

Ana Ostroški Anić

Institute for Croatian Language and Linguistics, Zagreb, Croatia
aostrosk@ihjj.hr

Definitional Patterns in Specialized Resources for Schoolchildren

Abstract

The paper describes the analysis of frequent definitional patterns in two specialized corpora, English and Croatian, compiled of educational texts in the field of climate change written for schoolchildren and young adults. The analysis is based on extracting knowledge-rich contexts from the corpora, followed by determining common lexical markers used to identify definitions, lexical knowledge patterns, and lexical-semantic relations between terms. Definitions and definitional patterns will be used in developing a children-oriented dictionary of climate change in Croatian.

Keywords: definitions, knowledge patterns, knowledge-rich contexts, lexical markers, specialized corpora

1. Introduction

When schoolchildren first encounter novel concepts in any specialized domain, e.g., in the field of biology or chemistry, they are often presented with a two-fold challenge: first, they need to understand the meaning of the concepts and put them into relation with previously acquired knowledge, and second, they need to memorize their definitions in order to be able to reproduce them later when needed. As expected, educational experts creating specialized resources for children must be aware of different linguistic skills children possess at different age, as well as their general cognitive development and categorization skills, which often hinder their understanding of very abstract concepts (Casadevante et al. 2019).

The issue of climate change and all related phenomena have been vastly discussed in all aspects and domains of modern life from various perspectives. The topic is therefore well known to children, as climate change topics are also part of the curricula of school subjects (e.g., science and geography), and are even approached in activities for pre-school children. However, climate change educational materials in Croatian are still very scarce, and mostly designed for teachers as additional instructions in preparing classroom activities.

This paper describes the process of identifying the most frequent definitional patterns in two specialized corpora, English and Croatian, consisting of educational

texts in the field of climate change written for schoolchildren and young adults. The method used to identify and formulate definitional patterns is corpus-driven, and it encompasses both the analysis of the lexical level of terms, the lexical markers used to identify definitions, and the lexical-semantic relations between terms, as well as the conceptual analysis applied to identify concept characteristics expressed in definitions. The motivation for the research is explained in the next section, followed by the detailed presentation of the methodology applied in the compilation of corpora, their analysis, and the annotation and analysis of knowledge patterns in the extracted definition. After the discussion of the definitional patterns in the analysed examples, a few recommendations for future research are given in the final section.

Results will be used in preparing material for experimental research investigating children's categorization development, especially in the context of developing classification schemes for children, as mentioned by Murray and Reuter (2005). In practical terms, definitions and definitional patterns will be used in developing a children-oriented dictionary of climate change in Croatian.

2. The Importance of Context in Children's Categorization

When creating educational or instructional resources for children, whether those are textbooks or leaflets and other similar short text formats used to educate children, it is important to consider the structure and format of the resource, as well as the language used to introduce and define new concepts. Using clear and age-appropriate language, e.g., formulating ideas in shorter sentences and avoiding the use of complex terms, ensures that both content and form resonates with children's language development and cognitive abilities. Murray and Reuter (2005) argue that many experimental researchers in children categorization mistake a difference in language use for a difference in ability to understand categories, whereas children can be fluent in less structured language, but not in applying adult-like categories (2005: 9). The ability to comprehend abstract concepts falls within that domain.

Research shows that children aged six to nine acquire abstract concepts to which they are more emotionally connected (Vigliocco et al. 2018). The emotional experience remains an important factor in concept acquisition even beyond the critical age, alongside the engagement of the sensorimotor system (Reggin et al. 2021). Moreover, Gelman and Meyer (2011) emphasize the contextual nature of children's categorization processes, indicating that the surrounding context strongly influences how they categorize information. Previous research has already shown that context is the dominant distinguishing element in eliciting conceptual relations by abstract and concrete concepts. Caramelli et al. (2004) report on the different development of concrete and abstract conceptual knowledge in children aged eight, ten, and twelve, i.e., on the different pattern of information elicited by concrete and

abstract concepts in schoolchildren. Their findings confirm that the conceptual information specific for abstract concepts refers to situations and events in which they are experienced, as opposed to concrete concepts that convey mostly information on the properties of the objects they refer to (2004: 31). As opposed to a wider range of conceptual relations elicited by concrete concepts, abstract concepts elicit thematic information that is contextually dependent. These results, as many other experimental findings, lead to doubts on the suitability of conventionally formulated definitions, e.g., intensional terminological definitions, in resources designed for schoolchildren.

3. Knowledge-Rich Contexts and Knowledge-Driven Approach to Definitions

The concept of knowledge-rich context is a well-known construct in terminology literature.¹ Starting from the Meyer's seminal paper in 2001, the idea of a knowledge-rich context (KRC) as a linguistic context providing information useful for conceptual analysis has seen numerous applications both in practical terminology work and in theoretical investigations. In order to detect these contexts, knowledge patterns are applied as recurrent linguistic structures "indicative of semantic relations" (Barrière 2004). Knowledge patterns usually include two elements linked by a relationship, e.g., a concept and one of its characteristics, a concept and its superordinate concept, etc. On the linguistic level, they are realized by terms or other linguistic expressions and a marker of their relationship (Marshmann 2006). Marshmann investigated lexical knowledge markers, which she defined as markers that take the "form of a lexical unit or sequence of lexical units" (2006: 2) in knowledge patterns for the cause-effect and association relations in medical texts. For our analysis, both relations are relevant as they often relate to the conceptual relations of events and activities, specific for the climate change domain.

A more recent example of using KRCs in text-mining and terminological analysis is Ramos and Costa (2023), who extracted definitional contexts as knowledge-rich contexts to establish lexical markers, lexical-semantic relations between terms, and interpret the term relations, which was followed by a conceptual analysis. A similar procedure of extracting definitional information from specialized corpora of unstructured texts (Fišer et al. 2010) revealed uncertainty in distinguishing between definitions and non-definitions, especially in texts where new concepts are introduced and described in many different manners (2010: 2934). The same was expected to be a distinguishing feature of semi-specialized texts prepared for children, which formed the corpora designed for this study.

¹ For an overview of the use of knowledge-rich contexts in terminology and knowledge engineering, see Condamines (2022).

When Croatian terminological literature is concerned, Grčić Simeunović and Vintar presented a list of sixteen lexico-syntactic knowledge markers to extract definition candidates from the corpus on karst (2015: 257–258). The markers are roughly grouped into three categories: those that represent typical intensional definitions (e.g., [noun in Nominative] + [verb *be*] + [noun in Nominative]), markers comprising structures used to introduce new concepts, including the verbs *represent*, *define*, and finally, markers that are specific for the analysed domain, e.g., [noun] + *process*, *consists of* + [noun in Genitive]. Grčić Simeunović and Vintar (2015) applied the lexico-syntactic patterns in order to identify definitions and extract knowledge-rich contexts from the corpus. The aim of this paper could be regarded as quite opposite: to start from definitional and knowledge-rich contexts to determine lexico-syntactic markers used in defining concepts of climate change and identify common knowledge patterns in definitional contexts.

4. Methodology

In order to identify the patterns of definitions in educational resources on climate change for schoolchildren, the analysis was carried out both in English and Croatian. Given the overall dominance of English materials in the domain, we started from English as the dominant language, and then applied the same methodology to Croatian. Two small, specialized corpora were initially compiled, both consisting of popular educational or instructional resources written for schoolchildren as the target audience. We understand popular educational resources to include all popular scientific books as well as glossaries, brochures or leaflets, which excludes textbooks used as the primary teaching source in regular school settings. Since the analysis described here presents the preliminary work that is part of a larger study, we focus on instructional materials only. The English corpus consists of 421,756 tokens. The lack of resources for the population of schoolchildren and young adults is evident in the size of the Croatian corpus, which consists of only 123,905 tokens.

First, lists of keywords and terms were extracted from both corpora. The list of keywords (or single unit terms) did not yield many results due to too much noise in the corpus because of the varieties of resources included, some of which are glossaries and not actual running texts. Therefore, we decided to focus on the list of terms to choose the candidates for the list of seed words for corpus querying. Additionally, as a type of validation, we manually analysed two popular educational resources in English on the topic of climate change, *Climate glossary for young people* (Cognuck González and Numer 2020) and *What's the issue? Climate change* (Jackson and Guitian 2020) to make sure that certain key concepts were not omitted from the list of candidates. The Croatian translation of Jackson and Guitian (2020) was analysed as well.

Having analysed the first one thousand term candidates from the corpus, we obtained the list of terms used for extracting definitions. We decided to omit the

terms for concepts that are not limited to the domain of climate change as well as those that are borderline with words in general language or those that, by being the core terminology of science, can be defined from several perspectives, e.g., *water, oxygen, atmosphere, polar ice*, etc. A list of forty-four key terms was eventually compiled on the basis of terms validated in other resources, e.g., *acid rain, air pollution, biodiversity, carbon footprint, climate, climate action, fossil fuel, global warming, renewable energy*, etc.

Definitions and knowledge-rich contexts of these key concepts have been extracted and first divided according to the type of the category of the *definiendum*, i.e., into events, activities, and entities, which were further categorized into concrete and abstract concepts. The same procedure was then applied to the Croatian corpus, starting from the Croatian terms for the forty-four English key concepts identified in the English corpus. Certain terms did not yield any definitions or knowledge-rich contexts (e.g., *klimatske akcije, klimatski uvjeti, klimatska pravda*, and *klimatska politika* in the Croatian corpus), and were therefore omitted from the final list. Being a much larger corpus, we had anticipated that the English corpus was to yield a more representative and interesting corpus examples.

Having extracted all relevant definitions and defining contexts, we analyzed the examples of verb patterns, as well as nominal and adjectival phrases in the sentences. All examples were then annotated for lexical markers, lexical knowledge patterns, and lexical-semantic relations expressed in the sentences in the same manner for English and Croatian (Appendix 1 and Appendix 2). Strategies used to define the concepts are then discussed according to the type of definition applied. A common strategy, especially in defining activities or processes, is to describe the cause or the most prominent aspect of the process, rather than rely on the intensional definition as the prototypical terminological definition. Different definitional patterns are then presented according to the types of concepts they identify, with the aim of suggesting definition templates, including various types of terminological definitions.

5. Results and Discussion

Given the register, style, and format of the texts comprising both corpora, especially the English one, it is difficult to determine whether certain extracted contexts could be classified as definitional contexts or as proper definitions. To illustrate this issue, we provide three extracted examples describing the concept of climate change:

- (1) Climate change is the global climate variation of the earth.
- (2) The adverse effects of climate change have resulted in impacts that people have not been able to cope with or adapt to, and that can lead to loss or damage, such as loss of biodiversity and ecosystem services, loss of income and damage to infrastructure.

(3) Climate change may be caused by natural internal processes or by external forces, such as volcanic eruptions or persistent anthropogenic actions.

Example (1) is the only example of the three that could be considered an intentional definition, modelled according to the traditional Aristotelian principle of defining a concept by referring to a superordinate concept and its delimiting characteristics. Regardless of its structure, it is still not the best example since both nouns, *change* and *variation*, are abstract nouns, and could be considered synonyms. Example (2) is an evident case of a knowledge-rich context because it contains conceptual information relevant for placing the concept in relation to other concepts in the network. However, it is not a proper definition as it elaborates on the effects of climate change, rather than defining climate change itself. Sentences like this one are classic examples of describing a concept by putting it in a broader context after it is initially identified. Example (3), on the other hand, although it does not define what climate change is, explains the causes of it and could therefore be considered a definitional context.

There is an overlap in the understanding of knowledge-rich contexts and definitional contexts. Having already defined KRCs as linguistic contexts that provide information useful for conceptual analysis, we understand *definitional context* in a narrower sense, as the “discursive context where relevant information to define a term could be found” (Sierra et al. 1998: 77). In unstructured texts such as popular scientific resources, it is a valid procedure to identify a specialized concept by a definitional context that provides the concept characteristics necessary to understand it.

In order to determine the most frequent definitional patterns used in educational resources on climate change, we first started from the English corpus. Appendices 1 and 2 list common lexical markers used to link the defined concept to other related concepts or its characteristics, among which the verb *be* is most often found in lexical knowledge patterns both in English and Croatian extracted examples. We distinguish between lexical markers and knowledge patterns in the sense that lexical markers are used as linguistic signals that introduce terms and/or link them to other terms or linguistic elements in the sentence, whereas lexical knowledge patterns are linguistic structures pointing to different conceptual or semantic relations and are therefore relevant for the conceptual analysis. Lexical markers are linguistic expressions, e.g., *is called*, *identified by*, *caused by*, or *attributed to*, while knowledge patterns are expressed as formulas, where NP stands for noun phrases expressing terms. Lexical-semantic relations expressed by the knowledge patterns are also identified for both languages, while the fourth column in each appendix contains examples of definitional contexts extracted from corpora.²

As is expected, the most typical lexical-semantic relation expressed between the two terms in definitional contexts in both languages is the relation between a

² Lexical-semantic relations are named after those in Marshmann (2006), whereas the lexical knowledge patterns are defined following Grčić Simeunović and Vintar (2015).

hyperonym and a hyponym, as in the following definition of *fossil fuels*, where *energy sources* is the hyperonym of *fossil fuels*:

(4) Fossil fuels are energy sources that are generated when plant and animal matter biodegrades.

The lexical-semantic relation between a term and its superordinate term is in the English corpus lexically signalled by markers such as *also known as*, *is called*, *refer to*, and by a not so common but interesting marker *should be seen as*. In addition to *is called* and *refer to*, the Croatian equivalents of markers for the same lexical-semantic relation include *represents* and the variants *the name for* and *the term for*.

As Table 1 shows, the *cause-effect* relation is another commonly detected lexical-semantic relation in the English data, as is *synonymy* and the *association* between the two terms expressed in a sentence. Using exemplification—signalled here by the lexical marker *such as*, but often also by markers *for example* or *that is*—is a useful approach used to further describe the defined concept in intensional definitions by means of enumerations (Nilsson 2015), such as in Example (5):

(5) Climate change may be caused by natural internal processes or by external forces, such as volcanic eruptions or persistent anthropogenic actions.

When enumeration is used as the governing principle to describe the concept, as in the definition “The three key greenhouse gases in our atmosphere are methane, water vapor, and CO₂” it is no longer the case of an intensional definition.

Table 1. Lexical markers and knowledge patterns used in English definitional contexts

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation
also known as	[NP] [be] also known as [NP]	synonymy hyperonymy/hyponymy
attributed to	[NP] [be] attributed to [NP]	association
is (a/the), are (a/the)	[NP] [be] ([article]) [NP] (of)	hyperonymy/hyponymy
caused by	[NP] [be] caused by [NP] [NP] [can/may] [be] caused by [NP]	cause-effect
is called, can be called	[NP] ([can]) [be] called [NP]	hyperonymy/hyponymy
due to	[NP] [be] due to [NP]	cause-effect
identified as/by, can be identified as/by	[NP] ([can]) [be] identified as/by [NP]	association
made from	[NP] [be] made from ([article]) [NP]	result
mean	[NP] means [NP]/ [VP]	cause-effect
refer to (as)	[NP] refers to [NP] (as)	hyperonymy/hyponymy
should be seen as	[NP] should be seen as [NP]	hyperonymy/hyponymy

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation
such as	[NP] ([punctuation mark]) such as [NP]	exemplification
that is	[NP] [punctuation mark] that is ([comma]) [NP]	synonymy

Although the Croatian corpus is significantly smaller than the English one, and not all knowledge-rich contexts of the seed terms could have been used for the extraction of definitional contexts, it nevertheless provided a significant number of lexical markers useful for reaching conclusions about the dominant lexical-semantic relations present on the term level. Apart from the mentioned relations of *hyperonymy* and *hyponymy*, as well as *synonymy* and the relation of *association*, the *causative* relation is interesting because of the variation in knowledge patterns, i.e., the valence patterns of the verbs used as lexical markers. Because the spatial preposition *do* ‘by’ in Croatian requires the Genitive case of nouns, the phrasal verbs *dovesti do* ‘lead to’ and *doći do* ‘occur’ always stand before a noun in Genitive, while the verb *uzrokovati* ‘cause’ asks for a noun in Nominative. Table 2 lists other markers and knowledge patterns, some of which also have variants, such as those that have the verb *be* in the copulative position.

Table 2. Lexical markers and knowledge patterns used in Croatian definitional contexts

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation
biti ‘be’	[NP] [biti] [NP]	hyperonymy/hyponymy
činiti ‘make’	[NP] čine [NP]	meronymy
dolazi do ‘occur’	do [NPGen] dolazi zbog [NPGen] do [NPGen] dolazi jer [NPNom]	cause–effect
dovesti do ‘lead to’	[NP] [dovesti] do [NPGen]	cause–effect
ime za ‘name for’	[NP] [ime] za [NP]	hyperonymy/hyponymy
koristiti ‘use’	[NP] [koristiti] [NPAcc]	purpose
nastati ‘form’	[NP] [nastati] [NPIns]	result
nastati od ‘made of’	[NP] [nastati] od [NPGen]	meronymy
naziv za ‘term for’	[NP] [biti] naziv za [NP] / [NP] naziv [biti] za [NP]	hyperonymy/hyponymy
nazivati se ‘called’	[NP] [nazivati] se [NP] / [NP] se [nazivati] [NP]	synonymy hyperonymy/hyponymy
naziva se još ‘also called’	[NP] [nazivati] se još (i) [NP] / [NP] se još [nazivati] (i) [NP]	hyperonymy/hyponymy
odnositi se na ‘refer to’	[NP] [odnositi] se na [NP]	hyperonymy/hyponymy

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation
posljedica ‘consequence of’	[NP] ([biti]) [posljedica] [NPGen] / [NP] [posljedica] ([biti]) [NPGen]	result
predstavljati ‘represent’	[NPAcc] [predstavljati] [NP]	hyperonymy/hyponymy
uzrokovati ‘cause’	[NP] [uzrokovati] [NP]	cause-effect

The Croatian data show the use of the relations of *purpose* and *result*, which were not detected in the English examples. However, these lexical-semantic relations could be broadly grouped as general causative relations, so it could certainly not be claimed that they are not present in the English data.

The function of lexical-semantic relations at the term level corresponds to the conceptual relations between concepts at the concept level or the level of knowledge organization. Therefore, at the concept level, the relations *type_of* and *part_of* are hierarchical concept relations used to express whether a subordinate concept is a type of or part of a superordinate concept. In addition to these relations, which make up the largest number of all concept relations in extracted contexts, non-hierarchical or associative relations are used to refer to the cause-effect, purpose, result, origin, attributes, or function of the concept. In the domain of climate change, where many activities and processes are linked by causal and temporal relations, non-hierarchical concept relations can offer more knowledge information than the traditional ontological relations hierarchically organized and presented.

6. Conclusion

The analysis of the definitional contexts of key terms in the domain of climate change began with extracting knowledge-rich contexts from English and Croatian corpora of popular educational resources written for schoolchildren aged eight or nine to approximately sixteen. After the initial analysis and downsizing the data to only those examples that could be used to define specialized concepts in case no definitions are available, a list of examples of definitional contexts was prepared for each language. The lexical markers and lexical knowledge patterns marked in them point to the conclusion that similar markers are used in both languages, with the difference that the knowledge patterns in Croatian have a more fixed structure in terms of valence patterns of certain causal verbs and overall a slightly greater variation in the form of lexical markers.

Causal relations are certainly the dominant type of semantic relations, which is not surprising given the format and register of the texts, as well as the target audience. *Cause-effect*, *purpose*, and *result* relations are to be expected in educational materials, the purpose of which is not merely to identify specialized concepts

(e.g., as a database or a glossary would), but also to explain how and why these concepts occur. What is not covered by this analysis, but could be elaborated in a future study, is the function of terminological variation on the text level. Alongside this, the comprehension of abstract versus concrete concepts is certainly a line of research worth investigating, focusing on different age groups of young users and applying different methods to test the categorization and classification skills of schoolchildren and young adults.

References

- Barrière, Caroline. 2004. "Knowledge-Rich Contexts Discovery." In *Proceedings of the Seventeenth Canadian Conference on Artificial Intelligence*. London, Ontario: CSCSI. 187–201.
- Caramelli, Nicoletta; Setti, Annalisa; Maurizzi, Donatella D. 2004. "Concrete and abstract concepts in school age children." *Psychology of Language and Communication* 8(2). 18-34.
- Casadevante, Cristina; Romero, Miriam; Fernández-Marcos, Tatiana; Hernández, José Manuel. 2019. "Category Learning in Schoolchildren. Its Relation to Age, Academic Marks and Resolution Patterns." *The Spanish Journal of Psychology* 22. 1–13.
- Cognuck González, Sara; Numer, Emilia. 2020. *Climate glossary for young people*. Panama: UNICEF.
- Condamines, Anne. 2022. "How the Notion of "Knowledge Rich Context" Can Be Characterized Today." *Frontiers in Communication* 7, 824711. doi: 10.3389/fcomm.2022.824711
- Fišer, Darja; Pollak, Senja; Vintar, Špela. 2010. "Learning to Mine Definitions from Slovene Structured and Unstructured Knowledge-Rich Resources." In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA).
- Gelman, Susan A.; Meyer, Meredith. 2011. "Child categorization." *Wiley Interdisciplinary Review: Cognitive Science* 2 (2). 95–105.
- Grčić Simeunović, Larisa; Vintar, Špela. 2015. "Modeliranje znanja: korpusna analiza definicijskih stilova." In Bratanić, M., Brač, I. & Pritchard, B. (Eds.) *Od Šuleka do Schengena: Terminološki, terminografski i prijevodni aspekti jezika struke*. Zagreb: Institut za hrvatski jezik i jezikoslovlje. 251–266.
- Jackson, Tom; Guitian, Cristina. 2020. *What's the issue? Climate change*. Mission Viejo, CA: QED Publishing.
- Jackson, Tom; Guitian, Cristina. 2021. *Klimatske promjene – O čemu je riječ?* Zagreb: Školska knjiga.
- Marinellie, Sally A. 2009. "The content of children's definitions: The oral-written distinction." *Child Language Teaching and Therapy* 25(1). 89–102.
- Marshmann, Elizabeth. 2006. "Lexical Knowledge Patterns for Semi-Automatic

- Extraction of Cause–effect and Association Relations from Medical Texts: A Comparative Study of English and French.” PhD thesis, University of Montreal.
- Meyer, Ingrid. 2001. “Extracting knowledge-rich contexts for terminography: a conceptual and methodological framework.” In Bourigault, D., Jacquemin, C. & L’Homme, M-C. (Eds.) *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins. 279–302.
- Murray, Craig G.; Reuter, Kara. 2005. “Children’s acquisition of categories and the implications for research in the development of classification schemes.” *ASIST* 1–13.
- Nilsson, Henrik. 2015. “Enumerations count: Extensional and partitive definition.” In Kockaert, H. J. & Steurs, F. (Eds.) *Handbook of Terminology. Volume 1*. Amsterdam: John Benjamins. 82–100.
- Ramos, Margarida; Rute Costa, Rute. 2023. “Extracting knowledge-rich information from definitions. A corpus-based approach to building a conceptual-based terminological resource.” In Di Nunzio, G., Costa, R. & Vezzani, F. (Eds.) *Proceedings of the 2nd International Conference on Multilingual Digital Terminology Today (MDTT 2023)*. RWTH Aachen University. 1–14.
- Reggin, Lorraine D.; Muraki, Emiko J.; Pexman, Penny M. 2021. “Development of Abstract Word Knowledge.” *Frontiers in Psychology* 12.
- Sierra, Gerardo E.; Alarcón, Rodrigo; Aguilar, Cesar Antonio; Bach, Carmen. 2008. “Definitional verbal patterns for semantic relation extraction.” *Terminology* 14. 74–98.
- Vigliocco, Gabriella; Ponari, Marta; Norbury, Courtenay. 2018. “Learning and Processing Abstract Words and Concepts: Insights From Typical and Atypical Development.” *Top Cogn Sci* 10(3). 533–549.

Appendix 1

Table 3. Lexical markers and lexical knowledge patterns in English definitional contexts

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation	Definitional context
also known as	[NP] [be] also known as [NP]	synonymy hyperonymy/ hyponymy	Fossil fuels are also known as «nonrenewable» fuels because the amount of time required to transform living tissue into coal, oil, or natural gas is tens of millions of years at a minimum. TROPICAL RAINFOREST Also known as jungles, these forests grow wherever it is warm and wet all year round. Carbon monoxide, sulfur dioxide, ozone, methane, fluorocarbons and black carbon are also known as shortlived climate pollutants.
attributed to	[NP] [be] attributed to [NP]	association	Climate change is attributed to human activities that may alter the composition of the atmosphere.
be is (a/the), are (a/the)	[NP] [be] ([article]) [NP] (of)	hyperonymy/ hyponymy	A carbon footprint is the measure of the amount of greenhouse gases that are produced directly or indirectly from your activities. Fossil fuels are energy sources that are generated when plant and animal matter biodegrades. Greenhouse gases are the gaseous component of the atmosphere.
caused by	[NP] [be] caused by [NP] [NP] [can/may] [be] caused by [NP]	cause-effect	Industrial pollution is caused by factories and industry, such as brick kilns, manufacturing companies or power generators, releasing pollutants into the air. Climate change may be caused by natural internal processes or by external forces, such as volcanic eruptions or persistent anthropogenic actions.
called, can be called	[NP] ([can]) [be] called [NP]	hyperonymy/ hyponymy	Energy from sources like wind, water, and the sun is called renewable energy because there is an endless supply of it. Every country is responsible for how much greenhouse gas it releases, which can be called its “carbon footprint.”
due to	[NP] [be] due to [NP]	cause-effect	Current climate change is due to global warming, which is caused by the increase in GHG emissions as a result of human activities.

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation	Definitional context
identified as/by, can be identified as/by	[NP] ([can]) [be] identified as/by [NP]	association	Climate change is identified by variability in climate properties that persists for a prolonged period
made from	[NP] [be] made from ([article]) [NP]	result	Fossil fuels are made from the remains of living things which have been squeezed and heated underground for millions of years.
mean	[NP] means [NP]/ [VP]	cause–effect	A bigger carbon footprint means more emissions of carbon dioxide and methane, and therefore a bigger contribution to the climate crisis.
refer to (as)	[NP] refers to [NP] (as)	hyperonymy/ hyponymy	Climate change refers to the dramatic warming of the planet caused by increased levels of carbon dioxide in the atmosphere.
should be seen as	[NP] should be seen as [NP]	hyperonymy/ hyponymy	Climate governance should be seen as a “multi-level” process that includes the 12 following levels:
such as	[NP] ([punctuation mark]) such as [NP]	exemplification	It is produced from the burning of fossil fuels (such as coal and diesel) and the smelting of mineral ores that contain sulfur. One symptom of climate change is the increased number of extreme weather events, such as intense hurricanes, extended and wildfires.
that is	[NP] [punctuation mark] that is ([comma]) [NP]	synonymy	When we burn fossil fuels –that is, long-dead plants–inside our car engines Many scientists were eager to learn whether the total amount of these gases in the atmosphere (particularly carbon dioxide, that is CO ₂) was lower in 2020 compared to 2019.

Appendix 2

Table 4. Lexical markers and lexical knowledge patterns in Croatian definitional contexts

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation	Definitional context
biti 'be'	[NP] [biti] [NP]	hyperonymy/hyponymy	Učinak staklenika je proces kojim zračenje Zemljine atmosfere zagrijava površinu na višu temperaturu nego što bi ona bila kada ne bi postojala atmosfera.
be'			"The greenhouse effect is the process by which radiation from the Earth's atmosphere heats the surface to a higher temperature than it would be if the atmosphere were not present."
činiti	[NP] čine [NP]	meronymy	Staklenik čine staklenički plinovi ugljik dioksid i metan.
'is composed of'			"The greenhouse is composed of greenhouse gases such as carbon dioxide and methane."
dolazi do	do [NPGen] dolazi zbog [NPGen]	cause-effect	Do klimatskih promjena dolazi zbog povišenja temperature Zemlje (globalnog zatopljenja) izazvanog dodavanjem neprirodno velike količine stakleničkih plinova u atmosferu.
'occur'			"Climate change occurs due to the Earth's rising temperature (global warming) caused by the addition of an unnaturally large amount of greenhouse gases into the atmosphere."
	do [NPGen] dolazi jer [NPNom]		Do efekta staklenika dolazi jer atmosfera sadrži plinove kao što su vodena para, ugljikov dioksid, metan i dušikov oksid.
			"The greenhouse effect occurs because the atmosphere contains gases such as water vapor, carbon dioxide, methane, and nitrous oxide."
dovesti do	[NP] [dovesti] do [NPGen]	cause-effect	Klimatske promjene dovode do većeg broja vremenskih ekstrema.
'result in'			"Climate change leads to an increased number of weather extremes."
			Klimatske promjene dovest će do povećanja broja toplijih dana i smanjenja broja hladnih dana.
			"Climate change will result in an increase in the number of warmer days and a decrease in the number of colder days."

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation	Definitional context
ime za	[NP] [ime] za [NP]	hyperonymy/hyponymy	Uragani, tajfuni i cikloni različita su imena za nasilne oluje koje nastaju iznad toplih tijela vode, kao što su Tihi ocean ili Karipsko more, kada postoji mnogo toplog, vlažnog zraka u atmosferi.
'name for'			"Hurricanes, typhoons, and cyclones are different names for violent storms that form over warm bodies of water, such as the Pacific Ocean or the Caribbean Sea, when there is a lot of warm, moist air in the atmosphere."
koristiti	[NP] [koristiti] [NPAk]	purpose	Klimatski modeli koriste matematičke jednadžbe kako bi opisali ponašanje elemenata Zemljina sustava koji utječu na klimu.
'use'			"Climate models use mathematical equations to describe the behavior of Earth's system elements that influence the climate."
nastati	[NP] [nastati] [NPIns]	result	Ti dodatni staklenički plinovi uglavnom nastaju spaljivanjem fosilnih goriva radi proizvodnje energije.
'produce'			"Those additional greenhouse gases are mainly produced by burning fossil fuels for energy production."
nastati od	[NP] [nastati] od [NPGen]	meronymy	Fosilna goriva nastala su od ostataka biljaka i životinja koji su živjeli prije milijun godina.
'formed from'			"Fossil fuels are formed from the remains of plants and animals that lived millions of years ago."
naziv za	[NP] [biti] naziv za [NP] /	hyperonymy/hyponymy	Globalno zagrijavanje je naziv za povećanje prosječne temperature zemljine atmosfere i oceana.
'term for'	[NP] naziv [biti] za [NP]		"Global warming is the term for the increase in the average temperature of the Earth's atmosphere and oceans."
nazivati se	[NP] [nazivati] se [NP] /	synonymy or hyperonymy /hyponymy	To se naziva "efektom staklenika" jer atmosfera funkcionira poput stakla u stakleniku – grije unutrašnjost.
'called'	[NP] se [nazivati] [NP]		"It is called the 'greenhouse effect' because the atmosphere functions like glass in a greenhouse, warming the interior."

Lexical marker	Lexical knowledge pattern	Lexical-semantic relation	Definitional context
naziva se još, nazivaju se još	[NP] [nazivati] se još (i) [NP] /	hyperonymy /hyponymy	Ugljen, nafta i prirodni plin nazivaju se još i fosilna goriva, a trenutno predstavljaju glavni izvor energije u svijetu.
'also called'	[NP] se još [nazivati] (i) [NP]		"Coal, oil, and natural gas are also called fossil fuels and currently represent the primary source of energy in the world."
odnositi se na	[NP] [odnositi] se na [NP]	hyperonymy /hyponymy	Klimatske promjene odnose se na mnoge različite učinke globalnog zagrijavanja na klimatski sustav Zemlje.
'refer to'			"Climate change refer to many different effects of global warming on the Earth's climate system."
posljedica	[NP] ([biti]) [posljedica] [NPGen] /	result	Ono je posljedica emisije ugljikovog dioksida i metana, tzv. stakleničkih plinova, u atmosferu većinom iz industrijskih postrojenja.
'consequence of'	[NP] [posljedica] ([biti]) [NPGen]		"It is a consequence of the emission of carbon dioxide and methane, known as greenhouse gases, into the atmosphere, mainly from industrial facilities."
predstavljati	[NPAk] [predstavljati] [NP]	hyperonymy/ hyponymy	Klimu predstavljaju prosječni vremenski uvjeti na nekom mjestu tijekom relativno dugih vremenskih razdoblja (npr. 30 godina).
'represent'			"Climate represents the average weather conditions in a location over relatively long periods (e.g., 30 years). The greenhouse is composed of greenhouse gases, carbon dioxide, and methane."
uzrokovati	[NP] [uzrokovati] [NP]	cause-effect	Utvrđeno je kako klimatske promjene već sada uzrokuju veće zagrijavanje.
'cause'			"It has been determined that climate change is already causing greater warming."