SnowTerm: a terminology database on snow and ice

Paolo Plini
CNR, Institute for Polar Sciences, Italy
Sabina Di Franco
CNR, Institute for Polar Sciences, Italy
Rosamaria Salvatori
CNR, Institute for Polar Sciences, Italy

Abstract

SnowTerm is the result of ongoing work on a structured reference multilingual scientific and technical vocabulary covering the terminology of a specific knowledge domain like the polar and the mountain environment. The terminological system contains around 2.500 terms and it is arranged according to the EARTh thesaurus semantic model. It is foreseen as an updated and expanded version of this system.

1.0 Introduction

The state of the environment and climate change have become crucial and urgent for the scientific, productive and political worlds. It is necessary to combine awareness of the phenomena, correct knowledge of the mechanisms that cause them, and the possibility of effectively mitigating the impacts of human activities.

In all these aspects, correct information is fundamental. Unorganised and uncontrolled data and information can lead to an 'information deluge' with more negative than positive effects.

The organisation of knowledge, starting with a language that is as shared and unambiguous as possible, is one of the central pillars for incisive action to facilitate access to validated and reliable information. Starting with words, concepts and their relationships help to better understand and organise scientific knowledge and reduce misunderstandings. The increasing complexity of scientific and technical subjects forms a network with numerous branches and connections, with frequent overlapping of knowledge domains, which, while being an invaluable asset, complicate the whole system.

Terminological tools (thesauri, glossaries, ontologies) are useful for improving the understanding of domain languages and the organisation of the associated knowledge. Thesauri, in particular, make it possible to share and

agree on scientific/technical terms in the target domain and to express them in several languages.

Moreover the use, management and diffusion of information is changing very quickly in the environmental domain, due also to the increased use of the Internet, which has resulted in people having at their disposal a large sphere of information and has subsequently increased the need for multilingualism.

To exploit the interchange of data, it is necessary to overcome problems of interoperability that exist at both the semantic and technological level and by improving our understanding of the semantics of the data. This can be achieved only by using a controlled and shared language.

The polar environment and related polar sciences are no exception to these communication needs.

The polar regions have unique characteristics, such as the phenomenon of polar amplification (Stuecker et al. 2018), which also make them unique in terms of describing their complexity. After research on the internet, several glossaries related to polar and mountain environment were found, written mainly in English. Typically these glossaries – with a few exceptions – are not structured and are presented as flat lists containing one or more definitions i.e. NSIDC Cryosphere Glossary,[1] IPCC "Special report: special report on the ocean and cryosphere in a changing climate Glossary" (Intergovernmental Panel on Climate Change 2019), or the GCW Cryosphere Glossary.[2]

The occurrence of multiple definitions might contribute to increasing the semantic ambiguity, leaving up to the user the decision about the preferred meaning of a term. On the contrary, providing a structure to the lexicon so that each term is placed within a semantic network allows specifying its meaning.

The preliminary results of this work of selection and classification of terms on polar and mountain environment are presented here, as a proposal of controlled and structured language with the goal to develop a prototype of a thesaurus on this specific sector.

The thematic areas, covered at present, deal with snow and ice physics, snow and ice morphology, snow and ice radiometry, remote sensing and GIS in cryosphere environment, sea ice, avalanches, glaciers, disaster management and risk prevention.

1 National Snow and Ice Data Center (NSIDC), "Cryosphere Glossary," last accessed October 14, 2021, https://nsidc.org/cryosphere/glossary.

2 World Meteorological Organization (WMO), Global Criosphere Watch (GCW), "Cryosphere Glossary," last accessed October, 14, 2021, https://globalcryospherewatch.org/reference/glossary.php.

2.0 Identification of terminological sources and selection of terms

The first sources used to collect the terminology consist of the "Glossario dei termini usati nei bollettini nivometeorologici",[3] the "Sea Ice Glossary",[4] the "Glossary of Selected Glacier and Related Terminology",[5] the "Sea Ice Nomenclature",[6] the trilingual "Glossary on snow and avalanches",[7] the "Večjezični Slovar – Sneg in plazovi" developed by Pavle Šegula (1995).

The terminology of these sources was analysed with respect to the degree of semantic relevance in the field. Terms too generic or considered as non-pertinent were excluded. Groups of terms that could be collected in specific appendices were also excluded.

At present the database contains 3,700 records; the identification of a certain number of non-descriptors have been performed, the final selection of terms is still ongoing.

3.0 Management of terms

It occurs quite often to find elements belonging to a parent concept which are expressed with terms like "small", "medium", etc. In such cases, we decided to modify the original string adding all the information that will make each term meaningful. For example, the "wind intensity" is declared as "weak", "tempered", "strong", etc. Such terms, if used out of context, are impossible to understand. Having modified these terms in "weak wind", "tempered wind", "strong wind" will allow any user to use the terms in any external application without losing information.

3 Associazione Interregionale di coordinamento e documentazione per i problemi inerenti alla neve e alle valanghe (AINEVA) and Friuli-Venezia Giulia Region, "Glossario dei termini usati nei bollettini nivometeorologici," last accessed October 14, 2021, http://www.aineva.it/bolletti/bollet5.html.

4 Scientific Committee on Antarctic Research (SCAR), "Sea Ice Glossary".

5 United States Geological Survey (USGS). 2007. "Glossary of Selected Glacier and Related Terminology," last accesed October 14, 2021, http://vulcan.wr.usgs.gov/Glossary/Glaciers/glacier_terminology.html.

6 "Sea ice nomenclature: English-Finnish-Swedish-Estonian-Russian". *Merenkulkulaitoksen julkaisuja* 5, Finnish Maritime Administration 2002.

7 Swiss Federal Institute for Snow and Avalanche Research, Working Group on Avalanches Warning Services. 2004, "Glossary on snow and avalanches," last accessed October 14, 2021, https://www.avalanches.org/glossary/?lang=it.

4.0 Classification of terms

The classification and relational structure are based on the EARTh (Environmental Applications Reference Thesaurus) semantic model (Mazzocchi and Plini 2005).

The terms are arranged according to a classification scheme that is founded on categories. At the first level, the system is structured into categories defined as "ENTITIES", "ATTRIBUTES", "DYNAMIC ASPECTS" and "DIMENSIONS". The "ENTITIES" describe material and immaterial objects; the "ATTRIBUTES" define the nature of the objects, at least as far as their static aspects are concerned; "DYNAMIC ASPECTS" define the activities, the processes and the conditions in which they are involved; the "DIMENSIONS" identify the spatio-temporal circumstances in which all this occurs.

The system is then organized in a framework of different levels and classification knots, and it comprises hierarchical relations. It continues into further levels as they obtain a greater specificity in order to allow a rational arrangement of objects.

At present around 1,100 terms have been put into the hierarchical structure.

The vertical structure can be used as a semantic reference system, stable and partially independent from the context.

The model envisages the possibility of complementing the faceted structure with a system of themes which by crossing with the vertical structure would form a matrix system.

In a thematic approach, the terms linked to a specific sector are reassembled, while the facet structure tends to scatter them under the more general referral concept.

Moreover, the system of themes, as it was conceived, should be developed by a user according to the specific needs of the applicative context.

One example of thematic setup is provided by the classification into sectors contained in the "Sea Ice Nomenclature" where the terms are clustered according to "ice development", "sky and air indications", "ice arrangement", "terms relating to surface shipping", "terms relating to submarine navigation".

5.0 Software details

All the terms are handled using the TemaTres software (Gonzales Aguilar et al. 2012). It is a tool for visual representation and management of controlled vocabularies. In the frame of an international cooperation some support has been provided to improve the functionalities of the software in particular

dealing with the relationships between terms. The web interface allows access to the system through the internet (https://vocabularyserver.com/cnr/ml/snowterm/en/index.php).

6.0 Multilingualism

Multilingualism is not the main interest of our working group. Nevertheless, in order not to waste important resources, the already available translations have been collected. The system now contains 2,700 English terms, the other languages are Italian (2,400), Estonian, Finnish, Russian and Swedish (94), French and German (1,900), Slovenian (1,300) and Spanish (1,800). The enlargement of the number of linguistic equivalents in French and German is mandatory due to the geographical and political position of the alpine area. Other languages will be updated following a direct interest and willingness to cooperate by other institutions.

7.0 Results, their use and future development

The results of this work is the production of a monolingual terminological system organized both in a vertical way -according to a classification system based on categories- and horizontally on the basis of the systems of themes.

SnowTerm could be considered as one of the first attempts to develop a thesaurus on the Polar and Mountain Environment domain.

In order to ensure a better and updated conceptual and terminological coverage, an extension and revision of the system are foreseen. Any other reliable glossary or term list will be considered as potential additional sources.

The semantic structure of the system will also be strengthened. In order to increase the efficiency of the system in information retrieval operations, a set of associative and equivalence relations will be implemented.

The organization of knowledge -through the support of a thesaurus- could bring a strong contribution to the management of the information in the specific domain: by suggesting a language that different institutions could share; ensuring higher semantic transparency to terminology; providing tools for indexing and retrieving the information and to interchange data and suggesting semantic maps usable for the conceptual description of the domain.

References

Associazione Interregionale di coordinamento e documentazione per i problemi inerenti alla neve e alle valanghe (AINEVA) and Friuli-Venezia Giulia Region, "Glossario dei termini usati nei bollettini nivometeorologici." Last accessed October 14, 2021. http://www.aineva.it/bolletti/bollet5.html.

Gonzales Aguilar Audilio, Maria Ramirez-Posada, and Diego Ferreyra. 2012, "TemaTres: software to implement thesauri". *El Profesional de la Informacion* 21: 319-25.

Intergovernmental Panel on Climate Change (IPCC). 2019. "Annex I: Glossary". In IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, edited by N. M. Weyer. https://www.ipcc.ch/srocc/chapter/glossary/.

Mazzocchi, Fulvio, and Paolo Plini. 2005, "Thesaurus classification and relational structure: the EARTh experience". In *7th International conference on Terminology and Knowledge Engineering "Terminology & Content Development"* TKE 2005 16-19 August 2005, Copenhagen, edited by B. Nistrup Madsen, H. Erdman Thomsen, 265-78.

National Snow and Ice Data Center (NSIDC), "Cryosphere Glossary." Last accessed October 14, 2021. https://nsidc.org/cryosphere/glossary.

Scientific Committee on Antarctic Research (SCAR), "Sea Ice Glossary".

Šegula Pavle. 1995. "Večjezični Slovar Sneg in Plazovi. Gorska resevalna sluzba pri Planinski zvezi Slovenije", ISBN: 9616156004.

Stuecker, Malte F., Cecilia M. Bitz, Kyle C. Armour, Cristian Proistosescu, Sarah M. Kang, Shang-Ping Xie, Doyeon Kim, Shayne McGregor, Wenjun Zhang, Sen Zhao, et al. 2018. "Polar amplification dominated by local forcing and feedbacks". *Nature Climate Change* 8, 1076-81, https://biblioproxy.cnr.it:2481/10.1038/s41558-018-0339-y.

Swiss Federal Institute for Snow and Avalanche Research, Working Group on Avalanches Warning Services. 2004, "Glossary on snow and avalanches." Last accessed October 14, 2021. https://www.avalanches.org/glossary/?lang=it.

United States Geological Survey (USGS). 2007. "Glossary of Selected Glacier and Related Terminology." Last accesed October 14, 2021. http://vulcan.wr.usgs.gov/Glossary/Glaciers/glacier_terminology.html.

World Meteorological Organization (WMO), Global Criosphere Watch (GCW), "Cryosphere Glossary." Last accessed October, 14, 2021. https://globalcryosphere watch.org/reference/glossary.php.