEX1, 2	COVID-19	TREC4	COVID-19	OWID1
	SARS	LEX1, 2	contagio	TREC1, 2, 3 OMW	Lockdown	OWID1
	asymptomatic	LEX1, 2 OMW	letalità	TREC1, 2	Impfausweis	–
	sanitization	LEX1, 2	quarantena	TREC1, 2, 3 OMW	Maskenpflicht	OWID1
	disinfectant	LEX1, 2, 3 OMW	linfedema	TREC1	Quarantäne	OWID2
	quarantining	–	droplet	TREC4	Letalität	–
	nonessential	LEX1, 2, 3	tampone	TREC1, 2, 3 OMW	R-Zahl	OWID1
	furlough	LEX2 OMW	sierologico	TREC1 OMW	durchimpfen	OWID1
	immune system	LEX1 OMW	paziente zero	TREC4	Herdenimmunität	OWID1, 3
	social distance	LEX1, 2	immunità di gregge	TREC3	Social distancing	OWID1
FIELD II: PHOTOVOLTAICS	solar film	–	pellicola solare	–	Solarfolie	–
	substructure	OMW	sottostruttura	TREC1	Unterbau	–
	highly efficient	–	ad alta efficienza	–	hocheffizient	–
	photovoltaic	LEX1, 2 OMW	fotovoltaico	TREC1 OMW	photovoltaisch	–
	voltage	LEX1, 2 OMW	voltaggio	TREC1, 2, 3 OMW	Spannung	–
	photovoltaic system	–	impianto fotovoltaico	–	Photovoltaikanlage	–
	in-roof	–	sopra tetto	–	Indach-	–
	solar cell	OMW	cella solare	–	Solarzelle	–
	solar inverter	–	invertitore fotovoltaico	–	Solarwechselrichter	–
	renewable energy	LEX1, 2	energia rinnovabile	–	erneuerbare Energie	–

A strictly lemmatic method was applied in the analysis, and cases in which one term appears in the entry of a further lemma were not taken into consideration (e. g. in Lexico.com, *solar cell* as a variant of *solar battery*, or *photovoltaic system* among the usage examples of *photovoltaic*). All orthographic lemma variants, such as *nonessential* and *non-essential*, have been accounted for.[19] As far as the Open Multilingual Wordnet is concerned, the focus was only on lexemes within synsets.

The results of the terminological analysis of lexicographic networks will now be summarised:

1. All portals record both single-word and multi-word terms as lemmas.
2. Coverage of Field I is larger than that of Field II. It is difficult to substantiate this difference with certainty, but it might be due to the fact that COVID-19 pandemic terminology has permeated common language more than photovoltaics terminology.
3. Extensive coverage of Field I, especially of newly coined terms (e. g. *COVID-19*), is a feature of regularly (or at least recently) updated resources. OWID displays the most extensive coverage of terminology of Field I in the same resource, while Treccani.it distributes this terminology across different resources. The Open Multilingual Wordnet records different established terms from Field I and Field II in English and Italian but no terms of recent creation.
4. Some portals, e. g. Lexico.com, allow for a broader coverage of terminology in general dictionaries (cf. LEX1 and LEX2), whereas others, e. g. OWID, tend to record terminology in special[20] dictionaries (cf. OWID1–3). Elexiko, the general dictionary of the OWID network, does not record any of the selected terms.
5. OWID is the only portal which does not cover any of the terms of Field II.
6. The amount and type of lexicographic information provided in lexicographic entries varies from resource to resource. The minimal microstructure of dictionaries includes a definition and a number of corpus citations, whereas the Open Multilingual Wordnet, as a wordnet-type resource, concentrates on synset descriptions and synset glosses.
7. Explicit reference to the specialised usage of the terms, e. g. through subject labels, is infrequent.

3.1.2 Specialised language in hybrid lexicographic-lexicological networks

In the case of hybrid networks, two portals will be taken into consideration, namely BabelNet and ConceptNet. They are relatively complex in terms of the number of integrated resources and the number of languages covered. As mentioned above, both of

19 For a definition of *orthographic variation* cf. Giacomini (2019).

20 The distinction between general and special dictionaries is based on Engelberg/Lemnitzer (2004).

these resources are part of the (Linguistic) Linked Open Data ecosystem and therefore meet very stringent formal criteria. As a consequence, they are linked to a large number of further lexical resources, such as Wikipedia, Wiktionary, Wikidata, DBPedia, and even the Open Multilingual Wordnet. These resources contribute to the lexicological and encyclopaedic enrichment of lexicographic data. Coverage of the term sample in BabelNet and ConceptNet is shown in Table 5:

Table 5: Coverage of selected COVID-19 and photovoltaics terminology in hybrid lexicographic-lexicological networks: BabelNet (BBN) and ConceptNet (CCN).

	TERM (English)	COVERAGE	TERM (Italian)	COVERAGE	TERM (German)	COVERAGE
FIELD I: COVID-19	COVID-19	BBN, CCN	COVID-19	BBN	COVID-19	BBN
	SARS	BBN, CCN	contagio	BBN, CCN	Lockdown	BBN, CCN
	asymptomatic	BBN, CCN	letalità	BBN, CCN	Impfausweis	BBN, CCN
	sanitization	BBN, CCN	quarantena	BBN, CCN	Maskenpflicht	BBN
	disinfectant	BBN, CCN	linfedema	BBN, CCN	Quarantäne	BBN, CCN
	quarantining	BBN, CCN	droplet	BBN	Letalität	BBN, CCN
	nonessential	BBN, CCN	tampone	BBN, CCN	R-Zahl	–
	furlough	BBN, CCN	sierologico	BBN, CCN	durchimpfen	–
	immune system	BBN, CCN	paziente zero	BBN, CCN	Herden-immunität	BBN, CCN
	social distance	BBN, CCN	immunità di gregge	BBN	Social distancing	BBN
FIELD II: PHOTOVOLTAICS	solar film	–	pellicola solare	–	Solarfolie	–
	substructure	BBN, CCN	sottostruttura	BBN, CCN	Unterbau	BBN, CCN
	highly efficient	BBN, CCN	ad alta efficienza	–	hocheffizient	CCN
	photovoltaic	BBN, CCN	fotovoltaico	BBN, CCN	photovoltaisch	BBN, CCN
	voltage	BBN, CCN	voltaggio	BBN, CCN	Spannung	BBN, CCN
	photovoltaic system	BBN, CCN	impianto fotovoltaico	BBN	Photovoltaikanlage	BBN, CCN
	in-roof	–	sopra tetto	–	Indach-	–
	solar cell	BBN, CCN	cella solare	BBN, CCN	Solarzelle	BBN, CCN
	solar inverter	BBN	invertitore fotovoltaico	–	Solarwechselrichter	BBN
	renewable energy	BBN, CCN	energia rinnovabile	BBN, CCN	erneuerbare Energie	BBN, CCN

A strictly lemmatic approach was also used for surveying terminology in hybrid portals. Since the two resources under examination are based on a wordnet structure, the lemma searched for belongs to one or more synsets, each identifying a single concept. As in the previous section, orthographic variants have been admitted.

The results of the terminological analysis carried out on hybrid portals can be summarised as follows:

1. Both portals record single-word and multi-word terms as lemmas.
2. In most cases, coverage through BabelNet and ConceptNet coincide. BabelNet has the most extensive coverage of the two portals.
3. Coverage of Field I is not noticeably larger than of Field II. Only nine terms are not recorded by either of the portals.
4. The amount and type of lexicographic information provided within the entries varies from resource to resource but also from term to term (see also Section 3.2.2). Entries are not to be understood in the traditional lexicographic sense. The usual entry structure in BabelNet includes a synset, a WordNet gloss, definitions from different resources, relations, sources, translations into a target language, and a network with semantically related forms and gloss-related forms. ConceptNet provides, among others, synonyms, related terms, hyponyms and hypernyms, antonyms, location, derived terms, word forms, in different languages, as well as links to other resources. The number of indications for each searched term is variable (cf. the difference between the maximal description for *disinfectant* and the minimal description for *social distance*).
5. Explicit reference to the specialised usage of the terms is typically given through a subject label and mentioned in the definitions of BabelNet, and can be inferred from different indications in ConceptNet.

3.1.3 General remarks

The main findings of the previous analysis will now be summarised in light of the similarities and differences between lexicographic network and hybrid network portals. Figure 1 displays an overview of the terminological coverage in the portals under discussion:

It should be noted once again that this is primarily a qualitative analysis, as the amount of data is too small to allow for quantitative generalisations. A look at this kind of data, however, can be useful to get a first glimpse of the differences found between the portals.

From the point of view of the terminology of the two different domains, it is evident that the coverage of virology terms related to the COVID-19 pandemic is very high in all portals and in all languages, with the only clear exception the Open Multilingual Wordnet portal, which does not record the most recently introduced and used terms. This outcome highlights the importance of regular updates of lexical resources on the web, as also highlighted by Fuertes-Olivera/Tarp (2014: 134).

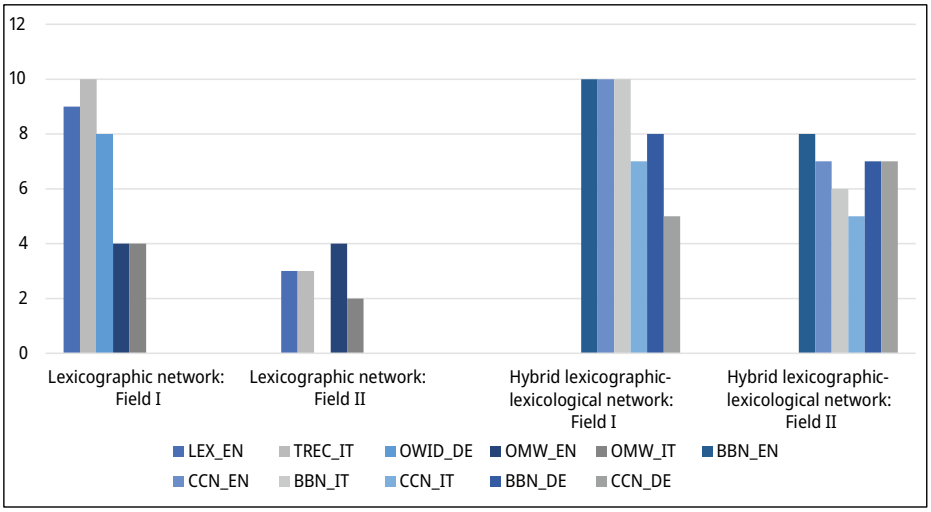

Figure 1: Overview of terminological coverage of Field I and Field II in lexicographic networks and hybrid lexicographic-lexicological networks: LEX_EN (Lexico.com for English), TREC_IT (Treccani.it for Italian), OWID_DE (OWID for German), OMW_EN and OMW_IT (Open Multilingual Wordnet for English and Italian), BBN_EN, BBN_IT and BBN_DE (BabelNet for English, Italian and German), CCN_EN, CCN_IT and CCN_DE (ConceptNet for English, Italian and German).

Terms from the field of photovoltaics, which are less recent and correspond to different communicative levels, are clearly more prominent in the hybrid portals. They therefore turn out to be more homogeneous in their coverage of terminology.

From the point of view of the three languages analysed, unexpectedly the difference is not substantial and indeed is cancelled in hybrid portals such as BabelNet, showing a high degree of alignment between the multilingual resources and enabling, as we have seen, a search in different languages.

For each language, a total of 20 terms were searched for, including several morphologically complex terms. Table 6 summarises the relative coverage of the two different types of terms depending on the portal, independent of the specialised field.

Table 6: Field-independent coverage of simple and complex terms in lexicographic networks and hybrid lexicographic-lexicological networks.

	English terms (20):		**Italian terms (20):**		**German terms (20):**	
	simple	**complex**	**simple**	**complex**	**simple**	**complex**
Lexicographic networks	9 / 7	5 / 13	7 / 7	6 / 13	1 / 4	7 / 16
Hybrid lexicographic-lexicological networks	7 / 7	11 / 13	7 / 7	9 / 13	3 / 4	13 / 16

Among complex terms, both compounds (e. g. *photovoltaic*, *linfedema*, *Maskenpflicht*) and multiword terms (e. g. *solar film*, *paziente zero*, *erneuerbare Energie*) were counted. The distribution of simple and complex terms in our term sample is random and therefore not representative of the terminology of the two chosen fields. However, it is indicative of the extent to which the different portals are able to record morphologically different terms in the three languages under consideration.

The portals under analysis are primarily concerned with the common language. They can therefore be defined as generalist portals. However, they seem to also contain a considerable number of terms with varying degrees of specialisation, not necessarily belonging to a non-expert communicative level.

3.2 Conceptual knowledge representation in LGP lexicographic portals

The exploration of terminology coverage in portals dedicated to the common language cannot be separated from the observation of the conceptual coverage. In both traditional Wüsterian terminology (Wüster 1991) and more recent approaches to terminology, such as Sociocognitive terminology (Temmerman 2001), Termontography (Temmerman/Kerremans 2003), and Frame-Based Terminology (Faber 2015), the idea of a connection between the terms and concepts of a domain is a fundamental assumption of both terminology work and terminographic practice.

In our analysis, it will not always be possible to distinguish the lexical level of terms from the conceptual level, not least because the concepts in question are themselves always lexicalised.

While lexicographic collections are not meant to offer a direct, joint semasiological or onomasiological access to the linked dictionaries, lexicographic search engines, on the other hand, may provide this option. This is, for instance, the case of OneLook, which also enables a thesaurus search. This is an onomasiological search that allows access to words related to the search word, mostly through synonymy relations. Although these are primarily lexical relations, it is also possible to consider them as part of a shallow and very simple conceptual system.

As in the previous section, we will concentrate on lexicographic networks and hybrid lexicographic-lexicological networks, assessing the extent of the conceptual data, the underlying descriptive models and the access possibilities made available to the user. The extension of conceptual data concerns both the quantity and granularity. Descriptive models can range from hierarchical models (taxonomies) to ontologies. We also include semantic relations (e. g. synonymy) specific to wordnets. The type of access to conceptual systems can also vary from external access (onomasiological) to internal access based on the microstructure of the resources of a portal.

The analysis of conceptual knowledge representation will be carried out on some of the portals presented in Section 3.1 with reference to the same specialist domains as previously discussed.

3.2.1 Conceptual knowledge representation in lexicographic networks

The degree of integration of resources in OWID is very high, but their structures are very different. The portal does not provide a unified conceptual system, but some of the dictionaries contain their own conceptual system.

Elexiko, the general dictionary, allows at the microstructure level, and specifically at the meaning indication level, access to semantically related words, linked by semantic relations such as hypernymy and hyponymy, synonymy and antonymy, or partonomy. Further relations covered by this resource are cause, condition, semantic gradation, modal opposition, and consequence (e. g. resultative and responsive relations). The coverage of lexical relations is extensive but it can be argued that conceptual data have not been systematically organised.

In OWID's dictionary of neologisms, the terms are attributed to twenty-one subject and specialist areas (e. g. employment and education, society, politics, banking and finance, food, sport) as well as to possible thematic groups within these areas (e. g. labour market, family, demography, communication). Some examples of German terms with their conceptual classification are given below:

SARS
Subject and specialist area: GESUNDHEIT/KÖRPERKULT (HEALTH/BODY CULTURE)

Disease-Management-Programm
Subject and specialist area: GESUNDHEIT/KÖRPERKULT (HEALTH/BODY CULTURE)
Thematic group: GESUNDHEITSWESEN (HEALTH CARE)

The model used for the representation of conceptual data is thematic and very shallow, displaying a maximum of three descriptive levels. The access to the conceptual data has a peculiarity: it does not take place externally from the home page of the dictionary (even the complex search does not provide onomasiological options), but neither does it take place internally through the microstructure. No mention is made of the subject and specialist areas within the dictionary entries. The conceptual system underlying this resource is only accessible via a separate page, entitled "Inhaltlich gruppierte Stichwörter" ('content-grouped words'), to which terms and concepts can be consulted, displayed in tabular form.

Treccani.it contains a conceptual system that constitutes the knowledge base of the encyclopaedic resource. This conceptual system revolves around hierarchically ordered themes called categories. These categories are also linked to a series of specific thematic-conceptual tags assigned to each entry. In the online encyclopaedia, for example, the Italian term *virus* is associated with the following conceptual elements:

virus
Tags:
APPARATO CARDIOCIRCOLATORIO (CARDIOVASCULAR SYSTEM),
SISTEMA NERVOSO CENTRALE (CENTRAL NERVOUS SYSTEM),
BARRIERA EMATOENCEFALICA (BLOOD-BRAIN BARRIER),
DETERMINANTE ANTIGENICO (ANTIGENIC DETERMINANT),
INFORMAZIONE GENETICA (GENETIC INFORMATION)
Categories:
MICROBIOLOGIA in Biologia (MICROBIOLOGY in Biology),
TEMI GENERALI in Informatica (GENERAL TOPICS in Information Technology)

The tags are links to further entries in the encyclopaedia, while the categories lead to pages containing the concepts relating to a certain conceptual category, presented in alphabetical order. The MICROBIOLOGY category includes the concepts (and terms) *Acetobacter*, *Actinomiceti*, *adenovirus*, *Aerobacter*, *aerobio*, etc. MICROBIOLOGY, as indicated above, is in turn a subcategory of BIOLOGY.

The conceptual system that has been described is relatively shallow, as it does not appear to exceed three levels of description (e. g. BIOLOGY > MICROBIOLOGY > adenovirus). It is not a taxonomy, as it does not only provide for hierarchical relations, rather all available relations can be described as thematic.

This type of conceptual structure is only visible in the online encyclopaedia, whereas it is not in the other resources of the portal. Moreover, this structure does not allow the user external access, i. e. access at the macrostructural level, but only internal access from individual entries. The usability of the conceptual data is therefore extremely limited.

The Open Multilingual Wordnet has, by its nature, a conceptual organisation of data, realised through synsets. The structure of the data is typical for wordnets: each synset collects lexical data linked together by synonymy and uniquely identifies a concept. Access to the data is therefore both semasiological and onomasiological. An example for conceptual coverage through synsets in the OMW will now be mentioned:

substructure (English, Italian)

03570372-n	***substructure**, infrastructure* - *infrastruttura*	the basic structure or features of a system or organization Relations: Hypernym: *structure* Holonym-Part: *system* Semantic Field: *artifact*$_n$ SUMO: Artifact

03387016-n	*base, foot, foundation, groundwork, fundament,* ***substructure****, understructure - piedistallo, piedestallo, base, sottofondo*	lowest support of a structure Relations: Hyponym: *bed raft foundation* Hypernym: *support* Holonym-Part: *structure* Semantic Field: $artifact_n$ SUMO: Region

The conceptual system is in fact a network, with multiple descriptive levels and a primarily hierarchical character. In addition to specific lexical relations (e. g. hypernymy, hyponymy, holonymy, and similarity), the assignment to a semantic field is also indicated. Finally, an important contribution to the conceptual data is provided by the link to SUMO, i. e. to categories of a top ontology.

3.2.2 Conceptual knowledge representation in hybrid lexicographic-lexicological networks

BabelNet and ConceptNet are also structured primarily as wordnets, and therefore have the same basic characteristics as the Open Multilingual Wordnet for what concerns the representation of conceptual data. However, conceptual systems in hybrid portals are supposed to be more extensive. If we look at the entry of the English term *virus* (virology), we find the following data on semantic relations in BabelNet:

virus
IS A infectious agent, microorganism, taxon, pathogen
HAS PART genome
HAS KIND animal virus, arbovirus, bacteriophage, plant virus, slow virus, ...
HAS INSTANCE HIV, Titi monkey adenovirus, JX-594, Schmallenberg orthobunyavirus, ...
CODE OF NOMENCLATURE International Code of Virus Classification and Nomenclature
DESCRIBED BY SOURCE Otto's encyclopedia
HAS EFFECT viral infection, viral pneumonia
HAS QUALITY viral life cycle
ON FOCUS LIST OF WIKIMEDIA -
PARENT TAXON biota
SAID TO BE THE SAME AS viral pathogen
TAXON RANK kingdom

The number of relations in BabelNet is very large and includes many other types besides those relevant to *virus* in virology, e. g. DIFFERENT FROM, NAMED AFTER, HAS CAUSE or HAS EFFECT. Many relations are extremely specific and only apply to certain terms or concepts: e. g. *virology* IS PRACTISED BY *virologist*, COVID-19 HAS LOCATION OF DISCOVERY *Wuhan*, *herd immunity* IS STUDIED BY *immunology*, *voltage* IS CALCULATED BY *electric potential* and IS FACET OF *electricity*. It is therefore evident that there are not only hierarchical but also other types of relationships. The conceptual system underlying BabelNet is ontological in nature and the granularity of the concepts appears to be highly marked.

Conceptual data in BabelNet are not limited to the written specification of the relations relating to a term (and a synset) but are also graphically made available through the 'Explore network' function. This function provides graphical access to the concepts, including an image and the gloss (definition). For a given term, the network displays semantically related forms, monosemous gloss related forms, as well as disambiguated gloss related forms, whereby each of the three categories is marked by a different colour. Conceptual data in BabelNet can be externally and internally accessed thanks to the semasiological and onomasiological structure of the portal.

ConceptNet contains a wide set of relations, but this is not as flexible as in BabelNet. For the English term *virus*, ConceptNet provides the following relations for all meanings of the term (no distinction is made at first between the medical and the IT domain):

virus
SYNONYMS: French: virus informatique, Portuguese : vírus, Purtuguese: vírus de computador, ...
TYPES OF VIRUS: AIDS, the flu, animal virus, ...
RELATED TERMS: antivirus, viral, disease, ...
DERIVED TERMS: adenovirus, antivirus, bacteriovirus, ...
VIRUS IS A TYPE OF: a program that replicates, malevolent program, ...
VIRUS IS CAPABLE OF: cause a cold, infect, destroy cells, ...
TERMS WITH THIS CONTEXT: inoculant, rheumatoid arthritis, ...
CONTEXT OF THIS TERM: virology, archaic, computing, ...
ETYMOLOGICALLY RELATED: arbovirus, bracovirus, capripoxvirus, ...
ETYMOLOGICALLY DERIVED TERMS: -
EFFECTS OF VIRUS: a disease, infection, a cold
WORD FORMS: virii, virus, viruses
LOCATION OF VIRUS: your computer, a body
THINGS CREATED BY VIRUS: sickness
THINGS WITH VIRUS: people get sick when they (HasA virus)
THINGS THAT DON'T WANT VIRUS: a person
PARTS OF VIRUS: a capsid
VIRUS WANTS: live
ETYMOLOGICALLY ROOTS OF VIRUS: virus

PROPERTIES OF VIRUS: lethal
VIRUS NotHasProperty: alive
VIRUS CAN BE: used as a weapon
DISTINCT TERMS: bacteria

This is a continuum of relationships ranging from the purely lexical (e. g. word forms) to the semantic-conceptual (e. g. parts of virus) to the conceptual (e. g. location of virus). The two subentries *virus* (n, communication) and *virus* (n, virology) are also available, each with a restricted number of relations. The conceptual system of this portal extends to linked external resources such as DBPedia and WordNet. The conceptual system is of an ontological nature, although with a lower degree of granularity, since conceptual relations are generally more loosely formulated. As in BabelNet, the semasiological access to conceptual data is at the same time onomasiological access.

3.2.3 General remarks

In this section, the conceptual coverage of lexicographic portals of the network and hybrid-network type was analysed. Of the portals observed in Section 3.1, only Lexico.com does not match any type of conceptual structure. Table 7 summarises the results of the analysis from the point of view of the three aspects under consideration: extent of conceptual data, conceptual data model, and access to conceptual data.

Table 7: Conceptual systems in lexicographic networks and hybrid lexicographic-lexicological networks.

	Conceptual data extent	**Conceptual data model**	**Conceptual data access**
Lexicographic networks:			
- OWID Elexiko	relatively high number of hierarchical and non-hierarchical relations	not systematic	internal access
Dictionary of neologisms	restricted number of relations	shallow thematic network	outer access through separate page
- Treccani.it	undefined number of relations, mostly hierarchical	shallow thematic network	internal access
- OMW	restricted number of relations, mostly hierarchical	lexical network + link to SUMO	external/internal access, semasiological/ onomasi-ological access

Table 7 (continued)

	Conceptual data extent	**Conceptual data model**	**Conceptual data access**
Hybrid lexicographic networks:			
- BabelNet	high number of hierarchical and non-hierarchical relations	ontology	external/internal access, semasiological/onomasi-ological access
- ConceptNet	relatively high number of hierarchical and non-hierarchical relations	ontology	external/internal access, semasiological/onomasiological access

This summary clearly shows the most significant differences between the various portals, but also a certain homogeneity between more lexicographic portals, for which a thematic structure may be the most appropriate, and portals based on a wordnet model, based on more or less complex ontological structures. In general, the type of conceptual relations is not purely hierarchical and therefore does not fit into taxonomies. There is also a correspondence with the specific type of portal as far as data access is concerned. Portals based on a wordnet model are extremely flexible, whereas lexicographically oriented portals tend to provide unique and therefore more limited access possibilities.

4 Conclusions

The analysis of the coverage of specialised terminology in generalist lexicographic portals was intentionally combined with the analysis of the coverage of conceptual knowledge. This makes it possible to assess the extent to which these portals may be complementary or even analogous to specialist lexicographic resources. As mentioned in the introduction, contemporary portals are overwhelmingly common language or generalist, while examples of LSP portals, such as the Croatian Terminology Portal, are sporadic. It can be assumed that a specialised lexicographic portal would reach a much smaller user base and may therefore be considered unprofitable in terms of design and implementation.

Nevertheless, terminology and terminographic work, as well as specialised translation and technical writing, constantly require reliable resources dedicated to terminology, in the absence of which it is necessary to resort to alternative instruments such as parallel texts, translation memories, and raw data extracted from specialised corpora.

The results of the analysis presented in this contribution show that general-language or generalist lexicographic portals can also be a useful tool for terminologists, translators and technical writers. This is especially true of hybrid lexicographic and

lexicological portals, which display both a high level of terminological coverage and detailed conceptual systems.

Against this background, it is possible to envisage the future emergence of new monolingual and multilingual lexicographic portals which, a) on the one hand, reflect this generalist profile, associating common language data with increasingly detailed terminological data relating to multiple domains, and, b) on the other hand, develop (ontological) conceptual systems which, by virtue of their complexity, can best reflect the various potential relations occurring between concepts as well as between domains.

5 Bibliography

5.1 Monographs and articles

Engelberg, S./Lemnitzer, L. (2004): *Lexikografie und Wörterbuchbenutzung*. 4th edition. Tübingen: Stauffenburg.

Engelberg, S./Storrer, A. (2016): Typologie von Internetwörterbüchern und -portalen. In: Klosa, A./Müller-Spitzer, C. (eds.): *Internetlexikografie. Ein Kompendium*. Berlin/Boston: De Gruyter, 31–63.

Faber, P. (2015): Frames as a framework for terminology. In: *Handbook of terminology* 1:14, 14–33.

Fuertes-Olivera, P.A./Tarp, S. (2014): *Theory and Practice of Specialised Online Dictionaries*. Berlin/Boston: De Gruyter.

Giacomini, L. (2019): *Ontology – Frame – Terminology. A method for extracting and modelling variants of technical terms*. Habilitation Thesis, University of Hildesheim.

Klosa, A./Gouws, R.H. (2015): Outer features in e-dictionaries. In: *Lexicographica* 31, 142–172.

Leroyer, P. (2011): Change of Paradigm in Lexicography: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools. In: Fuertes-Olivera, P.A./Bergenholtz, H. (eds.): *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum, 121–140.

Temmerman, R. (2001): Sociocognitive terminology theory. In: Cabré, M.T./Feliu, J. (eds.): *Terminología y Cognición*. Barcelona: Universitat Pompeu Fabra, 75–92.

Temmerman, R./Kerremans, K. (2003): Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description. In: *Proceedings, XVII. International Congress of Linguists. Prague, Czech Republic*, 1–10.

Wang, W. (2001): Zweisprachige Fachlexikographie: Benutzungsforschung, Typologie und mikrostrukturelle Konzeption. Frankfurt am Main et al.: Peter Lang.

Wiegand, H.E. (1989): Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In: Hausmann, F.J./Reichmann, O./Wiegand, H.E./Zgusta, L. (eds.): *Wörterbücher: Ein internationales Handbuch zur Lexikographie.* Berlin/New York: De Gruyter, 409–462.

Wüster, E. (1991): *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. 3rd edition. Bonn: Romanistischer Verlag.

5.2 Lexicographic portals

BabelNet. [<https://babelnet.org>; last access: October 10, 2021].

CLARIN glossaries. [<https://www.clarin.eu/resource-families/lexical-resources-glossaries>; last access: October 10, 2021].

ConceptNet. [<https://conceptnet.io>; last access: October 10, 2021].
Diccionarios.com. [<https://www.diccionarios.com>; last access: October 10, 2021].
DWDS = Digitales Wörterbuch der deutschen Sprache. [<https://www.dwds.de>; last access: October 10, 2021].
Hrvatski terminološki portal. [<http://nazivlje.hr>; last access: October 10, 2021].
Lexico.com. [<https://www.lexico.com>; last access: October 10, 2021].
MultiWordNet. [<https://multiwordnet.fbk.eu>; last access: October 10, 2021].
OneLook Dictionary Search. [<https://www.onelook.com>; last access: October 10, 2021].
Open Multilingual Wordnet. [<http://compling.hss.ntu.edu.sg/omw>; last access: October 10, 2021].
OWID = Online-Wortschatz-Informationssystem Deutsch. [<https://www.owid.de>; last access: October 10, 2021].
Svenska Akademiens Ordböcker. [<https://svenska.se>; last access: October 10, 2021].
TermCoord Glossary links. [<https://termcoord.eu/glossarylinks>; last access: October 10, 2021].
The European Dictionary Portal. [<http://www.dictionaryportal.eu>; last access: October 10, 2021].
The Free Dictionary. [<https://www.thefreedictionary.com>; last access: October 10, 2021].
Treccani.it. [<https://www.treccani.it>; last access: October 10, 2021].
Wortschatz Leipzig. [<https://wortschatz.uni-leipzig.de>; last access: October 10, 2021].